DIAZ VILLEGAS

TECNICAS ESTADISTICAS PARA INVESTIGACIONES SOCIALES

INSTITUTO DE INVESTIGACIONES SOCIALES DE LA
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

TECNICAS ESTADISTICAS

para investigadores sociales

por

OSCAR URIBE VILLEGAS

MÉXICO, D. F.

TÉCNICAS ESTADÍSTICAS PARA INVESTIGADORES SOCIALES

INSTITUTO DE INVESTIGACIONES SOCIALES DE LA
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

TECNICAS ESTADISTICAS
para investigadores sociales

por

OSCAR URIBE VILLEGAS

MÉXICO, D. F.

Impreso y hecho en México
Printed and made in Mexico

Al Dr. D. Lucio Mendieta y Núñez,
a quien tantas cosas mías, grandes y pequeñas, se deben.

Al Dr. D. José Gómez Robleda,
mi iniciador en el campo de la Estadística.

1.—INTRODUCCIÓN

LA INVESTIGACIÓN ESTADÍSTICA EN EL MARCO DE LA INVESTIGACIÓN SOCIAL

Definición de la Investigación Social

La investigación social es un proceso, mediante el cual el investigador trata de encontrar:

los elementos
- *1º*—determinantes ("causas" o "precipitantes")
- *2º*—concurrentes ("con causas")
- *3º*—influyentes, y
- *4º*—presentes

que intervienen en:

a.—un problema, o
b.—una situación social,

con el objeto de:

a.—resolver el problema, o
b.—mejorar la situación,

de acuerdo con:

1.—valoraciones, y
2.—finalidades claramente definidas.

Etapas de la Investigación Social

La investigación social, en cuanto proceso, está constituída por una serie ordenada de etapas entre las que, como principales, pueden distinguirse las siguientes:

I.—Planeación de la Investigación,
II.—Realización de la Investigación,
III.—Sugestión de soluciones y medios de mejoramiento.

La planeación de la pesquisa o investigación implica problemas metodológicos y tecnológicos.

La realización de la investigación implica problemas prácticos de destreza y experiencia en el manejo de ciertas técnicas.

La sugestión de soluciones y medios de mejoramiento implica problemas políticos o político-sociales de elección de fines y de medios relacionados con tales fines y, en última instancia, problemas de ideología.

Concepto de la Planeación

El concepto de la planeación de una investigación comprende, por lo menos, los tres elementos siguientes:

1º—un elemento de anticipación o previsión,
2º—un elemento volitivo y de decisión,
3º—un elemento representativo de captación.

Elemento anticipatori oy de previsión, en cuanto la planeación trata de prever las situaciones a las que tendrá que enfrentarse para poder controlarlas.

Elemento volitivo y de decisión por cuanto la planeación trata de decidir qué será lo mejor para evitar la dificultades y obtener el mayor éxito.

Elementos anticipatorio y de previsión, en cuanto la planeación trata de fijar sus previsiones y decisiones en forma coordinada, en un plan de acción consignado en forma escrita, diagramática o de otro tipo, para recordar cuál es el proceso que se ha decidido seguir, y para poder valorar el plan en su totalidad, o criticarlo.

Conforme a lo anterior, podemos decir que: planear una investigación equivale a anticipar situaciones y tomar decisiones que, consignadas por escrito o en diagramas, permitan recordar los pasos del proceso, valorarlo en su totalidad y controlar las situaciones a las que enfrente la investigación.

Una investigación está planeada metodológicamente, en cuanto:

A.—se le ha estudiado y valorado de antemano en su totalidad y en sus detalles, y
B.—se puede juzgar del procedimiento empleado en la obtención de ciertos resultados.

Primacía de la Planeación en la Pesquisa

En cuanto tratamos de la primacía de la planeación en la pesquisa o investigación social, en realidad lo que hacemos es plantearnos las preguntas de si ¿debe preceder la observación a la planeación? o ¿debe preceder la planeación a la observación? El problema se resuelve en cuanto se consideran los pasos siguientes:

1º—La observación más o menos burda hace presentir al investigador que hay un problema.
2º—La planeación permite que el investigador oriente su observación, determine si el problema realmente existe, y cuáles son los términos en que se plantea.
3º—La observación orientada y afinada (incluso complementada de la experimentación) da al investigador el conocimiento del problema.

La Planeación como Orientadora de la Observación

La observación, científicamente controlada, no puede existir sino a partir del momento en que decidimos:

$$\left.\begin{array}{l}\text{qué,}\\ \text{cuándo,}\\ \text{dónde,}\\ \text{cómo y}\\ \text{por qué}\end{array}\right\}\ \text{vamos a observar algo.}$$

La determinación de qué es lo que habrá necesidad de estudiar, en cuanto se trata de un problema social, deberá hacerla la sociología, la cual deberá:

1.—Señalar la forma en que pueda aislarse en forma *natural*, y *no arbitrariamente*, el problema social o la unidad social particular de que se trate, del complejo social del que forma parte.
2.—Indicar cuáles son los aspectos, *elementales* y *relacionales* que deberán estudiarse.

El cuándo y el dónde, o sean las determinaciones de lugar y tiempo, de delimitación espacial y temporal dependerán tanto de lo anterior como de:

1.—Los propósitos de la investigación,
2.—La precisión que se busque en los resultados,
3.—Los medios económicos y el tiempo de que se disponga.

Cómo hacer el estudio de un problema o situación sociales ya delimitados espacial y temporalmente es, sobre todo, el problema metódico de elección de métodos de:

A.—Inducción-deducción,

B.—Observación-experimentación,
C.—Tipificación.

Asimismo, se trata de determinar la utilidad que en el caso concreto puedan brindar los métodos:

a.—histórico,
b.—geográfico,
c.—antropológico,
d.—etnográfico,
e.—económico,
f.—psicológico,
g.—estadístico,
h.—literario, etc.

como auxiliares de la metodología sociológica estricta.

Por otra parte, importa determinar cuáles de entre las siguientes (que se listan a modo de ejemplo) habrán de ser las técnicas empleadas:

Cuestionario

Encuesta { Directa, Por Radio. Por Correo,

Entrevista { Libre, Controlada, Grupal.

Escalar sociométrica { de medición de opiniones, de medición de actitudes.

Sociogramática,
Del Observador Participante, etc.

Del por qué vamos a observar dependerá a su vez el valor que demos a ciertos aspectos y no a otras facetas del problema en relación con los propósitos de mejoramiento o solución de situaciones. La aclaración de los por qués es el único medio al travès del cuál un observador o un investigador distinto del que realizó la pesquisa puede juzgar de sus resultados y determinar si los mismos pueden colocarse en un plano de comparación con los propios o con los datos obtenidos por otro investigador; en cualquier otro caso queda la duda de si, por ejemplo, el cambio que parece poder afirmarse con respecto a una comunidad mediante la comparación de los resultados obtenidos por dos investigadores distintos deberá imputarse a una verdadera modificación de las condiciones sociales de la comunidad observada, o a que la misma comunidad ha sido estudiada por dos investigadores distintos que han hecho una selección distinta o han ponderado diversamente la probabilidad selectiva de ciertos aspectos de la realidad estudiada.

Dos Ventajas de la Planeación

La planeación permite, por una parte, determinar el grado de precisión que se requiere en la investigación, evitando con ello tanto una falta como un exceso (inútil y en ocasiones absurdo) de precisión; por otra parte, la planeación permite determinar el grado de adecuación de las diferentes técnicas de que dispone el investigador con respecto al objeto de estudio.

La Investigación

En toda investigación social es conveniente distinguir:

I.—Los elementos personales de la investigación,
II.—Las operaciones de la investigación, y
III.—Las fases de la investigación.

Elementos Personales de la Investigación

El investigador social necesita recordar siempre que no sólo es social la investigación que realiza por tener carácter social el objeto observado, sino que la investigación social misma es, de por sí, un problema social, un conjunto de situaciones en las cuales intervienen diferentes elementos personales, ligados entre sí por relaciones interhumanas; olvidarlo conduce a fracasos lamentables para la investigación social.

Los elementos personales de la investigación social son:

1.—El consumidor de la investigación social,
2.—El científico social que la planea,
3.—El observador (encuestador, censador, etc.) que la realiza, y
4.—El observado o la persona observada (si se toma "persona" en su sentido más lato de persona física o de persona moral) que deberá beneficiarse de ella.

Operaciones de la Investigación

Desde el punto de vista sociológico, puede establecerse una cierta equivalencia entre el rubro "operaciones de la investigación" y el de relaciones inter-humanas dentro de la investigación social.

Desde el punto de vista práctico, puede ponerse en función del rubro operaciones de la investigación y de la distinción clara de las mismas, el éxito o fracaso de la pesquisa ya que, del adecuado análisis de los diversos pasos de una investigación dependerá el que, en determinados casos, se encuentre o no el punto en el que se erró y la operación que hizo fracasara la pesquisa en su conjunto, con lo cual, y en caso de ser posible, en vez de invalidarse la investigación en su totalidad precisará rehacer sólo una parte de la misma.

Las operaciones principales de una investigación social son:

1º—Transmisión del problema que el consumidor hace al científico social,

2º—Entrenamiento y vigilancia de los encuestadores, que realiza el científico social,

3º—Estímulo de lo observado, que realizan los observadores (o experimentadores),

4º—Respuesta de lo observado al estímulo del observador,

5º—Registro y transmisión de respuestas, del observador al científico social,

6º—Informe acerca del problema y sugestiones para resolverlo que el científico social brinda al consumidor.

Tras el informe, será el consumidor quien, conocido el problema en sus términos precisos, y las soluciones alternativas propuestas por el científico social, habrá de seleccionar un curso de acción hacia una solución de mejoramiento. Aun cuando constituye tema que no nos corresponde abordar en este punto, podemos señalar que, a partir de este momento, y gracias al desarrollo de una moderna rama estadística, el consumidor para aplicar la política más conveniente de mejoramiento se ve precisado a recurrir nuevamente al científico social que mediante la hechura estadística de decisiones puede indicarle cual de los propuestos es el mejor curso de acción.

Fases de la Investigación

Como fases de la investigación pueden listarse:

I.—Reconocimiento de que el problema existe y de que no se trata de un problema ficticio o de una construcción artificial.

II.—Formulación del problema.

III.—Planeación Metódica para la investigación:

i.—Planeación ideal de la investigación o sea, instrumentación de la misma, determinación de la exactitud deseable, de la extensión, etc., que la misma debería tener de no enfrentarse a obstáculos prácticos en la realidad.

ii.—Planeación realista o práctica de la investigación, mediante la cual se hagan ajustes al plan ideal conforme lo requiera la realidad, pero *sin hacer que ésta prive* en más de lo indispensable sobre el plan ideal; de ahí la necesidad de distinguir en este punto entre: obstáculos difíciles de salvar, y dificultades insalvables o verdaderas imposibilidades (la frontera fluctuante entre una y otras dependerá en buena parte, del entrenamiento y experiencia previos del investigador, y en otra muy buena proporción de su brío y tenacidad —que no deben confundirse con ceguera y obstinación— para llevar adelante un proyecto).

iii.—Diseño Estadístico, o sea la especificación del grupo de objetos o personas por estudiar,
del número de los que se seleccionarán,
del criterio con que se elegirán,
de los rasgos que se han de estudiar
(numérica o cualitativamente).

iv.—Plan de Observación, mediante el cual

a.—se determinará en qué condiciones se harán las observaciones, y

b.—qué precauciones deberán tenerse al hacerlas.

v.—Plan de Operaciones, al cual corresponde la instrumentalización y tecnificación de la pesquisa misma.

Conforme hemos dicho, todo diseño estadístico debe iniciarse con la especificación del grupo de objetos o personas que han de ser estudiados, de los rasgos de esos objetos que van a fijarse numéricamente o mediante designación cualitativa, y con la delimitación espacial y temporal de la investigación; así, por ejemplo, si se trata de una investigación sobre accidentes de trabajo en la industria del cemento, será necesario:

1º—Especificar qué grupo va a incluirse, determinando:

a.—si ha de comprender a todos los accidentados o sólo a los que rebasen determinada edad, *b.*—si comprenderá todos los grados de incapacidad o sólo los que sean superiores o determinado porciento,

c.—si se incluirán todas las categorías de trabajadores o sólo algunas de ellas...

2º—Determinar el rasgo por estudiar o sea, en este caso, definir qué se entiende por accidente de trabajo, respondiendo a preguntas como las siguientes:

a.—¿para el cómputo deberá considerarse como unidad el hecho accidental perjuicioso, o el número de individuos afectados?

b.—¿se considerará como tal el accidente ocurrido durante la realización del trabajo, o el ocurrido como consecuencia de esa misma realización, o el ocurrido cuando el accidentado se disponía a realizarlo, o todos ellos?

3º—Delimitar espacial y temporalmente la investigación, resolviendo si:

a.—¿se van a incluir todas las fábricas de cemento de la República?

b.—¿se incluirán sólo las de las zonas industriales aledañas a la ciudad de México?

c.—¿se extenderá la investigación a un período más o menos amplio, o se reducirá a uno corto?

Como es fácil comprender, estas especificaciones y delimitaciones dependen de muchos factores; en el último apartado, sobre todo, de limitaciones de orden económico, de premura de tiempo para realizar la investigación, de carencia de datos para épocas anteriores o de inadecuación de los mismos para los propósitos de la investigación. Sin embargo, algunas de estas limitaciones pueden salvarse; así, por ejemplo, la falta de medios económicos para realizar una investigación muy extensa puede salvarse eligiendo una porción de la población que sea suficientemente representativa de la misma y que podrá ser menor conforme la población sea más homogénea; dependiendo por otra parte el tamaño de dicha "muestra" de la "población" o "universo" en estudio de las mismas limitaciones presupuestales así como del grado de exactitud que se busque en los resultados.

OPERACIONES ELEMENTALES DE LA INVESTIGACIÓN ESTADÍSTICA

La amplitud, complejidad y desarrollo alcanzado por la metodología estadística hace olvidar a menudo al investigador que esa complejidad a la que se enfrenta es el resultado de la combinación de un número relativamente corto de operaciones elementales, cada una de las cuales responde a una necesidad específica de la investigación ya sea social o de otro tipo. De otra parte, ese olvido de la relación existente entre los métodos, las operaciones elementales que los constituyen y las necesidades que tratan de satisfacerse al través suyo, impiden a quien se especializa en la metodología estadística orientada hacia el terreno de la investigación social el que, en cuanto se plantea una necesidad nueva de ésta, pueda encontrar los medios estadísticos de satisfacerla, con lo cual su actividad se estanca y vuelve infecunda o, en el mejor de los casos, se desarrolla, pero dentro de ciertos cauces lúdicos que pueden reportar beneficio a la investigación social, pero sólo a largo plazo y no en un futuro inmediato.

De ahí que sea importante, dentro de una labor analítica, señalar cuáles son las diferentes operaciones elementales de la investigación estadística y, en cuanto es posible, indicar asimismo las necesidades que satisfacen así como sus limitaciones.

Algunas de esas operaciones, se han mencionado en un contexto más amplio en el apartado anterior, no obstante lo cual, precisa volver a consignarlas en este nivel más específico.

Dentro de las mismas operaciones elementales de la investigación estadística existe una jerarquización que precisa no olvidar, y aun cuando resulte más breve la mención de algunas de dichas etapas (como puede ocurrir en el caso de la "definición", o de la "recolección de los datos") esto no significa que tengan importancia menor que aquellas otras cuya mención es obligadamente más extensa (como ocurre con "elaboración estadística" en sentido estricto). En este sentido, el uso de los números romanos para las operaciones principales, de las mayúsculas, las minúsculas y los arábigos para las subordinadas puede ser útil para llamar la atención hacia el hecho de que tanto monta una cuidadosa elaboración estadística como una precisa definición o una cuidadosa recolección, tanto más si se tiene en cuenta que la idea central de "proceso" como característica de toda investigación social se refleja en forma importantísima en la investigación estadística encuadrada en ella, de tal manera que una deficiente definición de su objeto, en cuanto inicial, puede hacer venir a tierra la investigación en su totalidad no obstante el cuidado que se haya puesto en las siguientes etapas, pudiendo darse el caso de investigaciones cuyos resultados, válidos en apariencia, carezcan totalmente de sentido por haber sido obtenidos de un conjunto de datos deficiente o descuidadamente recogidos, o de un colectivo vagamente definido, heterogéneo, etc.

En este sentido, es tan importante la definición, que incluso podría hablarse de dos grandes etapas de la investigación estadística: la definitoria y la propiamente elaborativa.

Conforme a lo anterior, consideraremos constituída toda investigación estadística por las siguientes etapas y operaciones elementales:

I.—Definición

A.—Del fin específico de la investigación social a cuyo servicio han de ponerse los resultados de la investigación estadística.

B.—del "universo", "población" o "colectivo" por estudiar (o sea, del grupo de personas u objetos que abarcará la investigación estadística).

C.—de los rasgos por estudiar en dicho colectivo.

D.—del grado de homogeneidad deseable o de heterogeneidad permisible en las unidades del universo por lo que respecta a todos aquellos rasgos que sean distintos del o de los rasgos que se deseen estudiar (de los que una de las cosas que precisamente importa estudiar es su heterogeneidad o variabilidad).

E.—de las unidades a emplear en la medida.

II.—Decisión de si conviene hacer:

A.—una investigación exhaustiva, o

B.—un muestreo;

en este último caso, elegir:

a.—un marco de referencia:

preciso,
completo,
adecuado,
actual.

b.—un tipo de muestreo:

simple aleatorio,
estratificado,
por afijación, etc.

c.—una forma de muestreo:
aislado,
repetitivo,
repetitivo periódico, etc.

d.—el tamaño de la muestra,
en vista de requerimientos de:
representatividad y
precisión de los resultados.

III.—Recolección de los datos:

A.—Directa, mediante

1.—cédulas y cuestionarios,

2.—escalas sociométricas.

B.—Indirecta:

1.—de fuentes primarias recolectoras de datos (ejemplo: de la Dirección General de Estadística y de sus Anuarios),

2.—de fuentes secundarias.

IV.—Elaboración estadística.

A.—Operaciones predominantemente críticas

1.—Completamiento de los datos faltantes cuando la misma sea posible en vista de los datos disponibles.

2.—Determinación de las inconsistencias entre los datos de una misma célula o cuestionario y corrección de las mismas de ser posible.

3.—Rechazo de los datos cuya inconsistencia no pueda salvarse.

4.—Rechazo de las cédulas o cuestionarios en los que abunden las inconsistencias.

B.—Operaciones predominantemente mecánicas

1.—Conjuntación de los materiales,

2.—Ordenación (constitución de arreglos o series).

3.—Clasificación:

a.—dicotómica,

b.—múltiple.

4.—Consignación:

a.—textual,

b.—tabular,

c.—gráfica.

C.—Operaciones predominantemente matemáticas.

1.—Condensación

a.—mediante recuento del número de veces que aparece un

dato *(i. e.* mediante determinación de su *frecuencia),*
b.—mediante la constitución de casilleros en los que colocar los datos *(i. e.* mediante la constitución de *clases* definidas por valores que les sirven de límites "inferiores" y "superiores" y entre los que existe un intervalo).

2.—Relievamiento de características por descripción estadística, al través de:
a.—cálculo de promedios centrales y laterales,
b.—medidas de variabilidad,
c.—medidas de asimetría,
d.—medidas de curtosis, aplanamiento o picudez,
e.—ajuste de curvas.

3.—Reducción a términos comparables, mediante:
a.—obtención de relativos,
b.—obtencióndе tantos porciento,
c.—expresión de resultados en términos de cuantilas (cuartilas, sextilas, etc.),
d.—expresión de datos o resultados en unidades de desviaciones promedio.
e.—reducción de medidas *n*-dimensionales a medidas *unidimensionales* por división entre las potencias correspondientes de las desviaciones promedio.

4.—Ponderación o reconocimiento matemático de la importancia relativa que cada dato juega dentro del conjunto, mediante:
a.—multiplicación por un "factor de ponderación" (según ocurre en el cálculo de medias ponderadas, y en el cálculo de números-índice).
b.—elevación a una potencia igual al factor de ponderación (caso que se reduce al anterior al usar logaritmos).

5.—Análisis, o descomposición de un movimiento en varios movimientos componentes, especialmente en el caso de: series temporales, dinámicas o cronológicas, mediante restas y divisiones sucesivas de:
A.—Movimientos permanentes:
Tendencia secular
rectilínea,
parabólica de 2º grado,

parabólica de 3er. grado,
hiperbólica,
logarítmica,
exponencial, etc.

B.—Movimientos oscilatorios:

α.—periódicos
variaciones estacionales,
variaciones cíclicas de alta o de baja frecuencia.

β.—no periódicos
variaciones accidentales, fluctuaciones.

6.—Síntesis
o recomposición del fenómeno previamente analizado para comprobar lo adecuado de dicho análisis, mediante sumas y productos de los elementos componentes:
Cálculo de los valores teóricos de una serie.
Determinación de errores probables y de zonas de estimación.

7.—Inferencia estadística.

a.—mediante la determinación de la probabilidad de que un hecho ocurra (relacionado con la adaptación de curvas, binomiales, de Poisson, hipergeométricas, normales, de Gram-Charlier, de Pearson, etc.).

b.—mediante "extrapolación" en series temporales, o sea, mediante el cálculo de lo que ocurrirá en el futuro de mantenerse cierto ritmo de crecimiento o decrecimiento, y repetirse como en el pasado ciertas variaciones periódicas.

c.—mediante asociación de series y determinación de:

α.—intensidad de la asociación,

β.—sentido de la asociación (¿al crecer una serie crece la otra y al decrecer decrece, o al crecer decrece la otra y al decrecer crece?),

γ.—ecuaciones de regresión que permitan calcular el valor de una serie en un momento determinado si se conoce el valor de la otra o de las otras series asociadas con ella,

δ.—determinación del retraso con el que una serie cronológica se asocia a otra.

d.—mediante extensión de los resultados de la "muestra" al "universo", al través de:

a.—pruebas o docimasias de hipótesis de que los valores de ciertas constantes encontradas para la muestra (estimadores) correspondan a los valores de las constantes correspondientes de la población dentro de ciertos límites de confianza previamente fijados.

RECOLECCIÓN DE DATOS

Entre las diversas técnicas de recolección de datos, se cuentan:

A.—El cuadro y el cuestionario,
B.—Las técnicas de opción sociométrica,
C.—Las escalas sociométricas,
D.—La entrevista.

El Cuadro

El cuadro constituye una herramienta de la que se valen las ciencias sociales con el objeto de hacer comparables las observaciones de diferentes investigadores individuales, o de un mismo investigador que lo aplica en diferentes ocasiones; o sea, que el cuadro garantiza:

1.—La recolección de los mismos datos o rasgos en todos los casos.
2.—El descargo de la memoria del investigador al consignar todos los datos por escrito (evita el "¿No se me habrá olvidado algo?").
3.—Una cierta comparabilidad de las observaciones de diferentes investigadores por:
  a.—separación analítica de diversos aspectos de la realidad estudiada, considerados como elementales,
  b.—especificación previa de las unidades que se han de emplear.
  Lundberg distingue tres clases de cuadros:
  1º—cuadro para el registro de hechos objetivos,
  2º—cuadro para determinación de opiniones o actitudes,
  3º—cuadro para determinar el funcionamiento de las organizaciones.

Estos cuadros pueden usarse separadamente o combinarse.

En la construcción de un cuadro hay que tener presente el objeto para el cual se construye y los elementos necesarios e indispensables para el conocimiento de ese objeto, a fin de eliminar del cuadro lo que sea superfluo para tal conocimiento. A fin de construirlo adecuadamente, se debe hacer previamente un cuadro "imaginario" o "fantasma" que el investigador debe tratar de llenar; en cuanto se presenta alguna dificultad al tratar de llenar uno de sus puntos, es necesario modificar o suprimir ese punto. Cuando es posible, conviene, tras esto, hacer una experiencia de prueba utilizando el cuadro con una porción pequeña de la población por estudiar para aprecias dificultades y reacciones. Además, es necesario cuidar la claridad y precisión de los términos empleados en el cuadro, los cuales deben ser, además, adecuados al nivel cultural medio del grupo que se investiga, debiendo recordarse, en ocasiones, que no se está redactando una pieza literaria, sino que se está elaborando un instrumental que debe ser iuncionalmente útil a la investigación y, por lo mismo, incluso adecuarse a los modos expresivos de los investigados. Asimismo, es necesario cuidar la presentación física, tamaño y forma de los cuadros, ya que en muchas ocasiones todo ello influye en los resultados.

Las fases por las que atraviesa la construcción de un cuadro son:

1ª—Formación de una lista de preguntas.
2ª—Constitución de una serie de cuadros fantasmas.
3ª—Ordenación de las preguntas conforme a:
 a.—una secuencia lógica,
 b.—una sucesión que evite el daño de una pregunta o su inutilización por alguna pregunta previa,
 c.—una presentación física adecuada.
4ª—Realización de algunas experiencias de prueba.
5ª—Elaboración del cuadro definitivo y
6ª—Instrucciones dadas a los recolectores.

Entre las precauciones que hay que tomar al llenar el cuadro, debe tenerse la de lograr respuestas inteligibles y asentarlas en forma legible, evitar las abreviaturas (a menos que se haya previsto la posibilidad de su utilización y se haya codificado su significado), y llenar con una línea aquellas preguntas a las que no se dio respuesta (para evitar que llegue a pensarse que no se llenó el espacio por descuido del observador o recopilador).

En la utilización de los cuadros, pueden presentarse dos alternativas:

a.—Investigación personal,
b.—Investigación por correo.

La investigación personal permite el que las preguntas sean más complejas y los resultados más apreciables, ya que da al investigador la posibilidad de aclarar el sentido de la pregunta, ampliarla o complementarla, en tanto que la investigación por correo no tiene esas ventajas, permitiendo en cambio resultados más amplios y la utilización de un número menor de recopiladores.

El cuestionario, a su vez, puede considerarse como un tipo especial de cuadro, sujeto en gran parte a las regulaciones de éste, debiendo tenerse en cuenta, además, que, en el caso del cuestionario:

Se ha de considerar que la persona a quien se dirige tiene escasa capacidad mental o dispone de poco tiempo para responder, de ahí que deba buscarse:

a.—que haya pocas preguntas,
b.—que sean sencillas,
c.—que puedan contestarse por un SI, por un NO o por un número,
d.—que cubran estrictamente la información requerida sin exceso ni defecto.
e.—que se pueda responder a ellas sin prejuicios (de ahí que deban evitarse las palabras cargadas emocionalmente, y buscarse la recolección de datos objetivos).

Escalas Sociométricas

La cédula y el cuestionario, a pesar de todas las precauciones que se tomen para aplicarlos, tienen el inconveniente de que muy a menudo resultan dañados de subjetivismo; de ahí que se haya tratado de evitar sus inconvenientes y de complementar los buenos resultados que mediante los mismos pueden obtenerse, por medio de técnicas de mensuramiento social como son las "escalas sociométricas". De ahí que pueda decirse que:

Una escala sociométrica es un instrumento usado en el mensuramiento social para categorizar o jerarquizar:

1.—Aspectos del medio social o cultural.—"escalas ecométricas".
2.—Conductas sociales individuales, actitudes y opiniones.—"psicométricas".

O sea, que las escalas sociométricas sirven principalmente para dar normas que hagan comparables dos o más comunidades o instituciones o bien para mostrar la separación de la conducta de una persona en relación con las normas grupales.

Para la construcción de una escala sociométrica, hay que tener en cuenta:

a.—la determinación de lo que se va a medir,
b.—la selección de los criterios y de los elementos que van a servir de base al mensuramiento,
c.—la hechura de una escala sencilla,
d.—la valoración y cuantificación de cada elemento.

Una escala sociométrica debe ser de aplicabilidad general dentro del campo elegido y conservar dentro del mismo, en todas sus aplicaciones su seguridad y validez, de tal modo que si la escala elegida no es generalmente aplicable por ser muy heterogéneo el grupo considerado debe preferirse la división del grupo en dos o más porciones homogéneas, y la elaboración de dos o más escalas (una para cada porción), y no la elaboración de una escala muy complicada y sólo parcialmente válida para todo el grupo.

Deben tomarse en cuenta criterios y elementos que por su objetividad y general repetición hagan comparables las instituciones, comunidades o personas medidos.

La escala sociométrica debe excluir factores que no sean utilizables, lo cual se logra en gran parte trabajando con grupos homogéneos (debe pensarse en el caso de una escala sociométrica que se quisiera aplicable lo mismo a los mexicanos que a los hindúes o a los bosquimanos, en cuanto a condiciones de habitación, por ejemplo, escala que contendría muchos renglones inútiles que entorpecerían su uso, y en mucho, la invalidarían). Sin embargo, la sencillez de la escala no debe estar reñida con la inclusión de todos los factores importantes para la medida o la comparación.

Por otra parte, para que se trate de una verdadera escala de medida, a cada factor se le debe asignar un valor numérico, el cual habrá de determinarse ponderalmente en relación con la importancia de cada elemento o renglón de la escala con respecto al conjunto de los que la forman.

Una escala sociométrica comporta, por lo menos, dos problemas fundamentales:

¿Cómo seleccionar los aspectos, factores o renglones de una unidad por medir?

¿Cómo determinar la medida o valor relativo que se ha de asignar a cada factor incluído?

¿Cómo seleccionar los renglones o factores a considerar dentro de la escala? Dejando la selección en manos de personas que conozcan el terreno de medición, a quienes se llama "jueces" y que pueden ser "generales" o "especiales", particularmente cuando se trata de problemas complejos, como el estudio integral de una comunidad; en este caso, los "jueces generales" especifican si se han de considerar los factores sanitarios, arquitectónicos, jurídicos, etc., en tanto que los jueces especiales son médicos, arquitectos, jurisconsultos que determinan en cada uno de esos aspectos lo que es necesario incluir en la escala.

El segundo problema es el relativo a la medida de los factores, o sea, que si al estudiar una institución se encuentra el rasgo *A* precisa responder a las preguntas: ¿cómo debe calificarse a esa institución en ese aspecto? ¿qué calificación (cuantitativa) debe otorgársele: 10, 8, 5, 0?, ya que, de la suma de las calificaciones dadas a todos y cada uno de los factores resultará la medida sociométrica de la institución que la estadística se encargará de elaborar.

La solución del problema de la calificación de los factores se ha logrado por uno de los siguientes medios:

1.—Calificación de existencia-carencia.
2.—Calificación de niveles de existencia.
3.—Calificación por diferencias.
4.—Calificación por puntuación sigmática.

La calificación de existencia-carencia es la más sencilla y consiste en asignar un cero al renglón cuando el rasgo considerado no existe, y 1 cuando existe.

La calificación de niveles de existencia asigna valores convencionales que van desde 0 cuando el rasgo no existe hasta un número *n* cuando existe en determinadas condiciones, o repetido cierto número de veces, etc.

La calificación por diferencia toma en cuenta los llamados "promedios laterales" (cuartilas principalmente que permiten calificar a un rasgo como "deficiente", "normal" o "excedente").

La calificación por la puntuación sigma tiene en cuenta la reducción a términos comparables lograda mediante la expresión de una serie de valores en términos de la desviación media cuadrática que, como las cuartillas, se estudiará posteriormente.

Supongamos que se trata de determinar ¿qué tan revolucionaria o qué tan conservadora es la gente del pueblo *A*, de la ciudad *B*, del país *C*?. El mejor procedimiento consiste en emplear una escala sociométrica referida al atributo "revolucionario" o al atributo "conservador" que se trata de estudiar.

A fin de construir la escala correspondiente, es necesario empezar por definir qué es lo que entendemos por "revolucionario" o por "conservador"; es decir, debemos dar una definición o concepto del atributo que queremos estudiar. Al final de la investigación, los resultados deberán ponerse en relación con —y deberá otorgárseles validez en relación y sólo en relación con— la definición de la que se ha partido; de ahí que la aplicación constituya un procedimiento, o sea algo que se integra por la sucesión de los diversos pasos y que sólo gracias a dicha sucesión ordenada tiene razón de ser.

Definido o conceptuado el atributo por estudiar, siguiendo un criterio conductista para descubrir las actitudes de los individuos, es necesario plantearse la pregunta de ¿qué actos o hechos de los individuos pueden tomarse como manifestaciones de la actitud estudiada o por estudiar? ¿qué actos de un individuo pueden juzgarse como demostrativos de que es un revolucionario (o de que es un reaccionario)? ¿Acepta la organización política del Estado, o piensa que sería más conveniente otra forma de gobierno; participa o no en movimientos sindicales, huelguísticos, etc.: acepta o rechaza las moras matrimoniales y sexuales de la sociedad en que vive, etc.? todo ello estudiado al través de hechos de su conducta.

Los actos de la conducta de un individuo considerados como manifestación del atributo o de la actitud que se estudia, reciben el nombre de "indicadores"; con ellos, debe formarse una lista *objetiva* y controlada por los límites de la definición que se haya dado en primer término (a menos que alguno de ellos muestre la conveniencia de ampliarla).

Como paso siguiente, debe someterse la lista a un conjunto de jueces cuidadosamente seleccionados:

1.—en razón de su capacidad científica, y
2.—habida cuenta de su objetividad de juicio,

a fin de que dichos jueces establezcan una jerarquía entre los diversos indicadores, señalando cuáles, según su parecer son los que apuntan a la existencia de una mayor intensidad del atributo estudiado.

El hecho de ser varios los jueces que jerarquizan independientemente los indicadores permite alcanzar un cierto grado de objetividad; en efecto, como paso siguiente, es necesario contrastar las diferentes jerarquizaciones para poner de manifiesto la igualdad, semejanza o diferencia entre la jerarquía dada a los diversos indicadores, a fin de que:

1.—cuando un indicador tiene una misma categoría en las jerarquizaciones hechas por los diversos jueces, dicha categoría se conserve en la escala.
2.—cuando un indicador tenga categorías próximas en las ordenaciones hechas por los diversos jueces se le dé como categoría el promedio de las mismas.
3.—cuando un indicador tenga categorías muy remotas en la clasificación de los distintos jueces, se le rechace por considerársele como indicador poco confiable.

Una vez asignados los valores promedio a cada indicador puede reducírseles a tantos porciento para hacer la escala más fácilmente manejable.

Preparada la escala, se puede aplicar a los diversos casos individuales, anotando la existencia o la ausencia del indicador. Sumando los valores de la escala correspondiente a los indicadores de cada caso individual, se tiene una cifra que indica un grado dentro de la escala: grado en que el individuo manifiesta con sus actos y opiniones, una opinión o actitud determinada.

Del estudio inicial y de la ejemplificación anterior resulta, en síntesis, que, para la aplicación de una escala sociométrica de este tipo son necesarios los pasos siguientes:

1º—Definición o conceptualización del atributo por investigar.

2º—Lista de actos u hechos que se consideren como indicadores de la existencia del atributo.

3º—Definición objetiva de los indicadores.

4º—Sometimiento de la lista a un cierto número de jueces con fines de jerarquización independiente de los indicadores por cada uno de ellos.

5º—Confrontación de las ordenaciones hechas por los jueces.

6º—Asignación de la puntuación dada por los jueces a un indicador cuando dicha puntuación coincida en las diferentes listas.

7º—Asignación del promedio de las puntuaciones cuando las puntuaciones de los diversos jueces difieren poco entre sí.

8º—Rechazo de los indicadores para los cuales difieren mucho las diversas puntuaciones.

9º—Sometimiento de la escala así construída a diferentes individuos cuya actitud u opinión quiera estudiarse.

ORDENACIÓN DE DATOS ESTADÍSTICOS

Con el objeto de analizar un conjunto de datos numéricos es útil, e incluso en la mayor parte de los casos indispensable, arreglar sistemáticamente los valores obtenidos, ordenándolos conforme a cierto criterio. Un conjunto ordenado de valores es lo que se conoce generalmente en Estadística como una "distribución" o como una "serie".

La ordenación de los datos puede hacerse de acuerdo con diferentes criterios: cuando el criterio seguido es el de localizar en un mapa el lugar en el que se presenta determinada magnitud del fenómeno, el arreglo resultante se conoce como "distribución espacial" o "distribución geográfica"; cuando el criterio seguido en el arreglo o en la ordenación de los datos es el de sus magnitudes (ya sea tomadas en forma creciente, decreciente o de otro tipo), se habla de "distribuciones o series estáticas"; cuando el criterio para la ordenación es el momento de aparición de las magnitudes en el tiempo, se habla de "series dinámicas, series temporales o series cronológicas". Aquí nos ocuparemos casi exclusivamente de los dos últimos tipos de distribución.

1.—el arreglo o serie sencilla,
2.—la serie o distribución de frecuencias,
3.—la serie de clases y frecuencias.

En el caso del arreglo, los valores se consignan comenzando por el menor (o por el mayor) tomando todos y cada uon de los que figuran en el conjunto de datos por ordenar, conforme a un orden creciente (o decreciente) hasta concluir con el mayor (el menor) de los datos del conjunto, y *sin evitar* las repeticiones.

En una distribución de frecuencias, también es necesario ordenar los valores consignándolos de menor a mayor o de mayor a menor, pero, a diferencia de lo que ocurre en el caso de formación de un arreglo o serie sencilla, en una distribución de frecuencias se evitan las repeticiones escribiendo frente a cada magnitud el número de veces que aparece en el conjunto. A esta última cantidad "número de veces que se repite un dato" se le da el nombre de "frecuencia". De este modo, una serie de frecuencias está constituída esencialmente por dos columnas: una en la que se consignan las magnitudes de los datos, y otra en la que se contienen las frecuencias o número de veces que se presenta cada magnitud.

En una distribución de clases y frecuencias, además de buscarse el orden como en el arreglo o serie sencilla, la condensación conseguida por medio de una distribución de frecuencias se lleva más adelante, mediante la constitución de grupos de valores definidos por un límite inferior y un límite superior; a cada uno de estos grupos se les da el nombre de "clases" y son a modo de casilleros en los cuales colocar los datos; el recuento del número de casos comprendidos entre los valores límites de cada clase (o sea el número de casos que tienen un valor superior al límite inferior, e inferior al límite superior de la clase correspondiente) da un valor que se conoce como "frecuencia" de clase. De este modo, una serie de clases y frecuencias está constituída esencialmente por tres columnas: una que contiene los límites inferiores de las clases, otra que contiene los límites superiores de las mismas y, frente a ellos una tercera conteniendo las correspondientes frecuencias de clase.

PROCEDIMIENTO DE FORMACIÓN DE UN ARREGLO, UNA DISTRIBUCIÓN DE FRECUENCIAS Y UNA DISTRIBUCIÓN DE CLASES Y FRECUENCIAS

1º—Cuéntese el número de casos que hay en el conjunto.

2º—Encabécese una columna de la tabulación que constituirá el arreglo con el rubro "número progresivo", anotando en dicha columna la serie de los números naturales comenzando con uno y terminando con el número de casos del conjunto.

3º—Recórranse los datos del conjunto hasta encontrar aquel cuya magnitud sea mínima (o máxima si el arreglo ha de ser decreciente).

4º—Consígnese dicho valor frente al 1 de la columna encabezada número progresivo y póngase una señal (√) delante del dato en el conjunto sin ordenar.

5º—Examínese de nuevo el conjunto de datos sin ordenar hasta descubrir otro caso cuyo valor sea igual al mínimo o que le siga inmediatamente; en uno o en otro caso, consígnese en la segunda columna del arreglo y póngase una señal (√) frente al dato consignado, en el conjunto que se ordena.

6º—Prosígase en la misma forma, sin evitar repetir en el arreglo los datos que se repitan en el conjunto hasta que todos los datos del conjunto aparezcan marcados con la señal y se haya llenado el espacio de la segunda columna correspondiente al último "número progresivo".

progresivo", ya que esto probará que no se ha dejado ningún caso fuera del arreglo.

El arreglo, como todas las demás ordenaciones de datos no cumple otra función que la de dejar preparado el ulterior trabajo de elaboración estadística; sin embargo, al través del arreglo es posible poner de relieve una de las características de la distribución: su "amplitud u oscilación".

La amplitud u oscilación de una distribución se define como la diferencia entre el valor máximo y el valor mínimo de la distribución. Cuando el valor mínimo es una cantidad negativa (esto ocurre cuando se recogen calificaciones de ciertas pruebas de sociabilidad, por ejemplo) es necesario recordar que esta diferencia entre el valor máximo y el valor mínimo debe entenderse como diferencia o resta algebraica, y que para efectuarla es necesario cambiar el signo del substraendo (o, puesto que el mínimo es una cantidad negativa que al cambiar de signo se convierte en positiva, sumar su valor al máximo en caso de que éste sea positivo, a fin de obtener la amplitud u oscilación).

Si representamos por *Amp.* la amplitud, por *Mn.* el mínimo de la distribución, y por *Mx* el máximo de la misma, la fórmula para la amplitud sería:

$$Amp. = Mx. - Mn.$$

Los casos que se presentarían en el cálculo serían:

1.—Máximo y mínimo positivos (réstense aritméticamente).
2.—Máximo y mínimo negativos (réstense algebraicamente cambiando el signo del mínimo, o sea, convirtiéndolo en positivo y restándolo del máximo).
3.—Máximo positivo y mínimo negativo (réstense algebraicamente cambiando el signo del mínimo o sea convirtiéndolo en positivo y sumándolo al máximo).
4.—Máximo negativo y mínimo positivo (réstense algebraicamente).

Para formar una serie de frecuencias, conviene seguir la siguiente rutina:

1º—Examínese el conjunto de datos sin ordenar, a fin de localizar el valor mínimo (o el máximo si se trata de series decrecientes).
2º—Consígnese dicho valor en el primer lugar de la primera columna de la tabla que constituirá la serie de frecuencias.
3º—Examínese el conjunto de datos a fin de determinar la diferencia mínima entre la magnitud de un dato y la del inmediato siguiente (si

se trata de salarios ¿es de un peso la diferencia mínima, o de 50 centavos, o de 25?).

4º—Agréguese al valor mínimo (o réstese del valor máximo en caso de series decrecientes) el valor de la diferencia mínima y anótese el valor resultante en la misma columna, 2 veces dicho valor, 3 veces dicho valor, etc., anotando siempre los resultados en esa columna hasta llegar al valor máximo (mínimo en el caso de series decrecientes) de entre los que figuran en la distribución.

5º—Examínese la serie original siguiendo sistemáticamente sus datos del primero al último, marcando cada uno de ellos con una señal (√) y anotando en la serie de frecuencias en formación una línea vertical (|) en cuanto el valor correspondiente haya aparecido por primera vez, una horizontal (L) que forme ángulo con la anterior al aparecer el mismo valor por segunda vez, otra vertical (⊔) al aparecer por tercera vez, otra horizontal (□) que completará un pequeño cuadrado, la cuarta vez en que el mismo valor aparezca, cruzando con una diagonal el cuadrado al aparecer ese mismo valor por quinta vez (⧅); de aparecer una sexta, séptima, etc. veces, se comenzará nuevamente el procedimiento cuidando de separar los cuadrados que sirven para llevar la cuenta, a fin de que la cercanía no haga que se confundan dos unidades contiguas y se cuenten por una.

6º—Tras el examen sistemático del conjunto de datos sin ordenar y el procedimiento de cuenta realizado en esta forma, todos los datos del conjunto deberán aparecer marcados; la cuenta de los cuadrados cruzados por diagonales (iguales a 5), los cuadrados no cruzados (iguales a 4) y las barras sueltas (cada una igual a 1) debe dar un resultado igual al número de datos del conjunto original.

7º—Determinar las frecuencias correspondientes a cada uno de los valores de la distribución equivale a contar el número de cuadrados cruzados y no cruzados y el de barras que figuran frente a cada valor. En casos en que no haya cuadrados o barras frente a un valor, la frecuencia es cero y dicho valor no tiene por qué figurar en la distribución de frecuencias puesta en limpio.

8º—Hecha la cuenta, la distribución de frecuencias se consignará en dos columnas:

una conteniendo los valores (no repetidos) a los que corresponda una frecuencia distinta de cero, o sea que aparezcan efectivamente en el conjunto, y

otra conteniendo los números (y no ya los cuadrados o barras del procedimiento de cuenta) correspondientes a la frecuencia o número de veces que aparece cada dato.

La rutina que puede utilizarse para la formación de una serie de clases y frecuencias consiste en:

1º—Examinar el conjunto de datos sin ordenar para determinar los valores mínimo y máximo.

2º—Iniciar la serie de clases tomando como límite inferior de la primera clase un valor que sea igual al mínimo o inferior a él (¿qué tan inferior? es pregunta a la que responderán otras consideraciones).

3º—Examinar el conjunto de datos a fin de determinar si hay cierto tipo de valores que por su mayor abundancia deban tener una representación más precisa (por ejemplo, ¿se repiten mucho las cifras en 0, los múltiplos de 5, los múltiplos de 3, etc.?).

4º—Elegir un intervalo de clases

a.—Grande si la amplitud de la distribución es grande, pequeña si la amplitud de la distribución es pequeña,

b.—Más pequeño conforme sea mayor la precisión buscada (esto hace que aumente el número de clases o grupos y con ello la dificultad de las elaboraciones).

c.—Más grande conforme sea mayor la facilidad buscada (esto hace que disminuya el número de clases o grupos y con ello la dificultad de las elaboraciones).

d.—Constante para todas las clases (para facilitar los cálculos).

e.—Que permita, en combinación con una buena elección del primer límite inferior el que los puntos medios de cada clase (obtenidos de sumar su límite inferior y su límite superior y dividir entre 2) sean representativos de los valores que resultan más frecuentes en la distribución, y de los que se habló en *3.*

f.—De ser posible, que permita obtener puntos medios que sean números enteros.

5º—Elegido el intervalo de clase conforme a todos estos criterios (o conforme a los que sea posible aplicar en el caso), es necesario, formar las clases eligiendo el primer límite inferior de la primera clase teniendo en cuenta las consideraciones hechas en *3* y en *4 e;* los límites inferiores siguientes se consignarán en la misma columna y se les obtendrá sumando al primero el valor del intervalo, *1, 2, 3, ... n* veces hasta obtener un límite inferior igual o inferior al valor máximo del conjunto; en seguida, será necesario:

6º—Consignar en otra columna frente a la que contiene los límites inferiores, los límites superiores de las clases para lo cual:

a.—si se trata de series discontínuas (o sea aquéllas en las que los valores pasan de una unidad a otra por "saltos" o sin transiciones, como ocurre con número de miembros de una familia), tomando la unidad inmediatamente anterior a la que representa el límite inferior del intervalo siguiente,

b.—si se trata de series continuas (o sea aquéllas en que el tránsito de un valor a otro se hace por pequeños incrementos, como ocurre con años, meses, semanas, días, horas, minutos y segundos de edad de un individuo), tomando como límites superiores de cada clase los mismos límites inferiores de la clase siguiente indicándose por un guión pospuesto el que dicho límite superior limita a la clase, pero "no está incluído en ella", de modo que un valor igual a dicho límite superior corresponderá a la clase siguiente y no a ésa (hay otros procedimientos alternativos de marcar tales límites superiores).

7º—Examinar el conjunto de datos sin ordenar, siguiendo sistemáticamente sus valores del primero al último, marcando cada uno de los que se adscriban a una clase y marcando, por el procedimiento empleado en el caso de las series de frecuencias ,frente a la clase correspondiente, el que dicho dato corresponde a una clase dada cuando es igual o superior al límite inferior de la misma e inferior (pero no igual) a su límite superior (dentro del supuesto de nuestra notación de los límites).

8º—Contar, cuando todos los datos del conjunto hayan sido marcados y adscritos a una clase, el total representado por los cuadrados cruzados, no cruzados y por las barras a fin de comprobar si se han consignado todos los datos del conjunto, y

9º—Determinar las frecuencias correspondientes a cada clase contando los cuadrados cruzados y no cruzados y las barras que figuran frente a cada clase.

10º—Tabular en limpio la distribución de clases y frecuencias consignando en dos columnas los límites inferiores y superiores de las clases y, frente a ellos las frecuencias correspondientes.

FORMAS ALTERNATIVAS DE PRESENTACIÓN DE LAS CLASES

Las clases se pueden representar en varias formas:

I.—Utilizando los límites inferiores y superiores para caracterizarlas, en cuyo caso, los límites pueden indicarse de los siguientes modos:

a.—tomando como iguales el límite superior y el límite inferior de dos clases contiguas (el límite superior de una clase igual al inferior de la subsecuente), sin ninguna otra especificación, en cuyo caso la clase se lee "de L_i a L_s", forma defectuosa que impide una adscripción unívoca de ciertos datos del conjunto a una clase determinada: cuando el valor coincide con el límite superior de la 6ª clase, por ejemplo, coincide también con el límite inferior de la 7ª y por lo mismo queda en duda si hay que consignarlo como dato correspondiente a la sexta clase, o contarlo entre los de la séptima (suele salvarse la dificultad dividiendo la frecuencia entre 2 y asignando media frecuencia a cada una de las clases contiguas).

b.—tomando como iguales el límite superior de una clase y el inferior al de aquélla en que están dados los valores que se ordenan los límites superiores; en este caso, las clases se leen "de L_i a L_s pero no incluyendo L_s"; de este modo, una valor que coincida con el del límite superior de una clase corresponde a la clase siguiente, con cuyo límite inferior también coincide. Es el uso que adoptamos haciendo preceder L_s de un guión.

c.—tomando como iguales el límite superior de una clase y el inferior de la siguiente pero haciendo preceder a los límites superiores de la expresión "a menos de". El procedimiento equivale al anterior, con la desventaja de una mayor vaguedad en la expresión y el inconveniente de ser menos práctico escribir toda una expresión que un simple guión.

d.—tomar como límites superiores de las clases los límites inferiores de las subsecuentes disminuídas en una unidad de orden inferior al de aquella en que están dados los valores que se ordenan (así, por ejemplo, si los valores están dados en pesos, puede hacerse que un límite superior difiera del inferior de la clase siguiente en un décimo; si los valores difieren de diez en diez centavos, pueden hacerse diferir los límites inferior de una clase y superior de la previa en un centavo, etc.). El procedimiento sigue equivaliendo al anterior, es más preciso, pero en ocasiones produce cierta confusión en el momento de los cálculos.

e.—tomar los límites inferiores como iguales a los límites superiores de las clases precedentes pero considerándolos como no incluídos, mediante los procedimientos correlativos de los anteriores de hacerles preceder de un guión (De L_i, no incluído, a L_s incluído), de la expresión "de más de" (de más de L_i a L_s) o de hacerlos diferir en una unidad de orden inferior, de los límites superiores de las clases precedentes.

II.—Mediante los puntos medios.

El procedimiento evita las dificultades de los anteriores en ciertos respectos, pero impide el uso de determinadas fórmulas ya consagradaspara el cálculo, en las que se utilizan ya límites superiores o ya límites inferiores (las cuales podrían modificarse pero con detrimento de la sencillez y brevedad) así como porque dificulta la adscripción de los datos del conjunto original a las clases correspondientes.

CONSIGNACIÓN TABULAR DE DATOS ESTADÍSTICOS

En la formación de tablas o cuadros estadísticos —especialmente en cuanto se destinan a la presentación de dasto o resultados— aun cuando no huelga tenerlo en cuenta en las mismas tablas o cuadros de trabajo— precisa considerar los siguientes elementos:

1.—Un *Número* que le corresponde al cuadro en el conjunto de la investigación, destinado a permitir el que se haga una referencia rápida al mismo y se le pueda relacionar con otras formas de presentación textual o gráfica.

2.—Un *Título* en el que se debe expresar:
- *a.*—cuál es el fenómeno de que se trata,
- *b.*—cuál su circunscripción espacial,
- *c.*—cuál el espacio temporal o período que abarca, y, en ciertos casos, las unidades empleadas.

3.—*Casillas-cabecera* destinadas a contener la denominación o encabezado de las columnas y, en ciertos casos, las unidades correspondientes.

4.—*Renglones,* o sean las divisiones horizontales que corresponden a las diversas categorías de un criterio clasificatorio.

5.—*Espacios* entre los renglones, que sirven para destacar ciertos agrupamientos de datos, ya sea de acuerdo con criterios de clasificación lógica de las especies en géneros, o ya sea según un criterio meramente pragmático de presentación más diáfana o clara de los materiales (de donde, en este último caso interlineado mecánico cada cierto número de renglones).

6.—*Columnas* o divisiones verticales, que permiten disponer los datos según un segundo criterio de clasificación. En los cuadros o tablas de trabajo, las columnas corresponden generalmente a etapas de cierto proceso de cálculo; sus encabezados generalmente indican, en este caso, la forma en que se obtuvieron los valores consignados en la columna.

7.—*Líneas intercolumnares,* generalmente sencillas, dobles cuando se quiere destacar el agrupamiento de ciertos datos (en estos casos, la línea doble corresponde generalmente a casillas-cabecera sub-divididas).

8.—Fuente de la que se obtuvieron los datos.

9.—Notas de pie, cuando sea necesario hacer aclaraciones con respecto a los datos consignados (por ejemplo, señalar que un valor corresponde a una estimación y no a algo verdaderamente observado, o que otro dato es poco confiable; que en determinado caso se ha utilizado un criterio de clasificación distinto, no incluyéndose a determinado grupo de edad, por ejemplo, que sí ha sido incluído en el dato de población de una fecha distinta, etc.).

PRINCIPALES FORMAS DE REPRESENTACIÓN ESTADÍSTICA

Las formas de representación gráfica de los datos y resultados estadísticos pueden agruparse como sigue:

1.—Lineales:
- *A.*—con escala aritmética
  - *a.*—polígonos de frecuencia (absoluta o relativa).
  - *b.*—ojivas (con frecuencias acumulativas absolutas o relativas).
    - α.—"más de",
    - β.—"menos de".
  - *c.*—series cronológicas.
- *B.*—con escala semilogarítmica,
  series cronológicas.
- *C.*—con escala logarítmica,
  series cronológicas.
- *D.*—con rayado probabilístico,
  distribuciones de frecuencias.
- *E.*—flechas de grosor variable para expresar la importancia de un movimiento (migraciones, tráfico citadino, etc.).

2.—Superficiales:
- *A.*—Gráficas de barras
  - *a.*—simples,
  - *b.*—subdivididas (en unidades absolutas o porcentualmente),
  - *c.*—agrupadas,
  - *d.*—pareadas (distribuciones por edad y sexo, etc.),
  - *e.*—de desviaciones o pérdidas y ganancias.
  - *f.*—deslizantes.
- *B.*—Histogramas y Gráficas de escalones.
- *C.*—Diagramas circulares,
  - *a.*—en pastel,
  - *b.*—en mariposa.

3.—Sólidos o de volúmenes.

4.—Diagramas de dispersión.

5.—Mapas estadísticos,

sombreados,
punteados,
de alfileres, banderas, etc.

De entre las anteriores formas de presentación gráfica, una gran mayoría se basa en el uso de un sistema de coordenadas cartesianas o rectangulares; dicho sistema está constituído por dos rectas (una vertical y una horizontal) que se cortan en ángulo recto en un punto llamado "origen". A la recta vertical se le designa generalmente como eje de las *yes* y sus extremos se designan en la gráfica por Y, Y'; a la recta horizontal se le llama eje de las *equis* y sus extremos se designan por X, X'; al punto en que las rectas se cortan, o sea al origen se le designa por 0.

Para localizar un punto P en el plano, basta con medir la distancia de dicho punto a cada un ode los ejes trazando a cada uno de ellos, desde el punto-dos perpendiculares que se conocen como las coordenadas de dicho punto. A la perpendicular que mide la distancia del punto al eje horizontal o eje de las equis se le conoce como "ordenada"; a la perpendicular que mide la distancia del punto al eje vertical o eje de las yes se le conoce como "abscisa".

Por convención, se considera que:

1.—Las ordenadas que se miden hacia arriba del eje horizontal son positivas.
2.—Las ordenadas medidas hacia abajo del eje horizontal son negativas.
3.—Las abscisas medidas a la derecha del eje vertical son positivas.
4.—Las abscisas medidas a la izquierda del eje vertical son negativas.

O sea, que:

En el cuadrante superior de la derecha, un punto tiene abscisa positiva y ordenada también positiva.

En el cuadrante superior de la izquierda, los puntos tienen abscisa negativa y ordenada positiva.

En el cuadrante inferior de la izquierda, los puntos tienen abscisa negativa y ordenada negativa.

En el cuadrante inferior de la derecha, los puntos tienen abcisa positiva y ordenada negativa.

En las representaciones estadísticas, de todos los cuadrantes, el más utilizado es el superior de la derecha, aun cuando ocasionalmente se presente el caso de utilización de los restantes (del superior de la izquierda se hace uso cuando se emplea el procedimiento de interpolación de líneas de tendencia valiéndose del año mediano; de los dos inferiores se puede hacer uso en caso de representación de pérdidas y ganancias, de variaciones cíclicas, etc.).

Con el objeto de faciliar la localización de los puntos en el plano, los dos ejes se gradúan generalmente tomando sobre ellos espacios iguales representativos de una unidad elegida; de este modo, para localizar un punto cuyas unidades se conocen se toman dichas coordenadas sobre los dos ejes, se levantan perpendiculares en los puntos correspondientes y en el cruce de ambas se localiza el punto; así, por ejemplo, si se trata de un punto de abscisa *m* y ordenada *n*,

1º—Se contarán *m* unidades a partir del origen sobre el eje horizontal, y en ese punto se levantará una perpendicular,

2º—Se contarán *n* unidades a partir del origen sobre el eje vertical, y en ese punto se levantará una perpendicular *(i. e.* se trazará una paralela al eje horizontal) hasta cortar a la perpendicular trazada en *1*. En donde se corten, se encontrará el punto.

El sentido en el que hayan de contarse las unidades de la abscisa y de la ordenada correspondientes dependerá del signo que tengan; así, si la abscisa es positiva, las unidades se contarán a la derecha y si, al mismo tiempo la ordenada es positiva en ese punto se levantará (en sentido estricto) la perpendicular correspondiente; en cambio, si la ordenada es negativa, desde ese punto se *bajará* la perpendicular correspondiente, etc.

Para trazar un polígono de frecuencias mediante el uso de las coordenadas rectangulares,

1.—Fórmese con los datos una distribución de frecuencias o una distribución de clases y frecuencias.

2.—En caso de tratarse de una distribución de frecuencias tómense como abscisas (o sea, como distancias al origen medidas sobre el eje horizontal), los valores de los datos; en caso de tratarse de una distribución de clases y frecuencias, tómense como abscisas los puntos medios o marcas de clase (véase adelante). En tales puntos, levántense las perpendiculares correspondientes.

3.—Ya sea que se trate de una simple distribución de frecuencias o de una serie de clases y frecuencias, tómense como ordenadas (o sea, como distancias al origen medidas sobre el eje vertical) las frecuencias correspondientes. Levántense las perpendiculares al eje vertical, y

4.—Encuéntrese la intersección de cada una de esas perpendiculares con la levantada en el punto correspondiente al valor del dato cuya es la frecuencia.

5.—Únanse los puntos así encontrados por medio de líneas rectas a fin de obtener el trazo de una línea quebrada que representa la distribución de las frecuencias y se conoce como polígono de frecuencias.

Lectura de un polígono de frecuencias: Para poder hacer lecturas en un polígono de frecuencias.

1.—Tómese un punto del mismo.
2.—Mídase sobre la perpendicular (bajada del punto al eje) su distancia al eje horizontal y léase como primer elemento de la oración "hay *tantos casos*" (aquí la distancia medida) "para los cuales la magnitud del fenómeno es de..."
3.—Léase sobre el eje horizontal la distancia entre el pie de la perpendicular y el origen y complétese la oración anterior "... tanto" (aquí la distancia recién medida).

De esta forma, si se cuenta con un polígono de frecuencias como única fuente de información y se desea saber cuántos casos de la distribución tienen una medida x dada:

1.—Se buscará el valor de x sobre el eje horizontal.
2.—En ese punto se levantará una perpendicular, y
3.—Se medirá la distancia entre el pie de la perpendicular y la intersección de la misma con el polígono de frecuencias; dicha medida dará la frecuencia correspondiente, o sea, el número de casos que en el conjunto estudiado corresponden a esa magnitud del fenómeno.

Como es fácil comprender, el trazo de un polígono de frecuencias no es sólo útil para "retratar" la forma de la distribución e incluso para tener una idea de cuál pudiera ser su distribución teórica en caso de que se pudieran eliminar errores accidentales, etc., sino que también permite tener una idea de cuál podrá ser la frecuencia con la que se presentarán magnitudes intermedias del fenómeno que la agrupación en una serie de clases y frecuencias subsume bajo una misma denominación de magnitud (la magnitud del punto medio).

Los polígonos de frecuencias sirven también para comparar dos o más distribuciones, especialmente, si se trata de polígonos para los cuales las frecuencias sean frecuencias relativas. Una frecuencia relativa se obtiene dividiendo la frecuencia absoluta correspondiente entre la suma de todas las frecuencias:

$$f_r = \frac{f}{\Sigma f_i}$$

La representación de un polígono de frecuencias relativas sigue la misma rutina que la representación de un polígono de frecuencias absolutas. La lectura de este tipo de gráfica es análogo; se tratará de que "hay tantos casos" (aquí la frecuencia relativa u ordenada) para los cuales la magnitud del fenómeno es tal (aquí la abscisa) si se considera al conjunto de casos como unitario.

Cuando la frecuencia relativa se multiplica por 100 se obtiene una frecuencia porcentual; el polígono de frecuencias que se represente tomando en cuenta estas frecuencias porcentuales permitirá una lectura del tipo siguiente: "de cada cien casos hay tantos (aquí la frecuencia porcentual u ordenada) en los cuales la magnitud del fenómeno es de tanto (aquí el valor de la abscisa)".

En la representación de polígonos de frecuencias relativas o porcentuales, a más de los ejes coordenados que siempre deben representarse con líneas más gruesas que las de la gráfica misma, deberá representarse asimismo con trazo más grueso la paralela al eje horizonal que corresponda a la unidad (en el caso de los relativos) o al 100% (en el caso de los tantos por ciento).

Ojivas.—Las ojivas son representaciones esiformes, sigmoides o en forma de *s*, correspondientes a distribuciones acumulativas. Para hacer este tipo de distribuciones:

1.—Fórmese una distribución acumulativa:

- *a.*—sumando las frecuencias en forma progresiva a partir de la correspondiente a los valores inferiores de la variable (o a las clases inferiores) y continuando hacia la correspondiente a los valores superiores de la variable (o a las clases superiores) si se trata de la ojiva llamada "menos de".
- *b.*—sumando las frecuencias en esa misma forma progresiva pero en sentido inverso: de los valores superiores (o clases superiores) hacia los valores inferiores (o clases inferiores), si se trata de la ojiva llamada "más de".

2.—Tómense los valores de la variable más o menos un medio (en series de frecuencias), o los límites de clase (en series de clases y frecuencias) sobre el eje horizontal:

- *a.*—tomando los valores de la variable *más* un medio (series de frecuencias), o los límites *superiores* (series de clases y frecuencias) en el caso de la ojiva *menos* de.
- *b.*—tomando los valores de la variable *menos* un medio o los límites *inferiores* (en los casos correspondientes) en el caso de la ojiva *más* de.

3.—Tómense las frecuencias acumulativas correspondientes sobre el eje vertical, trácense perpendiculares y en el cruce obténganse los puntos de intersección.

4.—Únanse los puntos de intersección para obtener la ojiva.

Lectura de las ojivas.—Si se trata de una ojiva "menos de":

1.—Tómese un punto sobre la ojiva.

2.—Mídase sobre la perpendicular bajada al eje horizontal la distancia del punto al eje, leyendo "hay tantos casos" (aquí la distancia medida), para los cuales la magnitud del fenómeno es de menos de...".

3.—Mídase la distancia del pie de la perpendicular al origen sobre el eje horizontal, y complétese la oración: "...tanto" (aquí el valor de la distancia medida).

En forma análoga, si se trata de una ojiva "más de" se leerá "hay tantos casos, para los cuales la magnitud del fenómeno es más de tanto".

Como en el caso de los polígonos de frecuencias, en el caso de las ojivas es también posible utilizar relativos o tantos por ciento como frecuencias; en tales casos las lecturas serán como sigue:

"Hay tantos casos para los cuales la magnitud del fenómeno es menos de (más de) tanto, si se considera el conjunto de la distribución como unitario."

"De cada cien casos, hay tantos para los cuales la magnitud del fenómeno es menos de (más de) tanto".

SIMBOLOGÍA ESTADÍSTICA

La simbología estadística es, fundamentalmente, la misma simbología matemática, constituída por un conjunto de símbolos que pueden agruparse en las siguientes categorías:

1.—Símbolos que carecen de valor como magnitudes:

A.—Signos de operación:

$+$, $-$, $\times$, $\div$, exponente, $\sqrt{}$

$+$ indica "cuéntense a partir de la última unidad del primer sumando o término tantas unidades como lo indique el segundo".

$-$ indica "cuéntense a partir de la última unidad del minuendo y *en sentido inverso* tantas unidades como indique el substraendo". De ahí la regla de que para restar es necesario cambiarle signo al substraendo (equivale a "ir en reversa").

$\times$ indica "tómese el multiplicando tantas veces como sumando como señale el multiplicador".

$\div$ equivale a "determínese cuántas veces (cociente) es necesario restar el divisor del dividendo para obtener cero como residuo (en el caso de la división exacta) o para obtener un residuo menor que el divisor (en caso de que la división no sea exacta).

Exponente equivale a "Tómese la cantidad afectada tantas veces como factor como indica el exponente".

$\sqrt{}$ vale tanto como "Búsquese el número que tomado como factor tantas veces como indica el índice del radical produce la cantidad colocada debajo del radical (cantidad sub-radical).

B.—Signos condensados o taquigráficos de operación u operadores. Generalmente se utilizan como operadores las letras griegas mayúsculas y, ocasionalmente, las griegas minúsculas o ciertas letras latinas modificadas:

Σ *Sigma* (mayúscula).—Equivale a la imperativa "súmese todo lo que sigue" o bien a la orden "súmense todos los *términos* del tipo siguiente".

Π *Pi* (Mayúscula).—Equivale a la orden de "multiplicar entre sí todos los *factores* del tipo siguiente".

Δ *Delta* (mayúscula.—Vale tanto como decir "auméntese o increméntese el valor siguiente en una cierta cantidad".

δ *Delta* (minúscula).—Se usa en ocasiones para indicar: "obténgase la derivada de lo sigüiente" o sea "tómese el límite del incremento de lo que sigue, dividido entre el incremento de su variable independiente cuando este último incremento tiende hacia cero".

$\int$ *Ese latina alargada.*—Vale por la orden de integrar todo lo que le sigue, o bien por "determínese cuál es la función que tiene por límite de la relación entre su incremento y el incremento de la variable independiente el siguiente, cuando el incremento de la variable independiente tiende hacia cero". Fundamentalmente, indica la necesidad de ejecutar una operación inversa de la derivación y, en el fondo, equivale a un proceso de suma en el que cada sumando es muy próximo al que le subsigue.

2.—Símbolos de valor distintivo ordinal.

Estos símbolos generalmente están representados por letras latinas minúsculas o por números romanos colocados como si fueran exponentes, pero sin serlo.

3.—Símbolos de valor único o constantes.

Generalmente estos símbolos se representan por:

A.—Las primeras letras del alfabeto latino.

B.—Las minúsculas del alfabeto griego.

Las constantes pueden ser:

A.—Constantes absolutas (como π minúscula, *e* minúscula) cuando tienen un solo valor, sea cual fuere el contexto en el que se encuentren, así:

π (*Pi* minúscula) vale siempre 3.1416...

e (*e* minúscula) vale siempre 2.7182...

B.—Constantes relativas, o sean aquéllas que tienen un solo valor dentro de determinado contexto o en determinada ecuación, pero que alcanzan valores diferentes para diferentes contextos o ecuaciones. Así, en la ecuación de la recta: $y = a + bx$, *a* valdrá 3 para la recta que corte al eje de las *yes* en el punto 3. Valdrá — 8 para la recta que corte a dicho eje 8 unidades por debajo del origen y será, por tanto, una constante relativa.

4.—Símbolos de valor múltiple o variables.

Para representar a las variables se utilizan, generalmente, las últimas letras del alfabeto latino. El signo representativo de una variable está en lugar de todo un género o conjunto de datos sin que represente a ninguno de ellos en particular; así, si x representa la variable "edad" para un conjunto de tres niños de 14, 11 y 10 años, x representará a todas y cada una de esas tres edades y no a cada una de ellas por separado ni a su suma; para *especificar* que se alude a la edad del primer niño se agregará al signo de la variable el subíndice $_1$; de este modo x_1 valdrá 14 años; para tener especificada la edad del segundo se agregará a la x el sub-índice $_2$: x_2 representará los 11 años de edad del segundo niño, y x_3 los 10 del tercero. En general, para indicar la variable no especificada se suele poner un subíndice, generalmente $_i$ minúscula, o $_j$ minúscula; de este modo, en el caso, x_i representa *genéricamente* edades del conjunto de tres niños.

EL OPERADOR SIGMA

Uno de los operadores más empleados en Estadística es el operador sigma mayúscula (Σ) que indica que deben sumarse todos los términos del género de los colocados delante de él. De este modo:

Σx_i indica: "Súmense todas las equis" o bien
"Súmense $x_1 + x_2 + x_3 + x_4 \ldots$".

Si x_i representa estaturas, Σx_i indicará:

"Súmente las estaturas de los individuos del siguiente universo, colectivo o población."

En ocasiones la suma indicada no se extiende a todo el universo; en tales ocasiones se especifican los límites entre los cuales ha de aplicarse la suma colocando debajo del operador sigma mayúscula el número de orden del valor asumido por la variable que ha de servir como primer sumando, y arriba del propio operador, el número de orden del valor de la variable que ha de servir como último término de la suma; de este modo:

$\sum_{i=3}^{5} x_i$ significará que se necesitarán sumar únicamente las estaturas de los individuos designados con los números 3, 4 y 5.

Cuando no se especifican los límites entre los cuales se extiende la suma, debe entenderse que la misma comprende todo el dominio de la variable, o todo el universo, colectivo o población estudiado.

En el uso del operador *sigma* deben considerarse ciertas reglas y precauciones elementales que no agotan, ni con mucho, la normatividad matemática que lo rige, pero que pueden contarse entre las de inmediata aplicación en la Estadística.

Primera regla.—Puesto que *sumar* una serie de cantidades constantes equivale a *multiplicar* la cantidad constante por el número de veces que aparece en la serie, cuando el operador sigma mayúscula aparece afectando a una constante, puede substituírsele por la magnitud que represente el número de veces que aparece la cantidad constante.

Sumar $5 + 5 + 5$ tres veces equivale a
Multiplicar 5×3.

O sea, si empleamos el operador:

$$\sum_{i=1}^{3} (5) = 3 \times 5 = 15.$$

En general, para una constante a que aparezca N veces:

$$\sum_{i=1}^{N} (a) = Na. \text{ (N veces a ó N multiplicado por a.)}$$

Segunda regla.—Si los diferentes valores de una variable / están multiplicados por una misma constante / y sus productos se suman / dicha suma / será igual al resultado que se obtendría / si se sumaran primero los diferentes valores de la variable / y se multiplicara la suma por la constante:

Diferentes valores de una variable:

x_i
$x_1 = 14$
$x_2 = 11$
$x_3 = 10$

Valores de la variable multiplicados por la constante $a = 2$.

ax_i
$ax_1 = 2 \times 14 = 28$
$ax_2 = 2 \times 11 = 22$
$ax_3 = 2 \times 10 = 20$

suma de esos productos: $\Sigma(ax_i) = 70$

Suma que se obtendría mediante la adición de los diferentes valores de la variable:

x_i
$x_1 = 14$
$x_2 = 11$
$x_3 = 10$

$\Sigma x_i = 35$

suma de los valores de la variable por la constante:

$$\Sigma x_i = 2 \times 35 = 70$$

O sea, que, como se había asentado anteriormente:

$$\Sigma a x_i = a \Sigma x_i$$

La regla puede expresarse prácticamente como sigue:

"Todo *factor constante* puede salir o entrar al operador Σ *sin* modificarse".

Tercera regla.—Multiplicar / los diferentes valores de una variable / por los correspondientes valores de otra variable / *no es igual* a / sumar los diferentes valores de la primera / sumar separadamente los de la segunda / y multiplicar en seguida las sumas obtenidas:

Valores de la variable número de casos y valores correspondientes de la variable estatura:

f_i	x_i.	
$f_1 x_1$	$= 19.60$	$(= 14 \times 1.40)$
$f_2 x_2$	$= 13.20$	$(= 11 \times 1.20)$
$f_3 x_3$	$= 11.00$	$(= 10 \times 1.10)$

Producto de los valores de la variable por los de la segunda:

	$f_i x_i$
	$f_1 x_1 = 19.60$
	$f_2 x_2 = 13.20$
	$f_3 x_3 = 11.00$
Suma de los productos:	$\Sigma f_i x_i = 33.80$

En cambio:

	fi
	14
	11
	10
Suma de los valores de la primera variable:	$\Sigma f_i = 35$

xi
1.40
1.20
1.10

Suma de los valores de la segunda variable: $\Sigma x_i = 3.70$

Producto de las sumas: $\Sigma f_i \times \Sigma x_i = 35 \times 3.70 = 129.50$

O sea, que como expresa la regla:

$$\Sigma f_1 x_1 \neq \Sigma f_i \, \Sigma x_i$$

Si se expresa esta regla en forma más concisa, podrá decirse que:

"La suma de los productos de las variables *no es igual* al producto de las sumas de las variables."

En general, puede asentarse asimismo que:

Si los diferentes valores de una variable se someten a cierta operación (se elevan al cuadrado, al cubo, se logaritman, etc.) y después se suman los resultados; el valor final obtenido *no es igual* al resultado que se obtendría sumando primero los diferentes valores de cada una de las variables y sometiendo después dicha suma a la operación de que se trate (elevándola al cuadrado, al cubo, etc., tomando su logaritmo, etc.).

II.—TÉCNICAS

MEDIAS

Media estadística es el valor representativo de una distribución, que se obtiene si:

1º—se sujeta a la variable a cierta operación,
2º—se suman los valores obtenidos,
3º—se divide la suma entre el número de datos, y
4º—se somete al cociente a la operación inversa de aquella a la que se sujetara a la variable.

Según la operación que se realice con los diversos valores de la variable (y la operación inversa a la que se sujete al cociente que resulta de dividir la suma de los resultados entre el número de datos), la media recibirá el calificativo de: aritmética, cuadrática, cúbica, armónica o geométrica.

La media aritmética se obtiene si:

1º—se dejan intactos los valores de la variable,
2º—se suman dichos valores,
3º—se divide la suma entre el número de datos, y
4º—se deja intacto el cociente así obtenido.

La media cuadrática se obtiene si:

1º—se elevan al cuadrado los valores de la variable,
2º—se suman dichos cuadrados,
3º—se divide la suma de los cuadrados entre el número de datos, y
4º—se extrae la raíz cuadrada (operación inversa de la potenciación) del cociente obtenido.

La media cúbica se obtiene si:

1º—se elevan al cubo los valores de la variable,
2º—se suman dichos cubos,
3º—se divide la suma de los cubos entre el número de datos, y
4º—se extrae la raíz cúbica del cociente.

La media armónica se obtiene si:

1º—se obtienen los recíprocos de los valores de la variable,
2º—se suman dichos recíprocos,
3º—se divide la suma de los recíprocos entre el número de datos, y
4º—se busca el recíproco del cociente.

La media geométrica se obtiene si:

1º—se buscan los logaritmos de los valores de la variable,
2º—se suman dichos logaritmos,
3º—se divide la suma de los logaritmos entre el número de datos, y
4º—se busca el antilogaritmo del cociente.

En este último caso (o sea, en el de la media geométrica), en cuanto las operaciones con logaritmos tienen sus equivalentes en operaciones realizadas con los números naturales correspondientes, siendo la suma de logaritmos equivalente al producto de los números a los que corresponden, la resta equivalente del cociente, el producto equivalente de la potenciación, y el cociente equivalente de la raíz, puede expresarse también que:

1º y *2º*—se multiplican entre sí los valores de la variable,
3º y *4º*—se extrae la raíz *enésima* de dicho producto (en cuanto *ene* represente el número de datos).

De acuerdo con lo anterior tendremos:

Para la media aritmética:

Valores de la variable:	$x_1, x_2 \ldots x_n$	x_i
Los mismos intactos:	$x_1, x_2 \ldots x_n$	x_i
Su suma:	$x_1 + x_2 + \ldots x_n$	Σx_i
Su suma dividida entre el número (N) de datos:		$\dfrac{\Sigma x_i}{N}$
La misma suma intacta:		$\dfrac{\Sigma x_i}{N}$

Fórmula de la media aritmética ($\bar{x}$):

$$\bar{x} = \frac{\Sigma x_i}{N}$$

En forma análoga, para la media cuadrática:

Valores de la variable: x_i

Cuadrados de dichos valores: x_i^2

Suma de dichos cuadrados: $\Sigma(x_i^2)$

Cociente de dicha suma entre el número (N) $\dfrac{\Sigma(x_i^2)}{N}$

Raíz cuadrada del Cociente: $\sqrt{\dfrac{\Sigma(x_i^2)}{N}}$

Fórmula de la media geométrica. ($\bar{x}_g$)

$$\bar{x}_g = \sqrt{\frac{\Sigma(x_i^2)}{N}}$$

Notación alterna de la fórmula para la media cuadrática:

$$\bar{x}_q = \frac{\Sigma(x_i^2)}{N}^{\overline{\;2}}$$

Esta notación se justifica si se tiene en cuenta que un exponente fraccionario equivale a un radical cuya cantidad subradical está elevada a la potencia indicada por el numerador del exponente fraccionario, y cuyo índice es el denominador de dicho exponente fraccionario.

Para la media cúbica:

Valores de la variable: x_i

Cubos de esos valores: x_i^3

Suma de esos cubos: $\Sigma(x_i^3)$

Cociente de la suma entre el número (N) de datos: $\dfrac{\Sigma(x_i^3)}{N}$

N

Raíz cúbica del cociente: $\sqrt[3]{\dfrac{(\Sigma x_i^3)}{N}}$

Fórmula para la media cúbica($\bar{x}_c$)

$$\bar{x}_c = \sqrt[3]{\frac{\Sigma(x_i^3)}{N}}$$

Notación alterna:

$$x_c = \left(\frac{\Sigma(x^3_i)}{N} \right)^{\frac{1}{3}}$$

Para la media armónica:

Valores de la variable:	$x_1, x_2 \ldots x_n$	x_i
Recíprocos de esos valores:	$\frac{1}{x_1}, \frac{1}{x_2} \ldots \frac{1}{x_n}$	$\frac{1}{x_i}$
Suma de los recíprocos:		$\Sigma\frac{1}{x_i}$
Cociente de la suma entre el número (N) de datos:		$\frac{\Sigma\frac{1}{x_i}}{N}$
Recíproco del cociente:		$\frac{1}{\frac{\Sigma\frac{1}{x_i}}{N}}$

La lectura de esta última fórmula nos indica que la media armónica es igual al recíproco (fracción principal) de la media aritmética (cociente de suma entre número) de los recíprocos (fracción terciaria precedida del operador sigma) de los datos.

El recíproco —resultado de dividir la unidad entre el valor cuyo recíproco se busca— se puede representar mediante un exponente "menos uno" que afecte al número cuyo recíproco se busca. Conforme a esta notación, se tendría para la media armónica:

Valores de la variable:	$x_1, x_2 \ldots x_n$	x_i
Recíprocos:	$(x_1)^{-1}, (x_2)^{-1}, \ldots (x_n)^{-1}$	$(x_i)^{-1}$
Suma de los recíprocos:		$\Sigma/(x_i^{-1})$
Cociente de la suma entre el número (N) de datos:		$\frac{\Sigma(x_i^{-1})}{N}$
Recíproco del cociente:		$\frac{\Sigma(x_i^{-1})^{-1}}{N}$

Fórmula de la media armónica $(\bar{x}_a)$

$$\bar{x}_a = \frac{\Sigma(x_i^{-1})^{-1}}{N}$$

Para la media geométrica:

Valores de la variable:	$x_1, x_2 \ldots x_n$	x_i
Logaritmos de esos valores:	$\log x_1, \log x_2 \ldots \log x_n$	$\log x_i$
Suma de los logaritmos:		$\Sigma(\log x_i)$
Cociente de la suma entre el número (n) de datos:		$\frac{\Sigma(\log x_i)}{N}$
Antilogaritmo del cociente:		$\text{antilog}\,\frac{\Sigma(\log x_i)}{N}$

Fórmula de la media geométrica ($\bar{x}_g$)

$$\bar{x}_g = \text{antilog}\,\frac{\Sigma(\log x_i)}{N}$$

Fórmula de la media geométrica en números naturales:

$$\bar{x}_g = \sqrt[N]{\Pi(x_i)}$$

OPERACIONES PARA LA OBTENCIÓN DE MEDIAS ESTADÍSTICAS

Media	*Operaciones que (1) deben ejecutarse con las variables*	*Operaciones (inversas) (2) que deben ejecutarse con la suma de los resultados (1) divididos entre N*
Aritmética ($\bar{x}$)	Ninguna	Ninguna
Cuadrática ($\bar{x}_q$)	Elevarlos al cuadrado	Extraerle la raíz cuadrada*
Cúbica ($\bar{x}_c$)	Elevarlos al cubo	Extraerle la raíz cúbica*
Armónica ($\bar{x}_a$)	Buscar sus recíprocos	Buscar su recíproco*
Geométrica ($\bar{x}_g$)	Buscar sus logaritmos	Buscar su antilogaritmo*

* En realidad, cada una de estas expresiones equivalen a: "Buscar el número del cual es cuadrado", "buscar el número del cual es cubo", "buscar el número del cual es recíproco", "buscar el número del cual es logaritmo".

En efecto, la suma de los logaritmos de x_i representa el producto de las x_i (de ahí la substitución del operador Σ por el operador Π); la división por N equivale a la extracción de la raíz enésima.

PONDERACIÓN DE LAS MEDIAS

Conforme ha quedado asentado anteriormente, ponderar es una de las operaciones elementales fundamentales de la Estadística. Hay necesidad de ponderar siempre que los diferentes datos de una distribución tienen distinta importancia relativa dentro del conjunto, o sea, siempre que unos pesan o son más densos que otros en la distribución.

El peso o la densidad estadística de cada uno de los datos de una distribución depende muy principalmente del número de veces que aparece el dato; o sea, que la densidad estadística de cada dato está expresada por su "frecuencia" (o sea, por el número de veces que dicho dato aparece).

Si conforme ya se ha dicho "ponderar" equivale fundamentalmente a multiplicar por un factor de ponderación, la ponderación de los datos de una distribución para el cálculo de una media ponderada equivaldrá a multiplicar cada uno de los valores *distintos* de la variable por la frecuencia correspondiente (o sea, por el número de veces que cada uno de esos valores diferentes se repite en la serie).

De acuerdo con esto, si x_1, x_2, $x_3 \ldots x_n$ no nos representan sino aquellos valores de la variable que son diferentes entre sí y el valor que corresponde a x_1 se repite f_1 veces, el de x_2 se repite $f_2 \ldots$ etc., en el conjunto de datos con que se trabaja, la ponderación nos obligará a multiplicar cada x_i por su frecuencia (f_i) correspondiente.

De este modo, la obtención de medias ponderadas nos obliga a seguir los mismos pasos a los que debíamos sujetarnos para la obtención de medias estadísticas no ponderadas, pero asimismo, nos obliga a incluir un paso adicional que estará comprendido entre el primero y el segundo del proceso de obtención general de la media y que modificará nuestra esquematización de dicho proceso en la forma siguiente:

Media estadística ponderada es el valor representativo de una distribución que toma en consideración la importancia relativa de cada dato dentro del conjunto, y que se obtiene si:

1º—se sujeta a la variable a cierta operación, se multiplican los valores obtenidos por sus frecuencias correspondientes (ponderación)
2º—se suman los productos así obtenidos,
3º—se divide la suma entre el número de datos, y

4º—se somete el cociente a la operación inversa de aquella a la que se sujetara a la variable.

En este caso, cabe aclarar que el número de datos es igual a la suma de las frecuencias.

Media aritmética ponderada:

$$\bar{x} = \frac{\Sigma(x_i f_i)}{\Sigma f_i}$$

El paréntesis —innecesario— que incluye a la x_i y a la f_i enfatiza el que se trata de la suma de los productos de los valores de la variable (x_i) por su frecuencia correspondiente (f_i) y no del producto de la suma de los valores de la variable $\Sigma(x_i)$ por la suma de los valores de las frecuencias $\Sigma(f_i)$ (lo cual, por otra parte, no podría pensarse, puesto que no existe un operador Σ delante de f_i), ni tampoco la suma de los valores de la variable $\Sigma(x_i)$ por las diferentes frecuencias (lo cual, de otro lado, no tendría sentido, ya que se obtendrían tantos productos como frecuencias, los cuales no estarían ligados por ninguna operación; en efecto $\Sigma(x_i)f_i$ no tendría sentido a menos que antes de la primera Σ se antepusiera otro operador cualquiera; otro operador Σ, por ejemplo, en cuyo caso se trataría de sumar todos los valores de la variable, multiplicar su suma por cada una de las frecuencias y en seguida adicionar esos productos). En todo caso, la fórmula de la media aritmética ponderada indica que:

1º—cada uno de los valores de la variable (ponderación) se multiplique por su frecuencia,
2º—se sumen los productos,
3º—se divida la suma entre la suma de las frecuencias (Σf_i) y
4º—se tome ese cociente como media aritmética ponderada de la *serie de frecuencias.*

Análogamente, para la media cuadrática, ponderada:

$$\bar{x}_q = \left[\frac{\Sigma(x_i^2)f_i}{\Sigma f_i}\right]^{\frac{1}{2}}$$

Para la media cúbica ponderada, se tendr:

$$\bar{x}_c = \left[\frac{\Sigma(x_i^3)f_i}{\Sigma f_i}\right]^{\frac{1}{3}}$$

Para la media armónica ponderada:

$$\bar{x}_a = \left[\frac{\Sigma (x_i^{-1}) f_i}{\Sigma f_i} \right]^{-1}$$

Para la media geométrica ponderada:

$$\bar{x}_g = \text{Alog} \frac{\Sigma (\log x_i) f_i}{\Sigma f_i}$$

MEDIAS PONDERADAS EN SERIES DE CLASES Y FRECUENCIAS

La diferencia fundamental que muestran las series de clases y frecuencias en cuanto se les contrasta con las series sencillas de frecuencias es la forma condensada de presentación de sus datos, la falta de individuación de sus valores en cuanto los mismos se encuentran subsumidos en una clase caracterizada por sus límites extremos pero con respecto a la cual no se dice cual sea la distribución que asumen los valores que la misma comprende.

De ahí que para trabajar con las series de clases y frecuencias se necesite tomar un valor como representativo de cada una de las clases y que generalmente se elija como tal valor representativo el llamado "punto medio" de cada clase.

En cuanto el "punto medio" o la "marca de clase" se toma como representativo de los valores contenidos en la clase de que se trate, se está procediendo sobre el supuesto que consiste en considerar como iguales al punto medio a todos aquellos valores de la variable que no se aparten (en más o en menos) del valor de dicho punto medio en más de cierta cantidad (la distancia del punto medio a los límites inferior y superior de la clase correspondiente). Así, por ejemplo, en una clase cuyos límites inferior y superior son 15 y 19 (incluído el 19), el punto medio o marca de clase 17 se toma como representativo de todos los valores que no sean inferiores a 15 ni superiores a 19 en la distribución; o sea, que si en dicha distribución existen: un 15, tres 16, dos 17, un 18 y cuatro 19, a todos ellos se les considerará iguales a 17; o sea que habrá (uno + tres + dos + uno + cuatro) once 17 en la distribución; o mejor aún que trabajaremos *como si hubiera* once diecisietes. De este modo, 17, marca de clase o punto medio de una de las clases de la distribución, (m_i) está en lugar de un conjunto de valores de la variable de la distribución primitiva (en lugar de x_1, x_2, x_3, x_4, x_5); por otra parte, la frecuencia asociada a m_1 será igual a la suma de las frecuencias asociadas a x_1, x_2, x_3, x_4, x_5, valores de la variable en cuyo lugar está m_i.

De acuerdo con todo lo anterior, si las x_i están representadas por las m_i correspondientes, y las frecuencias asociadas a las x_i sumadas equivalen a las frecuencias asociadas a las m_i, podremos substituir a las x_i por las m_i en las fórmulas correspondientes, y a las f_i por las F_i; de este modo, en el cálculo de las medias ponderadas en series de clases y frecuencias tendremos:

Media aritmética ponderada en series de clases y frecuencias:

$$\bar{x} = \frac{\Sigma(m_i F_i)}{\Sigma F_i}$$

Media cuadrática ponderada en series de clases y frecuencias:

$$\bar{x}_q = \frac{\Sigma(m_i^2 \mathrm{F}_i)^{\frac{1}{2}}}{\Sigma F_i}$$

Media cúbica ponderada en series de clases y frecuencias:

$$\bar{x}_c = \frac{\Sigma(m_i^3 F_i)^{\frac{1}{3}}}{\Sigma F_i}$$

Media armónica ponderada en series de clases y frecuencias:

$$\bar{x}_a = \frac{\Sigma(m_i^{-1} F_i)^{-1}}{\Sigma F_i}$$

Media geométrica ponderada en series de clases y frecuencias:

$$\bar{x}_g = \text{Antilog}\,\frac{\Sigma(\log m_1 F_i)}{\Sigma F_i}$$

En este último caso es pertinente aclarar que primero se buscarán los logaritmos de los puntos medios ($\log m_1$) y luego se multiplicarán los valores obtenidos por las frecuencias para después sumar los productos, dividirlos entre la suma de las frecuencias y buscar el antilogaritmo del cociente, y *no* se multiplicarán primero los puntos medios por las frecuencias para en seguida buscar los logaritmos del producto, etc.

Teniendo en cuenta la equivalencia entre las operaciones con logaritmos y las que se realizan con números naturales, la última fórmula equivale a:

$$\bar{x}_g = \sqrt[\mathrm{N}]{\Pi\,(m_i)^{f_i}}$$

Es decir, que, según esta fórmula, para calcular la media geométrica se deberá:

1º—elevar cada punto medio a una potencia igual a la frecuencia asociada con dicho punto medio,

2º—multiplicar cada uno de los resultados obtenidos por todos los restantes,

3º—extraer la raíz enésima (o sea, la raíz Σf_i) del producto.

DOS PROPIEDADES DE LA MEDIA ARITMÉTICA ÚTILES PARA LA SIMPLIFICACIÓN DEL CÁLCULO

Primera Propiedad de la Media Aritmética.—Cuando a los diferentes valores de una variable se les resta (/ se les suma) una cantidad constante, la media aritmética de los nuevos valores resultará igual a la media aritmética de los antiguos valores disminuída (/ aumentada) en la cantidad constante que se restó (/ sumó) a los diferentes valores de la variable.

Lenguaje Aritmético (Ejemplo ilustrativo de la primera propiedad):

Sean siete casos para cada uno de los cuales un rasgo determinado tiene las magnitudes de 2, 2, 3, 3, 3, 5, 5. Si calculamos la media aritmética de estos valores dividiendo la suma de $2 + 2 + 3 + 3 + 3 + 5 + 5 = 23$ entre el número de casos (7) obtendremos 3.28.

Si, en seguida, a cada uno de los valores de la serie anterior (2, 2, 3, 3, 3, 5, 5) le restamos la cantidad constante 3, obtendremos una nueva serie formada por los valores (—1,— 1, 0, 0, 0, 2, 2) ya que 2 menos 3 es —1, 3 menos 3 es cero y 5 menos 3 es 2. Si calculamos la media aritmética de estos nuevos valores dividiendo la suma de $(-1) + (-1) + 0 + 0 + 0 + 2 + 2 = 2$ entre el número de casos (7) obtendremos 0.28 como nueva media.

Si comparamos la nueva media (0.28) con la antigua (3.28) veremos que: la nueva media (0.28) es igual a la antigua (3.28) disminuida en la cantidad constante (3) que restamos de cada uno de los valores a partir de los cuales calculamos la media.

Traducción algebráica.—Si se tiene un rasgo de magnitud variable que adquiere los valores x_1, x_2, x_3, x_4, x_5, x_6, x_7 y se busca su media aritmética ésta se obtendrá sumando dichos valores ($x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 = \Sigma x_i$) y dividiendo la suma entre el número de datos (7 en este caso, o N en general). O sea, que la media aritmética será: $\frac{\Sigma x_i}{N}$

Si cada uno de los valores de la variable (x_1, x_2 . . .) restamos una cantidad constante x', tendremos una nueva serie formada por los valores $x_1 - x'$, $x_2 - x'$, $x_3 - x'$, $x_4 - x'$, $x_5 - x'$, $x_6 - x'$, $x_7 - x'$. Para obtener la media aritmética, sumaremos estos valores $(x_1 - x') + (x_2 - x') + (x_3 - x') + (x_4 - x') + (x_5 - x') + (x_6 - x') + (x_7 - x') = \Sigma(x_i - x')$, y, en seguida, dividiremos dicha suma entre el número de casos: $\frac{\Sigma(x_i - x')}{N}$

Conforme a lo dicho en lenguaje aritmético, la nueva media aritmética $\frac{\Sigma(x_i - x')}{N}$ es igual a la antigua media aritmética $\frac{\Sigma x_i}{N}$ disminuída en la cantidad constante x'.

$$\frac{\Sigma(x_i - x')}{N} - \frac{\Sigma x_i}{N} - x_i$$

Corolario:

$$\frac{\Sigma x_i}{N} = \frac{\Sigma(x_i - x')}{N} + x'$$

$$\bar{x} = \bar{x}_I + x'$$

$$\bar{x} = x' + \bar{x}_I$$

Es decir que:

La media aritmética de los valores originarios de una variable se puede obtener si a los valores originarios se les resta una cantidad constante (x'), se calcula la media aritmética de los valores obtenidos y a dicha media ($\bar{x}_I$) se le agrega la cantidad constante restada de los valores de la variable.

Pensamiento Algebraico (Demostración del principio).

$$\bar{x} = \frac{\Sigma x_i}{N}$$

Si de las x_i se resta una constante x', la nueva media aritmética ($\bar{x}_I$) será:

$$\bar{x}_I = \frac{\Sigma(x_i - x')}{N}$$

Si se tiene en cuenta que la suma de una serie de sumas (o restas) es igual a la suma (o resta) de las sumas de los sumandos (o de los minuendos y substraendos):

$$\Sigma(x_i - x') = \Sigma x_i - \Sigma x'$$

Por tanto, si se dividen los dos miembros entre N:

$$\bar{x}_I = \frac{\Sigma x_i - \Sigma x'}{N}$$

Como la fracción del segundo miembro puede descomponerse en dos fracciones que tengan el mismo denominador N y estén unidas por el signo menos:

$$\bar{x}_I = \frac{\Sigma x_i}{N} - \frac{\Sigma x'}{N}$$

Por ser x' una constante y ser la suma de una serie de cantidades constantes igual al producto de la cantidad constante por el número de veces en que aparece (en este caso, N), tendremos:

$$\bar{x}_I = \frac{\Sigma x_i}{N} - \frac{Nx'}{N}$$

En la segunda fracción, por figurar N en el numerador y en el denominador, se reducen a la unidad, quedando, por tanto:

$$\bar{x}_I = \frac{\Sigma x_i}{N} - x'$$

Pero como la suma de los valores de la variable (Σx_i) dividida entre el número de casos es la media aritmética originaria:

$$\bar{x}_I = \bar{x} - x'$$

En consecuencia:

$$\bar{x} = \bar{x}_I + x'$$

$$\bar{x} = x' + \bar{x}_I$$

El principio, no obstante haberse demostrado para el caso de series sencillas, es válido para series de frecuencias y para series de clases y frecuencias. A x' valor constante que se resta de los de la variable para obtener valores más fáciles de manejar se le conoce como "media arbitraria" o, mejor aún "media adivinada" porque, en efecto, conviene escoger un valor que se suponga cercano a la media (de este modo los residuos serán menores que en caso de elegir otro valor; en efecto, de coincidir la media adivinada con la media aritmética verdadera, los residuos serían mínimos, y su suma igual a cero). A $\bar{x}_I$, media aritmética de los residuos se le conoce como "primer momento con respecto a la media arbitraria o a la media adivinada (x')".

De este modo, puede decirse que:

La media aritmética de una serie es igual a la media adivinada arbitrariamente más el primer momento (o media aritmética de los residuos de los valores de la variable) con respecto a dicha media.

Aplicación del primer principio a series de frecuencias y a series de clases y frecuencias.—Hemos dicho en líneas anteriores que el principio demostrado se aplica a todo tipo de series sean estas simples, sean de frecuencias o de clases y frecuencias; en efecto, en las series de frecuencias, puede elegirse como media arbitraria o adivinada cualquier valor, pero, de preferencia, uno de los valores de la serie; en cuanto a la media aritmética de los residuos (o sea, en cuanto al primer momento) hay que recordar que si la media aritmética de los datos originarios ha de ser ponderada, será necesario que también sea ponderada la media aritmética de los residuos; o sea, que para calcular dicha media aritmética será necesario multiplicar cada residuo por la frecuencia correspondiente, sumar dichos productos y dividirlos por la suma de las frecuencias; es decir, que en series de frecuencias:

$$\bar{x}_I = \frac{\Sigma[(x_i - x')]f_i}{\Sigma f_i}$$

O sea que:

En series sencillas:

El primer momento con respecto a la media arbitraria se calcula:

1º—restando de cada valor de la variable el valor de la media arbitraria,

2º—multiplicando esos residuos (o "desviaciones") por la frecuencia correspondiente,

3º—sumando dichos residuos (o "desviaciones"),

4º—dividiendo la suma entre el número de casos.

El primer momento con respecto a la media arbitraria en series de frecuencias se calcula:

1º—restando de cada valor de la variable el valor de la media arbitraria. (Ponderación) multiplicando esos residuos o "desviaciones" por la frecuencia correspondiente,

2º—sumando dichos productos,

3º—dividiendo la suma entre la suma de las frecuencias.

En el caso de series de clases y frecuencias, en cuanto los puntos medios o marcas de clase son representativos de los datos comprendidos por cada clase, es conveniente tomar como media arbitraria o adivinada, uno de los puntos medios (m'). En cuanto al primer momento, será como en la serie de frecuencias (esta serie es de clases *y frecuencias*) la media aritmética ponderada de los residuos (o "desviaciones"); en este caso, dichos residuos serán los valores que resulten de restar de cada punto medio, la media arbitraria o adivinada elegida.

Conforme a esto, el primer momento en series de clases y frecuencias será:

$$\bar{x}_I = \frac{\Sigma(m_i - m')F_i}{\Sigma F_i}$$

O sea, que el primer momento en series de clases y frecuencias se calculará:

1º—Restando de cada punto medio la media arbitraria (uno de los puntos medios).
(Ponderación).—multiplicando cada uno de estos residuos (o "desviaciones") por la frecuencia asociada a la clase correspondiente.
2º—Sumando dichos productos, y
3º—Dividiendo dicha suma entre la suma de las frecuencias de clase.

El segundo principio relativo a la media aritmética nos obligará a precisar una distinción entre el primer momento calculado en esta forma para las series de clases y frecuencias, y al cual llamaremos "primer momento con respecto a la media arbitraria en unidades originarias", de otro al cual llamaremos "segundo momento con respecto a la media arbitraria en unidades del intervalo".

Segunda Propiedad de la Media Aritmética.—Cuando cada uno de los valores de una variable se divide entre una cantidad constante, la media aritmética de los cocientes resultará igual a la media aritmética de los valores originarios dividida entre la cantidad constante que sirvió de divisor de éstos.

Demostración.—Supongamos que calculamos la media aritmética de los valores originarios:

$$\bar{x} = \frac{\Sigma x_i}{N}$$

Si en seguida dividimos cada x_i entre un valor constante que (con vistas a su futura utilización) llamaremos i, obtendremos una media aritmética distinta:

$$\bar{x}_{II} = \frac{\Sigma\left(\frac{x_i}{i}\right)}{N}$$

Pero, si tenemos en cuenta que un cociente o una fracción (x_i/i) puede escribirse como el producto del numerador tomado como entero por el recíproco del denominador (o bien por una fracción que tenga por numerador la unidad y por denominador el mismo de la fracción: $1/i$), podremos escribir:

$$\bar{x}_{II} = \frac{\Sigma\left(x_i \frac{1}{i}\right)}{N}$$

Esto pone de manifiesto que en el numerador se trata de la suma de los productos de una variable x_i por una constante $1/i$, y como una suma de productos de este tipo es igual a la constante multiplicada por la suma de la variable (o bien, conforme a la otra expresión de la regla, como una constante puede salir o entrar al operador Σ sin modificación), tendremos:

$$\bar{x}_{II} = \frac{\frac{1}{i} \Sigma x_i}{N}$$

Pero $1/i$ puede pasar al denominador de la fracción principal multiplicando; o bien puede colocarse como factor de toda la fracción:

$$\bar{x}_{II} = \frac{1}{i}\left(\frac{\Sigma x_i}{N}\right)$$

Que es lo que se quería demostrar. En efecto, la expresión de dentro del paréntesis es la media originaria y, por lo mismo, se puede escribir:

$$\bar{x}_{II} = \frac{1}{i}\bar{x} = \frac{\bar{x}}{i}$$

Corolario:

$$\bar{x} = \bar{x}_{II}\, i$$

El segundo principio, como el primero, es aplicable tanto a series sencillas como a series de frecuencias y como a series de clases y frecuencias. Conforme a la última expresión, puede decirse en general que:

La media aritmética de una serie es igual a la media aritmética de los cocientes de los valores de dicha serie entre una cantidad constante, multiplicada por la cantidad constante.

APLICACIÓN CONJUNTA DE LOS DOS PRINCIPIOS A LA DETERMINACIÓN DE LA MEDIA ARITMÉTICA DE UNA SERIE

La aplicación conjunta de los dos principios anteriores tiene utilidad particularmente en cuanto se trata de encontrar la media aritmética de una serie de clases y frecuencias; en efecto:

Si de los valores de una serie se resta una cantidad constante (i.e. se calculan las desviaciones con respecto a una media arbitraria) y los residuos (o desviaciones) se dividen entre otra cantidad también constante (que en caso de ser el intervalo de clase convierte a las desviaciones en unidades de intervalo), la media aritmética que se obtenga de tales valores tendrá que multiplicarse por el valor de la cantidad constante divisora (por el intervalo) y a dicho producto deberá agregársele la cantidad constante (media arbitraria) restada a fin de obtener la media aritmética de los datos originarios.

Supongamos, en efecto, una serie de clases y frecuencias cuyos puntos medios sean $m_1, m_2 \ldots m_n$; si restamos de cada punto medio una cantidad constante (que por conveniencia puede ser uno de los mismos puntos medios) m', la nueva serie será $(m_1 - m')$, $(m_2 - m') \ldots (m_n - m')$ cuya media aritmética será igual a la de los datos originarios disminuída en m' unidades; si los nuevos valores de la serie los dividimos entre una constante (que por conveniencia puede ser el intervalo de clase) i, la segunda nueva serie será: $\frac{m_1 - m'}{i}, \frac{m_2 - m'}{i} \ldots \frac{m_3 - m'}{i} \ldots \frac{m_n - m'}{i}$ cuya media aritmética será menor que la media aritmética de la serie anterior en i unidades. O sea, que si a partir de la media aritmética de esta última serie queremos obtener la de la serie intermedia, tendremos que multiplicarla por i y si, una vez calculada así la de la serie intermedia buscamos la de la serie originaria, tendremos que agregarle m'. Simbólicamente, tendremos:

$$\bar{x} = \left[\frac{\Sigma \left(\frac{m_i - m'}{i} \right) f}{\Sigma f_i} \right] i + m'$$

1.—Media aritmética de las desviaciones con respecto a la media arbitraria en unidades del intervalo.

2.—Media aritmética de las desviaciones con respecto a la media arbitraria en unidades originales.

3.—Media aritmética.

O sea:

$$x = m' + \left[\frac{\Sigma\left(\frac{m_i - m'}{i}\right) f_i}{\Sigma f_i} \right] i$$

Es decir que:

Para determinar la media aritmética de una serie $(\bar{x})$

1º—se determina o elige una media arbitraria o adivinada (m') entre los puntos medios (uno de los m_i),

2º—se determinan las desviaciones de cada punto medio (m_i) con respecto a la media adivinada o elegida $(m_i - m')$ en términos del intervalo $\frac{m_i - m'}{i}$; o sea, más sencillamente, se determina cuántos intervalos dista cada punto medio del punto medio elegido como media arbitraria (de ahí que a la clase de la media arbitraria le corresponda una desviación 0, a las anteriores, desviaciones negativas (—1, —2, —3 . . . —n) y a las posteriores, desviaciones positivas (1, 2, 3 . . . n)

3º—se multiplican las desviaciones de cada clase por las frecuencias correspondientes y se dividen entre la suma de las frecuencias,

4º—se multiplica el cociente ("primer momento con respecto a la media arbitraria en unidades del intervalo") por el valor del intervalo,

5º—se suma al producto anterior la media elegida, adivinada o arbitraria.

Mecanización del procedimiento.—Para calcular la media aritmética de una serie:

1º—Se elige uno de los puntos medios de las clases como media arbitraria (media arbitraria m').

2º—Frente a la clase de la media arbitraria se coloca cero, hacia arriba —1, —2, —3, etc. y hacia abajo 1, 2, 3 . . . etc., por cada clase que nos alejemos de la que contiene a la media arbitraria (desviaciones con respecto a la media arbitraria en unidades del intervalo).

3º—Se multiplica cada una de las desviaciones de clase por la correspondiente frecuencia de clase.

4º—Se suman los productos de las frecuencias por las desviaciones.

5º—Se divide la suma de los productos de las frecuencias por las desviaciones entre la suma de las frecuencias (primer momento).

6º—Se multiplica el primer momento por el valor del intervalo.

7º—Se suma al producto del primer momento por el intervalo el valor de la media arbitraria.

DESVIACIONES Y DISPERSIÓN

Desviación estadística es, fundamentalmente, la diferencia (positiva o negativa) de un valor determinado con respecto a un valor constante o central tomado como término de comparación.

En páginas anteriores hemos tenido ocasión de hablar, por ejemplo, de desviaciones con respecto a una media arbitraria en el sentido de valores que resultan de restar de los datos la media arbitraria, elegida o adivinada; en estas páginas nos referiremos asimismo con frecuencia a desviaciones con respecto a la media aritmética o sea, a los valores que resultan de restar de los datos la media aritmética.

Frente a la desviación que es un carácter de cada uno de los datos de una distribución, la dispersión es una característica propia de la distribución.

La dispersión de una distribución indica la mayor o menor concentración de los datos en torno de un promedio central (principalmente en torno de la media) y, por consiguiente, indica asimismo el grado de representatividad del promedio.

La dispersión de una distribución se mide generalmente mediante el cálculo de promedios de desviaciones; más precisamente, mediante el cálculo de medias de las desviaciones.

En este lugar nos ocuparemos únicamente de dos de las medidas de dispersión más usuales: "la desviación media (aritmética)" y la "desviación media (cuadrática)".

Bastará con tener en cuenta los elementos formativos de estas denominaciones para saber de inmediato la forma de calcular estas medidas de dispersión.

Desviación media (aritmética) es la media aritmética de las desviaciones de los datos con respecto a la media aritmética de la distribución.

Si se tiene en cuenta que la suma de las desviaciones con respecto a la media aritmética es siempre cero, la desviación media aritmética sería siempre nula y, por lo mismo no tendría utilidad para comparar la mayor o menor concentración (la menor o la mayor dispersión) de los datos de varias distribuciones con respecto a sus correspondientes medias, o para comparar la mayor o menor representatividad de dichas medias. De ahí que necesite precisarse la definición de desviación media aritmética en la forma siguiente:

Desviación media aritmética es la media aritmética *(de los valores absolutos)* de las desviaciones de los datos con respecto a la media aritmética.

En efecto, en cuanto se eliminan los signos de las desviaciones (es decir, en cuanto se toman los valores absolutos de las desviaciones) con respecto a la media aritmética, no hay posibilidad de anulación de los valores positivos por los negativos y viceversa y la desviación media aritmética sirve a los propósitos para los que fue creada: la comparación de dispersiones / concentraciones de las series y la estimación de la representatividad de los promedios.

Conforme a la definición la desviación media aritmética será:

$$\bar{d} = \frac{\Sigma(d_i)}{N} \qquad \text{Serie sencilla.}$$

$$\bar{d} = \frac{\Sigma(d_i)f_i}{\Sigma f_i} \qquad \text{Serie de frecuencias.}$$

O bien, si expresamos las desviaciones con respecto a la media aritmética en términos de los datos originarios o en términos de los puntos medios o marcas de clase, tendremos:

$$\bar{d} = \frac{\Sigma(x_i - \bar{x})}{N} \qquad \text{Serie sencilla.}$$

$$\bar{d} = \frac{\Sigma(x_i - \bar{x})f_i}{\Sigma f_i} \qquad \text{Serie de frecuencias.}$$

$$\bar{d} = \frac{\Sigma(m_i - \bar{x})F_i}{\Sigma F_i} \qquad \text{Serie de clases y frecuencias.}$$

Desviación media cuadrática es la media cuadrática de las desviaciones de los datos con respecto a la media aritmética.

$$\bar{d}_q = \sqrt{\frac{\Sigma(d^2)}{N}}$$

O bien:

$$\bar{d}_q = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} \qquad \text{Serie sencilla.}$$

$$\bar{d}_q = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 f_i}{\Sigma f_i}} \text{ Serie de frecuencias.}$$

$$\bar{d}_q = \sqrt{\frac{\Sigma(m_i - \bar{x})^2 F_i}{\Sigma F_i}} \text{ Serie de clases y frecuencias.}$$

A la desviación media cuadrática se le acostumbra representar por la letra griega σ (sigma minúscula) que es una constante para cada distribución (variable en cuanto se investiga la distribución de dichas desviaciones medias cuadráticas en las muestras de una distribución) y que no debe confundirse con el operador Σ (signa mayúscula) que no tiene valor numérico sino operativo.

DOS PROPIEDADES DE LA DESVIACIÓN MEDIA CUADRÁTICA ÚTILES PARA LA SIMPLIFICACIÓN DE SU CÁLCULO

Primera Propiedad.—Si de los valores originales de una distribución se resta una cantidad constante, la desviación media cuadrática de la nueva distribución será igual a la desviación media cuadrática de los valores originales.

Una aclaración que no huelga en este punto es la que consiste en precisar que, en cuanto se habla de "desviación media cuadrática de la nueva distribución" se trata de la desviación media cuadrática de los nuevos valores con respecto a la media aritmética de los mismos.

Supongamos, en efecto, una serie de valores x_i con media aritmética $\bar{x}$. La desviación media cuadrática para esta serie estará dada por:

$$\sigma = \bar{d}_q = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}}$$

Si de cada uno de los valores de esta serie (x_i) restamos una cantidad constante (x'), tendremos una nueva serie de valores (los valores $x_i - x'$). La media aritmética de estos nuevos valores será distinta de la media aritmética de los antiguos valores (representaremos a la nueva media aritmética por $\bar{x}_I$, como hemos hecho en otras ocasiones). Las desviaciones de los nuevos valores ($x_i - x'$) con respecto a esta nueva media aritmética estarán dadas por la diferencia entre cada uno de esos nuevos valores y el valor de la nueva media. Conforme a esto:

Valores originales	x_i
Cantidad constante por restar de ellos:	x'
Valores obtenidos de la resta:	$x_i - x'$
Media aritmética de los nuevos valores:	$\bar{x}_I = \dfrac{\Sigma(x_i - x')}{N}$
	$\bar{x}_I = \bar{x} - x'$
Desviaciones de los nuevos valores con respecto a la nueva media ($\bar{x}_I$):	$(x_i - x') - \bar{x}_I$

Cuadrados de las desviaciones: $[(x_i - x') - \bar{x}_I]^2$

Desviación cuadrática media de los

nuevos valores: $\sigma_I = \sqrt{\dfrac{\Sigma[(x_i - x') - \bar{x}_I]^2}{N}}$

Pero como la nueva media aritmética $\bar{x}_I$ es igual a la antigua media aritmética (la media aritmética de los valores originarios) menos la media arbitraria o cantidad constante restada de los valores de la serie originaria (*i.e.*: $\bar{x}_I = \bar{x} - x'$) si en el valor de la desviación media cuadrática de los valores resultantes (σ_I) substituimos $\bar{x}_I$ por su valor, tendremos:

$$\sigma_I = \sqrt{\frac{\Sigma[(x_i - x') - (\bar{x} - x')]^2}{N}}$$

Si se eliminan los paréntesis curvos, los términos del primero quedan incambiados (puesto que van precedidos de signo $+$), en cambio los del segundo deben cambiarse (por ir precedidos de signo $-$). De este modo, tendremos:

$$\sigma_I = \sqrt{\frac{\Sigma[x_i - x' - \bar{x} + x']^2}{N}}$$

Como en el paréntesis rectangular x' aparece sumando y restando ($-x'$ y $+ x'$) se anula, quedando reducida la expresión de dicho paréntesis a $[x_i - \bar{x}]$, o sea, que el valor de la nueva desviación media cuadrática será:

$$\sigma_I = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}}$$

Si se compara esta expresión con la correspondiente a la desviación media cuadrática de la serie originaria dada en primer término, podrá verse que los segundos miembros son idénticos; de lo que se concluye que también serán iguales los primeros miembros:

$$\sigma = \sigma_I$$

O sea, que, como se quería demostrar, la desviación media cuadrática de la serie obtenida al restar la constante x' de los valores originarios (σ_I) es igual a la desviación media cuadrática de la serie original.

La demostración anterior es válida para la desviación media cuadrática de series sencillas como las anteriores, series de frecuencias o series de clases y frecuencias.

Segunda Propiedad de la Desviación Media Cuadrática.—Si los valores de una serie se dividen entre una cantidad constante, la desviación cuadrática media de los nuevos valores será igual a la desviación media cuadrática de los valores originarios dividida (la *d.m.c.*) por la cantidad constante que sirvió de divisor de tales valores originales.

Supongamos que tenemos una serie de valores x_i, cuya media aritmética sea x; su desviación media cuadrática será:

$$\sigma = \bar{d}_q = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}}$$

Si dividimos cada uno de los valores de la serie entre la constante i, obtendremos una serie de nuevos valores $\frac{x_i}{i}$. La media aritmética $(\bar{x}_{II})$ de los nuevos valores (conforme a la segunda propiedad de la media aritmética mencionada en estas páginas) será igual a la media aritmética de los valores originales $(\bar{x})$ dividida por la cantidad constante (i); o sea, que la media aritmética será igual a $\frac{\bar{x}}{i}$: Conforme a esto:

Valores originales: x_i

Valores de la nueva serie (Cocientes de los originales entre la constante i): $\frac{x_i}{i}$

Media aritmética de la nueva serie: $\bar{x}_{II} = \frac{\bar{x}}{i}$

Desviaciones de los valores de la nueva serie $\left(\frac{x_i}{i}\right)$ con respecto a su media $(\bar{x}_{II})$:

$$\frac{x_i}{i} - \bar{x}_{II} = \frac{x_i}{i} - \frac{\bar{x}}{i} = \frac{x_i - \bar{x}}{i}$$

Desviación media cuadrática:

$$\sigma_{II} = \sqrt{\frac{\Sigma\left(\frac{x_i - \bar{x}}{i}\right)^2}{N}}$$

En efecto, para escribir esta expresión, se tomaron las desviaciones $\left(\frac{x_i - \bar{x}}{i}\right)$ de los nuevos valores con respecto a su media, se elevaron al cuadrado $\left(\frac{x_i - \bar{x}}{i}\right)^2$; se sumaron $\Sigma\left(\frac{x_i - \bar{x}}{i}\right)^2$ se dividieron entre el número de casos y se extrajo al cociente la raíz cuadrada, conforme pide la obtención de la desviación media cuadrática.

Si en la expresión para la desviación cuadrática media de los nuevos valores (σ_{II}) elevamos al cuadrado el quebrado de dentro del paréntesis, tendremos que elevar al cuadrado por separado el numerador y el denominador, o sea:

$$\left(\frac{x_i - \bar{x}}{i}\right)^2 = \frac{(x_i - \bar{x})^2}{i^2}$$

Al aplicar el operador Σ a la expresión resultante, bastaría con que lo aplicásemos al numerador, con lo cual obtendríamos:

$$\frac{\Sigma(x_i - \bar{x})^2}{i^2}$$

En la expresión para σ_{II} la expresión anterior es numerador de la fracción principal cuyo denominador es N, pero tener la fracción numeradora como denominador a N equivale a estar dividida por él, y dividir una fracción entre un entero (N) equivale a multiplicar su denominador (i^2) por el entero; es decir:

$$\frac{\frac{\Sigma(x_i - \bar{x})^2}{i^2}}{N} = \frac{\Sigma(x_i - \bar{x})^2}{i^2} \div N = \frac{\Sigma(x_i - \bar{x})^2}{Ni^2}$$

Cuando se substituye este valor en el de σ_{II} se obtiene

$$\sigma_{II} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{Ni^2}}$$

La expresión anterior también puede escribirse como la raíz cuadrada del producto de dos factores, uno de ellos formado por $\frac{\Sigma(x_i - \bar{x})^2}{N}$ y el otro por $\frac{1}{i^2}$:

$$\sigma_{II} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N} \cdot \frac{1}{i^2}}$$

Esta expresión a su vez equivale a un producto de radicales:

$$\sigma_{II} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} \cdot \sqrt{\frac{1}{i^2}}$$

Pero como en el segundo radical se trata de extraer la raíz cuadrada de un cuadrado, el resultado es la cantidad misma sin elevación al cuadrado ni extracción de la raíz correspondiente:

$$\sigma_{II} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} \cdot \frac{1}{i}$$

Pero multiplicar por $1/i$ equivale a dividir por i:

$$\sigma_{II} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} \div i$$

Si se compara el dividendo de esta expresión con la expresión que da el valor de la desviación media cuadrática de los datos originales, se verá que son iguales; o sea:

$$\sigma_{II} = \frac{\sigma}{i}$$

Esto es lo que se quería demostrar, o sea, que la desviación media cuadrática de los valores originales divididos entre unas constante i (σ_{II}) es igual a la desviación media cuadrática de los valores originales (σ) dividida entre la constante.

De esto se deduce que:

Para obtener la desviación media cuadrática de los valores originales a partir de la desviación media cuadrática de la serie obtenida mediante su división por una constante, bastará con multiplicar esta segunda desviación cuadrática media por el valor de la constante entre la cual se dividieron los valores originales:

$$\sigma = \sigma_{II} \cdot i$$

APLICACIÓN CONJUNTA DE LOS DOS PRINCIPIOS A LA OBTENCIÓN DE LA DESVIACIÓN MEDIA CUADRÁTICA DE UNA SERIE

Conforme se dijo con respecto a la media aritmética, puede decirse en forma análoga para la desviación media cuadrática que la aplicación conjunta de los dos principios anteriores es especialmente útil cuando se trata de calcular la desviación media cuadrática de una serie de clases y frecuencias.

En el caso de una serie de clases y frecuencias, de los puntos medios de las clases, representativos de los datos, conviene restar como cantidad constante una media elegida entre los mismos puntos medios. Dicha resta no altera el valor de la desviación cuadrática media que se calcule tomando los nuevos valores y refiriéndola a la media aritmética de esos nuevos valores; en cambio, dicha resta permitirá el más fácil manejo por ser menores numéricamente los valores que se utlizan en el cálculo.

De este modo, si se tienen unos ciertos valores m_i de una serie y se trata de encontrar su desviación media cuadrática, podrá formarse una nueva serie con los valores $m_i - m'$ resultantes de restar de cada m_i una de las m_i a la que llamamos m'; en seguida, deberá calcularse la media de estas $m_i - m'$ (si suponemos que se trata de una serie de clases y frecuencias se tendrán que multiplicar las $m_i - m'$ por las correspondientes frecuencias, sumar los productos y dividir entre la suma de las frecuencias); a continuación tendría que restarse la media obtenida así, de cada una de las $m_i - m'$ para obtener las desviaciones de las $m_i - m'$ con respecto a su media; tras ello se elevarían cada una de esas desviaciones al cuadrado, se multiplicarían por las frecuencias, se sumarían los productos, se dividirían entre el total de las frecuencias y al cociente se le extraería la raíz cuadrada que sería al mismo tiempo la desviación media cuadrática de la $m_i - m'$ y de la m_i originarias:

$$\sigma = \sqrt{\frac{\Sigma \left[(m_i - m') - \frac{\Sigma (m_i - m') f_i}{\Sigma f_i} \right]^2 f_i}{\sqrt{\Sigma f_i}}}$$

O, si con el fin de escribir más brevemente la anterior expresión ponemos: d' por $m_i - m'$ "desviación respecto de una media elegida"

$\bar{d}'$ por $\dfrac{\Sigma(m_i - m')f_i}{\Sigma f_i}$ "media aritmética de las desviaciones con respecto a la media elegida"

escribiremos:

$$\sigma = \sqrt{\frac{\Sigma[d - \bar{d}]^2 f_i}{\Sigma f_i}}$$

Si suponemos que cada uno de los valores de la segunda serie $m_i - m'$ (o d_i') se han dividido entre una constante i (el valor del intervalo para dar una nueva serie $\dfrac{m_i - m'}{i}$ (o $\dfrac{d_i'}{i}$) y tratamos de calcular la desviación media aritmética de estos últimos valores ($\dfrac{d_i'}{i}$), dicha desviación tomada con respecto a la media aritmética de los nuevos valores (con respecto a la media de las $\dfrac{d_i'}{i}$) será igual a la desviación media artimética de los valores originarios dividida por el intervalo.

Para calcular la desviación media aritmética de las d_i'/i, se necesitará calcular la media aritmética de dichas d_i'/i multiplicando dichos valores por las correspondientes frecuencias $d_i f_i/i$ y dividiendo la suma de dichos productos $\Sigma[d_i' f_i/i]$ entre la suma de las frecuencias Σf_i. Calculada la media, habrá que restarla de cada una de las d_i'/i para obtener las correspondientes desviaciones; en seguida, se necesitará multiplicar dichas desviaciones elevadas al cuadrado por la correspondiente frecuencia, sumar los productos, dividirlos entre la suma de las frecuencias y al cociente extraerle la raíz cuadrada, para obtener la desviación media cuadrática de las desviaciones de las d_i'/i con respecto a su media. Esta desviación media cuadrática de las d_i'/i multiplicada por el intervalo nos dará la desviación media cuadrática de las d_i' (o de la $m_i - m'$) y, por consiguiente la de la m_i originarias:

$$\sigma = i\sqrt{\Sigma\left(\frac{d_i'}{i} \quad \frac{\dfrac{\Sigma d_i'}{i} f_i}{\Sigma f_i}\right)^2 f_i \div \Sigma f_i}$$

O, si con propósitos de escritura más breve representamos:

$\dfrac{d_i'}{i}$ por D' "desviación con respecto a la media arbitraria en unidades del intervalo"

por $\overline{D'}$ "media de dichas desviaciones"

$$\sigma = i\sqrt{\frac{\Sigma(D_i' - \overline{D'})^2 f_i}{\Sigma f_i}}$$

Conforme a esta expresión, para obtener el valor de la desviación media cuadrática de una serie basta con:

1º—Restar de cada una de las desviaciones con respecto a la media arbitraria en unidades del intervalo (de cada $D_i' = 0$ para la clase de la media arbitraria $-1, -2\ldots$ hacia arriba $1, 2\ldots$ hacia abajo) la media aritmética *de dichas desviaciones.*

2º—Elevar al cuadrado esos residuos.

3º—Multiplicar esos cuadrados por las frecuencias correspondientes.

4º—Sumar esos productos.

5º—Dividir dicha suma entre la suma de las frecuencias

6º—Extraer la raíz cuadrada del cociente.

7º—Multiplicar la raíz cuadrada por el intervalo.

EXPRESIÓN DEL CÁLCULO DE LA DESVIACIÓN σ EN TÉRMINOS DE LOS DOS PRIMEROS MOMENTOS

En el procedimiento anterior pueden hacerse algunas modificaciones teniendo en cuenta, por ejemplo, que la "media aritmética de las desviaciones con respecto a una media arbitraria en unidades de intervalo" es lo que se conoce como "primer momento" de los valores originarios. De este modo, el primer paso del procedimiento quedaría concebido en los siguientes términos:

1º—Restar de cada una de las desviaciones con respecto a la media elegida en unidades de intervalo, el primer momento.

Si en la expresión algebráica substituímos la "media de las desviaciones con respecto a la media arbitraria en unidades de intervalo" ($\overline{D'}$) por el "primer momento con respecto a la media arbitraria" (que representaremos por μ_1) tendremos:

$$\sigma = \sqrt{\frac{\Sigma(D'_1 - \mu'_1)^2 f_i}{\Sigma f_i}}$$

El desarrollo del cuadrado del paréntesis nos daría:

$$(D'_i - \mu'_1)^2 = D'^2_1 - 2\,D'_i\mu'_1 + \mu'^2_1$$

El producto por las frecuencias es:

$$(D'^2_i - 2\,D'_i\mu'_1 + \mu'^2_1)f_i = D'^2_i f_i - 2\,D'_i f_i \mu'_1 + \mu'^2_1 f_i$$

Si se aplica el operador Σ, se tiene:

$$\Sigma(D'^2_i f_i - D'_i f_i \mu'_1 + \mu'^2_1 f_i) = \Sigma(D'^2_i f_i) - 2\,\Sigma\, D'_i f_i \mu'_1 + \mu'^2_1 \Sigma f_i$$

μ_1^2 se escribió fuera del operador Σ en el último término por tratarse de una constante. Si todos los términos de la expresión anterior se dividen entre Σf_i como exige la expresión que sirve de punto de partida, se tendrá:

$$= \frac{\Sigma(D_i'^2 f_i)}{\Sigma f_i} - 2\,\mu_1' \frac{\Sigma D_i' f_i}{\Sigma f_i} + \mu_1'^2 \frac{\Sigma f_i}{\Sigma f_i}$$

El último término queda reducido a μ_1^2 puesto que la suma de las frecuencias como numerador y como denominador de la misma fracción se reduce a la unidad. En el segundo término, a más del 2 y de la μ_1 hay una fracción que tiene por numerador la suma de los productos de las desviaciones por las frecuencias y por denominador la suma de las frecuencias y que, por lo mismo, es igual al primer momento (μ_1); de este modo, el segundo término se reduce a dos veces el producto de μ_1 por sí mismo, o sea, a dos veces el cuadrado del primer momento. Con esto, la expresión se convierte en:

$$= \frac{\Sigma(D_i'^2 f_i)}{\Sigma f_i} - 2\,\mu_1'^2 + \mu_1'^2 = \frac{\Sigma(D_i'^2 f_i)}{\Sigma f_i} - \mu_1'^2$$

El primer término de esta expresión (que es en realidad el subradical de la expresión primitiva) se asemeja al primer momento, a no ser porque D_i' está elevado al cuadrado en vez de estarlo a la primera potencia; a dicho término o sea a la suma de los productos de los cuadrados de las desviaciones... por las frecuencias entre la suma de las frecuencias, se le conoce como "segundo momento" y se le representa por μ_2; según esto, lo anterior se reduce a:

$$\mu'_2 - \mu_1^2$$

Si ésta que es en realidad la cantidad subradical la reintegramos al radical para obtener el valor de σ, tendremos:

$$\sigma = i\,\sqrt{\mu'_2 - \mu'^2_1}$$

Esta expresión permite mecanizar el cálculo de la desviación media cuadrática conforme a los pasos siguientes:

1º—Se calcula el primer momento de la distribución:

a.—multiplicando las desviaciones con respecto a la media arbitraria en unidades de intervalo (— *n*. . . — 2, — 1, 0, 1, 2, . . . *n*) por las frecuencias correspondientes

b.—se suman esos productos

c.—se dividen entre la suma de las frecuencias.

2º—Se eleva al cuadrado el primer momento.

3º—Se calcula el segundo momento:

a.—elevando al cuadrado las desviaciones. . . (— *n*, . . . 2, 1, 0, 1, 2, . . . *n*)

b.—multiplicando dichos cuadrados por las frecuencias

c.—sumando dichos productos

d.—dividiendo la suma entre la suma de las frecuencias.

4.—Se resta del segundo momento, el cuadrado del primer momento.

5º—Se extrae la raíz cuadrada de la diferencia.

6º—Se multiplica la raíz cuadrada por el intervalo.

MOMENTOS Y MEDIDAS DE ASIMETRÍA Y CURTOSIS

Estadísticamente, "momento" de una distribución es la media aritmética de los valores obtenidos cuando se elevan a determinada potencia las desviaciones de los valores de la distribución tomadas con respecto a un valor fijo llamado "centro de momentos".

Dentro de esta conceptuación de "momento" hay posibilidad de distinguir entre los diferentes momentos obtenibles:

a.—según la potencia a la que se eleven las desviaciones con respecto al centro de momentos, y
b.—según el punto que se elija como centro de momentos.

En puridad, este segundo criterio debería mencionarse en primer término y el primero en segundo, ya que primero se elige el centro de momentos y después la potencia a que se elevaran las desviaciones y, consiguientemente el orden del momento; sin embargo, aparecen en el orden en que se presentan ambos criterios de diferenciación en la noción de momento que encabeza estas líneas.

Según el punto o valor de la distribución que se elija como centro de momentos se distinguen por lo menos los siguientes tipos:

1º—momentos con respecto al origen de la distribución,
2º—momentos con respecto a la media aritmética de la distribución,
3º—momentos con respecto a una media elegida.

En realidad el primer tipo puede reducirse a una categoría del último; de este modo, conviene distinguir sobre todo entre:

A.—Momentos con respecto a la media aritmética que representaremos por la minúscula griega μ y
B.—Momentos con respecto a una media arbitraria, que representaremos por la minúscula griega mu prima (μ').

De acuerdo con el segundo criterio de clasificación de los momentos, distinguiremos entre:

1º—Momentos obtenidos mediante la elevación de las desviaciones a la primera potencia, o primeros momentos.
2º—Momentos obtenidos mediante la elevación de las desviaciones al cuadrado, o segundos momentos.
3º—Momentos obtenidos mediante la elevación de las desviaciones a la tercera potencia, o terceros momentos, y
4º—Momentos obtenidos mediante la elevación de las desviaciones a la cuarta potencia, o cuartos momentos.

Generalmente, no se requiere de más, pero, en general, puede hablarse de:

Momentos obtenidos mediante la elevación de las desviaciones a la potencia N o enésimos momentos.

El orden de los momentos —o sea la potencia a la que se elevaron las desviaciones para obtenerlos— se indica generalmente por medio de un subíndice que afecta el correspondiente signo del momento (μ o μ').

Si se emplean combinadamente los dos criterios anteriores, podremos hablar de:

1º, 2º, 3º, 4º, ... enésimo momento con respecto a la media aritmética: $\mu_1, \mu_2, \mu_3, \mu_4, \ldots \mu_n$.
1º, 2º, 3º, 4º, ... enésimo momento con respecto a una media elegida: $\mu_1', \mu_2', \mu_3', \mu_4', \ldots \mu_n'$.

Si atendemos a la noción que se dio al principio, según la cual el momento es, fundamentalmente, una media aritmética de cierta potencia de las desviaciones de una distribución y aplicamos las convenciones ya aceptadas en estas páginas, podremos escribir las siguientes igualdades:

$$\mu_1 = \overline{d} \qquad \mu_1' = \overline{d'}$$

$$\mu_2 = \overline{d^2} \qquad \mu_2' = \overline{d'^2}$$

$$\mu_3 = \overline{d^3} \qquad \mu_3' = \overline{d'^3}$$

$$\mu_4 = \overline{d^4} \qquad \mu_4' = \overline{d'^4}$$

Debe notarse que se trata de la media de las potencias y no de la media aritmética de los valores elevada a determinada potencia.

Puesto que la media aritmética de las desviaciones (de los cuadrados de las desviaciones, de los cubos de las desviaciones, etc.) es igual a la suma de las desviaciones (de los cuadrados de las desviaciones, de los cubos de las desviaciones, etc.) dividida (dicha suma) entre el número de casos, podremos transformar las expresiones anteriores en:

$$\mu_1 = \frac{\Sigma d}{N} \qquad \mu_1' = \frac{\Sigma d'}{N}$$

$$\mu_2 = \frac{\Sigma (d^2)}{N} \qquad \mu_2' = \frac{\Sigma (d'^2)}{N}$$

$$\mu_3 = \frac{\Sigma (d^3)}{N} \qquad \mu_3' = \frac{\Sigma (d'^3)}{N}$$

$$\mu_4 = \frac{\Sigma (d^4)}{N} \qquad \mu_4' = \frac{\Sigma (d'^4)}{N}$$

En estas expresiones, debe recordarse que

d_i representa la desviación de cada dato (x_i) con respecto a la media aritmética $(\bar{x}) = x_i - \bar{x}$.

d_i representa la desviación de cada dato (x_i) con respecto a la media arbitraria $(x') = x_i - x'$.

Cuando la serie no es sencilla, sino es de frecuencias, los momentos tendrán que ser medias *ponderadas* de las desviaciones o de las potencias de las desviaciones; es decir, que dichas desviaciones deberán aparecer multiplicadas por la frecuencia correspondiente; el número de datos estará dado por la suma de las frecuencias.

$$\mu_1 = \frac{\Sigma (d_i f_i)}{\Sigma f_i} \qquad \mu_1' = \frac{\Sigma (d_i' f_i)}{\Sigma f_i}$$

$$\mu_2 = \frac{\Sigma (d_i^2 f_i)}{\Sigma f_i} \qquad \mu_2' = \frac{\Sigma (d_i'^2 f_i)}{\Sigma f_i}$$

Etc.

Cuando se trata de series de clases y frecuencias, tratándose como se trata de series de frecuencias, los momentos tendrán que ser medias *ponderadas* como en el caso anterior, pero, en cuanto son series de clases, las desviaciones no serán desviaciones de cada dato con respecto al centro de momentos, sino desviaciones de cada punto medio (representativo de todos los datos de cada clase) con respecto a dicho centro de momentos; o sea que:

d_i representa la desviación de cada punto medio (m_i) con respecto a la media aritmética $(\bar{x})$ o sea, $m_i - \bar{x}$.

d_i' representa la desviación de cada punto medio (m_i) con respecto a la media arbitraria (x') o sea, $m_i - x'$.

Según esto, los procedimientos inmediatos para calcular un momento de orden r serían los siguientes:

A.—Momentos con respecto a la media aritmética.

1º—Calcúlese la media aritmética.

2º—Réstese la media aritmética de cada uno de los datos de la serie (si es sencilla o si es de frecuencias) o de cada uno de los puntos medios (si es serie de clases y frecuencias).

3º—Elévense las desviaciones así obtenidas a la potencia r (1ª, 2ª, 3ª, 4ª, . . . enésima) que indique el número de orden del momento (1er. momento, 2º momento, 3er. momento, 4º momento . . . enésimo momento).

(*Ponderación*).—Si se trata de series de frecuencias o de series de clases y frecuencias, multiplíquense esas potencias de las desviaciones por las frecuencias correspondientes.

4º—Súmense los valores así obtenidos.

5º—Divídase dicha suma entre el número de datos (o entre la suma de las frecuencias, en caso de tratarse de series de frecuencias o de series de clases y frecuencias).

B.—Momentos con respecto a la media arbitraria o elegida.

1º—Elíjase una media (de preferencia entre los datos en casos de series sencillas o de series de frecuencias; entre los puntos medios, en series de clases y frecuencias)

2º—Réstese esta media elegida de los datos o de los puntos medios (según el tipo de serie).

3º—Continúese el procedimiento como para *A.*

RELACIÓN ENTRE LOS DOS TIPOS DE MOMENTOS

Es conveniente determinar la relación que existe entre los momentos de una distribución con respecto a la media aritmética de dicha distribución, y los momentos de la misma con respecto a una media libremente elegida, ya que, en general, son más fáciles de calcular los momentos del segundo tipo que los del primero, pero se emplean más comúnmente (por permitir expresiones más sintéticas o breves) los momentos con respecto a la media aritmética, en cuanto se trata de calcular medidas de asimetría o de curtosis.

Si, como hemos venido haciendo, representamos por d_i la desviación con respecto a la media aritmética y por d_i' la desviación con respecto a la media arbitraria, podremos escribir (teniendo en cuenta la noción de desviación que hace equivalentes "desviación con respecto a" y "residuo que resulta de restar...") :

$$d_i = x_i - \bar{x}$$

$$d' = x_i - x'$$

Si al segundo miembro de la primera igualdad, le restamos y agregamos una misma cantidad x' dicho miembro no se altera y la igualdad subsiste:

$$d_i = x_i - \bar{x} - x' + x'$$

Si se agrupan términos en determinada forma, se obtiene:

$$d_i = (x_i - x') + (x' - \bar{x})$$

Pero, de acuerdo con la segunda expresión, $(x_i - x')$ es igual a d_i' desviación con respecto a la media elegida. Hecha la substitución, se obtiene:

$$d = d_1 + (x' - \bar{x})$$

Pero, según la expresión de la media aritmética en términos de la media arbitraria: $\bar{x} = x' + \mu_1'$, o sea, que si se despeja a μ_1': $\bar{x} - x' = \mu_1'$. Si cambiamos signos: $x' - \bar{x} = -\mu_1'$ valor que podemos substituir en el valor de d en el lugar de $x' - \bar{x}$, con lo cual se obtiene:

$$d = d' + (-\mu_1')$$

A partir de esta última expresión, podemos estudiar la relación entre los momentos; para el efecto, seguiremos los diversos pasos del proceso de obtención de los momentos con respecto a la media aritmética operando sobre dicha expresión. En primer término, será preciso elevar las desviaciones d a la enésima potencia:

$$d^n = [d' + (-\mu_1')]^n$$

Si sumamos las enésimas potencias (procedemos siempre sobre el supuesto más simple de las series sencillas, ya que de tratarse de series de frecuencias se necesitaría ponderar previamente), tendremos:

$$\Sigma(d^n) = \Sigma[d' + (-\mu_1')]^n$$

El desarrollo del binomio nos da:

$$\Sigma(d^n) = \Sigma \left[\binom{n}{0} d'^n \mu_1'^0 + \binom{n}{1} d'^{n-1}(-\mu_1')^1 + \binom{n}{2} d'^{n-2}(-\mu_1')^2 \ldots \right.$$
$$\left. \binom{n}{3} d'^{n-3}(-\mu_1')^3 \ldots + \binom{n}{n-1} d'^n(-\mu_1')^{n-1} + \binom{n}{n} d'^0(-\mu_1')^n \right]$$

Si dividimos entre N, a fin de obtener en el primer miembro de la ecuación el enésimo momento, tendremos:

$$\frac{\Sigma(d^n)}{N} = \frac{\Sigma \binom{n}{0} d'^n(-\mu'_1)^0}{N} + \frac{\Sigma \binom{n}{1} d'^{n-1}(-\mu'_1)^1}{N} +$$
$$\frac{\Sigma \binom{n}{2} d'^{n-2}(-\mu'_1)^2}{N} \ldots + \frac{\Sigma \binom{n}{n-1} d'^1(-\mu'_1)^{n-1}}{N} +$$
$$\frac{\Sigma \binom{n}{n} d'^0 (-\mu'_1)}{N}$$

El primer miembro de esta ecuación es el momento de orden n con respecto a la media aritmética. El segundo miembro de la ecuación está formada por la serie de términos del desarrollo del binomio; el tipo de dichos términos está formado como sigue, en el presente caso:

1.—Un operador Σ aplicado al producto de tres factores, tomada la suma de dichos productos como numerador. Los tres factores del numerador son:

a.—Un factor constante representado por las combinaciones de n objetos tomados de $r-1$ en $r-1$ (si r representa el número de orden del término de que se trate en la serie). Como factor constante, podrá salir del operador Σ sin alteración.

b.—Un factor variable formado por la potencia $n-(r-1)$ (r orden del término) de las desviaciones con respecto a la media arbitraria (d'). Como factor variable deberá quedar dentro del operador.

c.—Un factor constante formado por la potencia $r-1$ (complemento (para n de $n-(r-1)$ del primer momento con respecto a la media arbitraria tomado con signo negativo ($-\mu'_1$). Como factor constante puede salir del operador.

2.—Un denominador N representativo del número de datos que puede afectar al producto o a uno de los factores solamente.

Conforme a lo anterior, el tipo de término en la serie sería:

$$\frac{\Sigma \binom{n}{r-1} d'^{n-(r-1)} (-\mu'_1)^{r-1}}{N}$$

De acuerdo con las apostillas finales de cada uno de los incisos anteriores: $\binom{n}{r-1}$ como factor constante puede salir del operador Σ; $\mu'_1{}^{(r-1)}$ en cuanto también es constante puede asimismo salir sin alteración del operador. De este modo, el operador Σ sólo afectará al factor $d'^{n-(r-1)}$. Como N puede colocarse como denominador de todo el producto o de uno solo de los factores, lo colocaremos como denominador del factor variable, con lo que tendremos que la expresión anterior equivaldrá a:

$$\binom{n}{r-1} (-\mu'_1)^{(r-1)} \frac{\Sigma d'^{n-(r-1)}}{N}$$

Pero, en esta expresión, el último factor (la fracción) es la media aritmética de las potencias de orden $n-(r-1)$ de las desviaciones con respecto a la media arbitraria, y dicha media aritmética recibe, según dijimos al principio, el nombre de momento de orden $n-(r-1)$ con respecto a la media arbitraria o elegida, el cual se representa por $\mu_{n-(r-1)}$. De acuerdo con todo lo anterior, el término típico de la serie equivalente al momento de orden n con respecto a la media aritmética será el que se obtenga al substituir en la expresión anterior el último factor por su equivalente "momento de orden $n-(r-1)$ con respecto a la media arbitraria"; esto da:

$$\binom{n}{r-1} \mu'_{n-(r-1)} (-\mu'_1)^{(r-1)}$$

Si substituímos este término típico en la serie que forma el segundo miembro de la ecuación que nos da el equivalente de $\Sigma(d^n)/N$ o sea, el equivalente del momento de orden n con respecto a la media aritmética (μ_n), tendremos:

$$\mu_n = \Sigma \left[\binom{n}{r-1} \mu'_{n-(r-1)} (-\mu'_1)^{(r-1)} \right]$$

Esta expresión señala la equivalencia entre un momento de orden n con respecto a la media aritmética y los momentos con respecto a la media arbitraria.

EXPRESIÓN DE LOS CUATRO PRIMEROS MOMENTOS CON RESPECTO A LA MEDIA ARITMÉTICA EN TÉRMINOS DE LOS MOMENTOS CON RESPECTO A LA MEDIA ARBITRARIA

Si aplicamos la anterior expresión al cálculo de los cuatro primeros momentos con respecto a la media aritmética a partir de los momentos con respecto a la media arbitraria, tendremos:

$$\mu_1 = \binom{1}{0}\mu'_{1-0}(-\mu'_1)^0 + \binom{1}{1}\mu'_{1-1}(\mu'_1)^1 = 1 \times \mu'_1 \times 1 - 1 \times 1 \times \mu'_1$$

$$= \mu_1 = \mu'_1 - \mu'_1 = 0$$

O sea, que el primer momento con respecto a la media aritmética es nulo.

$$\mu_2 = \binom{2}{0}\mu'_{2-0}(-\mu'_1)^0 + \binom{2}{1}\mu'_{2-1}(\mu'_1)^1 + \binom{2}{2}\mu'_{2-2}(-\mu'_1)^2 =$$

$$= \mu'_2 - 2\,\mu'_1\,\mu'_1 + \mu'^2_1 = \mu'_2 - 2\,\mu'^2_1 + \mu'^2_1 = \mu'_2 - \mu'^2_1$$

Es decir, que el segundo momento con respecto a la media aritmética es igual al segundo momento con respecto a la media arbitraria menos el cuadrado del primer momento con respecto a la media arbitraria. De paso, si se compara la expresión anterior con la de la desviación media cuadrática, podrá observarse que el segundo momento con respecto a la media aritmética es igual al cuadrado de la desviación media cuadrática.

Para el tercer momento, se tendrá:

$$\mu_3 = \binom{3}{0}\mu'_{3-0}(-\mu'_1)^0 + \binom{3}{1}\mu'_{3-1}(-\mu'_1)^1 + \binom{3}{2}\mu'_{3-2}(-\mu'_1)^2 +$$

$$\binom{3}{3}\mu'_{3-3}(-\mu'_1)^3 = \mu'_3 - 3\,\mu'_2\mu'_1 + 3\,\mu'_1\mu'^2_1 - \mu'^3_1 = \mu'_3 -$$

$$3\,\mu'_2\mu'_1 + 3\mu'^3_{1\,3} - \mu'^3_1 = \mu'_3 - 3\mu'_2\mu'_1 + 2\,\mu'_1$$

O sea, que el tercer momento con respecto a la media aritmética es igual al tercero con respecto a la media arbitraria menos 3 veces el producto del segundo por el primero con respecto a dicha media arbitraria, más dos veces el cubo del primero con respecto a esa misma media arbitraria o elegida.

Para el cuarto momento, se obtendría, en forma semejante:

$$\mu_4 = \mu'_4 - 4\,\mu'_1\mu'_3 + 6\,\mu'^2_1\mu'_2 - 3\,\mu'^4_1$$

Es decir, que el cuarto momento con respecto a la media aritmética es igual al cuarto con respecto a la arbitraria menos cuatro veces el primero por el tercero, más 6 veces el cuadrado del primero por el segundo menos tres veces la cuarta potencia del primero con respecto a la media arbitraria.

A fin de mecanizar las relaciones entre los momentos con respecto a la media aritmética y los momentos con respecto a la media arbitraria o elegida, podría delinearse el siguiente procedimiento:

Constituir, para la formación del equivalente del momento de orden n con respecto a la media aritmética una serie de n *términos* (puesto que los dos últimos de los $n + 1$ del desarrollo binomial se reducen a uno).

De dichos términos:

A.—Los $n - 1$ primeros términos estarán formados por:

1º—Los coeficientes binomiales (obtenibles simplemente de un triángulo de Pascal o Tartaglia).

2º—La serie de momentos con respecto a la media elegida, en orden decreciente (o sea, según el orden decreciente de sus sub-índices) a partir del momento de orden n que corresponderá al primer término.

3º—La serie de las potencias del primer momento con respecto a la media elegida, a partir de la potencia 0 que corresponderá al primer término.

B.—El último o enésimo término estará formado por:

1º—El coeficiente $n - 1$, y

2º—La potencia n del primer momento con respecto a la media arbitraria (Ver demostración de esto último).

De todo lo anterior, se obtienen como fórmulas para el cálculo de los momentos con respecto a la media aritmética en términos de los momentos con respecto a la media arbitraria:

$$\mu_2 = \mu'_2 - \mu'^2_1$$

$$\mu_3 = \mu'_3 - 3\,\mu'_1\mu'_2 + 2\,\mu'^3_1$$

$$\mu_4 = \mu'_4 - 4\,\mu'_1\mu'_3 + 6\,\mu'^2_1\mu'_2 - 3\,\mu'^4_1$$

MOMENTOS EN SERIES DE CLASES Y FRECUENCIAS

En series de clases y frecuencias, las relaciones anteriores subsisten; sin embargo, si en vez de buscarse los momentos en unidades del intervalo, como ocurre en los casos anteriores, se requieren esos momentos en unidades originales, se necesitará multiplicar el momento de orden n por la enésima potencia del intervalo.

Corrección de Sheppard.—En el caso de series de clases y frecuencias, los momentos se calculan sobre el supuesto de que los valores de cada clase se agrupan en torno del punto medio de la clase; supuesto sujeto a error, en virtud del cual se necesitan corregir los momentos, de acuerdo con Sheppard, en la siguiente forma:

Primer Momento corregido: 0

Segundo Momento corregido: $\mu_2 - \dfrac{i}{12}$

Tercer Momento corregido: μ_3

Cuarto Momento corregido: $\mu_4 - \dfrac{\mu_2}{2} + \dfrac{7i}{240}$

O sea, que de los cuatro momentos, requieren corrección el segundo y el cuarto.

ASIMETRÍA

El grado de asimetría de una distribución puede medirse convenientemente, haciendo uso del tercer momento con respecto a la media aritmética; en efecto:

$$\mu_3 = \frac{\Sigma(x_i - \bar{x})^3}{N} \quad \text{o bien:} \quad \mu_3 = \frac{\Sigma(d_i^3)}{N}$$

Si la distribución es simétrica, para cada valor positivo de $(x_i - \bar{x})$ habrá un valor igual pero negativo en la distribución. El cubo de cada valor positivo de la desviación será un valor positivo igual, pero de signo distinto al cubo de la correspondiente desviación negativa. De este modo, si la mitad de las desviaciones de la distribución valen d_i y la otra mitad, $-d_i$, la mitad de las desviaciones tendrá por cubo $+d_i^3$ y la otra mitad $-d_i^3$; o sea, que la suma de todos los cubos será igual a cero, puesto que cada $+d_i^3$ se anulará con la correspondiente $-d_i^3$. Si dicha suma es cero, la fracción que tiene a dicha suma por numerador será también cero, o sea, que el tercer momento será nulo. De este modo, resulta evidente que, cuando el tercer momento con respecto a la media aritmética es igual a cero, la distribución es simétrica.

En cambio, cuando existan más desviaciones positivas que desviaciones negativas en la distribución, predominarán los cubos positivos sobre los negativos y la suma algebraica de los cubos de las desviaciones será positiva, con lo que μ_3, el tercer momento con respecto a la media aritmética también lo será o sea, que cuando el tercer momento es positivo, la curva es positivamente asimétrica.

En el caso contrario, o sea cuando predominan las desviaciones negativas sobre las positivas, predominarán en la suma los cubos negativos sobre los positivos y la suma será negativa, con lo cual el tercer momento será negativo. O sea, que cuando el tercer momento con respecto a la media aritmética es negativo la curva es negativamente asimétrica.

El uso del tercer momento con respecto a la media aritmética tomado por sí solo no nos rendiría más servicio que el de darnos a conocer si la curva es simétrica o asimétrica y si es asimétrica en qué sentido lo es, si positiva o negativamente. Esto no basta, ya que, en general, en estadística se busca la comparación entre diferentes distribuciones, y no podrían compararse los resultados de distribuciones cuyos valores están dados en muy diferentes unidades.

Con el fin de poder comparar el grado de asimetría de las distribuciones, la burda medida de asimetría representada por el tercer momento con respecto a la media aritmética se reduce a términos comparables dividiéndola entre la medida de dispersión de la distribución, o sea entre la desviación media cuadrática a la que se eleva al cubo (con el fin de compensar el efecto de elevación al cubo de las desviaciones para la obtención del tercer momento). La nueva medida de asimetría (a la que llamaremos α_1) indica el grado de asimetría en términos abstractos (y por lo mismo comparables) en vez de en términos concretos; corresponde en realidad a un tercer momento en unidades sigmáticas.

$$\alpha_1 = \frac{\mu_3}{\sigma^3}$$

Otra medida que se suele emplear para la asimetría es la llamada β_1 que es en realidad el cuadrado de α_1. En efecto:

$$\beta_1 = \alpha_1^2 = \left(\frac{\mu_3}{\sigma^3} \right)^2 = \frac{\mu_3^2}{\sigma^6}$$

Pero como σ^2 es igual a μ_2 (el segundo momento con respecto a la media aritmética o "variancia"), podremos escribir:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Tanto α_1 como β_1 necesitan valer cero para que la distribución sea simétrica.

Una tercera medida de asimetría basada en los momentos.—Parece conveniente, con el fin de enmarcar el cálculo de las medidas de asimetría en un contexto más amplio, proponer una que siendo sustancialmente la misma que las anteriores, señala conexiones que las previamente mencionadas enmascaran. En efecto, si extraemos la raíz cúbica de α_1 y llamamos al nuevo valor γ_1, tendremos:

$$\sqrt[3]{\alpha_1} = \gamma_1 = \frac{\sqrt[3]{\mu_3}}{\sqrt[3]{\sigma^3}} = \frac{\sqrt[3]{\mu_3}}{\sigma}$$

Pero como $\mu_3 = \Sigma(d^3)/N$; $\sqrt[3]{\mu_3} = \sqrt[3]{\Sigma(d^3)/N}$ y esta raíz cubica del cociente que resulta de dividir a la suma de los cubos de las desviaciones entre el número de datos es lo que en la sección relativa a las medias estadísticas hemos definido como una media cúbica; en este caso, media cúbica de las desviaciones con respecto a la media aritmética, si dicha media cúbica la representamos por $\bar{d}_c$ en forma análoga a como representamos a la media cuadrática de las desviaciones por $\bar{d}_q$ encontraremos una cierta conexión entre la media de asimetría calculada con base en la desviación media cúbica, y la medida de dispersión calculada sobre la base de la desviación media cuadrática. Si en la expresión de γ_1 se substituye $\sqrt[3]{\mu_3}$ por $\bar{d}_c$ tendremos como medida de asimetría:

$$\gamma_1 = \frac{\bar{d}_c}{\sigma}$$

O sea, que con el fin de calcular la asimetría de una distribución habrá que:

1º—Calcular la desviación cúbica media de las desviaciones con respecto a la media aritmética (o bien la raíz cúbica del tercer momento con respecto a dicha media) y

2º—Dividir dicha desviación cúbica media entre la desviación cuadrática media.

CURTOSIS, ACHATAMIENTO O PICUDEZ

La curtosis de una distribución mide el grado de achatamiento o picudez de la misma.

Cuanto se ha dicho para el cálculo de la asimetría puede aplicarse, *mutatis mutandis,* al cálculo de las medidas de curtosis. De entre éstas, cabe mencionar principalmente la α_4 y la γ_2.

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

$$\gamma_2 = \frac{\sqrt[4]{\mu_4}}{\sigma} = \frac{\bar{d}_b}{\sigma}$$

O sea, que para determinar el grado de curtosis de una distribución, o bien se calcula el cuarto momento con respecto a la media aritmética y se le divide entre la cuarta potencia de la desviación cuadrática media (o entre la segunda potencia del segundo momento), o:

1º—Se calcula la desviación media bi-cuadrática de la distribución (igual a la raíz cuarta del cuarto momento), y

2º—Se le divide entre la desviación media cuadrática de la distribución.

Cuando se calcula el valor de α como medida de curtosis, se considera como valor medio el de 3 (así como se consideraba el valor de 0 como valor medio en el caso de la asimetría) y se clasifica a las curvas en la forma siguiente:

1.—Con curtosis inferior a 3 "platicúrticas" o achatadas,

2.—Con curtosis igual a 3 "mesocúrticas" o de achatamiento medio,

3.—Con curtosis superior a 3 "leptocúrticas" o picudas.

Puesto que γ_2 es la $\sqrt[4]{\ }$ de α_4, una curva será mesocúrtica cuando γ_2 valga $\sqrt[4]{3} = 1.316$, platicúrtica cuando tenga un valor menor al de $\sqrt[4]{3}$ y leptocúrtica cuando tenga un valor mayor que $\sqrt[4]{3}$.

CÁLCULO CONJUNTO DE LA MEDIA, LA DESVIACIÓN MEDIA CUADRÁTICA Y LOS ÍNDICES PEARSONIANOS DE ASIMETRÍA Y CURTOSIS

Cuando se trata de caracterizar una distribución mediante el cálculo de la media aritmética como promedio central representativo, de la desviación media cuadrática como medida de dispersión indicadora del grado de representatividad de la media aritmética, de un índice de asimetría que señale el predominio o falta de predominio de los valores inferiores sobre los superiores a la media aritmética o de éstos sobre aquéllos, y de un índice de curtosis que indique el mayor o menor aplanamiento (la menor o mayor picudez) de una distribución, puede procederse ordenadamente al cálculo de los momentos necesarios para la obtención de las medidas respectivas, en la forma siguiente:

1º—Calcúlense los momentos con respecto a la media arbitraria en unidades del intervalo, para lo cual:

A.—Obténganse las desviaciones con respecto a una media elegida, en unidades del intervalo:

a.—determinando los puntos medios de los intervalos,
b.—eligiendo uno de dichos puntos medios como media arbitraria,
c.—colocando 0 frente a la media elegida,

— 1, — 2, — 3 . . . hacia arriba,
1, 2, 3 . . . hacia abajo

(Estos últimos valores, colocados en una columna son las desviaciones buscadas.)

B.—Obténganse los productos de las potencias de las desviaciones con respecto a la media elegida, en unidades del intervalo por las frecuencias:

a.—multiplicando las desviaciones anotadas en la columna anterior por las frecuencias correspondientes y anotando los resultados en otra columna, a fin de obtener los productos de las primeras potencias de las desviaciones por las frecuencias (ponderación).
b.—multiplicando los productos de la columna anterior por las desviaciones, a fin de obtener los productos de los cuadrados de las desviaciones por las frecuencias, anotando los resultados en otra columna.
c.—multiplicando los productos de la columna anterior por las desviaciones, a fin de obtener los productos de los cubos de las desviaciones por las frecuencias, anotando los resultados en otra columna.
d.—multiplicando los productos de la columna anterior por las desviaciones, a fin de obtener los productos de las cuartas po-

tencias de las desviaciones por las frecuencias anotando los resultados en otra columna.

C.—Obténganse los momentos (primero, segundo, tercero y cuarto) con respecto a la media arbitraria mediante el cálculo de las medias de las potencias correspondientes:

a.—dividiendo la suma de la columna *a* del inciso anterior entre el efectivo de la distribución o sea la suma de sus frecuencias para obtener el primer momento.

b, c, d.—dividiendo las sumas de los productos contenidos en las columnas segunda, tercera y cuarta entre el efectivo de la distribución (o sea la suma de las frecuencias) a fin de obtener el segundo, el tercero y el cuarto momento con respecto a la media arbitraria en unidades del intervalo.

2º—Se calcularán las potencias crecientes del primer momento con respecto a la media arbitraria en unidades del intervalo,

A, B, C.—elevando al cuadrado, al cubo y a la cuarta potencia el cociente obtenido en *a* del inciso anterior.

3º—Se calcularán los momentos *con respecto a la media aritmética* en unidades del intervalo, substituyendo en las fórmulas correspondientes los valores de los momentos *con respecto a la media arbitraria* en unidades del intervalo y las potencias crecientes del primer momento con respecto a la media arbitraria en unidades del intervalo calculados en *1º* y *2º* (Ver fórmulas de relaciones entre unos momentos y otros en p. 100).

4º—Se calcularán los momentos con respecto a la media aritmética *en unidades originales,* multiplicando cada uno de los momentos con respecto a la media aritmética *en unidades del intervalo* obtenidos en *3º* por las potencias correspondientes del intervalo; para ello:

A.—Elévese el valor del intervalo a la segunda, a la tercera y a la cuarta potencia.

B.—Multiplíquese:

a.—el cuadrado del intervalo, por el segundo momento con respecto a la media aritmética en unidades del intervalo,

b.—la tercera potencia del intervalo por el tercer momento.

c.—la cuarta potencia del intervalo, por el cuarto momento con respecto a la media aritmética en unidades del intervalo.

En esta forma, se obtendrán el segundo, el tercero y el cuarto momento en unidades originales.

5º—Se aplicará la corrección por agrupamiento en caso necesario.

6º—La media aritmética se calculará haciendo las substituciones correspondientes en la fórmula:

$$\bar{x}_a = m' + \mu'_1 i \quad \text{o} \quad \bar{x}_a = m' + \mu_1$$

en donde M_1' representa el primer momento con respecto a la media arbitraria en unidades del intervalo.

7º—La desviación media cuadrática se calculará haciendo las substituciones correspondientes en la fórmula:

$$\sigma = \sqrt{\mu_2}$$

8º—El índice de asimetría $\sqrt{\beta_1}$ mediante la fórmula:

$$\beta_1 = \frac{\mu_3}{\sigma^3}$$

(en donde σ^3 se puede obtener multiplicando el valor de σ por μ_2).

9º—La medida de curtosis (β_2) se obtendrá haciendo las substituciones correspondientes en la fórmula:

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

(σ^4 puede obtenerse elevando al cuadrado μ_2, segundo momento con respecto a la media aritmética en unidades originales).

MEDIANA

Si suponemos una serie de clases y frecuencias, definida por los límites inferiores de las clases $l_1, l_2, l_3, \ldots l_n$, los límites superiores $L_1, L_2, L_3, \ldots L_n$ y las frecuencias de clase $F_1, F_2, F_3, \ldots F_n$, y queremos obtener la mediana de esa distribución, razonaremos como sigue:

Puesto que la mediana es el valor de la variable por debajo del cual y por encima del cual se encuentran respectivamente la mitad inferior y la mitad superior de la distribución, será necesario calcular en primer término cuantos casos comprende la distribución total (sumando las frecuencias) y cuál es la mitad de ese número (dividiendo la suma de las frecuencias entre 2). La mitad de la distribución estará dada por $\Sigma F_i/2$.

En seguida será necesario localizar cuál es la clase para la cual la suma de todas las frecuencias anteriores es igual a $\Sigma F_i/2$. Supongamos que al sumar $F_1 + F_2$ hubiésemos obtenido el valor $\Sigma F_i/2$ esto querría decir que L_2 o sea el límite superior de la segunda clase comprendería los F_2 casos de dicha segunda clase y los F_1 casos de la clase previa, o sea, que por debajo de L_2 se encontrarían $\Sigma F_i/2$ o sea la mitad de la distribución. Esto querría decir que L_2 sería el valor de la mediana.

Sin embargo, no siempre es tan sencillo el caso, y en las mismas series de clases y frecuencias puede ocurrir que, una vez determinado $\Sigma F_i/2$ no haya clase para la cual la suma de todas las frecuencias anteriores sea exactamente igual a la mitad de la distribución $\Sigma F_i/2$; en este caso, podemos suponer, por ejemplo que al sumar todas las frecuencias hasta la frecuencia de la tercera clase $(F_1 + F_2 + F_3)$ obtenemos un valor ligeramente inferior a $\Sigma F_i/2$, y que si sumamos la frecuencia de la clase siguiente $(F_1 + F_2 + F_3 + F_4)$ obtenemos un valor ligeramente superior a $\Sigma F_i/2$. Esto querrá decir que la mediana será un valor comprendido entre los límites superiores de las clases tercera y cuarta (L_3 y L_4). Si tomamos el límite superior de la clase tercera (que prácticamente se confunde con el límite inferior de la clase cuarta l_4) habrá que pensar en agregar un cierto número de frecuencias a la suma de las anteriores $F_1 + F_2 + F_3$ para tener las $\Sigma F_i/2$ que requiere la mitad de la distribución. Estas frecuencias habrá que tomarlas de la clase cuarta; o sea, que de las F_4 frecuencias de la cuarta clase habrá que tomar las $\frac{\Sigma F_i}{2} - (F_1 + F_2 + F_3)$. Finalmente, habrá que preguntarse a qué tantas unidades del intervalo entre l_4 y L_4 corresponden esas $\frac{\Sigma F_i}{2} - (F_1 + F_2 + F_3)$ y agregar dicho valor a l_4 para obtener la mediana.

En una simbología más comprimida, podríamos representar la suma total de las frecuencias por $\sum_1^N F_i$, y la suma de las frecuencias $(F_1 + F_2 + F_3)$ que acumuladas dan un valor cercano al de $\sum_1^N F_i/2$ por $\sum_1^{m-1} F_i$ (que leeríamos) suma de las frecuencias acumuladas hasta la anterior a la que contiene a la mitad de la distribución.

Conforme a la anterior simbología, lo que sería preciso agregar al límite inferior de la clase se obtendría resolviendo la proporción:

$$x : \frac{\sum_{1}^{N} F_i}{2} - \sum_{1}^{m-1} F_i :: i : F_4$$

Despejando el valor de x, se tendría:

$$x = \left(\frac{\dfrac{\sum_{1}^{N} F_i}{2} - \sum_{1}^{m-1} F_i}{F_4} \right) i$$

O sea, que la fórmula para la mediana sería:

$$Md = l + \left(\frac{\dfrac{\sum_{1}^{N} F_i}{2} - \sum_{1}^{m-1} F_i}{F_m} \right) i$$

Según esta fórmula, el cálculo de la mediana en series de clases y frecuencias puede hacerse conforme al siguiente procedimiento:

1º—Súmense todas las frecuencias y divídanse entre 2.

2º—Búsquese cuál es la clase para la cual la suma de las frecuencias se aproxima más (por debajo) a la mitad de la suma de las frecuencias.

3º—Obténgase la diferencia entre la mitad de la suma de las frecuencias y la suma acumulada de frecuencias que más se le aproxima por debajo.

4º—Divídase dicha diferencia entre la frecuencia de la clase en la que está comprendida la mitad de la suma de las frecuencias.

5º—Multiplíquese el cociente por el valor del intervalo.

6º—Agréguese el producto al límite inferior de la clase que contiene a la mitad de la suma de las frecuencias.

Si se tiene en cuenta que a la suma de las frecuencias de una clase con las frecuencias de las clases precedentes (o subsiguientes) se le conoce como "frecuencia acumulada", podremos delinear el mismo procedimiento anterior con las siguientes palabras:

1º—Determínese la mitad de las frecuencias.

2º—Búsquese la clase que contiene dicha mitad de la distribución.

3º—Réstese de la mitad de la distribución, la frecuencia acumulada anterior a la que contiene a dicha mitad.
4º—Divídase dicha diferencia entre la frecuencia no acumulada de la clase que contiene a la mitad de la distribución.
5º—Multiplíquese dicho cociente por el intervalo.
6º—Agréguese dicho producto al límite inferior de la clase que contiene a la mitad de la distribución.

En todo lo anterior, hemos explotado un supuesto: el de que se quiera trabajar con el límite inferior del intervalo de clase que contiene a la mediana; pero bien puede suceder que convenga trabajar con el límite superior; en efecto, si se trabaja con el límite superior, se hace innecesaria la mención de la clase *anterior* a la que contiene a la mediana y, con ello, se evita un escollo en la memorización del procedimiento de cálculo.

Efectivamente, es fácil comprender que si se usa el límite superior es necesario restar algo en vez de sumar algo al límite para obtener la mediana. De ahí que entre el límite superior y la fracción deba de haber un signo —, en tanto había un signo $+$ entre el límite inferior y la fracción. Por otra parte, para saber cuánto hay que restar será preciso saber cuántas son las frecuencias en que la suma de las frecuencias de los intervalos anteriores a la clase que contiene a la mediana $(\sum_{1}^{m} F_i)$ excede a la mitad de la distribución $\frac{\sum_{1}^{N} F_i}{2}$. O sea, que será necesario modificar también el numerador de la fracción, con lo cual la fórmula para el cálculo de la mediana con empleo del límite superior L será:

$$Md = L - \left(\frac{\sum_{1}^{m} F_i - \frac{\sum_{1}^{N} F_i}{2}}{F_m} \right) i$$

Según esto, el procedimiento consistiría en:

1º—Determinar cuál es la mitad de la distribución.
2º—Determinar cuál es la frecuencia acumulada que contiene a dicha mitad de la distribución.
3º—Restar de la frecuencia acumulativa que contiene a la mitad de la distribución, dicha mitad.

4º—Dividir dicha diferencia entre la frecuencia no acumulada, de la clase que contiene a la mitad de la distribución.

5º—Multiplicar dicho cociente por el intervalo.

6º—Restar dicho producto del límite superior de la clase que contiene a la mitad de la distribución.

Las Cuantilas.—La mediana no representa sino un tipo particular de promedio, especialmente importante por ser el central pero que no es el único, ya que hay otros promedios, que se calculan como la mediana, y que reciben el nombre de promedios laterales, o promedios de graduación.

A los promedios de graduación se les conoce genéricamente como "cuartilas" y por medio de ellos se divide a las distribuciones en partes iguales.

Es así como, a semejanza de la mediana que divide a una distribución en dos partes iguales (valores inferiores a la mediana y valores superiores al de la mediana) hay otros promedios que dividen a la distribución en cuatro, seis, diez, cien partes iguales, a los cuales se les denomina genéricamente cuantilas, llamándose cuartilas cuando dividen a la distribución en cuatro partes iguales, sextilas cuando la dividen en seis, decilas cuando las dividen en diez partes iguales y centilas (antiguamente las llamaban porcentilas) cuando la dividen en cien partes iguales.

Las cuartilas son especialmente importantes, pues sirven para determinar las llamadas:

1.—Zona de deficiencia.	de un fenómeno.
2.—Zona de normalidad, y	
3.—Zona de excedencia.	

En efecto, las cuartilas son los valores límites de cuatro zonas, dentro de cada una de las cuales está comprendida una cuarta parte (o un 25 %) de las frecuencias; es así como:

La Primera Cuartila	Es el límite superior de la primera cuarta parte de las frecuencias, o sea del primer	25 %
La Segunda Cuartila	Es el límite superior de la segunda cuarta parte de las frecuencias, o sea del segundo	25 %
La Tercera Cuartila	Es el límite superior de la tercera cuarta parte de las frecuencias, o sea del tercer	25 %
La Cuarta Cuartila	Es el límite superior de la cuarta cuarta parte de las frecuencias, o sea del cuarto (último)	25 %
O sea que las cuatro cuartilas cubren el		100 %

La zona (25 %)	limitada por la primera cuartila recibe el nombre de	Zona de deficiencia
La zona (50 %)	comprendida entre la primera y la tercera recibe el nombre de	Zona de normalidad
La zona (25 %)	comprendida entre la tercera y la cuarta recibe el nombre de	Zona de excedencia

Por simetría, puede hablarse de una cuartila cero como límite inferior de la zona de deficiencia.

Las cuartilas se pueden representar por una C (reservaremos la Q usada tradicionalmente para las cuantilas *in genere*) afectada de un sub-índice que indique el número de orden de la cuartila; así C_1, C_2, C_3, C_4 (representarán la primera, la segunda, la tercera y la cuarta cuartila).

De hecho, la C_0 o cuartila cero no tiene que calcularse, pues coincide siempre con el cero.

La C_4 o cuartila cuatro no tiene que calcularse, pues coincide con el total de la distribución (ΣF_i).

La C_2 o cuartila dos no tiene que calcularse, porque en cuanto la cuartila dos divide a la distribución en dos partes iguales (25 % entre C_0 y C_1 y 25 % entre C_1 y C_2) es, por definición la mediana que suponemos ya calculada.

Sea como fuere, la fórmula para el cálculo de las cuartilas será análoga a la empleada para calcular la mediana; así, mientras en el caso de la mediana se calculaba cuál era la mitad de la distribución, en el caso de las cuartilas será preciso determinar cuál es la cuarta parte de la distribución y tomar tantas cuartas partes como lo indique el número de orden de la cuartila para en seguida, utilizar el valor obtenido como guía para encontrar la frecuencia acumulativa que contiene a dicho valor, restar de dicha frecuencia ese valor, dividir el residuo entre la frecuencia no acumulativa de la clase que contiene a ese valor, multiplicar el cociente por el intervalo de clase y restar el producto del límite superior de la clase que contiene al valor calculado $n\frac{N}{4}$. O sea, que la fórmula para las cuartilas puede escribirse:

$$C_n = L - \left(\frac{\sum_1^m F_i - n \frac{\sum_1^N F_i}{4}}{F_m} \right) i$$

En forma general para las cuantilas, si Q representa cuantilas y q el número de partes iguales en que las cuantilas dividen la distribución, tendremos:

$$Q_n = L - \left(\frac{\sum_{1}^{m} F_i - n \dfrac{\sum_{1}^{N} F_i}{q}}{F_m} \right) i$$

Esta fórmula especialmente aplicable en el caso de una serie de clases y frecuencias es asimismo aplicable en el caso de una serie de frecuencias; en tal caso, L equivale a una x_i, e i vale 1.

DISPERSIÓN POR MEDIO DE CUANTILAS

En forma parecida a como la mediana cumple en la estadística cuantilar una función representativa de las distribuciones análoga a la función representativa de la media en la estadística de los momentos, las cuantilas han sido utilizadas a semejanza de la desviación media cuadrática (y de su cuadrado, la variancia) para medir la dispersión de una distribución.

Las medidas de dispersión calculadas mediante el uso de las cuantilas tratan de evitar los más serios inconvenientes de la más burda de las medidas de dispersión (la "amplitud" u "oscilación" de la distribución) en todos aquellos casos en los que no es posible calcular medidas más precisas de dispersión (como la desviación media cuadrática).

En efecto, la amplitud mide simplemente la diferencia (algebraica, puesto que puede darse el caso de valores negativos en la distribución y en tal caso habrá necesidad de sumar y no de restar) entre los valores máximo y mínimo.

Como puede comprenderse, la amplitud u oscilación resulta muy influída por los valores extremos, por lo cual se trata de eliminar dichos valores mediante el cálculo de medidas de dispersión que pueden agruparse bajo el rubro de "desviaciones intercuantilares".

A este tipo de desviaciones intercuantilares corresponden principalmente:

1.—La desviación inter-decilar.
2.—La desviación semi-inter-cuartilar.

Como su nombre lo indica, la desviación inter-decilar mide la distancia entre dos decilas (la primera y la novena) de una distribución, obviando con ello el inconveniente de los valores extremos que caen por debajo de la decila primera o por arriba de la decila novena.

Con el cálculo de la desviación semi-inter-cuartilar que mide la distancia media entre la cuartilas primera y tercera, no sólo se elimina el inconveniente de los casos extremos y no representativos (que caen por debajo de la primera cuartila o por arriba de la tercera), sino que además se salva el inconveniente de la asimetría, considerándose en realidad con la desviación semi-inter-cuartilar, la distancia media de los límites de la zona de normalidad al centro de la misma, medida mucho más significativa que la anterior.

ASIMETRÍA EN TÉRMINOS CUANTILARES

Si la distancia entre la cuartila tercera y la mediana fuese igual a la distancia entre la mediana y la cuartila primera, eso indicaría que la curva era simétrica. En este caso la diferencia de estas distancias $(C_3 - Md) - (Md - C_1)$ sería nula.

Si la distancia entre la tercera cuartila y la mediana fuese mayor que la distancia entre la mediana y la primera cuartila eso indicaría una asimetría de la distribución. En ese caso $(C_3 - Md) - (Md - C_1)$ sería positivo e indicaría predominio de los valores bajos.

Si la distancia entre la tercera cuartila y la mediana fuese menor que la distancia entre la mediana y la primera cuartila eso indicaría una asimetría de la distribución. En tal caso, $(C_3 - Md) - (Md - C_1)$ sería negativo e indicaría predominio de los valores altos.

Es decir, que puesto que $(C_3 - Md) - (Md - C_1)$ refleja según sea nulo, positivo o negativo si la distribución es simétrica o asimétrica y cuál es el sentido de la asimetría, puede considerársele como una buena medida de asimetría.

Sin embargo, como la diferencia $(C_3 - Md) - (Md - C_1)$ da por resultado un número concreto que no permite la comparación del grado de asimetría de diferentes distribuciones, conviene reducirlo a términos abstractos, dividiéndolo entre una medida de dispersión. En este caso, como nos movemos en el ámbito de la estadística cuantilar, la medida más conveniente sería la desviación inter-cuartilar o mejor la desviación semi-inter-cuartilar (d_Q).

$$\text{Asimetría} = \frac{(C_3 - Md) - (Md - C_1)}{d_Q}$$

DISTRIBUCIONES DE FRECUENCIAS

Distribución Binomial.—En lo que sigue, vamos a buscar con qué frecuencia se presentan 0, 1, 2, 3, . . . n águilas cuando se lanzan al aire 1, 2, 3, 4, . . . n monedas.

Partiremos del supuesto de que sea una la moneda que se lance. Teóricamente habrá la posibilidad de obtener o un águila o un sol, y la probabilidad que hay de obtener una u otro son iguales; o sea, que si la moneda se lanza un número creciente de veces, podrá observarse que cada vez se realiza más perfectamente la predicción de que "de cada dos tiradas" una corresponde a un sol y una a un águila. Es decir que hay:

1 posibilidad	de obtener 1 águila y 0 soles,	A
1 posibilidad	de obtener 0 águila y 1 sol,	S
—		
2 posibilidades	de obtener 1 águila o 1 sol.	S. A,

Si suponemos que en lugar de tratarse de una moneda se trata de dos monedas, para cada una de ellas se dará la posibilidad de obtener un sol o un águila, pero combinadamente, existen cuatro posibilidades: *1ª*—que las dos monedas den un águila cada una (*AA*); *2ª*—que la primera dé un águila y la segunda un sol (*AS*); *3ª*—que la primera dé un sol y la segunda un águila (*SA*); *4ª*—que las dos den un sol (*SS*). Si no tenemos en cuenta el orden en que aparecen el águila y el sol, podremos identificar como una misma combinación la 2ª y la 3ª; o sea, que podemos decir que hay:

1 posibilidad	de obtener 2 águilas y 0 soles	AA
2 posibilidades	de obtener 1 águila y 1 sol	AS SA
1 posibilidad	de obtener 0 águilas y 2 soles	SS
—		
4 posibilidades	de obtener 2, 1, 0 águilas o 0, 1, 2 soles.	

Podemos establecer algo semejante con respecto a tiradas de 3, 4, 5, ... 10 monedas, obteniendo los siguientes resultados:

Para 3 monedas:

1 posibilidad	de obtener 3 águilas y 0 soles	AAA
3 posibilidades	de obtener 2 águilas y 1 sol	AAS–ASA–SAA
3 posibilidades	de obtener 1 águila y 2 soles	ASS–SAS–SSA
1 posibilidad	de obtener 0 águilas y 3 soles	SSS
—		
8 posibilidades	de obtener 3, 2, 1, 0 águilas o 0, 1, 2, 3 soles.	$(8 = 2^3)$

Para 4 monedas:

1 posibilidad de obtener 4 águilas y 0 soles
4 posibilidades de obtener 3 águilas y 1 sol

6	posibilidades	de obtener	2	águilas	y	2 soles
4	posibilidades	de obtener	1	águila	y	3 soles
1	posibilidad	de obtener	0	águilas	y	4 soles.
—						
16	posibilidades.					$(16 = 2^4)$

Para 10 monedas:

1	posibilidad	de obtener	10	águilas	y	0 soles
10	posibilidades	de obtener	9	águilas	y	1 sol
45	posibilidades	de obtener	8	águilas	y	2 soles
120	posibilidades	de obtener	7	águilas	y	3 soles
210	posibilidades	de obtener	6	águilas	y	4 soles
252	posibilidades	de obtener	5	águilas	y	5 soles
210	posibilidades	de obtener	4	águilas	y	6 soles
120	posibilidades	de obtener	3	águilas	y	7 soles
45	posibilidades	de obtener	2	águilas	y	8 soles
10	posibilidades	de obtener	1	águila	y	9 soles
1	posibilidad	de obtener	0	águilas	y	10 soles.
—						
1 024	posibilidades.					$(1\,024 = 2^{10})$

Si representamos gráficamente la posibilidad de obtener 0, 1, 2, . . . *n* monedas, o sea, si representamos estas diferentes distribuciones, obtendremos una serie de curvas difícilmente comparables, ya que el total de posibilidades de cada distribución es muy distinto. En cambio, podremos comparar las formas de las curvas obtenibles si tomamos las frecuencias relativas dividiendo en cada distribución cada una de las frecuencias entre el total de las frecuencias.

Así, por ejemplo, la probabilidad o frecuencia relativa que hay de obtener un águila lanzando una moneda es de 1/2; siendo asimismo de 1/2 la probabilidad o frecuencia relativa que hay de obtener cero águilas. La frecuencia relativa o probabilidad de obtener dos águilas con dos monedas es de 1/4 (puesto que en total hay cuatro probabilidades distintas); la frecuencia relativa o probabilidad de obtener un águila es de 2/4; la frecuencia relativa o probabilidad de obtener cero águilas es de 1/4 . . .

Lo anterior podemos expresarlo en la forma siguiente:

Para una moneda:

$$\frac{1}{2} A^1S^0 + \frac{1}{2} A^0S^1$$

que se leería:

una de cada dos (1/2) posibilidades de obtener una (el exponente 1 de A) águila (A) y cero (exponente de S) soles (S), y
una de cada dos posibilidades de obtener cero águilas y un sol.

Para dos monedas:

$$\frac{1}{4}A^2S^0 + \frac{2}{4}A^1S^1 + \frac{1}{4}A^0S^2$$

que se leerá:

una de cada cuatro posibilidades de obtener 2 águilas y ningún sol,
dos de cada cuatro posibilidades de obtener 1 águila y un sol,
una de cada cuatro posibilidades de obtener ninguna águila y dos soles.

En la misma forma se podría continuar para el caso de 3, 4 o más monedas, y en todo caso, podría verse la forma en que este modo de notación refleja el desarrollo de las potencias [primera (para una moneda) segunda (para dos monedas), tercera (para tres monedas) ... enésima (para n monedas] del binomio:

$$\frac{1}{2}A + \frac{1}{2}S$$

Los coeficientes de los términos del desarrollo binomial indican la posibilidad que hay de que el atributo A se presente el número de veces que indica el exponente de A en dicho término cuando el número de elementos que tienen dicho atributo es igual al exponente del binomio; así 2/4 (dos de cuatro) son las posibilidades que hay de obtener un águila y un sol cuando se lanzan dos monedas, puesto que 2/4 es el coeficiente de A^1S^1 en el desarrollo de la segunda potencia de $\frac{1}{2}A + \frac{1}{2}S$.

En general un término de orden r del desarrollo de la enésima potencia del binomio tiene (como factores):

1.—El primer término del binomio, elevado a una [potencia igual a la que se ha elevado al binomio] disminuída del [número de orden del término menos uno] (o elevado a una [potencia igual a aquella a la que se ha elevado el binomio disminuída del número de orden del término], [más uno]). Si llamamos a al primer término del binomio, este primer factor será: $a^{n-(r-1)}$ o a^{n-r+1}.

2.—El segundo término del binomio elevado a una potencia igual al [número de orden del término] [menos uno]; o sea, que este segundo factor será: b^{r-1}.

3.—Un coeficiente fraccionario

a.—cuyo numerador está formado por la serie decreciente de los números naturales a partir del exponente del binomio y terminando con el exponente del primer término del binomio en este término del desarrollo menos 1 (de n a $n - (r - 2)$, tomados como factores y

b.—cuyo denominador estará formado por la serie creciente de los números naturales a partir de uno y terminando con el exponente del segundo término del binomio en este término del desarrollo (de 1 a $(r - 1)$, tomados como factores.

Según todo lo anterior, un término de orden r en el binomio será:

$$\frac{n(n-1)\ldots(n-[r-2])}{1.2\ldots\ldots\ldots\ldots\ldots(r-1)} \qquad a^{(n-r-1)} \qquad b^{(r-1)}$$

Coeficiente	Potencia del 1er. término del binomio	Potencia del 2º término

Si volvemos a nuestras probabilidades de obtener águilas o soles con un cierto número de monedas, puesto que la probabilidad de obtener un águila es 1/2 y la de obtener un sol 1/2 podremos substituir a y b por $1/2\,A$ y $1/2\,S$, con lo cual tendremos como probabilidad de obtener $n - (r - 1)$ águilas y $r - 1$ soles, la siguiente:

$$\frac{n(n-1)\ldots(n-[r-2])}{1.2\ldots\ldots\ldots\ldots\ldots(r-1)} \left(\frac{1}{2}A\right)^{n-(r-1)} \left(\frac{1}{2}S\right)^{r-1} =$$

$$\frac{n(n-1)\ldots(n-[r-2])}{1.2\ldots\ldots\ldots\ldots\ldots(r-1)} \left(\frac{1}{2}\right)^{n-(r-1)} \left(\frac{1}{2}\right)^{(r-1)} A^{n-(r-1)}S^{r-1}$$

De este término nos interesa particularmente el coeficiente de $A\,S$, puesto que, en cuanto A y S elevados a $(n - (r - 1)$ y a $r - 1$ respectivamente no hacen sino indicar el número de águilas $n - (r - 1)$ y de soles $r - 1$ que se esperan al tirar n monedas, y esos exponentes $n - (r - 1)$ y $r - 1$ ya aparecen afectando a 1/2 y 1/2 que son las probabilidades respectivas e iguales que corresponden a los dos sucesos posibles (águilas, soles). Por lo mismo, nos reduciremos siempre a considerar la expresión:

$$\frac{n(n-1)\ldots(n-[r-2])}{1.2\ldots\ldots\ldots\ldots\ldots(r-1)}\left(\frac{1}{2}\right)^{n-(r-1)}\left(\frac{1}{2}\right)^{r-1}$$

Esta expresión es, en realidad, el término de orden r del desarrollo potencial:

$$\left(\frac{1}{2}+\frac{1}{2}\right)^{n}$$

Consideremos un ejemplo distinto al de las águilas y soles, a fin de tener bases para establecer una expresión más general. Examinemos el caso del lanzamiento de dados, y supongamos que nos interesa determinar la posibilidad de obtener 0, 1, 2, 3, . . . n seises lanzando 1, 2, 3, . . . n dados.

Si tomamos un solo dado, habrá una oportunidad de cada seis de obtener seises, y 5 oportunidades de obtener cualquier cifra que no sea seis (1, 2, 3, 4, 5). Una de seis oportunidades de obtener seises se representaría como 1/6 y cinco de cada seis oportunidades de obtener otra cifra por 5/6. O sea que habría:

$$\frac{1}{6}+\frac{5}{6}$$

Si en vez de un dado se lanzan dos, para cada uno de ellos sigue habiendo una oportunidad de cada seis (1/6) de obtener un seis, y cinco de cada seis (5/6) de obtener una cifra que no sea seis; sin embargo, combinadamente, puede ocurrir que se obtengan: *1º*—un seis en el primer dado y un seis en el segundo. *2º*—un seis en el primer dado y otra cifra en el segundo. *3º*—una cifra que no sea seis en el primero y un seis en el segundo. *4º*—ningún seis ni en el primero ni en el segundo.

1º—Si se obtiene un seis en el primero no hay sino una posibilidad de que sea seis el del segundo (*1*), en tanto que el total de posibilidades de combinación de una cifra cualquiera del primer dado (sea o no seis) con otra cifra también cualquiera del segundo (sea o no seis) es de 6 (caras del primero) por 6 (caras del segundo); o sea, 36. Es decir que hay una posibilidad de cada 36 de que se obtenga un seis de cada dado (1/6).

2º—Si se obtiene un seis en el primero, hay cinco posibilidades (las cinco caras menos la marcada con seis) de que el segundo marque una cifra que no sea seis (*5*), en tanto que el total de posibilidades de combinación es (según vimos antes) de 36. O sea, que hay 5 posibilidades de cada 36 de que se obtenga un seis en el primer dado y un no-seis en el segundo (5/36).

3º—Si se obtiene un no-seis en el primero (y hay cinco posibilidades de obtener un no-seis con dicho dado) habrá por cada posibilidad de obtener un no-seis con ese dado, una sola posibilidad de obtener un seis con el segundo; como hay cinco oportunidades de obtener un no-seis con el primero, habrá también cinco oportunidades de que un seis del segundo entre en combinación con un no-seis del primero. Como el total de combinaciones posibles es de 36, podremos decir que hay 5 de cada 36 posibilidades de que el primer dado marque una cifra distinta de seis y el segundo un seis (5/36).

4º—Si se obtiene un no-seis con el primer dado (y hay cinco oportunidades de obtener con él un no-seis), habrá cinco oportunidades (por cada no-seis del primero) de obtener un no-seis con el segundo; o sea, que habrá (5×5) veinticinco oportunidades de obtener cifras que no sean seis con los dos dados; o sea, que habrá veinticinco de 36 oportunidades de obtener tal resultado (25/36).

Si tenemos en cuenta que para nuestros propósitos poco importa que el no-seis o el seis sean del primero o del segundo dado, tendremos las posibilidades siguientes:

2 seises	*1 seis y 1 no-seis*	*2 no-seises*
$\frac{1}{36}$	$2 \times \frac{5}{36}$	$\frac{25}{36}$

Lo anterior puede escribirse también, en términos de la probabilidad en el caso de un solo dado:

$$\left(\frac{1}{6}\right)^2 + 2 \times \frac{1}{6} \times \frac{5}{6} + \left(\frac{5}{6}\right)^2$$

Esto es el desarrollo de la segunda potencia del binomio $1/6 + 5/6$.

$$\left(\frac{1}{6} + \frac{5}{6}\right)^2$$

En forma análoga, si se tratara de tres dados, las probabilidades de obtener 3 seises, 2 seises, 1 seis, 0 seises serían dadas por el desarrollo de la tercera potencia del binomio $1/6 + 5/6$. En general, si se tratara de n dados y se trataran de determinar las posibilidades de obtener n, $n - 1$, $n - 2$, ... 2, 1, 0, seises se recurriría al desarrollo de la enésima potencia del binomio $1/6 + 5/6$.

$$\left(\frac{1}{6}+\frac{5}{6}\right)^n$$

El término de orden r (o sea el que indicaría con qué frecuencia relativa o con qué probabilidad se darían $n-(r-1)$ sieses) sería:

$$\frac{n(n-1)\,(n-2)\ldots(n-[r-2])}{1.2.3.\ldots\ldots\ldots\ldots(r-1)}\left(\frac{1}{6}\right)^{n-(r-1)}\left(\frac{5}{6}\right)^{r-1}$$

En forma análoga, si se tratara del mismo experimento de tirar estos dados ideales y saber de las oportunidades de obtener, por ejemplo, una cifra inferior a cuatro, la probabilidad para un dado sería de 4/6 para las tiradas de valor inferior o igual a 4, y de 2/6 para las tiradas de valor superior (5, 6) a cuatro; según esto si se lanzaran n dados, la expresión siguiente daría la posibilidad de obtener $n-(r-1)$ valores inferiores o iguales a cuatro:

$$\frac{n(n-1)\,(n-2)\ldots(n-[r-2])}{1.2.3.\ldots\ldots\ldots\ldots(r-1)}\left(\frac{4}{6}\right)^{n-(r-1)}\left(\frac{2}{6}\right)^{r-1}$$

Si, por ejemplo, se tratara de determinar la posibilidad de obtener 5 caras de valor inferior o igual a cuatro en caso de lanzarse 10 dados, habría que substituir en la expresión anterior n por 10 y $n-(r-1)$ por 5. De $n-(r-1)=5$, y $n=10$ se deduciría que $(r-1)=5$. Substituídos los valores en la expresión anterior, se tendría:

$$\frac{10\times9\times8\times7\times6}{1\times2\times3\times4\times5}\left(\frac{4}{6}\right)^5\quad\left(\frac{2}{6}\right)^5$$

que serían las probabilidades que habría de obtener en 5 de los 10 dados lanzados, valores inferiores a 4.

En general, si la probabilidad de obtener un águila con una moneda es de 1/2, la probabilidad de obtener un seis con un dado es de 1/6, y la probabilidad de obtener un valor menor o igual a cuatro con un dado es de 4/6, podremos representar 1/2, 1/6, y 4/6 o, en general, la probabilidad de que se dé el suceso que calificaríamos de éxito, por P, y 1/2, 5/6 y 2/6 o sean las probabilidades de obtener un sol con una moneda, un no-seis con un dado, o un valor superior a 4 con un dado o, en general, la probabilidad de que se realice el suceso que calificaríamos de "fracaso" por Q.

Con esta notación convencional, el desarrollo del binomio $(P+Q)^n$ o $(Q+P)^n$ representará las diferentes probabilidades (o frecuencias relativas) de obtener n, $n-1$, $n-2$, ... 3, 2, 1, 0 éxitos y complementariamente, de obtener 0, 1, 2, — ... $n-3$, $n-2$, $n-1$, n fracasos, siendo las probabilidades complementarias de éxito o de fracaso por unidad P y Q (las cuales pueden ser iguales como sucede en la probabilidad de obtener una águila o un sol de una moneda perfecta, o diferentes según ocurrió en nuestros ejemplos de los dados).

Media Aritmética de la Distribución Binomial.—Tratándose como se trata de una serie de frecuencias, referirse al cálculo de la media aritmética de la distribución binomial equivale a pensar en la determinación de la media aritmética del número de éxitos (0, 1, 2, 3, 4, ... n) o fracasos (n, $n - 1$, $n - 2$, $n - 3$, $n - 4$, ... 0) ponderada por medio de las frecuencias correspondientes, representada por los correspondientes términos del desarrollo del binomio $(Q + P)^n$. Y tomamos el binomio $Q + P$ y no $P + Q$ a fin de poder hacer coincidir el dato 0 éxitos con su frecuencia correspondiente P^0Q^n, 1 éxito con la respectiva frecuencia nP^1Q^{n-1}, etc.

Para calcular la media aritmética ponderada del número de éxitos, necesitaremos multiplicar el número de éxitos (0, 1, ... n), o sean los datos, por los términos correspondientes del desarrollo del binomio, o sean las frecuencias, sumar dichos productos y dividir la suma entre el total de las frecuencias o sea entre la suma de los términos del desarrollo del binomio.

El primer producto del número de éxitos (0) por su frecuencia (Q^n) se anula. De este modo, la suma de las frecuencias carecerá de término de grado n y comenzará con un término de grado ($n - 1$) al que le corresponderá como número de orden ($r - 1$) la cifra 2 y no 1.

El segundo producto será el resultado de multiplicar el número de éxitos (1) por su frecuencia $nQ^{n-1}P$. Este será el primer término de la suma de los productos de los datos por las frecuencias.

El tercer producto será el resultado que se obtenga de multiplicar el número de éxitos (2) por su frecuencia $\frac{n(n-1)}{1.\ 2} Q^{n-2}P$; el factor 2 (número de éxitos) se reduce a la unidad con el factor 2 del denominador del coeficiente binomial, con lo que el producto resulta ser: $\frac{n(n-1)}{1.} Q^{n-2} P = n(n-1)\, Q^{n-2} P$, que será el segundo término de la suma de los productos de los datos por las frecuencias.

En general, a partir del segundo producto (que se convierte en primer término de la suma de los productos) el factor número de éxitos reduce a la unidad al último factor que interviene en el denominador del coeficiente binomial y si para el producto de orden r, el número de éxitos es $r - 1$ y el valor de la frecuencia respectiva el del término de orden r del binomio $\frac{n(n-1)\ \ldots\ (n-[r-2])}{1.\ 2\ \ldots\ldots\ldots\ldots (r-1)} Q^{n-(r-1)} P^{(r-1)}$ el producto será:

$$(r-1)\,\frac{n(n-1\ (n-2\ \ldots\ (n-[r-2])}{1.\ 2.\ \ldots\ldots\ldots\ldots\ldots\ldots\ (r-1)}\,Q^{n-(r-1)}\,P^{(r-1)}$$

Si se elimina $(r-1)$ que aparece como factor y como divisor, y n se pone como factor de todo el quebrado a fin de que vuelva a haber el mismo número de factores en el numerador y en el denominador (puesto que la eliminación del $r-1$ había disminuído en uno los factores del denominador) y se tiene en cuenta que el factor inmediato anterior a $r-1$ es $(r-1)-1=r-2$, se tendrá:

$$n\,\frac{(n-1)\ (n-2)\ \ldots\ (n-[r-2])}{1.\ \ 2\ \ldots\ldots\ldots\ (r-2)}\,Q^{n-(r-1)}\,P^{(r-1)}$$

$P^{(r-1)}$ puede descomponerse en $P^{(r-1)-1}$ y P, o sea, en $P^{(r-2)}$ y P; si asociamos este segundo factor a n, tendremos:

$$nP\,\frac{(n-1)\ (n-2)\ \ldots\ (n-[r-2])}{1.\ \ 2\ \ldots\ldots\ldots\ (r-2)}\,Q^{n-(r-1)}\,P^{(r-2)}$$

Si agregamos un operador Σ a toda la expresión tendríamos la suma de los productos del número de éxitos por las frecuencias con las que se obtiene ese número; sin embargo nP en cuanto constante puede salir del operador que sólo se aplicará al resto de la expresión:

$$nP\,\Sigma\left[\frac{(n-1)\ (n-2)\ \ldots\ (n-[r-2])}{1\ .\ 2\ \ldots\ldots\ldots\ (r-2)}\right[\,Q^{n-(r-1)}\,P^{(r-2)}$$

Es fácil ver que la expresión de dentro del operador Σ es el término típico del desarrollo de un binomio; el primer factor del coeficiente binomial $(n-1)$ nos da la pista en el sentido de que se trata del binomio $(Q+P)^{n-1}$; en efecto, dicho coeficiente tiene en el numerador términos que van desde $n-1$ hasta $n-(r-2)$; en efecto, el último factor del denominador es el exponente de P, y el substraendo del último factor del numerador debe ser igual a dicho último factor del denominador disminuído en una unidad, o sea $r-2-1=r-3$, que restado de $n-1$ (equivalente de la n de la fórmula general) da: $n-1-(r-3)=n-1-r+3=$ $=n-[(r-2)]$, que es precisamente el último factor del numerador; en cuanto a los exponentes, el de Q es $n-(r-1)$ que sumado con el de P que es $r-2$ da $n-1$, o sea la potencia a la que está elevado el binomio correspondiente. Es decir, que toda la expresión previa se reduce al producto de nP por la suma de los términos de la potencia $n-1$ del binomio $Q+P$.

$$nP\,\Sigma\,(Q+P)^{n-1}$$

Como la suma de las probabilidades de éxito y de fracaso es igual a $1: P + Q = 1$, y la potencia $(Q + P)^{n-1}$ (o la suma de sus términos) es igual a 1, valor que multiplicado por nP da nP. Es decir, que la suma de los productos del número de éxitos por las frecuencias respectivas (o probabilidades de obtener tal número de éxitos) es nP.

Para obtener la media ponderada habrá que dividir ese producto nP por la suma de las frecuencias o sea la suma de los términos del desarrollo de $(Q + P)^n$; pero esta suma, por razones análogas es igual a 1, con lo cual:

$$\text{media binomial} = \frac{nP}{1} = nP$$

Mediante un proceso análogo, puede obtenerse como desviación media cuadrática de la distribución binomial $\sqrt{npq}$.

Distribución de Poisson.—Cuando las probabilidades de fracaso (Q) o de éxito (P) de un acontecimiento se hacen infinitamente pequeñas, pero, al mismo tiempo el número de probabilidades elementales n crece suficientemente como para que nq siga siendo una cantidad finita, los términos de la distribución binomial dan lugar a un tipo especial de distribución conocida como distribución de Poisson.

Para estudiar la distribución de Poisson desarrollemos matemáticamente lo dicho en el párrafo anterior, para lo cual partiremos de la expresión que da la probabilidad o frecuencia relativa de $(r - 1)$ éxitos:

$$\frac{n(n-1)\ (n-2)\ \ldots\ (n-[r-2])}{1.\ \ 2.\ \ \ 3.\ldots\ldots\ldots\ldots(r-1)} P^{n-(r-1)} Q^{r-1}$$

Si representamos Q, las probabilidades de fracaso, por m/n (lo que equivale a hacer a nq igual a la cantidad finita m y despejar el valor de Q) tendremos:

$$nQ = m$$

$$Q = \frac{m}{n}$$

Y, puesto que la suma de las probabilidades es igual a la unidad:

$$P + Q = 1$$

$$P = 1 - Q = 1 - \frac{m}{n}$$

Si substituimos los valores de P y de Q en la expresión de la frecuencia de $(r - 1)$ éxitos de la distribución binomial, tendremos:

$$\frac{n(n-1)\ (n-2)\ \ldots\ (n-[r-2])}{1.\ \ 2\ \ .\ \ 3\ \ldots\ldots\ldots\ldots\ (r-1)}\left(1-\frac{m}{n}\right)^{n-(r-1)}\left(\frac{m}{n}\right)^{r-1}$$

Pero elevar a la potencia $r-1$ a m/n equivale a elevar a m a $r-1$ y dividir dicha potencia entre la potencia $r-1$ de n. Puede asociarse m^{r-1} al producto de todos los factores $(r-1)!$ del denominador del coeficiente; además, el denominador n^{r-1} puede descomponerse en $r-1$ factores iguales a n, cada uno de los cuales puede asociarse con uno de los factores del numerador del coeficiente:

$$\frac{m^{r-1}}{(r-1)!}\left(\frac{n}{n}\ \frac{n-1}{n}.\frac{n-2}{n}\ \ldots\ \frac{n-[r-2])}{n}\right)\left(1-\frac{m}{n}\right)^{n-(r-1)}$$

Si se observan los factores de dentro del paréntesis, podrá verse que el primero se reduce a la unidad, y los siguientes a la unidad menos una fracción que tiene por numerador una constante, y por denominador n:

$$\frac{m^{r-1}}{(r-1)!}\left[1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{r-2}{n}\right)\right]\left(1-\frac{m}{n}\right)^{n-(r-1)}$$

Ahora bien, si el número de ensayos que se hacen (n) crece cada vez más hasta tender a hacerse infinito (lo que reduce cada vez más el valor de Q, probabilidad de fracaso), cada uno de los quebrados de dentro del paréntesis rectangular que tengan como denominador a n tenderán a anularse (puesto que conforme crece el denominador decrece el valor de la fracción), y cada uno de los factores contenidos en los paréntesis curvos de dentro del paréntesis cuadrado tenderán a hacerse iguales a uno, con lo que el producto de todos ellos tenderá a uno, o sea, que puede reemplazarse todo el paréntesis rectangular por uno, con lo cual se obtiene:

$$\frac{m^{r-1}}{(r-1)!}\left(1-\frac{m}{n}\right)^{n-(r-1)}$$

De estos factores, el segundo (o sea el contenido en el paréntesis potenciado) puede escribirse como un cociente ya que restar exponentes (n y $r-1$) equivale a dividir las potencias respectivas:

$$\frac{m^{r-1}}{(r-1)!}\frac{\left(1-\frac{m}{n}\right)^{n}}{\left(1-\frac{m}{n}\right)^{r-1}}$$

Del último factor no hemos tomado límites; o sea, no hemos visto cuál es su valor cuando n crece indefinidamente tendiendo a infinito. Cuando esto ocurre, la fracción m/n del denominador tiende hacia cero, $1 - \frac{m}{n}$ tiende a la unidad (1) y el denominador tiende a 1 elevado a la potencia $r - 1$ que es igual a 1; o sea, que todo el denominador tiende a la unidad, con lo cual, la expresión anterior se convierte en:

$$\frac{m^{r-1}}{(r-1)!}\left(1 - \frac{m}{n}\right)^n$$

En cuanto al segundo factor, puede demostrarse que cuando n (que figura en el denominador de la fracción substraendo y como exponente) tiende al infinito, todo ese factor tiende a la potencia m (numerador de la fracción substraendo) de una cantidad constante $e (= 2.7182)$, base de los logaritmos naturales; o sea, que la anterior expresión se convertirá en:

$$\frac{m^{r-1}}{(r-1)!} e^{-m}$$

Como lo que afirmamos de $r - 1$ éxitos podemos afirmarlo de r éxitos, la anterior expresión puede escribirse más simplemente como:

$$\frac{m^r}{r!} e^{-m}$$

Mediante esta expresión es posible calcular la probabilidad o frecuencia relativa de que se presente un número (r) determinado de éxitos o de fracasos en el caso de un acontecimiento para el cual la frecuencia relativa o probabilidad de un fracaso o de un éxito es muy pequeña.

Si tenemos en consideración que tomamos como punto de partida el hacer el producto $nQ = m$, y recordamos que nP es el valor de la media aritmética en la distribución binomial de la que la de Poisson es caso particular, y que r representa el número de *Ps*, podremos hacer las substituciones correspondientes y encontrar las frecuencias de este tipo de distribución.

Distribución Normal.—Señalaremos en seguida la forma en que la distribución binomial tiende a convertirse en un tipo de distribución particularmente importante, conocida como "distribución normal" cuando el número de ensayos n crece tendiendo al infinito.

Partiremos del caso en que las probabilidades de éxito sean iguales a las probabilidades de fracaso, o sea el caso en el que P pueda ser substituído por Q o Q por P; es decir, que tomaremos como punto de partida una distribución binomial simétrica. En seguida, calcularemos cuál es la probabilidad, frecuencia relativa u ordenada máxima (en este último caso en cuanto se piensa en términos de representación gráfica) de dicha distribución, y calcularemos la relación (relación = matemáticamente a cociente) que existe entre una probabilidad, frecuencia relativa u ordenada cualquiera y la probabilidad, frecuencia relativa u ordenada máxima de la distribución. Tras esto, calcularemos el límite hacia el cual tiende dicha relación entre dichas probabilidades, frecuencias relativas u ordenadas, y expresaremos el valor de la probabilidad, frecuencia relativa u ordenada cualquiera en términos de la máxima. La expresión obtenida nos dará la distribución de frecuencias conocida como "normal".

De acuerdo con lo dicho, si se tiene en cuenta que la distribución binomial de frecuencias está representada por la probabilidad de obtener $n - (r - 1)$ éxitos:

$$\frac{n(n-1)\ldots(n-[r-2])}{1.2\ldots\ldots\ldots\ldots(r-1)} P^{n-(r-1)} Q^{(r-1)}$$

O bien:

$$\frac{n!}{(r-1)!\,[n-(r-1)]!} P^{n-(r-1)} Q^{(r-1)}$$

En cuanto las probabilidades de éxito (P) o de fracaso (Q) son iguales (distribución binomial simétrica), las Qs se pueden substituir por Ps, con lo cual la expresión anterior se transforma en:

$$\frac{n!}{(r-1)!\,[n-(r-1)]!} P^{n-(r-1)} P^{(r-1)}$$

Como para multiplicar dos cantidades iguales elevadas a diferentes exponentes es necesario sumar los exponentes, y la suma de $n - (r - 1)$ exponente de una de las Ps con $r - 1$ exponente de la otra es igual a n, tendremos:

$$\frac{n!}{(r-1)!\,[n-(r-1)]!} P^{n}$$

Esta expresión corresponde a la distribución binomial simétrica y constituirá nuestro punto de partida.

En la expresión anterior, para un valor dado de n fijo, las variaciones que se produzcan dependerán del valor de r, número de orden del término de que se trate, y dicha variación afectará en forma inmediata al coeficiente del término. De este modo, buscar la probabilidad máxima equivale, en este caso, a buscar cual es el término para el cual el coeficiente es máximo.

$$\frac{n!}{(r-1)!\,[n-(r-1)]!}$$

Coeficiente del término de orden r del binomio es máximo (según puede verse por la distribución de los coeficientes binomiales en el triángulo de Pascal o Tartaglia) hacia el centro del desarrollo binomial. Dicho en otras palabras, el coeficiente binomial máximo corresponde al término binomial mediano, o sea, a aquel para el que r vale $n/2+1$, en tratándose de un exponente n par.

Cuando n es impar, hay dos coeficientes binomiales máximos que corresponden a los términos para los cuales r vale

$$\left[\frac{n}{2}+1+\frac{1}{2}\right]$$

Aquí nos ocuparemos sólo del caso de n par. En tal supuesto, si substituimos r por $\frac{n}{2}+1$ en la expresión que nos da el valor del coeficiente de un término de orden r del binomio, tendremos como coeficiente binomial máximo:

$$\frac{n!}{\left[\left(\frac{n}{2}+1\right)-1\right]!\left[n-\left(\frac{n}{2}+1\right)-1\right]!}.$$

En el cual, cada uno de los factoriales del denominador se reducen a $\frac{n}{2}!$ y a $\frac{n}{2}!$. Si representamos a $n/2$ por m y a n (consecuentemente) por $2m$, para facilitar el trabajo, el coeficiente binomial máximo estará dado por la expresión:

$$\frac{(2m)!}{m!\ m!}$$

Y el término binomial máximo, o sea, la probabilidad, frecuencia relativa u ordenada máxima que representaremos por y_{max}: quedará dada por:

$$y_{max}:=\frac{(2m)!}{m!\ m!}\,p^{2m}$$

Si en seguida calculamos una probabilidad, frecuencia relativa u ordenada de la distribución binomial simétrica para $m+x+1$ éxitos, tendremos que substituir en la expresión que nos sirvió de punto de partida, n por $2m$ y r por $m+x+1$.

Pero como $r = m + x + 1$, tendremos que $r - 1 = m + x + 1 - 1$, o sea, que habremos de substituir $r - 1$ por $m + x$:

$$y_{m+x+1} = \frac{2m!}{(m+x)\,!\,(2m - [m+x])\,!}\, P^{2m}$$

Hechas las reducciones en el segundo factorial, se tiene:

$$y_{m+x+1} = \frac{2m!}{(m+x)\,!\,(m-x)\,!}\, P^{2m}$$

Si dividimos entre y_{max}: (a la que en lo sucesivo representaremos por y_o) a y_{m+x+1} (a la que en lo sucesivo representaremos por y_t), tendremos:

$$\frac{y_t}{y_o} = \frac{(2m)\,!}{(m+x)\,!\,(m-x)\,!} \times P^{2m} \times \frac{m!\ m!}{(2m)\,!} \times \frac{1}{P^{2m}}$$

(Puesto que dividir un quebrado entre otro, o en general una cantidad entre otra equivale a multiplicar la primera por el recíproco de la segunda, y en la segunda parte se han tomado los recíprocos de todos los factores que intervienen en y_o.)

En esta última expresión $(2m)!$ que aparece como numerador y como denominador, se reduce a la unidad; P^{2m} que aparece asimismo como factor y como divisor (o denominador) se reduce a la unidad, reduciéndose la expresión a los factoriales $m!$ del numerador y a los factoriales $m + x$ y $m - x$ del denominador, que pueden asociarse formando dos fracciones:

$$\frac{y_t}{y_o} = \frac{m!}{(m+x)!}\ \frac{m!}{(m-x)!}$$

Si se tiene en cuenta que el factorial del denominador de la primera fracción $(m + x)!$ puede considerarse compuesto de dos partes: una formada por los m primeros números naturales 1. 2. 3. m (o sea por $m!$) y una segunda formada por el producto de los números de $m + 1$ a $m + x$, podemos darnos cuenta de que la primera de estas partes se reduce a la unidad con el numerador de la fracción que también es $m!$ quedando reducida dicha fracción a la unidad dividida entre el producto de los números de $(m + 1)$ a $(m + x)$:

$$\frac{m!}{(m+x)!} = \frac{m!}{m!\ (m+1) \ldots (m+x)} = \frac{1}{(m+1) \ldots (m+x)}$$

En forma análoga, si se tiene en consideración que el factorial del numerador de la segunda fracción $m!$ puede descomponerse en dos partes: una formada por los $m - x$ primeros números naturales tomados como factores (o sea por $(m-x)!$) y otra por los números que van de $(m-x+1)$ hasta m, podremos percibir que la primera de dichas partes se reduce a la unidad con el denominador de la fracción que también es $(m-x)!$, con lo que la fracción queda reducida al producto de los números que van de $(m-x+1)$ hasta m pasando por $(m-x+2)$, $(m-x+3)\ldots(m-x+[x-3])$, $(m-x+[x-2])$, $(m-x+[x-1])$ siendo en realidad m el caso de $(m-x+[x-0])$. Si en vez de tomar los factores de este producto resultado de simplificar la segunda fracción en sentido creciente, los ordenamos en sentido contrario, podremos decir que la fracción se reduce al producto de $m(m-1)\ (m-2)\ldots(m-x+1)$.

$$\frac{m!}{(m-x)!}=m(m-1)\ (m-2)\ \ldots\ (m-x+1)$$

Si substituimos estos valores en la relación de y_t y de y_o tendremos:

$$\frac{y_t}{y_o}=\frac{m(m-1)\ (m-2)\ \ldots\ (m-x+1)}{(m+1)\ (m+2)\ldots(m+x)}$$

Si dividimos todos y cada uno de los factores del numerador y del denominador por m (lo que equivale a dividir cada uno de los términos de cada paréntesis por m, tendremos:

$$\frac{y_t}{y_o}=\frac{1\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\cdots\left(1-\frac{x-1}{m}\right)}{\left(1+\frac{1}{m}\right)\left(1+\frac{2}{m}\right)\quad\left(1+\frac{x}{m}\right)}$$

Si tomamos los logaritmos naturales del numerador y del denominador, el logaritmo del cociente será igual a la diferencia entre esos logaritmos; a su vez el logaritmo del numerador (y, por otra parte del denominador) es el logaritmo de un producto y, por lo mismo igual a la suma de los logaritmos de los factores. En resumen, que el logaritmo de y_t/y_o será igual a la suma de los logaritmos de las expresiones del tipo $\left(1-\frac{z}{m}\right)$ menos la suma de los logaritmos de las expresiones del tipo $\left(1+\frac{z}{m}\right)$:

$$L\frac{y_t}{y_o}=L\,1+L\left(1-\frac{1}{m}\right)+L\left(1-\frac{2}{m}\right)\ldots+L\left(1-\frac{x-1}{m}\right)-$$
$$\left[L\left(1+\frac{1}{m}\right)+L\left(1+\frac{2}{m}\right)\ldots+L\left(1+\frac{x}{m}\right)\right]$$

Pero el logaritmo de 1 es cero, por lo cual desaparece el primer término; en cuanto a los siguientes hay que tener en cuenta que el logaritmo natural de una expresión como $\left(1 - \frac{Z}{m}\right)$ puede desarrollarse en forma de serie: $-z + \frac{z^2}{2} - \frac{z^3}{3} \ldots$ serie en la que z representaría a Z/m (o sea, a $1/m$ en el segundo término del desarrollo, a $2/m$ en el tercero, a $1 - \frac{x-1}{m}$ en el último de la suma). Pero, cuando m se hace muy grande tendiendo a infinito (o sea en el caso que nos interesa) $z^2/2$, $z^3/3\ldots$ etc., se hacen muy pequeños, o bien, que en lugar del logaritmo natural de las expresiones $1 - \frac{Z}{m}$, puede tomarse el valor del primer término del desarrollo en serie, o sea $-z$ (que representa a $-Z/m$). De este modo, en lugar del logaritmo natural de $1 - \frac{1}{m}$ podrá tomarse $-1/m$; en lugar de $1 - \frac{2}{m}$, $-2/m$; en lugar de $1 - \frac{x-1}{m}$ se podrá tomar $\frac{x-1}{m}$. Algo análogo puede decirse de los términos del substraendo, los que, sin embargo, por estar ligados 1 y la fracción $\frac{Z}{m}$ por signo $+$ tendrán como primer término de la serie una cantidad positiva $+z$ (es decir que, por ejemplo, el logaritmo natural de $1 + \frac{2}{m}$ podrá reemplazarse por $2/m$):

$$L\frac{y_t}{y_o} = -\frac{1}{m} - \frac{2}{m} \ldots - \frac{x-1}{m} - \left[\frac{1}{m} + \frac{2}{m} \ldots + \frac{x}{m}\right]$$

Como por cada término negativo del minuendo hay uno positivo del substraendo y para restar es preciso cambiar el signo de los términos del substraendo, en realidad por cada término del minuendo hay dos términos negativos en la expresión total, excepto por lo que se refiere a x/m que sólo figura en el substraendo con signo $+$ y que tendría que aparecer en la expresión final con signo $-$. Esto quiere decir que, en realidad, hay 2 términos $-1/m$, 2 términos $-2/m$, 2 términos $-\frac{x-1}{m}$ y, además, un término $-x/m$. Esto se puede expresar:

$$L\frac{y_o}{y_t}=2\left[-\frac{1}{m}-\frac{2}{m}\ldots-\frac{x-1}{m}\right]-\frac{x}{m}$$

Si sacamos como factor común a $-1/m$, tendremos:

$$L\frac{y_o}{y_t}=-\frac{2}{m}\left[1+2+\ldots(x-1)\right]-\frac{x}{m}$$

Pero el paréntesis contiene la suma de los $(x-1)$ primeros números naturales, y, según una regla bien conocida, para obtener la suma de los primeros números naturales, se multiplica el último de los que se desea sumar ($(x-1)$ en este caso) por el que le sigue en la serie natural $(x-1)+1=x$, y el producto se divide entre dos; o sea, que el paréntesis rectangular de la expresión equivale a:

$$\frac{(x-1)x}{2}=\frac{x^2-x}{2}$$

Substituído este valor, tenemos:

$$L\frac{y_t}{y_o}=-\frac{2}{m}\frac{x^2-x}{2}-\frac{x}{m}$$

Pero en esta expresión, 2 como numerador y como denominador se reduce a la unidad, con lo cual la primera fracción queda reducida a $-(x^2-x)/m$; al restar a ésta la segunda fracción se obtiene:

$$L\frac{y_t}{y_o}=\frac{-x^2+x-x}{m}=-\frac{x^2}{m}$$

O sea, que el logaritmo natural de la relación que existe entre una frecuencia cualquiera y la frecuencia máxima de la distribución que consideramos (una distribución binomial simétrica cuyo número de ensayos n tiende a hacerse infinita) es igual a la relación entre el cuadrado de lo que el número de éxitos excede a la mitad del número de ensayos (en efecto x es lo que excede de m, igual a $n/2$) entre la mitad del número de ensayos, tomada dicha relación con signo negativo.

Si se tiene en cuenta que el logaritmo de un número (y_t/y_o) es el número $(-x^2/m)$ al que es preciso elevar una cantidad constante llamada base (que en este caso por tratarse de un logaritmo natural será el número e) para obtener el número propuesto (y_t/y_o) podremos afirmar que:

Si el número e se eleva a la potencia $-x^2/m$ se obtendrá y_t/y_o, o sea que:

$$\frac{y_t}{y_o} = e^{-\frac{x^2}{m}}$$

En esta expresión, podemos despejar a y_t pasando a y_o que está como divisor del primer miembro en calidad de factor al segundo miembro de la ecuación:

$$y_t = y_o\, e^{-\frac{x^2}{m}}$$

Esta expresión puede tomarse como manera de indicar las relaciones que ligan a una probabilidad, frecuencia relativa, u ordenada con la probabilidad, frecuencia relativa u ordenada máxima de una distribución normal.

La expresión anterior cobra pleno sentido en cuanto se tiene en cuenta que $\underline{x}$ representa un exceso (resta, exceso, diferencia o, en terminología estadística, "desviación") y que para m tenemos todo lo que sigue:

La desviación media cuadrática de una distribución binomial cualquiera (no necesariamente simétrica) está dada por la raíz cuadrada del producto del número de ensayos por las probabilidades de éxito y por las probabilidades de fracaso (según se vio en páginas anteriores).

$$\sigma = \sqrt{nPQ}$$

En cuanto se trata de una distribución binomial simétrica, las probabilidades de éxito (P) son iguales a las de fracaso (Q), o sea, que las Qs pueden substituirse por las Ps, con lo que se obtiene:

$$\sigma = \sqrt{nP^2}$$

Si se elevan al cuadrado los dos miembros de la ecuación:

$$\sigma^2 = nP^2$$

Pero, por otra parte, como $n/2$ la hemos representado por m, $n = 2m$, y como, asimismo P la probabilidad de éxito es igual a la de fracaso y la suma de ambas tiene que ser 1, a P le corresponderá una mitad y a Q la otra mitad (recuérdese el ejemplo de las águilas y los soles); o sea:

$$n = 2m$$
$$P = 1/2$$

Substituídos estos valores en el de σ^2 se tiene:

$$\sigma^2 = 2m\left(\frac{1}{2}\right)^2 = 2m\left(\frac{1}{4}\right) = \frac{2m}{4} = \frac{m}{2}$$

Despejando a m, se tiene:

$$m = 2\,\sigma^2$$

Si este valor se substituye en la expresión que da el valor de una frecuencia cualquiera de la distribución (y_t) en términos de y_o y la exponencial e elevada a $(-x^2/m)$ se tendrá:

$$y_t = y_o\, e^{-\frac{x^2}{2\sigma^2}}$$

Como, según hemos dicho x representa la desviación de $m + x$ con respecto a m y m es la media de la distribución (en efecto: media $= np =$ $= n(1/2) = n/2 = m$), podemos representarla como desviación de un valor de la variable x (x_i que representaría a $m + x$) con respecto a la media de la variable ($\bar{x}$ que representaría a m); esto nos daría una forma más usual de la distribución normal:

$$y_i = y_o\, e^{-\frac{(x_i-\bar{x})^2}{2\sigma^2}}$$

En forma aún más común, y con el propósito de darle pleno significado al exponente puede escribirse el dos del denominador como un coeficiente $1/2$ delante del resto de la expresión que, por ser un cociente de cuadrados (el cuadrado de $x_i - \bar{x}$ entre el cuadrado de σ) puede escribirse como el cuadrado de la división o de una fracción $\dfrac{x_i - \bar{x}}{\sigma}$; esta fracción, a su vez, representa: "la desviación (—) con respecto a la media aritmética $(\bar{x})$ en unidades de la desviación cuadrática media ("entre σ") o bien "la desviación con respecto a la media artimética en unidades 'sigmáticas' ". La expresión para la distribución normal resultaría ser:

$$y_i = y_o\, e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}$$

Nota.—Hemos escrito y_i en lugar de y_t en cuanto apareció x_i a fin de relacionar la frecuencia y_i (variable dependiente) con la magnitud del dato x_i (variable independiente). En términos geométricos, se trata de relacionar la ordenada y_i con la abscisa x_i mediante una función de distribución en la que figuran el parámetro y_o probabilidad, frecuencia relativa, frecuencia modal u ordenada máxima, que hay que determinar para cada distribución particular, y una función exponencial (que tiene como base la de los logaritmos naturales) cuyo valor dependerá de dos medidas estadísticas de la distribución (la media y la distribución media cuadrática) y de las diferentes valores de la variable x_i.

Procedimiento para el cálculo de las frecuencias de una distribución normal.—De lo dicho, así como de trabajar cuidadosa y sistemáticamente la última expresión se deduce el siguiente procedimiento para ajustar una curva normal:

1º—Determínese por algún procedimiento (después se precisará cuál) la frecuencia modal u ordenada máxima y_o.

2º—Calcúlese el factor exponencial, en la siguiente forma:

a.—Determínese la desviación de cada dato (x_i) con respecto a la media aritmética ($\bar{x}$) de los mismos que deberá haberse determinado previamente, restando de cada dato la media,

b.—Divídase cada residuo o desviación entre la desviación cuadrática media (σ) que se habrá determinado previamente,

c.—Elévese al cuadrado dicho cociente,

d.—Tómese la mitad de dicha potencia,

e.—Cámbiesele signo a la mitad,

f.—Elévese 2.7182 (valor de e) a la mitad de la potencia con signo cambiado.

3º—Multiplíquense la frecuencia modal u ordenada máxima y el factor exponencial para obtener la frecuencia correspondiente.

Con respecto a este procedimiento deben hacerse varias observaciones:

1ª—Como lo indica ya el hecho de que hayamos empleado el símbolo y_i (una de las últimas letras del alfabeto latino afectada de un subíndice) habrá toda una serie de frecuencias, o sea, que habrá necesidad de repetir el segundo paso tantas veces cuantas x_i (o datos) haya, y asimismo habrá que hacer tantas multiplicaciones de y_o (común para toda la distribución) por el exponencial correspondiente como datos haya.

2ª—La obtención directa del exponencial es a menudo imposible, ya que la potencia resultante en el segundo paso (e) es a menudo fraccionaria. El obstáculo puede obviarse si se trabaja con logaritmos, se obtiene el logaritmo de 2.7182, se multiplica por el exponente obtenido en el paso *2º*, *e*, y se toma el antilogaritmo del resultado. Esta misma operación se simplifica si se trabaja con logaritmos naturales, pues en este caso no será necesario encontrar el logaritmo de 2.7182, puesto que éste es 1 por ser el logaritmo de la base que multiplicado por el exponente dará el exponente mismo, del que no habrá sino que encontrar el antilogaritmo. Pero incluso esta última simplificación puede evitarse, ya que hay tablas en las que están calculados los valores del factor exponencial para cada una de las desviaciones de los datos (x_i) con respecto a la media aritmética $(x_i - \bar{x})$ en unidades sigmáticas $\left[\frac{x_i - \bar{x}}{\sigma}\right]$ o sea, que todo el procedimiento implicado en el segundo paso, se detiene en *b*, o sea que se necesita calcular dicha desviación con respecto a la media aritmética en unidades sigmáticas y en seguida:

c.—Buscar en la tabla de "ordenadas (= frecuencias) de la curva normal" el valor correspondiente a dicha desviación con respecto a la media aritmética en unidades sigmáticas.

Expresión de la frecuencia modal de la distribución normal en términos de la desviación cuadrática media.—Partiremos de la expresión:

$$y_o = \frac{(2m)!}{m!\ m!}\, p^{2m}$$

Si aplicamos la fórmula de Stirling al cálculo de los diversos factoriales que figuran en la ecuación:

$$m! = \sqrt{2\pi}\, e^{-m} m^{m+\frac{1}{2}}$$

Como en el denominador figura $m!$ dos veces como factor, deberemos substituir el denominador por el cuadrado de esta expresión:

$$m!m! = (\sqrt{2\pi})^2\, e^{-2m} m^{2m+1}$$

En la que ha bastado con duplicar los exponentes de cada factor para obtener el cuadrado de la expresión.

Aplicando la misma fórmula de Stirling al factorial del numerador obtenemos:

$$(2m)! = \sqrt{2\pi}\, e^{-2m} (2m)^{2m+\frac{1}{2}}$$

Substituídos los valores de $(2m)!$ y de $m!m!$ en el numerador y denominador respectivamente, se obtiene:

$$y_o = \frac{\sqrt{2\pi}\, e^{-2m} (2m)^{2m+\frac{1}{2}}}{(\sqrt{2\pi})^2 e^{-2m} m^{2m+1}}\, p^{2m}$$

La fracción contenida en esta ecuación puede simplificarse teniendo en consideración que:

1º—Puesto que $\sqrt{2\pi}$ figura elevado al cuadrado en el denominador, y a la primera potencia en el numerador, el factor del numerador reduce a la unidad a uno de los del denominador, quedando solamente $\sqrt{2\pi}$ elevado a la primera potencia en el denominador.

2º—Como los exponenciales "elevado a —2 m" aparecen en numerador y denominador, se reducirán a la unidad.

3º—Como la expresión "$2m$ elevado a $2m + \frac{1}{2}$" que figura en el numerador puede descomponerse en "2 elevado a $2m + \frac{1}{2}$" y "m elevado a $2m + \frac{1}{2}$" y la expresión "m elevado a $2m + 1$" del denominador puede descomponerse en las expresiones "m elevado a $2m + \frac{1}{2}$" y "m elevado a $\frac{1}{2}$" las expresiones "m elevado a $2m + \frac{1}{2}$" que figuran en numerador y denominador se reducirán a la unidad, quedando de ellas sólo "2 elevado a $2m + \frac{1}{2}$" en el numerador y "m elevado a $\frac{1}{2}$" en el denominador.

$$y_o = \frac{2^{2m+\frac{1}{2}}}{\sqrt{2\pi}\, m^{\frac{1}{2}}} P^{2m}$$

Pero como P vale 1/2 por tratarse de una distribución simétrica, P^{2m} valdrá $(1/2)^{2m}$ o $1/2^{2m}$; o sea, que la expresión tiene ulterior simplificación, pues "2 elevado a $2m + \frac{1}{2}$" puede descomponerse en "2 elevado a $2m$" y "2 elevado a $\frac{1}{2}$" y de estas dos partes, la primera, que figura en el numerador se reduce con P^{2m} en el que "e elevado a $2m$" figura en el denominador. Con ello, la expresión resulta ser:

$$y_o = \frac{2^{\frac{1}{2}}}{\sqrt{2\pi}\, m^{\frac{1}{2}}} = \frac{\sqrt{2}}{\sqrt{2\pi}\,\sqrt{m}} = \frac{1}{\sqrt{2\pi}\,\sqrt{\frac{m}{2}}}$$

Finalmente, si se tiene en consideración que $m = 2\sigma^2$, $\sigma^2 = \frac{m}{2}$ y $\sigma = \sqrt{\frac{m}{2}}$ tendremos:

$$y_o = \frac{1}{\sqrt{2\pi}\,\sigma} \text{ o } = \frac{1}{\sqrt{2\pi\,\mu_2}}$$

Expresión que nos da el valor de la frecuencia modal u ordenada máxima y_o en términos de la desviación media cuadrática.

Si se substituye este valor en la expresión para el valor de una probabilidad, frecuencia relativa u ordenada cualquiera de la distribución se obtiene como expresión muy general:

$$y_i = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2}$$

Conforme a esto, el cálculo de la frecuencia modal se reduce a:

a.—Multiplicar la desviación media cuadrática de la distribución por $\sqrt{2\pi} = 0.3989$ (valor constante).

b.—Tomar el recíproco de dicho producto.

c.—En caso de no tratarse de una distribución unitaria será necesario multiplicar por el número de datos (en la expresión aparecerá N como numerador de raíz cuadrada de 2π, σ).

DESVIACIONES CON RESPECTO A LA MEDIA ARITMÉTICA EN UNIDADES DE LA DESVIACIÓN CUADRÁTICA MEDIA

Las desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media son las diferencias que se obtienen de restar de cada valor de la variable la media aritmética de la distribución y de dividir subsecuentemente los resultados entre la desviación media cuadrática de la propia distribución.

Representaremos las desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media por δ_i, y dichos valores quedarán definidos por las fórmulas:

$$\delta_1 = \frac{x_i - \bar{x}_a}{\bar{d}_q} = \frac{x_i - \bar{x}_a}{\sigma}$$

Cuando (con propósitos de cálculo de la media aritmética) se han obtenido las desviaciones con respecto a la media arbitraria en unidades del intervalo (D') y el primer momento con respecto a dicha media también en unidades del intervalo (M'_1) la obtención de las δ_i puede simplificarse; en efecto, si se trata de series de clases y frecuencias, se tendrá:

$$\delta_i = \frac{m_i - \bar{x}_a}{\sigma}$$

Si dividimos numerador y denominador de la fracción que forma el segundo miembro entre el intervalo, la fracción no se altera:

$$\delta_i = \frac{\dfrac{m_i - \bar{x}_a}{i}}{\dfrac{\sigma}{i}}$$

Como una fracción equivale a un cociente, y una división de fracciones equivale a la multiplicación de la fracción dividendo por el recíproco de la fracción divisora:

$$\delta_i = \frac{m_i - \bar{x}_a}{i} \div \frac{\sigma}{i} = \frac{m_i - \bar{x}_a}{i} \cdot \frac{i}{\sigma}$$

Si substituímos por $\bar{x}_a$ su valor $m' + M'_1 i$, tendremos:

$$\delta_i = \frac{m_i - m' - M'_1 i}{i} \cdot \frac{i}{\sigma}$$

Pero $\dfrac{m_i - m'}{i} = D'$ desviación con respecto a la media arbitraria en unidades del intervalo, y $M'_1 i/i$ es igual a M'_1 puesto que i en el numerador y en el denominador se reduce a la unidad. De este modo:

$$\delta_i = (D' - M'_1)\,\frac{i}{\sigma}$$

Según esto, ya con propósitos prácticos, si se han determinado las desviaciones con respecto a la media arbitraria en unidades del intervalo ($D' = -n \ldots -2, -1, 0, 1, 2, \ldots n$) y el primer momento con respecto a la media arbitraria también en unidades del intervalo M'_1, para encontrar las desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media (δ_i), según lo indica la fórmula anterior:

1º—Deberá restarse de cada desviación con respecto a la media arbitraria en unidades del intervalo el primer momento con respecto a dicha media en dichas unidades, anotando los resultados $(D' - M'_1)$ en una columna,

2º—Obtener el cociente del intervalo dividido entre la desviación cuadrática media de la distribución (σ),

3º—Deberá multiplicarse cada uno de los valores obtenidos en *1º* por el cociente obtenido en *2º*, a fin de obtener las desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media (δ_i).

AJUSTE SIMPLIFICADO DE UNA CURVA NORMAL POR EL MÉTODO DE LAS ORDENADAS

Calculadas las desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media, y si se cuenta con tablas que den el valor de *e* elevado a $-\delta_i$ sobre dos, dividido por raíz cuadrada de 2 π, el proceso se reduce a:

1º—Buscar los valores de las δ_i en la entrada de la tabla, y anotar en una columna de la tabulación el valor correspondiente obtenido en el cuerpo de la propia tabla de ordenadas de la curva normal en unidades de la raíz de 2π.

2º—Obtener el producto del efectivo de la distribución (suma de las frecuencias) por el cociente que resulta de dividir el intervalo entre la desviación cuadrática media de la distribución $(i/\sigma \times \Sigma f_i)$.

3º—Multiplicar el producto resultante en *2º* por los valores obtenidos de la tabla en *1º*.

EXPONENCIAL DE LA DESVIACIÓN SIGMÁTICA

Definiremos como exponencial de la desviación sigmática, la expresión:

$$e^{-\frac{\delta_i^2}{2}}$$

que, para mayor comodidad representaremos ocasionalmente por $\varphi_o(\delta_i)$.

La función exponencial de la desviación sigmática puede escribirse asimismo como:

$$\frac{1}{e^{\frac{\delta_i^2}{2}}}$$

puesto que una cantidad elevada a un exponente negativo es igual al recíproco de esa cantidad elevada al exponente positivo correspondiente.

El valor de $\varphi_o(\delta_i)$, en cuanto función de δ_i, variará con las variaciones en el valor de δ_i; debe notarse, sin embargo, que para cada valor de la función habrá dos valores posibles de la variante independiente; en efecto, sea que δ_i sea positiva o negativa, su cuadrado será positivo, la mitad de su cuadrado también será postivo y tanto *e* elevado a esa mitad del cuadrado de la desviación sigmática como su recíproco tendrán valores iguales ya sea que se trate de una desviación sigmática positiva o de una desviación sigmática negativa. De ahí que se diga que la exponencial de la desviación sigmática es una función doble.

Por otra parte, conforme sea mayor el valor de la desviación sigmática (δ_i) aumentará el valor de su cuadrado, y de la mitad de su cuadrado o sea, del exponente de *e*, con lo cual crecerá el denominador y, al crecer el denominador disminuirá el valor de la fracción, o sea, disminuirá el valor de la exponencial; asimismo, en cuanto disminuya el valor absoluto de la desviación sigmática aumentará el valor de la función exponencial de la desviación sigmática, llegando a ser máximo el valor de dicha exponencial cuando la desviación sigmática sea nula, caso en el cual se anulará todo el exponente del denominador y el denominador mismo será igual a 1, con lo cual, la función exponencial de la desviación sigmática será igual a la unidad; o sea, que la función exponencial de la desviación sigmática tiene un valor máximo igual a 1 cuando la desviación con respecto a la media aritmética expresada en términos de la desviación cuadrática media es nula.

Primera derivada de la exponencial de la desviación sigmática.—Esta primera derivada se obtendrá multiplicando la misma función por derivar ($\varphi_o(\delta_i)$) por la primera derivada del exponente (o sea, por la derivada de $-\delta_i^2/2$). Si a esta primera derivada la llamamos $\varphi_1(\delta_i)$, tendremos:

$$\varphi_1\delta_i = \varphi_o(\delta_i)\left(\frac{-2\delta_i}{2}\right) =$$

$$= \varphi_o(\delta_i)\ (-\delta_i) = -\delta_i\varphi_o(\delta_i)$$

En este caso, cuando δ_i es nula, si bien el segundo factor $\varphi_o(\delta_i)$ alcanza su valor máximo de 1, la expresión total se anula; asimismo, a diferencia de lo que ocurre con la función exponencial misma, en el caso de la primera derivada cada valor de $\varphi_1(\delta_i)$ corresponde a un solo valor de δ_i, ya que en la primera derivada δ_i aparece elevada a la primera potencia, por lo mismo conserva su signo que tiene que cambiar en cuanto, dentro de la expresión, está afectada de signo menos o sea, que como el valor del segundo factor no cambia su signo, siempre positivo, sea cual fuere el signo de δ_i, la primera derivada será negativa cuando δ_i sea positiva, y positiva cuando δ_i sea negativa.

Segunda derivada de la exponencial de la desviación sigmática.—La segunda derivada de la exponencial es igual a la primera derivada de la primera derivada; o sea, a la primera derivada del producto $-\delta_i\varphi_o(\delta_i)$. Como la derivada de un producto es igual al primer factor por la derivada del segundo más el segundo por la derivada del primero, si representamos esta segunda derivada de la exponencial por $\varphi_2(\delta_i)$ tendremos:

$$\begin{aligned}\varphi_2(\delta_i) &= -\delta_i\varphi_1(\delta_i) + \varphi_o(\delta_i)\frac{d(-\delta_i)}{d\,\delta_i}\\ &= -\delta_i(-\delta_i\varphi_o)(\delta_i) + \varphi_o(\delta_i)(-1)\\ &= \delta_i^2\varphi_o(\delta_i) - \varphi_o(\delta_i)\\ &= (\delta_i^2 - 1)\varphi_o(\delta_i)\end{aligned}$$

Cuando δ_i es nula, el segundo factor es igual a la unidad según se ha visto anteriormente, y el primero se reduce a -1, o sea que la segunda derivada de la exponencial de la desviación sigmática cuando ésta es nula, es igual a -1; la segunda derivada se anula cuando δ_i es igual a $+1$ o a -1, ya que el cuadrado es 1, y el cuadrado -1 es igual a 0, con lo cual, anulándose el primer factor se anula todo el producto o sea, que la segunda derivada es nula. Cuando δ_i tiene un valor absoluto mayor que 1, la segunda derivada es positiva.

Tercera derivada de la exponencial de la desviación sigmática.—La tercera derivada se obtiene derivando la segunda derivada, o sea multiplicando el primer factor $(\delta_i^2 - 1)$ por la derivada del segundo (o sea por $\varphi_1\delta_i$) y sumando el producto del segundo $(\varphi_o(\delta_i))$ por la derivada del primero (o sea, por $2\delta_i$). Si se llama $\varphi_3(\delta_i)$ a esa tercera derivada, se tendrá:

$$\varphi_3(\delta_i) = (\delta_i^2 - 1)\varphi_1(\delta_i) + \varphi_o(\delta_i)(2\delta_i)$$

Substituyendo el valor de $\varphi_1(\delta_i)$ y ejecutando operaciones:

$$\begin{aligned}\varphi_3(\delta_i) &= (\delta_i^2 - 1)[-\delta_i\varphi_o(\delta_i)] + \varphi_o(\delta_i)\,2\delta_i\\ &= (-\delta_i^3 + \delta_i)\varphi_o(\delta_i) + 2\,\delta_i\varphi_o(\delta_i) =\\ &= (-\delta_i^3 + 3\,\delta_i)\varphi_o(\delta_i) =\\ &= (3\,\delta_i - \delta_i^3)\varphi_o(\delta_i)\end{aligned}$$

Cuando δ_i es nula, el segundo factor es igual a la unidad y el primero se anula; o sea que, en total, la tercera derivada es nula.

Para que la tercera derivada sea nula se necesita que sea igualmente nulo el primer factor o sea, el binomio $3\,\delta_i - \delta_i^3$, y para que esto ocurra:

$$3\,\delta_i - \delta_i^3 = 0$$

$$3\,\delta_i = \delta_i^3$$

$$3 = \frac{\delta_i^3}{\delta_i} = \delta_i^2$$

$$\delta_i = \sqrt{3} = 1.732051$$

O sea, que la tercera derivada será nula cuando la desviación sigmática sea igual a 1.732051.

Cuando el valor absoluto de $3\,\delta_i$ (valor sin tomar en cuenta el signo) es mayor que δ_i^3, el valor absoluto de $3\,\delta_i$ menos δ_i^3 será mayor que cero; de ahí que, si la δ_i es positiva, la tercera derivada conserve, en este caso, su signo positivo; en cambio, si δ_i es negativa, será necesario cambiar el signo de la tercera derivada que será, entonces, negativa.

Cuando el valor absoluto de $3\,\delta_i$ sea menor que δ_i^3, el valor absoluto de $3\,\delta_i$ menos δ_i^3 será menor que cero; de ahí que si δ_i es positiva, la tercera derivada conserve, en este caso, el signo negativo; en cambio, si δ_i es negativa será necesario cambiar el signo de la tercera derivada que será, entonces, positiva.

O sea, en general, y ya como regla práctica, que:

A los valores generalmente tabulados de la tercera derivada —que consignan el signo que les corresponde de acuerdo con la relación entre $3\,\delta_i$ y δ_i^3— se les conservará el signo de la tabla si δ_i es positiva, y se les cambiará de signo si δ_i es negativa.

O bien, en otra forma:

La tercera derivada conserva los signos cuando corresponde a desviaciones sigmáticas positivas; los cambia cuando corresponde a desviaciones sigmáticas negativas.

Cuarta derivada de la exponencial de la desviación sigmática.—Si se deriva la tercera derivada, se obtiene como cuarta derivada:

$$\varphi_4(\delta_i) = (3\,\delta_i - \delta_i^3)\,\varphi_1\,(\delta_i) + \varphi_o(\delta_i)\,(3 - 3\,\delta_i^2) =$$
$$= (3\,\delta_i - \delta_i^3)\,\varphi_o(\delta_i)\ \ (-\,\delta_i) + \varphi_o(\delta_i)\,(3 - 3\,\delta_i^2)$$
$$\varphi_4(\delta_i) = (-\,3\,\delta_i^2 + \delta_i^4)\,\varphi_o\,(\delta_i) + \varphi_o(\delta_i)\,(3 - 3\,\delta_i^2) =$$
$$= \delta_i^4 - 6\,\delta_i^2 + 3)\,\varphi_o(\delta_i)$$

Cuando δ_i es igual a cero, el primer factor vale uno, el segundo 3 y la cuarta derivada en total, 3.

Para que la cuarta derivada sea nula, basta con que se anule el segundo factor:

$$\delta_i^4 - 6\,\delta_i^2 + 3 = 0$$

Si consideramos esta ecuación bicuadrática (no figuran sino las potencias pares de la incógnita, 4, 2, 0) como una ecuación bicuadrática en δ_i^2, obtendremos como raíces:

$$\delta^2_{i_1} = \frac{-(-6) \pm \sqrt{(6)^2 - 4 \times 1 \times 3}}{2 \times 1}$$

$$\delta^2_{i_1} = \frac{6 \pm \sqrt{36 - 12}}{2} = \frac{6 \pm \sqrt{24}}{2} = \frac{6 \pm 4.898979}{2}$$

$$\delta^2_{i_1} = .5505105 \qquad\qquad \delta^2_{i_1} = 5.449489$$

Para obtener los valores de δ_i bastará con extraer las raíces cuadradas a estos resultados; así se obtiene:

$$\delta_{i_1} = \pm\ .742 \qquad \delta_{i_2} = \pm\ 2.334$$

O sea, que cuando δ_i vale más o menos .742, y más o menos 2.334, la cuarta derivada es nula.

ADAPTACIÓN DE UNA CURVA NORMAL POR EL MÉTODO DE LAS ORDENADAS

La adaptación de una curva normal por el método de las ordenadas implica el cálculo de:

A.—Un valor constante (el de la ordenada máxima).

B.—Una serie de valores variables (los valores de la función exponencial *e* elevada a menos un medio del cuadrado de la desviación con respecto a la media aritmética en unidades sigmáticas).

Para calcular el valor de la ordenada máxima:

1º—Se obtiene el efectivo de la distribución sumando todas las frecuencias.

2º—Se obtiene el producto de 2.506628 por el valor de la desviación cuadrática media dada en unidades del intervalo (obtenida dividiendo la desviación media cuadrática entre el valor del intervalo).

3º—Se divide el efectivo de la distribución entre el producto anterior, obteniéndose así la ordenada máxima, valor constante por el que habrá que multiplicar los exponenciales que han de calcularse en seguida.

Para obtener el valor del factor exponencial $e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)}$

1º—Se resta de cada dato (/ de cada punto medio) la media aritmética de la distribución.

2º—Se divide dicha desviación entre la desviación cuadrática media de la distribución.

3º—Se busca el valor correspondiente a dicha desviación con relación a la media aritmética en unidades sigmáticas en la tabla de *ordenadas* de la curva normal, repitiéndose el proceso para cada uno de los valores de x_i (o de los puntos medios m_i).

Para encontrar las frecuencias teóricas de la distribución:

Se multiplica el valor de la ordenada o frecuencia máxima por todos y cada uno de los valores exponenciales encontrados en la tabla.

ÁREAS BAJO LA CURVA NORMAL

Cuando un conjunto de valores está distribuído según una curva normal, las magnitudes de los datos acostumbran ser representadas como abscisas, las frecuencias correspondientes a cada una de esas magnitudes, como ordenadas, de tal modo que el área total comprendida entre la curva y los ejes (más particularmente el eje de las equis) representa la suma de las frecuencias.

El área bajo la curva normal o sea la suma de las frecuencias puede obtenerse mediante la integración de la función correspondiente:

$$\frac{N}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}_a}{\sigma}\right)^2} dx$$

El área total obtenida mediante dicha integración llevada de — infinito a + infinito sería igual a N o sea el 100 % de la distribución. El área obtenible mediante la integración de esa función entre — infinito y 0 (la media) *o* entre 0 e infinito sería igual a $N/2$ o sea, al 50 % de la distribución (puesto que la media es, al mismo tiempo la mediana de la distribución o sea, la que la divide en dos partes iguales)... etc.

Como es fácil comprender, por simple inspección de la curva normal, a porciones iguales tomadas sobre el eje de las equis (o sea, a desviaciones iguales) no corresponden porciones iguales de área (o sea, igual porciento de frecuencias) sino que, por el contrario, a una distancia de una unidad tomada de la media-mediana-modo a la derecha o a la izquierda le corresponde una porción de área mayor que la que le corresponderá a una distancia de una unidad situada más lejos de aquel valor que divide a la distribución en dos partes iguales. De este modo, todo se reduce a mostrar que la media divide a la distribución en dos partes iguales (la procedencia de la curva normal de la simétrica binomial lo hace presentir ya).

En efecto, si $\bar{x}_i$ excede a x en una cierta cantidad a (positiva), al dividirla entre σ y elevarla al cuadrado se obtendrá una cantidad positiva que multiplicada por menos $1/2$ dará para el exponente de e una cantidad negativa igual a:

$$-\frac{1}{2}\frac{(+a)^2}{\sigma^2}=-\frac{1}{2}\frac{a^2}{\sigma^2}$$

Si, en vez de que x_i exceda a $\bar{x}$; x_i es excedida por $\bar{x}$ en la cantidad a, la resta de $x_i - \bar{x}$ será igual a la cantidad a (negativa) que, dividida entre σ y elevada al cuadrado dará como resultado una cantidad positiva que multiplicada por menos $1/2$ dará para el exponente de e una cantidad negativa igual a:

$$-\frac{1}{2}\frac{(-a)^2}{\sigma^2}=-\frac{1}{2}\frac{a^2}{\sigma^2}$$

O sea, que para cada valor x_i que difiera de la media en $+a$, habrá otro valor de x_i que difiere de la media aritmética en $-a$ y que tiene asociada la misma frecuencia que el primero y, si hay tantas x_i con desviación $(x_i - \bar{x})$ frecuencias f_i $(= y_i)$ iguales a uno y otro lado de la media aritmética (de donde el signo positivo o negativo de la desviación) esto quiere decir que la media aritmética divide a la distribución en dos partes iguales; o bien, que la media aritmética es, asimismo, la mediana de la distribución normal.

Promedios Centrales de la Curva Normal.—Entre los rasgos característicos de la distribución normal, se cuenta, en primer término, el que para dicha distribución la media, la mediana y el modo coinciden. En efecto, comenzaremos por considerar la igualdad que postulamos entre la media y el modo, para ocuparnos en seguida de la que hemos postulado asimismo entre la media y la mediana, para lo cual tomaremos como punto de partida la ecuación general para la curva normal:

$$y_i = \frac{N}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}$$

Determinar el modo equivale a determinar cuál es el dato asociado con la frecuencia máxima; o sea, determinar para qué valor de x_i, toda la expresión equivalente de y_i llega a ser máxima. Si siguiendo nuestra hipótesis, hacemos que x_i sea igual a la media aritmética:

$$x_i - \bar{x} = 0$$

$$\frac{x_i - \bar{x}}{\sigma} = 0$$

$$\left(\frac{x_i - \bar{x}}{\sigma}\right)^2 = 0$$

$$\frac{1}{2}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2 = 0$$

$$e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)} = e^0 = 1$$

Y si todo el factor (e elevado a. . .) se reduce a la unidad, en la expresión para y_i sólo queda el primer factor, o sea que:

Cuando $x_i = x$,

$$y_i = \frac{N}{\sqrt{2\pi}\,\sigma}$$

Este valor es, según hemos visto en páginas anteriores, la frecuencia máxima (frecuencia relativa máxima o máxima probabilidad en el caso de la distribución simétrica binomial de la que procede la distribución normal). Si esta es la frecuencia máxima el dato $x_i = \bar{x}$ asociado con dicha frecuencia será el modo; o sea, que la media es igual al modo.

El que el valor alcanzado es máximo puede también mostrarse considerando que e elevado a la potencia negativa menos un medio... etc., equivale a tener a e en el denominador de la expresión con signo positivo (e elevado a *más* un medio. . . etc.); de este modo, siempre que $(x_i - \bar{x})/\sigma$ sea distinto de 0, o sea, siempre que $x_i - \bar{x} \neq 0$; o $x_i + \bar{x} \neq 0$, o bien $x_i \neq \bar{x}$, del exponente ($x_i - x$ sobre σ) tendrá un valor, positivo o negativo, que, al elevarse al cuadrado, resultará positivo, de modo que, en el denominador figurará el factor e elevado a dicho cuadrado con lo cual al aumentar el valor del denominador, disminuirá el valor del quebrado $N/\sqrt{2\pi}\,\sigma$ en comparación con el que el mismo quebrado tiene cuando dicho factor e elevado a menos un medio. . . se reduce a la unidad por ser $x_i = \bar{x}$. O sea, que se confirma que la media de la distribución normal está asociada a la frecuencia máxima y es, por lo mismo, el modo de la distribución.

En todo caso, lo anterior mostraría la necesidad de calcular la integral correspondiente a determinados límites en cada caso particular a no ser que el problema hubiese sido (como ha sido) resuelto mediante la tabulación de los resultados correspondientes en las llamadas "tablas de áreas bajo la curva normal".

Las tablas de áreas bajo la curva normal, consignan las porciones de área calculadas para una curva normal típica, o sea, para una curva normal:

1º—cuya suma total de frecuencias, o cuya área total es igual a 1 (o a 100),

2º—cuya media-mediana-modo es cero, y

3º—cuya desviación media cuadrática es igual a 1.

De este modo, las tablas consignan la porción de área comprendida entre la curva, la ordenada máxima (correspondiente a la media-mediana-modo o sea a una desviación de cero), el eje de las equis y la ordenada levantada en el punto cuya desviación con respecto a la media consigna la tabla como entrada. El valor consignado en el cuerpo de la tabla, en cuanto indica una cierta porción de área total igual a 1 o igual a 100 deberá interpretarse como un número relativo o como un tanto por ciento.

Así, si frente a un valor de 1 como entrada de la tabla se encuentra en el cuerpo de la misma un valor de .3413 esto deberá interpretarse como que entre la curva, la ordenada máxima correspondiente a una desviación de 0, el eje de las equis y la ordenada levantada en el punto que dista de la media-mediana-modo una unidad (de desviación media cuadrática) existen .3413 datos si se trata de una distribución cuya suma de frecuencias sea igual a uno (caso de frecuencias fraccionarias que puede darse en cuanto se trate de frecuencias relativas como son las que se han encontrado en el caso de las distribuciones de probabilidad), o bien, que entre esos mismos límites existen unos 34 casos, en cuanto se trate de una distribución cuyo efectivo o suma de frecuencias sea igual a 100.

Lo anterior equivale a que si buscamos cuál es el área bajo la curva comprendida entre una desviación cuadrática media y la media-mediana-modo de la distribución normal debamos buscar 1 como entrada a la tabla y encontrar .3413 en el cuerpo de la misma, interpretando este valor en la forma en que lo hemos hecho en los renglones anteriores, *si* el efectivo de la distribución o la suma total de frecuencias es igual a 1 o igual a 100.

En caso de que el efectivo de la distribución no sea 1 o 100, lo anterior equivaldrá a que en cuanto buscando el área bajo la curva comprendida entre una desviación cuadrática media y la media-mediana-modo de la distribución normal encontremos .3413 debamos interpretar este resultado en el sentido de que el 34.13 % de esa distribución se encontrará entre esos límites o bien, que multiplicando el efectivo de la distribución por .3413 se encontrará cuántos datos están comprendidos entre una desviación media cuadrática y la media-mediana-modo de la distribución.

El significado de estas determinaciones se pone de manifiesto en cuanto los límites (número de desviaciones media cuadráticas y media de la distribución) se expresan en unidades originarias. Así, si la media de una distribución es 90 centímetros y la desviación media cuadrática 10 centímetros, el problema de determinar cuál es el área bajo la curva comprendida entre una desviación media cuadrática y la media-mediana-modo de la distribución se convierte en el más concreto de determinar cuántos casos de la distribución (área bajo la curva) son superiores a 90 centímetros (media-mediana-modo de la distribución) e inferiores a 100 (90 de la media más 10 de la desviación cuadrática media).

De este modo, si el problema que se nos plantea es el de determinar cuántos casos de una distribución tienen una magnitud superior a 90 e inferior a 100 y sabemos que 90 es la media aritmética de la distribución, que 10 es la desviación cuadrática media de la misma y que la distribución es normal, el problema equivaldrá a plantearse cuál es el área bajo la curva comprendida entre cero (90 — 90) y una $\left(\frac{100-90}{10}\right)$ desviación media cuadrática, o sea, a buscar 1 (una desviación media cuadrática de apartamiento de la media) como entrada de la tabla y a encontrar .3413 como valor del área bajo la curva en el cuerpo de la tabla, que se interpretará en la forma en que lo hemos venido haciendo: hay un 34 % del total de casos. Si la distribución comprende 600 casos, hay $600 \times 34.13\ \%$ o $600 \times .3413 = 274.78$ o casi 275 casos cuyo valor sea superior a 90 e inferior a 100.

En general, si se desea saber cuál es el número de casos de una distribución cuyo valor sea superior (/ inferior puesto que la distribución normal es simétrica) al de la media de la distribución pero inferior (/ superior) a un valor dado x_i, será necesario calcular cuantas desviaciones medias cuadráticas representa x_i con respecto al valor de la media $\bar{x}$ cuya desviación se considera igual a cerc, y en seguida buscar dicha desviación en la tabla a fin de determinar la correspondiente área bajo la curva.

El mecanismo sería el siguiente:

Si se quiere determinar el número de casos cuyo valor esté comprendido entre (sea inferior o superior a) la media de una distribución y (sea superior o inferior a) un valor determinado x_i en una distribución normal cuya media sea $\bar{x}$ y cuya desviación media cuadrática sea σ:

1º—Réstese de x_i el valor de la media aritmética $\bar{x}$.

2º—Divídase la diferencia (o desviación con respecto a la media aritmética) entre el valor de la desviación cuadrática media σ.

3º—Búsquese el cociente (desviación con respecto a la media aritmética en términos de desviación cuadrática media) en la entrada de la tabla de áreas bajo la curva normal y encuéntrese en el cuerpo de la misma el valor correspondiente.

4º—Multiplíquese el valor encontrado en el cuerpo de la tabla (porción de área bajo la curva o tanto por ciento del efectivo de la distribución comprendido entre la media aritmética y el valor de la desviación... etc.) por el efectivo de la distribución de que se trate, a fin de obtener el total de casos entre el valor de la media $\bar{x}$ y el valor de que se trate x_i.

Todo lo anterior es válido cuando uno de los límites es la media aritmética de la distribución. El problema que se plantea en seguida, es el de determinar el número de casos comprendidos entre ciertos valores cuando ninguno de ellos es la media-mediana-modo de la distribución normal.

Este problema presenta varias posibilidades:

1ª—Se trata de buscar cuál es el número de casos comprendidos entre un valor x_1 de una distribución que es inferior al valor de la media aritmética y el valor x_2 que es superior a la media aritmética.

Solución: Calcúlese el % de la distribución (la porción de área bajo la curva) comprendida entre x_1 y la media $\bar{x}$ conforme al procedimiento anterior; calcúlese el % de la distribución comprendida entre la $\bar{x}$ y x_2 conforme a ese mismo procedimiento; súmense ambos %s.

Multiplíquese la suma por el efectivo de la distribución.

2ª—Se trata de encontrar el número de casos inferiores al valor de x_1, siendo x_1 a su vez un valor inferior al de la media aritmética.

Solución: Puesto que el % de la distribución comprendido entre la media-*mediana*-modo y el extremo inferior de la misma es 50 %,

Réstese de 50 %, o de .50 la porción comprendida entre x_1 y la media que se calculará como en el caso general (restando de x_1 la media, dividiendo entre la desviación cuadrática media y buscando dicho valor en la tabla de áreas bajo la curva).

3ª—Se trata de encontrar el número de casos superiores al valor de x_2 que es a su vez superior al valor de la media.

Solución: Análoga a la anterior: habrá necesidad de restar de 50 % (0.50) el % o porción de área comprendida entre x_2 y la media.

4ª—Se trata de encontrar el número de casos comprendidos entre x_1 y x_2 cuando ambos valores son inferiores (o superiores) a la media.

Solución: Encuéntrese el porciento de casos comprendidos entre x_2 (que supondremos el más alejado de la media) y la media-mediana-modo de la distribución;

Encuéntrese el porciento de casos comprendidos entre x_1 y esa misma media-mediana-modo;

Réstese del primer porciento el segundo porciento;

Multiplíquese el residuo porcentual por el efectivo de la distribución para obtener el total de casos comprendidos entre x_2 y x_1.

5ª—Se trata de encontrar el número de casos de una distribución normal que sean inferiores a x_1 (que a su vez es inferior a la media) y superiores a x_2 (que a su vez es superior a la media).

Solución: Determínese el número de casos comprendidos entre x_1 y x_2 (situados a uno y otro lado de la media-mediana-modo) en la forma indicada para la primera posibilidad, y réstese de 100 % (o de 1), y multiplíquese el residuo por el efectivo de la distribución.

6ª—Obtener el número de casos superiores a cierto valor x_1 que es a su vez inferior a la media.

Solución: Obténgase el % de casos comprendidos entre x_1 y la media-mediana-modo y súmesele el 50 % (.50) correspondiente a la mitad superior de la distribución.

7ª—Obtener el número de casos cuyo valor es inferior a cierto valor x_2 que es a su vez superior a la media.
Solución: Obténgase el % de casos comprendidos entre x_2 y la media-mediana-modo y súmesele el 50 % (.50) correspondiente a la mitad inferior de la distribución.

ADAPTACIÓN DE LA CURVA NORMAL POR EL MÉTODO DE LAS ÁREAS

El raciocinio que nos puede permitir establecer un procedimiento de adaptación de la curva normal a una distribución empírica, valiéndonos de las áreas bajo la curva normal, puede seguirse resolviendo los problemas siguientes:

1º—¿Cómo obtener las frecuencias teóricas correspondientes a los diferentes valores de la variable (a los diferentes intervalos de clase) de una distribución cuyo efectivo (o cuya suma total de frecuencias es Σf_i) si se conocen las frecuencias teóricas respectivas de una distribución cuyo efectivo es igual a 1 o a 100?

Solución.—Multiplicar el efectivo de la distribución que se estudia por las frecuencias teóricas correspondientes a la distribución unitaria (cuyo efectivo es igual a 1 o a 100).

2º—¿Cómo encontrar las frecuencias teóricas correspondientes a cada uno de los valores de la variable (o a cada una de las clases) de la distribución unitaria si lo que se conoce por las tablas son las frecuencias comprendidas entre cada uno de esos valores (o de los límites inferiores y superiores de los intervalos) y la media aritmética?

Solución.—Restar de la frecuencia correspondiente a la desviación de cada valor de la variable con respecto a la media aritmética, la frecuencia correspondiente a la desviación del valor de la variable inmediato y más cercano a la media aritmética con respecto a ésta (de cada valor el siguiente para los inferiores a la media; de cada valor el anterior para los superiores a la media). En las series de clases y frecuencias, restar de la frecuencia comprendida entre el límite inferior de una clase y la media, la frecuencia comprendida entre el límite inferior de la clase siguiente y la media cuando dichas clases preceden a la clase que contiene a la media aritmética; restar de la frecuencia comprendida entre el límite superior de una clase y la media, la frecuencia comprendida entre el límite superior de la clase precedente y la media, cuando dichas clases subsiguen a la clase que contiene a la media aritmética; sumar a la frecuencia comprendida entre el límite inferior de la clase y la media aritmética, la frecuencia comprendida entre el límite superior de la clase y la media aritmética cuando se trata de la clase que contiene a la media aritmética.

3º—¿Cómo calcular las frecuencias comprendidas entre ciertos valores (entre ciertos límites de clase) de la variable y la media aritmética?

Solución.—Buscar en la "tabla de *áreas* bajo la curva" el valor correspondiente a tales valores (de tales límites de clase) expresados como "desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media".

4º—¿Cómo encontrar la equivalencia de los valores originales en términos de "desviaciones con respecto a la media aritmética en unidades de la desviación cuadrática media"?

Solución.—Restar de los valores originales de x_i el valor de la media aritmética y dividir los resultados entre la desviación cuadrática media. En caso de serie de clases y frecuencias, restar de los valores de los límites inferiores de las clases anteriores a las que contienen a la media aritmética y de la clase que la contiene, y de los límites superiores de las clases posteriores a la que contienen a la media aritmética y de la clase que la contiene, el valor de la media aritmética y dividir dichos valores entre la desviación cuadrática media.

El procedimiento que (como reflejo en un espejo del raciocinio anterior) nos permite ajustar una curva normal a una distribución empírica por el método de las áreas, reproduce, en sentido inverso, los pasos del raciocinio; en efecto, para hacer dicho ajuste se necesita:

1º—Calcular las desviaciones con respecto a la media aritmética en términos de la desviación cuadrática media:

a.—Restando de los valores de la variable (límites inferiores para las clases que contienen o son anteriores a la que contiene a la media, límites superiores para las que la contienen o son anteriores a la que contiene a la media en series de clases y frecuencias), el valor de la media, y

b.—Dividiendo el resultado de dicha resta (desviación con respecto a la media aritmética) entre la desviación cuadrática media de la distribución.

2º—Determinar la porción de área (% de frecuencia) comprendido entre las desviaciones con respecto a la media en unidades sigmáticas y la media.

a.—Buscando en la entrada de la tabla de *áreas* bajo la curva normal el valor de las desviaciones, *b.*—Encontrando en el cuerpo de la tabla, los valores de área correspondientes.

3º—Determinar la porción de área (% de frecuencias) correspondiente a cada valor (a cada clase).

a.—Restando del área comprendida entre la desviación correspondiente a un límite inferior de clase que preceda a la que contiene a la media aritmética y dicha media, el área comprendida entre el límite inferior de la clase siguiente y la media.

b.—Restando del área comprendida entre el límite superior de clase que subsiga a la que contiene a la media aritmética y dicha media, el área comprendida entre el límite superior de la clase precedente y la media.

c.—Sumando al área comprendida entre el límite inferior de la clase que contenga a la media aritmética y dicha media, el área comprendida entre el límite superior de la clase que contenga a la media aritmética y dicha media.

4º—Cálculo de las frecuencias teóricas de la distribución no unitaria:

a.—Multiplicando los valores anteriores (% de frecuencias) por el efectivo (o suma total de frecuencias de la distribución empírica que se estudia).

EQUIVALENCIAS DESIGNATIVAS DE RASGOS DE LA CURVA NORMAL Y CARACTERES ASOCIADOS

Totalidad de los casos.
Efectivo de la distribución.
Suma de las frecuencias.
Área total bajo la curva.
Integral de la función entre menos y más infinito.
Cien por ciento de las frecuencias.
Oscilación o Amplitud como distancia de Mínimo a Máximo medida sobre el eje de las abscisas o eje de las equis. De -3σ a $+3\sigma$ aproximadamente, a pesar de que la curva se prolonga de $-$infinito a $+$infinito, siendo asintótica.
Datos o valores de la variable como abscisas.
Frecuencias como ordenadas.

Mitad de los casos.
Mitad de la distribución.
Suma de las frecuencias divididas entre dos.
Mitad del área bajo la curva.
Integral de la función entre $\mp$, infinito y cero.
Cincuenta por ciento de las frecuencias.
Distancia entre el Mínimo/Máximo y la media medida sobre el eje de las abscisas. De $\mp 3\,\sigma$ a 0.
Datos o valores de la variable como abscisas, la media-mediana-modo, como abscisa central entre las correspondientes al máximo y al mínimo.
Frecuencias como ordenadas, la frecuencia máxima o modal como ordenada máxima, asociada a la media-mediana-modo.

Casos normales.
Mitad de la distribución que menos se aparta de la media.
Suma de las frecuencias de los datos que menos se apartan de la media, y de los que la mitad son inferiores a la media y la otra mitad superiores a la media.
Mitad del área bajo la curva distribuída por igual a un lado y otro de la ordenada correspondiente a la media.
Integral de la función que se iguala a 0.5 o 50 % para ciertos valores $-x$ y $+x$.
Veinticinco por ciento de las frecuencias de datos inferiores a la media, y
Veinticinco por ciento de las frecuencias de datos superiores a la media.
Distancia entre la primera y la tercera cuartilas medida sobre el eje de las abscisas o eje de las equis.
De -0.6745 a $+0.6745\,\sigma$.

AJUSTE DE CURVAS NO NORMALES. SISTEMA DE CURVAS DE GRAM-CHARLIER

Cuando el cálculo de las medidas de asimetría y de curtosis de una distribución muestran que la misma se aparta mucho de las condiciones de normalidad y que, por lo mismo resultaría abusivo el ajustar una curva normal a dicha distribución, es necesario recurrir a un sistema más general de curvas que permita ajustamientos para distribuciones cuya asimetría sea distinta de cero y cuya curtosis sea distinta de 3 (o cuyo exceso de curtosis sea diferente de cero). Uno de dichos sistemas es el de las curvas de Gram-Charlier cuyas ordenadas, para propósitos prácticos pueden considerarse representadas por la ecuación:

$$y_t = \Sigma f_i \frac{i}{\sigma} \left[\varphi_o(\delta_i) - \frac{\sqrt{\beta_1}}{6} \varphi_3(\delta_i) + \frac{\beta_2 - 3}{24} \varphi_4(\delta_i) \right]$$

En esta expresión, $\varphi_0(\delta_i)$ representa la exponencial e elevada a menos un medio de δ_i al cuadrado, dividida entre raíz cuadrada de 2π, o sean las ordenadas de la curva normal en unidades de la raíz cuadrada de 2π, dadas por la tabla; $\varphi_3(\delta_i)$ representa la tercera derivada de las ordenadas y $\varphi_4(\delta_i)$ la cuarta derivada, también tabuladas en forma semejante.

Conforme a lo anterior, podemos considerar a los dos últimos términos de dentro del paréntesis como correcciones por asimetría y por curtosis distintas de las normales; en efecto, cuando la asimetría es igual a cero, $\sqrt{\beta_1}$ vale cero y, todo el término en el que figura se anula; en forma semejante, cuando la curva es mesocúrtica el exceso de asimetría $(\beta_2 - 3)$ vale cero y el término en el que figura como factor se anula, con lo cual al ser la curva simétrica y mesocúrtica (es decir normal), volvemos a tener la expresión con la que ya hemos trabajado en el ajuste de la curva normal:

$$y_t = \Sigma f_i \frac{i}{\sigma} \varphi_0(\delta_i)$$

El ajuste de una curva de Gram-Charlier se hará conforme al siguiente procedimiento:

1º—Se calcularán como en el caso de ajuste de una curva normal, las desviaciones con respecto a la media aritmética en unidades de la desviación media cuadrática, por cualquiera de los dos procedimientos ya conocidos:

A.—(Directo)

a.—restando de cada punto medio la media aritmética y

b.—dividiendo los resultados entre la desviación media cuadrática o.

B.—(Indirecto)

a.—restando de cada desviación con respecto a la media arbitraria en unidades del intervalo, el primer momento con respecto a dicha media en dichas unidades,

b.—obteniendo el cociente del intervalo entre la desviación cuadrática media y

c.—multiplicando por dicho cociente los residuos obtenidos en *a.*

2º—Se buscan en la tabla y se anotan en tres columnas los valores de las ordenadas (φ_0), las terceras y las cuartas derivadas de las ordenadas (φ_3 y φ_4) contenidos en el cuerpo de la tabla frente a los valores de las desviaciones con respecto a la media aritmética en unidades de la desviación media cuadrática dadas en el inciso anterior.

3º—Se divide el índice de asimetría $\sqrt{\beta_1}$ que deberá haberse obtenido previamente, entre 6 y se multiplicará el cociente por los valores obtenidos para φ_3.

4º—Se divide el exceso de curtosis ($\beta_2 - 3$) entre 24 y se multiplicará el cociente por los valores obtenidos para φ_4.

Precaución en cuanto a signos.—Antes de proseguir con el procedimiento, debe tenerse cuidado:

a.—de copiar fielmente los signos de φ_3 y de φ_4 dados por la tabla

b.—de cambiar los signos de φ_3 (exclusivamente) cuando las desviaciones sean negativas.

c.—de recordar que el coeficiente de φ_3 es negativo y, por lo tanto se necesitará:

de un segundo cambio de signos para los valores correspondientes a desviaciones negativas,

de un primer cambio de signos para los valores correspondientes a desviaciones positivas.

5º—Se suman los valores

a.—de las ordenadas,

b.—de los productos de las terceras derivadas (φ_3) por su coeficiente $\left(\frac{\sqrt{\beta_1}}{6}\right)$ y

c.—de los productos de las cuartas derivadas (φ_4) por su coeficiente $\left(\frac{\beta_2 - 3}{24}\right)$ teniendo cuidado de hacer esa suma algebráicamente, o sea, teniendo en cuenta los signos.

6º—Calcúlese el producto del efectivo de la distribución (o suma de las frecuencias) por el cociente que resulta de dividir el intervalo entre la desviación media cuadrática.

7º—Multiplíquese el producto obtenido en *6º*, por los valores de las sumas algebraicas obtenidas en *5º*. Los resultados, contenidos en una última columna serán los valores teóricos de ajuste de la curva de Gram-Charlier correspondiente.

MOMENTOS DE UNA DISTRIBUCIÓN BIVARIADA

Los momentos de una distribución bivariada se definen como las medias aritméticas de los productos de las potencias de las desviaciones de cada una de las dos variables con respecto a dos valores constantes correspondientes.

Como en el caso de las distribuciones univariadas, puede distinguirse entre:

1º—Momentos con respecto a una media arbitraria, y
2º—Momentos con respecto a la media aritmética.

En el primer caso, se trata de medias aritméticas de los productos de las potencias de las desviaciones de cada variable con respecto a un valor arbitrariamente elegido para cada una de ellas; en el segundo caso, de medias aritméticas de los productos de las potencias de las desviaciones de cada variable con respecto a la media artimética correspondiente.

Los diferentes momentos de una distribución bivariada se distinguen por índices compuestos de dos cifras, cada una de las cuales indica la potencia a que es preciso elevar las desviaciones de cada una de las variables para obtener el momento correspondiente; así, si el índice del momento es 11, se tratará de la media de las primeras potencias de las desviaciones tanto de la primera como de la segunda variable; si el índice es 12 se tratará de la media de los productos de la primera potencia de las desviaciones de la primera variable por la segunda potencia de las desviaciones de la segunda, si el índice es 21, de la media de los productos de la segunda potencia de la primer variable por la primera potencia de las desviaciones de la segunda. Un índice 01 indicará la media de los productos de las potencias cero de las desviaciones de la primera, por las primeras potencias de las desviaciones de la segunda variable, o sea, que propiamente será el primer momento de la distribución univariada de la segunda variable; 10 indicará el primer momento de la distribución univariada de la primera variable; 02, el segundo momento de la distribución univariada de la segunda, etc. En general.

$$\text{MOMENTO}_{rs} = \frac{\Sigma(\text{desviación } x_i)^r\ (\text{desviación } x_2)^s}{N}$$

Cuando se trate de series de frecuencias:

$$\text{MOMENTO}_{rs} = \frac{\Sigma(\text{desviación } x_1)^r\ (\text{desviación } x_2)^s f_{12}}{\Sigma\, f_{12}}$$

Expresión en la que f_{12} representa las frecuencias de la distribución conjunta o sea, el número de casos que teniendo determinado valor para x_1 tienen un valor determinado de x_2.

Como en el caso de las distribuciones univariadas, en el de las bivariadas, si se desea obtener el equivalente de los momentos con respecto a la media aritmética en momentos con respecto a una media arbitraria, habrá necesidad de substituir el valor de las desviaciones con respecto a la media aritmética por su equivalente en desviaciones con respecto a la media arbitraria, desarrollar las potencias correspondientes y hacer reducciones. Buscaremos la equivalencia sólo para el momento de orden 11 con respecto a la media aritmética en unidades del intervalo.

$$M_{11} = \frac{\Sigma\left[\left(\frac{x_1 - \bar{x}_1}{i}\right)\left(\frac{x_2 - \bar{x}_2}{j}\right)\right]}{\Sigma f_{12}}$$

Si se expresan las desviaciones con respecto a la media aritmética en unidades del intervalo como desviaciones con respecto a la media arbitraria en esas mismas unidades se tendrá (expresando la media aritmética como suma de la media arbitraria más el primer momento de la distribución —univariada— en términos del intervalo por el intervalo):

$$\frac{x_1 - \bar{x}_1}{i} = \frac{x_1 - x'_1 - M'_{10}i}{i} = \frac{x_1 - x'_1}{i} - M'_{10} = D'_1 - M'_{10}$$

$$\frac{x_2 - \bar{x}_2}{j} = \frac{x_2 - x'_2 - M'_{01}i}{j} = \frac{x_2 - x'_2}{j} - M'_{01} = D'_2 - M'_{01}$$

Substituidos estos valores, tendremos:

$$M_{11} = \frac{\Sigma(D'_1 - M'_{10})(D'_2 - M'_{01})f_{12}}{\Sigma f_{12}}$$

$$M_{11} = \frac{\Sigma D'_1D'_2f_{12} - M'_{01}\,\Sigma D'_1f_{12} - M'_{10}\,\Sigma D'_2f_{12} - \Sigma M'_{10}M'_{01}f_{12}}{\Sigma f_{12}}$$

Pero como $\Sigma f_{12} = \Sigma f_1 = \Sigma f_2$, se tienen las siguientes equivalencias:

$$\frac{\Sigma D'_2f_{12}}{\Sigma f_{12}} = \frac{\Sigma D'_2f_2}{\Sigma f_2} = M'_{01}$$

$$\frac{\Sigma D'_1f_{12}}{\Sigma f_{12}} = \frac{\Sigma D'_1f_1}{\Sigma f_1} = M'_{10}$$

Y, puesto que:

$$\frac{\Sigma\, f_{12}M'_{10}M'_{01}}{\Sigma\, f_{12}} = M'_{10}M'_{01}$$

La expresión anterior se convertirá en:

$$M_{11} = \frac{\Sigma\, D'_1D'_2f_{12}}{\Sigma\, f_{12}} - M'_{01}M'_{10} - M'_{10}M'_{01} + M'_{10}M'_{01}$$

Los dos últimos términos de la expresión anterior se anulan. En cuanto al primer término de la misma es la suma (ponderada) de los productos de las desviaciones de cada variable con respecto a sendas medias arbitrarias en unidades de los intervalos correspondientes, o sea la M'_{11} con lo cual la expresión anterior se transforma en:

$$M_{11} = M'_{11} - M'_{01}M'_{10}$$

Para obtener el momento de orden 11 con respecto a la media aritmética en unidades originales, sería necesario multiplicar M_{11} por el producto ij de los intervalos. Sin embargo, este último paso no es por lo general necesario.

$$\mu_{11} = ijM_{11}$$

ÍNDICE DE CORRELACIÓN

En el estudio de las series bivariadas importa conocer el grado o la intensidad con que las variaciones de uno de los caracteres o fenómenos estudiados (las magnitudes de una de las variables) siguen las variaciones del otro carácter o fenómeno estudiado (las magnitudes de la otra variable). Dicho grado o intensidad se mide mediante el momento 11 con respecto a las medias aritméticas de las variables o sea, por medio de la media de los productos de las desviaciones de cada variable con respecto a su media aritmética:

$$M_{11} = \frac{\Sigma \left[\left(\frac{x_1 - x_1}{i} \right) \left(\frac{x_2 - \bar{x}_2}{j} \right) f_{12} \right]}{\Sigma\, f_{12}}$$

o $$\mu_{11} = \frac{\Sigma (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) f_{12}}{\Sigma\, f_{12}}$$

A fin de reducir esta medida de relación a términos abstractos y comparables de serie a serie, se acostumbra expresar dicha medida en unidades de las respectivas desviaciones cuadráticas medias, dividiendo el momento 11 entre el producto de la desviación media cuadrática de la primer variable por la desviación media cuadrática de la segunda variable. El cociente de la división, r se conoce como "índice de correlación". Si se trabaja con M_{11} momento 11 con respecto a las medias aritméticas en unidades de los intervalos, dicho momento deberá dividirse entre las desviaciones medias cuadráticas de las variables también tomadas en términos de los respectivos intervalos ($\overline{D}_{q1}$ $\overline{D}_{q2}$ o s_1 y s). Si se trabaja con μ_{11}, momento 11 con respecto a las medias aritméticas en unidades originales, dicho momento deberá dividirse entre

($\overline{d}_{q1}$, $\overline{d}_{q2}$ o sea σ_1 y σ_2)

Según esto, las fórmulas para r serán:

$$r = \frac{M_{11}}{s_1 s_2}$$

$$r = \frac{\mu_{11}}{\sigma_1 \sigma_2}$$

CÁLCULO DIRECTO DEL ÍNDICE DE CORRELACIÓN

Conforme a la última de las fórmulas dadas, el índice de correlación requiere, para su cálculo, del siguiente procedimiento:

1º—Consígnense las frecuencias combinadas (f_{12}) en un cuadro de doble entrada cuyos encabezados columnares serán los valores (o los puntos medios) de la primer variable, y cuyos encabezados de las hileras serán los valores (o los puntos medios) de la segunda variable.

2º—Súmense las frecuencias de cada columna para determinar las frecuencias de la primer variable, y súmense las frecuencias de cada renglón para determinar las frecuencias de la segunda variable.

3º—Calcúlese la media de la primer variable.

A.—Multiplicando cada valor o punto medio de dicha variable por la frecuencia correspondiente y anotando dichos productos en un renglón,

B.—Sumando todos los productos así obtenidos, y

C.—Dividiendo la suma entre el total de las frecuencias.

4º—Calcúlese en forma análoga (usando las columnas), la media aritmética de la segunda variable.

5º y *6º*—Determínense en los renglones y en las columnas, respectivamente, las desviaciones con respecto a la media aritmética,

A.—Restando de cada valor o punto medio de la primer variable la media aritmética de dicha variable y anotando los resultados en un renglón, y

B.—Restando de cada valor o punto medio de la segunda variable la media aritmética correspondiente, anotando los resultados en una columna.

7º—Obténganse los productos de cada frecuencia conjunta (f_{12} de las casillas del cuadro) por la desviación con respecto a la media aritmética contenida en la misma columna y por la desviación con respecto a la media aritmética contenida en el mismo renglón a que pertenezca la casilla, y anótese el producto de estas tres cantidades entre paréntesis en la casilla correspondiente. Súmense todas las cantidades entre paréntesis de todos y cada uno de los renglones anotando los resultados en una columna cuyos valores se sumarán al pie para obtener el numerador del índice de correlación.

8º—Obténganse las desviaciones medias cuadráticas de las dos variables. Para ello:

A.—Elévense al cuadrado las desviaciones obtenidas en *5º* y *6º* anotando los resultado en un renglón y una columna.

B.—Súmense los cuadrados de las desviaciones de cada una de las variables con respecto a sus correspondientes medias (por separado).

C.—Divídanse las sumas así obtenidas entre la suma de las frecuencias.

D.—Extráiganse las raíces cuadradas de dichos cocientes. Los resultados son las desviaciones medias cuadráticas de cada una de las variables.

9º—Multiplíquense las desviaciones medias cuadráticas de las dos medias entre sí, y por la suma de las frecuencias. El resultado es el denominador del valor del índice.

10º—Divídase el valor obtenido en *7º* entre el valor obtenido en *9º*.—El resultado es el índice de correlación *r*.

CÁLCULO SIMPLIFICADO DEL ÍNDICE DE CORRELACIÓN

El cálculo simplificado del índice de correlación hace uso de las siguientes igualdades:

$$M_{11} = M'_{11} - M'_{10}M'_{01}$$

$$s_1 = \sqrt{M'_{20} - (M'_{10})^2}$$

$$s_2 = \sqrt{M'_{02} - (M'_{01})^2}$$

El procedimiento es el siguiente:

1º—Consígnense las frecuencias conjuntas (f_{12}) en un cuadro de doble entrada cuyos encabezados columnares serán los valores (o los puntos medios) de la primer variable, y cuyos encabezados de las hileras serán los valores (o los puntos medios) de la segunda variable.

2º—Súmense las frecuencias de cada columna para determinar las frecuencias de la primer variable, y las frecuencias de cada renglón para determinar las frecuencias de la segunda variable.

3º—Calcúlese el momento 11 con respecto a las medias arbitrarias en unidades de los intervalos de las variables:

A.—Eligiendo de entre los puntos medios de cada una de las variables sendos valores a los que se considerará como medias arbitrarias,

B.—Colocando:
0 frente a cada una de las medias arbitrarias —1, —2, —3 . . ., —n hacia arriba o hacia la derecha, +1, +2, +3 . . . , +n hacia abajo o hacia la izquierda,

C.—Multiplicando cada una de las frecuencias conjuntas (f_{12}) por las desviaciones calculadas en *B* que se encuentren en su misma columna y en su mismo renglón, y anotando el producto entre paréntesis en la casilla correspondiente a la f_{12} con que se haya trabajado,

D.—Sumando todas las cantidades entre paréntesis contenidas en cada uno de los renglones, anotando las sumas en una columna cuyos valores se sumarán al pie para obtener el momento 11 con respecto a las medias arbitrarias en unidades de los intervalos (M'_{11}).

4º—Obténganse los primeros momentos con respecto a la respectiva media arbitraria en unidades del respectivo intervalo tanto de la primera como de la segunda variable; para ello:

A.—Multiplíquense las desviaciones obtenidas en *3º B* por las frecuencias de la distribución univariada de la primer variable (f_1) colocadas en un renglón, en tratándose de las desviaciones de esa variable con respecto a su media arbitraria, y las desviaciones de la segunda variable con respecto a su media arbitraria por las frecuencias de la distribución univariada de la segunda variable (f_2) colocadas en una columna (según quedó expresado en *2º*),

B.—Súmense separadamente los valores de los productos obtenidos para la primer variable y, por otra parte, los productos obtenidos para la segunda,

C.—Divídanse las dos sumas obtenidas en *B* entre la suma total de frecuencias a fin de obtener el primer momento de la primer variable con respecto a su media arbitraria en unidades de su intervalo (M'_{10}) y el primer momento de la segunda variable con respecto a su media arbitraria en unidades de su intervalo (M'_{01}),

D.—Multiplíquense los momentos obtenidos en *C*, y

5º—Réstese el producto obtenido anteriormente, del momento 11 con respecto a las medias aritméticas en unidades de los intervalos, obtenido en *3º*, con el fin de obtener el valor del numerador para el índice de correlación ($M_{11} = M'_{11} - M'_{10}M'_{01}$).

6º—Determínense los valores de las desviaciones cuadráticas medias en unidades del intervalo para las dos variables. Para ello:

A.—Calcúlense los segundos momentos de las respectivas distribuciones univariadas

a.—multiplicando cada uno de los productos de las desviaciones por las frecuencias, por las desviaciones, a fin de obtener los productos de los cuadrados de las desviaciones por las frecuencias de las respectivas distribuciones univariadas (f_1 y f_2), anotando los resultados en un renglón y una columna,

b.—sumando separadamente los productos obtenidos en *a* para cada una de las variables, y

c.—dividiendo ambas sumas entre el total de las frecuencias, con lo cual se habrán obtenido el segundo momento con respecto a la media arbitraria de la primera variable (M'_{20}) y de la segunda, (M'_{02}), ambos en unidades de los respectivos intervalos.

B.—Calcúlense los cuadrados de los primeros momentos (M'_{10} y M'_{01}) con respecto a las medias arbitrarias en unidades de los intervalos, elevando al cuadrado los valores obtenidos en *4º C*.

C.—Réstese de cada segundo momento, el cuadrado del primer momento correspondiente (de M'_{20}, $(M'_{10})^2$, de M'_{02}, $(M'_{01})^2$).

D.—Obténganse las raíces cuadradas de los residuos obtenidos en el paso anterior. Los resultados son las desviaciones medias cuadráticas de las distribuciones univariadas de las dos variables, en unidades de sus respectivos intervalos.

7º—Multiplíquense entre sí las desviaciones medias cuadráticas de las distribuciones dadas en unidades de los intervalos a fin de obtener el denominador del índice de correlación ($\sigma_1\sigma_2$).

8º—Divídase el valor obtenido en *5º* entre el valor obtenido en *7º* a fin de obtener el valor del índice de correlación r_{12}.

Si entre dos variables x_1 y x_2 existe una relación rectilínea esto equivale a afirmar: *a.*—que la primera variable es igual al producto (o al cociente) de la segunda (x_2) por una cantidad constante (a la que llamaremos b), en el caso más sencillo, o *b.*—que la primera variable es igual al producto de la segunda por una constante, aumentado (o disminuído) de otra cantidad constante (a la que llamaremos a) en el caso más complicado. O, sea simbólicamente, que si entre las variables x_1 y x_2 hay una relación rectilínea y a y b son cantidades constantes, puede escribirse:

$$x_1 = a + b\,x_2$$

A fin de encontrar la relación, que liga a las desviaciones de cada una de las variables con respecto a sus medias aritméticas, haremos las siguientes consideraciones:

Si de los dos miembros de una ecuación se resta una cantidad constante (por ejemplo, la media de la primera variable: $\bar{x}_1$), la ecuación no se altera:

$$x_1 - \bar{x}_1 = a - \bar{x}_1 + bx_2$$

Pero como $x_1 - \bar{x}_1$ por definición, representa a la desviación (diferencia entre cada x_1 y la media aritmética de las $\bar{x}_1$ (representable por d_1) y la diferencia $a - \bar{x}_1$ es una constante que podremos representar por a_1, tendremos:

$$d_1 = a_1 + bx_2$$

Si, basándonos en la misma propiedad, restamos de ambos miembros la media de las x_2 o sea, de la segunda variable, la ecuación no se alterará:

$$d_1 - b\bar{x}_2 = a_1 + bx_2 - b\bar{x}_2$$

Si se saca como factor común a b en los dos últimos términos del segundo miembro:

$$d_1 - b\bar{x}_2 = a_1 + b(x_2 - \bar{x}_2)$$

Pasando a $-b\bar{x}_2$ al otro miembro con signo contrario:

$$d_1 = a_1 + b\bar{x}_2 + b(x_2 - \bar{x}_2)$$

Pero los términos del paréntesis representan, por definición, las desviaciones de la segunda variable con respecto a su media aritmética, que se pueden representar por d_2 y, por otra parte, $a_1 + b\bar{x}_2$ es una suma de cantidades constantes, o sea, una constante que podemos representar por a_{12} (la elección de cuyos subíndices es arbitraria, buscándose con ella sólo indicar el camino recorrido)

$$d_1 = a_{12} + b\,d_2$$

Esta expresión muestra, en una primera aproximación, la relación que liga a las desviaciones de las dos variables con respecto a sus medias respectivas. Mostraremos en seguida, que esta expresión es simplificable; para ello tomaremos el momento de orden cero de toda la expresión (lo que equivale a tomar las sumas de cada término y dividirlas entre el efectivo):

$$\frac{\Sigma\, d_1}{N} = \frac{\Sigma\, a}{N} + b\,\frac{\Sigma\, d_2}{N}$$

En la expresión anterior, $\Sigma\, a$ es la suma de una constante, que es igual a tantas veces la constante como sea el número de veces que se le tome como sumando (es decir, es igual a Na). Como $\Sigma\, a$, igual a Na está dividida entre N, el primer término se reduce a a.

$$\frac{\Sigma\, \mathrm{d}_1}{N} = a_{12} + b\,\frac{\Sigma\, d_2}{N}$$

Si se despeja a a, se tiene:

$$a_{12} = \frac{\Sigma d_1}{N} - b\,\frac{\Sigma d_2}{N}$$

Pero, las dos fracciones que figuran en la expresión son los primeros momentos de cada variable con respecto a su media, y, según hemos visto anteriormente, el primer momento de una variable con respecto a su media es cero; o sea:

$$a_{12} = 0 + b.0 = 0 + 0 = 0$$

Siendo a_{12} igual a 0, la expresión anteriormente encontrada para la relación entre las desviaciones se convierte en

$$d_1 = b\, d_2$$

Si tomamos el primer momento de esta expresión considerando como factor de ponderación a d_2, tendremos:

$$\frac{\Sigma\, d_1 d_2}{N} = b\,\frac{\Sigma d_2^2}{N}$$

Despejando a b, se tendrá:

$$b = \frac{\Sigma\, d_1 d_2 \cdot N}{N \cdot \Sigma\, d_2^2}$$

Las Ns. del numerador y denominador se reducen a la unidad, con lo cual resulta:

$$b = \frac{\Sigma d_1 d_2}{\Sigma d_2^2}$$

Pero, como:

$$r = \frac{\Sigma d_1 d_2}{N\, \sigma_1 \sigma_2}$$

si en esta expresión se despeja a $\Sigma d_1 d_2$ pasando el denominador al otro miembro como factor, se tiene:

$$\Sigma\, d_1 d_2 = N\, r\, \sigma_1 \sigma_2$$

Valor de $\Sigma\, d_1 d_2$ que podemos substituir en el valor de b

$$b = \frac{N\, r\, \sigma_1 \sigma_2}{\Sigma\, d_2^2}$$

De la expresión anterior podemos tomar conjuntamente la N del numerador y la $\Sigma\, d_2^2$ o su recíproco, $\Sigma\, d_2^2$ entre N, pudiendo demostrarse que:

$$\frac{\Sigma\, d_2^2}{N} = \left(\sqrt{\frac{\Sigma\, d_2^2}{N}}\right)^2 = \sigma_2^2$$

O bien que:

$$\frac{N}{\Sigma\, d_2^2} = \frac{1}{\sigma_2^2}$$

O sea, que los dos valores que se han tomado conjuntamente pueden substituirse por σ_2^2 en el denominador de b:

$$b = \frac{r\,\sigma_1\sigma_2}{\sigma_2^2}$$

Pero σ_2 en el numerador y σ_2^2 en el denominador, se reducen a σ_2 en el denominador:

$$b = \frac{r\,\sigma_1}{\sigma_2}$$

Como b es el factor constante por el que es necesario multiplicar las desviaciones de la segunda variable para obtener las desviaciones de la primera, y hay otro factor constante por el que será necesario multiplicar las desviaciones de la primera variable para obtener las de la segunda, al que por conveniencia también designamos por b, los distinguiremos por medio de subíndices; de este modo:

$$b_{12} = r_{12}\frac{\sigma_1}{\sigma_2}$$

En forma análoga a como se hizo para b_{12}, podremos obtener para el otro coeficiente de regresión b_{21}:

$$b_{21} = r_{12}\frac{\sigma_2}{\sigma_1}$$

Obtenible mediante un cambio en el orden de los subíndices de todas las expresiones que intervienen exceptuada r, ya que r_{21} es igual a r_{12}.

Propiedad.—El coeficiente de correlación de dos variables (r_{12}) es la media geométrica del coeficiente de regresión de cada una de las variables en la otra (b_{12} y b_{21}).

En efecto:

La media geométrica de dos cantidades es el antilogaritmo del cociente de la suma de los logaritmos de esas cantidades dividida por dos, o bien:

La media geométrica de dos cantidades es la raíz cuadrada del producto de esas cantidades:

$$\sqrt{b_{12}\cdot b_{21}} = \sqrt{r_{12}\frac{\sigma_1}{\sigma_2}\,r_{12}\frac{\sigma_2}{\sigma_1}}$$

En el gran producto del subradical, la σ_1 del numerador de la primera fracción se reduce a la unidad con la σ_1 del denominador de la segunda fracción, y lo mismo ocurre con la σ_2 del denominador de la primera con el numerador de la segunda fracción, o sea, que queda sólo como cantidad subradical r_{12} tomada dos veces como factor, o sea r_{12} elevado al cuadrado. Como se trata de extraer el cuadrado de un cuadrado, el resultado final es r_{12}, según se trataba de demostrar:

$$\sqrt{b_{12} \cdot b_{21}} = \sqrt{r_{12}^2} = r_{12}$$

Era fácil esperar este resultado teniendo en cuenta que las *bes* están formadas por un factor común r_{12} y por sendos factores fraccionarios que son recíprocos uno del otro.

De la propiedad anterior puede obtenerse un método de cálculo del coeficiente de correlación a partir de los coeficientes de regresión.

PROCEDIMIENTO DE CÁLCULO DEL COEFICIENTE DE CORRELACIÓN MEDIANTE EL CÁLCULO PREVIO DE LOS COEFICIENTES DE REGRESIÓN

1º—Calcúlense las desviaciones de cada variable con respecto a su media aritmética correspondiente y anótense dichas desviaciones d_1 y d_2 en sendas columnas, o en columna y renglón si se trabaja con una tabla de doble entrada.

2º—Multiplíquese cada desviación de una de las variables por la correspondiente de la otra y súmense los productos, o tómese cada frecuencia conjunta (f_{12}) de la tabla de doble entrada y multiplíquese por las desviaciones de ambas variables que se encuentren en la misma columna y renglón que ella; anótense los productos de estos tres factores en una columna y súmense, para obtener el numerador (común) de los coeficientes de regresión.

3º—Elévense separadamente al cuadrado las desviaciones de cada variable con respecto a su media aritmética y súmense los cuadrados; para obtener los denominadores (diferentes) de los coeficientes de regresión.

4º—Divídase el numerador común obtenido en *2º* entre los denominadores obtenidos en *3º* para obtener los valores de los coeficientes de regresión.

5º—Obténgase la media geométrica de los dos valores obtenidos en *4º*. El resultado es el coeficiente de correlación (mejor llamado índice de correlación, r_{12}).

SERIES CRONOLÓGICAS

CONCEPTOS GENERALES

Se da el nombre de serie dinámica o serie cronológica a una sucesión ordenada de valores (de donde "serie") de un fenómeno cuya magnitud se modifica en el transcurso del tiempo (de donde los calificativos de "dinámica" o "cronológica").

En ocasiones, con propósitos de precisión, se distingue entre serie empírica y serie dinámica teórica, refiriéndose en el primer caso a un conjunto cronológicamente ordenado de magnitudes *observadas* de un fenómeno cuyas variaciones se suponen ligadas al transcurso del tiempo y, en el segundo caso, a un conjunto cronológicamente ordenado de magnitudes calculadas de ese mismo fenómeno cuyo cálculo se ha basado asimismo sobre el supuesto de que dichas variaciones de magnitud se encuentran ligadas al transcurso del tiempo.

La liga o vínculo que se postula entre los diversos valores del fenómeno estudiado mediante una serie dinámica y las unidades de tiempo transcurridas equivale a afirmar matemáticamente que la magnitud del fenómeno está en *función* (o es función) del tiempo. O sea, que las variaciones del fenómeno dependerán de las variaciones en el número de unidades de tiempo transcurridas a partir de determinado momento.

Por lo dicho, puede observarse que, al hablar de series dinámicas o series cronológicas nos estamos refiriendo y estamos poniendo en relación dos magnitudes (la magnitud del fenómeno y el número de unidades de tiempo) que pueden asumir valores muy diversos y no un valor único e incambiable, de tal modo que si bien el fenómeno puede alcanzar en un momento el valor r, en un momento distinto puede tener el valor s (superior, inferior o incluso igual como caso límite, con respecto a r), y, a su vez, las unidades de tiempo transcurridas al observarse el valor r del fenómeno pudieron ser n, y m las unidades de tiempo transcurridas cuando se observó el valor s del propio fenómeno. O sea, que las dos magnitudes con las que se trabaja en las series dinámicas o cronológicas no son cantidades constantes, sino *variables*, o sea, unos a modo de receptáculos o vasijas que, dentro de las limitaciones impuestas por su capacidad, pueden admitir diversos contenidos. De este modo, si por x representamos las unidades de tiempo transcurridas, x no representará específicamente el primer momento, el primer mes o el primer año, el segundo momento, el segundo mes o el segundo año, el enésimo momento, el enésimo mes o el enésimo año de nuestras observaciones considerado cada uno de ellos en particular, sino que x estará por todos y cada uno de esos momentos, meses o años, sin ser ninguno de ellos y siendo todos (en el mismo plan en que un género no es ninguna de las especies que lo integran y es o está formado por todas esas especies). En forma análoga, si por y representamos la magnitud del fenómeno estudiado, y no representará exclusivamente el valor r alcanzado por el fenómeno en la primera unidad de tiempo, ni tampoco el valor s alcanzado en la enésima unidad de tiempo, sino esos dos valores, y todos los demás valores observados (y los no observados) del fenómeno en las diversas unidades de tiempo. De ahí que en todo momento deba tenerse presente que x y y como simbólicas de tiempo transcurrido y de magnitud del fenómeno no representan una cantidad fija o constante, sino un conjunto de valores determinados por una definición o una ley matemáticas.

Todo lo anterior significa que los dos elementos principales con los que se trabaja en el estudio de las series dinámicas o series cronológicas son *variables* cuyo *dominio* está determinado por alguna definición, por la elección de un sistema de unidades o por alguna ley matemática.

El hecho de que se postule la vinculación entre las dos variables que intervienen en las dos series dinámicas, o sea la afirmación de que una de ellas es *función* de la otra o *depende* de ella permite distinguirlas con los calificativos de dependiente o independiente. De este modo, el tiempo se considera como variable independiente, y la magnitud del fenómeno como variable dependiente (ya que se supone —en forma un tanto convencional pero no privada de fundamento— que no es el tiempo el que transcurre en dependencia de las magnitudes que alcance el fenómeno, sino que éstas están condicionadas por el transcurso del tiempo).

En esta forma, resalta el hecho de que todo estudio de series dinámicas es un estudio continuamente ligado al tiempo transcurrido. Habida cuenta de esto, las finalidades principales en el estudio de las series cronológicas pueden delinearse como sigue:

1ª—Determinar si el fenómeno crece, decrece, o crece y decrece alternativamente (con sujeción o no a determinado período), conforme transcurre el tiempo.

2ª—Determinar con qué ritmo o velocidad crece, decrece o crece y decrece alternativamente el fenómeno por cada unidad de tiempo transcurrido.

3ª—Determinar la ley matemática del fenómeno, o sea encontrar cuáles son las operaciones mediante las cuales se puede obtener el valor de y (magnitud del fenómeno) en un momento determinado si se conoce el número de unidades de tiempo transcurridas.

4ª—Predecir la magnitud que alcanzará el fenómeno en un momento dado del futuro.

Si nos dejamos conducir por la secuencia de estas finalidades, podremos darnos cuenta de que, a partir de la serie dinámica empírica, llegamos a una serie dinámica teórica en la que se eliminan —conforme se verá más tarde— las variaciones accidentales (es decir, no atribuíbles directamente al transcurso del tiempo) que presentan los valores observados del fenómeno.

Para trabajar con una serie dinámica pueden emplearse los valores *numéricos* observados del fenómeno o puede utilizarse su representación *gráfica.*

La representación gráfica de una serie dinámica comporta previamente la necesidad de un sistema de localización de puntos en un plano (es decir, la utilización de un sistema de ejes coordenados perpendiculares o no entre sí, de un sistema de coordenadas polares, etc.). Resuelto el problema de localización, a cada par de valores de x y de y corresponderá un punto en el plano; mediante la unión de tales puntos se obtendrá la gráfica de la serie, en la cual está implicada la relación funcional de y con respecto a x ya que dichos puntos se encontraron de tal modo que satisficieran simultáneamente los requerimientos que (por definición, elección de unidades o ley matemática) se exigían de x y los que en forma análoga se exigían de y.

Cada uno de los puntos que en el espacio coordenado representan gráficamente las magnitudes del fenómeno en las diferentes unidades de tiempo, reciben el nombre de *polos.*

Interpolar una recta o una curva a una serie empírica equivale a hacer pasar por entre los polos de la misma una línea que, al librar a la serie de las variaciones accidentales que enmascaran las regularidades del fenómeno, represente la tendencia del fenómeno a aumentar o a disminuir cuando se consideran sus variaciones en un plazo más o menos largo *(tendencia secular),* la forma en que, dentro de cada unidad de tiempo ciertas modificaciones en su magnitud se repiten periódicamente *(variaciones estacionales),* o la manera en que la repetición de las variaciones periódicas se producen en lapsos mayores de tiempo *(variaciones cíclicas).*

Si, según hemos dicho, resulta equivalente trabajar con la serie numérica o con la representación gráfica de esa serie, podemos agregar a lo anterior que podemos trabajar asimismo con expresiones algebraicas que muestren la relación entre las variables, y movernos en un plan de paralelismo que nos lleva a afirmar que:

Algebraicamente, interpolar una recta o una curva equivale a encontrar una ecuación que ligue las variaciones del fenómeno a las unidades de tiempo transcurridas, lo cual, a su vez, vale tanto como determinar la ley matemática que rige el fenómeno. Dicha determinación se hace mediante la búsqueda de ciertas cantidades constantes (a, b, c, ... n) que figurarán en la ecuación-ley como coeficientes, como exponentes o simplemente como términos independientes de las diversas potencias de la variable independiente (x) y que relacionadas matemáticamente entre sí y con las potencias de x (por sumas, restas, multiplicaciones, divisiones, potenciaciones, radicaciones, etc.), permiten encontrar los distintos valores de y, o sea la variable dependiente.

Las cantidades a, b, c, ... n, constantes a cuya búsqueda se dirige principalmente o como fin inmediato el análisis de las series cronológicas reciben el nombre de *parámetros,* y su determinación no debe considerarse como fin en sí mismo, sino como medio para un fin: el de substituir sus valores en una ecuación de la forma:

$$y = a + bx + cx^2 + \dots nx^m$$

La interpolación de una recta o una curva determinada a la que naturalmente corresponderá una forma también determinada de ecuación depende,

a.—en primer lugar, de los datos mismos del fenómenos en estudio,
b.—en segundo lugar, del número de observaciones hechas o número de datos disponibles, y
c.—en tercer lugar, de la capacidad del estudioso para apreciar cuál es la recta o curva que representa con mayor aproximación la serie dinámica empírica en estudio.

Esto último que parece dejar librado el estudio de las series dinámicas a la apreciación subjetiva individual es un obstáculo que, sin embargo, se salva ulteriormente, según se verá en páginas posteriores. En cuanto al factor de indeterminación a que está sujeta la interpolación de un tipo determinado de curva o recta sè explica si se tiene en cuenta que, cuando dicha interpolación se hace con base en un número muy limitado de datos hay porciones de diferentes tipos de curva que son igualmente representativas de la serie en ese corto lapso de tiempo, pero con base en las cuales pueden hacerse predicciones diversas en las que puede depositarse confianza en muy diverso grado.

TENDENCIA SECULAR

INTERPOLACIÓN DE UNA RECTA

A fin de dar una idea de la relación existente entre una serie dinámica numérica, su gráfica y su ecuación, comenzaremos por considerar la más sencilla de las series que se nos podrían ocurrir, o sea aquella en la cual en la primera unidad de tiempo ($x = 1$) el fenómeno hubiera alcanzado el valor de una unidad (*un* metro, *un* litro, *una* tonelada, *un* millón de toneladas, *un* mil pesos o sea, $y = 1$), en la segunda unidad de tiempo ($x = 2$) el fenómeno alcanzara el valor de 2 unidades ($y = 2$) etc., tal y como se expresa en la pequeña tabulación siguiente:

x	y
1	1
2	2
3	3
4	4
5	5

Para representar gráficamente esta serie, necesitaremos algún procedimiento de localización de puntos en un plano, y, para resolver esta necesidad adoptaremos el sistema de dos ejes (o rectas) perpendiculares entre sí, que designaremos como eje de las *equis* (eje horizontal), y como eje de las *yes* (eje vertical), los cuales se cortan en un punto 0 llamado origen, determinando en el plano cuatro cuadrantes.

Dentro de este sistema, para determinar la posición de un punto en el plano bastará con medir la distancia de dicho punto a cada uno de los ejes (por medio de perpendiculares trazadas de ese punto a los ejes). En forma análoga, si conocemos el valor de las distancias ("coordenadas") del punto a cada uno de los ejes ("abscisa" la distancia al eje de las *yes*, "ordenada" la distancia al eje de las *equis*) podremos encontrar la posición del punto.

De este modo, si consideramos las unidades de tiempo transcurridas (x) como distancias tomadas desde el origen *sobre* el eje de las *equis*, o sea como distancias tomadas *con respecto* al eje de las *yes*, y las magnitudes del fenómeno estudiado como distancias tomadas del origen *sobre* el eje de las *yes*, o sea como distancias tomadas *con respecto* al eje de las *equis*, y en cada uno de esos puntos levantamos perpendiculares al eje correspondiente, obtendremos un punto en el que se cruzarán dichas perpendiculares y el cual representará un par de valores de x y y. Repitiendo el procedimiento, obtendremos puntos representativos de los diversos pares de valores consignados en la serie y, uniéndolos, una línea que representa a la serie.

En nuestro caso concreto, tomaremos sobre el eje de las *equis* una unidad arbitraria correspondiente a nuestra primera x ($x = 1$); sobre el eje de las *yes* esa misma unidad (con propósitos explicativos, ya que en otro caso la magnitud del segmento unidad puede ser distinto para cada eje) ya que la y correspondiente de la serie es 1; levantando perpendiculares en esos puntos, se determina la intersección o sea el punto P de la gráfica; para el segundo par de valores, tomaremos 2 unidades sobre el eje de las *equis* (ya que $x = 2$) y dos unidades sobre el eje de las *yes* (ya que $y = 2$) a partir del origen, y, por el mismo procedimiento, determinaremos el punto M. En la misma forma, determinaremos los puntos R, S, T; uniendo todos estos puntos obtendremos una recta PT que prolongada hacia abajo pasará por el origen. La recta OT representa la serie dinámica numérica, y trabajar con la serie o con la gráfica es equivalente (Figura 2).

Del simple examen de la serie numérica puede verse que cada y es igual a la x correspondiente; de este modo, en forma que sintetiza las relaciones entre los valores de la serie, poniendo en lugar de la colección de valores de la primera columna (1, 2, 3, 4, 5) x, y en lugar del colectivo de la segunda columna (1, 2, 3, 4, 5) y, podremos escribir:

$$x = y \qquad \text{o} \qquad y = x$$

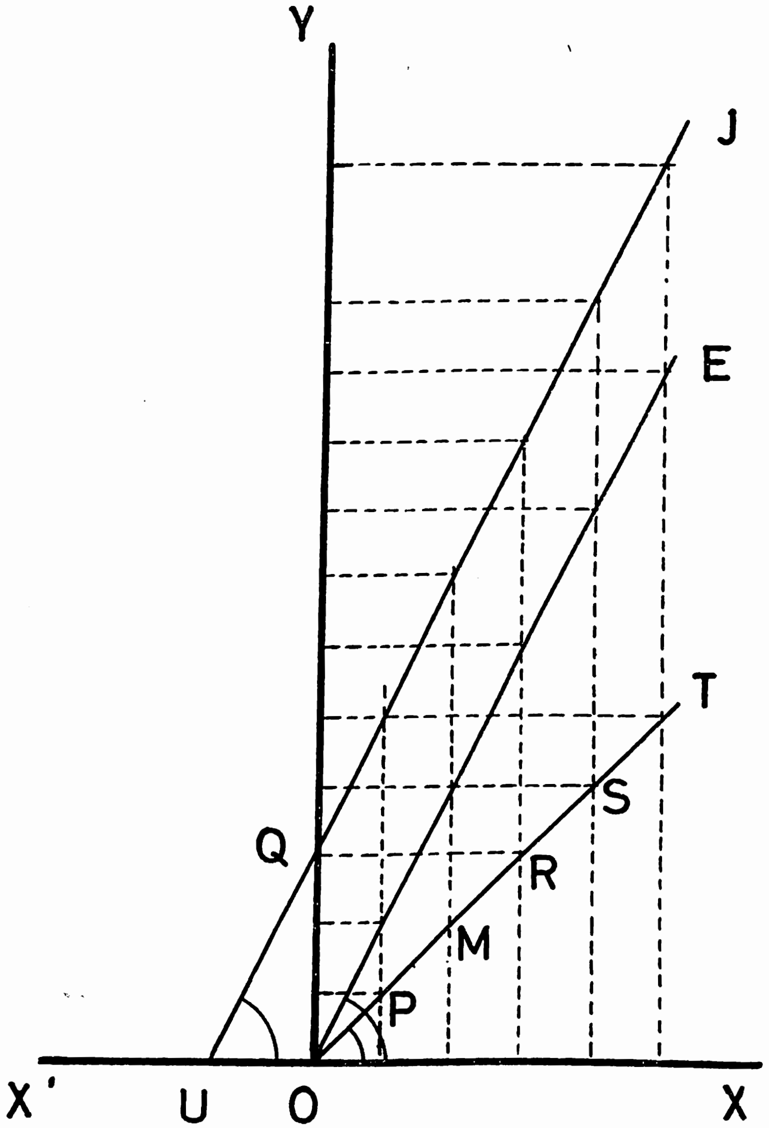
Esta expresión algebraica es equivalente a la serie dinámica numérica y a la representación gráfica por medio de la recta OT, y constituye la ley del fenómeno.

Consideremos el caso menos sencillo de la serie:

x	y
1	2
2	4
3	6
4	8
5	10

Si representamos gráficamente la serie por el mismo proceso empleado en el caso anterior y conservamos los valores de los segmentos unitarios sobre los ejes de las *equis* y de las *yes*, y sobreponemos la gráfica de esta segunda serie a la de la primera, obtendremos la recta OE que representa a esta nueva serie.

Si vemos, como en el caso anterior, cuál es la relación de cada una de las *yes* con su *equis* correspondiente en la serie numérica, tendremos las relaciones de la tabulación siguiente.



RECTAS

representación de

ecuaciones de 1er grado

Figura 2

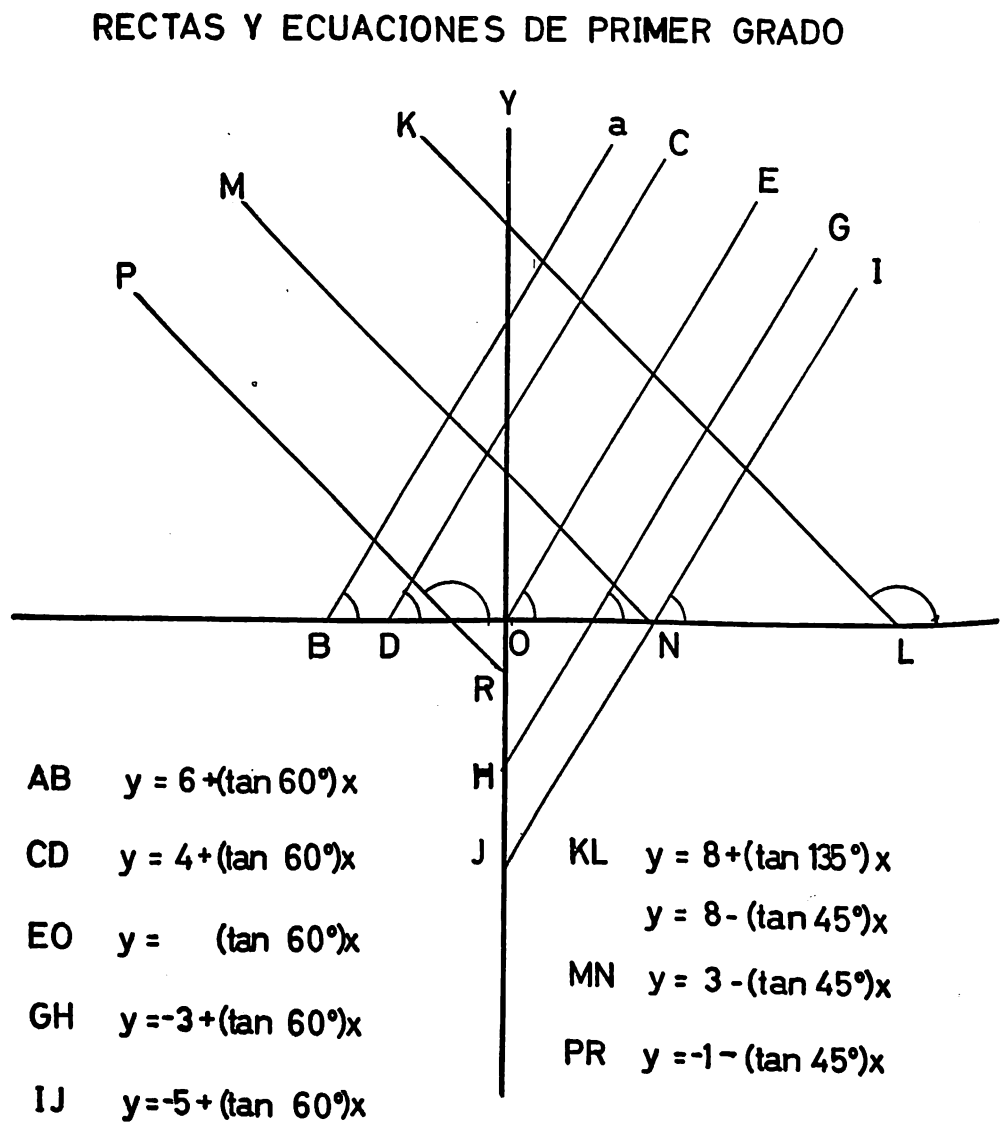

RECTAS Y ECUACIONES DE PRIMER GRADO
Y
K
a
C
M
E
G
P
I
B
D
O
N
L
R
H
J
AB y = 6 +(tan 60°) x
CD y = 4 + (tan 60°)x
EO y = (tan 60°)x
GH y = -3 + (tan 60°)x
IJ y = -5 + (tan 60°)x
KL y = 8 + (tan 135°) x
y = 8 - (tan 45°)x
MN y = 3 - (tan 45°)x
PR y = -1 - (tan 45°)x

Figura 3

x	y
1	$2 = 2 \times 1$
2	$4 = 2 \times 2$
3	$6 = 2 \times 3$
4	$8 = 2 \times 4$
5	$10 = 2 \times 5$

O sea, que cada y es igual a 2 multiplicado por la *equis* correspondiente (la primera y 2 es igual a 2 multiplicada por la primera *equis,* 1). Si en lugar del colectivo de la segunda columna (2, 4, 6, 8, 10) escribimos y, en lugar de la colección de números 2 que aparecen en todas las igualdad *2,* y en lugar del último colectivo (1, 2, 3, 4, 5) su representativo x, tendremos: $y = 2x$

Si comparamos las dos rectas representativas de las dos series numéricas podremos darnos cuenta de que difieren entre sí por la diferente inclinación de cada una de ellas con respecto al eje de las *equis* (y, consecuentemente también con relación al eje de las *yes*), ya que, en efecto, el ángulo α que forma la primera recta con el eje de las *equis,* es menor que el ángulo formado por la segunda recta y ese mismo eje.

Si comparamos las dos ecuaciones correspondientes, podremos percatarnos de que mientras en la primera ($y = x$), x estaba multiplicada por la unidad (ya que podría escribirse: $y = 1x$) en la segunda x aparece multiplicado por otro valor (2), de lo cual podemos deducir que:

La inclinación de una recta queda dada por el coeficiente de x en la ecuación correspondiente. O bien que: El coeficiente de x en la ecuación de primer grado indica cuál es el grado de inclinación de la recta que la representa.

En la figura 3 se han representado diferentes rectas y se han anotado sobre ellas sus ecuaciones a fin de que se vea la relación entre el valor del coeficiente de x y la inclinación de la recta.

El coeficiente de x para el caso de cada recta es un valor constante que podemos representar por b (b valió 1 en la primera recta, 2 en la segunda recta considerada). De este modo si en lugar de ecuaciones particulares de la forma $y = x$; $y = 2x$; $y = -2x$; $y = 0.5x$, escribimos $y = bx$ tendremos una expresión general representativa de todas estas rectas.

Consideremos finalmente el caso de la serie siguiente:

x	y
1	5
2	7
3	9
4	11
5	13

Si representamos gráficamente esta serie y la sobreponemos a la gráfica de las series anteriores, podremos darnos cuenta de que la recta QJ representativa de esta serie es paralela a la recta OE representativa de la serie anterior; o sea, que el ángulo γ formado por la recta QJ y el ángulo β formado por la recta OE y ese mismo eje son iguales (los ángulos formados por dos paralelas y una secante que se encuentran del mismo lado de la secante y en distinta paralela son iguales). Esto quiere decir que la inclinación de las rectas representativas de estas dos series es la misma; 2. Sin embargo, puede verse que, a diferencia de lo que ocurre con la serie representada por *OE*, la serie representada por *QJ* no está formada por un conjunto de *yes* que sean iguales al duplo (2) de la *equis* correspondiente, ya que:

x	y
1	$5 \neq 2 \times 1$
2	$7 \neq 2 \times 2$
3	$9 \neq 2 \times 3$
4	$11 \neq 2 \times 4$
5	$13 \neq 2 \times 5$

Para hacer desaparecer la desigualdad en cada caso deberemos agregar algo al segundo miembro de esas desigualdades, a fin de obtener el primer miembro, o sean las *yes*. En el caso concreto, podemos percatarnos de que si agregamos 3 unidades a cada segundo miembro de la desigualdad, la desigualdad se transforma en igualdad:

x	y
1	$5 = 3 + 2 \times 1$
2	$7 = 3 + 2 \times 2$
3	$9 = 3 + 2 \times 3$
4	$11 = 3 + 2 \times 4$
5	$13 = 3 + 2 \times 5$

Si comparamos las rectas *QJ* y *OE* representativas de las 2 series analizadas, nos daremos cuenta de que, si bien tienen la misma inclinación, difieren en que mientras la recta *OE* pasa por el origen, la recta *QJ*, no pasa por el origen, sino corta al eje de las *yes* por encima del origen.

Si debajo de cada columna de la serie colocamos la letra representativa del colectivo correspondiente contenido en la columna, o el valor constante que multiplica o se suma a esas cantidades, la serie últimamente examinada se convierte en:

$$y = 3 + 2x$$

Al comparar esta ecuación con la ecuación de la serie representada por *OE* ($y = 2x$) podemos darnos cuenta de que dichas ecuaciones difieren en que mientras en la ecuación de *OE* no aparece término independiente de x (o dicho término es cero), en la ecuación de *QJ* aparece un término independiente de x (3).

De estas dos comparaciones resulta que:

El punto en que la recta corta al eje de las *yes* queda dado por el término independiente de la ecuación de primer grado, o, lo que es lo mismo, el término independiente de la ecuación de primer grado indica el punto en el que la recta corta al eje de las *yes*.

El término independiente es un valor constante para cada recta, y puede representarse por a. En la figura puede verse la forma en que, a diferentes valores de a y un mismo valor de b corresponden diferentes rectas.

Todo lo anterior, quiere decir que, en forma general, podemos considerar a una recta representada por la siguiente forma de la ecuación de primer grado:

$$y = a + bx$$

A fin de hacer buen uso de esta ecuación general de la recta, es conveniente precisar —incluso como reiteración, en ciertos casos— el significado de sus literales. Para ello nos referiremos a tres niveles distintos: el geométrico, el de la geometría analítica y el propiamente estadístico.

1.—Nivel geométrico.

y, a, x.—representan distancias,

b.—representa una relación entre distancias, ya que b puede definirse como la tangente trigonométrica del ángulo que la recta forma con el eje de las *equis* y la tangente de un ángulo es la relación entre el cateto opuesto y el cateto adyacente a dicho ángulo. Por la figura 4 tenemos:

$$b = tan\ \alpha = tan\ XRP = tan\ SQP = \frac{PS}{QS}$$

Pero: $PS = PT - ST = PT - QO$

de donde:

$$b = \frac{PT - QO}{QS}$$

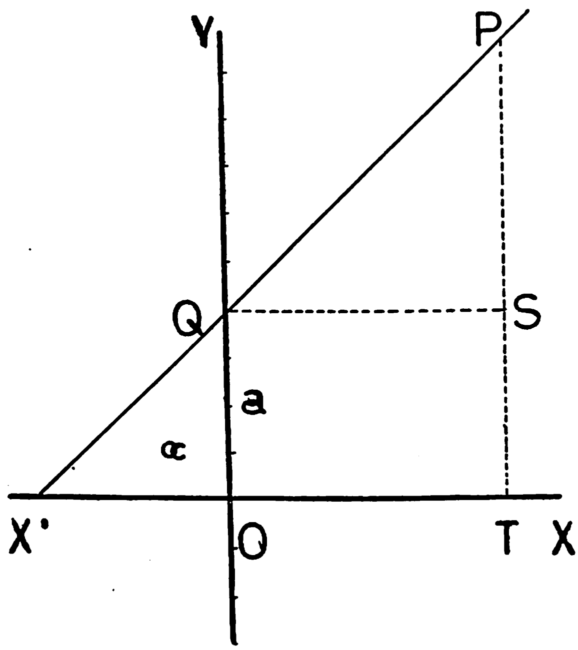
2.—*Nivel de la geometría analítica.*

y.—distancia *de un* punto cualquiera de la recta al eje de las *equis,* o sea la ordenada de un punto cualquiera de la recta.

a.—distancia *del* punto en que la recta corta al eje de las *yes* con respecto al origen (o sea a la intersección de dicho eje con el eje de las *equis*); recibe el nombre de ordenada en el origen.

x.—distancia de un punto al eje de las *yes* o sea la abscisa de un punto cualquiera de la recta.

b.—relación entre la distancia de un punto cualquiera de la recta al eje de las *equis,* y la distancia del pie de dicha perpendicular al punto en que la recta corta al propio ele de las *equis;* o bien: relación entre la distancia que se obtiene de restar a la ordenada del punto ($PT = y$) la ordenada en el origen ($QO = a$), y la



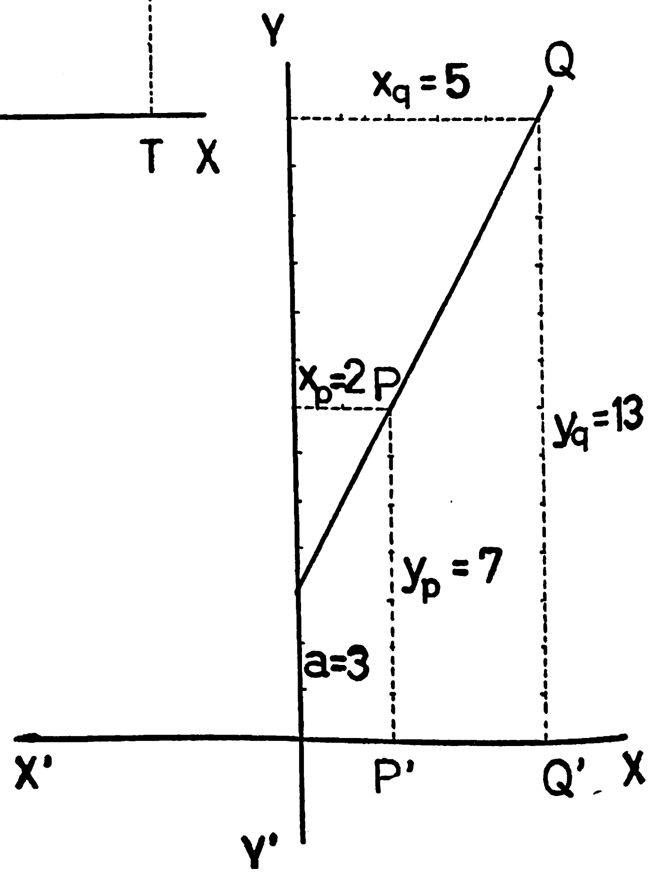


Figura 4

abscisa del punto ($OT = x$). Si volvemos sobre lo establecido en el nivel geométrico para b y substituimos las distancias por las literales, tendremos:

$$b = tan\ \alpha = \frac{y - a}{x}$$

El valor de b recibe el nombre de "pendiente" de la recta.

3.—Nivel estadístico.

y.—magnitudes variables del fenómeno en el tiempo,

a.—magnitud del fenómeno en el momento previo a aquel en que se iniciaron las observaciones (de poca importancia práctica *per se),*

x.—número de unidades de tiempo transcurridas desde el momento en que se iniciaron las observaciones (o desde aquél que se consideró como central en el estudio de la serie, conforme se verá más adelante),

b.—Velocidad o ritmo de crecimiento o decrecimiento del fenómeno. Crecimiento o decrecimiento del fenómeno por unidad de tiempo transcurrido.

Los casos particulares que pueden presentarse en la interpolación de una recta son:

Que a sea igual a *cero,* en cuyo caso, la recta pasa por el origen ($y = bx$).

Que a sea mayor que *cero,* o sea que a sea positiva, en cuyo caso, la recta corta el eje de las *yes* por encima del origen ($y = a + bx$).

Que a sea menor que *cero,* o sea que a sea negativa, en cuyo caso, la recta corta al eje de las *yes* por debajo del origen ($y = -a + bx$).

Que b sea igual a *cero,* en cuyo caso, la recta es horizontal ($y = a$).

Que b sea mayor que *cero,* o sea positiva, en cuyo caso la recta es ascendente, o sea que el fenómeno es creciente ($y = a + bx$).

Que b sea menor que *cero,* o sea negativa, en cuyo caso la recta es descendente, o sea que el fenómeno es decreciente ($y = a - bx$).

Que b sea igual a *1*, en cuyo caso, la recta forma un ángulo de 45 grados con los ejes. Si, además, a es igual a cero, la ecuación de la recta queda simplificada al máximo ($y = x$).

PROCEDIMIENTOS DE INTERPOLACIÓN

1er. Procedimiento Gráfico: Este procedimiento deriva directamente de las definiciones que se han dado de los parámetros y de las variables en la ecuación general de la recta, y consiste en:

1.—representar gráficamente la serie dinámica empírica (Fig. 4).
2.—hace pasar por entre los polos, a ojo, una recta que se encuentre aproximadamente equidistante de los más bajos y de los más altos,
3.—prolongar la recta así trazada hasta que corte el eje de las *yes*,
4.—medir la distancia entre el punto de intersección y el origen. Dicha distancia nos dará el valor de a,
5.—prolongar la recta hasta que corte al eje de las *equis*,
6.—medir, con un transportador, el valor del ángulo formado entre la recta y la dirección positiva del eje de las *equis* (ángulo α),
7.—consultar una tabla de funciones naturales para determinar el valor de la tangente del ángulo α. El valor consignado en la tabla dará el valor de b,
8.—determinadas a y b se substituirán en la ecuación general de la recta para obtener la ecuación que rige al fenómeno en estudio.

2º Procedimiento Gráfico: Este segundo procedimiento es mera variación del anterior, ya que la determinación del valor de a es idéntico (o sea, que son iguales hasta el 4º paso), variando en cuanto a la determinación del valor de b para el cual:

5.—desde $x = 1$ se levanta una perpendicular al eje de las *equis* hasta cortar a la recta en Q,
6.—desde Q se traza una perpendicular al eje de las *yes*; perpendicular cuyo pie es M,
7.—se mide la distancia entre M y el punto de intersección de la recta con el mismo eje de las *yes*. Así, se obtiene el valor de b. En caso de que M, pie de la perpendicular trazada del punto Q cuya x vale 1 caiga por debajo del punto I, intersección de la recta con el eje de las *yes*, la b debe afectarse del signo menos,
8.—obtenidos a y b, se sustituyen en la fórmula general.

3er Procedimiento Gráfico: La determinación del valor de a se hace como en los casos anteriores; para determinar el valor de b:

5.—se elige un punto P sobre la recta,
6.—desde P se baja una perpendicular al eje de las *equis* para determinar la ordenada (magnitud de la pérpendicular misma) y la abscisa (distancia del pie de la perpendicular al origen) del mismo punto,
7.—desde I, punto de intersección de la recta y el eje de las *yes* se traza una paralela al eje de las *equis* que cortará a la ordenada de P en un punto S,

8.—se miden las distancias *PS* y *SI*,
9.—se divide *PS* entre *SI* para obtener *b*,
10.—obtenidas *a* y *b* se substituyen en la fórmula general de la recta.

4º Procedimiento Gráfico: Es variante de todos los anteriores, pero especialmente del tercero, ya que sólo varía a partir del 7º paso:

7.—se miden las distancias *PT* y *TO*,
8.—se resta de *PT* el valor de *a* previamente determinado, y el resultado se divide entre *TO*, con lo que se obtiene el valor de *b*,
9.—obtenidos *a* y *b* se sustituyen.

En todos los procedimientos anteriores, es necesario tener en consideración que cuando la recta corta al eje de las *yes* por debajo del origen la *a* debe ir afectada de signo menos, y restar *a* quiere decir sumarla algebraicamente o sea, teniendo en cuenta su signo.

Procedimiento Elección de Puntos. El procedimiento de determinación de *a* y de *b* por elección de puntos se basa en el supuesto de que "si un punto pertenece a una recta, sus coordenadas satisfacen la ecuación de dicha recta", o bien que si el punto pertenece a la recta al substituir la y de la ecuación por la ordenada del punto y la x de la ecuación general de la recta por la abscisa del punto elegido se establece una igualdad entre los dos miembros de la ecuación. De otra parte, como son dos las cantidades desconocidas *a* y *b* cuyos valores se buscan, y la teoría general de las ecuaciones pide que para que sea definido el valor de las incógnitas haya tantas ecuaciones como incógnitas, necesitaremos dos ecuaciones en las que figuren *a* y *b* y, para tener esas dos ecuaciones necesitaremos elegir dos puntos sobre la recta, cuyas coordenadas substituídas en la ecuación general de primer grado la satisfarán. El procedimiento es el siguiente:

1.—Se elige un punto *P* sobre la recta y se determinan sus coordenadas; es decir:
  a.—desde *P* se traza una perpendicular al eje de las *equis* y se mide la distancia entre *P* y el punto de intersección de la perpendicular con el eje; así se obtiene la ordenada y_p. Ej.: $y_p = 7$,
  b.—desde *P* se traza una perpendicular al eje de las *yes* y se mide la distancia entre *P* y el punto de intersección de la perpendicular con el eje (o bien se mide la distancia entre el pie de la ordenada y_p y el origen); así se obtiene la abscisa x_p. Ej.: $x_p = 2$.

2.—En la ecuación general de la recta ($y = a + bx$) se substituyen los valores anteriormente encontrados y_p y x_p dejando intactas *a* y *b* que aún se desconocen. Así se obtiene una primera ecuación:

$y_p = a + bx_p$ o, en el ejemplo: $7 = a + 2b$.

3.—Se repite todo el proceso anterior para otro punto Q elegido arbitrariamente *sobre la recta,* obteniéndose así los valores y_q y x_q (13 y 5 en el ejemplo considerado), con lo que se obtiene la ecuación:

$y_q = a + bx_q$ o, en el ejemplo: $13 = a + 5b$,

4.—con las ecuaciones obtenidas en *2* y *3* se constituye un sistema de dos ecuaciones con dos incógnitas:

$y_p = a + bx_p$ o, en el ejemplo: $7 = a + 2b$
$y_q = a + bx_q$ $13 = a + 5b$

5.—se resuelve el sistema conforme a los principios estudiados en la "Revisión Matemática Indispensable", y así se obtienen los valores de a y de b.

Si resolvemos el sistema literal por sumas y restas (restando 2 de 1 o 1 de 2) tendremos:

$$\begin{aligned} y_p &= a + bx_p \\ -\ (y_q &= a + bx_p) \\ \hline y_p - y_q &= a - a + bx_p - bx_q \\ y_p - y_q &= bx_p - bx_q = b(x_p - x_q) \\ b &= \frac{y_p - y_q}{x_p - x_q} \end{aligned}$$

Esto quiere decir que podemos derivar un segundo procedimiento de elección de puntos, si queremos obtener los valores de a y de b directamente por la aplicación de una fórmula que evite la resolución del sistema de dos ecuaciones con dos incógnitas:

Procedimiento Elección de Puntos: Una vez determinados las coordenadas respectivas de P y Q (los dos puntos elegidos sobre la recta), substitúyanse esos valores en las siguientes fórmulas:

$$b = \frac{y_p - y_q}{x_p - x_q} \qquad a = y_p - bx_p$$

Esto quiere decir que, en este caso, la determinación del valor de b precede a la determinación del valor de a. Como precaución, debe tenerse cuidado de afectar a las diferencias $y_p - y_q$ y $x_p - x_q$ del signo que les corresponda (más o menos) y de hacerlo asimismo con el cociente de ambos.

Método de los Mínimos Cuadrados.—El método de interpolación conocido como "procedimiento de los mínimos cuadrados" responde a la necesidad de que la línea interpolada discrepe en tan poco como sea posible de la serie dinámica empírica a la que se interpola. Para conseguir esto, se busca el que la suma de los cuadrados de las diferencias entre las *yes* reales (valores de la serie empírica) y las *yes* teóricas (calculadas mediante la ecuación de la línea correspondiente) sea un mínimo, o sea $\Sigma(d^2)$ = un mínimo; expresión en la cual Σ (operador Σ) representa suma y d las diferencias o desviaciones entre las *yes* reales y las *yes* teóricas.[1]

En virtud de que:

$$d = y_r - y_t$$

En la que y_r representa a las *yes* reales y y_t a las yes teóricas, y de que:

$$y_t = a + bx$$

Conforme a la ecuación general de la recta, podemos establecer:

$$d = y_r - (a + bx)$$

y $$S(d^2) = S[y_r - (a + bx)]^2$$

Si desarrollamos el cuadrado de la expresión contenida dentro del paréntesis rectangular tendremos:

$$S(d^2) = S[y_r^2 - 2\,y_r(a + bx) + (a + bx)^2]$$

Si en el segundo miembro de la igualdad ejecutamos las operaciones indicadas en el segundo término, y desarrollamos el cuadrado del binomio del tercer término, tendremos:

$$S(d^2) = S[y_r^2 - 2y_ra - 2y_rbx + a^2 + 2abx + b^2x^2]$$

Conforme a la condición expresada en un principio: $S(d^2)$ debe hacerse un mínimo; esto equivale a que:

$$S[y - 2ya - 2ybx + a^2 + 2abx + b^2x^2]$$

sea un mínimo. A partir de este momento, para facilitar la escritura suprimimos el índice de y.

Como, según se demuestra en el Cálculo Diferencial, para obtener el mínimo de una expresión debe igualarse a cero la derivada de dicha expresión,

[1] *Advertencia.*—En el siguiente desarrollo, a causa de un descuido utilizamos S en donde por uniformidad con el resto debimos emplear Σ; sin embargo, en cuanto el fondo no se afecta con este cambio, no puede ocasionar tropiezo al lector atento.

procederemos a encontrar dicha derivada, para lo cual es necesario tener en cuenta que la derivada de una función es el límite de la relación entre el incremento de la función y el incremento de la variable independiente cuando el incremento de la variable independiente tiende hacia cero:

$$\text{Derivada de } u \text{ con respecto a } v = \lim_{v \to 0} \frac{\Delta u}{\Delta v}$$

Y, según esta definición, para derivar una función u con respecto a la variable inedependiente v:

1.—Incrementamos el valor de v, agregando Δv.

2.—Mediante la substitución de v por $v + \Delta v$ en la función y la ejecución de las operaciones obtenemos el incremento de u correspondiente al incremento de v.

3.—Restamos de la función incrementada la función no incrementada para obtener la relación entre los valores de los incrementos.

4.—Dividimos ambos miembros de la ecuación por el incremento de la variable independiente.

5.—Vemos cuál es el límite de los cocientes contenidos en ambos miembros de la ecuación cuando el incremento de la variable independiente tiende hacia cero.

Si aplicamos este procedimeinto de derivación a la expresión:

$$S[y^2 - 2ya - 2ybx + a^2 + 2abx + b^2x^2]$$

que igualaremos a v tomando como variable independiente a a (que, no obstante ser constante en el caso de una recta concreta, resulta variable en cuanto se consideran todas las rectas que sería posible interpolar a la serie dinámica empírica llenaran o no la condición de hacer mínima la suma de los cuadrados de las diferencias), tendremos:

1.—Incrementando a a (agregando Δa dondequiera aparezca a en la expresión):

$$S[y^2 - 2y(a + \Delta a) - 2ybx + (a + \Delta a)^2 + 2(a + \Delta a)bx + b^2x^2]$$

Ejecutando las multiplicaciones indicadas y desarrollando los cuadrados de los binomios de esta expresión, se tiene:

$$S[y^2 - 2ya - 2y\Delta a - 2ybx + a^2 + 2a\Delta a + (\Delta a)^2 + 2abx +$$
$$+ 2\Delta abx + b^2x^2]$$

2.—Al incrementar uno de los elementos de esta expresión no subsiste la igualdad con respecto a v que, por tanto, resultará incrementada en una cantidad Δv:

$$S[y^2 - 2ya - 2y\Delta a - 2ybx + a^2 + 2a\,a + (\Delta a)^2 + 2abx +$$
$$+ 2\Delta abx + b^2x^2 = v + \Delta v$$

3.—Si de esta función incrementada restamos la no incrementada, miembro a miembro y término a término (cancelando en ésta todos los términos que existen en la no incrementada con el mismo signo), tendremos:

$$S[y^2 - 2ya - \underline{2y\Delta a} - 2ybx + a^2 + \underline{2a\Delta a} + \underline{(\Delta a)^2} + 2abx +$$
$$+ \underline{2\Delta a\cdot bx} + b^2x^2] = v + \underline{\Delta v}$$

O sea, que quedan únicamente los términos subrayados:

$$S[2y\Delta a + 2a\Delta a + (\Delta a)^2 + 2\Delta a\cdot bx] = \Delta v$$

4.—Si se dividen los dos miembros de la igualdad por el incremento de la variable independiente (en este caso a) tendremos:

$$S\left[\frac{2y\Delta a}{\Delta a} + \frac{2a\Delta a}{\Delta a} + \frac{(\Delta a)^2}{\Delta a} + \frac{2\Delta a\cdot bx}{\Delta a}\right] = \frac{\Delta v}{\Delta a}$$

$$S[2y + 2a + a + 2bx] = \frac{\Delta v}{\Delta a}$$

5.—Si se hace que Δa tienda hacia cero, el tercer término de la suma contenida en el primer miembro tenderá a cero, y $\Delta v/\Delta a$ tenderá a la derivada de v con respecto a a (ya que, según la definición, la derivada de una función es el límite del cociente entre el incremento de la función y el incremento de la variable independiente cuando éste tiende a cero):

$$S[2y + 2a + 2bx] = \text{derivada de } v \text{ con respecto a } a = \frac{dv}{da}$$

Fuera ya del proceso de derivación, a fin de que la función alcance un mínimo, es necesario que:

$$S[2y + 2a + 2bx] = 0$$

O bien que:

$$S[2(y + a + bx)] = 0$$

O, en vista de que un factor constante puede entrar o salir del operador Σ ($= S$ en este caso) sin cambio alguno.

$$2\,S(y + a + bx) = 0$$

Como para que un producto (2 por S de...) sea cero basta que uno de sus factores sea cero, y el factor 2 no puede ser cero, tendrá que ser cero el otro factor (S de...):

$$S(y + a + bx) = 0$$

Pero como la suma de una serie de cantidades es igual a la suma de las sumas de los diversos colectivos representados por esas cantidades:

$$S(y + a + bx) = Sy + Sa + Sbx = 0$$

Si se despeja Sy, se obtiene:

$$Sy = Sa + Sbx$$

Pero como a ya en el caso concreto es una cantidad constante, la suma de N *aes* es igual a N multiplicado por a:

$$Sy = Na + Sbx$$

O bien, según es más usual:

$$Sy = Na + S(x)b$$

En esta forma, se obtiene una primera ecuación de interpolación mediante la cual obtener el valor de a que haga de la suma de los cuadrados de las diferencias un mínimo. Sin embargo, como existen dos incógnitas (a y b) se requieren de dos ecuaciones para obtener los valores correspondientes. Para obtener una segunda ecuación de interpolación, derivaremos la misma expresión que nos sirvió de punto de partida, sólo que en esta ocasión, la derivaremos con respecto a b en vez de hacerlo en relación a a, y, en seguida anularemos esa derivada para obtener un mínimo de la función:

$$S[y^2 - 2ya - 2ybx + a^2 + 2abx + b^2x^2] = v$$

De acuerdo con el procedimiento general de derivación:

1.—Incrementamos b (agregando Δb en dondequiera se encuentre b en la expresión):

$$S[y^2 - 2ya - 2y(b + \Delta b)x + a^2 + 2a(b + \Delta b)x + (b + \Delta b)^2x^2]$$

Si ejecutamos multiplicaciones y desarrollamos la segunda potencia del binomio $b + \Delta b$, tendremos:

$$S[y^2 - 2ya - 2yxb - 2yx\Delta b + a^2 + 2abx + 2ax\Delta b +$$
$$+ b^2x^2 + 2b\Delta b \cdot x^2 + (\Delta b)^2x^2]$$

2.—Las expresiones anteriores, como incrementadas que están, no podrán seguir siendo iguales a v, sino a $\underline{v + \Delta v}$.

3.—Si de las expresiones así incrementadas se restan las no incrementadas cancelando los términos comunes (o señalando con subrayado los que quedan después de la cancelación):

$$S[y^2 - 2ya - 2yxb - \underline{2yx\Delta b} + a^2 + 2abx + \underline{2ax\Delta b} + b^2x^2 +$$
$$+ \underline{2b\Delta b \cdot x^2} + \underline{(\Delta b)^2x^2]} = v + \underline{\Delta v}$$

O sea:

$$S(- 2yx\Delta b + 2ax\Delta b + 2b\Delta b \cdot x^2 + (\Delta b)^2x^2) = \Delta v$$

4.—Si dividimos ambos miembros de esta ecuación por el incremento de la variable independiente (en este caso b) tendremos:

$$S\left(\frac{- 2yx\Delta b}{\Delta b} + \frac{2ax\Delta b}{\Delta b} + \frac{2b\Delta b \cdot x^2}{\Delta b} + \frac{(\Delta b)^2x^2}{\Delta b}\right) = \frac{\Delta v}{\Delta b}$$

O sea:

$$S(- 2xy + 2ax + 2bx^2 + (\Delta b)x^2) = \frac{\Delta y}{\Delta b}$$

5.—Si hacemos que b tienda hacia cero, o sea, que se haga tan pequeño como sea posible, el último término de la suma contenida en el primer miembro desaparecerá, y el segundo miembro de la ecuación se convertirá (conforme a la definición de derivada) en la derivada de v con respecto a b.

$$S(-2xy + 2ax + 2bx^2) = \frac{dv}{db}$$

O bien (sacando a 2 como factor común en el primer miembro):

$$S[2(- xy + ax + bx^2)] = \frac{dv}{db}$$

O, como un factor constante puede salir del operador Σ sin sufrir cambio:

$$2\,S(- xy + ax + bx^2) = \frac{dv}{db}$$

O, aún:

$$2\,S[(-xy) + Sax + Sbx^2] = \frac{dv}{db}$$

Pero $S(-xy) = S(xy)$ ya que esto equivale a considerar que existía ante xy un factor -1 que salió fuera del operador Σ; por tanto:

$$2[-Sxy + Sax + Sbx^2] = \frac{dv}{db}$$

Ya fuera del proceso de derivación, si para obtener un mínimo anulamos el valor de esta derivada, tendremos:

$$2[-Sxy + S\,ax + S\,bx^2] = 0$$

Como para que un producto sea cero basta con que sea cero uno de sus factores y el primer factor (2) no puede ser cero,

$$-Sxy + Sax + Sbx^2 = 0$$

Pasando $-Sxy$ al otro miembro:

$$Sax + Sbx^2 = Sxy$$

O bien:

$$Sxy = Sax + Sbx^2$$

Esta es una segunda ecuación de interpolación que, unida a la encontrada anteriormente forma un sistema de dos ecuaciones con dos incógnitas, al través del cual pueden encontrarse los valores de a y de b de tal modo que la suma de los cuadrados de las diferencias entre las *yes* de la serie dinámica empírica y las *yes* de la recta que se le intepole sea un mínimo, de donde el nombre de procedimiento de los mínimos cuadrados.

El procedimiento en sí, puede resumirse como sigue:

1.—Se buscan los valores de Sy, N, Sx, Sxy y Sx^2 requeridos por las dos ecuaciones de interpolación:

$$Sy = Na + S(x)b$$

$$S(yx) = S(x)a + S(x^2)b$$

para lo cual:

a.—se suman las *yes* de la serie empírica al pie de la columna que las contiene (Sy), *b.*—se cuenta el número de observaciones hechas, o sea el número de valores consignados (N),

c.—se suman las *equis* correspondientes (ya sea que estén dadas como una sucesión de años: 1940·1941·1942·1943... o que, con propósitos de simplificación en las operaciones numéricas se haya puesto delante de cada año el número de orden correspondiente comenzando con 1 (1, 2, 3, 4...). ($S(x)$,

d.—se forman, en una columna adicional, los productos de cada una de las *equis* por cada una de las *yes* correspondientes, y se suman esos productos. ($S(xy)$),

e.—se buscan los cuadrados de las *equis* y se anotan en otra columna, cuidando de que sean los cuadrados de los años si con ellos se trabajó desde el principio, o los cuadrados del número de orden de esos años si fue con estos ordinales con los que se trabajó desde el principio, y se suman esos cuadrados ($S(x^2)$). Fijarse que se trata de la suma de los cuadrados y no del cuadrado de la suma de las *equis*: $S(x^2)$ y no $(Sx)^2$ pues no es lo mismo encontrar la suma de las *equis* y después elevarla al cuadrado.

2.—Se substituyen los valores encontrados para Sy, N, SX, Sxy y $S(x^2)$ en las dos ecuaciones de interpolación.

3.—Se resuelve el sistema de dos ecuaciones con dos incógnitas, determinando así los valores de a y de b.

4.—Los valores de a y de b se substituyen en la ecuación general de la recta:

$$y = a + bx$$

Variante: Lo dicho anteriormente se aplica sin mayores dificultades, cuando entre los diferentes años de observación media un mismo lapso de tiempo: cuando las observaciones se hacen de año en año, de década en década, etc. y a cada una de las observaciones se le puede asignar cada uno de los números de la serie de números naturales 1, 2, 3, 4... ; en cambio, cuando el lapso que media entre la primera observación y la segunda es distinto del que existe entre ésta y la tercera, etc., es necesario asignar a cada uno de los valores de y la x que le correspondería si la serie estuviese completa, y proceder, en todo lo demás como en el caso anterior; así, por ejemplo, si suponemos que nuestras observaciones correspondieron a los años de: 1940, 1941, 1943, 1947, 1948, consideraremos completa la serie de 1940 a 1948 asignando a cada uno de los años de esa serie uno de los números naturales del 1 al 9, con lo que tendremos:

Año	x
1940	*1*
1941	*2*
1942	3
1943	*4*
1944	5
1945	6
1946	7
1947	*8*
1948	*9*

Es decir, que en nuestra columna de las *equis,* frente a la *ye* correspondiente al año de 1940 aparecerá como *equis* correspondiente 1, frente a la de 1941, 2; frente a la de 1943, 4; frente a la de 1947, 8 y frente a la de 1948, 9; que la Sx será igual a $1 + 2 + 4 + 8 + 9 = 24$, y la N será 5 puesto que fueron cinco las observaciones hechas en el lapso de 1940 a 1948 por carecer de los datos correspondientes a varios años intermedios. Puede decirse, de paso, que, mediante el procedimiento de interpolación, una vez encontrados los valores de a y de b y substituidos en la ecuación general de la recta $y = a + bx$ pueden calcularse esos valores intermedios no observados poniendo en la ecuación, en lugar de x el valor de x que le hubiese correspondido al año intermedio del que no se hicieron observaciones.

Método del año mediano. Este procedimiento es una mera simplificación del método de los mínimos cuadrados ya que parte de sus mismos supuestos y, en términos generales consiste de los mismos pasos, consistiendo su diferencia en que al elegir el año mediano de las observaciones como origen de las mismas se obtiene igual número de valores positivos y negativos de x, con lo cual se anula la suma de las *equis* y se simplifican las fórmulas correspondientes, con lo que, de la necesaria solución de un sistema de dos ecuaciones con dos incógnitas, indispensable en el procedimiento clásico de los mínimos cuadrados, se pasa a la más simple solución de dos ecuaciones independientes de primer grado, cada una de ellas con una incógnita. En efecto si, gracias al procedimiento que exponemos a continuación, Sx se anula, tendremos que las ecuaciones de interpolación:

$$Sy = Na + S(x)b$$

$$S(yx) = S(x)a + S(x^2)b$$

Se convierten en:

$$Sy = Na + 0\,b = Na + 0 = Na \qquad Sy = Na$$

$$S(yx) = S(x)a + S(x^2)b = 0a + S(x^2)b = 0 + S(x^2)b$$

$$S(x^2)b = Syx$$

Si en estas dos ecuaciones despejamos a a y a b, tendremos:

Si $Sy = Na$ $\quad a = \dfrac{Sy}{N}$ (por definición esto es igual a la media aritmética de las *yes*)

Si $S(x^2)b = S(yx)$ $\quad b = \dfrac{S(yx)}{S(x^2)}$

Conforme a todo lo anterior, el procedimiento en sí consiste en:

1.—Determinar cuál es el año mediano, contando del primer año hacia el centro y del último año de las observaciones hacia el centro igual número de años, y señalando como mediano el año en el que las dos cuentas coinciden en un mismo número.

2.—Colocar frente al año mediano un cero,

hacia arriba de él $-1, -2, -3, -4, \ldots -n$

hacia abajo de él $1, 2, 3, 4, \ldots n$

siendo esta serie de valores que va de $-n$ (correspondiente al primer año) a n (correspondiente al último año) pasando por *cero* (que corresponde al año mediano) la que constituirá la serie de *equis* con las que habrá que trabajar aplicando el procedimiento habitual de los mínimos cuadrados o, puesto que hemos visto que, por este medio $S(x)$ se iguala a cero y las fórmulas de interpolación se simplifican, empleando las fórmulas simplificadas, que, por substitución de los valores de Sy, N, Sxy y Sx^2 dan inmediatamente los valores de a y de b; o sea que hay que:

3.—Buscar los valores de Sy, N, Sxy y Sx^2 y substituirlos en:

$$a = \frac{Sy}{N}$$

$$b = \frac{S(yx)}{S(x^2)}$$

4.—Substituir los valores de a y de b así encontrados en la ecuación general de la recta:

$$y = a + bx$$

Variante.—La aplicación del procedimiento, que es inmediata cuando el número de observaciones hechas es impar, necesita de una ligera modificación cuando el número de estas es par. Si suponemos que en el caso de un número impar de observaciones el año mediano fue 1940, a 1940 le corresponde cero, y, más precisamente, al mes mediano de 1940; es decir, a julio; en cambio si suponemos que se trató de un número par de observaciones, el año mediano estará encabalgado entre 1940 y 1941 (si agregamos un año) o entre 1940 y 1939 (si restamos un año a los de nuestras observaciones); o sea, tomando el caso de 1940-1941, que el cero corresponderá a los últimos días de diciembre de 1940 y a los primeros de enero de 1941; que los meses centrales de 1940 distarán en menos medio año de ese punto mediano (—0.5), y los meses centrales de 1941 distarán en más medio año de ese punto mediano (0.5) que los meses centrales de 1939 distarán en 1 año de los centrales de 1940, o sea 1 año y medio del punto mediano (1.5), etc. O sea, que la serie de las *equis* comenzará en —0.5 y en 0.5 atribuidos a los años en los que cabalga la mediana; —1.5, —2.5, —3.5, etc., hacia arriba de dicho punto, y 1.5, 2.5, 3.5, etc. hacia abajo constituirán los valores de las restantes *equis*.

CÁLCULO DE LA RECTA TEÓRICA

Conforme hemos asentado en un principio, al lado de una serie dinámica empírica, existe una serie dinámica teórica constituída por los valores de y calculados a partir de la ecuación que expresa la ley general del fenómeno.

Según esto, y teniendo en cuenta que la ecuación de una recta es de la forma:

$$y = a + bx$$

podremos siempre, a partir de ella, calcular los valores de y correspondientes a la recta teórica si substituimos en la ecuación general, en lugar de x los valores respectivos (es decir: 1940, 1941, 1942, 1943 si se trabajó con el número de los años, 1, 2, 3, 4 . . . si se usaron los números de orden de esos mismos años, —3, —2, —1, 0, 1, 2, 3 si se empleó el método del año mediano con datos impares, —3.5, —2.5, —1.5, —0.0, 0.5, 1.5, 2.5, 3.5 si se empleó el procedimiento del año mediano para un número par de datos, etc.).

En la práctica puede recurrirse a una tabulación, en cuya primera columna se copien las *equis* correspondientes, en tanto que en la segunda se forman los productos de cada x por el valor de b (con su signo), y en una tercera columna se consigna el valor que se obtiene de sumar a cada uno de los productos anteriores el valor de a. Los valores consignados en esta última columna serán los valores de la y teórica correspondiente.

Una vez obtenida la primera y teórica, el trabajo puede simplificarse si se tiene en cuenta que cada y es igual a la anterior más una b; o sea, que una vez calculada y_1 bastará con sumarle b para obtener y_2, sumarle a y_2 b para obtener y_3, etc.

Cuando se ha utilizado el método del año mediano para la interpolación, es conveniente determinar el valor de la tendencia para el año central (o sea, para cuando x vale cero), sumar al valor obtenido el valor de $1b$, $2b$, etc., como en el caso anterior para los años que siguen al año mediano, y restar del valor de y para el año mediano, $1b$, $2b$, etc., para los años que preceden a dicho año.

Asimismo, como hemos dicho anteriormente, cuando hay lagunas en los datos, se pueden calcular las *yes* correspondientes poniendo en la ley general del fenómeno la x que le correspondería y ejecutando las operaciones indicadas en la ecuación misma. En forma análoga, para determinar el valor de la tendencia correspondiente a un mes determinado de un año en el cual se han hécho observaciones, bastará con considerar a cada mes como un doceavo de año. Así, por ejemplo, si las observaciones se hicieron ordinariamente en los meses de enero a partir del año de 1940 y se busca el valor de la $\bar{y}$ correspondiente al mes dė abril de 1940 como de enero a abril han transcurrido 4 meses, la *equis* correspondiente será igual a $4/12$ o $4 \times 0.833(3)$ ya que a cada mes le corresponde $1/12$ de *equis*. En forma parecida se puede determinar matemáticamente el valor de y para cualquier fracción de año, aun cuando también se puede recurrir al trazo de la línea teórica y determinar gráficamente qué valor de la recta le corresponde al valor dado de x.

PRONÓSTICO DE LA TENDENCIA

El mismo cálculo de la recta teórica roporciona un medio de pronóstico en cuanto al continuar la serie de las *equis* más allá de las observaciones hechas, o de las unidades de tiempo realmente transcurridas, y substituir ese valor de las *equis* en la ecuación general de la recta, se obtienen los valores que probablemente alcanzará el fenómeno en un momento dado del futuro.

A esta *extrapolación* previsora puede oponerse una extrapolación retrospectiva que, dando a las *equis* valores inferiores a los consignados en la serie, permite conocer la forma en que probablemente se desarrolló el fenómeno en un pasado para el que se carece de datos realmente observados.

PARÁBOLA DE 2º GRADO

Si consideramos una serie como la siguiente,

x	y
-4	16
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	9
4	16

podremos ver, por su representación gráfica, que nos enfrentamos a una serie dinámica que no queda representada por una recta, sino por una curva que, en este caso particular, pasa por el origen y es simétrica con respecto al eje de las *yes*. Las curvas del tipo de la representada en la gráfica reciben el nombre de *parábolas de segundo grado* y tienen la propiedad (que no nos corresponde demostrar en este lugar) de que la distancia de cualquiera de sus puntos a un punto fijo llamado foco es igual a la distancia del punto que está sobre la curva a una recta fija llamada directriz.

Si, con objeto de estudiar las relaciones entre la serie dinámica numérica y su representación gráfica examinamos cuál es la relación que existe entre cada y y la *equis* correspondiente, podremos darnos cuenta de que:

x	y
-4	$16 = (-4)\ (-4) = (-4)^2$
-3	$9 = (-3)\ (-3) = (-3)^2$
-2	$4 = (-2)\ (-2) = (-2)^2$
-1	$1 = (-1)\ (-1) = (-1)^2$
0	
1	$1 = (\ 1)\ (\ 1) = (\ 1)^2$
2	$4 = (\ 2)\ (\ 2) = (\ 2)^2$
3	$9 = (\ 3)\ (\ 3) = (\ 3)^2$
4	$16 = (\ 4)\ (\ 4) = (\ 4)^2$

Cada y es igual al producto de dos factores iguales a *equis*, o sea, igual a la *equis* correspondiente elevada al cuadrado:

$$y = x^2$$

Tomemos en seguida, con propósitos de comparación, otras series semejantes en lo fundamental:

x	y
-4	$32 = 2 \times 16 = 2 \times (-4)^2$
-3	$18 = 2 \times 9 = 2 \times (-3)^2$
-2	$8 = 2 \times 4 = 2 \times (-2)^2$
-1	$2 = 2 \times 1 = 2 \times (-1)^2$
0	0
1	$2 = 2 \times 1 = 2 \times (\ 1)^2$
2	$8 = 2 \times 4 = 2 \times (\ 2)^2$
3	$18 = 2 \times 9 = 2 \times (\ 3)^2$
4	$32 = 2 \times 16 = 2 \times (\ 4)^2$

Caso en que las *yes* son iguales a los cuadrados de las *equis* correspondientes multiplicados por una cantidad constante, 2, o sea en que:

$$y = 2x^2$$

O, si representamos en forma más general, la cantidad constante por b:

$$y = bx^2$$

En la gráfica correspondiente, las ramas de la parábola de segundo grado están más próximas entre sí que las de la parábola anterior.

Considérese la serie:

x	y
-4	$35 = 3 + 32 = 3 + 2\ (-4)^2$
-3	$21 = 3 + 18 = 3 + 2\ (-3)^2$
-2	$11 = 3 + 8 = 3 + 2\ (-2)^2$
-1	$5 = 3 + 2 = 3 + 2\ (-1)^2$
0	$3 = \dots\dots\dots\dots$
1	$5 = \dots\dots\dots\dots$
2	$11 =$
3	$21 =$
4	$35 =$

caso en que las *yes* son iguales al cuadrado de las *equis* multiplicado por una cantidad constante (2), y adicionado de una cantidad también constante (3), o sea en que:

$$y = 3 + 2x^2$$

O, si representamos en forma más general, 2, por b, y 3 por a:

$$y = a + bx^2$$

En la gráfica correspondiente, la parábola corta al eje de las *yes* por encima del origen, pero la abertura de las ramas de la parábola sigue siendo la misma del caso anterior, en virtud de que en ambas series el coeficiente de x^2 es el mismo.

Si construímos una serie dinámica en la que además de figurar los anteriores elementos (término independiente de $x : a$, y término en $x^2 : bx^2$) hiciéramos figurar un término en $x : cx$) tendríamos una serie del tipo:

x	y
-4	$19 = 35 - 16 = 3 + 2\,(-4)^2 + 4\,(-4)$
-3	$9 = 21 - 12 = 3 + 2\,(-3)^2 + 4\,(-3)$
-2	$3 = 11 - 8 = 3 + 2\,(-2)^2 + 4\,(-2)$
-1	$1 = 5 - 4 = 3 + 2\,(-1)^2 + 4\,(-1)$
0	$3 = 3 - 0 = 3 + 2\,(0)^2 + 4\,(0)$
1	$9 = 5 + 4 = 3 + 2\,(1)^2 + 4\,(1)$
2	$19 = 11 + 8 = 3 + 2\,(2)^2 + 4\,(2)$
3	$33 = 21 + 12 = 3 + 2\,(3)^2 + 4\,(3)$
4	$51 = 35 + 16 = 3 + 2\,(4)^2 + 4\,(4)$

En este caso, cada *ye* es igual al cuadrado de la *equis* correspondiente multiplicado por una cantidad constante (2), adicionada de otra cantidad también constante (3), y del producto de la *equis* correspondiente multiplicada por una cantidad constante (4); es decir:

$$y = a + bx^2 + cx$$

en general, o, en el caso concreto:

$$y = 3 + 2x^2 + 4x$$

Podemos ver que, si representamos gráficamente esta serie, seguimos obteniendo una parábola de segundo grado, sólo que, en este caso no se mantiene la simetría con respecto al eje de las *yes*, sino que se desplaza hacia una recta paralela al propio eje.

Si la expresión $y = a + bx^2 + cx$ se ordena conforme a las potencias crecientes de x, se tiene:

$$\underline{y = a + cx + bx^2}$$

Que podemos considerar como expresión bastante general de la ecuación de segundo grado, o sea, de la ecuación representativa de una parábola de segundo grado, pudiendo afirmarse que:

Una parábola de segundo grado queda representada por una ecuación en la cual la variable independiente (x) está elevada al cuadrado como potencia máxima.

Todo lo anterior significa que además de los términos usuales pueden existir toda una serie de potencias de x siempre que ninguna de ellas sea superior a dos, y en todo caso la curva representada seguirá siendo una parábola de 2º grado, si, al mismo tiempo, dichas potencias son enteras y positivas.

Si en la ecuación de segundo grado anulamos el término en x^2, la ecuación se convierte en una ecuación de primer grado, y cada una de las ramas de la parábola se convierte en una recta.

Si en la ecuación de segundo grado se modifican convenientemente los signos (b negativa), la parábola en vez de abrirse hacia arriba se abrirá hacia abajo. Tanto en uno como en otro caso no puede afirmarse tan sencillamente como en el caso de la recta que el movimiento sea creciente o decreciente, ya que el fenómeno crecerá hasta determinado punto para después decrecer, o decrecerá hasta determinado punto para después crecer. En términos generales, cuando b es positiva, el fenómeno será primero decreciente y después creciente, y cuando b es negativa el fenómeno será primeramente creciente y después decreciente.

Velocidad en el caso de un fenómeno que obedece a una tendencia parabólica.—Si en un movimiento rectilíneo establecemos la relación entre el espacio, la velocidad y el tiempo, en un momento determinado el espacio recorrido (y) será igual a la velocidad (que llamaremos b) multiplicada por el tiempo (que representaremos por x); es decir:

$$y = bx$$

Si hacemos que el tiempo aumente en Δx, el espacio aumentará en Δy.

$$y + \Delta y = (x + \Delta x)b$$

$$y + \Delta y = bx + b\Delta x$$

Si restamos de esta ecuación la original no incrementada $y = bx$, tendremos:

$$\Delta y = b\Delta x$$

Si pasamos a Δx al otro miembro, tendremos:

$$\frac{\Delta y}{\Delta x} = b$$

Si hacemos que x tienda hacia cero, por definición, la relación entre el incremento de y y el incremento de x tenderá hacia la derivada de y con respecto a x:

$$\lim_{x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = b$$

> b no varía a pesar de que x tienda a cero porque b es una constante, independiente de los cambios que experimenta x.

Según la expresión anterior, teniendo en cuenta que b representa la velocidad del movimiento rectilíneo considerado, podremos establecer que:

La velocidad en el movimiento rectilíneo es igual a la derivada del espacio (y) con respecto al tiempo (x).

Como b es una constante, puede afirmarse que, en el movimiento rectilíneo la velocidad es constante (de ahí que las primeras diferencias entre las *yes* de dicho movimiento sean iguales en el caso de la tendencia rectilínea).

En términos más generales, se establece que:

Para cualquier clase de movimiento, rectilíneo y uniforme o no, se define la velocidad en un instante dado como el límite de la velocidad media cuando el incremento temporal tiende hacia cero, o bien:

La velocidad en un instante dado es la derivada de la distancia con respecto al tiempo.

Según esto, para encontrar la velocidad de un movimiento parabólico, buscaremos la derivada de y con respecto a x en la ecuación correspondiente:

$$y = a + cx + bx^2$$

Para derivar con respecto a x, incrementaremos las *equis* y anotaremos el correspondiente incremento de y:

$$y + \Delta y = a + c(x + \Delta x) + b(x + \Delta x)^2$$

$$y + \Delta y = a + cx + c\Delta x + bx^2 + 2bx\,x + b(\Delta x)^2$$

Al restar la primera ecuación de esta última, se obtiene:

$$y = c\,x + 2bx\,x + b(\Delta x)^2$$

Si dividimos ambos miembros de la ecuación por Δx:

$$\frac{\Delta y}{\Delta x} = c + 2bx + b\Delta x$$

Al tomar el límite de la relación $\Delta y/\Delta x$ cuando x tiende hacia cero, tenemos:

$$\lim_{x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = c + 2bx$$

O sea, que la velocidad del movimiento parabólico (derivada de y con respecto a x) está dada por la expresión: $c + 2bx$.

Si se examina la expresión anterior, se verá que está constituída por una cantidad constante c, y por el producto de dos constantes 2 y b multiplicado por una variable x; o sea, que a diferencia de lo que ocurre en el movimiento rectilíneo, en el movimiento curvilíneo (parabólico) la velocidad no es constante, sino variable, y que al preguntarse por ella debe especificarse a la velocidad de qué momento se refiere uno; así, si suponemos que se tratara de una de nuestras series representadas previamente en la que $y = 3 + 4x + 2x^2$, la velocidad estaría dada por una expresión de la forma: $c + 2bx$, o sea: $4 + 2 \times 2x = 4 + 4x$. De este modo:

En el año	*La velocidad será:*
-4	$4 + 4(-4) = 4 - 16 = -12$
-3	$4 + 4(-3) = 4 - 12 = -8$
-2	$4 + 4(-2) = 4 - 8 = -4$
-1	$4 + 4(-1) = 4 - 4 = 0$
0	$4 + 4(0) = 4 + 0 = 4$
1	$4 + 4(1) = 4 + 4 = 8$
2	$4 + 4(2) = 4 + 8 = 12$
3	$4 + 4(3) = 4 + 12 = 16$
4	$4 + 4(4) = 4 + 16 = 20$

Esto quiere decir que en el primer momento de nuestras observaciones (que hemos llamado -4 conforme al procedimiento del año mediano) la velocidad es de 12 unidades de y por cada unidad de *equis* y, como dicha velocidad es negativa, el fenómeno es decreciente (por ejemplo: la producción disminuye en 12 toneladas por año), en tanto que, en el segundo año de nuestras observaciones el decrecimiento en la producción es menor puesto que sólo es de 8 (la producción disminuye en ocho toneladas por año en ese momento), y a partir del quinto año de nuestras observaciones (marcado con 0) la producción aumenta en vez de disminuir (velocidades de signo positivo) en 4 unidades de y por cada año en el sexto año, en 8 unidades en el séptimo, etc.

Por las cifras dadas en la segunda columna de nuestra pequeña tabulación puede verse que la velocidad no es constante, sino variable, y que, en el caso concreto que estudiamos va en continuo aumento (pasar de -12 a -8 representa un aumento tanto como pasar de $+8$ a $+12$). Puede verse asimismo que ese aumento en la velocidad es constante (de 4 unidades por año) y que las cifras correspondientes, al representarse gráficamente quedan dadas por una recta.

La ecuación correspondiente la obtendremos de $v = c + 2bx$ que es la ecuación de una recta en la que v representa las magnitudes variables del fenómeno en el tiempo, c la ordenada en el origen, y $2b$ el ritmo de crecimiento de la velocidad por cada unidad de tiempo. Dicho ritmo de crecimiento de la velocidad por unidad de tiempo es lo que se conoce como aceleración. En virtud de que $2b$ es un producto de constantes (constante por ello mismo) podemos afirmar que la aceleración en un movimiento parabólico es una cantidad constante igual al duplo del coeficiente de x^2 en la ecuación general de la parábola.

$2b$ puede obtenerse buscando la derivada de v con respecto a x en la ecuación $v = c + 2bx$, ya que si incrementamos, buscamos la relación entre los incrementos y tomamos límites, obtendremos:

$$\frac{dv}{dx} = 2b$$

De ahí que se afirme que la aceleración ($2b$ en este caso) es la primera derivada de la velocidad.

Como, por otra parte:

$$\frac{dy}{dx} = c + 2bx = v$$

O sea que la velocidad es la primera derivada de *ye* con respecto a x, puede afirmarse que:

La aceleración es la primera derivada de la primera derivada de *ye* con respecto a x; o bien que:

La aceleración es la segunda derivada de *ye* con respecto a *equis.*

Interpolación de una parábola de segundo grado: En la interpolación de una parábola de segundo grado pueden utilizarse en su gran mayoría los métodos utilizados en la interpolación de la recta, especialmente:

1.—El Método de Elección de Puntos.

En este caso, los puntos a elegir serán 3, ya que existen *tres* parámetros (a, c, b) en la ecuación general de la parábola de 2º grado, cuyos valores es necesario determinar y, por lo mismo, se requieren *tres* ecuaciones, de la forma:

$$y_1 = a + cx_1 + bx_1^2$$

$$y_2 = a + cx_2 + bx_2^2$$

$$y_3 = a + cx_3 + bx_3^2$$

en las cuales y_1 y x_1 constituyen las coordenadas del primer punto, y_2 y x_2 las coordenadas del segundo punto, y y_3 y x_3 las coordenadas del tercer punto elegidos sobre la recta.

Con propósitos de simplificación, puede prescindirse del término en x, con lo que la ecuación de la parábola se reduce a $y = a + bx^2$, siendo por tanto necesarias sólo 2 ecuaciones, ya que sólo existen 2 parámetros (a, y b), y requiriéndose por lo mismo la elección de sólo 2 puntos sobre la parábola.

2.—El Método de los Mínimos Cuadrados.

El criterio para encontrar las ecuaciones de interpolación es el mismo que el utilizado en el caso de la recta: los valores encontrados al través de ellas deben ser tales que la suma de los cuadrados de las diferencias entre las *yes* reales y las *yes* teóricas sea un mínimo. Los resultados obtenidos tienen que ser distintos de los que se obtuvieron en el caso de la recta.

Obtención de las Fórmulas de Interpolación para una parábola de segundo grado.—Partiremos de nuestro criterio de los mínimos cuadrados expresados matemáticamente:

$$S(d^2) = \text{mínimo}$$
$$d = y_r - y_t$$
$$S(y_r - y_t)^2 = \text{mínimo}$$

Como las *yes* teóricas en este caso se calculan aplicando la fórmula: $y_t = a + cx + bx^2$, tendremos:

$$S(y_r - (a + cx + bx^2)^2 = S(y - a - cx - bx^2)^2 \text{ debe ser un mínimo.}$$

Para lograr que esa suma de cuadrados sea un mínimo es necesario encontrar la derivada de la expresión y, en seguida, anularla. Si derivamos con respecto al primer parámetro obtenemos:

$$S\,2(y - a - cx - bx^2)\,\frac{da}{da} = S\,2(y - a - cx - bx^2)$$

$$= 2(Sy - S\,a - Scx - Sbx^2) = 0$$

para que la expresión sea un mínimo.

$$= Sy - Sa - Scx - Sbx^2 = 0$$

$$Sy = Sa + Scx + Sbx^2$$

Como la suma de un conjunto de constantes iguales Sa, es igual a tantas veces la constante (a) como veces aparezca, tendremos:

$$\underline{Sy = Na + Scx + Sbx^2}$$

como primera ecuación de interpolación.

Si derivamos la misma expresión con relación al segundo parámetro c, obtenemos:

$$S\,2(y - a - cx - bx^2)\,\frac{dc}{dc}\,x = S\,2(y - a - cx - bx^2)\,x =$$

$$2\,S(yx - ax - cx^2 - bx^3) = 0$$

para que la expresión sea un mínimo.

$$Syx - Sax - Scx^2 - Sbx^3 = 0$$

$$\underline{Syx = Sax + Scx^2 + Sbx^3}$$

Esta última es nuestra segunda ecuación de interpolación.

Si derivamos la misma expresión con relación al tercer parámetro b, obtenemos:

$$S\,2(y - a - cx - bx^2)\,\frac{db}{db}\,x^2 = S\,2(y - a - cx - bx^2)\,x^2 =$$

$$2\,S(yx^2 - ax^2 - cx^3 - bx^4) = 0$$

para que la expresión sea un mínimo.

$$Syx^2 - S\,ax^2 - Scx^3 - Sbx^4 = 0$$

$$\underline{Syx^2 = Sax^2 + Scx^3 + Sbx^4}$$

Esta última constituye nuestra tercera ecuación de interpolación.

Reunidas las tres ecuaciones de interpolación necesarias para la aplicación del método general de los mínimos cuadrados, se tiene:

$$Sy = Na + S(x)c + S(x^2)b$$

$$Syx = Sxa + S(x^2)c + S(x^3)b$$

$$S(yx^2) = Sx^2a + S(x^3)c + S(x^4)b$$

Con propósitos nemotécnicos, puede observarse que la primera ecuación de interpolación se obtiene colocando un operador $\Sigma = S$ delante de cada uno de los términos de la ecuación general de segundo grado; que la segunda ecuación de interpolación se obtiene multiplicando cada uno de los términos de la ecuación general de segundo grado por la primera potencia de la variable independiente (por x) y agregando delante de cada uno de ellos un operador $\Sigma = S$; que la tercera ecuación de interpolación se obtiene multiplicando cada uno de los términos de la ecuación general de segundo grado por la segunda potencia de la variable independiente (por x^2) y agregando delante de cada producto un operador $\Sigma = S$.

Si para mayor simplicidad se utiliza en lugar de la forma completa de la ecuación de segundo grado $y = a + cx + bx^2$ la forma incompleta $y = a + bx^2$, se requerirá de sólo dos ecuaciones de interpolación, obtenidas por derivación y ulterior anulación de la derivada de $S(y - a - bx^2)^2$ con respecto a los parámetros a y b. Al realizar la derivación se encuentran como ecuaciones las siguientes:

$$Sy = Na + S(x^2)b$$

$$\underline{Syx^2 = S(x^2)a + S(x^2)^2b}$$

Con objeto de recordar estas ecuaciones si se recuerdan las de la recta, bastará con compararlas para darse cuenta de que, en dondequiera aparece x en las ecuaciones de la recta, aparece x^2 en las ecuaciones de la parábola de segundo grado.

Simplificaciones del método de los mínimos cuadrados en el caso de la interpolación de una parábola de segundo grado.

1.—Método del Año Mediano.

Si bien la elección de un año mediano reduce las magnitudes de los valores de x y facilita las operaciones aritméticas, a diferencia de lo que ocurre en el caso de la recta, dicho artificio no anula ninguno de los términos que intervienen en las ecuaciones de interpolación:

$$Sy = Na + S(x^2)b$$

$$S(yx^2) = S(x^2)a + S(x^4)b$$

ya que en ellas intervienen únicamente potencias pares de x, y las potencias pares tanto de números negativos como de números positivos son siempre positivas y, por lo mismo, no se anulan unas con otras. En cambio, cuando se utiliza la forma completa, y, consecuentemente el sistema de tres ecuaciones:

$$Sy = Na + S(x)c + S(x^2)b$$

$$S(yx) = S(x)a + S(x^2)c + S(x^3)b$$

$$S(yx^2) = S(x^2)a + S(x^3)c + S(x^4)b$$

puede mostrarse que tanto $S(x)$ como $S(x^3)$ se anulan, con lo cual desaparecen de esas ecuaciones los términos que las contienen, o sea:

$$Sy = Na + S(x^2)b$$

$$S(yx) = S(x^2)c$$

$$S(yx^2) = S(x^2)a + S(x^4)b$$

Como puede verse por estas expresiones, c se puede obtener directamente de la (2), despejando:

$$c = \frac{S(yx)}{S(x^2)}$$

y los otros dos parámetros, a y b, mediante la solución del sistema de dos ecuaciones con dos incógnitas:

$$Sy = Na + S(x^2)b$$

$$S(yx^2) = S(x^2)a + S(x^4)b$$

Puede observarse que el valor de c (coeficiente de la primera potencia de x en la parábola) está dado en términos análogos a aquellos en los que en el mismo procedimiento del año medaino está dado b (coeficiente de la primera potencia de x en la recta) en el caso de la recta.

2.—Valor de los parámetros en términos del número de datos.—A fin de encontrar la relación entre los parámetros de la ecuación de segundo grado y el número de datos, es necesario tener presentes las siguientes relaciones:

$$S(x) = \frac{N(N+1)}{2}$$

$$S(x^2) = \frac{N(N+1)}{2} \frac{(2N+1)}{3} = S(x) \frac{(2N+1)}{3}$$

$$S(x^3) = \left(\frac{N(N+1)}{2} \right)^2 = S(x)^2$$

$$S(x^4) = \frac{N^5 - N}{5} + 2S(x^3) - 2S(x^2) + Sx$$

Mediante la substitución de estos valores en las tres ecuaciones de interpolación correspondientes, y la resolución del sistema por medio de determinantes de tercer orden, pueden obtenerse los valores de los tres parámetros en términos del número de datos.

PARÁBOLA DE 3er. GRADO

Mediante la representación gráfica de la serie siguiente:

x	y
-4	-64
-3	-27
-2	-8
-1	-1
0	0
1	1
2	8
3	27
4	64

Obtenemos una curva, una de cuyas ramas se encuentra en el primer cuadrante, en tanto que la otra se encuentra en el tercero ya que, mientras para una corresponden abscisas positivas y ordenadas también positivas, en la otra se corresponden abscisas y ordenadas negativas. Si comparamos esta curva con alguna de nuestras parábolas de segundo grado, podremos notar que la rama que se encuentra en el primer cuadrante podría pasar por rama de una parábola de 2º grado, y que la rama que se encuentra en el tercer cuadrante podría tomarse por rama de otra parábola de segundo grado; o bien, que esta segunda rama podríamos imaginarla como la rama que se encontraría en el segundo cuadrante si se tratara de una parábola de 2º grado, y la cual se hubiese flexionado o doblado hasta llevarla a su posición actual en el tercer cuadrante. Quizás de estas semejanzas que pueden establecerse imaginativamente, proceda el nombre de esta curva a la que se denomina también "parábola" de tercer grado.

Si tratamos de ver cuáles son las relaciones que existen entre cada *ye* y la *equis* correspondiente de este serie, podremos darnos cuenta de que:

x	y
-4	$-64 = (-4)(-4)(-4) = (-4)^3$
-3	$-27 = (-3)(-3)(-3) = (-3)^3$
-2	$-8 = (-2)(-2)(-2) = (-2)^3$
-1	$-1 = (-1)(-1)(-1) = (-1)^3$
0	0
1	$1 = 1 \times 1 \times 1 = (1)^3$
2	$8 = 2 \times 2 \times 2 = (2)^3$
3	$27 = 3 \times 3 \times 3 = (3)^3$
4	$64 = 4 \times 4 \times 4 = (4)^3$

Según esto, en nuestra serie, cada una de las *yes* es igual al cubo o ter cera potencia de la *equis* correspondiente; es decir: $y = x^3$.

Como en el caso de la recta y la parábola de segundo grado, la potenci de x puede aparecer multiplicada por un factor constante distinto de uno (po ejemplo, b), lo cual hará variar la inclinación de las ramas de la curva; pue de asimismo cortar al eje por encima o por debajo del origen, a una distan cia a, y pueden además, intervenir en la formación de cada *ye* las potencia de *equis* inferiores a la tercera. De este modo, la ecuación general para est tipo de curvas (parábolas de tercer grado) será:

$$y = a + cx + dx^2 + bx^3$$

Cuando se pretende simplificar el proceso de interpolación de una pará bola de tercer grado, puede prescindirse de los términos en x y en x^2 y traba jar con la expresión:

$$y = a + bx^3$$

Puede observarse que, a diferencia de lo que ocurre con la parábola d segundo grado, la parábola de tercer grado es simplemente creciente o sim plemente decreciente según el signo de b.

Para determinar el ritmo de crecimiento o decrecimiento procederemo ateniéndonos a nuestra definición de velocidad, según la cual ésta es igual la derivada del espacio con respecto al tiempo, o, en series cronológicas, la derivada de *ye* con respecto a x.

Si derivamos:

$$y = a + cx + dx^2 + bx^3$$

con respecto a *equis*, obtenemos:

$$\frac{dy}{dx} = 0 + c\frac{dx}{dx} + 2dx\frac{dx}{dx} + 3bx^3\frac{dx}{dx}$$

$$\frac{dy}{dx} = c + 2dx + 3bx^2 = \text{velocidad.}$$

Por la anterior expresión, puede verse que la velocidad en este tipo de fenómenos no es constante sino variable, ya que en su valor intervienen x y x^2 que son valores variables.

Si tomamos la derivada de la velocidad con respecto al tiempo, obtendremos la aceleración (*i. e.* la segunda derivada de *ye* con respecto a *equis*).

$$\frac{dv}{dx} = 0 + 2d\frac{dx}{dx} + 3.2bx\frac{dx}{dx}$$

$$\frac{dv}{dx} = 2d + 6bx = \text{aceleración.}$$

Por esta expresión podremos darnos cuenta de que la aceleración en un fenómeno regido por un crecimiento parabólico de tercer grado no es constante sino variable (a diferencia de lo que ocurre en el caso de la parábola de segundo grado en la que si bien la velocidad es variable, la aceleración es constante). La aceleración del movimiento sujeto a una ley parabólica de tercer grado, según puede verse por la expresión, está sujeta a un crecimiento rectilíneo ($2d$ puede hacerse igual al término independiente de dicho crecimiento), en estas condiciones no es extraño el que la aceleración aumente en cantidades constantes: en $6b$ por cada aumento unitario de b; o sea, que si bien ni la primera derivada de y con respecto a x (velocidad) ni la segunda derivada de *ye* con respecto a *equis* (aceleración) son constantes, en cambio sí lo es la tercera derivada de *ye* con respecto a *equis* (o sea la primera derivada de la aceleración con respecto al tiempo), de tal manera que, en un fenómeno que crece conforme a una parábola de tercer grado, los terceros incrementos de la variable dependiente son constantes.

Interpolación de una parábola de tercer grado por el Método de los Mínimos Cuadrados.—Para obtener las ecuaciones de interpolación correspondientes partiremos como siempre, del criterio de que $S(d^2)$ sea un mínimo o sea que $S(y_r - y_t)^2$ sea un mínimo; en el caso concreto:

$$S(y_r - y_t)^2 = S(y - a - cx - dx^2 - bx^3)^2$$

Como para obtener un mínimo necesitamos anular las diversas derivadas de la función con relación a cada uno de los parámetros a, c, d, b, comenzaremos por derivar la función con respecto a a:

$$S\,2(y - a - cx - dx^2 - bx^3)$$

es la derivada correspondiente que, para producir un mínimo, debe de anularse:

$$2\,S(y - a - cx - dx^2 - bx^3) = 0$$

$$S(y - a - cx - dx^2 - bx^3) = 0$$

$$Sy - Sa - Scx - Sdx^2 - Sbx^3 = 0$$

$$Sy = Sa + Scx + Sdx^2 + Sbx^3$$

$$\underline{Sy = Na + S(x)c + S(x^2)d + S(x^3)b}$$

En forma análoga, si derivamos para c, tendremos:

$$S\,2(y - a - cx - dx^2 - bx^3)x\frac{dc}{dc} = 2\,S(y - a - cx - dx^2 - bx^3)x = 0$$

$$2(Syx - Sax - Scx^2 - Sdx^3 - Sbx^4) = 0$$

$$Syx - Sax - Scx^2 - Sdx^3 - Sbx^4 = 0$$

$$\underline{Syx = S(x)a + S(x^2)c + S(x^3)d + S(x^4)b}$$

Si derivamos para d, obtenemos:

$$S\,2(y - a - cx - dx^2 - bx^3)x^2\frac{dd}{dd} = 2\,S(y - a - cx - dx^2 - bx^3)x^2$$

$$2(Syx^2 - Sax^2 - Scx^3 - Sdx^4 - Sbx^5) = 0$$

$$\underline{Syx^2 = S(x^2)a + S(x^3)c + S(x^4)d + S(x^5)b}$$

Si derivamos para b, tendremos:

$$S\,2(y - a - cx - dx^2 - bx^3)x^3\frac{db}{db} = 2\,S(y - a - cx - dx^2 - bx^3)x^3$$

$$2(Syx^3 - Sax^3 - Scx^4 - Sdx^5 - Sbx^6) = 0$$

$$Syx^3 = S(x^3)a + S(x^4)c + S(x^5)d + S(x^6)b$$

Según esto, las ecuaciones de interpolación para una parábola de tercer grado serán:

$$\begin{aligned} Sy &= Na + S(x)c + S(x^2)d + S(x^3)b \\ Syx &= S(x)a + S(x^2)c + S(x^3)d + S(x^4)b \\ Syx^2 &= S(x^2)a + S(x^3)c + S(x^4)d + S(x^5)b \\ Syx^3 &= S(x^3)a + S(x^4)c + S(x^5)d + S(x^6)b \end{aligned}$$

Para obtener los valores de los parámetros, a, b, c, d, que será necesario substituir en la ecuación general de la parábola de tercer grado, habrá necesidad de buscar los valores de Sy, N, $S(x)$, $S(x^2)$, $S(x^3)$ ... Syx, Syx^2, etc. y substituirlos en estas ecuaciones. Una vez substituídos todos los valores que pueden obtenerse a partir de la elaboración de los datos de la serie empírica, será necesario resolver el sistema de cuatro ecuaciones con cuatro incógnitas (los parámetros a, b, c, d) mediante determinantes de cuarto orden.

Procedimiento del año mediano.—Las simplificaciones que se obtienen mediante la numeración de la serie de las *equis* a ambos lados del lado mediano se reflejan en la anulación de las sumas de potencias impares de *equis* y, por tanto, de los términos que contienen dichas sumas como coeficientes en las ecuaciones previamente estudiadas; de este modo, dichas ecuaciones se convierten en:

$$\begin{aligned} Sy &= Na + S(x^2)d \\ Syx &= S(x^2)c + S(x^4)b \\ Syx^2 &= S(x^2)a + S(x^4)d \\ Syx^3 &= S(x^4)c + S(x^6)d \end{aligned}$$

Puede notarse que, de las ecuaciones así simplificadas pueden formarse dos sistemas, el primero de los cuales está constituído por la primera y la tercera ecuaciones, de las que se pueden obtener a y d, en tanto que el segundo sistema está formado por la segunda y cuarta ecuaciones, de las que se pueden obtener c y d. O sea, que la simplificación consiste en que, en lugar de trabajar con un sistema de cuatro ecuaciones con cuatro incógnitas (que requiere uso de determinantes de cuarto orden) basta con resolver dos sistemas de dos ecuaciones con dos incógnitas, independientes (los sistemas de ecuaciones) entre sí: los sistemas resultan ser:

$$Sy = Na + S(x^2)d \qquad Syx = S(x^2)c + S(x^4)b$$

$$Syx^2 = S(x^2)a + S(x^4)d \qquad Syx^3 = S(x^4)c + S(x^6)b$$

Interpolación de la parábola de tercer grado en su forma incompleta.— A menudo resulta demasiado embarazoso trabajar con la forma completa de la ecuación de tercer grado, por lo cual se prefiere trabajar con su forma incompleta: $y = a + bx^3$.

En tal caso, es fácil demostrar que por derivación con respecto a los parámetros a y b se obtienen las dos ecuaciones de interpolación correspondientes.

$$S(d^2) = S(y_r - y_t)^2 = S(y - a - bx^3)^2$$

$$\frac{d}{da}S(y - a - bx^3)^2 = S\,2(y - a - bx^3)\,\frac{da}{da}$$

$2\,S(y - a - bx^3) = 0$ para obtener un mínimo.

$$S(y - a - bx^3) = 0$$

$$Sy - Sa - Sbx^3 = 0$$

$$Sy = Na - S(x^3)b$$

$$\frac{d}{db}S(y - a - bx^3)^2 = S\,2(y - a - bx^3)x^3\frac{db}{db}$$

$$2\,S(y - a - bx^3) = 0$$

$$S(yx^3 - ax^3 - bx^6) = 0$$

$$Syx^3 - Sax^3 - Sbx^6 = 0$$

$$Syx^3 = S(x^3)a + S(x^6)b$$

O sea, que el sistema de ecuaciones de interpolación estará formado por:

$$Sy = Na + S(x^3)b$$

$$Syx^3 = S(x^3)a + S(x^6)b$$

Si se emplea el procedimiento del año mediano, $S(x^3)$ se anula y, por lo mismo, los términos que lo contienen como coeficiente desaparecen de las ecuaciones anteriores que, de esta forma, quedan reducidas a:

$$Sy = Na \qquad \therefore\ a = \frac{Sy}{N}$$

$$Syx^3 = S(x^6)\,b \qquad\qquad b = \frac{Syx^3}{S(x^6)}$$

HIPÉRBOLA

Consideremos la siguiente serie:

x	y
1	1.000
2	0.500
3	0.333
4	0.250
5	0.200
6	0.166
7	0.142
8	0.125
9	0.111
10	0.100

La gráfica nos muestra la rama localizada en el cuadrante I, de una curva que posee una segunda rama (poco o nada utilizada en Estadística) localizada en el III cuadrante; o sea una curva simétrica con respecto a la bisectora de los cuadrantes II y IV; curva que tiene la propiedad de que al aumentar los valores de *equis*, *y* se acerca más y más al eje de las *equis* sin llegar a confundirse con él, en tanto que al disminuir los valores de *equis*, la propia curva se acerca más y más al eje de las *yes* sin llegar a confundirse con él, hecho que se expresa diciendo que la curva tiene como *asíntotas* los dos ejes coordenados. La curva que tiene dichas propiedades se conoce con el nombre de *hipérbola*.

Podemos observar en la serie dinámica cuya representación se ha obtenido por medio de esta hipérbola (equilátera), que si multiplicamos cada una de las *equis* por las *yes* correspondientes, obtenemos como producto la *unidad*. En efecto:

x	y
$1 \times$	$1.000 = 1$
$2 \times$	$0.500 = 1$
$3 \times$	$0.333 = 1$

Este hecho puede expresarse mediante la ecuación:

$$xy = 1$$

Y como, por definición, el recíproco de un número es el

$4 \times 0.250 = 1$
$5 \times 0.200 = 1$
$6 \times 0.166 = 1$
$7 \times 0.142 = 1$
$8 \times 0.125 = 1$
$9 \times 0.111 = 1$
$10 \times 0.100 = 1$

número que multiplicado por él produce la unidad, podemos afirmar que en esta serie (representada por una hipérbola) las *yes* son iguales al recíproco de las *equis*, lo cual se puede expresar asimismo, en la serie, como:

x	y
1	$1.000 = 1/1$
2	$0.500 = 1/2$
3	$0.333 = 1/3$
4	$4.250 = 1/4$
5	$0.200 = 1/5$
6	$0.166 = 1/6$
7	$0.142 = 1/7$
8	$0.125 = 1/8$
9	$0.111 = 1/9$
10	$0.100 = 1/10$

O sea, que cada una de las *yes* es igual a la unidad (1) dividida entre la *equis* correspondiente, por lo cual podemos escribir, en forma equivalente a la anterior:

$$y = \frac{1}{x}$$

Expresión que hubiéramos podido obtener de la anterior con despejar simplemente a y.

Pero $\frac{1}{x}$ también puede representarse por x^{-1} ya que si multiplicamos y dividimos x^{-1} por x, tenemos:

$$x^{-1} = \frac{(x^{-1})(x)}{x} = \frac{x^{-1+1}}{x} = \frac{x^0}{x} = \frac{1}{x}$$

En estas condiciones, en lugar de $1/x$ puede escribirse x^{-1}, con lo que la expresión se convierte en:

$$y = x^{-1}$$

Forma la más sencilla de representación algebraica de una hipérbola. Si suponemos una serie en la cual el producto de las *equis* por las *yes* sea una constante distinta de 1 (por ejemplo b) la forma de la curva no cambiará fundamentalmente, y las ecuaciones correspondientes resultarán ser:

$$xy = b$$

$$y = b\,\frac{1}{x} = \frac{b}{x}$$

$$y = bx^{-1}$$

Si en lugar de aparecer x elevada a -1 aparece elevada a potencias negativas constantes distintas de -1 (por ejemplo $-c$) se obtendrán curvas del mismo tipo, pero con diferente *pendiente.* Las expresiones correspondientes resultarán ser:

$$x^c y = b$$

$$y = b\frac{1}{x^c} = \frac{b}{x^c}$$

$$y = bx^{-c}$$

Puede suceder asimismo que aparezca un término adicional a, también constante, cuyo valor producirá el desplazamiento de la hipérbola hacia arriba o hacia abajo, con lo que se puede aceptar, como forma bastante general de la hipérbola, para usos estadísticos:

$$y = a + bx^{-c}$$

Interpolación de una Hipérbola por el Método de los Mínimos Cuadrados.—A esta altura de nuestro estudio, el único problema a resolver consistirá en la obtención de las correspondientes ecuaciones de interpolación. A fin de simplificar suficientemente el problema, consideraremos la interpolación de una hipérbola de la forma:

$$y = bx^{-c}$$

en la que $a = 0$, y, ulteriormente, de una ecuación de la forma:

$$y = a + bx^{-1}$$

en la que $c = 1$.

Fórmulas de interpolación para la hipérbola:

$$y = bx^{-c}$$

Dicha expresión puede transformarse por medio de los logaritmos en:

$$\log y = \log b + \log x^{-c} = \log b - c \log x$$

$$\log y = \log b - c \log x$$

Si representamos $\log y$ por Y, $\log b$ por A, $-c$ por B y $\log x$ por X tendremos la expresión:

$$Y = A + BX$$

O sea una expresión análoga a la ecuación de la recta. Si hacemos la interpolación correspondiente, obtendremos los valores de los parámetros

A y B, cuyos antilogaritmos corresponderán a los parámetros b y c de la expresión original.

Las fórmulas de interpolación para la expresión

$$Y = A + BX$$

serían:

$$S(Y) = NA + S(X)B$$

$$S(YX) = S(X)A + S(X^2)B$$

Si en estas fórmulas de interpolación substituímos por Y, X, A, B, los valores originales y, x, log b y $-c$, respectivamente, tendremos las dos ecuaciones de interpolación siguientes:

$$S(\log y) = N \log b + S(\log x)(-c)$$

$$S[(\log y)(\log x)] = S(\log x) \log b + S[(\log x)^2](-c)$$

Obtenido el valor de *log b* bastará con buscar su antilogaritmo para obtener el parámetro de la expresión original.

Expresado todo lo anterior en forma de procedimiento:

1.—Cuando se trabaja con la expresión $y = bx^{-c}$, se buscan los logaritmos de las *equis* y se anotan en una columna; se buscan los logaritmos de las *yes* y se anotan en otra columna.

2.—Se suman separadamente las columnas de los logaritmos de *equis* y la de los logaritmos de *ye*, para obtener $S(\log x)$ y $S(\log y)$ requeridas por las ecuaciones de interpolación.

3.—Se multiplica cada uno de los valores contenidos en la columna de los logaritmos de y, por cada uno de los correspondientes valores de la columna de los logaritmos de *equis* para obtener una columna de productos $(\log y)$ $(\log x)$ que sumados darán $S(\log y) \log x)$ requeridos por las fórmulas. *Precaución.*—No creer que se trata del logaritmo de un producto y convertirlo en suma de los logaritmos de los factores, ya que se trata del producto de dos logaritmos y, por lo mismo, es preciso encontrar primero el logaritmo y ulteriormente el producto y no primero el producto y después el logaritmo de ese producto.

4.—Cada uno de los valores contenidos en la columna de los logaritmos de x, se eleva al cuadrado sumando después los cuadrados para obtener $S[(\log x)^2]$. *Precaución.*—No confundir la suma de los cuadrados de los logaritmos con la suma de los logaritmos de los cuadrados de x; o sea, que es preciso buscar primero el logaritmo y en seguida elevar al cuadrado el valor encontrado y no buscar primero el cuadrado de *equis* y en seguida buscar su logaritmo.

5.—Se substituyen los valores de $S(\log y)$, N, $S(\log x)$, $S(\log y)$ $(\log x)$, $S(\log x)^2$ en las dos ecuaciones de interpolación dadas.

6.—Se resuelve el sistema de dos ecuaciones con dos incógnitas, con lo cual se obtienen los valores de $\log b$ y de $-c$.

7.—Se obtiene el antilogaritmo de $\log b$.

8.—Se substituyen los valores de b y $-c$ en la expresión:

$$\underline{y = bx^{-c}}$$

Fórmulas de interpolación para la hipérbola:

$$y = a + bx^{-1}$$

Conforme al criterio de los mínimos cuadrados:

$S(d^2) = S(y_r - y_t)^2 = S(y - a - bx^{-1})^2$, para ser un mínimo debe anular sus derivadas con respecto a a y a b; por lo que:

$$\frac{d}{da} S(y - a - bx^{-1})^2 = 2S(y - a - bx^{-1}) \frac{da}{da} = 0$$

$$2[Sy - Sa - Sbx^{-1})] = 0$$

$$Sy - Sa - Sbx^{-1} = 0$$

$$Sy = Na + S(x^{-1})b$$

Primera ecuación de interpolación para la hipérbola.

$$\frac{d}{db} S(y - a - bx^{-1})^2 = 2S(y - a - bx^{-1}) \frac{db}{db} x^{-1} = 0$$

$$2S[yx^{-1} - Sax^{-1} - Sbx^{-2}] = 0$$

$$Syx^{-1} - Sax^{-1} - Sbx^{-2} = 0$$

$$S(yx^{-1}) = S(x^{-1})a + S(x^{-2})b$$

$$S(yx^{-1}) = S(x^{-1})a + S(x^{-1})^2 b$$

Cualquiera de estas dos últimas expresiones puede tomarse como segunda ecuación de interpolación para la hipérbola.

La forma procesal de obtener los valores de a y de b conforme a lo anterior, será:

1.—Sumar los valores de las *yes* para obtener Sy.
2.—Buscar los valores de los recíprocos de x y sumarlos para obtener $S(x^{-1})$.
3.—Multiplicar cada uno de los valores obtenidos en esa forma por las *yes* correspondientes, para obtener $S(yx^{-1})$ mediante la suma de los productos respectivos.
4.—Elevar al cuadrado cada uno de los recíprocos de las *equis* y sumarlos para obtener $S(x^{-1})^2$.
5.—Substituir todos los valores anteriores en las dos fórmulas de interpolación.
6.—Resolver el sistema de ecuaciones para obtener (en este caso directamente) los valores de los parámetros a y b que deben substituirse en la ecuación:

$$y = a + bx^{-1}$$

LA CURVA DE PARETO DE DISTRIBUCIÓN DE LOS INGRESOS COMO CASO PARTICULAR DE CURVA HIPERBÓLICA

Una curva de tipo hiperbólico expresada por la ecuación

$$y = bx^{-c}$$

en la cual la pendiente c de la curva se ha encontrado empíricamente igual a 1.5 aproximadamente es lo que constituye la llamada Ley de Pareto (la curva correspondiente se llama Curva de Pareto) que rige la distribución del ingreso entre las diferentes capas de población; ley que, en términos corrientes expresa que muchos tienen pequeños ingresos, pocos tienen grandes ingresos y un número medio de personas tienen ingresos medios.

Pitman, usando los datos relativos al monto de los ingresos y al número de personas que disfrutan de determinados ingresos en Gran Bretaña e Irlanda del Norte ha determinado los siguientes valores que dan la ley del fenómeno en esos países:

$$\log y = 10.1560 - 1.565 \log x$$

1.565, valor de c da la pendiente de la curva hiperbólica correspondiente, o curva de Pareto.

LOGARÍTMICA

Si consideramos la siguiente serie dinámica, y la representamos gráficamente, obtenemos la curva logarítmica

x	y
1	0.0000
2	0.3010
3	0.4771
4	0.6020
5	0.6989
6	0.7781
7	0.8457
8	0.9030
9	0.9542
10	1.0000

La curva obtenida tiene la particularidad de que, en un principio, a incrementos dados de la variable independiente x corresponden grandes incrementos de la variable dependiente o función y, en tanto que, más tarde, a incrementos dados de la variable independiente x corresponden pequeños incrementos de la variable dependiente y; o sea, que a incrementos constantes de x corresponden incrementos decrecientes de y: el crecimiento se continúa, pero va haciéndose cada vez más lento. La forma de crecimiento representado por esta curva puede ponerse en relación de analogía con el crecimiento normal de un ser humano que, en las primeras etapas de su vida aumenta rápidamente de estatura y que, conforme avanza en edad crece cada vez menos sin que llegue a dejar de crecer totalmente.

Si, como en casos anteriores, examinamos las relaciones que ligan a las *yes* con las *equis* correspondientes, podremos darnos cuenta de que:

x	y
1	0.0000 = log. 1
2	0.3010 = log. 2
3	0.4771 = log. 3
4	0.6020 = log. 4
5	0.6989 = log. 5
6	0.7781 = log. 6

x	y
7	0.8451 = log. 7
8	0.9030 = log. 8
9	0.9542 = log. 9
10	1.0000 = log. 10

Es decir, que la serie de la que es representación la curva de cuyas características hemos hablado rápidamente en renglones anteriores está formada por una sucesión de valores en los que cada una de las *yes* es igual al logaritmo de la *equis* correspondiente, o sea:

$$y = \log.\ x$$

Una expresión más general de las series y curvas logarítmicas es:

$$y = a + b \log x$$

Puede demostrarse, mediante la aplicación del criterio de los mínimos cuadrados que las ecuaciones de interpolación aplicables al caso serán:

$$S(y) = Na + S(\log x)b$$
$$S(y \log x) = S(\log x)a + S[(\log x)^2]b$$

La simple búsqueda de las cantidades $S(y)$, $S(\log x)$, $S(y \log x)$ y $S[(\log x)^2]$, su substitución en estas ecuaciones, la resolución del sistema de ecuaciones para los parámetros a y b y la substitución de los valores de éstos en la ecuación de la logarítmica: $y = a + b \log x$ permite la interpolación de la curva correspondiente, el cálculo de los valores de las *yes* teóricas y la predicción matemática del valor que probablemente alcance el fenómeno en un momento dado del futuro.

EXPONENCIAL

Supongamos que se tiene la serie siguiente, y que se representa gráficamente:

x	y
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256

Si examinamos la curva que representa a la serie, podremos percatarnos de que en ella ocurre precisamente lo contrario de lo que sucede en el caso de la logarítmica, ya que, en un principio a incrementos dados de la variable independiente corresponden pequeños incrementos de la variable dependiente (compárense $y_1 - y_2$, $y_2 - y_3$, $y_3 - y_4$, etc.), en tanto que más tarde, a incrementos dados de la variable independiente corresponden grandes incrementos de la variable dependiente; o bien, que la función procede por incrementos crecientes (y no constantes) conforme la variable independiente aumenta por incrementos constantes.

El examen de los valores de y contenidos en la pequeña tabulación de la serie nos muestra que se trata de las sucesivas potencias de 2(2, su cuadrado, su cubo, su cuarta potencia, etc.), o sea que:

x	y
2	$4 = 2^2$
3	$8 = 2^3$
4	$16 = 2^4$
5	$32 = 2^5$
6	$64 = 2^6$
7	$128 = 2^7$
8	$256 = 2^8$

Si comparamos los exponentes de 2 con la *equis* correspondiente, podremos observar que para obtener el valor de *ye* fue preciso elevar a 2 al exponente x. O sea que:

$$y = 2^x$$

La serie podría haber estado formada no por las sucesivas potencias de 2, sino por las de 3, por las de 4, por las de b, en cuyo caso hubiéramos tenido:

$$y = 3^x$$

$$y = 4^x$$

$$y = b^x$$

O sea, que en esta clase de curvas cada *ye* puede considerarse igual a una constante b elevada a un exponente variable. Puede ocurrir, además, que los valores de ye procedan de esa potenciación de una constante el exponente variable x, pero multiplicado todo ello por otra constante a, lo cual no hace cambiar sustancialmente la forma de la relación entre las variables ni la forma de la curva. Es decir, que podemos aceptar como forma bastante general de la ecuación exponencial:

$$y = ab^x$$

Si, a fin de determinar el ritmo de crecimiento del fenómeno, buscamos su primera derivada, obtendremos:

$$\frac{dy}{dx} = \frac{d}{dx}(ab^x) = a\frac{d}{dx}(b^x) = ab^x L.b$$

En vista de que si bien a y $L.b$ (logaritmo natural de b) son constantes b^x es variable,, puede afirmarse que en este tipo de fenómenos la velocidad es variable y crece de unidad en unidad de tiempo en b veces (progresión geométrica).

A fin de obtener las fórmulas de interpolación para el caso de la exponencial, transformaremos la ecuación general original tomando los logaritmos correspondientes.

$$y = ab^x$$

$$\log. y = \log. (ab^x)$$

$$= \log. a + \log. (b^x)$$

$$\log. y = \log. a + x \log. b$$

Si comparamos esta ecuación con la de la recta, podremos notar la analogía entre ambas:

$$y = a + bx$$

$$\log. y = \log. a + x \log. b$$

En efecto, en donde aparece y en la ecuación de la recta, aparece log. y en la de la exponencial; en donde aparece a en la de la recta, aparece log. a en la exponencial, y en donde aparece b en la recta aparece log. b en la exponencial; es decir, que podemos interpolar una recta a una serie formada por las *equis* y por los logaritmos de las *yes*, obteniendo los parámetros correspondientes en forma logarítmica; o bien, que las fórmulas de interpolación utilizadas para la recta pueden transformarse convenientemente para que sirvan a la interpolación de una exponencial, con poner en los lugares correspondientes log. y por y, etc. Sería fácil demostrar que una substitución de este tipo satisfará el criterio de los mínimos cuadrados de hacer de la suma de los cuadrados de las diferencias entre las *yes teóricas* (logaritmos de las *yes* teóricas en este caso) y las *yes* reales (logaritmos de las *yes* reales en el caso de la exponencial) un mínimo.

De las ecuaciones de interpolación para la recta:

$$S(y) = Na + S(x)b$$

$$S(yx) = S(x)a + S(x^2)b$$

obtenemos:

$$S(\log. y) = N \log. a + S(x) \log. b$$

$$S(x \log. y) = S(x) \log. a + S(x^2) \log. b$$

Es fácil percatarse de la mayor facilidad que tiene trabajar con estas ecuaciones de interpolación que no con las correspondientes en números naturales:

$$Py = a^N P\, b^x$$

$$Py^x = Pa^x P\, b^{x^2}$$

En la que P (operador Π) indica "producto de..."

Para trabajar con las ecuaciones logarítmicas de interpolación en el caso de la exponencial se procede como sigue:

1.—Se buscan los logaritmos de cada una de las *yes* de la serie dinámica y se anotan en una columna cuyos valores se suman para obtener $S(\log. y)$.

2.—Se suman las *equis* tal y como aparecen en la serie para obtener $S(x)$.

3.—Se multiplica cada una de las *equis* por el correspondiente logaritmo de *ye*; se anotan los productos en una columna y se suman para obtener $S(x \log. y)$.

4.—Se calculan los cuadrados de cada una de las *equis* y se suman para obtener $S(x^2)$.

5.—Se substituyen los valores anteriores en las ecuaciones de interpolación, se resuelve el sistema de dos ecuaciones con dos incógnitas para los valores de log. a y log. b.

6.—Se buscan los antilogaritmos de log. a y log. b obtenidos mediante la resolución del sistema de ecuaciones, para obtener a y b.

7.—Se substituyen los valores de a y de b en la ecuación:

$$y = ab^x$$

Procedimiento del año mediano.—El procedimiento del año mediano permite simplificar la búsqueda de los parámetros a y b, ya que, al través suyo se anula la suma de las *equis* y desaparecen en las fórmulas de interpolación los términos que la contienen como coeficiente, las cuales se reducen a:

$$\therefore \quad \log. a = \frac{S(\log. y)}{N}$$

$$\therefore \quad \log. b = \frac{S(x \log. y)}{S(x^2)}$$

Parámetros en términos de N.—Cuando se emplea el procedimiento general, como en el caso de la recta, pueden substituirse:

$$S(x) \text{ por } \frac{N(N+1)}{2}$$

$$S(x^2) \text{ por } \frac{N(N+1)}{2} \frac{(2N+1)}{3}$$

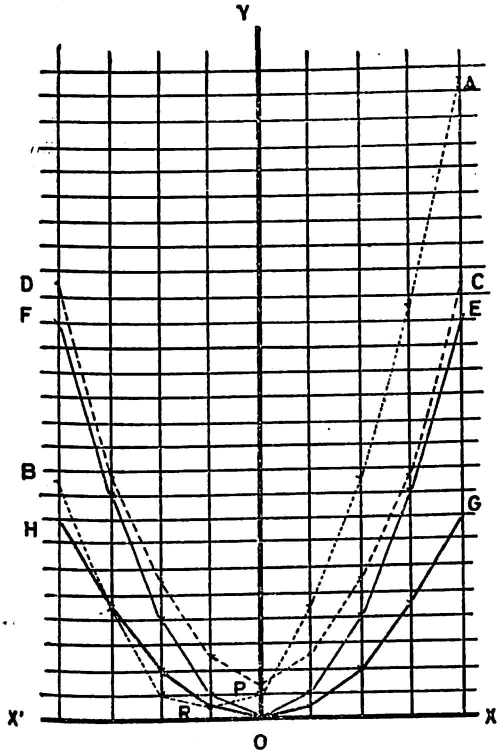
Ejecutando las operaciones correspondientes en forma análoga a como se hizo en el caso de la recta, pueden obtenerse los valores de a y de b en términos de N.

MÉTODO DE LOS MÍNIMOS CUADRADOS PARA EL CÁLCULO DE LA TENDENCIA

El método de los mínimos cuadrados busca hacer mínima la suma de los cuadrados de las desviaciones de cada valor de la serie dinámica teórica con respecto al valor correspondiente de la serie dinámica empírica. La aplicación del método permite obtener las necesarias ecuaciones de interpolación mediante un proceso de derivación parcial de la ecuación general con respecto a cada uno de sus parámetros y de ulterior igualación de los resultados a cero.

El procedimiento general de interpolación puede resumirse en las siguientes etapas:

1º—Representación Gráfica de la Serie Dinámica Empírica con el objeto de poder determinar el tipo de curva por interpolar.



PARABOLAS DE SEGUNDO GRADO

GOH $y = x^2$

EOF $y = 2x^2$

COD $y = 3 + 2x^2$

ARB $y = 3 + 4x + 2x^2$

PARABOLA DE TERCER GRADO

SOT $y = x^3$

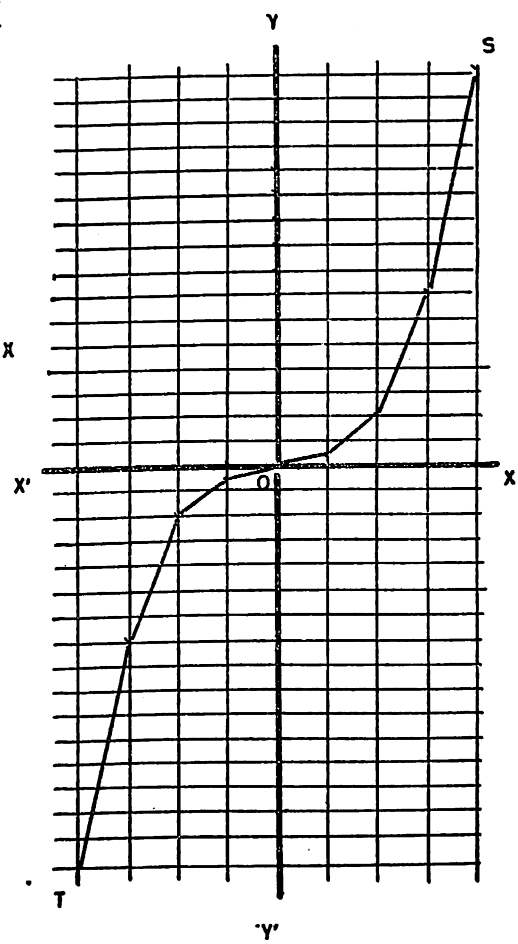


Figura 5

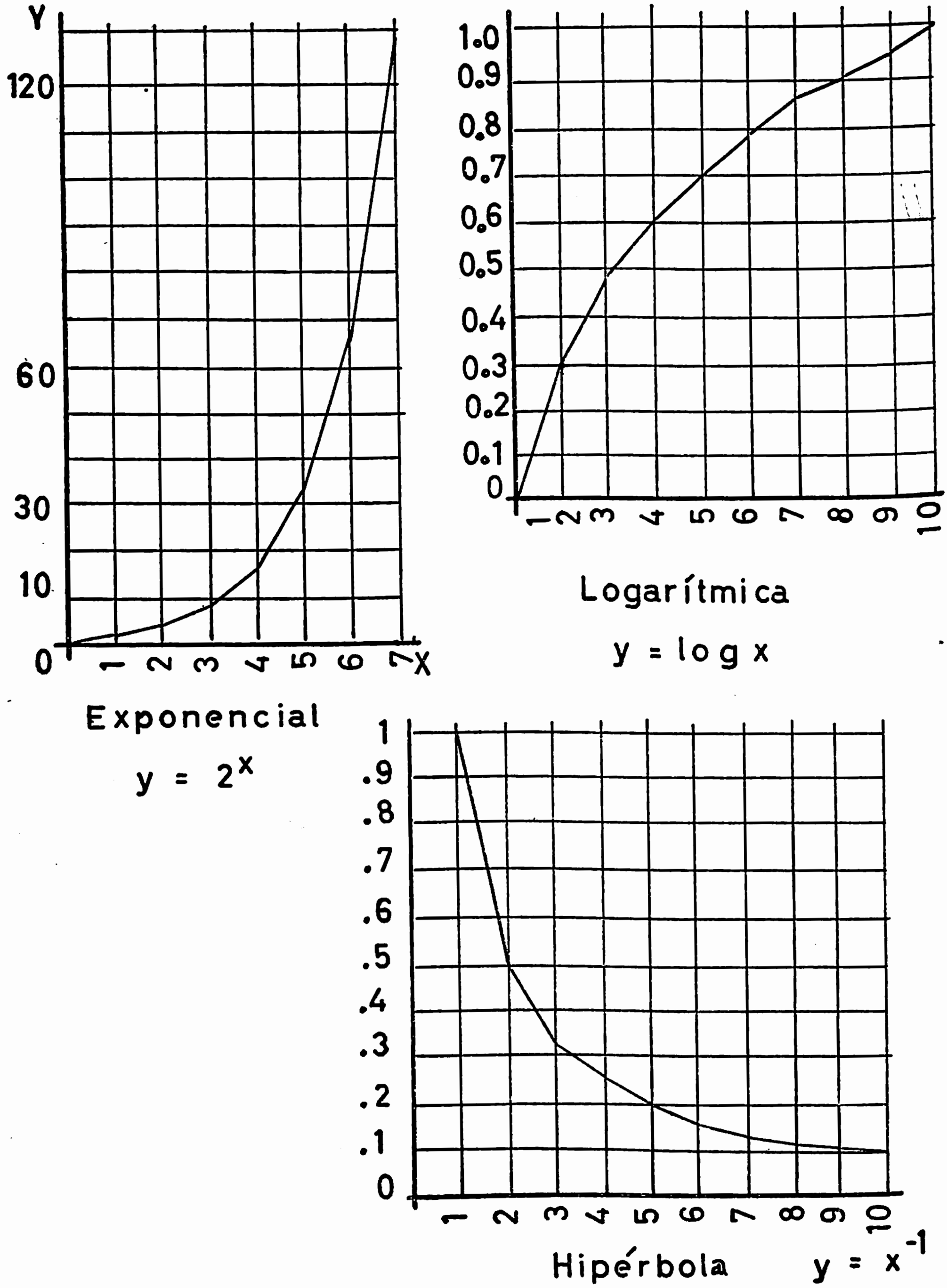

Y
120
60
30
10
0
1 2 3 4 5 6 7 X
Exponencial
y = 2^x
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10
Logarítmica
y = log x
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
1 2 3 4 5 6 7 8 9 10
Hipérbola
y = x^-1

Figura 6

2º—Elección —por inspección ocular— de una de las curvas siguientes como representativa de la serie dinámica empírica.

A.—Recta.
B.—Parábola de 2º Grado.
C.—Parábola de 3er. Grado.
D.—Hipérbola.
E.—Logarítmica.

3º—Consignación de la Fórmula General de la curva elegida:

A.—Recta:

$$y = a + bx$$

B.—Parábola de 2º Grado:

$$y = a + cx + bx^2$$

C.—Parábola de 3er. Grado:

$$y = a + cx + dx^2 + bx^3$$

D.—Hipérbola (Equilátera):

$$y = a + bx^{-1}$$

E.—Logarítmica:

$$y = a + b \log \mathrm{x}$$

4º—Obtención de tantas ecuaciones de interpolación como parámetros tenga la fórmula general de la curva elegida.

A.—Recta:

2 parámetros (a, b)
2 ecuaciones

B.—Parábola de 2º Grado:

3 parámetros (a, c, b)
3 ecuaciones

C.—Parábola de 3er. Grado:

4 parámetros (a, c, d, b)
4 ecuaciones

E.—Hipérbola (Equilátera):

2 parámetros (a, b)
2 ecuaciones

F.—Logarítmica:

2 parámetros (a, b)
2 ecuaciones.

Forma de las ecuaciones de interpolación:

I.—Para obtener la primera de interpolación, agréguese un operador Σ a cada uno de los términos de la ecuación general:

A.—Recta:

$$\Sigma y = Na + \Sigma(x)b$$

B.—Parábola de 2º Grado:

$$\Sigma y = Na + \Sigma(x)c + \Sigma(x^2)b$$

C.—Parábola de 3er. Grado:

$$\Sigma y = Na + \Sigma(x)c + \Sigma(x^2)d + \Sigma(x^3)b$$

D.—Hipérbola:

$$\Sigma y = Na + \Sigma(x^{-1})b$$

E.—Logarítmica:

$$\Sigma y = Na + \Sigma(\log.\ x)b$$

II.—Para obtener la última de interpolación, multiplíquense todos los términos de la general por la función correspondiente de x (por x en la recta, x^2 en la parábola, x^3 en la de tercer grado, x^{-1} en la hipérbola, log. x en la logarítmica), y agréguese un operador Σ:

A.—Recta:

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)b$$

B.—Parábola de 2º Grado:

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^4)b$$

C.—Parábola de 3er. Grado:

$$\Sigma(yx^3) = \Sigma(x^3)a + \Sigma(x^6)b$$

D.—Hipérbola:

$$\Sigma(yx^{-1}) = \Sigma(x^{-1})a + (x^{-2})b$$

E.—Logarítmica:

$$\Sigma(y \log x) = \Sigma(\log x)a + \Sigma(\log^2 x)b$$

III.—Para obtener la faltante de interpolación para la parábola de 2º grado, y una de las dos faltantes para la parábola de 3er. grado, multiplíquense todos los términos de la general por x y agréguese un operador Σ:

B.—Parábola de 2º Grado:

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)b$$

C.—Parábola de 3er. Grado:

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)d + \\ + \Sigma(x^4)b$$

IV.—Para obtener la faltante de interpolación para la parábola de 3er. grado, multiplíquense todos los términos de la general por x^2 y agréguese un operador Σ:

C.—Parábola de 3er. Grado:

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^3)c + \Sigma(x^4)d + \\ + \Sigma(x^5)b$$

5º—Fórmese con las ecuaciones anteriores un sistema de tantas ecuaciones como parámetros:

A.—Recta:

$$\Sigma(y) = Na + \Sigma(x)b$$

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)b$$

B.—Parábola de 2º Grado:

$$\Sigma(y) = Na + \Sigma(x)c + \Sigma(x^2)b$$

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)b$$

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^3)c + \Sigma(x^4)b$$

C.—Parábola de 3er. Grado:

$$\Sigma(y) = Na + \Sigma(x)c + \Sigma(x^2)d + \Sigma(x^3)b$$

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)d + \Sigma(x^4)b$$

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^3)c + \Sigma(x^4)d + \Sigma(x^5)b$$

$$\Sigma(yx^3) = \Sigma(x^3)a + \Sigma(x^4)c + \Sigma(x^5)d + \Sigma(x^6)b$$

D.—Hipérbola:

$$\Sigma(y) = Na + \Sigma(x^{-1})b$$

$$\Sigma(yx^{-1}) = \Sigma(x^{-1})a + \Sigma(x^{-2})b$$

E.—Logarítmica:

$$\Sigma(y) = Na + (\log. x)b$$

$$\Sigma(y \log. x) = \Sigma(\log. x)a + \Sigma(\log.^2 x)b$$

6º—Resuélvase el sistema de ecuaciones así formado (por ejemplo, por medio de determinantes)

7º—Substitúyanse los valores de los parámetros así obtenidos en las ecuaciones generales de las curvas correspondientes.

MÉTODO DE LOS MÍNIMOS CUADRADOS (SIMPLIFICACIÓN DEL AÑO MEDIANO)

A fin de simplificar el procedimiento de interpolación de una curva por el método de los mínimos cuadrados, se expresan las unidades de tiempo transcurridas como desviaciones con respecto a la mediana del período, en términos del intervalo. Como es fácil comprender, para que la simplificación sea practicable es necesario:

1º—Que los datos estén completos.

2º—Que exista un intervalo constante.

Al expresarse las unidades de tiempo como desviaciones con respecto a la media en unidades del intervalo, las sumas de potencias impares de x se anulan, simplificándose por lo mismo las ecuaciones de interpolación correspondientes, reduciéndose los sistemas formados por ellas a una serie de ecuaciones independientes, o a un conjunto de ecuaciones independientes y de sistemas más sencillos.

En efecto, al utilizarse las desviaciones con respecto al año mediano en unidades del intervalo, Σx, Σx^3, Σx^5, desaparecen de las correspondientes ecuaciones.

Así, en el caso de la recta, el sistema original de dos ecuaciones con dos incógnitas:

$$\Sigma y = Na + \Sigma(x)b$$

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)b$$

se reduce a las dos ecuaciones independientes:

$$\Sigma y = Na$$

$$\Sigma(yx) = \Sigma(x^2)b$$

Efectivamente, de la primera se puede obtener el valor de a, despejándolo, y de la segunda, en forma independiente, el valor de b:

$$a = \frac{\Sigma y}{N}$$

En forma análoga:

$$b = \frac{\Sigma(yx)}{\Sigma(x^2)}$$

El procedimiento en si, en general, consiste en lo siguiente:

1º—Determínese cuál es el año (mes, día, etc.) mediano del período estudiado, contando tantos años del primero al último como del último al primero a fin de determinar el central donde se cruzan las cuentas, o a fin de determinar los dos centrales (número par de datos) y tomar la media aritmética de los mismos.

2º—Determínense las desviaciones de cada año con respecto al año mediano, restando del año de que se trate el mediano y dividiendo entre el intervalo, o, mecánicamente:

A.—Si se trata de número impar de datos,
póngase 0 frente al año mediano,

$-1, -2, -3 \ldots -n$ frente a los anteriores,

$1, 2, 3 \ldots n$ frente a los posteriores.

B.—Si se trata de número par de datos,

póngase —0.5 y 0.5 frente a los años centrales,

—1.5, —2.5, —3.5 frente a los anteriores,

1.5, 2.5 3.5 frente a los posteriores.

3º—Consígnese la Ecuación General de la Curva por interpolar.

4º—Obténgase de las ecuaciones generales de interpolación anotadas en el procedimiento anterior, las correspondientes a este procedimiento simplificado haciendo desaparecer de ellas los términos que contengan a $\Sigma(x)$, $\Sigma(x^3)$, $\Sigma(x^5)$, como factor; en esta forma, se obtiene:

A.—Recta:

$$\Sigma(y) = Na$$
$$\Sigma(yx) = \Sigma(x^2)b$$

B.—Parábola de 2º Grado:

$$\Sigma(y) = Na + \Sigma(x^2)b$$
$$\Sigma(yx) = \Sigma(x^2)c$$
$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^4)b$$

C.—Parábola de 3er. Grado:

$$\Sigma(y) = Na + \Sigma(x^2)d$$
$$\Sigma(yx) = \Sigma(x^2)c + \Sigma(x^4)b$$
$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^4)d$$
$$\Sigma(yx^3) = \Sigma(x^4)c + \Sigma(x^6)b$$

D.—Hipérbola:

La simplificación no le es aplicable porque entre las desviaciones existe una nula cuyo recíproco es infinito.

E.—Logarítmica:

La simplificación no le es aplicable, ya que algunas de las desviaciones (la mitad de ellas para ser más precisos) son negativas, y no hay logaritmos de números negativos.

4º—Tómense separadamente las ecuaciones independientes, y fórmense uno o más sistemas de ecuaciones con las que no lo sean.

A.—Recta:

Dos ecuaciones independientes:

$$\Sigma(y) = Na; \qquad \Sigma(yx) = \Sigma(x^2)b$$

B.—Parábola de 2º Grado:

Una ecuación independiente:

$$\Sigma(yx) = \Sigma(x^2)c$$

y un sistema de dos ecuaciones con dos incógnitas:

$$\Sigma(y) = Na + \Sigma(x^2)b$$

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^4)b$$

C.—Parábola de 3er. Grado:

Dos sistemas, cada uno de dos ecuaciones con dos incógnitas:

$$\Sigma(y) = Na + \Sigma(x^2)d \qquad \Sigma(yx) = \Sigma(x^2)c + \Sigma(x^4)d$$

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^4)d \qquad \Sigma(yx^3) = \Sigma(x^4)c + \Sigma(x^6)d$$

5º—Resuélvanse por separado las ecuaciones independientes y los sistemas de ecuaciones por cualquier procedimiento conocido (por el de los determinantes, por ejemplo), a fin de obtener los valores de los parámetros.

6º—Sustitúyanse los valores de los parámetros en la ecuación general de la curva elegida a fin de obtener la ley matemática del fenómeno.

PROCEDIMIENTO DE LOS SUBTOTALES

AUXILIAR EN LA APLICACIÓN DEL MÉTODO DE LOS MÍNIMOS CUADRADOS

Fundamentalmente, el procedimiento de los subtotales tiende a simplificar los cálculos, obteniendo la suma de los productos de las potencias de x por y al través de las sumas acumulativas de las *yes* mismas.

Tomemos, como punto de partida, la serie:

x	y
1	y_1
2	y_2
3	y_3
4	y_4
5	y_5

Un primer subtotal puede obtenerse sumando todas las *yes* tal y como aparecen en la serie, de tal modo que:

$$S_1 = y_1 + y_2 + y_3 + y_4 + y_5$$
$$S_1 = \Sigma y_i$$

Obtendremos un segundo subtotal (S_2) si calculamos las sumas acumuladas de las *yes* y las sumamos al pie; o sea, si anotamos frente a y_1 de la segunda columna, y_1 en la tercera columna, si frente a y_2 de la segunda colocamos la suma de $y_1 + y_2$, frente a y_3 la de $y_1 + y_2 + y_3$ y en seguida sumamos:

x	y_i	
1	y_1	y_1
2	y_2	$y_1 + y_2$
3	y_3	$y_1 + y_2 + y_3$
4	y_4	$y_1 + y_2 + y_3 + y_4$
5	y_5	$y_1 + y_2 + y_3 + y_4 + y_5$
		$S_2 = 5y_1 + 4y_2 + 3y_3 + 2y_4 + y_5$

Por lo anterior, puede verse que en el segundo subtotal, la y_1 aparece 5 veces, o sea, tantas veces como observaciones se han hecho (N), la y_2 aparece 4 veces, o sea, tantas veces como observaciones se han hecho menos una ($N - 1$), la y_3, 3 veces o sea tantas como observaciones menos 2 ($N - 2$), etc. O sea, que, en general, el segundo subtotal está formado por tantas veces una de las y como número de observaciones se hayan hecho (N) menos el número de orden de la $y(r)$ menos uno; o sea que la enésima y se tomará N veces menos el número de orden de la misma (N) menos uno,, o bien, $N - (N - 1)$, lo cual es igual a 1; es decir, que la enésima y se tomará una sola vez. Un término general de orden r estará formado por la y_r tomada $N - (r - 1)$ veces:

$$S_2 = Ny_1 + (N - 1)y_2 + (N - 2)y_3 + (N - 3)y_4 + (N - 4)y_5 + \\ + \ldots N - (r - 1)y_r$$

O bien:

$$S_2 = \Sigma[N - (r - 1)y_r]$$

Pero, en cuanto se toman como x los números naturales, r número de orden de la y está dado precisamente por el valor de la x que le corresponde, por lo cual puede escribirse:

$$S_2 = \Sigma[N - (x - 1)y_i]$$

Si se ejecutan las operaciones de dentro del paréntesis, se tiene:

$$\begin{aligned} S_2 &= \Sigma[(N - x_i + 1)y_i] = \\ &= \Sigma[(N + 1) - x_i)y_i] = \\ &= \Sigma[(N + 1)y_i - x_iy_i] = \\ &= \Sigma(N + 1)y_i - \Sigma x_iy_i = \\ &= (N + 1)\Sigma y_i - \Sigma x_iy_i \end{aligned}$$

Si despejamos la suma de los productos de las *yes* por las *equis* correspondientes, tendremos:

$$\Sigma y_ix_i = (N + 1)\ \Sigma y_i - S_2$$

Pero como la suma de las *yes* es el primer subtotal:

$$\Sigma y_ix_i = (N + 1)\ S_1 - S_2$$

Conforme a esta fórmula, en cuanto se trata de buscar la suma de los productos de las *yes* por la primera potencia de las *equis* correspondientes para usarla en el método de los mínimos cuadrados:

1º—Se suman las *yes* de la serie dinámica empírica para obtener un primer subtotal, igual a la Σy_i requerida por las ecuaciones de interpolación del procedimiento general de los mínimos cuadrados.

2º—Se suman acumulativamente los valores de las *yes*, y estos valores acumulados se suman a su vez para obtener un segundo subtotal S_2.

3º—Se multiplica el primer subtotal por el número que sigue al de observaciones, en la serie natural.

4º—Del valor obtenido en 3 se resta el valor obtenido en 2 para obtener el valor de la suma de los productos de y por x que habrá de substituirse en las ecuaciones correspondientes del método general de los mínimos cuadrados.

Para poder obtener la suma de los productos de y por la segunda potencia de x, además de los anteriores, se necesitará del cálculo de un tercer subtotal que, en forma análoga a los anteriores, se obtendrá sumando acumulativamente los valores acumulativos usados en el segundo subtotal, y sumando al pie los valores obtenidos:

2º Subtotal	*3er. Subtotal*
y_1	y_1
$y_1 + y_2$	$2y_1 + y_2$
$y_1 + y_2 + y_3$	$3y_1 + 2y_2 + y_3$
$y_1 + y_2 + y_3 + y_4$	$4y_1 + 3y_2 + 2y_3 + y_4$
$y_1 + y_2 + y_3 + y_4 + y_5$	$5y_1 + 4y_2 + 3y_3 + 2y_4 + y_5$
	$15y_1 + 10y_2 + 6y_3 + 3y_4 + y_5 = S_3$

Cada uno de los términos de esta expresión es igual a la suma de los N primeros números (caso de y_1), de los $N - 1$ primeros números (caso de y_2) de los $N - 2$ primeros números (caso de y_3) ... de los $N - (N - 1)$ en el caso de y_N; como la suma de los N primeros números es igual al número de términos por el número más uno dividido entre dos, podemos escribir:

$$S_3 = \frac{5 \times 6}{2} y_1 + \frac{4 \times 5}{2} y_2 + \frac{3 \times 4}{2} y_3 + \frac{2 \times 3}{2} y_4 + \frac{1 \times 2}{2} y_5$$

en el caso concreto, y, en general:

$$S_3 = \frac{N(N+1)}{2} y_1 + \frac{(N-1)N}{2} y_2 + \frac{(N-2)(N-1)}{2} y_3 +$$

$$+ \ldots \frac{(N-(N-1)\ (N-(N-2)}{2} y_N$$

O bien:

$$S_3 = \Sigma \left[\frac{(N-(r-1)(N-(r-2)}{2} \right] y_r =$$

$$S_3 = \Sigma \left[\frac{(N-r+1)(N-r+2)}{2} \right] y_r$$

Pero, como cuando se toman como *equis* los números en su orden natural coincide el valor de la x con el número de orden, se tendrá:

$$S_3 = \Sigma \left[\frac{(N+1-x_i)(N+2-x_i)}{2} \right] y_r$$

Si comparamos el valor del tercer subtotal con el del segundo

$$S_2 = \Sigma[N+1-x_i)]y_i$$

Podremos percatarnos de la existencia de cierta regularidad que es fácil hacer extensiva a subtotales de orden superior; en efecto, es fácil probar las siguientes igualdades:

$$S_2 = \Sigma \left[\frac{(N+1-x_i)}{1} \right] y_i$$

$$S_3 = \Sigma \left[\frac{(N+1-x_i)}{1} \frac{(N+2-x_i)}{2} \right] y_i$$

$$S_4 = \Sigma \left[\frac{(N+1-x_i)}{1} \frac{(N+2-x_i)}{2} \frac{(N+3-x_i)}{3} \right] y_i$$

En general:

$$S_R = \Sigma \left[\frac{(N+1-x_i)}{1} \cdot \frac{(N+\overline{R-1}-x_i)}{R-1} \right] y_i$$

Conforme a lo anterior, pueden establecerse las siguientes relaciones:

$$S_2 = (N+1)\Sigma\, y - \Sigma xy$$

$$2S_3 = (N+1)(N+2)\Sigma\, y - [(N+1)(N+2)]\Sigma yx + \Sigma yx^2$$

$$\begin{aligned} 6S_4 = {} & (N+1)(N+2)(N+3)\Sigma\, y \\ & - [(N+1)(N+2)+(N+1)(N+3)+(N+2)(N+3)\Sigma yx] \\ & + [(N+1)+(N+2)+(N+3)]\Sigma yx^2 - \Sigma yx^3 \end{aligned}$$

Pero, puesto que Σy es el primer subtotal S_1

$$S_2 = (N+1)S_1 - \Sigma xy$$

$$\Sigma xy = (N+1)S_1 - S_2$$

Si se substituye este valor en el del tercer subtotal y se hacen las reducciones necesarias, despejando a Σyx^2 se tiene:

$$\Sigma yx^2 = (N+1)^2 S_1 - (2N+3)S_2 + 2S_3$$

En forma semejante, se puede obtener una expresión que de el valor de Σyx^3 si se hacen substituciones y reducciones en el valor del cuarto subtotal.

MÉTODO DE AGRUPACIÓN PARA EL CÁLCULO DE LA TENDENCIA

Supongamos una serie cronológica cuya tendencia sea rectilínea y de la que contemos con un número par de datos:

x_i	y_i
1	$y_1 = a + b$
2	$y_2 = a + 2b$
3	$y_3 = a + 3b$
4	$y_4 = a + 4b$
5	$y_5 = a + 5b$
6	$y_6 = a + 6b$

Si dividimos los valores de esta serie en dos grupos que contengan el mismo número de datos, y sumamos éstos dentro de cada grupo, obtendremos dos sumas que designaremos S_1 y S_2 respectivamente. Si en seguida restamos de la segunda suma la primera, obtendremos una diferencia D_1 que nos permitirá calcular el valor de b; en efecto.

$$\begin{array}{cl}
x_i & y_i \\
\hline
1 & y_1 = a + b \\
2 & y_2 = a + 2b \\
3 & y_3 = a + 3b \\
4 & y_4 = a + 4b \\
5 & y_5 = a + 5b \\
6 & y_6 = a + 6b
\end{array}
\quad
\begin{array}{l}
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} S_1 = y_1 + y_2 + y_3 = 3a + 6b \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} S_2 = y_4 + y_5 + y_6 = 3a + 15b
\end{array}
\quad
\begin{array}{r}
S_2 - S_1 = 3a + 15b \\
-3a - 6b \\
\hline
9b
\end{array}$$

$$D_1 = S_2 - S_1 = 9b$$

$$\therefore \quad b = \frac{D_1}{9}$$

Para obtener el valor de a, por ser b cantidad conocida en las expresiones para las sumas S_1 y S_2, bastará con substituir este valor y despejar el valor de a en la más sencilla (la expresión de S_1) para tener los valores de los dos parámetros de la ecuación que expresa la tendencia rectilínea; en efecto, si:

$$S_1 = 3a + 6b$$

$$S_1 - 6b = 3a$$

$$a = \frac{S_1 - 6b}{3}$$

Tanto en el caso de b, como en el de a, el denominador 9 del primero, y el denominador 3 del segundo así como el coeficiente 6 de b en el segundo dependen del número de datos con el que se trabajó. Si se desea obtener fórmulas más generales, válidas no sólo para 6, sino para 4, 2 o bien 8, 10, datos, tendremos que hacer las siguientes consideraciones:

Denominador de la expresión que permite obtener el valor de b.—El denominador de la expresión para b, (9) procede de la resta de los coeficientes (15 y 6) de b en cada una de las sumas, y dichos coeficientes son la suma de los valores de x que van del último al siguiente al que se encuentra en medio (de N a $\frac{N}{2} + 1$) y la suma de los valores de x que van del primero al que se encuentra en medio, o por mejor decir, al que es igual a la mitad del último (de 1 a $\frac{N}{2}$) respectivamente. Como se trata de sumas de términos que forman una progresión aritmética y dichas sumas se obtienen multiplicando la mitad del número de términos por la suma del primero y último término; estas sumas son, en el caso de los últimos valores:

$$\frac{\frac{N}{2}}{2}\left(N + \frac{N}{2} + 1\right) = \frac{N}{4}\left(N + \frac{N}{2} + 1\right)$$

Ya que dicha suma corresponde a la mitad del número de datos $N/2$ (a 3 datos en el ejemplo), valor que necesita dividirse entre dos para aplicar la fórmula de la suma de los términos de una progresión geométrica; $N/2$ entre 2 es igual a $N/4$. En el paréntesis figura el valor N del coeficiente de b para la última y, que es igual al número de datos (coeficiente de b en la y_6 de nuestro ejemplo de 6 datos, que es 6), en tanto que los otros dos términos del paréntesis $\frac{N}{2} + 1$ representan en conjunto al coeficiente de b correspondiente al dato cuya x es igual al número que sigue (de ahí el $+ 1$) a la mitad (de ahí el denominador 2) del número de términos (de ahí la N del numerador).

Con respecto a la segunda suma de los coeficientes de b para las expresiones de y se tiene:

$$\frac{N}{4}\left(1+\frac{N}{2}\right)$$

Ya que, como en el caso anterior, la suma comprende a la mitad del número de términos $N/2$; mitad que necesita dividirse entre 2 (de ahí $N/4$) para aplicar la fórmula para la suma de los términos de una progresión aritmética. De los términos en el paréntesis, 1 representa al coeficiente de b en la primera y, y $N/2$ al coeficiente de b en el dato cuya x es igual a la mitad del número de términos.

Si de la primera suma de coeficiente de b restamos esta segunda obtendremos el coeficiente de b en la expresión de D_1 o sea, el valor entre el que habrá que dividir D_1 para obtener b:

$$\frac{N}{4}\left(N+\frac{N}{2}+1\right)-\frac{N}{4}\left(1+\frac{N}{2}\right)=\frac{N}{4}\left(N+\frac{N}{2}+1-1-\frac{N}{2}\right)$$

Expresión en la que se ha sacado como factor común a $N/4$ y en la que pueden anularse $+1$ y -1, $+N/2$ y $-N/2$, con lo cual toda la expresión anterior se reduce a:

$$\frac{N}{4}\cdot N=\frac{N^2}{4}$$

O sea, que, en general, cuando se utiliza esta formación de dos grupos para el cálculo de b:

$$D_1=\frac{N^2}{4}b$$

o bien:

$$b=\frac{4D_1}{N^2}$$

Determinación del denominador de la expresión que permite obtener el valor de a.—El denominador de la expresión para $a(3)$ procede de la suma de los coeficientes de a (que siempre son 1) en la primera suma, la cual contiene la mitad del número de datos de la serie total; o sean $N/2$ datos; como el coeficiente de las *aes* es siempre 1, la suma de los $N/2$ coeficientes de a será $N/2$. O sea, que el denominador para la expresión de a será igual a la mitad del número de datos:

$$a = \frac{S_1 - 6b}{\frac{N}{2}} = \frac{2(S_1 - 6b)}{N}$$

La fórmula anterior sigue siendo válida solamente para el caso concreto estudiado, puesto que 6 (coeficiente de b) sigue dependiendo del número de datos considerado.

Determinación del coeficiente de b *en la expresión para obtener el valor de* a.—El coeficiente de b en esa expresión (6) procede de la suma de los $N/2$ primeros números naturales, o sea de la suma de $N/2$ términos de una progresión aritmética que comienza con 1, y cuya suma es igual, según hemos visto, a:

$$\frac{N}{4}\left(1 + \frac{N}{2}\right)$$

Según esto, la expresión general para a resulta ser:

$$a = \frac{2\left[S_1 - \frac{N}{4}\left(\frac{N}{2} + 1\right)b\right]}{N}$$

PROCEDIMIENTO DE CÁLCULO DE LA TENDENCIA RECTILÍNEA POR AGRUPAMIENTO

1º—Divídanse las *yes* observadas, por mitad, en dos grupos y obténgase la suma de las cantidades contenidas en cada uno de los grupos.

2º—Réstese de la suma de los valores del segundo grupo la suma de los valores del primero.

3º—Multiplíquese la diferencia por 4 y divídase el producto entre el cuadrado del número total de datos. El resultado es el valor de b.

4º—Multiplíquese el valor de b por el número que siga a la mitad del número de datos, y por la cuarta parte del número de datos.

5º—Réstese el valor obtenido en 4º de la suma de los valores del primer grupo.

6º—Multiplíquese el residuo por 2 y divídase entre el número de datos. El resultado de esta última operación es el valor de a.

7º—Substitúyanse los valores obtenidos en la expresión general de la recta.

PROCEDIMIENTO DE CÁLCULO DE LA TENDENCIA RECTILÍNEA POR AGRUPAMIENTO DE LAS *EQUIS* Y DE LAS *YES*

1º—Divídanse los datos, por mitad, en dos grupos y obténganse las sumas de los datos contenidos en cada grupo.

a.—por una parte, para las *yes* o magnitudes variables del fenómeno en el tiempo (S_{1y}, S_{2y})

b.—por otra parte, para las *equis* o sea para las unidades de tiempo transcurridas (S_{1x}, S_{2x}).

2º—Réstense de las segundas sumas (S_{2y}, S_{2x}) las primeras sumas (S_{1y}, S_{1x}) correspondientes, para obtener las respectivas diferencias (D_{1y}, D_{1x}).

3º—Divídase la diferencia correspondiente a las *yes* (D_{1y}) entre la diferencia que corresponda a las *equis* (D_{1x}). El resultado es el valor del parámetro *b*.

4º—Multiplíquese el valor así obtenido por la primera suma de las *equis* (bS_{1x}).

5º—Réstese el valor del producto así obtenido, de la primera suma de las *yes* ($S_{1y} - bS_{1x}$).

6º—Multiplíquese la diferencia así obtenida por 2 y divídase el producto entre el número de datos. El resultado es el valor de *a*.

7º—Obtenidos los parámetros *a* y *b*, substitúyanse en la ecuación general de la recta.

El procedimiento anterior se resume en las dos fórmulas siguientes:

$$b = \frac{D_{1y}}{D_{1x}}$$

$$a = \frac{2(S_{1y} - bS_{1x})}{N}$$

PROCEDIMIENTO DE CÁLCULO DE LA TENDENCIA LOGARÍTMICA POR AGRUPAMIENTO DE LAS *EQUIS* Y DE LAS *YES*

Previo.—Obténganse los logaritmos de las *equis*, o unidades de tiempo transcurridas.

Central.—Trabajando con los logaritmos de las *equis* como si fueran las *equis*, calcúlese la tendencia rectilínea por agrupamiento de las *equis* (de los logaritmos de *equis*) y de las *yes*, conforme al procedimiento anterior. Substitúyanse los valores de los parámetros a y b así obtenidos en la ecuación general de la logarítmica: $y = a + b \log x$.

PROCEDIMIENTO DE CÁLCULO DE LA TENDENCIA GEOMÉTRICA (O EXPONENCIAL) POR AGRUPAMIENTO DE LAS *EQUIS* Y LAS *YES*

Previo.—Obténganse los logaritmos de las *yes*, o sea de las magnitudes variables del fenómeno en el tiempo.

Central.—Trabajando con los logaritmos de las *yes* como si fueran las *yes*, calcúlese la tendencia rectilínea por agrupamientos de las *equis* y de las *yes* (de los logaritmos de las *yes*) conforme al procedimiento descrito con anterioridad.

Posterior.—Búsquense los antilogaritmos de los valores encontrados para los parámetros y substitúyanse dichos valores en la expresión general para la exponencial $y = ab^x$, o bien substitúyanse directamente los valores de los parámetros encontrados ($\log a$ y $\log b$) en la expresión logarítmica de la exponencial:

$$y = \log a + x \log b$$

PROCEDIMIENTO DE CÁLCULO DE LA TENDENCIA GEOMÉTRICA MODIFICADA POR AGRUPAMIENTO DE LAS *EQUIS* Y DE LAS *YES*

Se conoce como tendencia geométrica modificada aquella en la cual, para mostrar el carácter geométrico o exponencial del crecimiento, es necesario agregar a la expresión ordinaria de la tendencia geométrica o exponencial ($y = ab^x$, o $y = bc^x$ según nos conviene escribir aquí) un término (positivo o negativo) que se conoce como "factor o elemento de corrección" (que representaríamos por c en caso de elegir la forma ordinaria de escritura de la exponencial, pero que preferiremos representar por a eligiendo la segunda forma de escribir la exponencial, por las relaciones que nos permite establecer en cuanto a su procedimiento de cálculo, según se verá en seguida). De esta forma, la expresión para la tendencia geométrica modificada será:

$$y = a + bc^x$$

El cálculo de este tipo de tendencia constará, fundamentalmente, de 2 etapas: el cálculo de c, y el cálculo de a y de b.

Cálculo del parámetro c.—Para calcular el parámetro c, tómese un número de datos que sea múltiplo de 3; elegidos dichos datos, dénse a las *equis* valores que vayan de 0 (para la primera *equis*) en adelante; una vez hecho esto:

1º—Agrúpense los datos en tres grupos iguales en cuanto al número de valores que comprenda cada uno.
2º—Súmense los valores de cada grupo obteniéndose las tres sumas correspondientes: s_1, s_2, s_3.
3º—Réstense:
s_2 de s_1 para obtener una primera diferencia d_1.
s_3 de s_2 para obtener una segunda diferencia d_2.
4º—Divídase la segunda diferencia entre la primera.
5º—Extráigase del cociente una raíz que tenga por índice el total de datos divididos entre 3 (raíz cuadrada si son 6 datos, cúbica si son 9, cuarta si son 12, etc.). El resultado es el valor de c.

Cálculo de los parámetros a *y* b.—Para el cálculo de los parámetros recurriremos al procedimiento descrito con anterioridad, en la forma siguiente:

Previo.—Calcúlense las diferentes potencias de c indicadas por los valores de x, elevando el valor recién encontrado a las potencias 0, 1, 2, 3 . . . N.

Central.—Tomando los valores de c elevado a x como si fueran las *equis*, procédase a calcular una tendencia rectilínea mediante agrupamiento de las *equis* (de las *ces* elevadas a *equis*) y de las *yes*. Los valores de los parámetros obtenidos deberán substituirse junto con el valor del parámetro c en la ecuación de la tendencia geométrica modificada $y = a + bc^x$.

INTERPOLACIÓN DE UNA CURVA DE GOMPERZ POR AGRUPACIÓN

La ecuación general para la curva de Gomperz es:

$$y = ab^{c^x}$$

la cual, mediante los logaritmos se convierte en:

$$\log y = \log a + c^x \log b$$

En vista de que log y ocupa el lugar de las *yes* en la geométrica modificada y que cosa análoga ocurre con los logaritmos de a y de b que substituyen a los parámetros a y b de la geométrica modificada, el procedimiento de cálculo puede esquematizarse como sigue:

Previo.—Obténganse los logaritmos de las *yes.*

Central.—Tomando los logaritmos de las *yes* como si fueran las *yes,* interpólese una geométrica modificada.

Ulterior.—Substitúyanse los valores de los parámetros encontrados (c, log a y log b) en la forma logarítmica de la ecuación de Gomperz, o bien búsquense los antilogaritmos de los parámetros encontrados en segundo término (log a y log b) y junto con el parámetro calculado en primer lugar (c), substitúyanse en la forma ordinaria de la ecuación de Gomperz ($y = ab^{c^x}$).

INTERPOLACIÓN DE UNA CURVA LOGÍSTICA POR AGRUPACIÓN

La curva logística o de Pearl - Reed tiene por ecuación:

$$y^{-1} = a + bc^x$$

en la cual los recíprocos de las *yes* (y^{-1}) ocupan el lugar que corresponde a las *yes* en la geométrica modificada. Conforme a esto:

Previo.—Obténganse los recíprocos de las *yes.*

Central.—Tomando los recíprocos de las *yes* como si fueren *yes* interpólese una geométrica modificada.

Ulterior.—Substitúyanse los valores de los parámetros a, b, c en la ecuación general de la logística.

INTERPOLACIÓN DE UNA CURVA DE GOMPERZ POR AGRUPACIÓN

La ecuación general para la curva de Gomperz es:

$$y = ab^{c^x}$$

la cual, transformada por medio de logaritmos es:

$$\log y = \log a + c^x \log b$$

Conforme a lo anterior, el procedimiento de cálculo consistirá, como en el caso anterior, de dos etapas principales, el cálculo del parámetro c, y el cálculo de los parámetros a y b.

Cálculo del parámetro c.—Para obtener el valor de c en la curva de Gomperz:

Previo.—Obténganse los logaritmos de las *yes*.

Central.—Tomando los logaritmos de las *yes* como si fueran las *yes*, calcúlese una geométrica modificada.

OBTENCIÓN DE LAS FÓRMULAS PARA LOS PARÁMETROS DE UNA TENDENCIA GEOMÉTRICA MODIFICADA

En la interpolación de una tendencia geométrica modificada se hace uso de la ecuación:

$$y = a + bc^x$$

en la que a, b y c son constantes cuyo valor necesita determinarse a fin de obtener la ley general del fenómeno y, como ya es usual en nuestras series temporales, y representa las magnitudes variables del fenómeno en el tiempo y x las unidades de tiempo transcurridas (con la salvedad de que, en este caso, para aplicar las fórmulas que obtendremos a continuación, la serie de las *equis* debe iniciarse en el origen, o sea, que la primera *equis* debe hacerse igual a cero).

Supongamos la serie:

x	y
0	$y_0 = a + bc$
1	$y_1 = a + bc^1$
2	$y_2 = a + bc^2$
3	$y_3 = a + bc^3$
4	$y_4 = a + bc^4$
5	$y_5 = a + bc^5$

Si principiamos por formar tres grupos con los datos y sumar los valores contenidos en cada grupo (las dos *yes*, en este caso, que cada grupo contiene, obtendremos las tres sumas siguientes:

$$S_1 = y_0 + y_1 = a + bc^0 + a + bc^1 = 2a + b(c^0 + c^1)$$

$$S_2 = y_2 + y_3 = a + bc^2 + a + bc^3 = 2a + b(c^2 + c^3)$$

$$S_3 = y_4 + y_5 = a + bc^4 + a + bc^5 = 2a + b(c^4 + c^5)$$

Si en los segundos términos de los últimos miembros de estas igualdades se sacan como factores comunes c^0, c^2 y c^4, las expresiones para las tres sumas se convierten en:

$$S_1 = 2a + bc^0(1 + c)$$

$$S_2 = 2a + bc^2(1 + c)$$

$$S_3 = 2a + bc^4(1 + c)$$

El valor de cada una de las sumas, dado por el segundo miembro de cada una de estas igualdades, sigue conteniendo tres incógnitas (los parámetros a, b y c). Si de cada una de estas sumas restamos la suma precedente (de S_2 restamos S_1 y de S_3 restamos S_2) lograremos eliminar una de las incógnitas (a); en efecto, al hacer tales restas y obtener las diferencias D_1 y D_2 tenemos:

$$D_1 = S_2 - S_1 = 2a + bc^2(1 + c)$$
$$- [2a + bc^0(1 + c)] = b(1 + c)(c^2 - c^0)$$

$$D_2 = S_3 - S_2 = 2a + bc^4(1 + c)$$
$$- [2a + bc^2(1 + c)] = b(1 + c)(c^4 - c^2)$$

En las expresiones finales para las dos diferencias figuran sólo dos de las tres incógnitas originales. A fin de eliminar una incógnita más (b), dividiremos la segunda diferencia entre la primera, con lo cual b en el numerador y b en el denominador se reducen a la unidad. Además de esto y no como algo buscado sino subsidiario, el factor común $(1 + c)$ se reducirá asimismo:

$$\frac{D_2}{D_1} = \frac{b(1 + c)(c^4 - c_2)}{b(1 + c)(c^2 - c^0)} = \frac{c^4 - c^2}{c^2 - c^0}$$

Si se divide el numerador $c^4 - c^2$ entre el denominador $c^2 - c^0$ se obtendrá como cociente c^2; o sea que:

$$\frac{D_2}{D_1} = c^2$$

En esta igualdad, para obtener el valor de c bastará con extraer la raíz cuadrada:

$$c = \sqrt{\frac{D_2}{D_1}}$$

Conocido el valor de c, bastará con substituirlo en una expresión en la que figure con b como una de las dos incógnitas de la misma, para que se pueda calcular el valor de b. Si se toma el valor de la primera diferencia D_1 (en la que figuran las potencias menores de c) se tiene:

$$D_1 = b(1+c)(c^2 - c^0) = b(1+c)(c^2 - 1)$$

Expresión en la que podemos despejar a b:

$$b = \frac{D_1}{(1+c)\ (c^2 - 1)}$$

Si, con objeto de dar una forma más simétrica y manejable a esta expresión multiplicamos numerador y denominador por $c - 1$, obtendremos:

$$b = \frac{D_1(c-1)}{(1+c)(c-1)(c^2-1)} = \frac{D_1(c-1)}{(c^2-1)(c^2-1)} = D_1 \frac{c-1}{(c^2-1)^2}$$

Para obtener el valor de a, bastará con substituir los valores de b y de c en una expresión que contenga las tres incógnitas. Si se toma el valor de la primera suma S_1 (en la que figuran los valores numéricos más sencillos) se tiene:

$$S_1 = 2a + bc^0(1+c) = 2a + b(1+c)$$

(el último miembro ha sido escrito en vista de que una potencia elevada a la potencia cero es igual a la unidad, por lo cual c^0 puede desaparecer de la expresión).

Si en la expresión anterior despejamos a a pasando el término $b(1+c)$ al otro miembro con signo contrario, y el 2 que figura como coeficiente de a al otro miembro como divisor, se tendrá:

$$\frac{S_1 - b(1+c)}{2} = a$$

Si se quiere obtener el valor de a exclusivamente en términos de S_1 y de c, puede substituirse el valor de b, con lo cual se obtiene:

$$= \frac{S_1 - D_1 \dfrac{c-1}{(c^2-1)^2}(1+c)}{2} = = \frac{S_1 - D_1 \dfrac{c^2-1}{(c^2-1)^2}}{2} = \frac{S_1 - \dfrac{D_1}{c^2-1}}{2}$$

No debe olvidarse que, en todo lo anterior, hemos obtenido las fórmulas de c, b y a para un ejemplo bastante general, pero no todo lo general que sería de desear, ya que el ejemplo se refiere a un caso en el que hay seis datos o sean tres grupos de dos datos y las fórmulas calculadas sobre esta base se aplican sólo a casos de seis datos. Sin embargo, sería fácil mostrar o demostrar que las fórmulas anteriores dejan de ser particulares para convertirse en generales si: *1º*—en lugar del índice 2 del radical que permite calcular el valor de c, *2º*—en lugar del exponente 2 de c en la expresión para el cálculo de b, y *3º*—en lugar del denominador 2 y el exponente 2 de c en la expresión para a, se coloca $N/3$ o sea el valor que se obtiene de dividir el número de datos entre 3 (grupos). Con esto, las expresiones se convierten en:

$$c = \sqrt[\frac{N}{3}]{\frac{D_2}{D_1}} = \sqrt[n]{Q}$$

$$b = D_1 \frac{c - 1}{(c^{\frac{N}{3}} - 1)^2} = D_1 \frac{c - 1}{(c^n - 1)^2}$$

$$a = \frac{S_1 - \frac{D_1}{c^{\frac{N}{3}} - 1}}{\frac{N}{3}} = \frac{S_1 - \frac{D_1}{c^n - 1}}{n}$$

INTERPOLACIÓN DE LAS LÍNEAS DE TENDENCIA POR EL PROCEDIMIENTO DE LOS MOMENTOS FACTORIALES

Si en una serie dinámica cuyo crecimiento sea rectilíneo representamos la serie de unidades de tiempo transcurridas por la serie de los números naturales ($x = 1, 2, 3, 4, 5, \ldots$) y designamos a la magnitud del fenómeno correspondiente a la primera unidad de tiempo transcurrida (y_1) con la letra A, podremos afirmar que y_2 es igual a A (o sea, a la y anterior) más una vez el ritmo de crecimiento constante propio de la recta, que representaremos por b; o sea:

$$y_2 = A + 1b$$

En forma análoga, la y_3 será igual a la y_2 más el ritmo de crecimiento b, o sea, será igual a la y_1 ($= A$) más 2 veces el ritmo de crecimiento:

$$y_3 = A + 2b$$

De la misma manera, se obtendrían las igualdades:

$$y_4 = A + 3b$$

$$y_5 = A + 4b$$

....

O bien, que podríamos escribir:

x	y
1	A
2	$A + 1b$
3	$A + 2b$
4	$A + 3b$
5	$A + 4b$

Si sumamos acumulativamente la columna de las *yes*, obtendremos la siguiente columna:

x	y	
1	A	A
2	$A + 1b$	$2A + 1b$
3	$A + 2b$	$3A + 3b$
4	$A + 3b$	$4A + 6b$
5	$A + 4b$	$5A + 10b$

A cada una de las expresiones contenidas en esta última columna se les conoce como "primer momento factorial" (para la recta) de una, dos, tres, cuatro y cinco observaciones respectivamente; así $5A + 10b$ es el primer momento factorial (para la recta) de 5 datos u observaciones.

El examen de los momentos factoriales anteriores nos permite observar que:

1º—Todos ellos son expresiones binomiales, o sea, expresiones de 2 términos.

2º—En todos ellos figura:

un término en A, y
un término en b (precedido en el primer caso por un coeficiente cero y anulado por ello mismo).

3º—El coeficiente de A es igual al valor de la última x

(igual a 1 en el primer momento factorial de un dato,

igual a 2 en el primer momento factorial de dos datos,
igual a N en el primer momento factorial de N datos).

4º—El coeficiente de b es igual a la suma de todas las x anteriores a la última.
(igual a la suma de cero números naturales para un dato,
igual a la suma del primer número natural para dos datos,
igual a la suma de los dos primeros números naturales para tres datos,
igual a la suma de los $N - 1$ primeros números naturales para N datos).

Como la suma de los primeros números naturales es igual al producto del último de ellos (en este caso $N - 1$ por el que le seguiría en la serie $(N - 1 + 1 = N)$ dividido entre 2, podremos representar el coeficiente de este segundo término, en b, por:

$$N \frac{N-1}{2}$$

Si se tiene en cuenta todo lo anterior, podremos representar el primer momento factorial de N observaciones por:

$$\overset{N}{F_1} = NA + N \frac{N-1}{2} b$$

Esta expresión nos permitirá encontrar el valor de los parámetros A y b en cada caso particular siempre y cuando se asocie esta expresión a otra, ya que habiendo dos parámetros A y b, son necesarias dos ecuaciones; de ahí la necesidad de un segundo momento factorial de N observaciones.

Si en el ejemplo que sirvió de punto de partida sumamos acumulativamente los valores de donde obtuvimos el primer momento factorial, tendremos:

A	A
$2A + 1b$	$3A + 1b$
$3A + 3b$	$6A + 4b$
$4A + 6b$	$10A + 10b$
$5A + 10b$	$15A + 20b$

A cada una de estas expresiones se les conoce como "segundo momento factorial" (para la recta) de *dos*, tres, cuatro cinco, y *seis* observaciones, respectivamente; así, A es el segundo momento factorial (para la recta) de dos observaciones, $3A + 1b$ el segundo momento factorial de tres observaciones, $6A + 4b$ el segundo momento factorial de cuatro observaciones, $10A + 10b$, el segundo momento factorial de cinco observaciones y $15A + 20b$, el segundo momento factorial de seis observaciones; o sea, que en la suma acumulativa basta con detenerse en el penúltimo renglón para obtener el segundo momento factorial correspondiente a la última observación.

Si se comparan los coeficientes de A en el segundo momento factorial con los coeficientes de b en el primer momento factorial correspondiente (o sea del mismo número de observaciones), podremos darnos cuenta de que dichos coeficientes son iguales o sea, que se trata en uno y otro caso de las sumas de los $N - 1$ primeros números naturales. En cuanto al coeficiente de b en el segundo momento factorial, el mismo puede obtenerse multiplicando el coeficiente de A por $(N - 2)/3$. O sea que, con respecto a los segundos momentos factoriales puede observarse que, en el caso de la recta:

1º—Todos son expresiones binomiales.

2º—En las que figuran:
un término en A, y
un término en b.

3º—En ellos, el coeficiente de A (suma de los $N - 1$ primeros números naturales) es igual a:

$$N\frac{N-1}{2}$$

4º—El coeficiente de b se obtiene de multiplicar el coeficiente de A por $(N - 2)/3$; o sea, que es igual a:

$$N\frac{N-1}{2}\,\frac{N-2}{3}$$

Según esto, el segundo momento factorial de N observaciones para el caso de la recta puede expresarse como:

$$\overset{N}{F_2} = N\frac{N-1}{2}A + N\frac{N-1}{2}\,\frac{N-2}{3}b$$

En cuanto se tienen estas dos fórmulas (suficientes) para el cálculo de los parámetros que figuran en la ecuación de la recta, es posible delinear el siguiente:

Procedimiento de interpolación de una recta mediante el cálculo de los momentos factoriales

1º—Súmense acumulativamente las *yes* de la serie dinámica empírica; la última suma acumulativa es el primer momento factorial de la serie $(= \overset{N}{F_1})$.

2º—Substitúyanse en la fórmula del primer momento factorial para la recta los valores de:

a.—el primer momento factorial de la serie dinámica empírica encontrado en *1.*

b.—el número de observaciones que la serie comprende (N).

3º—Ejecútense las operaciones necesarias en la ecuación del primer momento factorial.

4º—Súmense acumulativamente las sumas acumuladas o primeros momentos factoriales encontrados en 1, deteniéndose en esta suma acumulativa en el penúltimo renglón; la suma correspondiente a este penúltimo renglón es el segundo momento factorial de la serie dinámica empírica $(\overset{N}{F_2})$.

5º—Substitúyanse en la fórmula del segundo momento factorial:

a.—el segundo momento factorial de la serie dinámica empírica encontrado en *4.*

b.—el número de observaciones.

6º—Ejecútense las operaciones necesarias en la fórmula substituída del segundo momento factorial.

7º—Resuélvase el sistema de dos ecuaciones con dos incógnitas (A, b) formado por las expresiones obtenidas en *3* y *6.*

8º—Substitúyanse los valores de los parámetros A y b en la ecuación general de la recta:

$$y = A + b(x - 1)$$

Esta expresión utilizable en el caso de los momentos factoriales, puede reducirse a la forma más común si se ejecutan operaciones en la anterior:

$$y = A + bx - b$$

Como b es una constante, se puede asociar con la A que es como ella una constante independiente de cualquier potencia de la variable:

$$y = (A - b) + bx$$

Si comparamos esta expresión con la más general de la recta, veremos que el lugar que en ella ocupa a está ocupado en esta otra por $A - b$, o sea, que $a = A - b$, lo cual coincide con el supuesto del que partimos, ya que aceptamos que A representaba la primera magnitud del fenómeno y_1, o sea aquella que puede hacerse igual a la ordenada en el origen (a) más una vez el ritmo de crecimiento constante $b(A = a + 1b)$.

O sea, que la ley general del fenómeno puede volver a expresarse en la forma ordinaria si

8º—Se resta el valor de b del valor de A $(A - b)$ y tanto el valor resultante de esa resta (a) como el valor de b se substituyen en la ecuación general de la recta:

$$y = a + bx$$

EXTENSIÓN DEL PROCEDIMIENTO DE LOS MOMENTOS FACTORIALES A LA INTERPOLACIÓN DE OTRAS LÍNEAS POLINOMIALES

El procedimiento de los momentos factoriales es aplicable en la interpolación de líneas de tendencia de la forma polinomial:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 \ldots + a_nx^n$$

Para ello:

I.—Se necesitan tantos momentos factoriales como parámetros (a_0, a_1, a_2, a_3, $\ldots a_n$) figuren como coeficientes de las potencias de la variable (x, x^2, x^3, $\ldots x^n$) en la ecuación (dos en el caso de la recta en el que sólo existen los parámetros a_0 y a_1, tres en el de la parábola de segundo grado para la que existen los parámetros a_0, a_1 y $a_2 \ldots$ etc.).

II.—Cada momento factorial es un polinomio formado por tantos términos como parámetros haya.

III.—El coeficiente de un término de orden r en el primer momento factorial estará dado por r fracciones cuyos numeradores constituyan una serie decreciente de números naturales a partir de N, número de observaciones, y cuyos denominadores estén formados por la serie creciente de los números naturales a partir de 1 (y terminando en r); o sea, que el conjunto de los denominadores forma "factorial r".

Según esto, el coeficiente del primer término será $\frac{N}{1}$

el del segundo $\frac{N}{1}\,\frac{N-1}{2}$

el del tercero $\frac{N}{1}\,\frac{N-1}{2}\,\frac{N-2}{3}$

etc.

IV.—El coeficiente de orden r del segundo momento factorial será igual al coeficiente de orden $r+1$ en el primer momento factorial (o sea, que el coeficiente del parámetro a_0 en el segundo momento factorial es igual al del parámetro a_1 en el primer momento, el del parámetro a_1 del segundo igual al del a_2 del primero, etc.).

V.—El coeficiente de orden r del tercer momento factorial será igual al coeficiente de orden $r+2$ en el primer momento factorial, o al coeficiente de orden $r+1$ del segundo.

VI.—En general, el coeficiente de orden r del momento factorial de orden s será igual al coeficiente de orden $r+(s-1)$ del primer momento factorial o al momento de orden $r+1$ del momento de orden $s-1$.

VII.—Formados los momentos factoriales necesarios, será necesario resolver el sistema de ecuaciones formado por ellos (mediante determinantes, por ejemplo), y

VIII.—Substituir los valores de los parámetros así encontrados en la expresión polinomial correspondiente (del *tipo* de la anotada al principio de este parágrafo).

PROCEDIMIENTO DE LOS POLINOMIOS ORTOGONALES PARA LA DETERMINACIÓN DE LÍNEAS DE TENDENCIA

Supongamos dos polinomios p_1 y p_2 en la variable u. Si se obtienen los valores de p_1 y p_2 para todos los valores de u, se multiplica cada valor de p_1 por el correspondiente de p_2 y al obtener la suma de dichos productos, dicha suma resulta nula, se dice que los polinomios p_1 y p_2 son *ortogonales*.

En general, para que un conjunto de polinomios (y no ya sólo dos) sea ortogonal, es preciso que las sumas de los productos de los polinomios tomados de dos en dos, sean nulas, o bien que:

$$\Sigma p_1 p_2 = 0 \qquad \Sigma p_1 p_3 = 0 \ldots \qquad \Sigma p_1 p_n = 0$$

$$\Sigma p_2 p_3 = 0 \qquad \Sigma p_2 p_4 = 0 \ldots \qquad \Sigma p_2 p_n = 0$$

$$\Sigma p_n p_1 = \Sigma p_1 p_n = 0$$

Una línea de tendencia puede expresarse como la suma de una serie de términos, cada uno de los cuales está formado por el producto de un factor constante por un polinomio que con los restantes forma un conjunto de polinomios ortogonales (de orden 0, de 1º, 2º, 3º, ... enésimo orden):

$$v = L_0 p_0 + L_1 p_1 + L_2 p_2 + L_3 p_3 + \ldots L_n p_n$$

Mediante el criterio de los mínimos cuadrados, se obtienen (para el caso más sencillo de $L_0 p_0 + L_1 p_1 + L_2 p_2$), las ecuaciones del tipo siguiente:

$$\Sigma v p_0 = L_0 \Sigma p_0^2 + L_1 \Sigma p_1 p_0 + L_2 \Sigma p_2 p_0$$

$$\Sigma v p_1 = L_0 \Sigma p_0 p_1 + L_1 \Sigma p_1^2 + L_2 \Sigma p_2 p_1$$

$$\Sigma v p_2 = L_0 \Sigma p_0 p_2 + L_1 \Sigma p_1 p_2 + L_2 \Sigma p_2^2$$

Como 1., p_1, p_2 se requiere que formen un conjunto de polinomios ortogonales, esto significa que:

$$\Sigma p_0 p_1 = \Sigma p_1 p_0 = 0$$

$$\Sigma p_0 p_2 = \Sigma p_2 p_0 = 0$$

$$\Sigma p_1 p_2 = \Sigma p_2 p_1 = 0$$

O sea, que en las ecuaciones de interpolación se anulan todos los términos en los que aparecen esas sumas de productos de polinomios, con lo cual las ecuaciones se reducen a:

$$\Sigma v p_0 = L_0 \Sigma p_0^2$$

$$\Sigma v p_1 = L_1 \Sigma p_1^2$$

$$\Sigma v p_2 = L_2 \Sigma p_2^2$$

De estas expresiones se pueden despejar las constantes L.

$$L_0 = \frac{\Sigma v p_0}{\Sigma p_0^2}$$

$$L_1 = \frac{\Sigma v p_1}{\Sigma p_1^2}$$

$$r = \frac{\Sigma v p_2}{\Sigma p_2^2}$$

Si, como hemos convenido, p_0 representa un polinomio de grado cero, p_1 un polinomio de grado uno, p_2 un polinomio de segundo grado, etc., y representamos por medio de *kas* afectadas de subíndices los coeficientes constantes de estos polinomios en la variable u, tendremos:

$$p_0 = u^0 = 1$$

$$p_1 = k_{10}\, u^0 + u^1 = k_{10} + u$$

$$p_2 = k_{20}\, u^0 + k_{21}\, u^1 + u^2 = k_{20} + k_{21}\, u + u^2$$

(Convencionalmente, se ha destinado el primer guarismo del subíndice doble que afecta a la k para indicar el grado del polinomio al que corresponde, en tanto que el segundo guarismo del propio subíndice se ha destinado a indicar la potencia a la que el coeficiente afecta.)

La condición de ortogonalidad de estos polinomios impone el que las sumas de los productos de los mismos tomados de dos en dos sean nulas; es decir:

$$\Sigma p_0 p_1 = \Sigma(1 \times p_1) = \Sigma p_1 = 0$$

$$\Sigma p_0 p_2 = \Sigma(1 \times p_2) = \Sigma p_2 = 0$$

$$\Sigma p_1 p_2 = 0$$

Si en estas expresiones se substituyen los valores dados anteriormente para p_1 y para p_2, tendremos:

$$\Sigma p_1 = \Sigma(K_{10} + u) = \Sigma k_{10} + \Sigma u = 0$$

$$\Sigma p_1 = \Sigma(k_{10} + u) = \Sigma k_{10} + \Sigma u = 0$$

$$\Sigma p_1 p_2 = \Sigma(k_{10} + u)\, p_2 = k_{10} \Sigma p_2 + \Sigma p_2 u = 0$$

La forma de los polinomios que cumplen estas condiciones de ortogonalidad puede simplificarse si u se expresa como desviación de las unidades de tiempo con respecto a su media en unidades del intervalo. En estas condiciones, las sumas de las potencias impares de u se anulan, anulando también a los términos en los que aparecen como factores, reduciéndose las expresiones anteriores a:

$$\Sigma p_1 = \Sigma k_{10} = N k_{10} = 0$$

$$\Sigma p_2 = \Sigma k_{20} + \Sigma u^2 = N k_{20} + \Sigma u^2 = 0$$

$$\Sigma p_1 p_2 = \Sigma p_2 u = 0$$

En la última expresión, se eliminó el término $k_{10} \Sigma p_2$ porque, conforme a las condiciones de ortogonalidad, Σp_2 debe ser nulo; no se eliminó el segundo término porque $\Sigma p_2 u$ representa una suma de los productos de cada una de las p_2 por cada una de las *ues* y si bien tanto la suma de las p_2 individualmente es nula como lo es la suma de las *ues* tomadas individualmente, la suma de los productos (distinta del producto de las sumas) *no* tiene por qué ser necesariamente nula.

De las expresiones anteriores, podemos despejar a los coeficientes constantes k_{10}, k_{20} y obtener:

$$k_{10} = \frac{0}{N} = 0$$

$$k_{20} = \frac{-\Sigma u^2}{N}$$

Para el valor de k_{21}, substituiremos en la última ecuación el valor de p_2 y despejaremos:

$$\Sigma p_2 u = \Sigma(k_{20} + k_{21}u + u^2)u = k_{20}\Sigma u + k_{21}\Sigma u^2 + \Sigma u_3 = k_{21}\Sigma u^2 = 0$$

Despejando a k_{21}, se obtiene $k_{21} = 0$

Si, en estas condiciones substituimos los valores de las constantes en los polinomios correspondientes, obtendremos:

$$p_1 = k_{10} + u = u$$

$$p_2 = k_{20} + k_{21}u + u^2 = -\frac{\Sigma u^2}{N} + u^2 = u^2 - \frac{\Sigma u^2}{N}$$

Pero Σu^2 es igual a dos veces la suma de los cuadrados de los $(N+1)/2$ números naturales, cuyo valor substituído en el de p_2 da:

$$p_2 = u^2 - \frac{N^2 - 1}{12}$$

En forma semejante, pueden obtenerse las fórmulas para los polinomios de tercero, cuarto, quinto . . . grado:

$$p_3 = u^3 - \frac{3N^2 - 7}{20} u$$

$$p_4 = u^4 - \frac{3N^2 - 13}{14} u^2 + \frac{3(N^2 - 1)(N^2 - 9)}{560}$$

$$p_5 = u^5 - \frac{5(N^2 - 7)}{18} u^3 + \frac{15N^4 - 230\,N^2 + 407}{1\,008} u$$

PROCEDIMIENTO DE LOS POLINOMIOS ORTOGONALES PARA LA DETERMINACIÓN DE LÍNEAS DE TENDENCIA

Para determinar líneas de tendencia mediante el procedimiento de los polinomios ortogonales:

I.—Calcúlense los coeficientes L_0, L_1, $L_2 \ldots L_n$ que habrán de afectar a los polinomios de primero, segundo, tercero o enésimo grado en la expresión orto-polinomial de la tendencia,

1º—calculando los valores de λ_0, λ_1, λ_2, λ_3, ... λ_n, para lo cual:

a.—deberán obtenerse los subtotales primero, segundo, tercero, ... enésimo de la distribución, y

b.—dividirse dichos subtotales entre las combinaciones del número de datos tomados de uno en uno, el número de datos más 1 tomados de dos en dos, el número de datos más 2 tomados de tres en tres, el número de datos más 3 tomados de cuatro en cuatro, el número de datos más n tomados de $n + 1$ en $n + 1$.

2º—substituyendo los valores de λ_1, λ_2, λ_3, ... λ_n en las fórmulas dadas en el cuadro adicional para el cálculo de las λ primas correspondientes.

3º—substituyendo los valores de las λ prima en las fórmulas para el cálculo de las L, o bien:

a.—multiplicando cada λ' por el valor correspondiente contenido en la diagonal subrayada del cuadro, y

b.—dividiendo el producto entre el resultado de multiplicar tantos factores decrecientes a partir de $N - 1$ como indique el sub-índice de la L cuyo valor se calcula.

II.—Determinar las formas de los polinomios de primero, segundo, tercero, ... enésimo grado, substituyendo el valor de N en las fórmulas para p_1, p_2, p_3, ...

III.—Reducir la ley del fenómeno a la forma ordinaria:

1º—Multiplicando cada L por el polinomio correspondiente.

2º—Agrupando y reduciendo términos semejantes (los que contengan las mismas potencias de la variable).

3º—Ordenando los términos resultantes según el orden creciente de las potencias de la variable.

FÓRMULAS DEL PROCEDIMIENTO DE LOS POLINOMIOS ORTOGONALES PARA LA INTERPOLACIÓN DE TENDENCIAS

1º—Fórmulas de las λ:

General:

$$\lambda_r = \frac{S_r + 1}{\binom{N+r}{r+1}}$$

Particulares:

$$\lambda_0 = \frac{S_1}{\binom{N}{1}} \qquad \lambda_1 = \frac{S_2}{\binom{2}{N+1}}$$

$$\lambda_2 = \frac{S_3}{\binom{N+2}{3}} \qquad \lambda_3 = \frac{S_4}{\binom{N+3}{4}}$$

$$\lambda_4 = \frac{S_5}{\binom{5}{N+4}} \qquad \lambda_5 = \frac{S_6}{\binom{N+5}{6}}$$

$$\lambda_6 = \frac{S_7}{\binom{N+6}{7}} \qquad \lambda_7 = \frac{S_8}{\binom{N+7}{8}}$$

$$\lambda_8 = \frac{S_9}{\binom{N+8}{9}} \qquad \lambda_9 = \frac{S_{10}}{\binom{N+9}{10}}$$

2º—Fórmulas para las λ':

	λ_0	$+\lambda_1$	$+\lambda_2$	$+\lambda_3$	$+\lambda_4$	$+\lambda_5$	$+\lambda_6$	$+\lambda_7$	$+\lambda_8$	$+\lambda_9$
λ'_0	1									
λ'_1	1	—1								
λ'_2		—3	2							
λ'_3	1	—6	10	—5						
λ'_4	1	—10	30	—35	14					
λ'_5	1	—15	70	—140	126	—42				
λ'_6	1	—21	140	—420	630	—462	132			
λ'_7		—28	252	—1050	2310	—2772	1716	—429		
λ'_8	1	—36	420	—2310	6680	—12012	12012	—6435	1430	
λ'_9	1	—45	660	—4620	6930	—42042	60060	—51480	24310	—4862

3º—Fórmulas para las L:

$$L_0 = \lambda'_0$$

$$L_1 = \frac{6}{N-1}\lambda'_1$$

$$L_2 = \frac{30}{(N-1)(N-2)}\lambda'_2$$

$$L_3 = \frac{140}{(N-1)(N-2)(N-3)}\lambda'_3$$

$$L_4 = \frac{630}{(N-1)(N-2)(N-3)(N-4)}\lambda'_4$$

RITMO DE CRECIMIENTO DE UN FENÓMENO

Una vez que, por cualquiera de los procedimientos reseñados brevemente en páginas anteriores, se ha determinado la ley matemática que rige un fenómeno, conviene determinar, operando sobre esa misma expresión matemática, si:

1.—el ritmo de crecimiento es constante o variable, y
2.—de ser variable, el modo en que varía.

A fin de determinar el ritmo de crecimiento de un fenómeno, se procede a encontrar la primera derivada de la magnitud del mismo con respecto al tiempo, en la ley matemática que lo rige.

Consideremos, por ejemplo, el caso de la recta:

$$y = a + bx$$

Si calculamos la derivada de y con respecto a x, tenemos:

$$\frac{dy}{dx} = \frac{d(a + bx)}{dx} = \frac{da}{dx} + \frac{dbx}{dx} = 0 + \frac{bdx}{dx} = b$$

Como la primera derivada de la función con respecto al tiempo es la velocidad o ritmo de crecimiento del fenómeno, y en este caso esa primera derivada es igual a b, siendo como es ésta una constante, podemos afirmar que:

Un movimiento rectilíneo tiene un ritmo de crecimiento constante; o sea, que el fenómeno aumenta en una misma cantidad por cada unidad de tiempo transcurrida. Dicho ritmo está dado por el coeficiente de x (tiempo) en la ley general del fenómeno o por la pendiente de la recta que representa tal fenómeno.

De ahí también que, a intervalos de tiempo constantes correspondan diferencias constantes entre cada uno de los valores del movimiento rectilíneo y los valores previos correspondientes o bien que, en tal tipo de movimiento sean constantes las primeras diferencias.

En el caso de una parábola de segundo grado, la ley general del fenómeno puede representarse por la ecuación:

$$y = a + cx + bx^2$$

Si obtenemos la primera derivada de dicha expresión, tendremos:

$$\frac{dy}{dx} = \frac{d(a + cx + bx^2)}{dx}$$

$$\frac{dy}{dx} = \frac{da}{dx} + \frac{d(cx)}{dx} + \frac{dbx^2}{dx}$$

$$= 0 + c\frac{dx}{dx} + b\frac{dx^2}{dx}$$

$$= c + 2bx$$

El hecho de que en esta primera derivada de la función con respecto a la variable dependiente (tiempo), siga figurando esta última, indica que en la parábola de segundo grado, el crecimiento del fenómeno no es constante sino variable; o sea que el incremento en la magnitud del fenómeno al pasar de una unidad de tiempo a la siguiente (de un año a otro) no es la misma sean cuales fueren las unidades de tiempo consideradas, según ocurrió en el caso de la recta, sino que dicho incremento será distinto según que se trate del paso de la primera unidad de tiempo a la segunda, de la segunda a la tercera o de la penúltima a la última.

La expresión de la velocidad instantánea en un movimiento parabólico: $c + 2bx$, presenta la estructura propia de un movimiento rectilíneo ya que, en efecto, existe un término independiente de la variable x, y x aparece multiplicada por un coeficiente también constante $2b$.

En el caso de una parábola de 3er. grado, la ley general del fenómeno está dada por:

$$y = a + cx + dx^2 + bx^3$$

El cálculo de la primera derivada da como resultado:

$$\frac{dy}{dx} = \frac{d(a + cx + dx^2 + bx^3)}{dx}$$

$$\frac{dy}{dx} = \frac{da}{dx} + c\frac{dx}{dx} + d\frac{dx^2}{dx} + b\frac{dx^3}{dx}$$

$$\frac{dy}{dx} = c + 2dx + 3bx^2$$

La existencia de términos en los que figuran la primera y la segunda potencia de la variable independiente tiempo muestra que la velocidad o ritmo de crecimiento de un movimiento parabólico de tercer grado no es constante sino variable; el hecho de que la potencia máxima que figura en dicha expresión sea el cuadrado de la variable indica que la velocidad aumenta a su vez, conforme a un movimiento parabólico de segundo grado.

Si se calculan las segundas derivadas para los movimientos parabólico de 2º grado y parabólico de 3er. grado o sea, si se obtienen las primeras derivadas de sus velocidades, se obtiene:

Para la parábola de 2º grado:

Velocidad: $c + 2bx$

Primera derivada de la velocidad o Aceleración:

$$\frac{d(c + 2bx)}{dx} = 2b$$

Para la parábola de 3er. grado:

Velocidad: $c + 2dx + 3bx^2$

Primera derivada de la velocidad o Aceleración:

$$\frac{d(c + 2dx + 3bx^2)}{dx} = 2d + 6bx$$

O sea, que en el caso de la parábola de 2º grado, la aceleración es constante, y el movimiento, uniformemente acelerado. En la parábola de 3er. grado, la aceleración es variable, y el movimiento acelerado (sin ulterior adverbialización).

Ritmo o Velocidad Media de Crecimiento en los Movimientos Rectilíneo, Parabólico de 2º Grado y Parabólico de 3er. Grado.—En el caso del movimiento rectilíneo, la velocidad media de crecimiento sería:

$$\frac{\Sigma b}{N} = \frac{Nb}{N} = b$$

En el caso de una parábola de 2º grado:

$$\frac{\Sigma(c + 2bx)}{N} = \frac{\Sigma c + 2b\Sigma x}{N} = \frac{Nc}{N} + 2b\frac{\Sigma x}{N} = c + 2b\bar{x}_a$$

Si se ha trabajado con x expresadas como desviaciones con respecto al año mediano, conviene calcular la velocidad media de crecimiento mediante la media cuadrática:

$$\sqrt{\frac{\Sigma(c + 2bx)^2}{N}} = \sqrt{\frac{\Sigma(c^2 + 4cbx + 4b^2x^2)}{N}}$$

$$\sqrt{\frac{\Sigma c^2 + 4cb\Sigma x + 4b^2\Sigma x^2}{N}}$$

$$\sqrt{c^2 + 4cb\bar{x}_a + 4b^2\frac{\Sigma x^2}{N}}$$

Pero $\bar{x}_a$ es en este caso igual a cero, o sea, que desaparece todo el segundo término, y la expresión de la velocidad media en el movimiento parabólico se reduce a:

$$\sqrt{c^2 + 4b^2 \frac{\Sigma x^2}{N}}$$

En el caso de la parábola de tercer grado, se obtiene como expresión para la velocidad media de crecimiento:

$$\frac{\Sigma(c + 2dx + 3bx^2)}{N} = \frac{\Sigma c + 2d\Sigma x + 3b\Sigma x^2}{N} =$$

$$= c + 2d\bar{x}_a + 3b \frac{\Sigma x^2}{N}$$

Velocidad instantánea en el caso de un Movimiento Hiperbólico.—Sea la expresión:

$$y = a + bx^{-1}$$

Si tomamos su primera derivada:

$$\frac{dy}{dx} = \frac{d(a + bx^{-1})}{dx} = 0 + b\frac{dx^{-1}}{dx} = b(-1)x^{-2} =$$

$$= -bx^{-2} = -\frac{b}{x^2}$$

Según esta expresión en el caso de la hipérbola equilátera de que se trata, el ritmo de crecimiento es negativo, o sea, se trata de un ritmo de decrecimiento, y dicho ritmo es inversamente proporcional al cuadrado del tiempo transcurrido (que figura en el denominador); o bien, que en el caso de la hipérbola hay siempre decrecimiento (nos referimos siempre a la parte positiva de la hipérbola que es la única de interés estadístico), pero dicho decrecimiento que es rápido al principio (cuando x es pequeño), se va haciendo más y más lento (conforme x aumenta).

En general, para una hipérbola, sea o no equilátera, puede trabajarse con la expresión:

$$y = bx^{-n}$$

Si se calcula la primera derivada de dicha expresión, se obtiene:

$$\frac{dy}{dx} = b(-n)x^{-n-1} = -nbx^{-(n+1)} = \frac{-nb}{x^{n+1}}$$

O sea, que se trata de un ritmo de *decrecimiento,* inversamente proporcional al tiempo según la potencia $n + 1$.

En el caso de una curva logística del tipo:

$$\frac{1}{y} = \frac{bc^x}{1}$$

Si se toma la primera derivada, se obtiene:

$$\frac{dy}{dx} = \frac{d(b^{-1}c^{-x})}{dx} = b^{-1}\frac{dc^{-x}}{dx} = b^{-1}c^{-x}(-1)Lc = -b^{-1}c^{-x}Lc$$

En esta expresión de la velocidad de un movimiento logístico, puede verse que $-b^{-1}$ y Lc son constantes y que c^{-x} es variable en cuanto su exponente lo es; o sea, que la velocidad de este movimiento es variable y depende de las variaciones de c^{-x}; esta última expresión según es fácil reconocer (constante elevada a un exponente negativo variable) es una exponencial negativa que puede representarse asimismo por $(1/c^x)$. De este modo, conforme crezca x, crecerá exponencialmente (o sea con un ritmo de crecimiento acelerado) c^x y, en cuanto c^x (denominador) crezca, disminuirá el valor de la fracción $1/c^x$, disminuyendo con ella el ritmo de crecimiento del fenómeno logístico. Todo esto indica que un crecimiento logístico procede por *incrementos aceleradamente decrecientes* que tienden a anularse (cuando x es infinitamente grande, lo es mucho más c^x y $1/c^x$ tiende a hacerse infinitamente pequeño).

OBTENCIÓN DE LOS VALORES TEÓRICOS DE LA TENDENCIA, POR SUBSTITUCIÓN

Para obtener los valores que teóricamente debería alcanzar un fenómeno en el caso de que estuviese sujeto solamente a tendencia secular, se procede a substituir los valores de las *equis* empleadas en el proceso de interpolación, en la ley matemática del fenómeno. Para proceder ordenadamente y mecanizar el procedimiento, conviene servirse de una tabulación auxiliar, en la cual figuren serialmente los valores necesarios para la substitución; así, por ejemplo, en el caso de la recta, cuya expresión matemática general es: $y = a + bx$, se necesitarán dos columnas principales (una para cada término de la expresión) y de ellas una dividida en dos sub-columnas (una destinada al coeficiente y otra a la variable). En la tabulación conviene proceder del núcleo de la ecuación (punto en el que se encuentra la variable independiente elevada a su máximo exponente) hacia la periferia; o sea, que en el caso de la recta, la tabulación se hará como sigue:

PRIMERA COLUMNA.—Para el término que contiene a x.

Primera Subcolumna.—Valores de las x con que se trabajó la interpolación.

Segundo Subcolumna.—Producto de cada x por el valor encontrado para el parámetro b.

SEGUNDA COLUMNA.—Para el término independiente de x.

TERCERA COLUMNA.—Suma de la segunda * y la segunda subcolumna de la primera.—Resultado: *yes* teóricas.

Los encabezados serían los siguientes:

Término en x		a	$a + bx$
x	bx		

En el caso de la parábola de segundo grado, cuya expresión es:

$$y = a + cx + bx^2$$

la tabulación se hará como sigue:

PRIMERA COLUMNA.—Para el término en x^2.

Primera Subcolumna.—Para las *equis* de la interpolación.

Segunda Subcolumna.—Para los cuadrados de las *equis*.

Tercera Subcolumna.—Para los cuadrados de las *equis* multiplicados por el parámetro b.

SEGUNDA COLUMNA.—Para el término en x.

Primera Subcolumna.—Para las *equis* multiplicadas por el parámetro c.

TERCERA COLUMNA.—Para el término independiente de x.

* En todo lo que sigue, después de los ordinales "primera", "segunda", "tercera", "cuarta", se da por supuesto el término "columna".

CUARTA COLUMNA.—Para la suma de la tercera, la segunda y la tercer subcolumna de la primera. *R*: *yes* teóricas.

Los encabezados de la tabulación serían:

Término en x^2			*Término en* x	*Término Ind.*	*Y teórica*
x	x^2	bx^2	cx	a	$a + cx + bx^2$

En el caso de la parábola de tercer grado, cuya expresión es:

$$y = a + cx + dx^2 + bx^3$$

la tabulación se hará como sigue:

PRIMERA COLUMNA.—Para el término en x^3.

Primera Subcolumna.—Para las *equis* de la interpolación.

Segunda Subcolumna.—Para los cubos de las *equis.*

Tercera Subcolumna.—Para los cubos de las *equis* por el parámetro *b.*

SEGUNDA COLUMNA.—Para el término en x^2.

Primera Subcolumna.—Para los cuadrados de las equis.

Segunda Subcolumna.—Para los cuadrados de las *equis* por el parámetro *d.*

TERCERA COLUMNA.—Para el término en x.

Primera Subcolumna.—Para las *equis* multiplicadas por el parámetro *c.*

CUARTA COLUMNA.—Para el término independiente.

QUINTA COLUMNA.—Para la suma de la cuarta, tercera, segunda subcolumna de la segunda y tercera subcolumna de la primera. Resultado: *yes* teóricas.

Los encabezados de la tabla son:

Término en x^3			*Término en* x^2		*Término en* x	*T. Ind.*	*Yes teóricas*
x	x^3	bx^3	x^2	dx^2	cx	a	$a+cx+dx^2$ $+ bx^3$

En el caso de la hipérbola, cuya expresión matemática es:

$$y = a + bx^{-1}$$

La tabulación se hará como sigue:

PRIMERA COLUMNA.—Para el término recíproco de x.
Primera Subcolumna.—Equis de la interpolación.
Segunda Subcolumna.—Recíprocos de las *equis*.
Tercera Subcolumna.—Recíprocos de las *equis* por el parámetro b.

SEGUNDA COLUMNA.—Para el término independiente.

TERCERA COLUMNA.—Para la suma de la segunda columna y de la tercera subcolumna de la primera. Resultados: *yes* teóricas.

Los encabezados de la tabla son:

Término en x^{-1}			*Término independiente*	*Yes teóricas*
x	x^{-1}	bx^{-1}	a	$a + bx^{-1}$

Si la hipérbola que se ha interpolado no es equilátera, la expresión de la misma será:

$$y = bx^{-n}$$

de la que es equivalente la forma logarítmica:

$$\log.\ y = \log b - n \log x$$

Si se trabaja con la primera de estas expresiones, la tabulación deberá hacerse en la forma siguiente:

PRIMERA COLUMNA.—Para el término en x^{-n}.
Primera Subcolumna.—Para las *equis* de la interpolación.
Segunda Subcolumna.—Para las *equis* elevadas al exponente n.
Tercera Subcolumna.—Para los recíprocos de esas potencias multiplicados por el parámetro b. Resultados: *yes* teóricas.

Los encabezados de la tabulación serán entonces:

Término en x^{-n}			
x	x^n	$x^{-n} = 1/x^n$	bx^{-n}

Si se usa la forma logarítmica, la tabulación se hará como sigue:

PRIMERA COLUMNA.—Para el término en log x.
Primera Subcolumna.—Para las *equis* de interpolación.
Segunda Subcolumna.—Para los logaritmos de las *equis*.
Tercera Subcolumna.—Para los logaritmos de las *equis* por n.

SEGUNDA COLUMNA.—Para el logaritmo de b.
TERCERA COLUMNA.—Para la primera, menos la tercera subcolumna de la tercera columna. Resultado: logarítmos de las *yes* teóricas.

CUARTA COLUMNA.—Para los antilogaritmos de los valores de la tercera columna. Resultado: *yes* teóricas.

Los encabezados serían los siguientes:

Término en log x			log b	log b — n log x	*Antilog.*
x	n log x	log x	*T. Ind.*	*Log de la Y teórica*	*Y teórica*

En el caso de la logarítmica, cuya expresión es:

$$y = a + b \log x$$

la tabulación se hará como sigue:

PRIMERA COLUMNA.—Para el término en log x.
Primera Subcolumna.—Para las *equis* de interpolación.
Segunda Subcolumna.—Para los logaritmos de esas *equis*.
Tercera Subcolumna.—Para los logaritmos de las *equis* por el parámetro b.

SEGUNDA COLUMNA.—Para el término independiente.

TERCERA COLUMNA.—Para la suma de la segunda columna y de la tercera subcolumna de la primera columna. Resultado: *yes* teóricas.

Los encabezados son los siguientes:

Término en log x			*T. Ind.*	*Y teórica*
x	log x	b log x	a	$a + b$ log x

En el caso de una exponencial, la expresión matemática es:

$$y = ab^x$$

La tabulación de la exponencial se hará como sigue:

PRIMERA COLUMNA.—Para el factor exponencial.
Primera Subcolumna.—Para las *equis* de interpolación.
Segunda Subcolumna.—Para el parámetro b elevado a una potencia igual a cada una de las *equis*.

SEGUNDA COLUMNA.—Para el factor no exponencial, o coeficiente.

TERCERA COLUMNA.—Producto de la segunda, por la segunda subcolumna de la primera. Resultados: *yes* teóricas.

Encabezados de la tabulación:

Factor exponencial		*Coeficiente*	*Y teórica*
x	b^x	a	ab^x

Si se trabaja la forma logarítmica de la exponencial,

$$\log y = \log a + x \log b$$

la tabulación se hará en la siguiente forma:

PRIMERA COLUMNA.—Para el término en x.
Primera Subcolumna.—Para las *equis* de interpolación.
Segunda Subcolumna.—Para las *equis* de interpolación por el logaritmo del parámetro b.

SEGUNDA COLUMNA.—Para el término independiente.
Primera Subcolumna.—Para los logaritmos del parámetro a.

TERCERA COLUMNA.—Para la suma de la segunda, y la segunda subcolumna de la primera.

CUARTA COLUMNA.—Para los antilogaritmos de la tercera. Resultados: *yes* teóricas.

Término en x	*Término Independiente*		

OBTENCIÓN DE LOS VALORES DE LA TENDENCIA POR MEDIO DE DIFERENCIAS FINITAS

Si consideramos una serie cronológica cuya tendencia sea rectilínea, podremos reemplazar cada una de las y_i por una expresión de la forma $a + bx_i$; si, además, igualamos la serie de las x_i a la serie de los números naturales, tendremos, en lugar de la serie de las y_i, $(a + 1b)$, $(a + 2b)$, $(a + 3b) \ldots$ $(a + (n - 1)b)$, $(a + nb)$. Si de cada una de estas y_i restamos la anterior, obtendremos las "primeras diferencias" de esta serie rectilínea:

x_i	y_i	*Primeras diferencias*	Δ^1
1	$a + 1b$	$(a + 1b) - (a + 0b)$	b
2	$a + 2b$	$(a + 2b) - (a + 1b)$	b
3	$a + 3b$	$(a + 3b) - (a + 2b)$	b
$(n - 1)$	$a + (n - 1)b$		
....	....	....	..
n	$a + n\ b$	$(a + nb) - (a + (n-1)b)$	b

El último caso tipifica lo que representa buscar las primeras diferencias de una serie rectilínea, ya que n puede representar el valor de cualquiera de las *equis*; en efecto, al restar de $a + nb$, valor de la y_n, el valor de la ye anterior y_{n-1} igual a $a + (n - 1)b$ las aes se anulan, puesto que son iguales y de signo contrario, obteniéndose: $nb - (n - 1)b$, expresión en la que puede sacarse como factor común a b: $b(n - (n - 1$, lo cual es igual a $b \times 1$ o sea, a b, sea cual fuere el valor de n.

Todo lo anterior quiere decir que en este caso —de tendencia rectilínea— las primeras diferencias son constantes.

Y así como podemos afirmar que las primeras diferencias de un movimiento rectilíneo son constantes, la afirmación inversa también es cierta, o sea, que, cuando las primeras diferencias son constantes, el movimiento es rectilíneo.

En forma análoga, si consideramos una serie cronológica cuya tendencia esté representada por una parábola de segundo grado, la serie de las y_i quedará representada por expresiones de la forma $(a + 1b)$, $(a + 4b)$, $(a + 9b) \ldots [a + (n - 1)^2]b$, $(a + n^2)b$, en las que los coeficientes de b son los cuadrados de los números naturales ya que, tratándose de una parábola de segundo grado, una expresión general (incompleta puesto que le falta término en x) es: $y_i = a + bx_i^2$. Si de cada y_i se resta la anterior, se obtiene una serie de "primeras diferencias" cuyos valores no son constantes sino variables (3, $5b$). De estas diferencias podría mostrarse que el crecimiento es rectilíneo; en efecto, si se toman las primeras diferencias de estas primeras diferencias de la serie original (o sea, si se calculan las "segundas diferencias" de la serie original, se obtiene una serie de constantes ($2b$).

En efecto, tomemos los valores de dos *yes* contiguas:

$$y_n = a + n^2b$$
$$y_{n-1} = a + (n-1)^2b$$

La primera diferencia entre estas *yes* se obtendrá restando de y_n, y_{n-1}, lo cual produce

$$[n^2 - (n-1)^2]b = [n^2 - (n^2 - 2n + 1)]b = [n^2 - n^2 + 2n - 1]b = \\ = (2n-1)b$$

Si, en forma análoga, tomamos la diferencia entre y_{n-1} y y_{n-2}, obtendremos como primera diferencia:

$$[(n-1)^2 - (n-2)^2] = [(n^2 - 2n + 1) - (n^2 - 4n + 4)]b = \\ = (2n-3)b$$

Asimismo, la diferencia entre y_{n+1} y y_n será:

$$[(n+1)^2 - n^2]b = [(n^2 + 2n + 1) - n^2]b = (2n+1)b$$

Tomadas en su orden, estas tres diferencias formarían la serie $(2n-3)b$, primera diferencia entre y_{n-1} y y_{n-2}; $(2n-1)b$, primera diferencia entre y_n y y_{n-1}; $(2n+1)b$, primera diferencia entre y_{n+1} y y_n.

Si se resta de $(2n-1)b$, la diferencia anterior, $(2n-3)b$, se obtiene como segunda diferencia (primera diferencia de las primeras diferencias):

$$(2n-1)b - (2n-3)b = [2n-1) - (2n-3)]b = \\ = [2n - 1 - 2n + 3]b = 2b$$

Este es el valor de la segunda diferencia de la serie y_{n-1}, y_n, y_{n+1}.

Si se resta de $(2n+1)b$, la diferencia anterior $(2n-1)b$, se obtiene como segunda diferencia:

$$(2n+1)b - (2n-1)b = [(2n+1) - (2n-1)]b = \\ = [2n + 1 - 2n + 1]b = 2b$$

Como puede verse, la segunda diferencia, independientemente del valor que se tenga para n (o, lo que es equivalente, para x_i) es constante en este caso.

O sea, que en el caso de una parábola de segundo grado, las segundas diferencias son constantes. La inversa también es cierta, o sea, que cuando las segundas diferencias son constantes, el movimiento queda representado por una parábola de segundo grado.

Por razonamientos análogos, podría mostrarse que:

Para un polinomio de grado N, las diferencias de orden N son constantes.

También podría mostrarse que:

Cuando las diferencias de orden N de una serie son constantes, la serie puede representarse por un polinomio de grado N.

De estas propiedades, en tanto que la segunda sirve para determinar en muchas ocasiones el tipo de curva a interpolar o de polinomio a perecuar, la primera tiene una utilidad muy grande en el cálculo simplificado de los valores teóricos de la tendencia.

En efecto, hemos visto que el cálculo de los valores de la tendencia requiere la substitución de los diversos valores de x_i en la ecuación general correspondiente o ley del fenómeno, y que esta substitución da lugar a cálculos embarazosos que tratamos de mecanizar mediante la elaboración de cuadros que nos den los diferentes valores requeridos para el cálculo final, pero que no nos libran de tener que ejecutar varias operaciones diversas, como la elevación de x_i a las diferentes potencias que se requieran, la multiplicación de dichas potencias por los coeficientes correspondientes, y la suma o resta de los diferentes valores obtenidos. Las propiedades de las diferencias finitas de orden N permiten reducir estos cálculos al mínimo, y transformar todas las operaciones restantes en simples procesos de suma.

Supongamos, para el caso, que la tendencia fuese rectilínea y hubiésemos obtenido como ley general una ecuación de la forma $y_i = a + bx_i$. Si calculamos los dos primeros valores y_1 y y_2 mediante la substitución de x_i por su valor y la ejecución de las operaciones correspondientes, y en seguida restamos del valor de y_2 el de y_1 obtendremos una primera diferencia de primer orden. Como sabemos que esta diferencia de primer orden será constante (o sea que la segunda diferencia de primer orden será la misma, la tercera diferencia de primer orden será la misma, etc.), sabemos de antemano que la diferencia entre y_3 y y_2 tendrá ese mismo valor, o sea que sumando dicho valor a y_2 obtendremos y_3. En forma análoga, sumando la diferencia constante de primer orden a y_3 obtendremos $y_4 \ldots$ sumándola a y_{n-1} obtendremos y_n.

Si suponemos el caso de una tendencia parabólica de segundo grado y calculamos los tres primeros valores de y_i (y_1, y_2, y_3) restando de y_2 y_1 y de y_3 y_2 obtendremos la primera y la segunda diferencia de primer orden que no tienen por qué ser iguales, ya que mientras el polinomio es de *2º* grado, las diferencias son de *1er.* orden. Si de la segunda diferencia de primer orden se resta la segunda diferencia de primer orden, se obtiene una primera diferencia de segundo orden cuyo valor será constante, ya que coinciden el grado del polinomio representativo de la tendencia (parábola de *2º* grado) y el número de orden de la diferencia (diferencia de *2º* orden). Esto quiere decir que, sin conocer el valor de la tercera diferencia de primer orden, ya sabemos que su diferencia con respecto a la segunda diferencia de primer orden es el valor recientemente obtenido, o sea, que si sumamos dicho valor a la segunda diferencia de primer orden obtendremos la tercer diferencia de primer orden, y que sumando ese mismo valor a esta tercer diferencia obtendremos la cuarta diferencia de primer orden o, en general sumando la diferencia de segundo orden a la diferencia $n - 1$ de primer orden se obtendrá la diferencia n de primer orden. Obtenidas las diferencias de primer orden y conocidas y_1, y_2 y y_3 que se obtuvieron por el procedimiento ordinario de cálculo, bastará con sumar a y_3 la correspondiente diferencia de primer orden para obtener y_4, a y_4 la correspondiente diferencia de primer orden para obtener y_5 o, en general, sumar a y_{n-1} la correspondiente diferencia de primer orden para obtener y_n.

En general, para obtener los valores teóricos de la tendencia por medio de las diferencias finitas:

1º—Calcúlense, por el procedimiento ordinario de substitución de los valores de x_i en la ecuación general, tantas y_i como sean necesarias para encontrar una diferencia de orden n (en general, si el grado del polinomio es N, se necesitará calcular por el procedimiento ordinario $N + 1$ *yes* teóricas).

2º—Mediante restas sucesivas entre cada *ye* y la precedente, calcúlense las primeras diferencias, mediante restas entre estas diferencias de primer orden, las de segundo orden, y así sucesivamente hasta obtener una diferencia de orden igual al grado del polinomio.

3º—Mediante suma de la diferencia de orden igual al grado del polinomio con las diferencias correspondientes de orden inmediato inferior, obténgase toda la serie de éstas, mediante la suma de éstas con las correspondientes de orden inmediato inferior, obténganse estas otras, hasta encontrar las de primer orden que, sumadas a las *yes* calculadas previamente, permitirán completar la serie de las *yes* calculadas.

VARIACIONES ESTACIONALES

Un gran número de fenómenos, a más de la tendencia general a aumentar o a disminuir en el curso de los años, muestran combinadamente una tendencia al alza en determinados meses o estaciones del año que contrasta con la tendencia que a la baja muestran en otros meses o estaciones. Esas fluctuaciones del fenómeno en el curso de los diversos meses o estaciones del año se conocen con el nombre de "variaciones estacionales".

Las variaciones estacionales se expresan, en general, como promedios de relativos (o de tantos por ciento) de los valores observados con respecto a ciertos valores centrales o normales, ajustados en tal forma que el promedio de tales promedios de relativos sea igual a 1 (o a 100 en caso de haberse trabajado en tantos por ciento).

Los índices así calculados, expresan en qué tanto (por ciento) se alejan proporcionalmente en más o en menos los valores reales de los valores que sería dable esperar si el fenómeno no estuviera sujeto a variación estacional.

Conforme al concepto que hemos dado de variaciones estacionales, cuatro resultan ser los pasos esenciales para el cálculo de los índices de variación estacional:

1º—Determinar el valor central o normal que servirá de base para calcular los relativos o los tantos por ciento: valor central del tipo de una media estadística, la mediana, etc., o valores normales como los de las magnitudes calculadas mediante la tendencia secular, por medio de promedios móviles, etc.

2º—Calcular los relativos (o los tantos por ciento) de cada valor observado respecto del valor central o de los valores normales elegidos. Para ello, habrá que dividir cada valor observado entre el valor elegido como base (con lo cual se obtendrá el relativo), o hacer dicha división y multiplicar el cociente por 100 (con lo cual se obtendría el tanto por ciento).

3º—Promediar todos los relativos o tantos por ciento correspondientes a todos los eneros, a todos los febreros, a todos los marzos, etc. (ya sea mediante el cálculo de la media aritmética de todos los valores de un mismo mes obtenidos para los diferentes años, o ya sea mediante la determinación de la mediana).

4º—Ajustar los promedios a fin de que su suma, dividida por 12 (meses del año) resulte igual a 1 (si se trabajó con relativos) o a 100 (si se trabajó con tantos por ciento).

En seguida distinguiremos, dentro de este proceso general que todos siguen, según trataremos de mostrar, cuatro técnicas para el cálculo de los índices de variación estacional:

a.—Procedimiento de las medias mensuales.
b.—Procedimiento de relativos de la tendencia.
c.—Procedimiento de relativos de las medias móviles.
d.—Procedimiento de los eslabones relativos.

Procedimiento de las medias mensuales.—Este es uno de los procedimientos más sencillos y rápidos, y consiste fundamentalmente en (seguimos los lineamientos del proceso general de cálculo de las variaciones estacionales, delineado anteriormente):

1º—Elegir como base de comparación un valor central (particularmente la media); para ello:
  a.—Calcular la media de cada mes, sumando los valores observados y dividiéndolos entre el número de años en que se hayan hechos las observaciones.
  b.—Sumar las medias de todos los meses y dividir el resultado entre 12 (meses) para obtener la media general.

2º—Calcular los relativos (o %s) de los valores observados con respecto a dicho valor central o media, dividiendo cada uno de los valores observados entre la media general (y multiplicando el resultado por 100 si se desea obtener en forma de porciento).

3º—Promediar todos los relativos correspondientes a un mismo mes, sumándolos y dividiéndolos entre el número de años o determinar el valor mediano de dichos relativos contando tantos del año más remoto hacia el más próximo como del más próximo hacia el más remoto y eligiendo el central (donde las cuentas se cruzan).

4º—Ajustar los promedios correspondientes a cada uno de los meses, para lo cual:
  a.—Se suman dichos promedios mensuales y se dividen entre 12 (meses).
  b.—Se ve si este promedio (puesto que puede determinarse también la mediana en lugar de calcular la media aritmética) es:
  Igual a 1 (o 100 si son %s) en cuyo caso no necesitan ajuste,
  Mayor que 1 (o 100) caso en el que el exceso deberá dividirse entre 12 (meses) y restarse de cada uno de los promedios mensuales obtenidos (que ya son índices estacionales pero

"no ajustados"), a fin de obtener los índices estacionales ajustados.

Menor que 1 (o que 100) caso en el que el defecto deberá dividirse entre 12 (meses) y sumarse a cada uno de los promedios mensuales obtenidos, a fin de obtener índices estacionales ajustados.

Procedimiento de relativos de la tendencia.—Este es uno de los procedimientos que traban más lógicamente dentro del marco analítico de las series temporales o cronológicas. Fundamentalmente, consiste en:

1º—Elegir como término de comparación una serie de valores normales (los valores calculados de la tendencia), para lo cual:

a.—Se obtienen los promedios de cada año, ya sea sumando los valores de todos los meses y dividiendo entre 12 (media aritmética) o, ya sea determinado el valor mediano para lo cual será necesario ordenar los valores mensuales en orden creciente o decreciente, determinar cuáles ocupan los lugares sexto y séptimo, sumarlos y dividirlos entre 2 (mediana).

b.—Con los promedios de cada año (sean medias aritméticas o medianas) se forma una serie cronológica de valores a los cuales se les interpola la curva correspondiente.

c.—Se determinan los valores mensuales de la tendencia, substituyendo la x_t de la ley general del fenómeno por $1/12$, $2/12 \ldots 12/12 = 1$, $1 + 1/12 \ldots 2 \ldots 3 \ldots 4 \ldots n$ (o utilizando cualquier otro procedimiento de simplificación como el de obtención de los valores de la tendencia mediante diferencias finitas).

2º—Calcular los relativos (o %s) de los valores observados con respecto a los valores normales correspondientes dividiendo el valor observado en un mes específico entre el valor de la tendencia calculado para ese mes de ese año determinado.

3º—Promediar todos los relativos (o %s) correspondientes a un mismo mes en los diferentes años, ya sea sumándolos y dividiendo la suma entre el número de años (media aritmética) o ya sea determinando la mediana.

4º—Ajustar los promedios de relativos (o de %s) de cada mes, para lo cual:

a.—Se calcula el promedio (media aritmética o mediana) de dichos promedios de relativos mensuales, o sea el promedio mensual general (sumando y dividiendo entre 12 u ordenando, determinando los valores sexto y séptimo, sumándolos y dividiendo la suma entre dos).

b.—Se divide cada promedio mensual de relativos (que es ya un índice estacional no ajustado) entre el promedio general, con lo cual se obtiene un índice ajustado de variación estacional.

Procedimiento de los relativos de las medias móviles.—Si se tiene una serie de x_i, se conoce como "suma móvil de 2 en 2" la que se obtiene al sumar x_1 con x_2, x_2 con x_3, x_3 con x_4, ... o, en general x_n con x_{n+1}; como "suma móvil de 3 en 3" se conocería a la que se obtuviera de sumar x_1 con x_2 y x_3, a x_2 con x_3 y x_4 ... o, en general x_n con x_{n+1} y con x_{n+2}. En general, las "sumas móviles de m en m" serán las que se obtengan al sumar m valores a partir del primero, m valores a partir del segundo, m valores a partir del tercero, etc. Con la notación del operador Σ, dichas sumas móviles podrían representarse como sigue:

$$\sum_{i=1}^{N} (x_i + x_{i+1}) \text{ para la suma móvil de 2 en 2.}$$

$$\sum_{i=1}^{N} (x_i + x_{i+1} + x_{i+2}) \text{ para la suma móvil de 3 en 3.}$$

$$\sum_{i=1}^{N} (x_i + x_{i+1} \ldots + x_{i+m}) \text{ para la suma móvil de } m \text{ en } m.$$

O bien, puesto que la suma del paréntesis puede escribirse simplificadamente mediante un operador Σ:

$$\sum_{i=1}^{N} \left(\sum_{i=1}^{N=m} \right) \text{para la suma móvil de } m \text{ en } m.$$

En forma análoga a como hemos dado la noción de suma móvil puede darse la de media móvil, si se tiene en cuenta que si se divide cada "suma móvil de m en m" entre m (número de datos considerado) se obtendrá la correspondiente "media móvil de m en m".

Cuando el valor de m para la suma móvil y para la media móvil es impar, el resultado de cada media móvil es equiparable al valor central de los que han servido para obtenerlo; así, si se ha calculado una media móvil de 3 en 3, la primera media móvil no corresponderá ni al primer dato ni al tercero, sino que será comparable con el segundo, la segunda media móvil no lo será ni con el segundo ni con el cuarto sino con el tercero (ya que para obtenerla se sumaron el segundo, tercero y cuarto). Es fácil comprender el que al calcular las medias móviles queden valores inequiparables a una media móvil: en el caso de medias móviles de 3 en 3, no hay posibilidad de equiparar con una media móvil ni el primer dato ni el último; en el caso de medias móviles de 5 en 5, no hay posibilidad de equiparar con una media móvil ni los dos primeros ni los dos últimos, o sea, que, en general, se pierden $m - 1$ datos, de los cuales $(m - 1)/2$ corresponden al principio y $(m - 1)/2$ al fin de la distribución en cuanto m es impar. Cuando m es par se pierden asimismo $m - 1$ datos, de los cuales $(m - 1)/2$ corresponden al principio y $(m - 1)/2$ al fin de tal distribución; sin embargo, a diferencia de lo que ocurre en el caso de m impar, en que las medias móviles obtenidas son equiparables con alguno de los valores de la serie, en el caso de m par, las medias móviles no son equiparables a los valores de la serie, ya que, por ejemplo, si se trata de medias móviles de 2 en 2, la primera media móvil no puede equipararse ni al primero ni al segundo valor de los que sirvieron para calcularla, sino a un punto intermedio entre ambos, la segunda media móvil no corresponde ni al segundo ni al tercero sino a un punto intermedio, etc., lo cual hace necesaria —en cuanto tratan de equipararse las medias con los valores de la serie— una nueva operación llamada "de centrar", y la cual consiste en calcular la media móvil de 2 en 2 de las medias móviles previamente obtenidas. Como es fácil comprender, esta operación de centrar las medias móviles sólo tiene razón de ser en caso de m par, ya que en caso de m impar, la operación de centrar se convertiría en operación descentradora.

Con estos conceptos previos, será fácil entender la forma en que las medias móviles permiten el cálculo de las variaciones estacionales de acuerdo con el procedimiento siguiente:

1º—Se eligen ciertos valores normales (medias móviles de 12 en 12 datos, por ser doce los meses, centradas para el séptimo mes) como base de comparación, calculando para ello:

a.—La suma móvil de 12 en 12 meses, sumando de enero a diciembre del primer año, de febrero del primer año a enero del segundo, de marzo del primero a febrero del segundo... como si los meses constituyeran (como en realidad constituyen) una serie contínua.

b.—La media móvil correspondiente, dividiendo cada suma móvil entre 12.

c.—La media móvil centrada, sumando la primera media móvil con la segunda, la segunda con la tercera, etc., y dividiendo cada suma entre dos. Esto deberá hacer que cada media móvil de 12 en 12 centrada para el 7º mes haga coincidir la primer media móvil centrada con julio del primer año, la segunda media móvil centrada con agosto... la séptima media móvil centrada con enero del segundo año..., etc.

2º—Se calculan los relativos (o %s) de los valores observados con respecto a las medias móviles centradas correspondientes dividiendo dichos valores observados entre dichas medias móviles.

3º—Se promedian los relativos de los valores observados respecto de las medias móviles centradas correspondientes a un mismo mes (los relativos de todos los julios se suman y se divide la suma entre el número de los relativos sumados, que será igual al número de años menos 1; se hace lo mismo con los relativos de los agostos... de los eneros, de los febreros, ... etc.).

4º—Los promedios obtenidos para cada mes (índices no ajustados de variación estacional) se ajustan por el procedimiento ya indicado en el caso de los relativos de la tendencia, a fin de obtener los correspondientes índices ajustados de variación estacional.

Hay quienes consideran (Rigleman and Frisbee, por ejemplo) que los índices de variación estacional calculados por el procedimiento de los relativos de las medias móviles son más representativos por seguir la media móvil de doce meses bastante de cerca la curva de los ciclos.

Procedimiento de los eslabones relativos.—En el procedimiento de los eslabones relativos, el proceso general delineado en las primeras páginas como esquematización de la forma de calcular las variaciones estacionales, se convierte en un proceso complejo, ya que, como se verá en seguida, es necesario obtener relativos en dos ocasiones: al determinar los valores de los "eslabones" y al formar la "cadena de relativos"; por otra parte, el ajuste comprende dos operaciones principales: la eliminación de la discrepancia de los valores obtenidos para un mismo mes en dos años sucesivos dentro de la cadena de relativos, y la expresión de cada elemento de dicha cadena como relativo (o como %) de la media de la cadena ajustada por discrepancia de los valores de un mismo mes en dos años sucesivos.

Esto muestra la necesidad que hay de usar dos ordenaciones de las etapas del procedimiento de los eslabones relativos: una que muestre la sucesión de las diferentes fases del mismo con independencia de cualquier otra consideración (y dentro de la cual usaremos la notación 1º, 2º, 3º, etc.), y otra que señale la relación de esas diferentes fases con las etapas del proceso general de cálculo de las variaciones estacionales (para lo cual usaremos los romanos I, II, III, ... etc. en correspondencia con las cuatro etapas del proceso general).

Conforme a estas indicaciones previas, el procedimiento de cálculo de las variaciones estacionales puede delinearse en la forma siguiente:

1º—(I).—Elíjanse como base de comparación los valores correspondientes al mes inmediato anterior a aquel del cual se va a calcular el eslabón relativo (Diferencia con respecto a los anteriores procedimientos: ni elección de un valor central ni de una serie de valores normales calculados, sino de una serie de valores empíricamente observados).

2º—(II).—Cálculo de los relativos (o de los %s) de cada valor obserrrespondiente a febrero entre el de enero, del correspondiente a marzo entre el de febrero... del correspondiente a enero del segundo año entre el que corresponda a diciembre del primer año, etc.) para obtener los "eslabones" correspondientes que indican lo que representa el valor de cada mes con respecto al valor del precedente.

3º—(II).—Determinación del promedio (en este caso de la mediana y no cálculo de la media aritmética) de todos los eslabones correspondientes a un mismo mes a fin de obtener las "medianas mensuales de eslabones" que indiquen cuánto, en promedio, representa el valor de cada mes con respecto al anterior, en la serie de años estudiada.

4º—(I).—Elección de la mediana de eslabones correspondiente a enero como base fija de comparación de las demás medianas mensuales.

5º—(II).—Cálculo de los relativos (o %s) de cada promedio mensual con respecto a la base fija (mediana de enero), teniendo en cuenta que:

a.—El relativo de enero con respecto a si mismo es igual a 1 ($x_e/x_e = 1$) o a 100 (si se trabaja con %s).

b.—El relativo de febrero con respecto a enero (x_f/x_e) es la mediana misma de los eslabones de febrero con respecto a enero (x_f/x_e).

c.—El relativo de marzo con respecto a enero es igual a la mediana de los eslabones de marzo con respecto a febrero (x_m/x_f) multiplicada por el relativo de febrero con respecto a enero (x_f/x_e). En efecto, el producto de estos relativos produce ($x_m/x_f.\ x_f/x_e$) x_m/x_e.

d.—El relativo de abril con respecto a enero es igual a la mediana de los eslabones de abril con respecto a marzo

(x_a/x_m) multiplicada por el relativo de marzo con respecto a enero acabado de obtener $(x_a/x_m \cdot x_m/x_e = x_a/x_e)$.

e.—En general, el relativo de un mes con respecto a enero será igual a la mediana de los eslabones del mes con respecto al mes precedente, multiplicada por el relativo del mes precedente con respecto a enero.

f.—El relativo de enero con respecto a enero será igual a la mediana de los eslabones relativos de enero con respecto a diciembre del año anterior (x_e/x_d) multiplicada por el relativo del mes precedente (diciembre) con respecto a enero (x_d/x_e). Este producto debería dar 1 $(x_e/x_d \cdot x_d/x_e = 1)$, pero esto ocurre raras veces, lo cual es de esperar, si se tiene en cuenta que el diciembre con respecto al cual se tomaron los eslabones y está dada la mediana de los eslabones de enero es el diciembre del año anterior, y que el relativo de diciembre con respecto a enero contiene como numerador el diciembre del año mismo (o sea llamando x_{d-1} al del año previo y x_d al del mismo año, se tendrá: $x_e/x_{d-1} \cdot x_d/x_e \neq 1$ e igual a x_{d-1}/x_d). Esta diferencia con respecto a 1 (o a 100 si se trata de %s) impone el ajuste.

6º—(IV).—Ajuste en dos etapas:

a.—Reducción del segundo relativo de enero (el encontrado hace un momento) al valor 1 (o 100) del primer relativo de enero, considerando la discrepancia entre ambos como debida a la tendencia (por tanto, igual a *b* si la tendencia es rectilínea) por lo que: se dividirá la discrepancia entre 12 y se (restará) (o sumará) 1/12 al valor de febrero, 2/12 al de marzo, etc., . . . 12/12 o sea la discrepancia entera al segundo relativo de enero que quedará en 1 (o en 100).

5º—(II).—

b.—Reducción de los valores anteriores (que dan cada valor mensual con respecto a enero) a relativos (o %s) del promedio de los mismos, para lo cual: se encontrará la mediana (o se calculará la media aritmética de los valores obtenidos) y se dividirá cada uno de esos valores entre dicho promedio, obteniéndose así los índices ajustados de variación estacional por el método de los eslabones relativos.

AJUSTE POR TENDENCIA DE LOS ÍNDICES ESTACIONALES CALCULADOS MEDIANTE ESLABONES RELATIVOS, CUANDO LA TENDENCIA NO ES RECTILÍNEA

En el procedimiento de cálculo de los índices de variación estacional mediante los eslabones relativos se habló ejemplificativamente del ajuste por tendencia, haciendo mención del caso de una tendencia rectilínea. En lo que sigue, trataremos de señalar la forma de ajuste en términos más generales para cuando la tendencia está representada por una expresión de grado n.

Para el efecto, consideraremos el valor del primer enero como resultante de considerar que no ha transcurrido una sola unidad de tiempo, resultando por lo mismo igual al parámetro (a) "ordenada en el origen" de la curva representativa de la tendencia. Consideraremos asimismo que el paso del primero al segundo enero representa el transcurso de una unidad de tiempo (o bien una diferencia de 1 en las abscisas de la curva), correspondiendo por tanto 1/12 a febrero, 2/12 a marzo, $m/12$ al enésimo mes, como diferencia de abscisas o como lapso transcurrido de enero al mes que se considera.

Conforme a lo anterior y si pensamos en una forma simplificada de expresión de la tendencia, los valores para los diferentes meses, serán:

enero:	$a + b(0/12)^n$	a
febrero	$a + b(1/12)^n$	
mes de orden m:	$a + b(m/12)^n$	
enero_2:	$a + b(12/12)^n$	$a + b$

O sea, que los dos eneros discrepan en b, y para lograr que el segundo sea igual al primero se necesitará restar de su valor $(12/12)^n$ de b; en el caso del mes de orden m, si se quiere eliminar la tendencia habrá que restar $(m/12)^n$ de b.

Así si se trata de un fenómeno con tendencia rectilínea y se busca el valor de marzo una vez eliminada la tendencia, por ser marzo el tercer mes (*segundo* después de enero) habrá que buscar la diferencia b entre los eneros, y multiplicar su valor por $(2/12)^1 = 2/12$. Si la tendencia fuese parabólica de segundo grado y se buscara el valor de abril, por ser abril el cuarto mes (*tercero* después de enero) sería necesario multiplicar dicha diferencia b por

$(3/12)^2 = 9/144$ y restar dicho valor del que se tenía para abril. Si la tendencia fuese parabólica de tercer grado y se buscara el valor de septiembre, siendo septiembre el noveno mes (*octavo* después de enero) se necesitaría multiplicar dicha diferencia b por $(8/12)^3 = 512/1728$... etc., restando la diferencia del valor dado para el mes en cuestión si el valor del segundo enero es superior al del primero, o sumándola si el valor del segundo enero es inferior al del primero.

Ajuste cuando la tendencia es exponencial: Si como expresión de una tendencia exponencial aceptamos $y = ab^x$, conforme a los supuestos anteriores, tendremos:

enero $_1$:	ab^0	a
febrero:	$ab^{\frac{1}{12}}$	
marzo:	$ab^{\frac{2}{12}}$	
mes de orden m:	$ab^{\frac{m}{12}}$	
enero $_2$:	$ab^{\frac{12}{12}}$	ab

Como puede comprenderse fácilmente, de lo que se trata no es ya de una diferencia entre el segundo y el primer enero, sino propiamente de un cociente entre ambos; lo que interesa entonces es dividir el valor del segundo enero entre el del primero para obtener b. O sea que si los dos eneros son uno respecto de otro b veces mayor, para lograr que el segundo sea igual al primero se necesitará dividir al segundo entre b; en el caso del mes de orden m si se quiere eliminar la tendencia habrá también que dividir entre la raíz doceava de la potencia emésima de b, o entre b elevado a m sobre doce (que son una misma cosa).

Como extraer raíces superiores a la tercera requiere un proceso laborioso, pueden utilizarse los logaritmos, y en vez de dividir entre la potencia m sobre doce de b, se puede encontrar el logaritmo de b, multiplicarlo por m y dividirlo entre 12, obtener el antilogaritmo del resultado y entre este antilogaritmo dividir el valor del mes emésimo.

Así, en tratándose de un valor cuya tendencia sea exponencial, si se busca el valor de marzo una vez eliminada la tendencia, por ser marzo el segundo mes después de enero, habrá que buscar el cociente b entre los eneros (o la diferencia entre los logaritmos de los eneros) buscar su logaritmo (que si se buscó la diferencia entre los logaritmos de los eneros se habrá obtenido ya directamente), dividir log b entre 12 y multiplicarlo por 2 (o multiplicar por 2 y dividirlo entre 12), buscar el antilogaritmo del resultado y dividir el valor que se tenía para marzo entre el antilogaritmo así obtenido. En caso de que la tendencia fuese exponencial decreciente sería necesario multiplicar los valores por el antilogaritmo en lugar de dividirlos entre el mismo.

VARIACIONES CÍCLICAS

Las variaciones cíclicas son movimientos de ascenso y descenso alternativo de los valores de una serie cronológica, la cual pasa de una situación normal a un período de prosperidad, de éste a uno de receso, a uno de depresión, a uno de recuperación, y a un nuevo período normal para volver a iniciar un nuevo ciclo.

Las variaciones cíclicas se miden fundamentalmente como desviaciones porcentuales con respecto a los valores normales de la tendencia corregida por la variación estacional.

Conforme a lo anterior, el cálculo de la variación cíclica comporta:

1º—Calcular los valores de la tendencia en la forma que ya es habitual.

2º—Calcular los índices de variación estacional (12 índices en caso de tratarse de datos mensuales, que serán comunes a todos los años considerados).

3º—Calcular los valores normales de la tendencia corregida por la variación estacional, multiplicando el valor de la tendencia para un mes determinado de un año también determinado por el índice de variación estacional correspondiente a dicho mes en la serie.

4º—Expresar los valores observados como relativos (o como %s) de los valores normales, dividiendo cada valor de la serie empírica entre el correspondiente valor normal.

5º—Restar de los relativos de los valores normales con respecto a los valores normales, 1 (en caso de haberse trabajado con relativos) o 100 (en caso de haberse trabajado con tantos por ciento). Los valores obtenidos son las variaciones cíclicas, ya que, por este procedimiento, se han eliminado la tendencia y las variaciones estacionales.

En caso de necesidad de comparación de las variaciones cíclicas de un fenómeno A con las variaciones cíclicas de otro fenómeno B, conviene reducir las variaciones de uno y otro a términos comparables mediante la expresión de las mismas en unidades sigmáticas, lo cual requiere que para cada uno de los fenómenos estudiados:

a.—Se eleven al cuadrado las desviaciones cíclicas obtenidas.
b.—Se sumen dichos cuadrados.
c.—Se dividan entre el número de meses o estaciones que figuren en la serie.
d.—Se extraiga la raíz cuadrada del cociente, con lo cual se habrá calculado el valor de δ para el fenómeno.
e.—Se divida cada desviación cíclica entre el valor encontrado, con lo cual dichas variaciones serán comparables con las obtenidas en forma análoga en el caso de otros fenómenos.

NÚMEROS ÍNDICES

Los números índices son, fundamentalmente, relativos o medias de relativos que sirven para apreciar las variaciones conjuntas de un grupo de variables con respecto a la magnitud alcanzada por las mismas en cierto período, lugar o institución, tomados como base.

Si u_i, v_i, w_i es el conjunto de variables por estudiar, podremos representar por u_0, v_0, w_0 los valores correspondientes alcanzados por dichas variables en el período, lugar o institución que ha de servir como base de comparación, y por u_1, v_1, w_1; u_2, v_2, w_2, . . . los valores alcanzados por esas mismas variables en períodos distintos, en lugares distintos o en instituciones distintas de aquellos que sirven como base de comparación.

En cuanto los números índices son fundamentalmente relativos de variaciones conjuntas, será necesario pensar en la forma de conjuntar los valores de las variables consideradas tanto en el período, lugar o institución que se estudie como en el período, lugar o institución base y tras ello relacionar matemáticamente esos conjuntos de variables.

La forma más sencilla e inmediata de conjuntación que se ofrece es la suma de los valores de la variable, de una parte, en el período, lugar o institución de estudio, y, de otra, en el período lugar o institución base. Si u_i, v_i, w_i nos representan los valores de las variables en el período, lugar o institución de estudio (valores que, como caso particular pueden ser iguales a los del período, lugar o institución de base), su suma $u_i + v_i + w_i$ nos representará el valor conjunto de esas variables en dicho período, lugar o institución. A su vez, $u_0 + v_0 + w_0$ representará el valor conjunto de esas mismas variables en el período, lugar o institución elegidos como base. Si por p representamos a las diferentes variables del conjunto (u_i, v_i, w_i), la magnitud conjunta del grupo en el período, lugar o institución de estudio podrá representarse por Σp_i, y la magnitud conjunta del grupo en el período, lugar o institución base por Σp_0. Como el número índice es fundamentalmente el relativo que mide la variación entre la magnitud conjunta del grupo de variables en el período, lugar o institución de estudio y la magnitud conjunta del mismo grupo de variables en el período, lugar o institución de base, podremos obtener dicho valor dividiendo $\Sigma p_i / \Sigma p_0$.

Este índice —el más sencillo de los que se puedan considerar— es el que se conoce como "índice agregativo" o "índice aditivo".

Pero los números índices pueden ser no sólo relación entre las variaciones conjuntas de un grupo de variables en dos lugares, períodos o instituciones distintas, sino considerarse también como medias del conjunto de las variaciones de un grupo de variables, en dos lugares, períodos o instituciones distintas. Y, esta nueva concepción de los números índices, no obstante su aparente semejanza con la anterior (en efecto, sin forzar mucho las cosas se podría incluso hablar en el primer caso de "las medias de las variaciones conjuntas"... que convertiría la apariencia de semejanza en aparente identidad), es una concepción distinta de las cosas: noción que marca la primacía de la variación de cada variable individual conjuntada posteriormente con la variación de las otras variables, frente a la noción previa que marca la primacía de la variación conjunta de todas las variables previamente conjuntadas (y hacemos valer aquí el pleonasmo por su calidad de subrayante).

Conforme a esta nueva noción, es necesario obtener, en primer término, la variación relativa de cada variable con respecto a la magnitud de la misma en el período, lugar o institución que sirven como base y, ulteriormente, conjuntar dichas variaciones mediante el cálculo de las medias correspondientes. O sea, en primer término, calcular los relativos u_i/u_0, v_i/v_0, w_i/w_0 de cada variable y, en seguida ejecutar con ellos las operaciones necesarias para el cálculo de una media o sea, si ejemplificamos con la media aritmética, sumar dichos relativos y dividir entre el número de ellos (que es al mismo tiempo el número de las variables):

$$\frac{u_i}{u_0} + \frac{v_i}{v_0} + \frac{w_i}{w_0} N \quad \text{o bien} \quad \frac{\Sigma\left(\frac{p_i}{p_0}\right)}{N}$$

Como es fácil comprender, podrán calcularse tantos números índices como medias distintas según las necesidades prácticas lo impongan. Aquí mencionaremos sólo los cálculos mediante:

1.—Media aritmética de los relativos.
2.—Media geométrica de los relativos.
3.—Media armónica de los relativos.

Media aritmética de los relativos.—Para obtener números índices mediante este procedimiento:

1º—Divídase el valor alcanzado por cada variable en el período, lugar o institución estudiado entre el valor alcanzado por esa misma variable en el período, lugar o institución de base, con lo que se obtendrán los relativos correspondientes.
2º—Súmense los relativos de las diversas variables del grupo.
3º—Divídase dicha suma entre el número de variables.

La fórmula correspondiente sería:

$$\frac{\Sigma\left(\frac{p_i}{p_0}\right)}{N}$$

Media geométrica de los relativos.—El procedimiento de cálculo podría resumirse como sigue:

1º—Divídase el valor de cada variable en el período, lugar o institución de estudio entre el valor alcanzado por la misma en el período, lugar o institución de base.
2º—Encuéntrese el logaritmo de cada uno de los relativos así obtenidos.
3º—Súmense los logaritmos de los relativos.
4º—Divídase la suma de los logaritmos entre el número de variables consideradas en el grupo.
5º—Búsquese el antilogaritmo del cociente (cf. obtención de la media geométrica en general).

La fórmula correspondiente a este procedimiento sería:

$$\textit{Indice}_g = \text{Antilog}\,\frac{\Sigma(\log\left(\frac{p_i}{p_0}\right)}{N}$$

Media armónica de los relativos.—No obstante que en éste como en los otros casos, el título mismo indica cuáles son las operaciones por realizar, esquematizaremos el procedimiento en la siguiente forma:

1º—Calcúlense los relativos del valor de la variable en el período, lugar o institución de estudio con respecto al de esa misma variable en el período, lugar o institución de base, dividiendo el primero entre el segundo.

2º—Obténganse los recíprocos de dichos relativos.

3º—Súmense los recíprocos de los relativos.

4º—Divídase la suma de esos recíprocos entre el número de variables consideradas.

5º—Calcúlese el recíproco de dicho cociente.

La fórmula correspondiente es:

$$Índice_a = \left(\frac{\Sigma \left(\frac{p_i}{p_0} \right)^{-1}}{N} \right)$$

Si se toma en cuenta que el recíproco de p_i/p_0 es una fracción que tiene por numerador el denominador de su recíproca y por denominador el numerador de la misma (p_0/p_i) ya que multiplicadas dan 1, y que, asimismo el recíproco de la suma de los recíprocos de los relativos entre el número de variables es igual a una fracción que tenga por numerador el número de variables y por denominador la suma de los recíprocos de los relativos podrán hacerse en la fórmula y en el procedimiento, las transformaciones siguientes:

$$Índice_a = \frac{\Sigma \left(\frac{p_0}{p_i} \right)^{-1}}{N}$$

$$Índice_a = \frac{N}{\Sigma \left(\frac{p_0}{p_i} \right)}$$

En forma de procedimiento, el cálculo de este índice se reduciría a:

1º—Calcular los relativos de los valores de la variable en el período, lugar o institución *de base* con respecto a los valores de dicha variable en el período, lugar o institución *de estudio.*

2º—Sumar dichos relativos, y

3º—Dividir *entre dicha suma,* el número de variables consideradas.

Ponderación de los números índices.—En el cálculo de los números índices es necesario tener en consideración que no todas las variables tienen la misma importancia dentro del conjunto, sino que hay unas que tienen mayor importancia que otras, unas que tienen mayor peso o que pesan más que otras, o mejor aún, unas que son más densas que otras.

La necesidad de tomar en cuenta la diferente importancia de las variables dentro del conjunto se refleja matemáticamente en la precisión que hay de ponderar cada una de esas variables, o sea, de multiplicar cada una de esas variables por un factor adecuado de ponderación que, en la mayoría de los casos es la frecuencia con la que se presenta la magnitud considerada de la variable.

Así, por ejemplo, en cuanto se tratan de calcular índices de precio ("patria de origen" de los números índices), el factor de ponderación de los precios por unidad es el número de unidades (producidas o consumidas, según el caso) y, ese número de unidades es, fundamentalmente, la frecuencia correspondiente a los precios unitarios.

Sin embargo, para que la ponderación en vez de ser mecánica tenga sentido en el caso del cálculo de los números índices deberá dársele una orientación un tanto diferente de la acostumbrada; en efecto, es indispensable considerar que los números índices pretenden poner de relieve una relación entre las variables mismas, y que si los valores de las variables se multiplican por los factores de ponderación correspondientes a los períodos, lugares o instituciones de estudio (en tratándose de la magnitud de las variables para dicho período, lugar o institución de estudio) y de base (en tratándose de las magnitudes para período, lugar o institución de base), lo que medirá el índice será la variación total de fenómeno.

En cambio, si un mismo factor de ponderación (frecuencia correspondiente al período, lugar o institución de base *O* frecuencia correspondiente al período, lugar o institución de estudio) se aplica a la magnitud de la variable en el período, lugar o institución de base *Y* a la magnitud de la variable en el período, lugar o institución de estudio, esto equivaldrá a hacer que permanezca constante uno de los factores que intervienen en el total y que, por lo mismo, se midan únicamente las variaciones del factor que interesa (el conjunto de variables).

Así, si por ejemplo, nos referimos a índices de precios, multiplicar los precios del año de estudio por las cantidades del año base y los precios del año base por las cantidades del año base equivale a hacer el estudio de la variación de precios *como si* en el transcurso del tiempo entre el año base y el año de estudio la producción no hubiese aumentado ni disminuido. En forma análoga multiplicar tanto los precios del año de estudio como los precios del año base por las cantidades del año de estudio, equivale a estudiar la variación de precios *como si* o *en el supuesto de que* en el año base se hubiesen producido las mismas cantidades que se han producido en el año de estudio.

Como podrá comprenderse, en un caso se obtienen valores ligeramente más altos, en otro ligeramente más bajos, y la solución que de inmediato se ofrece es promediar los obtenidos mediante un sistema de ponderación con los obtenidos mediante el otro sistema, llegándose por este procedimiento a fórmulas como la llamada "Fórmula ideal de Fisher", que es fundamentalmente una media geométrica ($\sqrt{}$) de índices ponderados mediante cantidades en el año base $\left(\frac{\Sigma p_i q_0}{\Sigma p_0 q_0}\right)$ y mediante cantidades en el de estudio $\left(\frac{\Sigma p_i q_i}{\Sigma p_0 q_i}\right)$ lo que da como fórmula:

$$I = \sqrt{\frac{\Sigma p_i q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_i q_i}{\Sigma p_0 q_i}}$$

AJUSTE PREVIO DE SERIES CRONOLÓGICAS

En muchas ocasiones, el verdadero carácter de una serie cronológica está enmascarado por la influencia variable que sobre los datos de la misma ejercen otros fenómenos concomitantes que, en veces hacen aparecer como mayor y en otras como menor la magnitud del fenómeno estudiado correspondiente a una fecha determinada. Las influencias de este tipo se ejercen principalmente sobre datos de carácter económico (ventas, producción, etc.) y requieren que se hagan con las series cronológicas por estudiar ajustes previos, entre los que cuentan como principales:

1.—Ajuste deflactario, y
2.—Ajuste calendárico.

El ajuste deflactario tiene por objeto eliminar de las series por estudiar todas aquellas variaciones de magnitud que no dependen del mismo y puro desarrollo del fenómeno, sino que son producto de las variaciones de precios registradas en el período; o sea, que deflactar los datos de una serie cronológica consiste fundamentalmente en eliminar de ella lo que son variaciones conjuntas del sistema al que pertenece o por el que se encuentra directamente influída, para dejar como material de estudio las variaciones que le son propias.

Para ajustar deflactariamente los datos de una serie se hace uso de los índices de precios que, fundamentalmente, son números relativos que expresan la relación matemática entre el sistema de precios de un año dado y ese mismo sistema en un año que se toma como base de comparación. El afirmar que a un año dado corresponde un índice de 1 o de 100 equivale a indicar que el sistema de precios en ese año es el mismo (en conjunto) del año base; el que a un año corresponda un índice menor que 1 o menor que 100 quiere decir que en ese año los precios descendieron por debajo del nivel correspondiente al año de base; en forma análoga, el que a un año corresponda un índice mayor que 1 o mayor que 100 significa que en dicho año el nivel del sistema general de precios fue mayor que el nivel correspondiente al año base.

De este modo, en el primer supuesto, o sea en el de un año que tenga por índice de precios 1 o 100 %, el ajuste deflactario dejará sin alteración el dato registrado originalmente para la serie considerada; en el segundo supuesto, o sea en el caso de un índice inferior a 1 o a 100 % el hecho se interpretará —con vistas al ajuste— en el sentido de que, a no haber sido por la variación conjunta del sistema general de precios, el valor o magnitud del fenómeno para la fecha considerada hubiera sido mayor que el valor o magnitud registrados en dicha fecha, siendo por lo mismo la función y efecto del ajuste deflactario los de elevar la cifra registrada; en el tercer supuesto, o sea cuando el índice es superior a 1 o a 100 %, deberá pensarse en que, a no haber sido por la variación conjunta del sistema general de precios, el valor o magnitud del fenómeno para la fecha considerada hubiese sido menor que la registrada, debiendo ser el efecto del ajuste deflactario el de dar dicha cifra, menor que la registrada.

Matemáticamente, los resultados anteriores pueden obtenerse expresando los datos originales como relativos o porcientos del número índice correspondiente, o sea, dividiéndolos entre dicho número índice, ya que las variaciones de los resultados se producirán en sentido inverso a aquel en que se producen las variaciones del índice de precios, denominador de la fracción.

Conforme a lo anterior, el ajuste deflactario puede concretarse en la siguiente forma:

$$\textit{Valor deflactado} = \frac{\textit{Valor no deflactado}}{\textit{Número índice}}$$

El segundo de los ajustes previos importantes en el estudio de las series cronológicas es el ajuste calendárico, mediante el cual se eliminan los efectos que sobre las serie tienen las diferencias de períodos de tiempo que, de acuerdo con su denominación, podrían parecer equivalentes y que, en la realidad, no lo son. Es así como el ajuste calendárico deberá considerar en ciertas ocasiones si el año es o no bisiesto, o bien que dos períodos de tiempo, no obstante su común denominación de "meses" del año no comprenden un mismo número de días, o bien que en cada uno de dichos meses, el número de días efectivamente trabajados es distinto según el número de domingos (y de sábados en caso de tratarse de la llamada "semana inglesa") y según el número de fiestas fijas y movibles, etc.

En esta parte, nos ocuparemos específicamente del ajuste calendárico de tipo más general o sea, de aquel que tiene en cuenta el carácter bisiesto o no del año y la diferencia en número de días de los diversos meses del año, ya que el número de días efectivamente trabajados estará en función de las fiestas que se celebren y en las que se suspenda el trabajo en cada país o región.

El ajuste calendárico puede hacerse de acuerdo con dos procedimientos generales, uno aplicable con antelación a la etapa recolectora de los datos; otro utilizable cuando dicha etapa se ha cumplido ya.

El procedimiento aplicable antes de la recolección equivale propiamente a evitar el ajuste calendárico y consiste fundamentalmente en pasar por sobre las unidades convencionales de división del año y constituir dentro del mismo un cierto número de períodos de igual duración. Así, hay quienes acostumbran dividir el año en 13 períodos de 28 días cada uno, dejando sin registrar un día en los años no bisiestos y 2 en los bisiestos.

El procedimiento de ajuste aplicable a datos ya registrados responde fundamentalmente a las mismas necesidades y consideraciones hechas en el caso de ajuste deflactario; dichas necesidades y consideraciones son, en el caso, las siguientes:

Si, prescindiendo de consideraciones prácticas inmediatas hubiésemos de pensar en un año dividido en 12 períodos exactamente iguales de días ¿cuántos días corresponderían a cada uno de esos períodos? En el caso de un año no bisiesto, cada período constaría de $365/12 = 30.41667$ días; en el caso de un año bisiesto, cada período estaría formado por $366/12 = 30.5$ días.

Si tomamos como base (igual a 1 o 100 %) la duración de este "mes" teórico o ideal para el cálculo de un número índice, un mes cuya duración fuera exactamente de 30.41667 (o de 30.5 en un año bisiesto) tendría por número índice 1 o 100 %; un mes de más de 30.41667 (o 30.5 según el caso) tendría un número índice superior a 1 o a 100 %; un mes de menos de 30.41667 (o 30.5) tendría un número índice inferior a 1 o a 100 %. O sea, que a ninguno de los meses en que realmente está dividido el año le correspondería un índice de 1 o 100 %, que a enero, marzo, mayo, julio, agosto, octubre y diciembre les correspondería un índice superior a 1 o a 100 % por tener 31 días, que a abril, junio, septiembre y noviembre les correspondería un índice de menos de 1 o 100 % por tener 30 días, y a febrero un índice también menor de 1 o 100 % por tener 28 o 29 días según sea el año bisiesto o no.

Obtenidos estos números índices, será necesario tener en consideración que, en cuanto el índice sea de 100 % la cifra correspondiente no necesitará ajuste calendárico, que, de ser el índice inferior a 100 % esto querrá decir que en el mes considerado hubo menor número de días que el número medio de días por mes y que, por lo mismo, la cifra mensual del fenómeno es menor que la ajustada, o sea, que el ajuste deberá dar una cifra mayor, o que, finalmente, de ser el índice superior a 100 % en el mes considerado habrá habido un número de días menor que el número medio de días y, por lo tanto, que la cifra dada para el mes es mayor que la que deberá obtenerse tras el ajuste; o sea, que, como en el caso anterior, será necesario dividir cada una de las cifras dadas entre el número índice de número de días en el mes con respecto al número medio de días. Es decir, que el ajuste calendárico podría concretarse en la expresión:

$$\textit{Datos ajustados calendáricamente} = \frac{\textit{Datos no ajustados}}{\textit{N}^{\text{o}}\ \textit{índice de días en el mes}}$$

Los números índices de días en el mes con respecto al número medio de días puede calcularse como ya se dijo, dividiendo el número de días que realmente tiene el mes entre el número medio de días por mes en el año correspondiente; de esta forma:

En un año no bisiesto:

Número medio de días por mes 30.41667

Índice para los meses de 31 días .... $\dfrac{31}{30.41667} = 1.017177$

Índice para los meses de 30 días .... $\dfrac{30}{30.41667} = .986304$

Índice para el mes de 28 días .... $\dfrac{28}{30.41667} = 1.920547$

En un año bisiesto:

Número medio de días por mes 30.5

Índice para los meses de 31 días .... $\dfrac{31}{30.5} = 1.016393$

Índice para los meses de 30 días .... $\dfrac{30}{30.5} = 0.983606$

Índice para el mes de 29 días .... $\dfrac{29}{30.5} = 0.950819$

En vista de que dividir por un número equivale a multiplicar por el recíproco de dicho número, los dos ajustes pueden resumirse en forma de procedimientos del modo siguiente:

1.—Ajuste deflactario:
1º—Calcúlense los recíprocos de los números índices de precios correspondientes a cada una de las unidades de tiempo en las que se hayan hecho las observaciones.
2º—Multiplíquense los valores de la serie por ajustar por dichos recíprocos. Los valores resultantes son los que constituyen la serie ya deflectada.

2.—Ajuste calendárico:
1º—Calcúlense los recíprocos de los índices del número de días en el mes con respecto al número medio de días por mes (llámase a tales recíprocos "factores de corrección) *
2º—Multiplíquense los valores no ajustados por variación calendárica por dichos recíprocos o factores de corrección. Los valores obtenidos son los valores ajustados por variación calendárica.

* Estos factores de corrección son los siguientes:

Meses de	*Año ordinario*	*Año bisiesto*
31 días	.981183	.983271
30 días	1.013889	1.016667
29 días	....	1.051724
28 días	1.086310	....

CORRELACIÓN

Cuando al comparar dos o más series estadísticas el examen de sus magnitudes nos permite suponer que existe una relación definida entre las variaciones de magnitud de una de ellas y las variaciones que en el mismo sentido o en sentido contrario experimentan las magnitudes de las otras series, podemos hablar de correlación.

Así, por ejemplo, si se observa que al aumentar la estatura de un individuo (primer fenómeno) aumenta su peso corporal (segundo fenómeno), podremos decir que existe una correlación entre la estatura y el peso corporal; si observamos que en cuanto es menor el ingreso de una familia es mayor la proporción del mismo que se consume en alimentos, podremos afirmar que existe una correlación entre el monto del ingreso y el consumo alimenticio; si en un país, conforme aumenta el número de hospitales disminuye el coeficiente de mortalidad de la población, podremos postular una correlación entre hospitalización y mortalidad; si en el conjunto de los países del mundo, en cuanto aumenta la proporción de individuos dedicados a la agricultura aumenta el coeficiente de mortalidad, podremos señalar una correlación entre dedicación a la agricultura y mortalidad, etc.

La existencia de una correlación matemática entre dos fenómenos (o sea, la posibilidad de predecir el valor de uno de los fenómenos mediante el conocimiento que se tenga del valor correspondiente del otro) no significa necesariamente la existencia de una relación de hecho entre los fenómenos correlacionados. Puede ocurrir, en efecto, que la correlación matemática establecida no sea sino el resultado de la convergencia numérica en las series de sendos conjuntos de causas cuyas ligas postulables no son inmediatas sino mediatas o remotas, vinculándose esos conjuntos en lo que, en términos generales, podría denominarse como "orden universal". Conforme a lo anterior, puede indicarse que el hecho de que exista correlación entre dos o más fenómenos puede carecer de significación científica y que por lo mismo, se puede hablar de:

A.—Una correlación aparente, como ocurre con la relación observada en Suecia entre el aumento en el número de nacimientos y el aumento en el número de cigüeñas, y

B.—Una correlación real.

En caso de que la correlación entre dos o más fenómenos sea real y no aparente, la correlación misma puede interpretarse de muy distintas manera, ya que:

I.—El fenómeno *A* puede ser causa del fenómeno *B* (en caso de correlacionarse dos fenómenos) o ser el fenómeno *A* causa de los fenómenos *B, C, D, E*... (en caso de ser varios los fenómenos entre los cuales se busca la correlación).

II.—El fenómeno *A* puede ser efecto del fenómeno *B* (correlación entre dos fenómenos) o ser el fenómeno *A* efecto del conjunto de los fenómenos *B. C, D, E*..., o bien ser *A, B, C, D*... efectos de *E*.

III.—El fenómeno *A* tanto como el fenómeno *B* pueden ser efectos de una misma causa *C* (en el caso de la correlación entre dos fenómenos), o *A, B, C, D, E*... pueden ser efecto de una misma causa común y distinta de ellos que podemos llamar *X*.

Todo lo anterior muestra que el hecho de encontrar matemáticamente que existe una correlación entre dos o más fenómenos sólo da pistas al investigador para continuar su investigación. Al cómputo de la correlación debe seguir, por lo tanto, un análisis lógico del problema; o sea, que se necesitará examinar si es lógicamente posible hacer una imputación causal entre los dos o más fenómenos correlacionados, y en qué sentido debe hacerse tal imputación.

Cuando se ha establecido el que entre dos o más fenómenos existe una correlación real y no aparente es preciso determinar:

1º—El tipo de correlación.
2º—El grado o intensidad de la correlación.
3º—El sentido de la correlación.

El tipo de correlación se refiere a la determinación de si la correlación es:

a.—rectilínea, o
b.—curvilínea.

Cuando se trata de una correlación rectilínea, a aumentos constantes en el fenómeno *A* corresponden aumentos constantes en el fenómeno *B*. Cuando se trata de una correlación curvilínea, a aumentos constantes del fenómeno *A* corresponden aumentos variables del fenómeno *B*, o bien, a aumentos variables del fenómeno *A* corresponden aumentos constantes del fenómeno *B*.

En el caso de la correlación curvilínea, ésta podrá ser parabólica, logarítmica, hiperbólica, etc. Así, cuando al pasar *A* de 2 a 3 (aumento de una unidad) B pasa de 4 a 9 (aumento de 5 unidades) y cuando A pasa de 3 a 4 (aumento de una unidad) B pasa de 9 a 16 (aumento de 7 unidades), la correlación entre A y B será curvilínea y, más específicamente, parabólica, ya que mientras A crece por unidades, el crecimiento de B se produce conforme a la sucesión de los cuadrados correspondientes.

El grado o intensidad de la correlación nos indica hasta qué grado podemos, a partir de un valor dado de A calcular un valor previsto de B y con qué grado de seguridad admitir dicha previsión. En el caso de la correlación rectilínea, la correlación se mide por un coeficiente de correlación r, cuyo valor no puede ser inferior a 0 ni superior a 1. De obtenerse para r el valor cero, esto indicaría que no existe correlación entre los fenómenos comparados, o que la correlación es nula; dado un valor de A no podremos prever cuál será el valor correspondiente de B científicamente. De obtenerse para r el valor 0.5 esto indicará que hay entre los fenómenos una correlación débil y que, por tanto, si conocemoes un valor de A podremos prever cuál será el correspondiente de B aún cuando con cierto riesgo de equivocarnos, ya que 0.5 está a medio camino entre el cero de la correlación nula, y el 1 de la correlación perfecta. En caso de obtenerse como valor de r 1, podrá hablarse de correlación perfecta: dado un valor de A podemos prever cuál será el correspondiente valor de B.

El sentido de la correlación nos permite distinguir entre:

a.—una correlación directa, y

b.—una correlación inversa.

Correlación directa es aquella en que las variaciones de un fenómeno se producen en el mismo sentido que las variaciones de otro fenómeno. Correlación inversa es aquella en que las variaciones del primero se producen en sentido contrario a las variaciones del segundo.

O sea, que se habla de correlación directa cuando a aumentos en la magnitud de A corresponden aumentos del fenómeno B y a disminuciones del fenómeno A corresponden disminuciones de B; en cambio se habla de correlación inversa cuando a aumentos del fenómeno A corresponden disminuciones del fenómeno B, y a disminuciones del fenómeno A corresponden aumentos de B.

EL COEFICIENTE DE CORRELACIÓN RECTILÍNEA COMO COMPARACIÓN ENTRE MEDIAS DE VARIABILIDAD

El coeficiente de correlación rectilínea puede considerarse fundamentalmente como el resultado que se obtiene de comparar matemáticamente la bondad de las estimaciones hechas cuando se toma como base para las mismas la media aritmética de uno de los fenómenos por correlacionar con las que se hacen tomando como base la ecuación de regresión de dicho fenómeno con respecto al otro con el cual se relaciona o, dicho en otra forma, que el coeficiente de correlación rectilínea puede concebirse como el resultado que se obtiene de comparar la variabilidad de las observaciones de un fenómeno con respecto a su media aritmética con la variabilidad de dichas observaciones con respecto a los valores obtenibles mediante la interpolación de una recta a los valores de dicho fenómeno conjuntado con los del otro fenómeno con el cual quiere relacionársele.

La variabilidad de un fenómeno con respecto a su media aritmética puede medirse mediante el cálculo del cuadrado de la desviación media cuadrática. La variabilidad de un fenómeno con respecto a la línea de regresión del mismo en relación con el otro fenómeno puede medirse al través del cálculo del cuadrado del error de estimación. La comparación entre ambas medidas de variabilidad puede hacerse a base de una relación matemática o cálculo del cociente de una con respecto a la otra medida de variabilidad, o sea, en la práctica, dividiendo el cuadrado del error de estimación entre el cuadrado de la desviación cuadrática media.

Como las variaciones del cociente que resulta de dividir el cuadrado del error de estimación entre el cuadrado de la desviación cuadrática media se producen en sentido contrario al seguido por las variaciones de magnitud de la correlación entre las variables, conviene introducir alguna modificación que haga que la medida de correlación sufra variaciones en el mismo sentido que ésta.

En efecto, cuando la correlación entre los fenómenos es máxima, la variabilidad en torno de la línea de regresión es mínima y, consecuentemente el error de estimación y su cuadrado son nulos; o sea, que la relación entre el cuadrado del error de estimación y el cuadrado de la desviación cuadrática media es también nula, dándose por tanto el caso de que, de utilizarse sin modificación esta medida de la correlación, al ser máxima la correlación sería mínimo el valor de su medida y máximo al ser la correlación mínima. Con el objeto de modificar esta situación, se acostumbra tomar como medida de la correlación la desviación del cociente del error de estimación y la desviación cuadrática media, con respecto a la unidad; en esta forma, cuando dicho cociente es nulo por no haber variabilidad en torno de la línea de dispersión, la desviación de dicho cociente con respecto a la unidad es uno o sea, que al ser máxima la correlación es asimismo máxima (igual a 1) su medida; conforme aumenta la dispersión con respecto a la línea de regresión, aumentará el valor del cociente entre el error de estimación elevado al cuadrado y el cuadrado de la desviación cuadrática media; cuando el error de estimación alcance su valor máximo (igual a la desviación cuadrática media), el valor del cociente será asimismo máximo e igual a uno y, consecuentemente, la desviación de dicho cociente con respecto a la unidad será nulo; o sea, que al ser mínima la correlación, su medida alcanzará un valor también mínimo (igual a cero).

A fin de reducir el valor de la medida de la correlación a unidades lineales se acostumbra tomar la raíz cuadrada de la desviación del cociente entre las medidas de variabilidad con respecto a la unidad.

Según lo dicho, la fórmula para el coeficiente de correlación rectilínea entre dos fenómenos x_1, x_2 sería:

$$r_{12} = \sqrt{1 - \frac{s^2_{12}}{\sigma^2_1}}$$

Conforme a todo lo anterior, el procedimiento para obtener el coeficiente de correlación rectilínea constará de los siguientes pasos:

1º—Obtener el cuadrado del error de estimación, para lo cual:

A.—Adáptese a los datos consignados para los dos fenómenos, una línea recta.

B.—Con base en la ecuación de la recta en la que se hayan substituido los valores de los parámetros encontrados mediante la adaptación, calcúlense los valores que teóricamente serían de esperar para uno de los dos fenómenos dado cada uno de los valores del otro fenómeno (mediante substitución de estos últimos en la ecuación de regresión).

C.—Obténganse las desviaciones de los valores teóricos así encontrados con respecto a los valores reales del fenómeno.

D.—Elévense al cuadrado dichas desviaciones.

E.—Súmense dichos cuadrados.

F.—Divídase dicha suma entre el número de observaciones a fin de obtener el cuadrado del error de estimación.

El resultado es s^2_{12}

2º—Obtener el cuadrado de la desviación cuadrática media. Para ello:

A.—Obténgase la media aritmética de los valores del fenómeno cuyas magnitudes teóricas se calcularon mediante la línea de regresión.

B.—Calcúlense las desviaciones de los valores reales de dicho fenómeno con respecto a su media aritmética restando de cada uno de ellos la media aritmética obtenida.

C.—Elévense al cuadrado dichas desviaciones.

D.—Súmense dichos cuadrados.

E.—Divídase dicha suma entre el número de observaciones, a fin de obtener el cuadrado de la desviación cuadrática media.

El resultado es σ_1^2

3º—Calcular el cociente del valor obtenido en 1 entre el valor obtenido en 2.

4º—Restar de uno dicho cociente.

5º—Extraer la raíz cuadrada del resultado a fin de obtener el coeficiente de correlación rectilínea.

CORRELACIÓN EN SERIES TEMPORALES

A fin de correlacionar dos series dinámicas, temporales o cronológicas, pueden seguirse varios procedimientos, todos los cuales tienen por objeto eliminar la exageración del índice de correlación producida por la semejanza de las líneas de tendencia.

Primer Procedimiento

1º—Ajústese a cada una de las series por correlacionar su correspondiente línea de tendencia.

2º—Calcúlense los valores teóricos de la tendencia para cada una de las series que se tratan de correlacionar.

3º—Determínense las desviaciones de los valores reales de cada serie con respecto a los valores teóricos correspondientes de la misma serie.

4º—Determínense la curva de regresión y los índices de correlación (o, en su caso, la recta de regresión y el coeficiente de correlación) de las desviaciones de las series que se relacionan.

Segundo Procedimiento

1º—Calcúlense las diferencias de cada observación en cada una de las series con respecto a la observación anterior.

2º—Determínense la curva de regresión y los índices de correlación de dichas diferencias.

Tercer Procedimiento

1º—Ajústense las curvas de tendencia correspondientes a cada una de las series.

2º—Calcúlense los valores teóricos de la tendencia para cada una de las series.

3º—Exprésense los valores reales de cada serie como tantos por ciento de los valores teóricos correspondientes.

4º—Búsquese la correlación entre tales tantos por ciento.

III.—*DOS SECTORES DE ESTUDIO ESTADÍSTICO*

CENSOS DE POBLACIÓN

La estadística social mundial se enfrenta a varios problemas, entre los que tienen singular importancia:

1.—el problema de la falta de informes, y
2.—el problema de la falta de comparabilidad de los datos de que se dispone.

La falta de informes constituye un problema tanto en el campo de la estadística demográfica general como en el de la más especializada, pero el mismo se agudiza especialmente en esta última.

En efecto, en 1948, sólo 147 de las 245 áreas identificables en el *Demographic Yearbook* de las Naciones Unidas contaban con datos al través de los cuales seguir el desarrollo de su población desde 1900; de las 98 restantes algunas no habían realizado sino un solo censo de población desde esa fecha. En campos más especializados como el de la distribución de la población por edad y sexo, ocupación, vida rural o urbana, etc., la situación era más crítica, ya que sólo 41 de las 245 áreas tenían estadísticas de población por edad, sexo y estado matrimonial.

Al problema de la falta de información hay que agregar el de la falta de comparabilidad entre los datos de los que se dispone, ya que esto imposibilita su utilización conveniente por el investigador social.

El reconocimiento de ambos problemas no es, sin embargo, reciente, ya que en 1872, el International Statistical Institute, en su reunión de San Petesburgo señalaba la necesidad de utilizar ciertos métodos uniformes, de interesarse por unos mismos problemas, de emplear unas mismas definiciones en la obtención de los datos demográficos a fin de hacerlos comparables. En una sesión posterior (1897) del propio Instituto, el húngaro Körösi propuso la realización simultánea de un censo por todos los países en el año de 1900 y su repetición periódica. Resoluciones posteriores, no siempre acatadas por todos los países, señalaron la conveniencia de realizar los censos de población cada década, en los años terminados en cero.

Recientemente, las Naciones Unidas, al ocuparse de estos problemas, solicitaron de los diferentes gobiernos el envío de informes acerca de métodos empleados y tópicos estudiados en su último censo de población; con base en estos informes, se formó una lista de tópicos investigados por todos los países en sus censos; sin embargo, dicha lista no podía adoptarse como un mínimo de información requerida de todos los países, ya que no todos ellos cuentan con las mismas posibilidades de investigación estadística ni son iguales las necesidades a que todos ellos se enfrentan; de ahí la necesidad de seleccionar ciertos tópicos especialmente importantes, y de los que resultaba deseable obtener datos precisos de todos los países.

Dicha lista incluye:

1.—Población total,
2.—Sexo,
3.—Edad,
4.—Estado matrimonial,
5.—Lugar de nacimiento,
6.—Nacionalidad legal,
7.—Lengua materna,
8.—Características educativas,
9.—Fertilidad,
10.—Características económicas.
  a.—Población económicamente activa e inactiva.
  b.—Ocupación, industria, estatus industrial.
  c.—Población dependiente.
  d.—Población agrícola.
11.—Distribución urbano-rural,
12.—Familia y hogar (incluyendo relación con el jefe de familia).

Como para algunos países aún esta lista pudiera resultar excesiva, las propias Naciones Unidas subrayaron que en caso de países que tuvieran poca o ninguna experiencia censal, sería deseable se concretaran a obtener los datos relativos a población total, sexo, edad, estado matrimonial y características económicas, pero que ya fuera que se redujeran a esta lista más corta o tomaran la más extensa, proporcionaran datos *completos* y *confiables*.

En estas sugestiones y recomendaciones para obtener datos completos y comparables acerca de la población mundial las Naciones Unidas tuvieron muy especialmente en cuenta las experiencias que por esas mismas fechas recogían los países americanos al planear y realizar el Censo de las Américas (1950), así como las recomendaciones del Comité de Estadígrafos de la Liga de las Naciones (1937-1939), y de las VI y VII Conferencias Internacionales de Estadígrafos del Trabajo (1947-1949).

Sin embargo, es necesario tener en cuenta que en los censos mundiales de población, para los propósitos de comparabilidad internacional, importan no sólo los datos mismos, sino la forma de recolectarlos, de definirlos, clasificarlos y tabularlos; de ahí la preocupación mostrada por las comisiones estadísticas de la ONU respecto de los problemas metodológicos.

Población total.—Con respecto a las estadísticas de población total, la falta de comparabilidad entre los datos de diversos países procede de:

a.—Diferencias en el modo de definir la población total,
b.—Sub- y sobre-enumeración.

Entre los modos de definir la población total, se pueden reconocer dos tipos:

definición *de facto,* y
definición *de jure.*

Según la definición *de facto,* integran la población total de un lugar todas las personas que se encuentran físicamente en él en el momento del censo.

Según la definición *de jure,* forman parte de la población total de un lugar aquellas personas y sólo aquellas que residen habitualmente en él, sea cual fuere el lugar en el que se encuentren en el momento del censo.

La definición de facto tiene la ventaja de la simplicidad y la objetividad; simplicidad, pues censadores y censados no tienen que hacer otra cosa que contar a los presentes en el momento del censo; objetividad, porque no presenta casos dudosos que deba resolver el censador según su criterio. La definición de jure implica instrucciones más complicadas, diferencias en la definición de "domicilio" por los diversos países, problema de personas que no tienen residencia fija; sin embargo, la definición de jure puede dar mejores resultados cuando la enumeración censal debe hacerse en un lapso relativamente grande de tiempo y, muy especialmente, cuando se hace con poblaciones de una gran movilidad.

El problema de la diferencia entre las definiciones de facto y de jure se manifiesta en el campo nacional o local en cuanto se toman en cuenta ciertas necesidades; en efecto:

Una definición *de jure* será más útil en la planeación de servicios sociales tales como los programas de alojamiento o de educación de la población residente. Una definición *de facto* será más útil en los programas de protección policíaca, frente a las necesidades del transporte, etc.

En el campo internacional dichas diferencias pueden provocar dificultades en aquellos países que tienen fuertes núcleos de población radicada en el extranjero en calidad de fuerzas de ocupación o de aquellos otros que atraen fuertes núcleos de turistas, etc.

De 53 países consignados en el Anuario de las Naciones Unidas,

31 usaron ambos tipos de definición,

11 emplearon definiciones *de jure*, y

De entre los que usaron ambos tipos de definición, se pueden hacer dos grupos:

a.—el de los que enumeraron sobre una base de facto recogiendo en la cédula no sólo el lugar ocupado físicamente por la persona, sino también el de su residencia habitual, y

b.—el de los que, en una misma cédula, pero en renglones separados, consignaron en cada lugar: las personas que residen usualmente en él pero están temporalmente ausentes, y las personas que residen usualmente en otro lugar pero están temporalmente presentes.

A este último grupo, que permite un control mayor corresponden principalmente los censos de Dinamarca, Noruega y Polonia.

En el caso de las definiciones *de facto,* algunos países excluyeron de la enumeración las fuerzas armadas extranjeras localizadas en su territorio; algunos incluyeron y otros no a sus propias fuerzas armadas que se encontraban en el extranjero; algunos excluyeron asimismo a los civiles acompañantes de las fuerzas armadas extranjeras; otros excluyeron a los miembros del servicio diplomático extranjero (incluyendo unos, y otros no, a los miembros de su propio servicio diplomático en el extranjero); en las regiones marítimas, se incluyó a los marineros cuyos barcos estaban anclados en un puerto nacional, a quienes viajaban en barcos que navegaban entre puertos nacionales, etcétera.

En el caso de las definiciones *de jure,* las estadísticas de los países que la utilizan hablan de "población residente", o "población domiciliada".

Debido a la diversidad de aplicaciones que tienen ambos tipos de definición de la población total resulta conveniente el uso simultáneo de ambos. Sin embargo, en todo caso, deben tenerse en cuenta las recomendaciones hechas por los organismos internacionales.

1º—Se recomienda la obtención de datos sobre el total de la población en el momento del censo.

2º—Debe excluirse al personal militar y diplomático extranjero radicado en el país, e incluirse al personal militar y diplomático propio radicado en el extranjero.

3º—Los grupos que tengan núcleos importantes de habitantes a los que no se pueda enumerar individualmente (por ejemplo, grupos que vivan fuera de la estructura socio-económica del país) deberán estimar su número y apreciar sus características por los mejores medios disponibles, y, en caso de imposibilidad, indicar en los informes censales que los datos del total de población no incluyen estos grupos.

La segunda gran dificultad que enfrentan los censos de población total son los de sub- y sobre-enumeración que dependen, entre otros, de los siguientes factores:

1.—De la tradición estadística del país,
2.—De la planeación y organización del censo,
3.—Del entrenamiento, experiencia y honradez de los censadores,
4.—De los recursos económicos y técnicos de que se dispone,
5.—Del nivel educativo general y especial de la población censada,
6.—De las facilidades de transporte y comunicación, etc.

Todas estas condiciones provocan sub- o sobre-enumeraciones; presentándose las sub-enumeraciones principalmente en el caso:

a.—de núcleos aislados física, social o culturalmente (habitantes de lugares de difícil acceso, personas proscritas socialmente, individuos que no hablan el idioma principal del país).
b.—de personas o núcleos móviles (trabajadores migratorios, habitantes de hoteles, etc.).
c.—de ciertos núcleos de edad y sexo: niños, hombres jóvenes, muchachas solteras.

Sin embargo, las cifras censales no pueden tomarse sin crítica, y aceptarse sin más los peligros de la sub- o la sobre-enumeración sino que debe planearse una verificación que se lleve a cabo inmediatamente después del censo por medio de un muestreo científico del área censada.

SEXO Y EDAD

La clasificación por edad y sexo de los habitantes de un país es muy importante en los censos de población, ya que las informaciones acerca de la estructura de una población de acuerdo con las edades de sus componentes sirve:

1.—con propósitos de política demográfica para prever los cambios en la composición de la población,
2.—con vistas a la política sanitaria del estado habida cuenta de las tasas de morbilidad y mortalidad,
3.—en relación con fines comerciales y cálculos actuariales de las compañías de seguros (determinación de probabilidades de supervivencia, etc.),
4.—con respecto al suministro de fuerza de trabajo y de contingentes militares,
5.—en relación con los problemas de la dependencia social infantil y senil, etc.

La clasificación sexual de la población es sencilla, en cambio, la clasificación por edades presenta ciertas dificultades, entre las cuales se cuentan las derivadas de una mala información que puede depender de:

a.—ignorancia de la edad correcta o descuido,
b.—preferencia por ciertos números (especialmente las cifras llamadas "redondas" en cero, cinco o en números pares) y la aversión por otros guarismos (especialmente por los nones y, entre ellos, por el 7 y el 13 muy particularmente),
c.—motivos económicos, sociales, políticos, etc. (como temor a la conscripción o servicio militar, etc.),
d.—la existencia de culturas que viven fuera del mundo numérico (en Oceanía, Asia y África principalmente) y que distinguen burdamente entre "niños" y "adultos", etc.

La definición de la edad en los censos tiene un cierto campo de variación, ya que en tanto hay países que toman en cuenta la edad (generalmente en años simplemente) del interrogado en el cumpleaños inmediatamente anterior, en otros casos se considera el cumpleaños más inmediato posterior, en otros el más inmediato (sea anterior o posterior) y, en algunos otros casos hay que tener en consideración que, socialmente, la edad se cuenta no a partir del momento del nacimiento, sino que el punto de partida de la misma se retrotrae a un año antes del nacimiento.

Se usan generalmente dos métodos para obtener información respecto de la edad:

1.—se pregunta la fecha de nacimiento, o
2.—se pregunta la edad.

En el primer caso, se evitan las dificultades derivadas de la diferencia social en la definición de la edad; en cambio, el segundo procedimiento es más sencillo. El primer procedimiento es típicamente europeo; el segundo, típicamente americano. Al primer procedimiento se le han encontrado ventajas *no* estadísticas en los países en que los datos censales se emplean para fines judiciales o de otro tipo, en cambio, tiene el inconveniente de requerir no sólo la recolección de las fechas de nacimiento, sino el ulterior cómputo de edad.

En algunos censos, además de la edad de los adultos y de los niños mayores de un año en años, se registra la edad de los niños menores de un año en meses completos, y la de los niños menores de un mes en días.

La Comisión de Población de las Naciones Unidas propuso, en relación con la clasificación de la población por edades:

1.—Tomar la edad del último cumpleaños.
2.—Tabular edades para cada sexo separadamente.
3.—Hacer, por lo menos, los siguientes grupos de edades:
  a.—menores de un año,
  b.—de 1 a 4 años,
  c.—grupos de 5 en 5 años hasta el fin de la vida.

El Comité para el Censo de las Américas, en lugar de considerar grupos de 5 en 5 años hasta el fin de la vida, limita dichos grupos hasta los 84 años, formando un último grupo con los individuos de 84 años o más.

En la práctica ha sido frecuente formar grupos de 10 en 10 años, según dos sistemas:

1.—el sistema que forma grupos con edades cuyo límite inferior es un número terminado en cero (10-9, 20-29, etc.), o
2.—el sistema que agrupa las edades de 10 en 10 comenzando con un número cuyo último guarismo sea 5 (15-24, 25-34, etc.).

A fin de permitir la comparación o comparabilidad de los datos recogidos se han recomendado las clasificaciones de 5 en 5, ya que esto permite:

La conversión entre los diversos países, y
La conversión entre los datos que acerca de la estructura de su población según edad que obtenga cada uno de ellos en el futuro, y los que a ese mismo respecto obtuvo en el pasado conforme a otros sistemas.

Al lado de los grupos con intervalos de 5 o 10 años, importan otros grupos de edades, particularmente, los definidos por los siguientes criterios:

1.—Edades económicamente productivas, frente a edades económicamente dependientes.

2.—Edades demográficamente reproductivas, frente a las que no lo son. (Esto implica un criterio mixto de edad y sexo que fija atención preferente en las mujeres).
3.—Edades políticamente significativas (votantes) frente a las que no lo son.
4.—Edades militarmente importantes.
5.—Edades escolares.

Todo esto, sin embargo, representa en general, la elección más o menos arbitraria de límites para cada grupo, o una elección de límites que tenga en cuenta las condiciones de estructura económica, demográfica, sanitaria, legal, etc., del país del que se trata.

ESTADO MATRIMONIAL

El estado matrimonial permite clasificaciones de la población que importan desde los ángulos:

1.—demográfico,
2.—médico,
3.—económico,
4.—sociológico.

Demográficamente, la clasificación estadístico-censal de acuerdo con el estado matrimonial importa por cuanto la soltería, el divorcio o la viudez de capas más o menos extensas de la porción reproductiva de la población constituyen algunos de los factores que pueden explicar o que permiten prever cambios en la cuantía y estructura de la misma.:

Médica, asistencialmente y desde el punto de vista de la seguridad social dicha clasificación es especialmente importante en relación con el establecimiento de servicios de maternidad, etc.

Económicamente, dicha clasificación permite prever ciertas características de la demanda de artículos para el hogar y otros bienes semejantes, y, desde otro punto de vista, las posibilidades de suministro de fuerza de trabajo para la industria.

Socialmente, orienta al investigador a interesarse en los problemas que implica la soltería, la viudez y el divorcio. En los países en que existe la poligamia legal, los datos obtenidos en esta forma ofrecen la posibilidad de estudiar los caracteres que en la realidad tiene dicha costumbre.

La Comisión de Población de las Naciones Unidas recomendó el que en la clasificación por estado matrimonial se hicieran las siguientes agrupaciones:

1.—casados, con inclusión de las personas que viven en uniones estables *de facto,*
2.—viudos no vueltos a casar,
3.—divorciados no vueltos a casar,
4.—solteros.

La comisión señaló como deseable el que en la primera categoría se hiciera una distinción entre:

a.—casados *de jure* (o legalmente),
b.—unidos *de facto* (o en unión real no sancionada por la ley).

La comisión no señaló pero puede agregarse que en algunos países sería conveniente hacer una distinción más fina que tuviera en cuenta la unión o falta de unión eclesiástica, con lo cual vendrían a constituirse las cuatro categorías siguientes:

casados *de jure,* casados también eclesiásticamente,
casados *de jure,* no unidos eclesiásticamente,
unidos eclesiásticamente (en unión *de facto,* según el concepto legal y estadístico),
unidos *de facto* sin unión eclesiástica.

La propia comisión señaló la conveniencia de clasificar en categorías separadas a

a.—divorciados legalmente,
b.—separados de sus cónyuges sin mediación de divorcio.

El que estas categorías que parecen tan obvias no lo son tanto se demuestra por el hecho de que algunas naciones subsumen en una misma categoría solteros, viudos y divorciados (así como separados, ocurriendo esto generalmente en países en donde no se reconoce legalmente el divorcio).

La separación en una categoría especial de las personas que viven en uniones matrimoniales *de facto* es especialmente importante en los países latinoamericanos; de otra parte, permite le flexibilidad de los datos en la adaptación a diferentes tipos de análisis estadístico; así, desde el punto de vista legal, las personas incluídas en esa categoría se asimilarán a los solteros, en tanto que en un análisis demográfico (de tasas de reproductividad, etc.), en uno económico (que estudie demanda de artículos domésticos, por ejemplo), etc., se asimilarán a los casados.

En forma análoga, la formación de una categoría especial con las personas "separadas" que no lo están legalmente permite distinguirlas en análisis relativos a la reproductividad, demanda de artículos domésticos, servicios de maternidad, etc.

Además de señalar las categorías matrimoniales, según las recomendaciones de la Comisión de Población, debe agregarse una categoría de "estado matrimonial no determinado", y hacer las especificaciones en cada categoría para los distintos grupos de edad señalados en el capítulo correspondiente.

Las principales deficiencias notadas en los censos de acuerdo con las normas de la Comisión, fueron:

1.—falta de totales para las diferentes categorías matrimoniales de individuos menores de 15 años (deficiencia importante en particular en países en los que se acostumbra el matrimonio en edad muy temprana),
2.—falta de agrupación conveniente de las edades (falla en la que incurrieron muy particularmente Guatemala y México, en el que "el agrupamiento por edades no fue lo suficientemente fino como para formar ni siquiera grupos de 10 en 10 años").

Otros aspectos importantes en la clasificación del estado matrimonial se refieren a:

1.—Edad en el momento del matrimonio.
2.—Duración del matrimonio (para los casados).
3.—Duración de la viudez o el carácter de divorciados (para los viudos o divorciados en el momento del censo).
4.—Estado matrimonial previo al actual (con propósitos de estudio de la fertilidad, y de los índices de probabilidad de re-casamiento).
5.—Número de veces que se ha estado casado.
6.—Monogamia o poligamia (en los países en que la poligamia o, más propiamente poliginia o poliandria, están legalmente reconocidas), etc.

FERTILIDAD

Los datos estadísticos sobre fertilidad se refieren a la frecuencia de los nacimientos en una población dada. Importan particularmente desde los ángulos:

1.—demográfico,

2.—biológico,
3.—sociológico.

Desde el punto de vista demográfico, los datos sobre fertilidad permiten prever el posible desarrollo de las poblaciones así como su posible composición futura por grupos de edad, etc., con todo lo que esto implica para las transformaciones en la economía y en la mentalidad del grupo.

En el campo de la biología, dichos datos son especialmente útiles para los estudios de genética humana, de antropología y de medicina.

Sociológicamente, la importancia de las cifras sobre fertilidad destaca cuanto se considera que, al través del estudio de la fertilidad diferencial de los diversos grupos sociales (grupos étnicos, clases sociales, grupos religiosos, etc.), se pueden predecir ciertos cambios en la estructuración social del grupo, en el predominio de ciertos modos de pensar y de actuar dentro del mismo y, consiguientemente, puede preverse cuál será el proceso mismo del llamado "cambio social".

Las cifras de fertilidad se obtienen usando combinadamente dos fuentes:

1.—los censos de población, y
2.—el registro de nacimientos.

A partir de los datos consignados en estas fuentes, se obtienen los índices de fertilidad mediante una relación entre el número de nacimientos ocurridos en un grupo y el número de personas enumeradas en el censo de ese mismo grupo (en busca de una mayor precisión, dicha relación deberá obtenerse no con respecto al total de la población, sino en referencia a la porción reproductiva de la misma).

Cuando faltan los registros de nacimientos, se recurre a varios procedimientos para obtener cifras de fertilidad a partir de los censos, entre las cuales se cuenta:

La obtención de relaciones matemáticas entre el número de niños enumerados en el censo, y el número de adultos en edad reproductiva.

En otros casos, se prevé dicha obtención al planear el censo, registrándose en las cédulas correspondientes el número de niños y el número de adultos o parejas de adultos casados existentes en cada casa calculando la relación matemática entre ambas cifras.

O sea, que pueden reconocerse dos formas de obtención de datos sobre fertilidad:

1.—Uso de tabulaciones especiales basadas en el análisis de la composición hogareña.

2.—Uso de datos acerca del número de hijos nacidos.

En el uso del análisis de la composición hogareña, se parte de la inspección del número de niños menores de 5 años enumerados en cada hogar, puestos en relación con los adultos que se presume sean sus padres con base en el estudio de las relaciones familiares que la cédula recoge. Las tabulaciones se hacen:

a.—por grupos de edad y sexo de los padres,
b.—por estado matrimonial de los propios padres,
c.—por ocupación,
lengua,
pertenencia a determinado grupo étnico,
lugar de nacimiento,
ciudadanía,
educación, etc.

Las tabulaciones deben consignar el número total de adultos correspondientes a cada grupo de los antes señalados, y el número total de sus hijos.

Con respecto al segundo procedimiento, o sea el que utiliza el interrogatorio acerca del número de hijos nacidos, es necesario consignar:

a.—número de mujeres que informaron acerca del número de hijos que habían tenido (incluídas las que informaron no haber tenido),
b.—número de hijos reportados,
c.—número de mujeres que reportaron 0, 1, 2, 3, 4, 5 o más hijos, todo ello clasificado por edades de la madre, estado matrimonial, etc.

Las tabulaciones especiales sobre fertilidad tienen la ventaja sobre las relaciones basadas en datos proporcionados por los censos de población y registros de nacimientos, de que se evitan los problemas que surgen del hecho de que el padre haya cambiado de ocupación entre la fecha de registro de su hijo y la fecha del censo, ya que dichos problemas dificultan el estudio de la fertilidad diferencial de los varios grupos ocupacionales, de ingreso, etc.

Otro de los problemas que se presenta en los censos actualmente existentes que consignan datos utilizables para el estudio de la fertilidad de la población estriba en que en ocasiones se descuida la posibilidad de estudiar y comparar la fertilidad diferencial de la porción masculina y de la porción femenina de la población, lo cual es particularmente interesante en cuanto éstas se relacionan con la posición ocupacional de los padres (ya que algunas ocupaciones son desempeñadas más frecuentemente por las mujeres y otras por los hombres), o con las zonas de migración (a las que llegan más generalmente hombres que mujeres en edad reproductiva), etc.

Además de esto, se presentan problemas especiales en cuanto se trata de la enumeración de los hijos de las madres solteras que es conveniente tabular separadamente de los que lo son de personas casadas, ya que representan un problema social importante.

LUGAR DE NACIMIENTO

Con objeto de determinar la composición étnica de la población, los censos consignan generalmente datos sobre:

1.—Lugar de nacimiento,
2.—Nacionalidad legal, y
3.—Lengua.

Además se suelen consignar datos relativos a "raza" y "religión"; sin embargo, estos últimos resultan más difíciles de obtener y por lo mismo son menos utilizados.

De los tres tipos de datos arriba mencionados, el del lugar de nacimiento no resulta suficiente para la determinación de la pertenencia étnica de un individuo a un grupo, ya que hay muchas ocasiones en que ciertos países han recibido fuertes núcleos de inmigrantes extranjeros y los hijos o nietos de los miembros de dichos núcleos demográficos no obstante haber nacido en el país, conservan muchos de los rasgos culturales de sus progenitores. De otra parte, esos mismos datos no llegan a reflejar el grado de asimilación del nacido en el extranjero que, no obstante tener por lugar de nacimiento uno distinto del del país en que está radicado, se encuentra asimilado por completo a la población nativa.

Los datos sobre el lenguaje resultan más útiles al respecto desde varios puntos de vista:

1.—el lenguaje es un criterio más sensible por ir generalmente acompañadas las diferencias lingüísticas de otras importantes diferencias culturales,
2.—el conocimiento del lenguaje es conocimiento casi indispensable para la asimilación de los inmigrantes, de tal modo que quien no posee la lengua nacional puede juzgarse como no asimilado a la población nativa,

3.—el lenguaje proporciona criterios más flexibles que el lugar de nacimiento, etc.

Las cuestiones relativas a lugar de nacimiento y a lenguaje permiten reconocer no sólo los grupos de inmigrantes, sino también los diferentes grupos étnicos integrantes de la población nativa y su procedencia.

Los datos relativos a la nacionalidad son menos útiles a los propósitos antes señalados, ya que las posibilidades de naturalización de los extranjeros les hacen valiosos únicamente como datos utilizables en investigaciones legales.

La clasificación de la población por lugar de nacimiento permite apreciar:

1.—volumen, y
2.—fuente de la inmigración.

Esto puede referirse tanto a la inmigración propiamente dicha como a las migraciones internas entre diversos lugares de una misma área geográfica.

Sin embargo, la apreciación de las migraciones que se haga sobre la base de las respuestas al interrogatorio sobre el lugar de nacimiento de los censados es muy burda, ya que:

a.—no toma en cuenta las muertes de inmigrantes ocurridas entre dos censos,
b.—no toma en consideración el regreso de ciertos inmigrantes a sus lugares de origen en el lapso que media entre dos censos.

Para obtener un ligero refinamiento de los datos al través de los cuales se pueda apreciar el fenómeno de las migraciones tanto internas como externas, se puede recurrir a:

a.—una clasificación de los inmigrantes por edad y sexo,
b.—una clasifiación de los propios inmigrantes que tenga en cuenta no sólo el lugar de partida, sino la fecha de la misma, y la fecha de arribo al lugar de la inmigración.

Un estudio de los problemas migratorios y de la fecundidad diferencial entre los varios grupos étnicos componentes de un país obliga a la obtención de datos complementarios y separados de los del total de la población de:

1.—número de mujeres en edad de procrear dentro de la población inmigrante,
2.—número de hijos de las mujeres inmigrantes de preferencia clasificados según que hayan nacido o no en el país de la inmigración.

En la clasificación que cada país haga de sus inmigrantes por lugar de procedencia, deberá tener en cuenta los límites geográficos de las diversas áreas políticas *en el momento del censo,* para lo cual deberá tener a la vista el registro más reciente de los países y territorios no autónomos, publicado por las Naciones Unidas.

En caso de ser pequeñas las cifras de inmigración y haber necesidad de combinar las correspondientes a varios países se ha recomendado por la Comisión de Población de las Naciones Unidas que dichas combinaciones se hagan dentro de los países de un mismo continente, a fin de posibilitar los estudios de migraciones intercontinentales.

Con fines de estudio de las migraciones internacionales, es necesario definir si la información referente a lugar de nacimiento tiene en cuenta la localidad en que residían los padres en el momento de ocurrir éste, o aquélla en la que realmente ocurrió, especialmente si se tiene en cuenta que en los últimos decenios y en muchos países se recurre a los servicios de maternidades no localizadas en el lugar de residencia de los padres. En este caso de las migraciones internas debe subrayarse que cuanto más específico sea el dato (localidad más que gran división geográfica ya se trate de "provincia", "estado", etc.), resultará más útil en las investigaciones. Sin embargo, la especificidad no debe hacer olvidar la precaución de que el dato sea unívoco (evitar, que por haber en el país dos o más lugares con un mismo nombre no pueda adscribirse al inmigrante a una determinada división territorial dentro del mismo Estado o área política).

Con respecto a la población nativa, la mayor especificidad del lugar de nacimiento que se reporte tiene varias ventajas, especialmente.

1.—cuando los límites de los Estados, a causa de guerras internacionales, reclamaciones, etc., son muy fluctuantes,

2.—cuando, dentro de un mismo país, existen diferencias étnicas o culturales importantes dentro de las diversas zonas que lo constituyen.

Con fines de comparabilidad internacional, las Naciones Unidas han preparado un agrupamiento de países por zonas diversas. Dicho agrupamiento puede consultarse en:

United Nations Statistical Office: *Nomenclature of Geographic Areas for Statistical Purposes.* Statistical Papers. Series M. Nº 1.1. Enero de 1949.

NACIONALIDAD

Bajo el rubro "nacionalidad" los diversos países estudian y consignan datos estadísticos acerca de las cosas más diversas, entre las que muchas veces se cuentan el lugar de nacimiento y el lenguaje, así como los grupos formados por individuos con antepasados o con costumbres comunes, etc. Sin embargo, en la mayoría de ellos, los datos se refieren a la nacionalidad legal, única de la que nos ocuparemos.

Los datos sobre nacionalidad legal se utilizan principalmente para estudiar:

1.—problemas de estatuto legal, y
2.—problemas de derechos civiles de los inmigrantes.

Las recomendaciones hechas por las comisiones internacionales con respecto a los problemas de la nacionalidad legal, señalan la necesidad de distinguir:

a.—quiénes son nacionales del Estado ya sea por nacimiento, naturalización u otro concepto, y
b.—quiénes son extranjeros, debiendo clasificarse a estos últimos —de ser posible— según el Estado del cual son nacionales.

Tanto nacionales como extranjeros deben clasificarse asimismo por edad y sexo, formándose grupos de:

los menores de 5 años,
de 5 a 14 años,
de 14 años en adelante por intervalos de 10 años hasta 64 años,
de 64 años o más.

En los Estados en los que las tasas de inmigración sean altas se considera conveniente el que se clasifiquen además tanto a los nacionales como a los extranjeros,

según estado matrimonial,
conforme a sus características socio-económicas,
de acuerdo con su residencia urbana o rural,
teniendo en cuenta su nivel educativo,
habida cuenta de la ciudadanía del cónyuge, etc.

Las comisiones internacionales consideran igualmente útil el que, en aquellos casos en que, como en el del Imperio Británico, los Estados están formados por zonas muy diversas, se clasifiquen sus súbditos según la zona de los mismos a la que pertenezcan.

Frente a estos *desiderata,* los censos realizados hasta ahora muestran la siguiente realidad:

algunos países incluyen en sus cédulas la distinción entre nacionales y extranjeros, pero no la clasificación según el Estado del que estos últimos son nacionales (Argentina, Chile, Estados Unidos),

algunos otros países, no obstante incluir en sus cédulas, a más de la distinción entre nacionales y extranjeros el Estado del cual éstos son nacionales, no vierten las cifras obtenidas en los resultados censales (Cuba, Inglaterra y Gales),

algunos países distinguen entre nacionales por nacimiento y por naturalización (Chile),

algunos otros forman una categoría especial con los extranjeros que han dado pasos para su naturalización (Estados Unidos),

algunos preguntaron a los naturalizados acerca de nacionalidad o nacionalidades previas a la naturalización, lo cual hace que estos datos resulten útiles desde el punto de vista de la adscripción de los censados a un grupo étnico (México, Nicaragua),

algunos países preguntaron, además, la nacionalidad de los padres (Argentina, Panamá, Paraguay, Venezuela).

Entre los problemas que se presentan en los censos sobre nacionalidad legal, se cuenta el de las personas sin nacionalidad legal o apátridas. En algunos países se ha formado con ellas un grupo aparte; en otros, se han adscrito al grupo de la nacionalidad que tuvieron previamente.

Como segundo problema se presenta el de quienes tienen doble nacionalidad (generalmente personas nacidas en un país en el que rige el *jus soli* de padres nacidos en un país en el que rige el *jus sanguinis*). En estos casos, si el censo se realiza en uno de esos dos países, se les cuenta como nacionales; en caso de que el censo se realice en un país distinto de los dos, se interrogará al extranjero acerca de la nacionalidad en la que prefiere entrar en el censo.

Un tercer problema es el planteado por el cambio de fronteras, con respecto al cual no se ha previsto nada internacionalmente.

La comparabilidad internacional de los resultados se enfrenta a otros problemas de designación, como es el representado por el caso de los súbditos del Imperio Británico que en algunos censos se clasifican como "ingleses", en otros como "británicos", en otros bajo "Gran Bretaña", en otros como "Gran Bretaña e Irlanda del Norte", en otros como "súbditos y colonos británicos", etc., sin que estas denominaciones cubran exactamente la misma extensión y sin que, por lo mismo, sean comparables los datos consignados en cada una de ellas.

Entre los datos adicionales que es deseable tomar con respecto a la nacionalidad legal, además de los ordinarios de sexo, edad y estado matrimonial (con especificación de ser posible, de la nacionalidad del cónyuge), se cuentan los relativos al número de años de permanencia de los extranjeros en el país, los cuales sirven para estudiar los progresos de la naturalización y la relación entre la mayor o menor permanencia y la asimilación de los extranjeros. Además es importante hacer la clasificación de los extranjeros por ocupaciones, ya que las leyes nacionales dejan abiertas a los mismos, sólo algunas de ellas y esto puede explicar cierta selección de las migraciones, etc.

Entre las preguntas relativas a naturalización se cuentan principalmente:

a.—si ha habido o no naturalización del extranjero,
b.—cuál ha sido la nacionalidad previa,
c.—cuál ha sido el lugar de nacimiento,
d.—cuál ha sido la última residencia.

ESTADÍSTICAS DE EXTRANJEROS EN BÉLGICA

Hace 20 años aproximadamente que las estadísticas oficiales belgas contienen datos numerosos y sugerentes acerca de los extranjeros.

En la práctica administrativa, el extranjero se define de acuerdo con su estatuto jurídico de ciudadano, o de acuerdo con su nacionalidad legal, y no por raza o lugar de origen. Las estadísticas enumeran sólo a los extranjeros radicados en Bélgica con exclusión de los agentes diplomáticos, los turistas o los militares estacionados en el territorio.

Con respecto al país de nacionalidad, frecuentemente se incluye en la misma clasificación a quienes provienen de los territorios metropolitanos y a quienes provienen de las colonias en casos como el de Francia, Inglaterra, etc.

Las informaciones estadísticas no proporcionan todos los informes de los que se dispone acerca de los extranjeros, ya que ciertos datos no son reunidos o analizados, y otros permanecen inéditos y son utilizados únicamente por los organismos oficiales; en muchos casos, los resultados son publicados en forma incompleta y se reunen en clases estadísticas demasiado amplias. Los agrupamientos de datos se hacen por circunscripción administrativa y no por unidades sociales como sería de desear.

De ahí que, en una investigación, deban clasificarse las informaciones correspondientes según:

1.—época en que se obtuvieron los datos,
2.—organismo que intervino en la obtención,
3.—el hecho de ser las informaciones generales o específicamente referidas a los extranjeros,
4.—el hecho de que las mismas se refieran a todos los nacionales y a todos los extranjeros o que sólo tomen en cuenta ciertas categorías: adultos, trabajadores, etc.,
5.—el que la fuente sea un censo realizado en fechas determinadas periódicamente, o que sean documentos administrativos que se llevan permanentemente en las oficinas del gobierno.

Censos generales de la población, la industria, el comercio y la agricultura.—El método de obtención de estos censos que son las fuentes estadísticas oficiales más importantes consiste en:

1.—señalar una fecha determinada para realizar el Censo,
2.—enviar a cada hogar (o a cada empresa industrial o comercial, según el caso) a grupos de censadores designados por los ayuntamientos.
3.—dotar a dichos censadores de cédulas impresas conteniendo un cierto número de preguntas que se llenan con las declaraciones hechas por los habitantes,
4.—verificar y transcribir los boletines, formando, mediante el despojo de los mismos, tabulaciones especiales.

Desde 1910, la verificación, transcripción y despojo de las cédulas censales ha dejado de efectuarse por los ayuntamientos, desarrollándose en cambio la centralización de dichas operaciones, de tal modo que ahora sólo la recolección de los datos corresponde a las administraciones comunales en tanto que las restantes operaciones se realizan por los servicios centrales de estadística, por medio de máquinas estadísticas, gracias a lo cual:

a.—se eliminan los errores,
b.—se acelera y facilita el despojo.

En esta forma, los censos permiten una enumeración más exacta de los extranjeros que la que pueden proporcionar ciertos registros administrativos que también pueden resultar útiles como auxiliares, según es el caso del "Registro de Extranjeros".

Mientras en algunos censos, como el de 1920 que estuvo sujeto a limitaciones presupuestales en su publicación, los datos son fragmentarios, en el censo de 1930 se consigna:

1º—repartición de habitantes por nacionalidad distinguiéndose la nacionalidad de origen de los extranjeros, el lugar de nacimiento (Bélgica o el extranjero), y el sexo, por cada jurisdicción administrativa, y dentro de cada una de ellas, por grupos de comunas,

2º—datos sobre la población extranjera en las secciones correspondientes a:

estado civil,
edad,
profesión,
familias,
jefes de familia.

En los censos de la industria, la agricultura y el comercio, en ocasiones se omite la nacionalidad de los trabajadores (1930), mientras que en otros casos se señala la repartición de los extranjeros ocupados según su sexo, el grupo de actividad, y la provincia.

El censo económico y social de 1937 tiene en cuenta:

1.—el país de nacionalidad de los trabajadores extranjeros,
2.—su repartición por edad y sexo,
3.—su repartición por grupos de industria o de comercio,
4.—su distribución según el monto de sus ingresos,
5.—su clasificación en obreros, empleados y dirigentes obreros.

Los censos relativos a los desocupados también tienen en cuenta la nacionalidad de éstos, y señalan:

1.—duración de la desocupación,
2.—la edad, etc.

También se toma en cuenta la nacionalidad de los diplomados de la enseñanza superior en los datos censales concernientes.

Las investigaciones que conciernen especialmente a los extranjeros están constituídas por:

I.—El Registro de Extranjeros, ya que:

todo extranjero que resida por más de 15 días en Bélgica, debe obtener un *titre de sejour*, inscribirse en el *registre des étrangers* y obtener un certificado de inscripción en ese registro, y, al prolongarse su estancia por más de 6 meses, obtener una carta de identidad.

La utilización de estos datos debe hacerse, sin embargo, con ciertas reservas, ya que las entradas y salidas de los extranjeros no dejan traza en dichos documentos.

II.—Censos especiales de extranjeros como el efectuado en 1939 en vísperas de la movilización, y cuyos datos se obtuvieron por medio de censadores enviados a cada hogar a registrar la presencia de algún extranjero en él. Los datos así obtenidos se verificaron por medio de preguntas a empleadores, huéspedes, etc.

Los resultados se refieren a extranjeros de más de 15 años por:

nacionalidad,
sexo,
edad,
lugar de residencia,
país de nacimiento,
fecha de llegada a Bélgica,
carácter migratorio,
ocupación,
estado social, etc.

Entre los extranjeros censados se consideraron aquellos que habían sido movilizados por su país y habían dejado Bélgica, pero cuyas familias permanecían en Bélgica.

III.—Estadísticas de permisos de trabajo en el caso de los extranjeros que ejercen sus actividades conforme a las estipulaciones de un contrato de trabajo o de un empleo, y estadísticas de cartas profesionales en el caso de quienes ejercen una actividad profesional independiente.

Otras estadísticas oficiales que contienen datos importantes acerca de los extranjeros son:

Las estadísticas de matrimonios, basadas en los datos contenidos en las actas matrimoniales y comunicadas por los oficiales del estado civil, las cuales, desde 1947 tienen en cuenta la nacionalidad de los cónyuges.

Las estadísticas de naturalización y de declaraciones de indigenato (comprendiendo estas últimas las declaraciones de opción, las declaraciones de renuncia a la nacionalidad belga por quienes adquieren una extranjera, las declaraciones de las mujeres que desean conservar o recobrar la nacionalidad belga a pesar de su matrimonio con un extranjero, etc.).

Las estadísticas de comerciantes ambulantes, hechas por el Ministerio de Asuntos Económicos, que se obtienen a partir de las cartas de las que los comerciantes son titulares obligados. Dichas estadísticas señalan: distribución por nacionalidad, por tipo de comercio ejercido, por división administrativa, etc.

Las estadísticas de aprendizaje artesanal, realizadas por el mismo Ministerio, y

Las estadísticas de la Fundación Universitaria, relativas a la población de las universidades y centros de enseñanza superior.

EDUCACIÓN

Los datos sobre educación importan para:

1.—la planeación de programas educaitvos,
2.—la difusión informativa,
3.—los programas de mejoramiento económico, político y social.

En efecto, los datos sobre educación indican el nivel cultural y la posición de los individuos en la comunidad.

Los problemas de estandarización de los datos sobre educación provienen principalmente de:

1.—diferencia en los sistemas educativos de diferentes países (problema del contenido social),
2.—diferencia en cuanto a las posibilidades estadigráficas de diversos países (problema de aproximación metodológica).

En los censos de población recientes se incluyen tres variedades principales de datos acerca de características educativas:

a.—datos sobre alfabetismo y analfabetismo,
b.—datos sobre el nivel educativo, número de años de escuela y tipo de instrucción recibida por cada individuo,
c.—datos de asistencia escolar.

De acuerdo con el nivel general educativo de un país, puede enfatizarse uno sobre los otros dos tipos de datos; a una población atrasada le interesarán sobre todo los datos sobre alfabetismo y analfabetismo; a una adelantada en este aspecto (Canadá, Estados Unidos, Suecia, Finlandia, Holanda), le interesarán más las diferencias en el nivel educativo, dándose por sentado que la mayoría de la población está alfabetizada y que, en caso de no estarlo alguno de sus miembros este hecho (excepcional en tales países) podrá deducirse de los mismos datos proporcionados acerca del nivel de instrucción.

La Comisión de Población de las Naciones Unidas propuso que, para fines censales se definiera el alfabetismo como "la capacidad para leer y escribir un mensaje sencillo en cualquier idioma", y que los datos sobre el mismo se tomaran en relación con la población mayor de 15 años. A partir de ésta, los grupos por edad y sexo para la tabulación de datos sobre analfabetismo deberán ser los siguientes:

De menos de 15 (si el censo llega a cubrir estas edades).
De 15 a 19 años.
De 20 a 24 años.
Grupos de diez años de 25 a 64.
De 65 años o más.

En los censos realizados por diferentes países, en la obtención de datos sobre analfabetismo se han utilizado tres procedimientos:

1º—Preguntar si la persona sabe leer (exclusivamente) lo cual no satisface los requerimientos de la definición de persona alfabeta (que lee y escribe).
2º—Preguntar si la persona sabe leer y escribir (procedimiento en el que precisa cuidar de no considerar alfabetas a personas que sólo leen y escriben su nombre, ya que esto tampoco se ajusta a lo requerido por la definición).
3º—Preguntar separadamente:
a.—si la persona sabe leer,
b.—si sabe escribir.

El último de estos procedimientos —desde luego más fino que los dos anteriores— tiene la ventaja de permitir la constitución de un grupo o categoría de "semi-alfabetos" constituído por quienes saben leer pero no escribir.

En cuanto a la tabulación de los datos sobre alfabetismo de la población hay tantas variedades que pueden contrastarse, en un extremo, países como Guatemala y la Dominicana que excepto por la distinción entre población "menor de" y población "mayor de", no hacen otra distinción, y, en el otro extremo, países que, como Portugal, señalan el alfabetismo de la población de 7 a 119 años, de año en año.

Nivel educativo.—La Comisión propuso que se indicara el nivel educativo alcanzado en términos de años completos de instrucción cursada, haciendo una división simultánea por grupos de edad de los interrogados:

personas de menos de 25 años.
Grupos de 10 años de intervalo de personas entre 25 y 64.
Personas de más de 65 años.

En algunos censos (entre los que se incluyen los de México, Nicaragua, Perú, Chile, Grecia, Bulgaria, Polonia, Hungría, Suecia y Dinamarca), se menciona asimismo el *tipo* de instrucción al que se refieren los años de escuela cursados: instrucción universitaria, técnica, comercial, por correspondencia, etc.

En cuanto a los datos censales sobre nivel educativo alcanzado puede observarse que, no obstante que las cédulas censales muestran a menudo gran riqueza en cuanto a informes que al respecto se exigen a la población censada, las publicaciones tabulares de los datos recogidos resultan a menudo demasiado pobres debido, probablemente, a la complejidad de las tabulaciones que sería necesario hacer de tales datos, así como al costo de la impresión correspondiente. En el caso de México, por ejemplo, en el censo de 1940 la información se redujo a la educación primaria no obstante que en las cédulas había referencias a educación superior alcanzada.

Asistencia escolar.—Son varios los problemas que, con respecto a la asistencia escolar se presentan para la comparación censal internacional; en algunos países se interroga acerca de las personas que asisten a la escuela en el momento del censo; en otros, se refiere esa asistencia o falta de asistencia a un período determinado (del tantos del mes tal al día tantos del mes tal); en algunos países se adopta para esta pregunta cierta amplitud en las edades de las personas a las que hay que referirla (¿cuántas personas mayores de tantos años y menores de tantos años asisten o no asisten a la escuela en este momento? o ¿cuántas entre esas edades han asistido entre el día tantos y el día tantos?); en algunos países hay referencia al tipo de institución al que se asiste (México, Polonia y Suecia hacen una categoría separada de las personas asistentes a instituciones especiales del tipo de las academias comerciales y las escuelas de idiomas); en algunos países se señala el nivel de la escuela o curso al que se asiste; en algunos (Venezuela) se interroga acerca de las razones que hay para la falta de asistencia a la escuela de personas en edad escolar. Las tabulaciones a este respecto sufren, a menudo, serios cortes.

La clasificación por grupos de edad de los asistentes o faltantes a las escuelas aparece en casi todos los censos, excepto en el mexicano de 1940, lo que limita la utilidad de los datos al faltar la referencia a la edad de quien asiste o no a la escuela.

POBLACIÓN ECONÓMICAMENTE ACTIVA

Las datos relativos a la población económicamente activa interesan principalmente, desde los ángulos:

1.—Económico, y
2.—Sociológico.

Económicamente, importan con el fin de determinar:

a.—Las disponibilidades de mano de obra,
b.—La distribución de la mano de obra por ramas de actividad,
c.—La importancia relativa de diversas ramas de actividad en la producción total y el ingreso nacional,
d.—El curso del desarrollo económico.

Sociológicamente, interesan dichos datos para determinar:

a.—La constitución de diferentes estratos económico-sociales,
b.—La proporción de asalariados frente a no asalariados,
c.—La proporción de trabajadores manuales frente a trabajadores intelectuales, etc.

Las principales fuentes de obtención de datos acerca de la población económicamente activa son:

1.—Los censos de población.
2.—Los registros de seguridad social (seguros de desempleo, de enfermedad, vejez, etc.).
3.—Ocasionalmente, ciertos registros de conscripción.
4.—Ocasionalmente también, ciertos muestreos que se realizan simultáneamente con el censo.

La necesidad de empleo de los censos de población en el estudio de la población económicamente activa se pone de manifiesto en cuanto se considera que, a diferencia de otras fuentes, son los únicos que consignan datos acerca de:

a.—personas empleadas en organizaciones caritativas,
b.—trabajadores familiares que no reciben paga,
c.—trabajadores domésticos,

d.—trabajadores agrícolas, y
e.—empleados gubernativos.

Se define, en general, como población económicamente activa, "la parte de población que suministra el trabajo para la producción de bienes y servicios económicos", incluyéndose a:

a.—empleadores,
b.—trabajadores por cuenta propia,
c.—trabajadores miembros de la familia que no reciben pago en metálico, ya que se considera como "ocupación lucrativa" aquella por la que es remunerada la persona que la realiza ya sea en dinero o en especie, directa o indirectamente.

Deben incluirse:

a.—quienes se encuentran realizando esos tipos de trabajo en el momento del censo, y
b.—los desempleados.

Diferentes reuniones internacionales (El Comité de Expertos de la Liga de las Naciones y la Sexta Conferencia Internacional de Estadísticos Laborales) han adoptado dos criterios distintos para la identificación de los miembros de la población económicamente activa:

1º—El criterio de la "ocupación remunerada" *(gainful occupation)*, y
2º—El criterio de la "fuerza de trabajo" *(labour force)*.

El criterio de ocupación remunerada se basa en la idea de que cada persona tiene un papel funcional más o menos estable, por lo cual se le interroga acerca de su ocupación y subsecuentemente se tabulan los datos que caen dentro del concepto de ocupación remunerada indicado más arriba. Las ventajas que presenta este criterio estriban en la caracterización que permite hacer de las poblaciones desde el punto de vista económico y social; sus desventajas son principalmente las derivadas de falta de precisión en cuanto se trata de personas que no tienen una ocupación única y bien definida (amas de casa, estudiantes, etc.), así como las que derivan de una cierta falta de objetividad.

Conforme al criterio de fuerza de trabajo, se interroga a los individuos acerca de la ocupación que realizan en el momento del censo, o que han realizado en un corto período inmediatamente anterior. La principal ventaja en la adopción de este criterio consiste en su objetividad; su principal desventaja en que los datos obtenidos están afectados seriamente por condiciones atípicas que pueden imperar en el momento de recoger los datos censales, siendo esto particularmente importante en países en que se producen amplias variaciones ocupacionales en cortos períodos de tiempo.

De lo anterior parece desprenderse el que los dos criterios no se excluyen sino se complementan y que, utilizados adecuadamente y en conexión con otras fuentes informativas acerca de los cambios de ocupación en cortos períodos pueden dar valiosas indicaciones acerca de la estratificación y la movilidad ocupacionales y económicas.

NIVEL DE VIDA

Internacionalmente se ha reconocido la conveniencia de distinguir tres conceptos que a menudo se confunden o usan indistintamente como si fueran sinónimos.

1.—nivel de vida,
2.—estándar de vida, y
3.—norma de vida.

Nivel de vida sirve para indicar las condiciones reales de vida de un grupo humano.

Estándar de vida sirve para designar aquellas condiciones de vida que un grupo cree deseable, posible y justo alcanzar.

Norma de vida se refiere a ciertas condiciones de vida que se consideran convenientes por lo que respecta a fines determinados, y que se alcanzan mediante acuerdos o convenios nacionales o internacionales (ejem.: salarios, horas de trabajo, etc.).

El concepto de nivel de vida, para ser internacional y humanamente útil debe relacionarse con la satisfacción de necesidades y aspiraciones de índole muy diversa, que van desde las puramente materiales (de bienestar físico, posibilidades de consumo, etc.), a otras de carácter inmaterial como las culturales y educativas o las relativas al goce de derechos políticos.

Sin embargo, si bien es cierto que conceptualmente el nivel de vida debe integrarse no sólo por la consideración de los factores materiales sino también por la de los inmateriales, no es menos cierto que éstos en ocasiones son difícilmente manejables; así, si bien es cierto que puede tenerse un índice del disfrute estético mediante los datos relativos a la asistencia a los museos y galerías artísticas éste resultará insuficiente ya que no incluye las motivaciones de los asistentes; en forma semejante, si bien el número de votantes puede ser un indicador político, su validez y utilidad estarán sujetas a que se precise si quienes votán lo hacen por propia voluntad o por evitar el que, de no hacerlo, se les imponga una multa.

De otra parte, la difícil comparabilidad de los niveles de vida entre los pueblos depende de la relatividad cultural, o sea, de la diferencia de patrones culturales (de la diversidad de matrices valorativas o de la diferencia de mentalidad) entre los distintos grupos humanos tanto como de la diversidad de sus necesidades (en relación con el medio físico en el que se desarrollan, etc.).

Así, por ejemplo, todas las personas necesitan mantener una cierta temperatura corporal, pero, para lograrlo, tiene diferentes requerimientos el esquimal que el habitante de las zonas tropicales; las calorías necesarias variarán asimismo de acuerdo con múltiples factores entre los que se cuentan la edad, el sexo, etc.

Por otra parte, aumentan los problemas de comparabilidad en cuanto se tiene en cuenta que, con los cambios de la moda o el gusto dentro de cada sociedad, cambian asimismo sus necesidades.

A fin de salvar, en una primera aproximación, estas dificultades, el Comité de las Naciones Unidas encargado de la medición de los niveles de vida, decidió considerar como denominador común el de las necesidades a las que se ha tratado de dar satisfacción mediante la fijación de objetivos político-sociales hecha por organismos nacionales e internacionales.

Según esto, el Comité concluyó que la forma más satisfactoria de medir los niveles de vida en el orden internacional era medir aspectos o partes claramente delimitados de las condiciones generales de vida capaces de ser representados cuantitativamente y de expresar objetivos internacionales.

Dichos aspectos o partes delimitadas de las condiciones generales de vida (nutrición, salud, vivienda, empleo, enseñanza, etc.), reciben el nombre de *componentes* del nivel de vida, y dentro de cada uno de ellos se toman en cuenta ciertos factores concretos (calorías consumidas, grado de alfabetización, etc.), que se utilizan en la medición estadística y a los que se conoce como *indicadores.*

El uso de los indicadores plantea, a su vez, ciertos problemas que dependen, entre otras cosas, de:

a.—Que los diversos indicadores sean independientes o interdependientes entre sí (o sea, que estén o no, en cierta forma, implicados unos en otros).

b.—Que dos o más indicadores pueden resultar contradictorios entre sí (o sea que mientras uno señala un nivel alto otro señala un nivel bajo de vida en la población).

c.—Que los indicadores se refieran a realidades o a posibilidades (así número de escuelas es un indicador de posibilidades, como lo es también el indicador número de maestros en tanto no se precise número de matriculados (y aún éste es un indicador imperfecto si no se tiene en cuenta asistencia media, resultados de la instrucción, etc.).

d.—Que los indicadores se refieran a medios o a fines (así, por ejemplo, ¿el alfabetismo se considera como un fin en sí o como un medio de lograr la elevación técnica y cultural de la población?).

Los indicadores, a su vez, pertenecen a muchos tipos distintos, debiendo procurarse, para los fines de la comparabilidad internacional, que se usen los de un mismo tipo en la comparación; así, se establecen distinciones entre:

1.—Macro-indicadores (ejem.: renta nacional).
2.—Micro-indicadores (ejem.: presupuesto familiar).

De otra parte, se distingue entre:

1.—indicadores personales,
2.—indicadores colectivos.

O bien, según un tercer criterio, entre:

1.—indicadores estáticos (ejem.: calorías per capita),
2.—indicadores dinámicos o tendenciales (ejem.: aumento de la producción agrícola, industrial, etc., en relación con el aumento de población, o relación entre inversiones y renta nacional, etc.).

El procedimiento de los componentes difiere substancialmente del procedimiento que se basa en las cantidades de dinero gastado en el consumo, para apreciar el nivel de vida; en efecto, para valorar el nivel de vida de alguien importa conocer no sólo cuánto ha gastado, sino en qué ha gastado, pues si el gasto se ha destinado a elementos perjudiciales, no obstante haber gastado lo mismo o más que antes, habrá hecho que su nivel de vida descienda.

El procedimiento monetario no carece, con todo, de interés, pero es mucho más laborioso e impreciso; así, por ejemplo, se presenta el problema de la convertibilidad de las monedas, y, de otra parte, el que la diversidad de necesidades en razón de clima, precios, etc. dentro de los que se desarrolle la vida de dos personas o grupos por comparar, hagan necesario un procedimiento como el siguiente:

Calcular los gastos de la primera persona y determinar en seguida qué cantidad de dinero necesitaría la segunda persona para adquirir los mismos bienes y servicios o su equivalente; calcular a continuación los gastos reales de la segunda persona, y determinar la cantidad de dinero que en bienes y servicios tendría que emplear la primera persona para conseguir lo logrado mediante tales gastos por la segunda persona.

Pero, aún en el caso del procedimiento de los componentes, es necesario tener en cuenta que los diversos tipos de indicadores dan lugar a distintos tipos de comparaciones; así, puede haber:

1.—comparaciones directas entre individuos (escolaridad, consumo de calorías, etc.),
2.—comparaciones entre grupos (tasas de natalidad, índices de analfabetismo, etc.).

En este último caso, cuando los indicadores presenten cifras individuales, se puede recurrir a los promedios; sin embargo, la comparación de promedios encierra ciertos peligros si no se tienen en cuenta, además, medidas de dispersión o de distribución de la característica estudiada entre distintos grupos o capas de la población.

Los problemas técnicos que se presentan en la medición de los niveles de vida son, en realidad, problemas estadísticos relativos a:

1.—La necesidad de obtención de informes anuales acerca de la variación en los niveles de vida y, consecuentemente, la necesidad de progreso de las técnicas estadísticas que permitan rendir tales informes.
2.—La necesidad de que las variaciones anuales puedan medirse con gran precisión, ya que los cambios en las condiciones de vida son, en general, excesivamente lentas.
3.—La necesidad de que en la interpretación de las variaciones anuales se tengan en consideración las alteraciones producidas por las variaciones estacionales, cíclicas y seculares.
4.—La necesidad de tener en cuenta que los ciclos económicos no tienen la misma significación ni el mismo efecto en los diversos países del mundo y que, consiguientemente, afectan en forma diversa los distintos indicadores del nivel de vida.
5.—La necesidad de determinar (estadísticamente) los márgenes de error (y, por lo tanto los límites de precisión) dentro de los que pueden moverse las cifras relativas a cada uno de los indicadores del nivel de vida.

De otra parte, es necesario considerar la necesidad de que, cuando no se recurra a los censos, se reconozca:

6.—La precisión de que la parte de la población estudiada sea representativa del total de la misma al que se van a hacer extensivos los resultados; o sea, que se obtengan *muestras* representativas de la población o *universo* en estudio mediante una rigurosa aplicación de los métodos del muestreo basados en el cálculo de probabilidades.

Además, con fines de comparabilidad, resalta la necesidad de uniformizar:

a.—definiciones,
b.—métodos de recolección,
concentración,
elaboración e
interpretación de datos.

Sin embargo, más allá de los problemas técnicos estadísticos, existen algunos otros que es necesario enfrentar en la estimación de los niveles de vida desde el punto de vista social y cultural; en efecto, es necesario tener en cuenta el tipo de estructura social en el que se da determinado nivel de vida para poder apreciarlo en su verdadero significado; así, por ejemplo, los estudios sobre "salario real" pueden dar lugar a problemas y dudas con respecto al número y situación de los "trabajadores no remunerados" si se aplican exclusivamente los puntos de vista de una economía en el que la remuneración se entiende como entrega periódica de una cantidad precisa de dinero por un servicio prestado, y se olvida aquellos tipos de sociedad y economía en los cuales los servicios se remuneran mediante contra-prestaciones no monetarias, solidaridad, etc., ya que en tales casos aparecerá un alto porcentaje de trabajadores no remunerados, no obstante lo cual el nivel de vida no será tan bajo como esto haría suponer, en cuanto existen esas otras contraprestaciones no monetarias que equivalen a dicha remuneración.

Componentes del nivel de vida:

Los componentes del nivel de vida de la población propuestos por el Comité de las Naciones Unidas es una síntesis selectiva de las listas proporcionadas por los diversos organismos especializados de las propias Naciones Unidas, hecha teniendo en cuenta la particular estructura de este organismo. Sin embargo, conviene señalar que, cada organismo especializado al formular su lista obedeció a ciertos criterios, entre los que destacan los establecidos por la Organización Internacional del Trabajo que, al señalar los componentes que consideró indispensables para medir niveles de vida tuvo en cuenta:

1.—Su importancia para el bienestar del individuo, conforme a normas objetivas generalmente aceptadas.

2.—La medida en que su deficiencia constituye un problema en relación con aspiraciones sentidas.

3.—La medida en que depende de la intervención humana el remedio de la deficiencia.

4.—La posibilidad de su medición estadística.

Los componentes propuestos por el Comité de las Naciones Unidas fueron:

1º—salud, con inclusión de las condiciones demográficas,
2º—alimentación y nutrición,
3º—educación, con inclusión de alfabetismo y enseñanza técnica,
4º—condiciones de trabajo,
5º—empleo y desempleo,
6º—consumo y ahorro globales,
7º—transporte,
8.—vivienda e instalación doméstica,
9º—vestido,
10º—esparcimiento y recreo,
11º—seguridad social,
12.—libertades humanas.

Indicadores:

1º—Indicadores del componente "Salud, con inclusión de condiciones demográficas",

a.—esperanza de vida al nacer,
b.—tasa de mortalidad infantil,
c.—tasa de mortalidad bruta anual,
d.—número de camas de hospital en relación con la población total,
e.—número de médicos en relación con la población total.

Todos estos indicadores son indirectos, y los tres primeros dependen, en cierto modo, de los dos últimos; éstos, además, muestran posibilidades más que realidades.

2º—Indicadores del componente "Alimentos y nutrición",

a.—Promedio nacional de la distribución al por menor de alimentos expresado en *calorías.*

b.—Promedio nacional de la distribución al por menor de alimentos expresado en *proteínas.*

c.—Promedio nacional de la distribución al por menor de alimentos expresado en *proteínas animales.*

d.—Educación y propaganda, y legislación relativa a la alimentación.

Así como en el caso del elemento "salud y condiciones demográficas" cabía establecer las dificultades de comparación de la tasa de mortalidad bruta entre poblaciones con una distinta distribución por edad y sexo, cabe decir lo mismo con respecto a las necesidades de consumo alimenticio, ya que este consumo varía en relación con la edad, el sexo y el peso corporal por lo que "la distribución de las poblaciones por edad y sexo es un elemento importante para llegar a un juicio válido sobre la suficiencia en el consumo de alimento".

3º—Indicadores del componente "Educación",

a.—Proporción de niños entre 5 y 14 matriculados y asistentes a las escuelas.

b.—Proporción de matriculados o asistentes a las escuelas pos-primarias en relación con:

α.—el total de la población.

β.—el total de niños entre 5 y 14 años.

c.—Proporción de escuelas primarias por cada 100,000 niños entre 5 y 14 años.

d.—Proporción de alumnos por maestro en las primarias.

e.—Número total, porcentaje, y distribución de la población alfabetizada, a partir de cierta edad.

f.—Porcentaje de matriculados en instituciones de enseñanza técnica por cada 100,000 habitantes.

g.—Proporción de diarios circulantes por cada 1,000 habitantes.

h.—Proporción de libros publicados anualmente por cada 100,000 personas.

La UNESCO fija 15 años como edad límite para la apreciación del alfabetismo, y añade a este indicador el de la "mediana de los años completos de instrucción escolar cursados por la población de 25 o más años de edad, acreditados oficialmente".

4º—Indicadores del componente "Condiciones de Trabajo".

a.—Horas de trabajo por semana,

b.—salario semanal de los trabajadores industriales,

c.—salario real de los trabajadores industriales,

d.—horas de trabajo por semana establecidas como normales por la ley o los contratos,
e.—días de vacaciones pagadas por año,
f.—edad mínima de admisión al trabajo.

En este sector, las principales limitaciones dependen del hecho de que los datos de que se dispone se refieren casi exclusivamente a los trabajadores industriales que no constituyen sino una parte de la población trabajadora total.

De otra parte, las comparaciones en cuanto a horas de trabajo deben hacerse en relación con la economía total; la disminución de las mismas puede representar un aumento del tiempo libre de los trabajadores, pero puede ser indicador, asimismo, de niveles de desempleo.

5º—Indicadores del Componente "Empleo y desempleo".

a.—Proporción de la población económicamente activa en relación con la población total (distribución).
b.—Proporción de menores de 20 años que forman parte de la población económicamente activa.
c.—Proporción de mayores de 65 años que forman parte de la pobl. económicamente activa desocupada.
d.—Proporción de la población económicamente desocupada.
e.—Distribución porcentual por ocupaciones,
empleadores,
trabajadores por cuenta propia
miembros de la familia que trabajan sin remuneración,
empleados.
f.—Distribución por ramas ocupacionales.

Debe señalarse que existe una correlación inversa entre los niveles de vida y la proporción de personas mayores de 65 y menores de 20 años incluídas dentro de la población total; algo análogo ocurre en relación con la proporción de mujeres; sin embargo, es necesario tener en cuenta también el carácter predominantemente agrícola o industrial de la economía.

6º—Indicadores del componente "Consumo y ahorro globales":

a.—Proporción del ingreso nacional gastado en productos alimenticios.
b.—Proporción de los gastos públicos invertidos en servicios sociales.
c.—Proporción entre la inversión en tales servicios y el ingreso nacional.

d.—Índice de "consumo personal" *per capita.*

e.—Proporción entre el consumo personal y el ingreso nacional (y variaciones correspondientes).

f.—Índice de las inversiones y los ahorros *per capita* y su coeficiente de variación.

g.—Proporción entre tal índice y el ingreso nacional.

Para el estudio de las variaciones en el nivel de vida precisa que se trabaje con números índices (en cadena, etc.), expresados en términos de precios de un año base.

7º—Indicadores del componente "Transporte":

a.—Kilómetros de vía férrea por cada 100 kilómetros cuadrados de superficie.

b.—Número de pasajeros-kilómetro anual por cada 100,000 habitantes.

c.—Toneladas-kilómetros de carga anual por cada 100,000 habitantes.

d.—Kilómetros de carretera por cada 100 kilómetros cuadrados de superficie.

e.—Número de vehículos por cada 100,000 habitantes.

f.—Número de pasajeros-kilómetro transportados por aire por cada 100,000 habitantes.

En el caso del número de kilómetros de carretera, se recomienda la mención de los diversos tipos de carretera y la longitud correspondiente.

En el caso del número de vehículos, deben incluírse en el mismo:

A.—Automotores, y

B.—Vehículos de tracción animal, y entre los automotores, debe especificarse el número de los camiones y el de los automóviles.

De estos datos los más difícilmente obtenibles son los relativos al kilometraje de las carreteras y al número de vehículos, sin embargo, la obtención de los mismos no es imposible.

En relación con la diferencia de medios de transporte hay que hacer notar, en forma análoga a como se llamó la atención hacia la relatividad cultural en otros respectos, que en este sector al juzgar del nivel de vida al través de los medios de transporte hay que tener en cuenta la relativización que a dicho juicio impone la geografía de la región; en efecto, un medio de transporte que podría considerarse como indicador de un nivel de vida inferior puede no significar eso, sino el hecho de que las condiciones geográficas del país, lo convierten en el único medio de transporte utilizable en él.

Una limitación reconocida por el Comité Internacional encargado de estudiar los niveles de vida es la que se refiere, en este renglón, al hecho de que entre los indicadores no se incluyen medios de transporte acuático, u otros más personales (bicicletas, mulas, caballos, etc.).

8º—Indicadores del componente "Viviendas e instalación doméstica". El Comité consideró muy difícil el establecimiento de indicadores adecuados para este componente que permitieran la comparabilidad internacional, ya que, el concepto mismo de vivienda no está bien definido y, de otra parte, la apreciación del nivel de vida al través de la vivienda se enfrenta a la relativización de dos tipos:

1.—Geográfica (relacionada con requerimientos del clima, etc.) y
2.—Sociológica (vinculada a tipos de parentesco, etc.).

Sin embargo, el Comité propuso indicadores como los siguientes:

a.—Clases de vivienda y materiales empleados en su construcción.
b.—Superficie por ocupante.
c.—Personas por unidad de vivienda.
d.—Suministro de agua:
potable,
para otros usos.
e.—Servicio sanitario y alcantarillado.
f.—Servicios públicos e instalaciones comunales.

9º—Indicadores del componente "Vestido":

Este componente se encuentra relativizado por:

a.—el clima,
b.—el empleo,
c.—la situación social,
d.—la tradición,
e.—la moda.

No obstante, se considera conveniente que, para apreciar el nivel de vida de una población se incluyan datos no sólo sobre el vestido, sino sobre sombrero y calzado.

10º—Indicadores del componente "Esparcimiento y Recreo" (Ejemplificación):

a.—Número de butacas en cinematógrafos y teatros por 100,000 habitantes,

b.—Número de aparatos de radio por 100,000 habitantes.
c.—Número de telerreceptores por 100,000 habitantes, etc.).

A los indicadores que como ejemplificación se dieron respecto al componente "esparcimiento y recreo" es necesario agregar descripciones y datos numéricos tan detallados como se pueda sobre diversas formas de esparcimiento y recreo existentes en diversas poblaciones ya que, fuera del mundo occidental, la carencia de cines, teatros, radio y tele-receptores, etc., podría deformar la visión del nivel de vida de las poblaciones, en caso de no considerarse las formas alternativas de satisfacción de las necesidades de esparcimiento.

11º—Indicadores del componente "seguridad social". Incluyen principalmente informes acerca de los diferentes tipos de seguro de ancianidad, de desempleo, etc., de subsidios familiares y de todas aquellas prestaciones ofrecidas por los servicios de seguridad social o sus equivalentes; sin embargo, el Comité no llegó a especificar los indicadores pertinentes en dicho renglón de la medición del nivel de vida.

12º—Indicadores del componente "libertades humanas". Como ejemplificación de los posibles indicadores de este componente, el Comité menciona la participación de la mujer en diversas actividades, su situación legal y su situación política.

IV.—EJEMPLOS

FORMACIÓN DE UNA SERIE SENCILLA:

Datos sin ordenar:

4	4	5	4	3
5	5	4	3	2
6	6	6	6	2
7	7	7	7	7

1.—Número de datos: 20

2.—Columna de número progresivo:

3.—Mínimo: 2

4.—Consignación en una segunda columna

4	4	5	4	3
5	5	4	3	2
6	6	6	6	2√
7	7	7	7	7

N	x
1	2
2	
3	
...	
20	

5.—Señalar en los datos sin ordenar los que ya se hayan registrado en la columna x. (estado de la tabulación hasta $N = 10$)

4√	4√	5	4√	3√
5√	5√	4√	3√	2√
6	6	6	6	2√
7	7	7	7	7

N	x
1	2
2	2
3	3
4	3
5	4
6	4
7	4
8	4
9	5
10	5
....	...
20	—

N	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

Serie resultante:

N	x
1	2
2	2
3	3
4	3
5	4
6	4
7	4
8	4
9	5

6.—Proceder en la misma forma hasta llenar la casilla de enfrente del último número progresivo y haber checado todos los datos del conjunto.

4√ 4√ 5√ 4√ 3√
5√ 5√ 4√ 3√ 2√
6√ 6√ 6√ 6√ 2√
7√ 7√ 7√ 7√ 7√

N	x
10	5
11	5
12	6
13	6
14	6
15	6
16	7
17	7
18	7
19	7
20	7

FORMACIÓN DE UNA SERIE DE FRECUENCIAS:

Datos sin ordenar:

4	4	5	4	3
5	5	4	3	2
6	6	6	6	2
7	7	7	7	7

4√	4√	5√	4√	3√
5√	5√	4√	3√	2√
6√	6√	6√	6√	2√
7√	7√	7√	7√	7√

1.—Mínimo: 2

2.—Consignado en primer término en la columna de las x_i:

x_i
2

3.—Diferencia entre la magnitud de un dato y la del siguiente:

$$3 - 2 = 1$$

4.—Mínimo + 1, 2, 3, veces la diferencia mínima:

x_i
2
3
4
5
6
7 (valor máximo)

5.—Recuento mediante cuadretes y señal de los ya contados frente a los datos

2	
3	
4	□
5	
6	□
7	

6.—Cuenta de cuadretes =
= 20

debe ser igual a total de datos = 20

7.—Recuento de los cuadretes correspondientes a cada valor de x_i (frecuencia — F)

x_i	f
2	2
3	2
4	4 Tabulación resultante,
5	3
6	4
7	5

CÁLCULO DIRECTO DE LA MEDIA ARITMÉTICA EN SERIES DE CLASES Y FRECUENCIAS

Edad de los individuos varones de una muestra de la población filipina tomada en el año de 1939

Clases de edad		f_i	m_i	m_if_i
5 a menos de	15	27	10	270
15	25	19	20	380
25	35	14	30	420
35	45	9	40	360
45	55	6	50	300
55	55	4	60	240
		79		1 970

$$\bar{x}_a = \frac{\Sigma m_if_i}{\Sigma f_i}$$

$$\bar{x}_a = \frac{1\,970}{79} = 24.9$$

Media aritmética 24 años 10 meses.

CÁLCULO DE LA MEDIA CUADRÁTICA

Puntos en que difirieron (por exceso o por defecto) con respecto a la puntuación media las personas sometidas a una prueba

Diferencias de puntuación		f_i	m_i	m_i^2	$m_i^2f_i$
— 45	— 35.1	6	— 40	1 600	9 600
— 35	— 25.1	11	— 30	900	9 900
— 25	— 15.1	11	— 20	400	6 000
— 15	— 5.1	18	— 10	100	1 800
— 5	4.9	20	0	0	0
5	14.9	19	10	100	1 900
15	24.9	16	20	400	6 400
25	34.9	12	30	900	10 800
35	44.9	7	40	1 600	11 200
		124			57 600

$$\bar{x}_q = \sqrt{\frac{\Sigma m_i^2f_i}{\Sigma f_i}}$$

$$\bar{x}_q = \sqrt{\frac{57\,600}{124}}$$

$$\bar{x}_q = \sqrt{464}\ 21.54$$

Media cuadrática de las diferencias de puntuación con respecto a la puntuación media: 21.54 puntos en más o en menos.

CÁLCULO DE LA MEDIA ARMÓNICA EN EL CASO DE SERIES DE CLASES Y FRECUENCIAS

Distribución de los ingresos de un grupo de personas

Monto de los ingresos	f_i	m_i	m_i^{-1}	$m_i^{-1}f_i$
50 a menos de 150	800	100	.01000	8.00
150 250	500	200	.00500	2.50
250 350	350	300	.00333	1.16
350 450	275	400	.00250	.68
450 550	225	500	.00200	.45
550 650	200	600	.00166	.33
650 750	175	700	.00142	.25
750 850	150	800	.00125	.18
850 950	120	900	.00111	.13
	2 795			13.68

$$\bar{x}_h = \left(\frac{\Sigma m_i^{-1} f_i}{\Sigma f_i} \right)^{-1}$$

$$\bar{x}_h = \left(\frac{13.68}{2\,795} \right)^{-1}$$

$$\bar{x}_h = \frac{2\,795}{13.68} = 204.3$$

Media armónica de los ingresos del grupo: 204.3

CÁLCULO DE LA MEDIA GEOMÉTRICA DE UNA SERIE DE CLASES Y FRECUENCIAS

Índices de precios de un grupo de artículos durante 3 años (Índices mensuales)

Índices de precios	f_i	m_i	$\log m_i$	$(\log m_i) f_i$
90 a menos de 95	4	92.5	1.96614	7.86456
95 100	10	97.5	1.98900	19.89000
100 105	12	102.5	2.01072	24.12864
105 110	10	107.5	2.02141	20.21410
	36			72.09730

$$\bar{x}_g = \text{Antilog}\, \frac{\Sigma (\log . m_i) f_i}{\Sigma f_i}$$

$$\bar{x}_g = \text{Antilog}\, \frac{72.09730}{36}$$

$$\bar{x}_g = \text{Antilog}\; 2.00270$$

$$\bar{x}_g = 100.63$$

Media geométrica de los índices de precios de los tres años: 100.63.

CÁLCULO DE LA MEDIA ARITMÉTICA CON APLICACIÓN DE LOS DOS PRINCIPIOS SIMPLIFICADORES

Distribución por edades de los individuos varones de una muestra de la población filipina tomada en 1939

Clases de edad		f_i	m_i	$m_i - m'$	$\dfrac{m_i - m'}{i} = d'$	$\dfrac{m_i - m'}{i} f_i$
5	15	27	10	— 30	— 3	— 81
15	25	19	20	— 20	— 2	— 38
25	35	14	30	— 10	— 1	— 14
35	45	9	40	0	0	0
45	55	6	50	10	1	6
55	65	4	60	20	2	8
65						— 133
						14
						— 119

m' media adivinada

m_i puntos medios

i intervalo

f_i frecuencia de clase

$\bar{x}_a$ media aritmética

$m_i - m'$ desviación de los puntos medios (m_i) con respecto a la media adivinada (m').

$\dfrac{m_i - m'}{i}$ desviación de los puntos medios (m_i) con respecto a la media adivinada (m') en unidades del intervalo (i).

$$\bar{x}_a = m' + \left[\frac{\Sigma \left(\dfrac{m_i - m'}{i} \right) f_i}{\Sigma f_i} \right] i$$

$$\bar{x}_a = 40 + \left[\frac{-119}{79} \right] 10$$

$$\bar{x}_a = 40 + [(-1.5) \times 10] = 40 + (-15) = 40 - 15 = 25$$

CÁLCULO SIMPLIFICADO DE LA MEDIA ARITMÉTICA EN SERIES CON INTERVALOS DIFERENTES

Clases		f_i	m_i	$m_i - m'$	$(m_i - m')f_i$
1	— 15	43	8	— 17	— 731
15	— 35	34	25	0	0
35	— 65	19	50	25	475
		96			— 256

$$\bar{x}_a = m' + \frac{\Sigma(m_i - m')f_i}{\Sigma f_i}$$

$$\bar{x}_a = 25 + \frac{-256}{96} = 25 - 2.67 = 22.33$$

CÁLCULO CONJUNTO DE LA MEDIA ARITMÉTICA, LA DESVIACIÓN MEDIA CUADRÁTICA Y LAS MEDIDAS DE ASIMETRÍA Y DE CURTOSIS EN UNA SERIE DE CLASES Y FRECUENCIAS

Número de personas expuestas a enfermarse durante el próximo año en un grupo de asegurados

Edades		*Expuestos*	m_i	d_i'	$d_i'f_i$	$d_i'^2f_i$	$d_i'^3f_i$	$d_i'^4f_i$
15 a	19.9	35	17	—4	—140	560	—2 240	8 960
20	24.9	145	22	—3	—435	1 305	—3 915	11 745
25	29.9	160	27	—2	—320	640	—1 280	2 560
30	34.9	150	32	—1	—150	150	— 150	150
35	39.9	125	*37*	0	0	0	0	0
40	44.9	100	42	1	100	100	100	100
45	49.9	90	47	2	180	360	720	1 440
50	54.9	75	52	3	225	675	2 025	6 075
55	59.9	50	57	4	200	800	3 200	12 800
60	64.9	35	62	5	175	875	4 375	21 875
65	69.9	25	67	6	150	900	5 400	32 400
70	74.9	15	72	7	105	735	5 145	36 015
75	19.9	5	77	8	40	320	2 560	20 480
80	84.9	3	82	9	27	243	2 187	19 683
85	89.9	1	87	10	10	100	1 000	10 000
		1 014			—1 045	7 763	—7 585	184 283
					1 212		26 712	
Sumas					167		19 127	
Divididas entre 1 014					.1646	7.655	18.862	181.590
Momentos con respecto a la media arbitraria					μ_1'	μ_2'	μ_3'	μ_4'

Potencias crecientes del primer momento con respecto a la media arbitraria:

Primera: $\mu_1' = .1646$
Cuadrado: $\mu_1'^2 = .0271$
Cubo: $\mu_1'^3 = .0045$
Cuarta: $\mu_1'^4 = .0007$

Momentos con respecto a la media aritmética:

Primero: $\mu_1 = 0$

Segundo: $\mu_2 = \mu_2' - \mu_1'^2 = 7.655 - .0271 = 7.384$

Tercero: $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + \mu_1'^3 = 18.862 - 3(.1646)(7.655 +$
$+ 2(.0045)$

$\mu_3 = 18.862 - 3.78 + .0090 = 15.172.$

Cuarto: $\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2 - 3\mu_1^4 =$

$\mu_4 = 181.59 - 4(.1646)(15.172) + 6(.0271)(7.384) -$
$- 3(.0007) =$

$\mu_4 = 181.59 - 9.9892448 + 1.2006384 - .0021 = 172.7992936$

Momentos con respecto a la media aritmética en unidades originales.

Potencias del intervalo de clase:

Primera: $i = 5$

Segunda: $i^2 = 25$

Tercera: $i^3 = 125$

Cuarta: $i^4 = 625$

Multiplicación de los momentos con respecto a la media aritmética en unidades del intervalo por las potencias de éste:

Primer momento: $\mu_1 = .1646 \times 5 = .8230$

Segundo momento: $\mu_2 = 7.384 \times 25 = 184.6$

Tercer momento: $\mu_3 = 15.172 \times 125 = 1\,896.5$

Cuarto momento: $\mu_4 = 172.8 \times 625 = 108\,000$

Obtención de la media aritmética:

$$\bar{x}_a = m' + \mu_1 = 37 + .8230 = 37.823$$

Obtención de la desviación media cuadrática:

$$\sigma = \sqrt{\mu_2} = \sqrt{184.6} = 13.58$$

Obtención del índice pearsoniano de asimetría:

$$\beta_1 = \frac{1896.5}{(13.58)^3} = \frac{1\,896.5}{2\,506.9} = .756$$

Obtención del índice pearsoniano de curtosis:

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{108\,000}{34\,077} = 3.16$$

INTERPOLACIÓN DE UNA CURVA NOR

Distribución por edades de l

Clases de edad		f_i	$l_i - \bar{x}_a$	$L_i - \bar{x}_a$	$\frac{l_i - \bar{x}_a}{}$
0	10 —	10	— 49.07		— 2.53
10	20 —	45	— 39.07		— 2.02
20	30 —	125	— 29.07		— 1.50
30	40 —	210	— 19.07		— 0.98
40	50 —	280	— 9.07	.93	— 0.47
50	60 —	295		10.93	
60	70 —	185		20.93	
70	80 —	105		30.93	
80	90 —	40		40.93	
90	100 —	5		50.93	

Valores calculados previamente:

$\bar{x}_a = 49.07$

$\sigma = 19.37$

MAL POR EL MÉTODO DE LAS ÁREAS

os pobladores de un caserío

$L_i - \bar{x}_a$	*Áreas entre la* $\bar{x}_a$ *y l*	*y L*	*Áreas en cada intervalo*	*Áreas de dist. normal unitaria* $x\Sigma f_i$
	.4943		.0160	20.8
	.4783		.0451	58.6
	.4332		.0967	125.7
	.3365		.1557	202.41
0.05	.1808	.0199	.2007	260.9
0.56		.2123	.1924	250.1
1.08		.3599	.1476	191.9
1.60		.4452	.0853	110.9
2.11		.4826	.0374	48.6
2.63		.4957	.0131	17.0

Las cifras de la última columna son las que teóricamente deberían esperarse de ser la distribución normal; si se comparan con las correspondientes frecuencias reales f_i podrá verse en qué clases de edad hay deficiencias y en cuales excedencias con respecto a lo que debería esperarse que ocurriera normalmente.

INTERPOLACIÓN DE UNA CURVA DEL SISTEMA DE

Número de personas expuestas a enfermarse du

Puntos medios de las clases de edad m_i	*Expuestos* f_i	$m_i - \bar{x}_a$	$\frac{m_i - \bar{x}_a}{\sigma} = \delta_i$	φ_0	φ_3	*signo*
17	35	— 20.823	— 1.53	.1238	.1248	—
22	145	— 15.823	— 1.17	.2012	.3840	—
27	160	— 10.823	— 0.80	.2897	.5469	—
32	150	— 5.823	— 0.43	.3637	.4403	—
37	125	— .823	— 0.06	.3982	.0716	—
42	100	4.177	+ 0.31	.3802	.3423	+
47	90	9.177	0.68	.3166	.5463	+
52	75	14.177	1.04	.2323	.4635	+
57	50	19.177	1.41	.1476	.2107	+
62	35	24.177	1.78	.0818	— .0245	—
67	25	29.177	2.15	.0396	— .1380	—
72	15	34.177	2.52	.0167	— .1408	—
77	5	39.177	2.88	.0063	— .0962	—
82	3	44.177	3.25	.0020	— .0499	—
87	1	49.177	3.62	.0006	— .0208	—

Valores obtenidos anteriormente:

$\bar{x}_a = 37.823$

$\sigma = 13.58$

$\beta_1 = .756$

$\beta_2 = 3.16$

GRAM-CHARLIER A UNA SERIE DE CLASES Y FRECUENCIAS

ante el próximo año en un grupo de asegurados

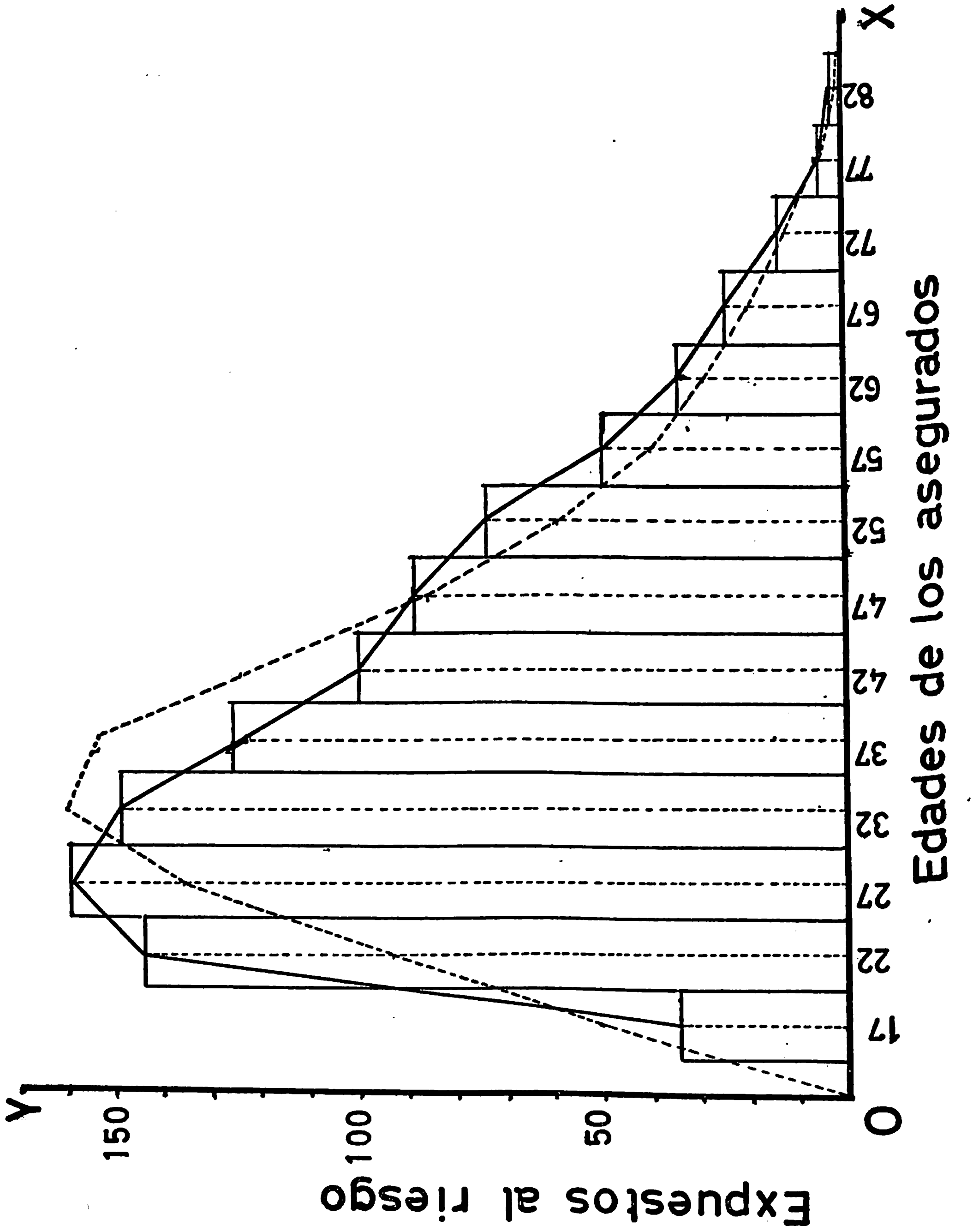
$-.167\sqrt{\beta_1}\varphi_3$	φ_4	$.42(\beta_2-3)\varphi_4$	*Suma*	$\frac{\Sigma f_i i}{\sigma}$ *Suma*
+.0181	—.6888	—.0046	.1373	51.25
+.0557	—.6720	—.0045	.2524	94.22
+.0793	—.1247	—.0008	.3682	137.44
+.0638	.7001	+.0047	.4322	161.34
+.0104	1.1894	+.0079	.4165	155.47
—.0496	.9250	+.0062	.3368	125.72
—.0792	.1391	+.0009	.2383	88.95
—.0672	—.5389	—.0036	.1615	60.28
—.0306	—.7347	—.0049	.1121	41.84
+.0036	—.4887	—.0033	.0821	30.65
+.0200	—.1332	—.0009	.0587	21.91
+.0204	.0871	+.0006	.0377	14.07
+.0139	.1389	+.0009	.0211	7.88
+.0072	.1039	+.0007	.0099	3.70
+.0030	.0547	+.0004	.0040	1.49

Ordenada máxima: $\dfrac{\Sigma f_i i}{\sigma} = \dfrac{1\,014 \times 5}{13.58} = 373.3$

Valores necesarios para el cálculo:

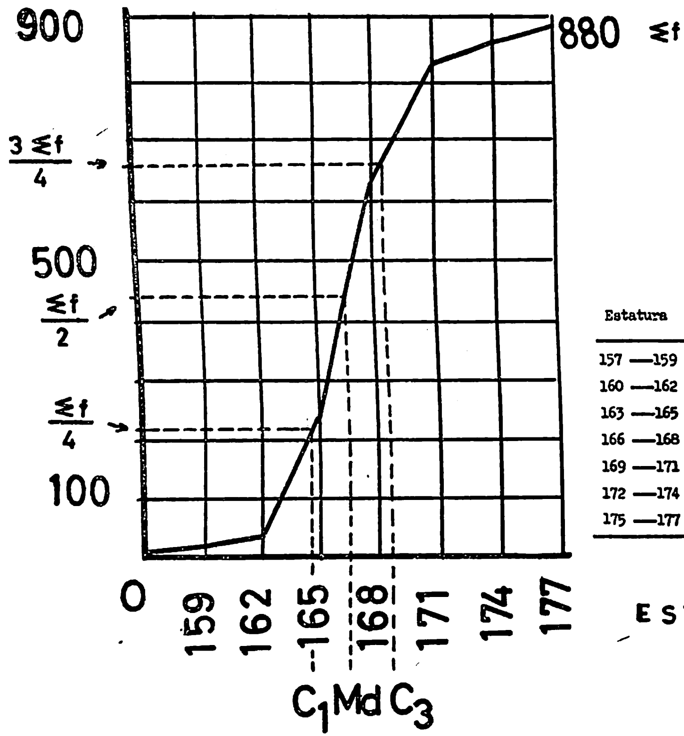
$$\sqrt{\beta_1} = \sqrt{.756} = .8694; \ .167\sqrt{\beta_1} \doteq .1449 \doteq .145$$

$$\beta_2 - 3 = 3.16 - 3 = .16: \ .42(\beta_2 - 3) \doteq .0067$$



Interpolación de una Curva de Gram-Charlier

Ojiva menos de

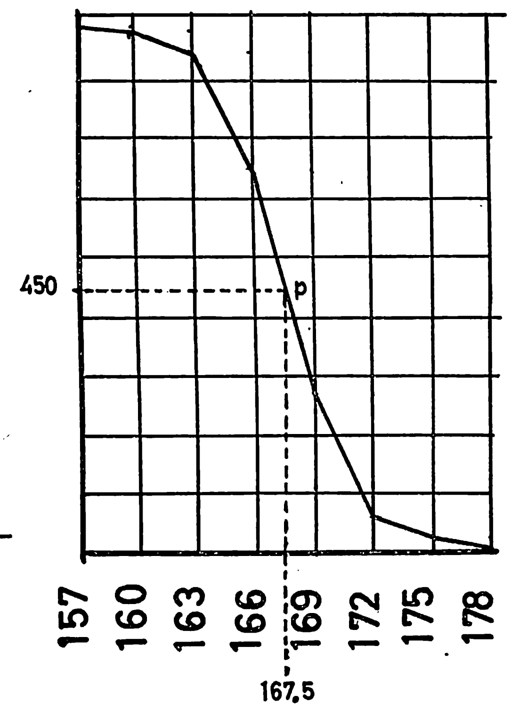


Determinación gráfica de Mediana y Cuartilas

Estatura	F_i	F_{a-}	F_{a+}
157 —159	2	2	880
160 —162	28	30	878
163 —165	200	230	850
166 —168	380	610	650
169 —171	210	820	270
172 —174	35	855	60
175 —177	25	880	25

ESTATURAS

Ojiva más de



LECTURA

450 individuos tienen más de 167.5 de estatura

POBLACIÓN ECONÓMICAMENTE ACTIVA EN KALABÁ

Años	x	*Pobl. en millones* y	xy	x^2
1946	1	5.2	5.2	1
1947	2	5.5	11.0	4
1948	3	5.5	16.5	9
1949	4	5.8	23.2	16
1950	5	6.0	30.0	25
1951	6	6.2	37.2	36
1952	7	6.5	45.5	49
1953	8	6.5	52.0	64
1954	9	6.8	61.2	81
1955	10	7.0	70.0	100
	55	61.0	351.8	385
	Sx	Sy	$S(yx)$	$S(x^2)$

Ecuación general:

$$y = a + bx$$

Ecuaciones de interpolación:

$$Sy = Na + S(x)\,b$$

$$S(yx) = S(x)\,a + S(x^2)\,b$$

Substitución de valores:

$$61 = 10a + 55b$$

$$351.8 = 55a + 385b$$

Resolución del sistema por determinantes:

Eliminante:

$$\begin{vmatrix} 10 & 55 \\ 55 & 385 \end{vmatrix} = 3\,850 - 3\,025 = 825$$

Determinante para a:

$$\begin{vmatrix} 61 & 55 \\ 351.8 & 385 \end{vmatrix} = 23\,485 - 19\,349 = 4\,136$$

Determinante para b:

$$\begin{vmatrix} 55 & 351.8 \\ 10 & 61 \end{vmatrix} = 3\,518 - 3\,355 = 163$$

Valor de a:

$$a = \frac{\textit{Determinante para } a}{\textit{Eliminante}} = \frac{4\,136}{825} = 5.01$$

Valor de b:

$$b = \frac{\textit{Determinante para } b}{\textit{Eliminante}} = \frac{163}{825} = 0.1975$$

Ley del fenómeno:

$$y = 5.01 + 0.1975x$$

POBLACIÓN ECONÓMICAMENTE ACTIVA EN KALABÁ

Años	x'	*Pobl. en millones* y	$x'y$	x'^2
1946	—4	5.2	—20.8	16
1947	—3	5.5	—16.5	9
1948	—2	5.5	—11.0	4
1949	—1	5.8	— 5.8	1
1950	0	6.0	0	0
1951	1	6.2	6.2	1
1952	2	6.5	13.0	4
1953	3	6.5	19.5	9
1954	4	6.8	27.2	16
	0	54.0	—54.1	60
			65.9	
			11.8	
	$S(x')$	$S(y)$	$S(x'y)$	$S(x'^2)$

Mediana de los años con respecto a la cual se calcularon los valores de *equis* como desviaciones: 1950

Ecuación general:

$$y = a' + bx'$$

Ecuaciones de interpolación:

$$Sy = Na' + S(x')b$$

$$S(yx') = S(x')a + S(x'^2)b$$

Substitución de valores, previa simplificación:

$$Sy = Na'$$

$$S(yx') = S(x'^2)b$$

Substitución:

$$54 \quad = 9a'$$

$$11.8 = 60b$$

Resolución de las dos ecuaciones independientes:

$$a' = \frac{54}{9} = 6 \qquad\qquad b = \frac{11.8}{60} = .196$$

Ley del fenómeno:

$$y = 6 + .196x'$$

POBLACIÓN ECONÓMICAMENTE ACTIVA EN KALABÁ

Años	x''	*Pobl. en millones*	$x''y$	x''^2
1946	—4.5	5.2	—23.40	20.25
1947	—3.5	5.5	—19.25	12.25
1948	—2.5	5.5	—13.75	6.25
1949	—1.5	5.8	— 8.70	2.25
1950	—0.5	6.0	— 3.00	.25
1951	0.5	6.2	3.10	.25
1952	1.5	6.5	9.75	2.25
1953	2.5	6.5	16.25	6.25
1954	3.5	6.8	23.80	12.25
1955	4.5	7.0	31.50	20.25
	0	61.0	—68.10	82.50
			84.40	
			16.30	

Mediana de los años, con respecto a la cual se tomaron las *equis* como desviaciones: 1950.5

Ecuación general:

$$y = a'' + bx''$$

Ecuaciones de interpolación:

$$Sy = Na'' + S(x'')b$$
$$S(yx'') = S(x'')a + S(x''^2)b$$

Ecuaciones simplificadas:

$$Sy = Na''$$
$$S(yx'') = S(x''^2)b$$

Substitución de valores:

$$61 \quad = 10a''$$
$$16.30 = 82.50b$$

Resolución de las dos ecuaciones independientes:

$$a'' = \frac{61}{10} = 6.1 \qquad\qquad b = \frac{16.30}{82.50} = .197$$

Ley del fenómeno:

$$y = 6.1 + .197x''$$

POBLACIÓN ECONÓMICAMENTE ACTIVA EN KALABÁ

Años	X''	*Pobl. en millones*	$X''y$	X''^2
1946	—9	5.2	— 46.8	81
1947	—7	5.5	— 38.5	49
1948	—5	5.5	— 27.5	25
1949	—3	5.8	— 17.4	9
1950	—1	6.0	— 6.0	1
1951	1	6.2	6.2	1
1952	3	6.5	19.5	9
1953	5	6.5	32.5	25
1954	7	6.8	47.6	49
1955	9	7.0	63.0	81
	0	61.0	—136.2	330
			168.8	
			32.6	
	$S(X'')$	$S(y)$	$S(yX'')$	$S(X''^2)$

Mediana de los años, con respecto a la cual se tomaron las *equis* como desviaciones en medios años o semestres: 1950.5

Ecuación general:

$$y = a'' + b'X''$$

Ecuaciones de interpolación:

$$S(y) = Na'' + S(X'')\,b'$$

$$S(yX'') = S(X'')\,a + S(X''^2)\,b'$$

Ecuaciones simplificadas:

$$S(y) = Na''$$

$$S(yX'') = S(X''^2)\,b'$$

Substitución de valores:

$$61 \quad = 10a''$$

$$32.6 = 330b'$$

Resolución de las dos ecuaciones independientes:

$$a'' = \frac{61}{10} = 6.1$$

$$b' = \frac{32.6}{330} = .0987$$

Ley del fenómeno:

$$y = 6.1 + .0987\, X''$$

X'' (como desviación semestral)

o $$y = 6.1 + .1974\, x''$$

x'' (como desviación anual)

CRECIMIENTO DE LA POBLACIÓN DE KALABÁ *

(Población en centenas de millar)

Años	X''	*Habitantes*	X''^2	yX''^2	X''^4	yX''
1830	—9	5	81	405	6 561	— 45
1860	—7	15	49	735	2 401	— 105
1870	—5	25	25	625	625	— 125
1880	—3	45	9	405	81	— 135
1890	—1	65	1	65	1	— 65
1900	1	90	1	90	1	90
1910	3	120	9	1 080	81	360
1920	5	155	25	3 875	625	775
1930	7	190	49	9 310	2 401	1 330
1940	9	230	81	18 630	6 561	2 070
	0	940	330	35 220	19 338	— 475
						4 625
						150
	$S(X'')$	$S(y)$	$S(X''^2)$	$S(yX''^2)$	$S(X''^4)$	$S(yX'')$

Ecuación general: $y = a'' + c''X'' + b'X''^2$

Ecuaciones de Interpolación:

$$\begin{aligned} S(y) &= Na'' + S(X'')c'' + S(X''^2)b' \\ S(yX'') &= S(X'')a'' + S(X''^2)c'' + S(X''^3)b' \\ S(yX''^2) &= S(X''^2)a'' + S(X''^3)c'' + S(X''^4)b' \end{aligned}$$

Simplificación:

$$\begin{aligned} S(y) &= Na'' + S(X''^2)b' \\ S(yX'') &= S(X'')c'' + \\ S(yX''^2) &= S(X''^2)a'' + S(x''^4)b' \end{aligned}$$

Substitución de valores en el sistema formado por las ecuaciones primera y última, y ulterior substitución en la ecuación independiente:

* Kalabá es un sitio inexistente, creado con propósitos pedagógicos a semejanza del "idioma de Kalabá" que aparece en los problemas controlados del libro de PIKE, K. L.: *Phonemics.*

$$940 = 10\,a'' + 330\,b'$$
$$35\,220 = 330\,a'' + 19\,338\,b'$$

Resolución del sistema mediante determinantes de 2º orden:

Eliminante:

$$\begin{vmatrix} 10 & 330 \\ 330 & 19\,338 \end{vmatrix} = 193\,380 - 108\,900 = \underline{84\,480}$$

Determinante para a:

$$\begin{vmatrix} 940 & 330 \\ 35\,220 & 19\,338 \end{vmatrix} = 18\,177\,720 - 11\,622\,600 = $$
$$= \underline{6\,555\,120}$$

Valor de a:

$$a = \frac{\textit{Determinante para a}}{\textit{Eliminante}} = \frac{6\,555\,120}{84\,480} = 77.59$$

Determinante para b:

$$\begin{matrix} 10 & 940 \\ 330 & 35\,220 \end{matrix} = 352\,200 - 310\,200 = \underline{42\,000}$$

Valor de b:

$$b = \frac{\textit{Determinante para b}}{\textit{Eliminante}} = \frac{42\,000}{84\,480} = .497$$

Para encontrar el valor de c, utilícense la ecuación independiente:

$$SyX'' = S(X''^2)\,c$$

Substitución de valores:

$$150 = 330\,c$$

$$c = \frac{150}{330} = .45$$

Ley del fenómeno:

$$y = 77.59 + .45\,X'' + .497\,X''^2$$

CRECIMIENTO DE LA POBLACIÓN DE KALABÁ

Años	x'	*Pobl. en millones*	x'^2	yx'^2	x'^4	yx'
1860						
1870	—4	1.5	16	24.0	256	— 6.0
1880	—3	2.5	9	22.5	81	— 7.5
1890	—2	4.5	4	18.0	16	— 9.0
1900	—1	6.5	1	6.5	1	— 6.5
1910	0	9.5	0	0.0	0	0.0
1920	1	12.0	1	12.0	1	12.0
1930	2	15.5	4	62.0	16	31.0
1940	3	19.0	9	171.0	81	57.0
1950	4	23.0	16	368.0	256	92.0
	0	94.0	60	684.0	708	— 29.0
						192.0
						163.0

Ecuación general:

$$y = a' + c'x' + b'x'^2$$

Ecuaciones de interpolación:

$$S(y) = Na' + S(x')c' + S(x'^2)b'$$

$$S(yx') = S(x')a' + S(x'^2)c' + S(x'^3)b'$$

$$S(yx'^2) = S(x'^2)a' + S(x'^3)c' + S(x'^4)b'$$

Simplificación:

$$S(y) = Na' + S(x'^2)b'$$

$$S(yx') = S(x'^2)c'$$

$$S(yx'^2) = S(x'^2)a' + S(x'^4)b'$$

Substitución de valores en las ecuaciones primera y tercera que forman sistema, y después en la independiente:

$$94 = 9a' + 60b'$$
$$684 = 60a' + 708b'$$

Resolución del sistema:

Eliminante:
$$\begin{vmatrix} 9 & 60 \\ 60 & 708 \end{vmatrix} \quad 6\,372 - 3\,600 = \underline{2\,772}$$

Determinante de a:
$$\begin{vmatrix} 94 & 60 \\ 684 & 708 \end{vmatrix} = 66\,552 - 41\,040 = \underline{25\,512}$$

Valor de a:

$$a = \frac{\textit{Determinante para } a}{\textit{Eliminante}} = \frac{25\,512}{2\,772} = 9.2$$

Determinante para b:

$$\begin{vmatrix} 60 & 684 \\ 9 & 94 \end{vmatrix} = 6\,156 - 5\,940 = \underline{516}$$

Valor de b:

$$b' = \frac{\textit{Determinante para } b'}{\textit{Eliminante}} = \frac{516}{2\,772} = .0861$$

Para encontrar el valor de c, utilícese la ecuación independiente:

$$S(yx') = S(x'^2)c$$

$$163 = 60c$$

$$c = \frac{163}{60} = 2.71$$

Ley del fenómeno:

$$y = 9.2 + 2.71\,x + .0861\,x^2$$

DISTRIBUCIÓN DE INGRESOS EN KALABÁ EN 1850

Cientos de unidades monetarias de ingreso mensual x	*Millones de personas que perciben ese ingreso* y	x^{-1}	yx^{-1}	$(x^{-1})^2$ x^{-2}
1	1.08	1.000	1.0800	1.000
2	.73	.500	.3650	.250
3	.65	.333	.2167	.110
4	.47	.250	.1175	.062
5	.37	.200	.0740	.040
6	.41	.166	.0680	.027
7	.36	.142	.0511	.020
8	.31	.125	.0387,	.015
9	.28	.111	.0310	.012
10	.26	.100	.0260	.010
	4.91 $S(y)$	2.927 $S(x^{-1})$	2.0680 $S(yx^{-1})$	1.546 $S(x^{-1})^2$

Ecuación general:

$$y = a + bx^{-1}$$

Ecuaciones de interpolación

$$\begin{aligned} S(y) &= Na + S(x^{-1})b \\ S(yx^{-1}) &= S(x^{-1})a + S(x^{-1})^2 b \end{aligned}$$

Substitución de valores:

$$\begin{aligned} 2.0680 &= 2.927a + 1.546b \\ 4.91 &= 10a + 2.927b \end{aligned}$$

Resolución del sistema de ecuaciones por medio de determinantes:

Eliminante:

$$\begin{vmatrix} 10 & 2.927 \\ 2.927 & 1.546 \end{vmatrix} = 15.46 - 8.567329 = \underline{6.892671}$$

Determinante para a:

$$\begin{vmatrix} 4.91 & 2.927 \\ 2.0680 & 1.546 \end{vmatrix} = 7.59086 - 6.0319616 = \underline{1.5688984}$$

Valor de a:

$$a = \frac{\textit{Determinante para } a}{\textit{Eliminante}} = \frac{1.5688984}{6.892671} = .22$$

Determinante para b:

$$\begin{vmatrix} 10 & 4.91 \\ 2.927 & 2.0608 \end{vmatrix} = 20.608 - 14.37157 = \underline{6.23643}$$

Valor de b:

$$b = \frac{\textit{Determinante para } b}{\textit{Eliminante}} = \frac{6.23643}{6.892671} = 0.94$$

Ley del fenómeno:

$$y = .22 + 0.94\, x^{-1}$$

CRECIMIENTO DE LA POBLACIÓN DE LA REPÚBLICA MEXICANA
(1895 - 1950)

Años	x	*Pob. en millares*	$\log y$	$x \log y$	x^2
1895	5	12 632	4.10147	20.50735	25
1900	10	13 607	4.13376	41.33760	100
1910	20	15 160	4.18070	83.61400	400
1921	31	14 335	4.15640	128.84840	961
1930	40	16 552	4.21885	168.75400	1 600
1940	50	19 653	4.29343	214.67150	2 500
1950	60	25 791	4.41147	264.68820	3 600
	216		29.49608	922.42105	9 186
	Sx		$S(\log y)$	$S(x \log y)$	$S(x^2)$

Ecuación general:

$$y = ab^x$$

Transformación por logs.:

$$\log y = \log a + x \log b$$

Ecuaciones de Interpolación:

$$S \log y = N \log a + S(x) \log b$$
$$S(x \log y) = S(x) \log a + S(x^2) \log b$$

Substitución de valores:

$$29.49608 = 7 \log a + 216 \log b$$
$$922.42105 = 216 \log a + 9\,186 \log b$$

Resolución del sistema por medio de determinantes de segundo orden:

$$\log a = 4.04080$$
$$\log b = 0.005604$$

Ley general del fenómeno en forma logarítmica:

$$\log y = 4.04080 + 0.005604\, x$$

(x con origen en 1890 como se indica en la columna correspondiente).

AUTOMÓVILES REGISTRADOS POR CADA 100 PERSONAS EN LOS ESTADOS UNIDOS DE AMÉRICA (1905–45)

Años	*Automóviles por 100 personas* y	y^{-1}			
1905	.09	11.110			
1910	.50	2.000	$S_1 = 13.513$		
1915	2.48	0.403			
1920	8.72	0.114		$D_1 = -13.296$	
1925	17.50	0.057	$S_2 = 0.217$		$Q = 0.0062$
1930	21.57	0.046			
1935	20.57	0.048		$D_2 = -0.083$	
1940	24.26	0.041	$S_3 = 0.134$		
1945	21.95	0.045			

Ecuación general:

$$y^{-1} = a + bc^{x}$$

Fórmulas para los parámetros:

$$c = \sqrt[n]{Q}$$

$$b = D^{1} \frac{c-1}{(c^{n} - 1^{2})}$$

$$a = \frac{S_1 - b(1+c)}{n}$$

Substitución de valores:

$$c = \sqrt[3]{.0062}$$

$$\log .0062 = \bar{3}.79239$$

$$\text{entre: } 3 = \bar{1}.26413$$

$$c = \text{antilog} \frac{\log .0062}{3} = \text{antilog } \bar{1}.26413 = .1837$$

$$b = -13.296 \frac{.1837 - 1}{(.0062 - 1)^2} = -13.296 \frac{-.8163}{(-.9938)^2}.$$

$$b = 10.98$$

$$a = \frac{13.513 - 10.98(1.1837)}{3}$$

$$a = \frac{13.513 - 12.997026}{3} = \frac{.515974}{3} = .171991$$

Ley general del fenómeno:

$$y^{-1} = .171991 + 10.98(.1837)^x$$

POBLACIÓN ECONÓMICAMENTE ACTIVA

Método de los Momentos Factoriales

Años	x	y	F^N_1	F^N_2
1946	1	5.2	5.2	5.2
1947	2	5.5	10.7	15.9
1948	3	5.5	16.2	32.1
1949	4	5.8	22.0	54.1
1950	5	6.0	28.0	82.1
1951	6	6.2	34.2	116.3
1952	7	6.5	40.7	157.0
1953	8	6.5	47.2	204.2
1954	9	6.8	54.0	258.2
1955	10	7.0	61.0	
		61.0		
			F^{10}_1	F^{10}_2

Ecuación general:

$$y = \mathrm{A} + b(x - 1)$$

Ecuaciones de interpolación:

Momentos factoriales:

$$F^N_1 = NA + \frac{N(N-1)}{2} b$$

$$F^N_2 = \frac{N(N-1)}{2} A + \frac{N(N-1)}{2} \frac{(N-2)}{3} b$$

Substitución de valores:

$$61 = 10A + 10 \times \frac{9}{2} b$$

$$258.2 = 10 \times \frac{9}{2} A + 10 \times \frac{9}{2} \times \frac{8}{3} b$$

$$61 = 10A + 45b$$

$$258.2 = 45A + 120b$$

Resolución del sistema de ecuaciones mediante determinantes:

Eliminante:

$$\begin{vmatrix} 10 & 45 \\ 45 & 120 \end{vmatrix} = 1\,200 - 2\,025 = \underline{-825}$$

Determinante para A:

$$\begin{vmatrix} 61 & 45 \\ 258.2 & 120 \end{vmatrix} = 7\,320 - 11\,619 = \underline{-4\,299}$$

Valor de A:

$$A = \frac{\textit{Determinante para } A}{\textit{Eliminante}} = \frac{-4\,299}{-\;825} = 5.21$$

Determinante para b:

$$\begin{vmatrix} 10 & 61 \\ 45 & 258.2 \end{vmatrix} = 2\,582 - 2\,745 = -163$$

Valor de b:

$$b = \frac{\textit{Determinante para } b}{\textit{Eliminante}} = \frac{-163}{-825} = 0.197$$

Ley del fenómeno:

$$y = 5.21 + 0.197(x - 1)$$

$$y = 5.21 + -0.197 + 0.197x$$

$$y = 5.013 + \ .197x$$

MILLARES DE PERSONAS TRANSPORTADAS MENSUALMENTE POR LOS FERROCARRILES DE KALABÁ EN EL PERÍODO 1910–1955

Años	y	F_1	F_2	F_3
1910	6	6	6	6
1915	15	21	27	33
1920	28	49	76	109
1925	45	94	170	279
1930	66	160	330	609
1935	91	251	581	1 190
1940	120	371	952	2 142
1945	153	524	1 476	3 618
1950	190	714	2 190	
1955	231	945		

Fórmula general:

$$y = a_0 + (N-1)a_1 + \frac{(N-1)(N-2)}{2}a_2$$

Momentos factoriales:

$$F_1^N = Na_0 + N\frac{N-1}{2}a_1 + N\frac{N-1}{2}\frac{N-2}{3}a_2$$

$$F_2^N = N\frac{N-1}{2}a_0 + N\frac{N-1}{2}\frac{N-2}{3}a_1 + N\frac{N-1}{2}\frac{N-2}{3}\frac{N-3}{4}a_2$$

$$F_3^N = N\frac{N-1}{2}\frac{N-2}{3}a_0 + N\frac{N-1}{2}\frac{N-2}{3}\frac{N-3}{4}a_1 + N\frac{N-1}{2}\frac{N-2}{3}\frac{N-3}{4}\frac{N-4}{5}$$

Substitución de valores:

$$945 = 10a_0 + 45a_1 + 120a_2$$
$$2\,190 = 45a_0 + 120a_1 + 210a_2$$
$$3\,618 = 120a_0 + 210a_1 + 252a_2$$

Eliminante del sistema:

$$\begin{vmatrix} 10 & 45 & 120 \\ 45 & 120 & 210 \\ 120 & 210 & 252 \end{vmatrix} = 108\,900$$

Determinante para a_0

$$\begin{vmatrix} 945 & 45 & 120 \\ 2\,190 & 120 & 210 \\ 3\,618 & 210 & 252 \end{vmatrix} = 653\,400$$

Valor de a_0:

$$a_0 = \frac{\textit{Determinante para } a_0}{\textit{Eliminante}} - \frac{653\,400}{108\,900} - 6$$

Substitución del valor de a_0 en las dos primeras ecuaciones:

$$945 = 60 + 45a_1 + 120a_2$$
$$2\,190 = 270 + 120a_1 + 210a_2$$

Paso de términos independientes a un solo miembro.

$$945 - 60 = 885 = 45a_1 + 120a_2$$
$$2\,190 - 270 = 1\,920 = 120a_1 + 210a_2$$

Eliminante:

$$\begin{matrix} 45 & 120 \\ 120 & 210 \end{matrix} = -4\,950$$

Determinante para a_1:

$$\begin{matrix} 885 & 120 \\ 1\,920 & 210 \end{matrix} = -44\,550$$

Valor de a_1:

$$= \frac{\textit{Determinante para } a_1}{\textit{Eliminante}} = \frac{-44\,550}{-4\,950} = \text{o}$$

Sustitución del valor de a_1 en la ecuación más sencilla:

$$885 = 45 \times 9 + 120a_2$$
$$885 = 405 + 120a_2$$
$$885 - 405 = 120a_2$$
$$480 = 120a_2$$

$$a_2 = \frac{480}{120} = 4$$

Ley del fenómeno:

$$y = 6 + (N - 1)9 + \frac{(N - 1)(N - 2)}{2} 4$$

$$y = 6 + 9(x - 1) + 4 \frac{(x - 1)(x - 2)}{2}$$

Transformación a la forma más usual:

$$y = 6 + 9x - 9 + 2(x^2 - 3x + 2)$$
$$y = 6 + 9x - 9 + 2x^2 - 6x + 4$$
$$y = (6 - 9 + 4) + (9 - 6)x + 2x^2$$
$$y = 1 + 3x + 2x^2$$

ESCUELAS CONSTRUÍDAS EN UNA PROVINCIA DE KALABÁ
(1895–1955)

Años	u	v	S_2	S_3
1895	—3	7	7	7
1905	—2	12	19	26
1915	—1	17	36	62
1925	0	22	58	120
1935	1	27	85	205
1945	2	32	117	322
1955	3	37	154	476
		154	476	1 218

Fórmula general:

$$v = L_0 + L_1 p_1$$

Fórmulas para la obtención de los coeficientes:

$$\lambda_0 = \frac{S_1}{\binom{N}{1}}$$

$$\lambda_1 = \frac{S_2}{\binom{N+1}{2}}$$

Valores de las λ:

$$\lambda_0 = \frac{154}{\binom{7}{1}} = \frac{154}{7} = 22$$

$$\lambda_1 = \frac{476}{\binom{8}{2}} = \frac{476}{\frac{8 \times 7}{1 \times 2}} = \frac{476}{28} = 17$$

Valores las las λ':

$$\lambda'_0 = \lambda_0$$
$$\lambda'_1 = \lambda_0 - \lambda_1$$
$$\lambda'_0 = 22$$
$$\lambda'_1 = 22 - 17 = 5$$

Valores de las L:

$$L_1 = \frac{6}{N-1}\lambda'_1$$
$$L_0 = 22$$
$$L_1 = \frac{6}{6}5 = 5$$

Valor de p_1 (polinomio ortogonal de primer grado en u):

$$p_1 = u$$

Ley del fenómeno:

$$v = 22 + 5u$$

u (desviación con respecto al año mediano, según aparece en la tabla).

CRECIMIENTO DE LA POBLACIÓN DE MÉXICO (1900–1950)

Años	*x*	*Población en miles de habitantes*			
1900	0	13 607	$S_1 = 28\,767$		
1910	1	15 160		$D_2 = 14\,558$	
1920 *	2	14 334	$S_2 = 30\,886$		$Q_1 = \frac{D_2}{D_1} = \frac{14\,558}{2\,119}$
1930	3	16 552		$D_1 = 2\,119$	$= 6.87$
1940	4	19 653	$S_3 = 45\,444$		
1950	5	25 791			

Ecuación General de la Geométrica Modificada:

$$y = a + bc^{x}$$

Fórmulas para los parámetros:

$$c = \sqrt[n]{Q}$$

$$b = \frac{D_1}{(1 + c)(c^n - 1)}$$

$$a = \frac{S_1 - b(1 + c)}{n}$$

Substitución de valores para la obtención de los parámetros:

$$c = \sqrt{6.87} = 2.62$$

$$b = \frac{2\,119}{(1 + 2.62)(6.87 - 1)} = \frac{2\,119}{3.62 \times 5.87} = \frac{2\,119}{21.25} = 99.71$$

$$a = \frac{S_1 - c(1 + c)}{n} = \frac{28\,767 - 99.71(1 + 2.62)}{2} =$$

$$= \frac{28\,767 - 99.71 \times 3.62}{2} = \frac{28\,767 - 360.95}{2} = \frac{28\,407.05}{2} =$$

$$= 14\,203.525$$

* Se ha tomado como dato para 1950 el del Censo de 1921 con el propósito de simplificar pedagógicamente la ejemplificación.

Substitución de los valores de los parámetros en la ecuación general:

$$y = 14\,203.525 + 99.71\ (2.62)^{x}$$

Estimación para 1960, según esta interpolación:

$$y_{1960} = 14\,203.525 + 99.71(2.62)^{6}$$

$$y_{1960} = 46\,447.745 \text{ mil habitantes}$$

$$y_{1960} = 46\,447\,745\,000$$

VOLUMEN DE LAS VENTAS TRIMESTRALES DE UN ALMACÉN DURANTE 3 AÑOS

Trimestre	1950	1951	1952
Primero	90	120	150
Segundo	145	190	240
Tercero	140	190	230
Cuarto	125	160	200
	——	——	——
Sumas anuales	500	660	820
Medias	125	165	205

Interpolación de una recta a las medias obtenidas (125, 165, 205) procedimiento del año mediano:

Años	x	y	xy	x^2
1950	-1	125	-125	1
1951	0	165	0	0
1952	1	205	205	1
Sumas		495	80	2

$$y = a + bx$$

$$Sy = Na \qquad 495 = 3a \qquad a = 165$$

$$Syx = Sx^2b \qquad 805 = 2b \qquad b = 40$$

Ritmo anual de crecimiento:

$$b = 40$$

Ritmo trimestral:

$$b/4 = 40/4 = 10$$

Ritmo semi-trimestral:

$$b/4/2 = 10/2 = 5$$

Valor del fenómeno para el centro del año de 1951:

$$a = 165$$

(valor colocado entre el segundo y el tercer trimestre)

Valores teóricos de la tendencia:

Trimestre	1950	1951	1952
Primero	110	150	190
Segundo	120	160	200
		165	
Tercero	130	170	210
Cuarto	140	180	220

(— 5) (— 10)

(+ 10) (+ 5)

Relativos de los datos con respecto a los valores de la tendencia:

Trimestre	1950	1951	1952	*Medias trimes-trales*	*Índices ajustados* — .05
Primero	82	80	80	80.6	80.55
Segundo	121	119	120	120	119.95
Tercero	108	112	110	110	109.95
Cuarto	89	89	91	89.6	89.55
				400.2	400.00
Medias de las medias trimestrales:				100.05	100.00
Exceso por ajustar:				.2	
Valor por disminuir de cada media trimestral:				.05	

V.—*APÉNDICES*

APÉNDICE

REVISIÓN MATEMÁTICA INDISPENSABLE

Igualdad es la expresión matemática de que:

a.—Dos cantidades son equivalentes: ejemplo: 1 dólar $=$ 12.60 pesos,
b.—Dos operaciones tienen el mismo resultado: $5 \times 8 = 30 + 10$.

Ecuación es una igualdad en la que intervienen cantidades conocidas (llamadas "datos") y cantidades desconocidas (llamadas "incógnitas"); ejemplo: $8 \times 5 = 30 + x$. Esta expresión equivale a la interrogante: "¿Qué cantidad (x) es necesario agregar ($+$) a 30 para obtener el mismo resultado que se obtiene de multiplicar 8 por 5?".

Miembro de una ecuación es cada una de las partes de la misma que están separadas por el signo igual. Hay, por tanto, dos miembros en cada ecuación:

Primer miembro de la ecuación es el que se encuentra antes del signo igual.

Segundo Miembro de la ecuación es el que está después del signo igual; ejemplo:

$$\underbrace{8 \times 5}_{\text{primer miembro}} = \underbrace{30 + x}_{\text{segundo miembro}}$$

Término es toda expresión separada por los signos más o menos (cuando una cantidad no va precedida de signo se le considera antecedida por el signo más). En la ecuación anterior:

($+$) 8×5 es un término (término único del primer miembro),
($+$) 30 es otro término (primer término del segundo miembro),
$+\ x$ es otro término (segundo término del segundo miembro).

Factor es toda expresión separada de otra por el signo por. En el caso anterior, el término único del primer miembro consta de dos factores 8 y 5.

Divisor es toda expresión separada de otra por el signo entre, o bien, todo denominador de una fracción.

Despejar una incógnita de una ecuación equivale a encontrar el valor de la incógnita, haciendo que:

I.—La incógnita quede en un solo miembro de la ecuación.
II.—Los datos queden en el otro miembro de la ecuación.

Para lograr lo anterior:

1º—Se reducen o sea, se suman o se restan según el signo que les preceda), todos los términos semejantes (o sean aquellos en los que figuran las mismas literales afectadas de los mismos exponentes), en cada miembro de la ecuación. Ejemplo:

$$2x + 3x + 13 - 5 = 5x - x + 15 + 1$$

En esta ecuación, son términos semejantes: por contener a x elevado a la primera potencia $2x$ y $3x$ en el primer miembro, y $5x$ y $-x$ en el segundo; por no contener a x, 13 y -5 en el primero y 15 y 1 en el segundo.

La reducción de estos términos semejantes nos permite escribir:

$$5x + 8 = 4x + 16$$

2º—Se hace que los términos que contienen a la incógnita queden en un solo miembro de la ecuación (en el primero, generalmente) y los que no la contienen queden en el otro para lo cual:

a.—Los términos que ya se encuentran en el miembro deseado ($5x$, término que contiene a la incógnita en el primero y 16 término que no contiene a la incógnita en el segundo) se dejan en él sin alteración.

b.—Los términos que no se encuentran en el miembro deseado se pasan al otro miembro cambiándoles el signo; o sea, que si tienen signo más en el miembro en el que se encuentran, deberán pasar al otro miembro con signo menos, y si tienen signo menos deberán pasar con signo más. (8, término positivo que no contiene a la incógnita y se encuentra en el primer miembro de la ecuación, deberá pasar al segundo con signo menos; $4x$, término también positivo que contiene a la incógnita y se encuentra en el segundo miembro deberá pasar al primero con signo negativo):

$$5x - 4x = 16 - 8$$
$$1x = 8$$

En este caso la incógnita está despejada y sabemos que x tiene que ser igual a 8 para cumplir las condiciones de la ecuación original.

Sin embargo, puede ocurrir que tras ejecutar las operaciones, x quede (ya sea en el primero o en el segundo miembro), precedida por un factor o por un divisor; en este caso, para obtener el valor de x,

3º—El factor de x se pasa como divisor al otro miembro de la ecuación.

Supongamos, por ejemplo, que tras seguir todo el proceso anterior tuviéramos:

$$4x = 12$$

$$x = \frac{12}{4}$$

$$x = 3$$

Si x tuviera un divisor en lugar de un factor, el divisor pasaría al otro miembro como factor:

$$\frac{x}{7} = 14$$

$$x = 7 \times 14 = 98$$

Sistema de ecuaciones: Cuando en una sola ecuación existen dos incógnitas, no puede precisarse el valor de cada una de ellas, ya que hay muchas combinaciones de valores que pueden satisfacer la ecuación. Así, si nos preguntamos ¿qué par de números enteros, sumados, dan 4 como suma?, podremos representar algebraicamente la pregunta por la ecuación:

$$x + y = 4$$

En la que x y y representan los valores que se buscan y que pueden ser los siguientes:

$x = 0$, $x = 4$ puesto que $0 + 4 = 4$

$x = 1$, $y = 3$ puesto que $1 + 3 = 4$

$x = 2$, $y = 2$ puesto que $2 + 2 = 4$

Y lo mismo podría decirse de x igual a 3 y y igual a 1 o de x igual a 4 y y igual a 0.

En cambio, si agregamos otra condición que diga que de esos números "el duplo del primero ($2x$) más el triple del segundo ($3y$) debe ser igual a 9" sólo habrá un par de valores que satisfaga ambas condiciones:

Cuando

x es igual a	*y es igual a*	*$x + y$ es igual a*	*$2x$ es igual a*	*$3y$ es igual a*	*$2x + 3y$ es igual a*
0	4	4	0	12	12
1	3	4	2	9	11
2	2	4	4	6	10
3	1	4	6	3	9
4	0	4	8	0	8

Por esta tabulación puede verse que, si bien todos los pares de valores consignados para x y para y en las dos primeras columnas satisfacen la condición de que sumados den como resultado 4, no todos ellos satisfacen la segunda condición de que el duplo del primero más el triple del segundo sumen 9; o sea, que sólo el par de valores 3 y 1 cumplen *simultáneamente* la condición de que, siendo enteros, sumados den 4 y la otra condición de que el duplo del primero y el triple del segundo, sumados, den 9.

Las dos expresiones ($x + y = 4$ y $2x + 3y = 9$) que expresan las dos condiciones que deben satisfacerse simultáneamente se dice que constituyen un "sistema de 2 ecuaciones (dos expresiones en las que figuran cantidades conocidas y desconocidas) simultáneas (en cuanto los valores de las incógnitas deben satisfacer a ambas) con dos incógnitas (dos valores desconocidos x y y).

Como el método de tanteo representado por las tabulaciones anteriores no podría seguirse en todos los casos por ser demasiado embarazoso (piénsese, por ejemplo, en el problema que representaría buscar diez sumandos que produzcan 245 como suma y cuyos duplos sumados den como resultado 2 600), se recurre a un procedimiento como el siguiente:

1.—Se busca un número que multiplicado por el coeficiente o factor numérico de una incógnita en una de las dos ecuaciones produzca el coeficiente de la misma incógnita en la otra ecuación. Ejemplo: sea el sistema anterior:

$$x + y = 4$$
$$2x + 3y = 9$$

¿Qué número, al multiplicarse por 1 (coeficiente de x en la primera) produce 2 (coeficiente de x en la segunda)? La respuesta es 2.

$$2 \times 1 = 2$$

2.—Se multiplica el número encontrado por todos y cada uno de los términos de la primera ecuación a fin de que la misma no se altere. En el ejemplo, se multiplicará 2 por x, por y y por 4:

$$2x + 2y = 8$$

3.—Si la incógnita cuyos coeficientes se han igualado por este procedimiento (x en este caso) tiene el mismo signo en las dos ecuaciones, la ecuación recién multiplicada y la no modificada se restan término a término; en caso contrario (o sea, si tienen signos iguales) se suman término a término.

$$\begin{array}{r} 2x + 2y = 8 \\ 2x + 3y = 9 \\ \hline (2x - 2x) + (2y - 3y) = 8 - 9 \\ -y = -1 \end{array}$$

Tener $-y$ es como tener $-1y$; para despejar a y, bastará con pasar -1 que está como factor al otro miembro como divisor de -1:

$$y = -1/-1 = 1$$

4.—Obtenido en esta forma el valor de una de las incógnitas, para obtener el valor de la otra, basta con substituir ese valor ($y = 1$ en la ecuación más sencilla del sistema y despejar, en la ecuación resultante a la otra incógnita:

$$x + y = 4$$

(ecuación más sencilla del sistema)

Substituyendo y por su valor:

$$x + 1 = 4$$

Despejando a x:

$$x = 4 - 1$$
$$x = 3$$

O sea, que, conforme habíamos encontrado previamente, $y = 1$ y $x = 3$ son los valores que satisfacen las dos condiciones del sistema.

Forma más general de solución de un sistema de dos ecuaciones con dos incógnitas.—Puede ocurrir que no exita un número entero que multiplicado por el coeficiente de la incógnita en una de la ecuaciones dé como producto el coeficiente de esa misma incónita en la otra ecuación; en tal caso, podría encontrarse un número fraccionario como factor, pero su uso tiene algunos inconvenientes en cuanto la manipulación de los números fraccionarios ya sean comunes o decimales es más embarazosa y, más aún, en cuanto se pierde precisión; de ahí que, en tales casos se siga otro procedimiento (que, además, es más general que el anterior, que no es sino caso particular de éste en un cierto sentido). Supongamos el sistema:

$$3x + 5y = 19$$

$$5x + 7y = 29$$

Puede observarse en el que no existe ningún número *entero* que multiplicado por 3, coeficiente de x en la primera, produzca 5, coeficiente de x en la segunda. Tampoco hay número entero que multiplicado por 5, coeficiente de y en la primera, produzca 7 coeficiente de y en la segunda. En casos como este:

1.—Se multiplican todos los términos de la primera ecuación por el coeficiente que tiene la incógnita por eliminar en la segunda ecuación. en el ejemplo, toda la primera ecuación por 5.

$$3x \times 5 + 5y \times 5 = 19 \times 5$$

$$15x + 25y = 95$$

2.—Se mulitplican todos los términos de la segunda ecuación por el coeficiente que tiene la incógnita por eliminar en la primera ecuación. En el ejemplo, toda la segunda ecuación por 3:

$$5x \times 3 + 7y \times 3 = 29 \times 3$$

$$15x + 21y = 87$$

3.—Estas dos ecuaciones transformadas se restan una de otra (en caso de que los términos que contienen a la incógnita cuyos coeficientes se han igualado tengan el mismo signo) o se suman (en caso de que sean contrarios los signos de los términos cuyos coeficientes se han igualado). En nuestro ejemplo, será necesario restar:

$$\begin{array}{r} 15x + 25y = 95 \\ -15x + 21y = 97 \\ \hline -4y = -2 \end{array}$$

Continuándose las operaciones como en el ejemplo anterior.

POTENCIAS Y RAÍCES

Cuadrado de un número es el resultado que se obtiene de tomar al número dos veces como factor. El cuadrado se expresa mediante un pequeño 2 colocado arriba del número cuyo cuadrado se desea indicar. Elevar a 3 al cuadrado equivale a tomar a 3 dos veces como factor, o sea a multiplicar a 3 por sí mismo una vez:

$$3^2 = 3 \times 3$$

$$3^2 = 9$$

Elevar a un número al cubo equivale a tomar al número tres veces como factor, o sea, a multiplicar al número por sí mismo y, en seguida, el producto obtenido por el mismo número:

$$5^3 = 5 \times 5 \times 5$$

$$5^3 = 125$$

Inversamente, si el exponente (2 en el primer caso y 3 en el segundo) de una potencia (3 al cuadrado, 5 al cubo) indica el número de veces que hay que tomar como factor al número afectado por ese exponente, si tenemos el producto de un cierto número de factores iguales (3×3, $5 \times 5 \times 5$) podemos representar el producto tomando el factor igual (3 en el primer caso, 5 en el segundo) y poniendo como exponente el número que indica las veces que se repite como factor:

$$3 \times 3 = 3^2$$

$$5 \times 5 \times 5 = 5^3$$

$$7 \times 7 \times 7 \times 7 = 7^4$$

Supongamos que queremos elevar al cuadrado a un número que ya está elevado al cuadrado (3^2 al cuadrado), entonces tendremos:

$$(3^2)^2 =$$

Como elevar al cuadrado equivale a tomar dos veces como factor la cantidad afectada del exponente, esto querrá decir que deberemos tomar a 3^2 dos veces como factor:

$$(3^2)^2 = (3^2) \times (3^2)$$

Por otra parte, cada una de las cantidades incluídas en los paréntesis es un cuadrado (o sea un producto de dos factores iguales, 3×3), substituyendo estos valores, tendríamos:

$$(3^2)^2 = (3 \times 3) \times (3 \times 3)$$
$$(3^2)^2 = \ 3 \times 3 \ \times \ 3 \times 3$$

En el segundo miembro de esta última ecuación figura un producto de cuatro factores iguales a 3; o sea, 3 elevado a la 4ª potencia:

$$(3^2)^2 = 3^4$$

Si se ve cuál es la relación que existe entre el exponente 4 del resultado y los exponentes del cuadrado del cuadrado (2 y 2) podremos ver que $4 = 2 \times 2$, de donde podemos decir que:

Para elevar al cuadrado $(\ )^2$ una cantidad elevada previamente al cuadrado (3^2), basta con tomar la cantidad primitiva (3) y elevarla a un exponente igual al producto de 2 (exponente de fuera del paréntesis) por 2 (exponente de dentro del paréntesis). Así:

$$(4^2)^2 = 4^4$$
$$(5^2)^2 = 5^4$$
$$(a^2)^2 = a^4$$

O sea, en general, que el cuadrado de una cantidad previamente elevada al cuadrado es igual a la cuarta potencia de dicha cantidad.

Podrímos mostrar igualmente que el cuadrado $(\ )^2$ de un cubo (4^3) será la sexta potencia ($2 \times 3 = 6$); el cuadrado $(\ )^2$ de una cuarta potencia (4^4) será la octava potencia (2×4) y, en general que:

Para elevar al cuadrado una cantidad previamente elevada a otra potencia, será necesario duplicar el exponente de dicha cantidad:

$$(a^n)^2 = a^{2n}$$

En forma semejante, puede afirmarse que:

Para elevar a una potencia m una cantidad previamente elevada a una potencia n, es necesario elevar la cantidad a la potencia que se obtiene de multiplicar m por n, o sea, a mn.

$$(a^n)^m = a^{mn}$$

Esto es aplicable incluso a los exponentes negativos cuyo significado explicamos adelante.

Raíz cuadrada de un número.—La raíz cuadrada de un número es otro número que multiplicado por sí mismo una vez da como producto el primero. Así, la raíz cuadrada de 9 ($= 3 \times 3$) es igual a 3 porque 3 es el número que multiplicado por sí mismo una vez (o que tomado dos veces como factor) produce 9.

Procedimiento de cálculo de la raíz cuadrada.—Para el cálculo de la raíz cuadrada es necesario tener en cuenta la relación que existe entre los números naturales y sus cuadrados:

Número	*Cuadrado*
1	1
2	4
3	9
4	16
5	25
6	36
7	49
8	64
9	81

Si se trata de obtener la raíz de un número cuadrado perfecto, bastará con encontrarlo en esta lista y ver a qué número corresponde; ese número será su raíz cuadrada exacta; así, de 25 la raíz cuadrada es 5, de 81 es 9.

Si el número del cual se busca la raíz cuadrada no es un cuadrado perfecto, se busca cuál es el mayor cuadrado perfecto que el mismo número contiene; la raíz de este cuadrado perfecto, seguido de una parte decimal será la raíz cuadrada del número. El cálculo se hará en la siguiente forma:

1º—Sepárense períodos de dos en dos cifras a partir del punto decimal en ambas direcciones. Si el número es entero, a partir de la cifra de las unidades hacia las cifras de orden superior. Si el número es puramente decimal y no tiene enteros, a partir del punto en dirección de las cifras correspondientes a órdenes inferiores.

En todos los casos anteriores puede ocurrir una o varia de las siguientes situaciones:

a.—Que el último período entero tenga 2 cifras o solamente una, lo cual no constituye problema, pues se trabaja, en el último caso, igual que si se tratara de un período completo.

b.—Que el último período decimal tenga 2 o una cifra. En este último caso, se completa el último período agregando un cero.

Ejemplo: Encuéntrese la raíz de 6 32. 53 2

La separación de dos en dos cifras a partir del punto en ambas direcciones da: 6/32./53/2

El último período entero (primero del número en total) está constituído por una sola cifra 6, pero esto no molesta; el último período decimal, por constar de una sola cifra, 2, necesita complementarse con un cero:

6/32./53/20

La iniciación de la operación presenta el siguiente aspecto:

$$\begin{array}{l|l} \sqrt{6/32./53/20} & \\ & \end{array}$$

2º—Véase cuál es el mayor cuadrado contenido en el primer período (¿cuál es el mayor cuadrado contenido en 6?), consultando la lista de números naturales y sus cuadrados (el mayor cuadrado contenido en 6 es 4 que se anota debajo de él y se resta), anotándose la raíz cuadrada del mismo (2) al lado del número cuya raíz se busca:

$$\begin{array}{l|l} \sqrt{6/32./53/20} & 2 \\ 4 & \text{—} \\ \text{—} & \\ 2 & \end{array}$$

3º—Bájese al lado de la resta del mayor cuadrado comprendido en el primer período y éste, el segundo período

$$\begin{array}{l|l} \sqrt{6/32./53/20} & 2 \\ 4 & \text{—} \\ \text{—} & \\ 2\ 32 & \end{array}$$

4º—Divídase la cantidad formada por la resta y el período recién bajado entre el duplo de la raíz consignado en una línea horizontal inferior a la que contiene a la raíz

$$\begin{array}{l|l} \sqrt{6/32./53/20} & 2 \\ 4 & \text{—} \\ \text{—} & 4 \\ 2\ 32 & \text{—} \end{array}$$

La división de 2 32 entre 4 (duplo de la raíz 2) da 5 como cocienciente que se anota al lado de la raíz hallada 2 y al lado del duplo 4

$$\begin{array}{l|l} \sqrt{6/32./53/20} & 25 \\ 4 & \text{—} \\ \text{—} & \\ 2\ 32 & 45 \\ & \text{—} \end{array}$$

5º—La cifra recién encontrada se multiplica por sí misma y por el duplo de la raíz restando el producto de la cantidad formada por la resta y el siguiente período. En el ejemplo, la cifra recién encontrada 5 se multiplicará por el 45 de la segunda línea restándose el producto de 232 (ya sea que se anote el resultado de la multiplicación y después se reste, a la manera estadounidense, o que ambas operaciones se realicen simultáneamente, al modo nuestro):

$$\begin{array}{r|l} \sqrt{6/32./53/20} & 25 \\ 4\quad\quad\quad\ \ & \text{—} \\ \text{—}\quad\quad\quad\ \ & \\ 2\ 32\quad\quad\ \ & 45 \\ 2\ 25\quad\quad\ \ & \text{—} \\ \text{——}\quad\quad\ \ & \\ 7\quad\quad\ \ & \end{array}$$

6º—A partir de este momento la operación repite cíclicamente los pasos 4º y 5º, o sea, que se duplicará 25 en una tercera línea, se bajará el siguiente período 53 y se dividirá 753 entre el duplo de 25 (50); el resultado se anotará en la primera línea al lado de 25, y en el duplo, al lado de 50; se multiplicará el resultado (1) por 501 (el duplo seguido de la cifra recién hallada) y se restará de 735, para obtener un nuevo residuo con el que se continuará la operación. Con respecto al punto decimal, debe decirse que en el momento en que se baja el primer período decimal y antes de obtener el cociente del valor resultante entre el duplo, se pone el punto decimal después de la última cifra obtenida en la raíz (después del 5).

RESOLUCIÓN DE UN SISTEMA DE DOS ECUACIONES CON DOS INCÓGNITAS POR MEDIO DE LOS DETERMINANTES

Un sistema de n ecuaciones con n incógnitas se puede resolver mediante el uso de determinantes de orden n. De este modo, un sistema de *2* ecuaciones con *2* incógnitas, puede resolverse mediante determinantes de *2º* orden; un sistema de *3* ecuaciones con *3* incógnitas, mediante determinantes de *3er.* orden, etc.

Un determinante de segundo orden es un binomio (expresión algebraica constituída por dos términos), cada uno de cuyos términos están constituídos por sendos productos de dos factores, y que están separados (los términos) por el signo menos. Es decir, se trata de una expresión de la forma:

$$ab - cd$$

Esta expresión acostumbra escribirse convencionalmente colocando las cuatro cantidades (a, b, c, d) que forman el determinante, entre dos líneas verticales, formando *dos* columnas y *dos* renglones, de tal manera que los dos factores que forman cada producto se encuentren en la misma diagonal, en la siguiente forma:

$$\begin{vmatrix} a & d \\ c & b \end{vmatrix}$$

De acuerdo con todo lo anterior, el valor de un determinante de segundo orden se calcula formando los productos en cruz de las cuatro cantidades, y restando del producto ab el producto cd.

Así, por ejemplo, si se trata de obtener el valor del determinante:

$$\begin{vmatrix} 5 & 3 \\ 2 & 4 \end{vmatrix}$$

Multiplicaremos 5 por 4 (= 20), y a este producto le restaremos 6 (pro ducto de 2×3); o sea:

$$\begin{vmatrix} 5 & 3 \\ 2 & 4 \end{vmatrix} = 5 \times 4 - (2 \times 3) = 20 - 6 = 14$$

O sea, que 14 es el valor de este determinante.

Uso de los determinantes de segundo orden en la resolución de un sistema de dos ecuaciones con dos incógnitas.—Sea el sistema:

$$3a + 5b = 42$$

$$7a + 9b = 82$$

Si entresacamos los coeficientes de las incógnitas a y b y los colocamos en dos columnas y dos renglones en el mismo orden en que aparecen en las ecuaciones, tendremos:

$$\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix}$$

que constituyen un determinante de 2º orden que —por permitirnos la eliminación de una incógnita— recibe el nombre de *eliminante* del sistema.

El eliminante nos permite encontrar los valores de las incógnitas (a y b), ya que sus valores estarán dados por sendas fracciones, cuyo denominador común es el eliminante; *i. e.*:

$$a = \frac{\textit{Numerador I}}{\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix}} \qquad b = \frac{\textit{Numerador II}}{\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix}}$$

Los numeradores (I y II) de estas fracciones se encuentran substituyendo en el eliminante los coeficientes de la incógnita cuyo valor se trata de determinar, por los términos independientes (42, 82) en el orden en que aparecen en el sistema de ecuaciones. Así, en el caso de a, el 3 y el 7 del eliminante se substituirán por 42 y por 82 respectivamente; en el caso de b, serán el 5 y el 9 (elementos de la segunda columna) del eliminante, los que se substituyan por 42 y 82 respectivamente.

$$\begin{array}{cc} \textit{Numerador I} & \textit{Numerador II} \\ a = \dfrac{\begin{vmatrix} 42 & 5 \\ 82 & 9 \end{vmatrix}}{\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix}} & b = \dfrac{\begin{vmatrix} 3 & 42 \\ 7 & 82 \end{vmatrix}}{\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix}} \end{array}$$

Mediante la obtención de los valores de todos estos determinantes, se obtiene:

$$a = \frac{42 \times 9 - 82 \times 5}{3 \times 9 - 7 \times 5} = \frac{-32}{-8} = 4$$

$$b = \frac{3 \times 82 - 7 \times 42}{3 \times 9 - 7 \times 5} = \frac{246 - 294}{-8} = \frac{-48}{-8} = 6$$

Según esto, el procedimiento para resolver un sistema de dos ecuaciones con dos incógnitas por el método de los determinantes puede resumirse como sigue:

1º—Se forma un determinante de segundo orden con los coeficientes de las incógnitas (este determinante se conoce como eliminante).

2º—Se substituyen los coeficientes de la primera incógnita en el eliminante por los términos independientes (aquellos que no contienen a la incógnita).

3º—Se divide el determinante así formado entre el eliminante; así se obtiene el valor de la primera incógnita.

4º—Se substituyen los coeficientes de la segunda incógnita en el eliminante, por los términos independientes de las ecuaciones.

5º—Se divide el determinante formado mediante esta substitución entre el eliminante; así, se obtiene el valor de la segunda incógnita.

El procedimiento puede simplificarse, eliminándose los pasos cuarto y quinto si, una vez encontrado el valor de la primera incógnita:

4º—Se substituye el valor de la primera incógnita en una de las ecuaciones del sistema, y se resuelve la ecuación resultante en la que aparece ya una sola incógnita (la segunda).

Nota.—Esta simplificación tiene la desventaja de que los valores de las dos incógnitas no se obtienen independientemente, de modo que un error cometido al encontrar una no perjudique a la otra, sino que, por depender el valor de la segunda incógnita del encontrado para la primera, un error cometido al calcular el valor de la primera incógnita repercute en el que se calcule para la segunda.

Determinantes de 3er. Orden

Así como un determinante de 2º orden está constituído por 4 valores arreglados en *dos* columnas y *dos* renglones, un determinante de *3er.* orden está constituído por 9 valores arreglados en *tres* columnas y *tres* renglones. De esta forma:

$$\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix} \text{ es un determinante de 3er. orden.}$$

Para obtener el valor de un determinante de 3er. orden no podemos aplicar la regla dada para encontrar el valor de los determinantes de 2º orden (multiplicaciones en cruz y resta de los resultados). Podríamos, en cambio, utilizar la llamada regla de Sarrus que consiste en repetir dos de las columnas o dos de los renglones del determinante, formar los productos en diagonal de tres factores, afectar a los que están en un sentido de signo positivo, y a los que están en sentido contrario, de signo negativo y sumarlos algebraicamente:

5	1	8		$-9 \times 6 \times 8 = -432$
4	6	3		$-5 \times 2 \times 3 = -\ 30$
9	2	7	$= 5 \times 6 \times 7 = 210$	$-4 \times 1 \times 7 = -\ 28$
5	1	8	$= 4 \times 2 \times 8 = 64$	
4	6	3	$= 9 \times 1 \times 3 = 27$	
			301	$-490 = -189$

O sea, que el valor de este determinante de tercer orden es -189.

Sin embargo, esta regla tiene el inconveniente de no ser aplicable a determinantes de orden superior al tercero; de ahí que sea preferible aplicar un procedimiento suficientemente general, como es el caso del procedimiento de los menores o de los cofactores, que consiste, fundamentalmente, en reducir los determinantes de orden superior a determinantes más sencillos, hasta llegar a los de segundo orden.

Procedimiento de los menores o de los cofactores en la resolución de determinantes

Para aplicar este procedimiento es necesario fijar terminológicamente qué entendemos por *menor* y qué por *cofactor* de un determinante de orden n.

Menor de un determinante de orden n es el determinante de orden n—1 que se obtiene al eliminar una columna y un renglón del determinante primitivo.

De este modo, el menor de un determinante de 3er. orden, será el determinante de 2º orden ($n - 1$) que se obtiene al eliminar una de las tres columnas y uno de los tres renglones del determinante de 3er. orden.

Si tenemos:

$$\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}, \text{ determinante de 3er. orden,}$$

$\begin{vmatrix} 6 & 3 \\ 2 & 7 \end{vmatrix}$ determinante de 2º orden obtenido mediante la supresión de su primera columna (5, 4, 9) y de su primer renglón (5, 1, 8), será uno de los menores del determinante anterior.

En forma análoga, si en el determinante de 3er. grado dado anteriormente, suprimimos la segunda columna (4, 6, 2) y el tercer renglón (9, 2, 7), tendremos:

$\begin{vmatrix} 5 & 8 \\ 4 & 3 \end{vmatrix}$ como otro de los menores del determinante dado, de 3er. orden.

Cofactor de un determinante de orden n, es el menor correspondiente multiplicado por -1 elevado a una potencia igual a la suma del número de orden de la columna y el número de orden del renglón suprimidos:

$$\text{Cofactor} = (-1)^{c+r} \text{ por Menor.}$$

De esta forma:

Si tenemos el mismo determinante de 3er. grado anotado anteriormente:

$$\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}$$

y queremos encontrar el cofactor del determinante correspondiente a 5 (elemento de la primera columna y del primer renglón, tendremos que este cofactor será igual al menor de 5, multiplicado por -1 elevado al cuadrado (porque se trata de la 1ª columna y del 1er. renglón y $(1 + 1 = 2)$; de donde:

$$\text{Cofactor de } 5 = (-1)^{1+1} \begin{vmatrix} 6 & 3 \\ 2 & 7 \end{vmatrix} = (-1)^2 |(6 \times 7 - 2 \times 3)$$

$$= 1\ (42 - 6) = 1 \times 36 = 36$$

O sea, que 36 es el cofactor de 5.

En forma semejante, podemos determinar en este mismo determinante los cofactores de 4, de 9, de 1, de 6, de 2, etc., o sea de todos y cada uno de los elementos del propio determinante.

Una vez definido lo que entendemos por *menor* y por *cofactor,* puede explicarse fácilmente cómo se procede para encontrar el valor de un determinante de orden superior al segundo, por medio de los cofactores.

Procedimiento: Para encontrar el valor de un determinante de 3er. grado por medio de sus cofactores: se multiplican todos y cada uno de los términos de una de sus columnas *o* de uno de sus renglones por cada uno de sus respectivos cofactores.

O sea, que, si en nuestro caso concreto elegimos la primera columna (5, 4, 9), el valor del determinante (representado generalmente por Δ mayúscula) quedará dado como sigue:

$$\Delta = 5 \times \text{Cofactor de } 5 + 4 \times \text{Cofactor de } 4 + 9 \times \text{Cofactor de } 9.$$

O, en forma más comprimida:

$$\Delta = 5\,C_5 + 4\,C_4 + 9\,C_9$$

C_5, según hemos calculado antes, es igual a 36; C_4 que podemos calcular en forma semejante, es igual a 9 y C_9 es igual a -45; por lo cual, substituyendo estos valores en nuestra igualdad, tendremos:

$$\begin{aligned} &= 5 \times 36 + 4 \times 9 + 9 \times -45 \\ &= 180 + 36 + (-405) \\ &= 180 + 36 - 405 = -189 \end{aligned}$$

El valor -189 del determinante, encontrado por este procedimiento es el mismo que obtuvimos por la regla de Sarrus.

La manera práctica de proceder a buscar el valor de un determinante consiste en:

1º—Elegir una columna o un renglón del determinante.
2º—Buscar los menores de los tres elementos de la columna o el renglón elegidos, y encontrar su valor.
3º—Conocidos los menores, multiplicarlos por $(-1)^{c+r}$ para obtener los cofactores correspondientes.
4º—Multiplicar esos cofactores por los elementos de la columna o del renglón elegidos a los que corresponden.
5º—Sumar algebraicamente los resultados.

Ejemplo: Buscar el valor del determinante:

$$\begin{vmatrix} 45 & 1 & 8 \\ 38 & 6 & 3 \\ 52 & 2 & 7 \end{vmatrix}$$

Para encontrar su valor:

1º—Elegimos la primera columna (45, 38, 52).
2º—Determinamos los menores (Mr) de 45, de 38 y de 52, que son:

$$Mr_{45} = \begin{vmatrix} 6 & 3 \\ 2 & 7 \end{vmatrix} = 6 \times 7 - 2 \times 3 = 42 - 6 = 36$$

$$Mr_{38} = \begin{vmatrix} 1 & 8 \\ 2 & 7 \end{vmatrix} = 1 \times 7 - 2 \times 8 = 7 - 16 = -9$$

$$Mr_{52} = \begin{vmatrix} 1 & 8 \\ 6 & 3 \end{vmatrix} = 1 \times 3 - 6 \times 8 = 3 - 48 = -45$$

O sea, destacando linealmente estos resultados:

$$Mr_{45} = 36$$
$$Mr_{38} = -9$$
$$Mr_{52} = -45$$

3º—Para encontrar los cofactores, multiplicaremos el menor de 45 (36) por -1 elevado al cuadrado, por tratarse de un elemento de la primera columna, y del primer renglón $(1 + 1 = 2)$; el menor de 38 (-9) lo multiplicaremos por -1 elevado al cubo por ser 38 un elemento de la primera columna y del segundo renglón $(1 + 2 = 3)$;

el menor de 52 (-45), lo multiplicaremos por -1 elevado a la cuarta potencia, por ser 52 un elemento de la primera columna y del tercer renglón ($1+3=4$).

$$C_{45} = (-1)^{1+1} \times 36 = (-1)^2 \times 36 = 1 \times 36 = 36$$
$$C_{38} = (-1)^{1+2} \times -9 = (-1)^3 \times -9 = -1 \times -9 = 9$$
$$C_{52} = (-1)^{1+3} \times -45 = (-1)^4 \times -45 = 1 \times -45 = -45$$

O sea, destacando linealmente los resultados:

$$C_{45} = 36$$
$$C_{38} = 9$$
$$C_{52} = -45$$

Como puede verse por la comparación de los menores y los cofactores, los cofactores son iguales a los menores, sólo que, en ciertas ocasiones, cambian sus signos; de ahí que podamos establecer también prácticamente:

Todo cofactor puede calcularse a partir de su menor si a éste:

a.—Se le deja el mismo signo cuando la suma del ordinal de la columna y el del renglón es positiva, par, o

b.—Se le cambia signo cuando la suma del ordinal de la columna y el del renglón es impar.

4º—En seguida, multiplicaremos el cofactor de 45 (36) por su elemento correspondiente (45); el cofactor de 38 (9) por 38; el cofactor de 52 (-45) por 52:

$$45\, C_{45} = 45 \times 36 = 1\,620$$
$$38\, C_{38} = 38 \times 9 = 342$$
$$52\, C_{52} = 52 \times -45 = -2\,340$$

5º—Si sumamos algebraicamente estos valores, obtendremos el valor del determinante:

$$1\,620 + 342 + (-2\,340) = 1\,620 + 342 - 2\,340 = -378$$

De este modo, -378 es el valor del determinante dado.

Resolución de un sistema de 3 ecuaciones con 3 incógnitas por medio de determinantes

Para resolver un sistema de 3 ecuaciones con 3 incógnitas por medio de determinantes, se procede en forma semejante a como se procedió en el caso de un sistema de 2 ecuaciones con 2 incógnitas; así:

1º—Se busca el eliminante del sistema, que será un determinante de 3er. grado formado por los coeficientes de las 3 incógnitas arregladas en 3 columnas y 3 renglones en el mismo orden en que aparecen en las ecuaciones.

2º—Para cada una de las incógnitas, se forma un determinante mediante la substitución de la columna de los coeficientes de la incógnita cuyo valor se va a determinar, por los términos independientes.

3º—Cada uno de los determinantes así formados, se dividen entre el eliminante, con lo cual se encuentran los valores correspondientes de cada una de las incógnitas;

Ejemplo:

$$\begin{aligned} 5a + 1b + 8c &= 45 \\ 4a + 6b + 3c &= 38 \\ 9a + 2b + 7c &= 52 \end{aligned}$$

Eliminante del sistema:

$$\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}$$

Determinante formado por la substitución de los coeficientes de *a* (5, 49) por los términos independientes (45, 38, 52):

$$\begin{vmatrix} 45 & 1 & 8 \\ 38 & 6 & 3 \\ 52 & 2 & 7 \end{vmatrix}$$

Valor de *a*, obtenido por la división de este determinante entre el eliminante:

$$a = \frac{\begin{vmatrix} 45 & 1 & 8 \\ 38 & 6 & 3 \\ 52 & 2 & 7 \end{vmatrix}}{\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}} = \frac{-378}{-189} = 2$$

(Los valores de determinante y eliminante se obtuvieron en la ejemplificación del cálculo de determinantes).

En forma análoga, para b y para c:

$$b = \frac{\begin{vmatrix} 5 & 45 & 8 \\ 4 & 38 & 3 \\ 9 & 52 & 7 \end{vmatrix}}{\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}} = 3$$

$$c = \frac{\begin{vmatrix} 5 & 1 & 45 \\ 4 & 6 & 38 \\ 9 & 2 & 52 \end{vmatrix}}{\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}} = 4$$

REFERENCIAS BIBLIOGRÁFICAS

ACKOFF, Rusell I.: *The Design of Social Research.* The University of Chicago Press. Chicago, 1953.

LUNDBERG, George A.: *Técnica de la Investigación Social.* Trad. de José Miranda. Fondo de Cultura Económica. México, 1ª Ed. 1949.

MILLS, Frederick Cecil: *Métodos Estadísticos Aplicados a la Economía y a los Negocios.* Versión castellana de Juan Ruiz Magan y Enrique Gastardi. Nueva e íntegra traducción de la última edición norteamericana por Juan Ruiz Magan y Juan José Ruiz Rubio. Con un suplemento sobre Estadística de Atributos por José Juan Forns y Emilio Gimeno Rascon. Aguilar, S. A. de Ediciones. Madrid.

NACIONES UNIDAS (Organización de las): *Informe sobre la Definición y Medición Internacional del Nivel de Vida.* Nueva York, 1954.

RIETZ, Henry Lewis: *Mathematical Statistics.* The Carus Mathematical Monographs. Number Three. Published for The Mathematical Association of America by the Open Court Publishing Company. Chicago. Illinois, 1927.

RIGGLEMAN, John R., and FRISBEE, Ira N.: *Business Statistics.* Mc Graw-Hill Book Company, Inc. Nueva York. Toronto. London. Third Edition, 1951.

YOUNG, Pauline V.: *Métodos Científicos de Investigación Social.* Traducción por Angela Müller Montiel. Instituto de Investigaciones Sociales de la Universidad Nacional Autónoma de México.

YULE, G. Udny y KENDALL, M. G.: *Introducción a la Estadística Matemática.* Traducción de la 13ª edición inglesa y prólogo por José Ros Jimeno. M. Aguilar Editor. Madrid, 1947.

TESTIMONIO DE RECONOCIMIENTO

El responsable de la redacción de estas notas sobre técnicas estadísticas desea hacer público su reconocimiento hacia las Sritas. Mª Teresa Aznar y Aurora Patricia Vela, bibliotecarias del Instituto Anglo-Mexicano de Cultura, quienes les facilitaron la consulta de libros tan estimulantes para el estudioso de la Estadística como *An Advanced Theory of Statistics* de M. G. Kendall; para el Srita. Lucila Elsa Flammand, antigua y excelente discípula que sostuvo con su esfuerzo de cálculo la elaboración de las ejemplificaciones; para el Sr. Vicente Polo y el Sr. Salvador Mittón, gerente y linotipista de la Gráfica Panamericana, cuyo empeño y buena voluntad permitieron zanjar muchas dificultades; para el Sr. José Mª Avilés, sin cuya ayuda se hubiesen deslizado gruesos errores al aprendiz-corrector que es quien esto escribe; para la Sra. Profa. Enriqueta Baz de Velarde y la Srita. Profa. María Lomelí, catedráticas de Matemáticas Financieras y de Estadística en la Facultad de Ciencias de la U. N. A. M., en cuyos cursos obtuvo estímulos y enseñanzas inapreciables.

El Autor.

ÍNDICE

I.—Introducción 9

II.—Técnicas 57

III.—Dos Sectores de Estudio Estadístico 307

IV.—Ejemplos 347

V.—Apéndices 391

Se terminó la impresión de este libro el día 14
de diciembre de 1957 en los talleres de
Gráfica Panamericana, S. de R. L.,
Parroquia 911. México 12, D. F. Se
imprimieron 1.000 ejemplares

VILLEGAS

...ICAS
ES...TICAS PARA
INVESTIGACIONES
SOCIALES