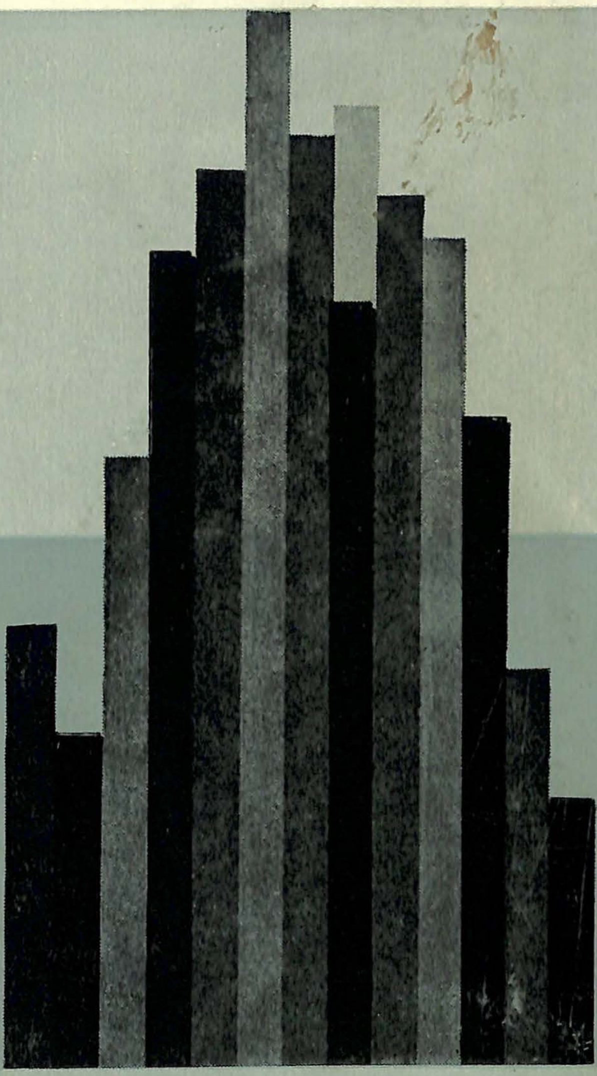
Óscar Uribe Villegas

Los elementos de la estadística social



Instituto de Investigaciones Sociales

ÓSCAR URIBE VILLEGAS: LOS ELEMENTOS DE LA ESTADÍSTICA SOCIAL

INSTITUTO DE INVESTIGACIONES SOCIALES

Óscar Uribe Villegas

Los elementos de la estadística social

Universidad Nacional Autónoma de México. *México, 1971*

Primera edición: 1971

Ciudad Universitaria. México 20, D. F.

Dirección General de Publicaciones

Impreso y hecho en México

A la memoria de

Don Manuel María Contreras
figura descollante de la didáctica
matemática mexicana

PRÓLOGO

No es éste ni un texto para "estudiantes a secas" ni un tratado para "especialistas sin más". Para unos y otros —a más de la abundante bibliografía internacional— existen ya (o pronto existirán) obras mexicanas que o representan el enfoque tradicional (como los Métodos estadísticos *de Andrés García Pérez) o el innovador (de la instrucción programada que prepara nuestra Universidad).*

No es libro para estudiantes porque no trata de trasmitir ni sólo ni principalmente un "saber hacer" estadístico: una tecnología. No es para especialistas porque se mueve en nivel elemental, a ras de tierra, fuera de cualquier exploración del "espacio exterior".

Pero es para estudiantes, especialistas y hombres cultos, si todos ellos se cuentan entre los estudiosos de la ciencia y de sus métodos.

Lo es porque trata de mostrar por qué razón se hace la estadística, cómo se hace, y dice al estudiante cuáles son los elementos que estructuran las técnicas disponibles. Lo es porque retrotrae al especialista a la tierra nutricia —de los primeros principios— de su tecnología; porque le recuerda cuáles son los elementos cuya combinación es indispensable para estructurar nuevas técnicas. Lo es porque permite que el culto ubique las técnicas dentro de un sistema de conceptos y operaciones, librándolo de la desorientación que producen las congeries de operacionès dispares sin punto alguno de convergencia.

El autor siempre ha deseado escribir para párvulos y eso le movió a redactar estas páginas. En ellas recoge materiales de sus Técnicas estadísticas para investigadores sociales *(hoy agotadas), pero no las reproduce sin cambio. Ha modificado el viejo material para que sirva a unos párvulos que no lo son ni por la edad ni por la inteligencia, pero que sí lo son por su actitud humilde frente al conocimiento.*

O sea, que estas páginas están dedicadas a quienes ni pretenden que lo saben todo ni fingen que todo lo ignoran: a quienes saben que el conocimiento científico se consigue, más que por acumulación, por intususcepción. Han sido escritas para quienes saben que conforme *crece el árbol y multiplica sus ramas, más hondamente debe hacer que penetren sus raíces si ha de lograr con esto mayor sustentación y mejor sustento.*

En estadística, la semilla, que encierra entre sus cotiledones la raicecilla es el cálculo de promedios. Lo es, hasta tal punto, que así como "el que conjuga y declina sabe la lengua latina", quien sabe promediar y entiende lo que es promediar, tiene abiertas las puertas de la estadística.

El autor pretende —así— que el estudioso vislumbre el cuerpo de arquitectura al que la promediación sirve de piedra angular; que el estudiante ubique, en un campo metodológico, las técnicas que aprende; que el hombre de cultura general aprecie a qué propósitos sirve, desnudamente, la estadística.

Más que desarrollar un automatismo estadístico, intenta despertar la sensibilidad estadística de quienes —como estudiosos de las ciencias sociales— sienten el apremio de someter a cuantificación los fenómenos que estudian, así tengan que reconocer que el enfoque cuantitativo es —en ciencias sociales— simultáneamente indispensable e insuficiente.

En busca de la sencillez, se eliminaron de este libro algunos materiales de las Técnicas estadísticas *capaces de difumar el diseño con el arabesco. En ánimo de completud, se desarrollan aquí aspectos que ahí aparecían embrionariamente. Con deseo unificador, los que eran apéndices ocupan su sitio y forman parte proporcionada del texto.*

Dentro de esa orientación, incluye nuevos ejemplos que gravitan más sobre la secuela conceptual que sobre la rutina manipulativa de signos. Los viejos ejemplos —en cambio— permiten, también, la adquisición de las rutinas ineludibles.

Todo ello responde al convencimiento de que para avanzar en cualquier estudio, hay que enriquecer contenidos y simplificar formas, particularmente si se quiere servir a quienes cuentan simultáneamente entre los más necesitados y los más ambiciosos.

Ojalá y este libro encuentre los lectores que tiene en potencia. Aunque éstos no lleguen a millones y ni siquiera alcancen a ser millares, si los encuentra, si les es útil, habrá satisfecho los anhelos de

El autor

México, D. F., marzo de 1970

AGRADECIMIENTOS

El autor agradece al doctor Pablo González Casanova, director del Instituto de Investigaciones Sociales de la Universidad Nacional Autónoma de México, la simpatía con que vio la reestructuración de unos materiales publicados antes (gracias a la benevolencia del maestro Lucio Mendieta y Núñez) con fines de creciente rigorización.

Agradece a quienes en diferentes formas y ocasiones le estimularon con comentarios y frases de aprecio hacia *Técnicas estadísticas* obligándole así —por la vía más amable— a hacerse digno de su interés, mediante la reestructuración que hoy presenta: al licenciado Jorge Martínez Ríos, antiguo discípulo y hoy distinguido colega; al licenciado Agustín Cue Cánovas, inapreciable amigo, al doctor Jiri Kolaja, de su intelectual aprecio, al doctor Nathan Keifitz, benévolo juez; a Adolfo Santone, estadígrafo argentino; a quienes agotaron la edición de *Técnicas* y, en general, a cuantos —de una u otra manera— le hicieron pensar que valía la pena realizar un esfuerzo adicional y verter el material antiguo en un nuevo molde, más idóneo para las necesidades presentes y futuras.

Reitera aquí su agradecimiento al maestro José Gómez Robleda, su iniciador en la estadística, a la maestra Enriqueta Baz de Velarde, de cuya cátedra derivó estímulos y enseñanzas, a la licenciada Lucila Flammerand que, en la ocasión anterior, sostuvo con su esfuerzo la elaboración de los ejemplos; a la señorita María Teresa Aznar, bibliotecaria del British Council en México, gracias a cuya ayuda consultó y aprovechó la enseñanza de importantes obras británicas dedicadas a la estadística.

Al licenciado Jorge Moreno, secretario del Instituto de Investigaciones Sociales, le agradece la buena disposición y la eficacia con que le ha proporcionado medios administrativos adecuados, para la realización de sus trabajos.

A los anteriores quiere agregar su agradecimiento para la señorita Margarita Sánchez Vélez, quien copió cuidadosa y diligentemente el nuevo original y supo encontrar medios de suplir —con una máquina de escribir ordinaria— la falta de varios símbolos matemáticos empleados en estadística.

A todos ellos, el agradecimiento sincero de

ÓSCAR URIBE VILLEGAS

O. LA INVESTIGACIÓN SOCIAL COMO MARCO DE LA INVESTIGACIÓN ESTADÍSTICA

Definición de la investigación social

La investigación social es un proceso, mediante el cual el investigador trata de encontrar los elementos: 1) determinantes; 2) concurrentes; 3) influyentes, y 4) presentes, que intervienen: *a*) en un problema o *b*) en una situación social, así como sus relaciones, con objeto de plantear y resolver el problema o conocer la situación y mejorarla, de acuerdo con: 1) valoraciones y 2) finalidades, claramente definidas.

A los elementos determinantes se les suele llamar "causas" o "precipitantes"; a los concurrentes, se les designa también como "con-causas".

Etapas de la investigación social

La investigación social, en cuanto *proceso*, está constituida por una serie ordenada de etapas. Pueden distinguirse como principales etapas de toda investigación social:

1. Su planeación.
2. Su realización.
3. Su obtención de conclusiones.
4. La sugestión de soluciones y medios de mejoramiento.

La planeación de la pesquisa o investigación implica problemas epistemológicos, metodológicos y —apendicularmente— técnicos.

La realización de la misma plantea problemas técnicos y prácticos de experiencia y destreza en el manejo de ciertas técnicas.

La obtención de conclusiones equivale a un planteamiento científico del problema o a un conocimiento científico de la situación que, antes de la investigación, sólo se hacía en forma intuitiva o impresionista.

La sugestión de soluciones y medios de mejoramiento implica una enumeración de posibilidades alternativas de resolución del problema o de mejoramiento de la situación, afectadas de índices de factibilidad y riesgo.

La elección de una de esas soluciones o de uno de esos medios representa una problemática político-social de elección de los resultados más deseables, de determinación de los riesgos aceptables, de búsqueda de los medios más idóneos para conseguir tales fines.

En última instancia, en este último sector que no es ya el del sociólogo o el investigador social (dedicado a la pesquisa pura o a la aplicada), se plantean problemas de ideología.

Planeación de la pesquisa

La planeación de una investigación comprende, por lo menos, los tres elementos siguientes:

1. Un elemento de anticipación o previsión.
2. Un elemento volitivo y de decisión.
3. Un elemento representativo de captación.

El elemento anticipatorio, de previsión, en cuanto la planeación trata de prever las situaciones a que tendrá que enfrentarse la pesquisa para poder señorearlas: eludirlas si su dificultad es extrema, amortiguar una dificultad seria e ineludible, o preparar los medios necesarios para resolver cualquier otra dificultad.

El elemento volitivo, de decisión, consiste en que la planeación trata de decidir qué será lo mejor (cuando hay alternativas) para evitar dificultades y obtener éxito, para lograr éxito, superando las dificultades que no conviene eludir.

El elemento anticipatorio y el volitivo van subseguidos de uno representativo, de captación, 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 habrá de seguirse, cuáles son las conexiones entre los diferentes elementos, etapas y operaciones del proceso, y para poder valorar —primero— el plan en su totalidad, o criticarlo, y para vigilar —después— la realización correcta y oportuna de sus diversas etapas.

Planear una investigación equivale —así— a anticipar situaciones y tomar decisiones que, consignadas por escrito o en diagramas, marquen los pasos del proceso y permitan valorarlo en su totalidad y en sus partes y vigilar su realización.

Una investigación está planeada metodológicamente, cuando 1) se han estudiado y valorado de antemano sus posibilidades en total y en detalle, y 2) se puede enjuiciar el procedimiento por el que se han obtenido ciertos resultados.

Primacía y sitio de la planeación

En toda pesquisa científica, la planeación es primordial; es de primera importancia. Sin embargo, ¿debe ser la primera operación de una pesquisa? El problema se resuelve si se considera que: *primero*, es la observación más o menos burda —cuando no un clima de inquietud social— lo que hace que el investigador sienta que hay un problema; *segundo*, la planeación permite que el investigador oriente su observación, determinando si el problema existe realmente, y en qué términos se plantea y *tercero,* que es la observación orientada y afinada (complementada incluso por la experimentación) la que le da al investigador el conocimiento de la situación, o la que le permite plantear científicamente el problema. Esto pone de relieve la importancia de la planeación pero, también, el orden que le correspende en la secuela de operaciones investigadoras.

La planeación como orientadora de la observación

La observación, científicamente vigilada y señoreada, no puede existir sino a partir del momento en que decidimos por lo menos cinco cuestiones: qué, cuándo, dónde, cómo y por qué vamos a observar algo (a ellas hay que agregar frecuentemente, sobre todo si se trata de problemas sociales, el para qué).

Qué estudiar

La determinación de qué es lo que hay que estudiar deberá hacerlo la filosofía social y la sociología, en cuanto al tema de estudio es, o una situación o un problema social; para ello hay que:

1º Señalar la forma en que la unidad o el problema social particular de que se trate puede aislarse del complejo social del que forma parte. Deberá indicar también cómo puede hacerse esta separación en forma *racional* (más que "natural") y no arbitraria.

2º Indicar cuáles son los aspectos *elementales y relacionales* que deberán estudiarse, mediante un cuidadoso análisis de la unidad o el problema en estudio. Si se trata de una unidad social, los individuos, los grupos y sus relaciones cuentan —sobre todo— para el análisis. Si se trata de un problema, girará —pricipalmente— en torno de los valores, las actitudes, las opiniones pero, también, se relacionará con los individuos, los grupos y sus relaciones.

Cuándo y dónde investigar

El "cuándo" y el dónde", o sean las determinaciones de lugar y tiempo (determinación por punto) y de delimitación espacial y temporal (determinación por ámbito) dependerán tanto de lo anterior como de: 1) la amplitud buscada en vista de consideraciones teóricas, 2) las posibilidades de tiempo, dinero, experiencia y energía de que se disponga, 3) las técnicas e instrumentos de los que pueda disponerse, 4) los propósitos de la investigación, y 5) la precisión que busquen los resultados.

Cómo investigar

Cómo hacer el estudio de una situación o problemas sociales ya delimitados teórica, espacial y temporalmente, es, sobre todo, problema de elección de métodos y técnicas. Puede optarse, en efecto, por un método predominantemente inductivo o por uno predominantemente deductivo (aunque al fin de cuentas ambos tengan que usarse). Hay que determinar, también, hasta qué punto hay que reducirse a observar simplemente, a emplear técnicas de observador participante, de participante que observa o de experimentador (caso en que la intervención es activa y el investigador anticipa sus propias observaciones y las comprueba o reprueba). Asimismo hay que determinar hasta qué punto conviene emplear criterios clasificatorios o criterios tipificadores (la clasificación obedece por lo general a criterios simples, en tanto la tipificación se hace sobre la base de un criterio complejo o de la conjuntación de varios criterios).

Colateralmente conviene determinar la utilidad que en su caso concreto pueden brindar —como *auxiliares* y sin desplazar la específica metodología sociológica— la metodología histórica, geográfica, antropológica, etnográfica, psicológica, estadística, económica, lingüística, literaria o filológica, etcétera.

Los grandes problemas de elección de método conducen a los problemas subordinados de elección de técnica.

Por qué y para qué investigar

El realce que demos a ciertos aspectos y, no a otros, de la situación o el problema que se investigue dependerá de los motivos de la investigación y de sus finalidades. Los propósitos de resolver el problema o mejorar la situación, arrojarán más tarde cierta luz sobre los resultados de la misma. Al hacer explícitos los "por qués" y los "para qués" de la pesquisa, el investigador brinda a otros investigadores la posibilidad de juzgar sus resultados y determinar si, por una parte, son comparables con otros resultados obtenidos por él u otro investigador en relación con esa misma situación o problema, un aspecto pasado de los mismos, o una situación o problema análogos que se produzcan en otra sociedad o grupo.

Cuando las razones o motivos y las finalidades de una investigación social no se vuelven explícitos, los resultados mismos son anfibológicos, pueden decir cosas muy distintas, según el marco en el que se les coloque. Por otra parte, y en concreto, lo que en ciertas ocasiones puede parecer cambio social en una comunidad puede no serlo, si las diferencias entre dos situaciones no son reales sino puramente aparentes, en cuanto se obtuvieron por métodos diferentes, sujetos a motivaciones y finalidades distintas. Las diferencias pueden ser el resultado de una investigación, por dos sociólogos distintos, de una misma situación, en la que cada uno de ellos haya hecho una *selección diferente* de ciertos aspectos de la realidad estudiada, o en la que cada uno de ellos haya dado diferente importancia o peso a unos sobre otros aspectos, de entre los seleccionados para el estudio de esa realidad.

Realización de la pesquisa

En toda investigación social conviene distinguir:

1. Sus elementos personales.
2. Sus operaciones.
3. Sus fases.

Elementos personales de la investigación

El investigador social necesita recordar que la pesquisa no es social sólo por el carácter del objeto por investigar (la sociedad o sus grupos o la comunidad internacional), sino que la investigación social misma plantea un problema social en cuanto es un conjunto de situaciones en las cuales intervienen diferentes elementos personales, ligados entre sí por relaciones interhumanas y sobre los que gravita el peso de toda una serie de convenciones desarrolladas por la convivencia humana. Olvidar esto conduce al fracaso del investigador social.

Los elementos personales de la investigación social son:

1. El consumidor.
2. El investigador científico.
3. El investigador técnico.
4. El investigado.

El consumidor es aquel que tiene necesidad de conocer en forma imparcial, realista, segura, precisa, una situación o problema social y que incapacitado por sus prejuicios o intereses, su impresionismo, y su atecnia está impedido para satisfacer por sí mismo esa necesidad de conocimiento.

El investigador científico es aquel que relaciona ciertos conocimientos de orden teórico con ciertas necesidades de orden práctico, a fin de aproximarse a esa realidad y conocerla y con el fin de derivar de ella nuevos conocimientos teóricos, metodológicos y prácticos que utilizar ulteriormente.

El investigador técnico es aquel que recoge directamente sobre el terreno o sobre el documento los datos pertinentes para comprobar o reprobar las hipótesis que haya planteado el investigador científico, a partir del conocimiento sociológico previo. En ciencias sociales, el investigador científico es, con máxima frecuencia, un observador (más que un experimentador) y un recolector de datos (encuestador, censador, entrevistador, etcétera).

El elemento investigado (raras veces una persona física, en la mayoría de los casos un grupo, una sociedad, la humanidad) es sólo en parte *objeto* de la investigación pues, en buena parte —aunque en medida variable según toda una serie de circunstancias— reacciona como *sujeto* a los estímulos introducidos por el investigador científico (estímulos entre los que se cuentan los propios investigadores técnicos).

Hay ocasiones en que el investigador científico es, simultáneamente, su técnico (particularmente si la investigación no es grande); en menor número de casos puede ocurrir que el investigador (técnico y científico) se confunda con el consumidor (porque sea un científico dispuesto a intervenir en una actividad político-social, o porque haya un político que tenga capacitación científica para descubrir la realidad que quiere transformar). La coincidencia entre el elemento investigador y el investigado es más compleja: ni siempre existe ni es siempre aconsejable o desaconsejable que exista. La coincidencia no puede ser simple: porque el investigador (especialmente el científico) es generalmente un individuo y el elemento investigado es, por lo general, un grupo; porque el individuo investigador puede pertenecer o no pertenecer al grupo que investiga (lo que en cada caso plantea problemas metodológicos distintos); porque incluso en caso de que el investigador sea un grupo (un "equipo" como se dice ordinariamente) es difícil que ese grupo se investigue pura y simplemente a sí mismo.

Operaciones de la investigación

Desde el ángulo sociológico, se puede establecer una equivalencia entre el rubro "operaciones de la investigación" y el de "relaciones interhumanas" de la investigación social ya que, generalmente, en cada operación intervienen por lo menos dos de las personas que forman parte de la misma (consumidor, investigadores, investigados).

Desde el ángulo práctico, el éxito o fracaso de la pesquisa depende de las operaciones de la investigación; de que se establezca una distinción clara entre cada una de ellas, así como una secuencia pertinente.

Del adecuado análisis de las operaciones de una investigación, dependerá el que, en algunos casos, se encuentre el punto en que se erró, la operación que hizo fracasar la pesquisa en su conjunto. Con esto muchas veces podrá rehacerse o corregirse sólo esa parte y evitar que se invalide toda la investigación. En una situación menos buena que la anterior, habrá que rehacer sólo aquellas operaciones que subsigan a la operación equivocada o inadecuada y, en el peor de los casos, será posible rehacer toda la investigación o compararla con la que se dañó en forma tal que los resultados obtenidos permitan descubrir cuáles son los factores que dañan la pesquisa, a fin de tomarlos en consideración —metodológicamente— al realizar investigaciones ulteriores.

Las operaciones principales de una investigación social son:

1ª Planteamiento, lo hace el consumidor, quien lo trasmite al investigador científico.

2ª Replanteamiento científico e implementación técnica que hace el investigador científico, quien los trasmite a los investigadores técnicos, mediante instrucción, entrenamiento, alertamiento, vigilancia.

3ª Estímulo de lo que se ha de investigar, que realizan los investigadores técnicos, por sí mismos (estímulo subjetivo consciente e inconsciente), y mediante el uso de ciertos instrumentos (estímulo objetivo).

4ª Respuesta a los estímulos de la investigación, dada por quienes son investigados.

5ª Registro de las repuestas, que realizan los investigadores-técnicos y que trasmiten al investigador-científico.

6ª Elaboración de los datos y redacción del informe acerca del problema, junto con las sugerencias alternativas que para resolverlo brinda el investigador científico al consumidor.

Gracias al informe el consumidor conocerá el problema en sus términos precisos, así como las soluciones alternativas propuestas por el científico social. A partir de ellos, habrá de seleccionar un curso de acción: una de las varias soluciones que se le ofrecen, considerada como mejor, tanto con base en los criterios científicos (factibilidad, riesgo) como con base en criterios extra-científicos que él reintroduce, políticos y de otro tipo.

Aun cuando en este punto no podemos tratar el tema, se debe indicar que, gracias al desarrollo de una moderna rama de la estadística, el consumidor puede recurrir al científico para que le indique (basado en técnicas como la "hechura de decisiones", o la programación lineal) cuál de los propuestos es el mejor curso de acción (el óptimo); cuál es el más económico, de mayor rendimiento dados ciertos medios, el menos riesgoso.

Fases de la investigación

En cuanto proceso, la investigación está constituida por una serie de fases. De ellas, las principales son:

1ª Reconocimiento de que el problema existe realmente. En esta primera fase es indispensable determinar si no se trata de un problema ficticio, de una construcción artificiosa, de un mero fantasma.

2ª Formulación del problema. Esta etapa implica la aceptación de ciertas premisas: un problema social se plantea dentro de determinado marco social. Existe sólo en el grado en que se aceptan ciertos valores y se rechazan otros. Y si bien, en términos muy amplios, la formulación debe hacerse en términos de valores humanos compartidos o compartibles por toda la humanidad, en términos más restringidos, la formulación debe considerar las condiciones reales de la sociedad en que el problema se plantea y ha de resolverse.

3ª Planeación metódica de la investigación. Esta fase abarca tres etapas: 1. Planeación ideal o teórica de la investigación; o sea, la determinación de la extensión, profundidad, exactitud apetecibles. 2. Planeación realista o práctica de la investigación; o sea, el conjunto de ajustes que hay que hacer al plan ideal, en vista de las limitaciones que impone la realidad. En esto hay que distinguir entre obstáculos *difíciles* de salvar y limitaciones *superables*, por una parte y dificultades *insalvables* y limitaciones *insuperables*, por otra. Entre unos y otros existe una frontera fluctuante; el sitio preciso de ésta, en cada caso, dependerá de la experiencia y entrenamiento previos de los investigadores y de su brío y tenacidad (que no deben confundirse con ceguera y obstinación) para sacar adelante un proyecto. 3. Diseño estadístico. Mediante el diseño estadístico se especifica ya, con base en la conjugación de los dos criterios anteriores, el grupo de personas por estudiar que, para fines estadísticos, se convierte en un "colectivo"; en una colección de individuos y de relaciones que poseen ciertos rasgos o conjuntos de rasgos. Dentro del diseño estadístico hay que precisar: *a*) el número de los que se seleccionarán, *b*) el criterio con el que se elegirán, *c*) los rasgos que se han de estudiar. 4. Plan de observación. Mediante éste: *a*) se determinará en qué condiciones se harán las observaciones, *b*) se señalará qué precauciones se deberán tener al hacerlas. 5. Plan de operaciones. A éste corresponde la instrumentalización y tecnificación de la pesquisa misma. Deberá fijar cuál es el instrumental de recolección de datos que deberá utilizarse pero, en forma no menos importante deberá tener en consideración cuáles son las elaboraciones a que se piensa sujetar los datos que con él se obtengan. De ahí que cada instrumento deba obedecer a las dificultades y resistencia que haya que vencer para obtener el dato, y a los resultados que su elaboración deba proporcionar para la comprobación o reprobación de ciertas hipótesis teóricas.

4ª Realización de la investigación. Esta etapa comprende la recolección misma de los datos sobre el terreno, su crítica sistemática, su tabulación, su elaboración, su interpretación, su uso en la comprobación o reprobación de las hipótesis planteadas al formular el problema.

1. OPERACIONES ELEMENTALES DE LA INVESTIGACIÓN ESTADÍSTICA

La amplitud, complejidad y desarrollo alcanzado por la metodología estadística hace que el investigador olvide a menudo que esa complejidad 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 y las operaciones elementales —por una parte— y los métodos y las necesidades que tratan de satisfacer —por otra— impiden que, en cuanto se plantea una necesidad nueva, se puedan encontrar medios estadísticos —nuevos— para satisfacerla, a base de combinar las operaciones elementales conocidas. Con ello, su actividad se estanca y vuelve infecunda o, en el mejor de los casos, se desarrolla, pero dentro de cauces que pueden conducir al fin buscado sólo a largo plazo.

De ahí que sea importante —dentro de una labor analítica— señalar cuáles son las diferentes operaciones elementales de la investigación estadística e indicar las necesidades que satisfacen, así como sus limitaciones.

Algunas de esas operaciones se han mencionado ya en el apartado anterior; pero su tratamiento —ahí— era genérico y ahora se buscará —en cambio— que sea específico y detallado.

Jerarquía y secuela de las operaciones

Las operaciones estadísticas elementales tienen una jerarquía y secuelas que no hay que olvidar, pues, si bien es cierto que algunas de dichas etapas (como puede ocurrir con la "definición" o la "recolección de datos") se mencionan más brevemente, esto no significa que tengan menor importancia que aquellas cuya mención es obligadamente más extensa como ocurre con "elaboración estadística" en sentido estricto).

Es hacia esto hacia lo que hemos querido llamar la atención mediante el uso de números enteros y fraccionarios, en la siguiente lista. En ella, los enteros se emplean para las operaciones principales y los decimales (décimos, centésimos, milésimos, en su orden) para las subordinadas de primero, segundo, tercer orden, etcétera.

Esta necesidad es tanto más evidente en cuanto se considera que la idea central de "proceso" es característica de toda investigación social y que la misma se refleja en forma importantísima en la investigación estadística que en ella se encuadra. En efecto, una deficiente definición del objeto —en cuanto inicial— puede derrumbar la investigación en su totalidad, aunque se hayan cuidado mucho las siguientes etapas, y puede darse el caso de que haya investigaciones cuyos resultados —válidos en apariencia— carezcan totalmente de sentido, por haber sido obtenidos de un conjunto de datos deficientes o recogidos en forma descuidada, o de un colectivo delimitado vagamente, más heterogéneo de lo debido, etcétera.

En este sentido, es tan importante la definición que podría hablarse, en justicia, de tres grandes etapas de la investigación estadística: la definitoria, la elaborativa y la interpretativa que —nuevamente— se correlaciona con las definiciones iniciales y, por decirlo así, cierra el círculo.

Lista de operaciones

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

1. Definición.

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

   1.2. Del "universo", "población" o "colectivo" por estudiar (o sea, del grupo de objetos, personas o relaciones que abarcará la investigación estadística).

   1.3. De los rasgos por estudiar en dicho colectivo.

   1.4. Del grado de homogeneidad deseable o de heterogeneidad permisible en el universo, por lo que se refiere a los rasgos distintos de los que se desean estudiar particularmente, o en lo que respecta a éstos (ya que uno de los aspectos que precisamente importa estudiar es su heterogeneidad o variabilidad).

   1.5. De las unidades por emplear en la medida.

2. Decisión censo-muestreal.

   Es decir, decisión de si conviene hacer:

   2.1. Una investigación exhaustiva, de tipo censal, o

   2.2. Una investigación restringida, representativa, de tipo muestreal y, en este último caso, elegir:

      2.21. Un marco de referencia
      preciso,
      completo,
      adecuado,
      actual.

2.22. Un tipo de muestreo
  2.221. simple,
  2.222. estratificado,
  2.223. por racimos.

2.23. Una forma de muestreo
  2.231. aislado,
  2.232. en dos etapas,
  2.233. en múltiples etapas,
  2.234. repetido periódicamente.

2.24. El tamaño de la muestra, en vista de requisitos de:
  2.241. representatividad,
  2.242. precisión de los resultados,
  2.243. economía en el costo de las operaciones.

3. Recolección de los datos.

3.1. Directa, mediante:
  3.11. Cuestionarios y cédulas.
  3.12. Tarjeteros y escalas de fines sociométricos.

3.2. Indirecta.
  3.21. De fuentes primarias recolectoras de datos, como la Dirección General de Estadística, sus publicaciones y sus datos primarios no publicados.
  3.22. De fuentes secundarias.

4. Elaboración estadística.

4.1. Operaciones predominantemente críticas:
  4.11. Se completan los datos faltantes cuando esto es posible en vista de lo disponible, cuando se pueden realizar investigaciones complementarias, cuando se puede recurrir a otras fuentes de información.
  4.12. Se determinan las inconsistencias entre los datos de una misma cédula o cuestionario, y se corrigen.
  4.13. Se rechazan los datos, las respuestas de cédulas o cuestionarios, cuyas inconsistencias no puedan salvarse en forma que no ofrezca duda.
  4.14. Se rechaza todo el material obtenido en caso de que el porciento de cuestionarios con inconsistencias sea muy alto, pues esto califica el valor de los instrumentos con los que se obtuvieron las respuestas.

4.2. Operaciones predominantemente mecánicas
  4.21. Conjuntación de los materiales.
  4.22. Ordenación, mediante la formación de arreglos o series.
  4.23. Clasificación:
    4.231. dicotómica,
    4.232. multicotómica.
  4.24. Consignación:
    4.241. textual,
    4.242. tabular,
    4.243. gráfica.

4.3. Operaciones predominantemente matemáticas
  4.31. Condensación:
    4.311. mediante el recuento del número de veces que aparece un dato (o sea, mediante la determinación de su frecuencia).
    4.312. mediante la constitución de "casilleros" en los que colocar los datos (o sea, mediante la constitución de clases definidas por valores que les sirven de límites "inferiores" y "superiores") y entre los que existe un "intervalo" igual o diferente para toda la extensión de la serie.
    4.313. mediante el uso del operador sigma, que indica que todas las expresiones del tipo de la que siga a sigma mayúscula (Σ) deben sumarse.
  4.32. Relievamiento de características por descripción estadística al través de:
    4.321. cálculo de promedios centrales, como las medias, la mediana y el modo, laterales, como las cuartilas, decilas, porcentilas.
    4.322. medidas de variabilidad,
    4.323. medidas de asimetría,
    4.324. medidas de aplanamiento o picudez (curtosis).
    4.325. ajuste de curvas, mediante diferentes sistemas como:
      4.3251. el de Gram-Charlier,
      4.3252. el de Pearson.
  4.33. Reducción a términos comparables, mediante:
    4.331. obtención de relativos,
    4.332. obtención de tantos porciento o porcentajes,
    4.333. expresión de resultados en términos de cuantila (un valor está comprendido entre la cuantila de orden "tanto" y la de orden "tanto más uno" o sea, la siguiente),

4.334. expresión de datos o resultados en unidades del promedio de las desviaciones (con máxima frecuencia en "unidades sigmáticas" que, con todo, es uno de los casos posibles de uso de un promedio de desviaciones),

4.335. reducción de medidas n-dimensionales a medidas uni-dimensionales por división entre las potencias correspondientes del promedio de las desviaciones (un resultado como la suma del *cubo* de las desviaciones, que es *tri*-dimensional se vuelve o-dimencional, al dividirla entre el *cubo* o *tercera* potencia del promedio de las desviaciones).

4.34. Ponderación, o reconocimiento matemático de la importancia que cada dato juega dentro del conjunto, mediante:

4.341. multiplicación por un "factor de ponderación" (según ocurre en el cálculo de las medias ponderadas y en el de los números índice, donde los índices de cantidad de un producto se ponderan por su precio o los de precio por su cantidad),

4.342. elevación a una potencia igual al factor de ponderación (que es básicamente lo mismo que el anterior, según se pone de manifiesto cuando se usan logaritmos).

4.35. Análisis o descomposición de un movimiento en varios movimientos componentes.
Este análisis (que se aplica particularmente en el caso de las series temporales, dinámicas o cronológicas) se logra:

4.351. identificando los movimientos más amplios o notables,

4.352. restando, los ya identificados, del movimiento que se analiza (o dividiendo éste entre aquéllos).

4.353. identificando movimientos menos amplios (en lo que resta). Los pasos anteriores identifican:

4.3511. movimientos permanentes: tendencia secular,

4.35111. rectilínea,

4.35112. parabólica de segundo grado,

4.35113. parabólica de tercero u otros grados,

4.35114. hiperbólica,

4.35115. logarítmica,

4.35116. exponencial, etcétera.

4.3512. movimientos oscilatorios:

4.351211. periódicos, variaciones estacionales (o sean las que se producen en el curso de un año) variaciones cíclicas de alta o de baja frecuencia,

4.351212. no periódicos, variaciones accidentales, fluctuaciones.

4.36. Síntesis o recomposición del fenómeno previamente analizado, para comprobar que el análisis ha sido adecuado; mediante sumas y productos (operaciones inversas de la resta y la división) de los elementos componentes, y mediante cálculo de los valores teóricos de una serie y determinación de los errores probables y las zonas de estimación.

4.37. Inferencia estadística.

4.371. mediante la determinación de la probabilidad de que ocurra un hecho (en relación con la adaptación de curvas probabilitarias como las binomiales, de Poisson o hipergeométricas o con las que de ellas resultan, según es el caso de las normales, las de Pearson y otras semejantes).

4.372. mediante "extrapolación" en series temporales; o sea, mediante el cálculo de lo que ocurrirá en el futuro, en caso de que se mantenga cierto ritmo de crecimiento o decrecimiento de un fenómeno y se repitan, como en el pasado, ciertas variaciones periódicas,

4.373. mediante asociación de series y determinación de:

4.3731. la intensidad de la asociación para determinar si dos o más series están muy vinculadas entre sí, si tienen vínculos lejanos o no tienen vínculo alguno,

4.3732. el sentido de la asociación para determinar si al crecer una serie crece la otra y al decrecer, decrece, o si al crecer la primera decrece la segunda y al decrecer la primera crece la segunda,

4.3733. la forma matemática de la asociación o sean las llamadas "ecuaciones de regresión" que expresan por qué valores numéricos hay que multiplicar los valores de una serie y qué otros valores numéricos hay que sumar a los resultados (en el ejemplo más simple) para obtener los valores correspondientes de la otra serie, 4.3734. determinación del retraso con que los datos de una serie aparecen, después de que los similares de otra han hecho su aparición,

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

4.3741. el estudio de la forma en que una determinada medida estadística (la media, por ejemplo) se distribuye en el total de las muestras posibles (se distribuye normalmente, en el caso) para determinar qué probabilidad hay (conforme a esa forma de distribución) de que la medida calculada para la muestra sea la medida del universo. Si la probabilidad es alta, se dice que la media de la muestra puede aceptarse como media del universo; si es baja, se rechaza como media del universo,

4.3742. pruebas o docimasia de la hipótesis de que los valores de ciertas constantes encontradas para la muestra (estimadores) correspondan a los valores de las constantes de la población dentro de ciertos límites de confianza previamente fijados.

Como es fácil comprender este es sólo otro aspecto de lo anterior.

1.1. ANÁLISIS PREVIO A LA MEDICIÓN ESTADÍSTICA

En el terreno de las ciencias sociales las unidades por estudiar son, frecuentemente, unidades complejas. Como tales, no se les puede sujetar a medición estadística en forma inmediata. Esto impone la necesidad de sujetarlas a un análisis antes de intentar dichas mediciones. El análisis de la unidad socio-económica compleja tiene que dar como resultado un conjunto de "componentes" principales, cuyo orden de complejidad sea menor que el de la unidad original. La medición —con todo— no es posible sino cuando el análisis se ha llevado hasta un punto tal que se han llegado a identificar componentes suficientemente simples (unidimensionales, de ser posible). Por razones de orden práctico, no es posible medir todos los componentes de orden más bajo que haya identificado el análisis; es por ello por lo que se hace una selección de ellos, aceptándose sólo aquellos que, para el componente de orden inmediato superior, sean sintomáticos. A estos componentes de orden más bajo y de carácter sintomático, susceptibles ya de medida, es a lo que se conoce como "indicadores".

A fin de mostrar cómo opera, en la práctica, el procedimiento de los componentes e indicadores (o sea, el análisis previo a la medición estadística de una unidad socio-económica compleja) nos referiremos a la medición internacional de los niveles de vida tal y como ha sido propuesta por la Organización de las Naciones Unidas.

Análisis de un complejo estadístico-social: el nivel de vida

Distinciones conceptuales: nivel, estándar y norma de vida

Internacionalmente, se ha reconocido la conveniencia de distinguir los tres conceptos siguientes: 1) nivel de vida, 2) estándar de vida, y 3) norma de vida.

El nivel de vida indica las condiciones reales de existencia de una colectividad humana. El estándar de vida designa aquellas otras que el grupo cree deseable, posible y justo alcanzar. La norma de vida se refiere a las condiciones que se consideran convenientes en relación con fines determinados y que tratan de asegurarse mediante acuerdos o convenios nacionales o internacionales (y que se refieren, por ejemplo, a niveles de salarios, horas de trabajo, etcétera).

Criterios analíticos: sectores del "nivel de vida"

El concepto de "nivel de vida" para ser humanamente significativo, internacionalmente utilizable y de utilidad práctica, debe relacionarse con la satisfacción de necesidades y el logro de aspiraciones muy diversas. Abarcan éstas desde las puramente materiales (de bienestar físico, posibilidad de consumir, etcétera), a las de carácter inmaterial (como las culturales y educativas, o las que se refieren al ejercicio de los derechos políticos).

Pero, aunque conceptualmente el nivel de vida deba de integrarse no sólo por los factores materiales sino también por los no materiales, éstos, generalmente, son más difíciles de manejar, o más fáciles de mal interpretar. Así, si bien puede tenerse un índice del disfrute estético mediante los datos sobre asistencia a los museos y a las galerías artísticas, éste resultará insuficiente, ya que no incluye las motivaciones de quienes a ellos asisten. En forma parecida, si bien el número de votantes es un indicador político, su significación dependerá de que se precise cuántos de quienes votan lo hacen por propia voluntad y cuántos lo hacen por evitar el que —de dejar de hacerlo— se les imponga una multa.

Hay, como se ve, una mayor dificultad práctica en la medición de los aspectos no materiales del nivel de vida que en el de los aspectos materiales. Pero, en forma no menos acusada, existe la necesidad teórica de incluirlos si se quiere valorar auténticamente el nivel general de vida y no sólo medir el nivel material de vida.

En términos del procedimiento de análisis y medición de las unidades socio-económicas complejas, la dificultad de incluir los aspectos no materiales en la apreciación de los niveles de vida se descompone en dificultades de orden inferior, como las siguientes: dificultad para reconocer los componentes principales de cada uno de esos aspectos; necesidad de profundizar los análisis descendiendo de los de primero a los de segundo y aun a los de orden *n* antes de encontrar unidades que sean medibles en forma directa y que, por lo mismo, puedan funcionar como "indicadores", y necesidad de seleccionar un número mayor de indicadores que el que es necesario para uno de los aspectos materiales, a fin de que cumplan adecuadamente con su función.

Problemas de comparabilidad

En relación con la medida y comparación de los niveles de vida de diferentes poblaciones, hay que señalar que los problemas de comparabilidad dependen de que éstas vivan en diferentes lugares de la tierra, que tienen diferentes climas, y de que han desarrollado diferentes formas de vida y de cultura. La diferencia del medio físico impone, así, una primera relativización a los indicadores del nivel de vida; la diversidad de mentalidades y de matrices valorativas relativizan, asimismo, esos indicadores. De este modo, un mismo indicador puede tener significaciones diferentes para pueblos que viven en dos lugares diferentes o que tienen diferentes formas de organización o diferente cultura; en cambio, indicadores aparentemente diferentes pueden ser si no iguales sí equivalentes, en función del contexto geográfico, social o cultural de los pueblos o colectividades a los que se refieren.

Así, por ejemplo, todos los hombres necesitan mantener una cierta temperatura corporal, pero, el esquimal tiene necesidades de abrigo distintas de las del habitante de las zonas tropicales y la desnudez de éste puede que no signifique una especial privación y un bajo nivel de vida, en tanto que la desnudez de aquél puede representar no sólo privación y penuria, sino un nivel de vida tan bajo que pone en peligro su existencia misma. En forma parecida, las calorías necesarias para un individuo varían de acuerdo con su sexo, con su edad, etcétera.

Los problemas de comparabilidad aumentan en cuanto se consideran no sólo los elementos sincrónicos, sino los diacrónicos: en cuanto se considera que, dentro de cada sociedad, hay cambios en la moda o el gusto (en los valores y en su jerarquización) que cambian o modulan las necesidades básicas y que transforman —a veces completamente— los anhelos o aspiraciones de cada sociedad.

A fin de resolver estas dificultades, en una primera aproximación, 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 han tratado de dar satisfacción los organismos nacionales e internacionales mediante la fijación de objetivos político-sociales.

Componentes e indicadores

El comité llegó a la conclusión de que la forma más satisfactoria de medir los niveles de vida era tomar aspectos o partes claramente delimitados de las condiciones generales de vida, capaces de expresar objetivos internacionales y susceptibles de medición y representación cuantitativa.

Esos aspectos o partes bien delimitados de las condiciones generales de vida (nutrición, salud, vivienda, empleo, enseñanza, etcétera) reciben el nombre de "componentes" del nivel de vida. Para cada uno de ellos se consideran ciertos factores concretos (calorías consumidas, grado de alfabetización, etcétera) que se utilizan en la medición estadística, y a los que se conoce como "indicadores".

El procedimiento de los componentes y el procedimiento monetario

Antes de referirnos más detenidamente al procedimiento de los componentes, conviene que mencionemos otro procedimiento para la medición de los niveles de vida: el procedimiento monetario.

El procedimiento de los componentes difiere sustancialmente del procedimiento que se basa en las cantidades de dinero gastado en el consumo, y se considera que es mejor que éste. En efecto, para valorar el nivel de vida de alguien, importa conocer no sólo cuánto ha gastado, sino en qué lo ha gastado, pues si el gasto se ha destinado a elementos perjudiciales, aunque la persona o el grupo hayan gastado lo mismo o más que antes, habrán hecho descender el nivel de vida.

El procedimiento monetario no carece de interés, pero es mucho más impreciso y laborioso. Entre otras cosas, plantea, por ejemplo, el problema de la convertibilidad de las monedas (que, como se sabe, es, a más de todo, muy variable a través del tiempo). En vista de esa necesidad de convertir unos gastos hechos en unos términos monetarios a otros términos monetarios distintos, se necesita: 1º calcular los gastos de la primera persona o grupo; 2º determinar qué cantidad de dinero necesitaría la segunda persona (en su propia moneda) para adquirir los mismos bienes y servicios o sus equivalentes, y 3º reducir los gastos de la segunda persona a términos comparables con los de la primera, mediante la equivalencia entre las monedas empleadas por una y otra para la adquisición de bienes y servicios iguales o equivalentes.

Tipos de indicador en el procedimiento de los componentes

En el procedimiento de los componentes, los indicadores pueden ser de muy diversos tipos. Por sí mismos, los indicadores pueden ser:

indicadores personales o colectivos; pueden ser micro-indicadores (como el presupuesto familiar) o macro-indicadores (como la renta nacional); pueden ser estáticos (como consumo de calorías por persona en el año tal) o dinámicos (como aumento de la producción agrícola del año tal al año tal).

Pero, no sólo pueden variar los tipos de indicadores de por sí, sino también los tipos de relación de cada uno de ellos con cada uno, con un grupo o con todos los restantes. En efecto, los indicadores pueden ser más o menos independientes de los demás, interdependientes o dependientes; pueden ser (por pares) contradictorios entre sí, pues mientras uno puede indicar un alto nivel de vida el otro puede señalar uno bajo; pueden referirse a medios o a fines (¿el alfabetismo, fin en sí o medio para lograr la elevación técnica y cultural de la población?); pueden referirse a realidades o posibilidades ("número de escuelas" y "número de maestros", por sí solos son indicadores de las posibilidades de instrucción en una colectividad, sólo el "número de alumnos" y, más aún el "número de asistentes" es un indicador de realidad en cuanto a la instrucción de la misma).

Tipos de comparación

Dentro del mismo procedimiento de los componentes es necesario asegurar la comparabilidad y, por ello, es útil emplear sólo un mismo tipo de indicadores o emplearlos distintos, pero con la conciencia de que difieren entre sí, para ponderar la contribución que cada uno de ellos haga al conjunto. Asimismo, conviene tener en cuenta que los diversos tipos de indicadores dan lugar a distintos tipos de comparación, pues puede haber: comparaciones directas de datos individuales (escolaridad, consumo de calorías, etcétera) y comparaciones de datos grupales (tasas de natalidad, índices de analfabetismo, etcétera).

Cuando los indicadores se refieran a cifras individuales, se puede recurrir a los promedios. Pero, como la comparación a base de promedios encierra ciertos peligros (a los que se aludirá más adelante) deben agregarse a ellos ciertas medidas de dispersión (que se estudian en la porción técnica) o la forma en que se distribuye estadísticamente la característica estudiada entre las distintas capas o grupos de la población (véase la sección referente a las distribuciones estadísticas).

Problemas técnicos de medición de los niveles de vida

Los problemas técnicos que se presentan en la medición de los niveles de vida son, en realidad, problemas estadístico-sociales relativos a:

1. La necesidad de obtener informes anuales sobre las variaciones del nivel de vida, lo que implica la necesidad de que avancen las técnicas estadísticas necesarias para recoger, elaborar y presentar los datos correspondientes a 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, por lo general, excesivamente lentas y por ello, los instrumentos de medición deben ser extraordinariamente sensibles a fin de descubrir esos cambios tan pronto como se producen o incluso tan pronto como apuntan.

3. La necesidad de que al interpretar las variaciones anuales se tengan en consideración las alteraciones producidas en la tendencia general del fenómeno por las variaciones estacionales y cíclicas.

4. La necesidad de considerar que los ciclos económicos no tienen ni la misma significación ni idéntico efecto en diferentes sistemas económico-sociales y que, en consecuencia, afectan en diversas formas los distintos indicadores de los niveles de vida.

5. La necesidad de determinar (estadísticamente) los márgenes de error (y, por tanto, los límites de precisión) dentro de los que pueden moverse las cifras relativas a cada uno de los indicadores del nivel de vida.

6. La necesidad de precisar —cuando no se recurre a datos censales sino muestreales— cuál es la parte de la población estudiada que se ha tomado como representativa de toda la población.

Con fines de comparabilidad, resalta la necesidad de uniformizar: definiciones y métodos de recolección, concentración, elaboración e interpretación de los datos. El problema de las interpretaciones uniformes es, con todo, de los más arduos y no corresponde propiamente al sector estadístico social sino que plantea problemas propios de la sociología del conocimiento en cuanto la diversidad de interpretaciones de un mismo dato depende, en buena parte, de la diversidad de intereses socio-políticos y económicos de quienes lo interpretan.

Otros problemas de comparación de niveles de vida

Más allá de los problemas técnico-estadísticos, en la estimación de los niveles de vida, se plantean algunos otros desde los ángulos social y cultural. En efecto, es necesario considerar cuál es el tipo de estructura social en el que se da determinado nivel de vida, para apreciar 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 la que la remuneración se entiende como una entrega periódica de una cantidad determinada de dinero por un servicio prestado y se olvida —en cambio— que hay sociedades en las que los servicios se remuneran mediante contraprestaciones no monetarias (como las manifestaciones de solidaridad entre los co-societarios, etcétera). En tales casos pueden aparecer, en los recuentos muchos "trabajadores no remunerados" y, con todo, el nivel de vida no será tan bajo como parecería indicarlo la cifra correspondiente, pues esas contraprestaciones no monetarias equivalen a dicha remuneración y elevan, por consiguiente, dicho nivel.

Componentes del nivel de vida

El Comité de las Naciones Unidas para la Medición de los Niveles de Vida, solicitó de los organismos especializados de dicha organización una lista de posibles componentes, y en seguida hizo una selección y una síntesis de las diversas listas. Así elaboró su propia lista de componentes del nivel de vida.

Antes de listar los componentes aceptados, conviene mencionar algunos de los criterios que sirvieron de base para la elaboración de las listas primitivas. En este sentido son paradigmáticos los utilizados por la Organización Internacional del Trabajo. Este organismo incluyó ciertos componentes de acuerdo con: 1) la importancia que tienen para el bienestar del individuo (conforme a normas objetivas generalmente aceptadas); 2) el grado en que su falta o deficiencia constituye un problema (en relación con necesidades y aspiraciones); 3) la medida en que esa falta o deficiencia puede remediarse mediante la intervención humana, y 4) la posibilidad de medirlos estadísticamente.

Lista de los componentes del nivel de vida

Los componentes que el Comité de las Naciones Unidas propuso para la medición del nivel de vida son: 1) salud (con inclusión de condiciones demográficas); 2) alimentación y nutrición; 3) educación (con inclusión de alfabetismo y enseñanza técnica); 4) condiciones de trabajo; 5) empleo y desempleo; 6) consumo y ahorro globales; 7) transporte; 8) vivienda e instalación doméstica; 9) vestido; 10) esparcimiento y recreo; 11) seguridad social, y 12) libertades humanas.

Lista de indicadores

Los indicadores del componente "salud" (con inclusión de condiciones demográficas) son: esperanza de vida al nacer; tasa de mortalidad infantil, tasa de mortalidad bruta anual, número de camas de hospital en relación con la población total, número de médicos en relación con la población total. Todos ellos son indirectos; los tres primeros están relacionados con los dos últimos; éstos muestran, además, posibilidades más que realidades.

Los indicadores del componente "alimentación y nutrición" son: promedio nacional de distribución de alimentos al por menor (en calorías); promedio nacional de distribución al por menor (en proteínas); promedio nacional de distribución de alimentos al por menor en proteínas animales; educación, propaganda y legislación relativa a la alimentación.

Así como en el caso del componente "salud y condiciones demográficas", podía señalarse la dificultad que representa comparar la tasa de mortalidad de poblaciones con distinta distribución por edad y sexo, se puede decir algo semejante en relación con las necesidades del consumo alimenticio, ya que éste varía con la edad, el sexo y el peso corporal. Por ello "la distribución de las poblaciones por edad y sexo es un elemento importante, para hacer juicios válidos sobre la suficiencia en el consumo de alimentos".

Los indicadores del componente "educación" son: proporción de niños de entre 5 y 14 años asistentes a las escuelas; proporción de asistentes a escuelas posprimarias en relación con el total de la población y el total de niños entre 5 y 14 años; proporción de escuelas primarias por cien mil niños de esas edades; proporción de alumnos por maestro en las primarias; total, porciento y distribución de alfabetizados a partir de cierta edad; matriculados en instituciones de enseñanza técnica por cien mil habitantes; diarios circulantes por mil habitantes; libros publicados anualmente por cien mil habitantes.

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 veinticinco años de edad o más, acreditados oficialmente". La mediana, como se verá más tarde, es uno de los promedios estadísticos centrales.

Los indicadores del componente "condiciones de trabajo" son: horas de trabajo por semana; salario semanal de los trabajadores industriales; salario real de dichos trabajadores; horas de trabajo por semana establecidas como normales por la ley o los contratos; días de vacaciones anuales pagadas; edad mínima de admisión al trabajo. La principal limitación depende de que los datos de que se dispone se refieren a los trabajadores industriales que no constituyen sino parte de la población trabajadora. Por otra parte, las comparaciones en cuanto a horas de trabajo deben relacionarse con la economía total, pues una disminución de éstas puede representar aumento del tiempo libre de los trabajadores o conllevar incrementos en el desempleo.

Los indicadores de "empleo y desempleo" incluyen: proporción de los económicamente activos en la población (y su distribución); proporción de menores de 20 años que forman parte de la población económicamente activa; proporción de los mayores de 65 años que forman parte de la población económicamente activa, desocupada; proporción de la desocupada respecto del total; distribución porcentual por ocupaciones (empleadores, trabajadores por cuenta propia, miembros de la familia que trabajan sin remuneración, empleados); distribución de empleados por ramas ocupacionales.

En general se considera que existe una correlación inversa entre lo alto de los niveles de vida y la proporción de mayores de 65 y menores de 20 años incluidos en la población, así como en relación con la proporción de mujeres. Sin embargo, aunque exista dicha correlación la misma no es tan intensa en una economía industrial como en una agrícola.

Los indicadores del componente "consumo y ahorro globales" incluyen: proporción del ingreso nacional que se gasta en productos alimenticios; proporción de los gastos públicos invertidos en servicios sociales; proporción entre la inversión en tales servicios y el ingreso nacional; índice de "consumo personal" *per capita*; proporción entre el consumo personal y el ingreso nacional (y las variaciones correspondientes); índice de las inversiones y los ahorros, *per capita* y su coeficiente de variación; proporción entre tal índice y el ingreso nacional.

En este sector, como puede verse claramente, se enfatizan las variaciones en el nivel de vida. Para el estudio de dichas variaciones es necesario trabajar con números-índices (véase la sección correspondiente) que expresan los niveles alcanzados en una fecha dada en términos (generalmente como tantos porciento) de los niveles alcanzados en una fecha base.

Los indicadores del componente "transporte" incluyen: kilómetros de vía férrea por cada cien kilómetros cuadrados de superficie; número de pasajeros-kilómetro anual por cada cien mil habitantes; toneladas-kilómetro de carga anual transportada por cada cien mil habitantes; kilómetros de carretera por cada cien kilómetros cuadrados de superficie; número de vehículos por cada cien mil habitantes; número de pasajeros-kilómetro transportados por aire por cada cien mil habitantes.

En el caso del número de kilómetros de carretera, se recomienda mencionar los diversos tipos de carretera y la longitud correspondiente. En el número de vehículos se deben incluir: automotores, vehículos de tracción animal, y entre los automotores se debe especificar el número de camiones y el de automóviles.

De estos datos, los más difíciles de obtener son los relativos al kilometraje de las carreteras y al número de vehículos. Sin embargo, su obtención no es imposible.

En forma parecida a como se llamó la atención hacia la relatividad cultural en otros respectos, en relación con la diferencia de los medios de transporte, hay que hacer notar que su valoración está relativizada por la geografía de la región. En efecto, aunque un determinado medio de transporte pueda considerarse como indicador de un nivel de vida inferior, en ciertos casos puede que no signifique esto, porque bien puede darse el caso de que las condiciones geográficas del país lo conviertan 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, así como otros más personales (como las bicicletas, las mulas y los caballos) que desempeñan un papel importante en muchos países, tanto en los medios urbanos como en los rurales.

En relación con el componente "viviendas e instalación doméstica" el comité consideró muy difícil establecer indicadores adecuados que permitieran la comparabilidad internacional. El concepto mismo de vivienda no está bien definido, y apreciar el nivel de vida a través de la vivienda se enfrenta a una doble relativización: 1) geográfica (relacionada con requerimientos climáticos de protección contra la intemperie), y 2) sociológica (vinculada con tipos de parientes que habitan una misma vivienda). Sin embargo, el comité propuso indicadores como los siguientes: clases de vivienda y materiales empleados en su construcción; superficie por ocupante; personas por unidad de vivienda; suministro de agua (potable y para otros usos); servicio sanitario y alcantarillado; servicios públicos e instalaciones comunales.

En relación con el componente "vestido" se considera útil incluir también datos sobre sombrero y calzado. Debe de considerarse, sin embargo, que todos ellos se encuentran relativizados por: el clima, el empleo, la situación social del usuario, la tradición y la moda.

Los indicadores del componente "esparcimiento y recreo" pueden incluir, a título ejemplificativo: número de asientos en cines y teatros, por cien mil habitantes; número de aparatos de radio por cien mil habitantes; número de telerreceptores por cien mil habitantes. A estos indicadores hay que agregar descripciones y datos numéricos tan detallados y precisos como se pueda, sobre diversas formas de esparcimiento y recreo de diversas poblaciones, ya que, fuera del mundo occidental, la falta de cines, teatros, radio y telerreceptores podría deformar la visión del nivel de vida de las poblaciones, en caso de que no se considerasen las formas alternativas de satisfacción de las necesidades de esparcimiento y recreo (o de diversión y entretenimiento, de acuerdo con una distinción grata al doctor José Gómez Robleda).

Los indicadores del componente "seguridad social" incluyen principalmente informes acerca de los tipos de seguro (ancianidad, desempleo, etcétera) sobre subsidios familiares y sobre todas aquellas prestaciones que ofrecen los servicios de seguridad social o sus equivalentes. Sin embargo, el comité no llegó a especificar cuáles podrían ser los indicadores pertinentes para este componente del nivel de vida.

Entre los indicadores del componente "libertades humanas" se mencionan: la participación de la mujer en diversas actividades, su situación legal y su situación política.

En todo lo anterior, hemos mencionado los indicadores del nivel de vida más o menos en la forma en que los presenta el Comité de la Organización de las Naciones Unidas. Sin embargo, conviene explorar formas de presentación más ceñidas y analíticas y es lo que nos proponemos hacer en seguida.

Análisis destinado a asegurar la comparabilidad: los censos de población

Problemas generales de los censos de población

La estadística social mundial se enfrenta a varios problemas. Entre ellos, tienen particular importancia metodológica: 1) el de la falta de informes; 2) el de la falta de comparabilidad de los datos. La falta de informes constituye un problema tanto de la estadística demográfica general como de la más especializada, pero se agudiza especialmente en esta última.

Falta de informes. En 1948, 147 de las 245 áreas identificables en el *Demographic Yearbook* de la Organización de las Naciones Unidas eran las que contaban con datos sobre 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 edades, sexos, ocupaciones, ruralidad) la situación era más crítica: sólo 41 de las 245 áreas tenían estadísticas de población por edad, sexo y estado matrimonial, por ejemplo.

Falta de comparabilidad. La falta de comparabilidad entre los datos de que se dispone imposibilita su utilización conveniente por el investigador social o le conduce a conclusiones aparentemente válidas y, en realidad, erróneas.

El problema no es reciente. Ya en 1872, el Instituto Internacional de Estadística (en su reunión de San Petesburgo) señalaba la necesidad de interesarse por los mismos problemas, emplear las mismas definiciones y usar los mismos métodos en la obtención de 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 en todos los países en el año 1900, y su repetición periódica. Resoluciones posteriores —no siempre acatadas por todos los países— señalaron la conveniencia de realizar los censos de población cada diez años, en aquéllos cuya cifra terminara en cero.

La Organización de las Naciones Unidas, al ocuparse de estos problemas solicitó de los gobiernos que le enviaran informe sobre los tópicos y los métodos de su último censo de población. Con base en esos informes se formó una lista de tópicos. En el momento de hacer sugestiones, se vio que dicha lista no podía proponerse como mínimo de información porque los países no tienen ni las mismas necesidades ni idénticas posibilidades de investigación estadística. Esto señaló la utilidad de seleccionar sólo los tópicos más importantes para todos los países.

Sectores que debe cubrir un censo

Lista de tópicos deseables en un censo de población

De acuerdo con la selección hecha por los técnicos de la Organización de las Naciones Unidas, es de desear que todos los censos incluyan los siguientes tópicos:

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, posición o *status* industrial,
    c) población dependiente,
    d) población agrícola.
11. Distribución urbano-rural.
12. Familia y hogar (incluyendo la relación con el jefe de familia).

Para algunos países, aún esta lista puede resultar excesiva; por ello, la Organización de las Naciones Unidas subrayó que aquellos que tuvieran poca o ninguna experiencia censal, deberían concretarse a obtener datos sobre población total, sexo, edad, estado matrimonial y características económicas. Subrayó, sin embargo, que ya fuere que se hiciera esta lista más corta o que tomaran la más extensa, deberían proporcionar datos *completos* y *confiables*.

En las sugestiones y recomendaciones correspondientes, la organización tomó muy particularmente en consideración 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 Conferencias Internacionales (VI y VII) de Estadígrafos del Trabajo (1947-1949).

Como en los censos mundiales de población importan no sólo los datos mismos, sino la forma de definirlos y recolectarlos, clasificarlos y tabularlos, las comisiones estadísticas de la organización subrayaron la importancia de los problemas metodológicos.

Análisis de las definiciones, clasificaciones y unidades empleadas en el estudio de esos sectores

Población total

Con respecto a las estadísticas de población total, la falta de comparabilidad entre los datos de diversos países procede de: 1) diferencias en el modo de definirla, y 2) de sub y sobre-enumeraciones.

Definiciones

Hay dos modos de definir la población total: 1) definición *de facto,* y 2) definición *de jure.*

Para 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. Para 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 que se encuentren en el momento del censo.

La definición de facto tiene como ventajas 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 se presentan casos dudosos que deba de resolver el censador según su criterio.

La definición de jure implica diferencias en la definición de "domicilio" por los diversos países. En el momento de la recolección, plantea el problema de las personas que no tienen residencia fija, y la necesidad de elaborar instrucciones más complicadas para los enumeradores. La definición de jure puede dar mejores resultados cuando la enumeración censal deba hacerse en un lapso relativamente grande y, muy especialmente, cuando se haga con poblaciones de gran movilidad.

Las definiciones *de facto* y *de jure* pueden complementarse, en cuanto cada una cumple una función en la investigación estadístico-social. Más concretamente, esto se manifiesta en el terreno nacional o local, pues una definición de jure será más útil para planear servicios sociales (alojamientos, educación de la población residente), mientras una definición de facto puede ser más útil para los programas de protección policiaca, frente a las necesidades del transporte, etcétera.

En el campo internacional, las diferencias de definición pueden provocar dificultades en los países que tienen muchos nacionales radicados en el extranjero (como fuerza de ocupación), o en los que atraen muchos turistas.

De 53 países consignados en el Anuario de las Naciones Unidas: 11 emplearon definiciones *de jure,* y 31 usaron ambos tipos de definición. De los que usaron ambas definiciones: 1) unos enumeraron sobre base fáctica recogiendo en la cédula no sólo el lugar ocupado físicamente por la persona, sino también el de su residencia habitual, y 2) otros, en una misma cédula, pero en renglones separados, consignaron a quienes residen usualmente en el lugar pero están ausentes en el momento del censo, y a quienes residen usualmente en otro lugar, pero están temporalmente presentes en el del censo. A este último grupo, que permite mayor precisión, corresponden los censos de Dinamarca, Noruega y Polonia.

Fuentes de discrepancia

Para mostrar de dónde procede la falta de comparabilidad en casos como éstos, puede mencionarse que en las definiciones de facto, algunos países excluyeron a las fuerzas armadas extranjeras localizadas en su territorio y otros las incluyeron; algunos incluyeron y otros excluyeron a sus propias fuerzas armadas que estaban en el extranjero; algunos incluyeron y otros no a los civiles asimilados a las fuerzas armadas; unos incluyeron y otros no a los miembros del servicio diplomático extranjero; unos incluyeron y otros no a los de su propio servicio en el extranjero. En las regiones marítimas, a veces se incluyó en el censo a los marineros cuyos barcos estaban anclados en puerto nacional, a quienes viajaban en barcos que navegaban entre puertos nacionales, y así sucesivamente.

En el caso de las definiciones de jure, las estadísticas de los países que las utilizan hablan de "población residente" y de "población domiciliada".

Debido a la diversidad de aplicaciones que tienen ambos tipos de definición de la población total, resulta conveniente usar ambas simultáneamente. En todo caso deben atenderse las recomendaciones de los organismos internacionales en el sentido de:

1º Obtener datos sobre el total de la población en el momento del censo.

2º Excluir al personal militar y diplomático extranjero radicado en el país e incluir al propio radicado en el extranjero.

3º Estimar o indicar que no se incluye el número de aquellos grupos que no se puedan enumerar individualmente. Esto abarca a los grupos que viven fuera de la estructura económico-social del país, a los que hablan otras lenguas, etcétera.

Sub y sobre-enumeraciones

La segunda gran dificultad que enfrentan los censos de población total son los de las sub y sobre-enumeraciones. Éstas dependen de: 1) el nivel educativo general y especial de la población censada; 2) la tradición estadística del país; 3) la planeación y organización del censo;

4) el entrenamiento, la experiencia y la honradez de los censadores; 5) los recursos técnicos y económicos de que se dispone; 6) las facilidades de comunicación y de transporte.

Las sub-enumeraciones se presentan principalmente en el caso de: 1) los núcleos aislados física, social o culturalmente (habitantes de lugares inaccesibles, proscritos sociales, individuos que no hablan el idioma oficial); 2) las personas o núcleos móviles (trabajadores migràtorios, habitantes de hoteles); 3) ciertos núcleos de edades y sexos (niños, hombres jóvenes, muchachas solteras).

Cómo subsanar en parte estos problemas

Todo lo anterior, muestra que ni las cifras censales deben tomarse sin crítica ni aceptar —sin hacer esfuerzos por valorarlos— los peligros de la sub o de la sobre-enumeración. Debe planearse, en efecto, una verificación de los resultados obtenidos. Ésta ha de realizarse inmediatamente después del censo, por medio de un muestreo del área censada. En este punto, la postulación estadística de hipótesis y su comprobación o reprobación juegan un papel muy importante. Por otra parte, una vez comparados los resultados censales con los de origen muestreal, puede apreciarse el grado de sobre o de sub-enumeración. Si se guarda un registro de esta estimación a lo largo de un periodo considerable puede tenerse un medio adicional de apreciar el grado en que en ese país, en una etapa determinada, tiende a sobre-enumerarse o a sub-enumerarse, y con ello se podrá contar, en el futuro, con un auxiliar para obtener cifras más próximas a la realidad en los censos de población.

Sexo y edad

La clasificación por edad y sexo de los habitantes de un país es muy importante en los censos de población. Las informaciones acerca de la estructura de una población de acuerdo con las edades de sus componentes sirve: 1) para prever los cambios en la composición de la población; 2) para prever las disponibilidades de fuerza laboral y bélica; 3) para estudiar la dependencia social infantil y senil; 4) con vistas a la política sanitaria del Estado (en relación con la morbilidad y mortalidad), y 5) en relación con fines comerciales y cálculos actuariales de las compañías de seguros (como la determinación de las probabilidades de supervivencia).

La clasificación sexual de la población es sencilla; en cambio, la clasificación por edades presenta ciertas dificultades. Entre éstas cuentan las derivadas de una mala información. Puede depender ésta de: 1) ignorancia de la edad correcta, o descuido; 2) motivaciones económicas, políticas, sociales, sicológicas (como el temor a la conscripción); 3) la preferencia irracional por ciertos números ("números redondos" en cero, en 5, pares) y la aversión por otros (los nones, el 7, el 13); 4) la existencia de culturas que viven fuera del mundo numérico corriente (en Oceanía, África y Asia, principalmente) y que distinguen burdamente entre "niños" y "adultos".

Definición de la edad

Aunque por lo general la edad se mide en años completos, la definición de la edad en los censos tiene —dentro de este ámbito— cierto campo de variación internacional. En tanto hay países que toman en cuenta la edad del interrogado en los cumpleaños inmediatamente anterior, otros consideran el cumpleaños inmediatamente posterior, otros el más inmediato (sea anterior o posterior) y algunos cuentan la edad no a partir del momento del nacimiento, sino que lo retrotraen a un año antes del nacimiento.

Métodos

Generalmente, se usan dos métodos para obtener información sobre la edad: o se pregunta la fecha de nacimiento, o se pregunta directamente la edad. El segundo procedimiento es más simple, pero equívoco: el primero, evita las dificultades derivadas de la diferencia social en la definición de la edad. El primero es típicamente europeo; el segundo, americano. Al primero, se le han encontrado ventajas no estadísticas en los países en que los datos censales se emplean para fines judiciales o de otro tipo; pero, tiene el inconveniente de que requiere no sólo la recolección de las fechas de nacimiento, sino el ulterior cómputo de la edad.

En algunos censos, además de la edad de los adultos y de la de los niños mayores de un año, dadas en años, se registra la edad de los menores de un año en meses completos, y la de los menores de un mes, en días.

Respecto de la clasificación de la población por edades, la Comisión de las Naciones Unidas para la Población propuso: 1) tabular las edades para cada sexo separadamente; 2) tomar la edad del último cumpleaños; 3) hacer por lo menos los siguientes grupos de edad: el de los menores de un año, el de los de uno a cuatro años; los de cinco en cinco años hasta el fin de la vida.

El Comité para el Censo de las Américas, en lugar de considerar grupos quinquenales hasta el fin de la vida, limitó dichos grupos hasta los 84 años, pues formó un último grupo con los individuos "de 84 años o más".

En la práctica, ha sido frecuente formar grupos de diez en diez años, siguiendo dos sistemas: 1) el que forma grupos con edades cuyo límite inferior es un número terminado en cero (10 a 19, 20 a 29... 80 a 89, etcétera), y 2) el que agrupa las edades de diez en diez comenzando con un número cuyo último guarismo es 5 (de 15 a 24, de 25 a 34, ... de 85 a 94, etcétera).

Para asegurar la comparabilidad, se han recomendado las clasificaciones quincenales, pues éstas permiten: 1) una conversión entre diversos países, y 2) la conversión entre los que obtenga cada uno de ellos en el futuro y los que haya obtenido en el pasado con otros sistemas.

Al lado de los grupos con intervalos de 5 o 10 años, importan otros grupos de edades. Particularmente, interesan los definidos por los siguientes criterios: 1) 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); 2) edades escolares frente a las extraescolares; 3) edades económicamente productivas, frente a edades económicamente dependientes; 4) edades políticamente significativas (ciudadanos) frente a las que no lo son.

Todo esto, sin embargo, representa —en general— una elección más o menos arbitraria o convencional de límites para cada grupo; o una elección de límites que tiene que tener en cuenta las condiciones de estructura económica, demográfica, sanitaria, etcétera. Algunas de estas consideraciones son las que permiten hablar, por ejemplo, de la "prolongación social de la juventud" y la "anticipación social de la vejez" en contextos determinados que pueden ser de interés más que para el puro estadístico, para el estadístico-social.

Estado matrimonial

El estado matrimonial permite clasificaciones de la población que interesan desde los ángulos: 1) demográfico; 2) médico; 3) económico, y 4) sociológico.

Demográficamente, la clasificación estadístico-censal, según estado matrimonial, importa porque la soltería, el divorcio o la viudez de capas más o menos extensas de la porción reproductiva de la población, constituyen factores que pueden explicar o que permiten prever cambios en la cuantía y estructura de la misma.

Desde el ángulo médico, asistencial y de la seguridad social, dicha clasificación es importante cuando se trata de establecer servicios de maternidad y otros análogos.

Económicamente, esa clasificación permite prever ciertas características de la demanda de artículos para el hogar y de otros bienes semejantes, así como valorar las posibilidades de suministro de fuerza laboral.

Socialmente, interesa al investigador que examina los problemas propios de la soltería, la viudez o el divorcio y le permiten medirlos volumétricamente. En los países en los que existe la poligamia legal, los datos obtenidos en esta forma ofrecen la posibilidad de estudiar los caracteres con que ésta se da en la realidad.

La Comisión de la Organización de las Naciones Unidas sobre Población recomendó que se hicieran las siguientes agrupaciones, en este aspecto: 1) casados (con inclusión de quienes han formado uniones estables *de facto*); 2) viudos que no se han vuelto a casar desde que enviudaron la última vez; 3) divorciados que no se han casado desde su último divorcio, y 4) solteros.

Aunque estas categorías parecen obvias, no lo son tanto. Esto lo demuestra el hecho de que algunas naciones subsumen en una misma categoría a solteros, viudos y divorciados (así como a los separados, en donde no se reconoce legalmente el divorcio), o sea, que establecen propiamente una dicotomía entre los casados y los no casados.

Casados

La comisión consideró deseable que en la primera categoría se distinguiera entre: *a*) casados *de jure* (o legalmente), y *b*) unidos *de facto* (o en unión real, no sancionada por la ley).

La separación, en una categoría especial, de quienes viven en uniones matrimoniales *de facto* es especialmente importante en los países latinoamericanos. Por otra parte, esta sub-clasificación da flexibilidad a los datos que, así, se adaptan a diferentes tipos de análisis estadístico. Es así como, desde el ángulo legal, las personas incluidas en esa categoría se asimilarán a los solteros, en tanto que, para un análisis demográfico realista (de tasas de reproductividad, etcétera) o en uno económico (que estudie, por ejemplo, la demanda de artículos domésticos) se asimilarán a los casados.

Divorciados y separados

En forma análoga a la anterior, la formación de una categoría especial con las personas "separadas" de hecho, que no lo están legalmente, permite distinguirlas, en análisis relativos a la reproductividad, demanda de artículos domésticos, servicios de maternidad y otros parecidos. La propia Comisión de la Organización de las Naciones Unidas señaló la conveniencia de clasificar en categorías separadas a: *a*) los divorciados legalmente; *b*) los separados de sus cónyuges sin mediar divorcio.

Propuestas para clasificación más fina de los casados

Puede agregarse que, en algunos países —con fines de interpretación sociológica— sería conveniente hacer una distinción más fina que tuviera en cuenta la unión o falta de unión eclesiásticas pues aun cuando no va acompañada de unión legal, la misma representa un principio de respeto a la norma social, de gran importancia para el sociólogo. Así vendrían a constituirse las cuatro categorías siguientes: *a*) casados legal y eclesiásticamente; *b*) casados legalmente, no unidos eclesiásticamente; *c*) unidos eclesiásticamente (en unión *de facto* para la ley y estadísticas corrientes), y *d*) unidos *de facto*, sin unión eclesiástica.

Además de señalar las categorías matrimoniales, según las recomendaciones de la comisión, se debe agregar una categoría de "estado matrimonial no determinado" y hacer las especificaciones correspondientes en cada categoría, para los distintos grupos de edad señalados en este capítulo.

Principales deficiencias observadas

De acuerdo con las normas de la comisión, las principales deficiencias que se observaron en los registros del estado matrimonial en los diferentes países del mundo consistieron en: 1) falta de totales para las diferentes categorías de los de menos de 15 años, y 2) falta de agrupación conveniente de las edades. El primer aspecto es de particular importancia para los países —como India— en que se acostumbra el matrimonio en edad muy temprana; el segundo aspecto se presentó particularmente en los datos de Guatemala y México en los que "el agrupamiento por edades no fue lo suficientemente fino como para formar siquiera grupos de 10 en 10 años".

Otros aspectos estadísticos del estado matrimonial

Otros aspectos que conviene registrar para su uso ulterior, se refieren a: la edad en el momento del matrimonio; la duración del matrimonio; la duración de la viudez o del estado de divorcio; el estado matrimonial previo al registrado en el momento del censo (para estudios de fertilidad y determinación de probabilidades de recasamiento), el número de veces que se ha estado casado, y el número de cónyuges (en los países que permiten la poligamia o, con mayor precisión, la poliginia y la poliandria).

Fertilidad

Los datos estadísticos sobre fertilidad se refieren a la frecuencia de los nacimientos en una población dada, con respecto al volumen de esa población. Importan particularmente desde los ángulos: 1) demográfico, 2) biológico, y 3) sociológico.

En el campo de la biología, esos datos son útiles para los estudios de genética humana y de antropología, y en el terreno aplicado, para los de medicina.

Desde el ángulo demográfico, los datos sobre fertilidad permiten prever el posible desarrollo de las poblaciones, su posible composición futura por edades, etcétera. Todo esto importa para prever y encauzar las transformaciones en la economía y en la mentalidad del grupo (pues no es igual ésta si en él predominan los jóvenes que si son los adultos, o los ancianos quienes predominan en él).

Sociológicamente, la importancia de las cifras sobre fertilidad destaca en cuanto se considera que, a través del estudio de la fertilidad diferencial de los diversos agrupamientos sociales (etnias, clases sociales, grupos religiosos, etcétera) se pueden predecir ciertos cambios en la estructura social, en el predominio de ciertos modos de pensar y actuar en el ritmo y sentido del cambio social.

Las cifras de fertilidad se obtienen usando combinadamente dos fuentes: 1) los censos de población, y 2) los registros de nacimientos.

A partir de los datos consignados en estas fuentes, se obtienen los índices de fertilidad relacionando matemáticamente el número de nacimientos ocurridos en un grupo con el número de personas miembros del mismo. Para mayor precisión, se debe obtener dicha relación dividiendo el número de nacimientos no entre el total de la población, sino entre la porción reproductiva de la misma (los adultos en edad de procrear, en forma muy burda todavía).

Cuando faltan los registros de nacimientos, se recurre a varios procedimientos para obtener cifras de fertilidad a partir de los censos; entre ellas, se cuenta la obtención de relaciones matemáticas entre el número de niños enumerados en el censo y el número de adultos censados, en edad reproductiva.

Hay casos en los que dicha obtención se prevé desde la planeación del censo, registrándose en las cédulas correspondientes el número de niños y el número de adultos o parejas de adultos casados que existen en cada casa y calculándose la relación matemática entre ambas cifras.

O sea, que pueden reconocerse dos formas de obtención de datos sobre fertilidad: 1) el uso de tabulaciones especiales que se basan en el análisis de la composición hogareña, y 2) el uso de datos acerca del número de hijos nacidos.

Análisis de la composición hogareña

En el análisis de la composición hogareña, se parte de la inspección del número de menores de cinco años enumerados en cada hogar, puesto en relación con el de los adultos que se presume sean sus padres, con base en el estudio de las relaciones familiares tal y como aparecen en la cédula correspondiente. Las tabulaciones se hacen: 1) por grupos de edad y sexo de los padres; 2) por estado matrimonial de éstos; 3) por ocupación, lengua, etnia, lugar de nacimiento, ciudadanía, educación, etcétera.

Las tabulaciones deben consignar el total de los adultos que correspondan a cada uno de los grupos antes señalados, así como el total de sus hijos.

Interrogatorio sobre el número de hijos nacidos

Con respecto al segundo procedimiento —o sea el que interroga acerca del número de hijos nacidos— es necesario consignar :1) el número de mujeres que informaron acerca de cuántos hijos habían tenido (incluyendo a las que informaron que no habían tenido); 2) el número de hijos de quienes se informó; 3) el número de mujeres que informaron haber tenido 0, 1, 2, 3, 4, 5 o más hijos; todo ello, clasificado por edades de la madre, estado matrimonial, etcétera.

Las tabulaciones sobre fertilidad tienen, sobre las relaciones basadas en datos censales y en registros de nacimientos, una ventaja: 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. Problemas como ése dificultan el estudio de la fertilidad diferencial de los diferentes colectivos ocupacionales, de ingreso, etcétera.

En los censos actuales, que consignan datos para estudiar la fertilidad de la población, se descuida —a veces— la posibilidad de estudiar y comparar la fertilidad diferencial de las porciones masculina y femenina de la población. Esto es particularmente interesante en cuanto se establece una relación con los puestos ocupacionales de los padres (en cuanto ciertas ocupaciones son desempeñadas más por hombres que por mujeres, y otras más por éstas que por aquéllos), o en cuanto se relacionan con ciertas zonas de migración (a las que pueden llegar, por ejemplo, más hombres que mujeres en edad reproductiva).

Además de esto, se presentan problemas especiales, en cuanto se trata de enumerar a los hijos de las madres solteras. Conviene que a ellas y a sus hijos se les consagre lugares separados de aquellos que se destinan a las personas casadas, ya que representan un problema social, de gran interés para el sociólogo.

Lugar de nacimiento. Lengua y nacionalidad

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, se registran menos.

De los tres tipos de datos arriba mencionados, el del lugar de nacimiento no resulta suficiente para determinar la pertenencia étnica de un individuo a un grupo, ya que muchos países reciben fuertes núcleos de inmigrantes extranjeros y los hijos o nietos de los inmigrantes, no obstante haber nacido en el país, conservan muchos rasgos físicos y culturales de sus progenitores. De otra parte, esos mismos datos no llegan a reflejar el grado de asimilación del nacido en el extranjero quien, aunque haya nacido en país distinto de aquel en que está radicado, puede estar ya asimilado completamente a la población nativa.

Los datos sobre el lenguaje resultan más útiles al respecto pues: 1) el lenguaje es criterio más sensible en cuanto generalmente las diferencias lingüísticas van acompañadas de otras importantes diferencias culturales, y 2) el conocimiento del lenguaje del país es casi indispensable para los inmigrantes como base de asimilación, pues sin la lengua nacional es difícil que alguien pueda asimilarse a la población nativa. Y aunque ese conocimiento no es condición suficiente, sí es condición necesaria de asimilación. A más de esto, el lenguaje es un criterio más flexible que el del lugar de nacimiento y otros semejantes.

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 para los propósitos antes señalados, ya que las posibilidades de naturalización de los extranjeros hacen que los mismos sean valiosos —casi únicamente— en relación con investigaciones sociojurídicas.

Clasificación de la población por lugar de nacimiento

La clasificación de los pobladores por lugar de nacimiento permite apreciar los volúmenes y las fuentes de la migración. Esto se puede referir tanto a la inmigración propiamente dicha, como a la intramigración o migración interna entre los diversos lugares de una misma área; entre diversas entidades de un mismo Estado.

La medida de las migraciones hecha a base de los datos aportados por las respuestas sobre lugar de nacimiento de los censados es muy burda. Es una mera aproximación, ya que, en el cálculo no se pueden considerar: 1) las muertes de inmigrantes, ocurridas entre dos censos; 2) el retorno de ciertos migrantes a su lugar de origen en el lapso que media entre dos censos.

Para refinar un poco los datos a través de los que se puede apreciar el fenómeno de las migraciones tanto internas como externas, se puede recurrir a clasificaciones de los migrantes: 1) por edad y sexo; 2) por lugar de partida, fecha de la misma y fecha de arribo al lugar de destino.

Migraciones y fertilidad

Para estudiar los problemas migratorios y de fecundidad diferencial entre los varios grupos étnicos de un país, hay que obtener, a más de los del total de la población, datos complementarios sobre: 1) número de mujeres en edad de procrear dentro de la población migrante; 2) número de hijos de las mujeres que migran. Estos últimos deberán estar clasificados en dos categorías según que hayan nacido o no en el país al que inmigraron sus madres.

Áreas de procedencia

Para clasificar a los inmigrantes por lugar de procedencia, cada país deberá tener en consideración los límites geográficos de las diversas áreas político-administrativas *en el momento del censo*. Para ello, deberá tener a la vista el registro más reciente de los países y territorios no autónomos, publicado por las Naciones Unidas, y la división territorial pertinente, publicada por los órganos oficiales del país de que se trate.

Hay ciertos casos en que las cifras de inmigración son pequeñas. Se acostumbra, en tales casos, sumar las que corresponden a varios países. En relación con esta práctica, la comisión de las Naciones Unidas consagrada a los problemas demográficos ha recomendado que dichas adiciones se hagan siempre con las de los países de un mismo continente, a fin de que al menos puedan realizarse estudios sobre las migraciones intercontinentales.

Con fines de comparabilidad internacional, la Organización de las Naciones Unidas ha preparado, también, un agrupamiento de países por zonas. Esta división zonal del mundo aparece en:

United Nations Statistical Office: *Nomenclature of Geographic Areas for Statistical Purposes*. Statistical Papers, series M, núm. 1.1, enero de 1949.

Algunos problemas de detalle

En el estudio de todas las migraciones, pero particularmente en el de las internacionales, se suele presentar en nuestros días un problema adicional en cuanto los nacimientos no siempre se producen en la residencia paterna, sino en hospitales o maternidades que, en ocasiones no se encuentran en el mismo lugar de residencia. De ahí que se requiera que la información especifique si el lugar de nacimiento se refiere a: 1) la localidad en que residían los padres cuando ocurrió el nacimiento, independientemente del lugar en que éste se haya producido realmente, o 2) la localidad en que ocurrió el nacimiento, con independencia del lugar en que hayan podido residir los padres en ese momento.

En el caso de las migraciones internas, en cuanto más específico sea el dato (localidad más que provincia, Estado, etcétera) resultará más útil para las investigaciones. A más de ser específico el dato debe ser unívoco, pues, en muchos casos no basta mencionar el lugar sino que hay que referirlo a la división político-administrativa a la que corresponde, pues en muchos países existen varios lugares que tienen el mismo nombre.

Ventajas apendiculares

En relación con la población nativa, la especificidad de su adscripción a un lugar de nacimiento determinado tiene ventajas, especialmente: 1) cuando fluctúan mucho de límites de los Estados, a causa de las guerras y las reclamaciones territoriales correspondientes o por otras causas semejantes, y 2) cuando, dentro de un mismo país, hay diferencias étnicas o culturales importantes dentro de las diversas zonas que lo constituyen (pues con ello se facilita, adicionalmente, el estudio de los recambios étnico-culturales y los progresos que logra la integración étnica, social y cultural del país).

Nacionalidad

Bajo el rubro "nacionalidad", los países consignan datos estadísticos muy diversos (muchas veces sobre lugar de nacimiento, lenguaje, grupos formados por individuos con antepasados o costumbres comunes). Sin embargo, en su mayoría se refieren a la nacionalidad legal que es la única de la que nos ocuparemos.

Estos datos sobre nacionalidad legal se utilizan principalmente para estudiar: 1) los problemas de estatuto legal, y 2) los de los derechos civiles de los inmigrantes.

Recomendaciones sobre los datos de nacionalidad

Las recomendaciones hechas por las comisiones internacionales con respecto a los problemas de la nacionalidad legal, señalan la necesidad de distinguir: 1) a quienes son nacionales del Estado (por nacimiento, naturalización u otro concepto), y 2) a quienes son extranjeros (clasificados de acuerdo con el Estado del que sean nacionales).

Tanto los nacionales como los extranjeros deben de clasificarse por edad y sexo, formándose grupos de: menores de 5 años, de 5 a 14 años, de 14 años en adelante (por intervalos de 10 años) hasta los 64 años, y de 64 años o más.

En los Estados en los que las tasas de inmigración sean altas, se debe clasificar tanto a los nacionales como a los extranjeros según: 1) estado matrimonial; 2) características socio-económicas; 3) residencia urbana o rural; 4) nivel educativo, y 5) ciudadanía del cónyuge.

Las comisiones internacionales también consideran útil que, en casos como el de la Comunidad Británica, que está formada por zonas muy diversas, se clasifique a los individuos de acuerdo con la zona a la que pertenezcan.

Situación actual de los datos sobre nacionalidad

Hasta ahora, los datos sobre nacionalidad muestran que: 1) en algunos países se distingue entre nacionales y extranjeros, pero no se clasifica a éstos según su nacionalidad; 2) en otros existe la distinción en la cédula censal, por lo que se refiere a los extranjeros, pero esa distinción no pasa a los recuentos; 3) otros distinguen, además, en el grupo de los nacionales, a los que lo son por nacimiento y los que lo son por naturalización; 4) otros establecen también una distinción para quienes han iniciado el proceso de naturalización; 5) hay algunos que preguntan sobre la nacionalidad o nacionalidades previas a la naturalización, y otros, 6) también interrogan acerca de la nacionalidad de los padres.

Problemas generales sobre nacionalidad

Un problema de designación que dificulta las comparaciones internacionales se refiere a casos como los de los súbditos del Imperio Británico, que en algunos censos se clasifican como "ingleses", en otros como "británicos", en otros, bajo el rubro "Gran Bretaña" y en otros como "Gran Bretaña e Irlanda del Norte", o, en otros como "súbditos y colonos británicos" sin que estas denominaciones cubran exactamente la misma extensión y sin que, por lo mismo, sean comparables los datos consignados bajo cada una de ellas.

Pero hay otros problemas que son aún de mayor importancia y más difíciles de resolver, como son: 1) el de los apátridas; 2) el de quienes tienen doble nacionalidad, y 3) los que resultan de los cambios de fronteras.

En algunos países se ha formado un grupo aparte con las personas que carecen de nacionalidad (o apátridas); en otros, se les ha adscrito al grupo de la nacionalidad que tuvieron previamente.

En el caso de quienes tienen doble nacionalidad (generalmente, personas nacidas en un país en el que rige el *jus soli* de padres nacidos en un país regido por el *jus sanguinis*) hay que diferenciar según que: 1) el censo se realice en uno de los países que dan base a la aparición de la doble nacionalidad del censado, o 2) que el censo se realice en un país distinto de los dos implicados en la doble nacionalidad del censado. En general, en el primer caso, el censo incluye a quien tiene la doble nacionalidad entre sus nacionales. En el segudo caso, se acostumbra preguntar a quien se censa cuál es la nacionalidad en la que prefiere entrar en el censo.

El problema del cambio de fronteras aún no ha dado lugar a previsiones internacionales.

Algunos aspectos de detalle en los censos sobre nacionalidad

Además de los datos ordinarios de sexo, edad y estado matrimonial (con especificación de la nacionalidad del cónyuge) en los censos de nacionalidad se cuentan los relativos al número de años que los extranjeros han permanecido en el país. Estos datos sirven para estudiar los progresos en la asimilación de los mismos y, en su caso, el correspondiente progreso de su naturalización. Además, es importante clasificar a los extranjeros por ocupaciones, ya que las leyes de cada país sólo les permiten dedicarse a algunas de ellas y no a otras. Estos datos pueden ponerse en relación con cierta selectividad de las migraciones, pues si bien no migran sólo quienes pueden practicar su mismo oficio o profesión en el lugar de destino (en cuanto hay quienes están dispuestos a cambiarlo, en función de otras consideraciones), sí es ése el caso más general.

En relación con la naturalización se suele preguntar: si se han iniciado o no trámites de naturalización, cuál ha sido la nacionalidad previa, cuál fue el lugar de nacimeinto, cuál fue la última residencia del censado.

Estadísticas de extranjeros en Bélgica

Como un ejemplo de los estudios censales sobre extranjería, mencionaremos los que se hacen hace veinte años aproximadamente, en Bélgica.

En la práctica administrativa, al extranjero se le define de acuerdo con su estatuto jurídico de ciudadano belga, o de acuerdo con su nacionalidad legal, y no por su raza o lugar de origen. Las estadísticas enumeran sólo a los extranjeros radicados en Bélgica; excluyen a los agentes diplomáticos, a los turistas y a los militares estacionados en el territorio.

Con respecto al país del extranjero —en Bélgica— es frecuente que se subsuman bajo un mismo rubro a quienes proviene de los territorios metropolitanos y quienes provienen de las colonias (en casos como los de Francia, Inglaterra, etcétera).

No todos los datos recogidos acerca de los extranjeros están disponibles en forma inmediata, ya que ciertos datos no se analizan o elaboran y otros, aunque se elaboren, son utilizados únicamente por los organismos oficiales y quedan, por lo demás, inéditos. En muchos casos los resultados se publican en forma incompleta y se reúnen en clases estadísticas demasiado amplias. Los datos se agrupan de acuerdo con las circunscripciones administrativas y no por unidades sociales, como sería deseable.

De ahí que, en una investigación, se deben clasificar las informaciones correspondientes: 1) según la época en que se obtuvieron los datos; 2) el organismo que intervino en la recopilación; 3) el carácter general o específico de la información con respecto a los extranjeros; 4) según que se refieran a todos los nacionales y a todos los extranjeros o que consideren sólo ciertas categorías de éstos (adultos, trabajadores, etcétera), y 5) según que la fuente sea un censo realizado en fechas determinadas periódicamente, o que sean documentos administrativos compilados en forma permanente por las oficinas del gobierno.

Censos generales de la población, la industria, el comercio y la agricultura

Estos censos, que son las fuentes estadísticas oficiales de mayor importancia, se realizan: señalando una fecha determinada para levantar el censo; enviando a cada hogar, empresa industrial o comercial, ciertos grupos de censadores designados por los ayuntamientos; dotando a dichos censadores de cédulas impresas que contengan preguntas a las que habrán de responder los habitantes, y verificando los datos y transcribiéndolos a fin de formar tabulaciones especiales.

Desde 1910, la verificación, transcripción y despojo de las cédulas censales ya no las realizan los ayuntamientos. Dichas operaciones se han centralizado. Ahora ya sólo la recolección de las datos es la que corresponde a las administraciones comunales. Las operaciones restantes las realizan los servicios centrales de estadística por medio de máquinas. Gracias a esto: se eliminan los errores y 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 (como el "registro de extranjeros"). Sin embargo, esos registros pueden resultar útiles como medios de ratificar o rectificar resultados o para obtener mayores detalles.

Algunas características de los censos de diferentes fechas

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: 1) la distribución de los habitantes por nacionalidad (distinguiéndose la nacionalidad de origen de los extranjeros, el lugar de nacimiento y el sexo) en relación con las diferentes jurisdicciones administrativas y, dentro de cada una de ellas, por grupos de comunas, y 2) los datos sobre la población extranjera en las secciones correspondientes a: estado civil, edad, profesión, familia, jefatura familiar.

Aunque en algunos censos de la industria, la agricultura y el comercio, se omite en ocasiones la nacionalidad de los trabajadores (según ocurrió en 1930), en otros se señala la distribución de los extranjeros ocupados, según su sexo, el grupo de actividad y la provincia a las que correspondan.

El censo económico y social de 1937 tiene en cuenta: 1) el país del que son nacionales los trabajadores extranjeros; 2) su distribución por edad y sexo; 3) su repartición por grupos industriales o comerciales; 4) su distribución de acuerdo con el monto de sus ingresos, y 5) su clasificación en obreros, empleados, empleadores, etcétera.

Los censos relativos a los desocupados también consideran su nacionalidad y señalan: su edad, cuánto dura su desocupación, y otros datos semejantes. La situación de extranjería también se considera en los registros de diplomados de la enseñanza superior.

El registro de extranjeros, los censos especiales de extranjeros y las estadísticas de permisos de trabajo son las principales fuentes de datos para las investigaciones que se refieren a los extranjeros en Bélgica.

El registro de extranjeros

Todo extranjero que resida por más de 15 días en Bélgica, debe obtener un *titre de sejour,* inscribirse en el *régistre des étrangers* y obtener un certificado de inscripción en ese registro. Al prolongarse su permanencia por más de seis meses, debe obtener una carta de identidad.

Los datos que proceden de esta fuente deben de usarse con ciertas reservas, ya que las entradas y salidas de los extranjeros no dejan huella en dichos documentos.

Censos especiales de extranjeros

Un ejemplo de los censos especiales de extranjeros es el que se realizó en 1939, en vísperas de la movilización. Sus datos se obtuvieron por medio de censadores que se enviaron a cada hogar para registrar la presencia de cualquier extranjero que pudiese haber en él. Los datos así obtenidos se ratificaron mediante preguntas análogas que se dirigieron a los empleadores, hospederos, etcétera.

Los resultados correspondientes se refieren a extranjeros de más de quince años, y los clasifican según su nacionalidad, su sexo, su edad, su lugar de residencia, su país de nacimiento, la fecha de llegada a Bélgica, su carácter migratorio, su ocupación, su estado social, etcétera.

Entre los extranjeros censados se consideraron aquellos que aunque hubieran sido movilizados por su país y hubiesen dejado Bélgica continuaban teniendo familiares que permanecían en Bélgica.

Estadísticas de permisos de trabajo

Estas estadísticas se llevan en el caso de los extranjeros que ejercen sus actividades de acuerdo con las estipulaciones de un contrato de trabajo o de un empleo, y mediante las estadísticas de cartas profesionales en el caso de quienes ejercen una actividad profesional independiente.

Otras estadísticas

Entre las otras fuentes para el estudio de la extranjería, en Bélgica, se cuenta: 1) las estadísticas de matrimonios, que se basan en los datos contenidos en las actas matrimoniales y comunicados por los funcionarios del estado civil (pues desde 1947 tienen en cuenta la nacionalidad de los cónyuges); 2) las estadísticas de naturalización y las declaraciones de indigenato (comprendiendo estas últimas las declaraciones de opción o de renuncia a la nacionalidad belga, por adquisición de una extranjera, y las de las mujeres que desean conservar o recobrar la nacionalidad belga a pesar de su matrimonio con un extranjero); 3) las estadísticas de comerciantes ambulantes, hechas por el Ministerio de Asuntos Económicos, que se obtienen a partir de las credenciales de las que los comerciantes son titulares obligados, señalándose en esas estadísticas las distribuciones por nacionalidad, tipo de comercio, división administrativa, etcétera; 4) las estadísticas de aprendizaje artesanal que lleva el mismo ministerio, y 5) las estadísticas de la fundación universitaria, relativas a la población de las universidades y centros de enseñanza superior.

Educación

Los datos sobre educación son importantes para: 1) planear y programar la educación; 2) difundir informaciones, y 3) programar en general el mejoramiento económico, político y social.

La falta de comparabilidad en materia de educación proviene principalmente de: 1) diferencias en los sistemas educativos de diferentes países (problema del contenido social), y 2) diferencias en cuanto a las posibilidades estadigráficas de esos países (problema de aproximación metodológica).

En los recientes censos de población, se incluyen tres variedades principales de datos acerca de la educación: 1) datos sobre alfabetismo y analfabetismo; 2) datos sobre nivel educativo, número de años de escuela y tipo de instrucción recibida por cada individuo, y 3) datos sobre la asistencia a las escuelas.

Los datos anteriores reciben mayor o menor atención relativa de acuerdo con el nivel educativo general de cada país; en países de bajo nivel educativo general habrá mayor interés en recoger los datos referentes al analfabetismo y la alfabetización, sobre todo; en países de nivel educativo general más elevado esos datos casi no tendrán interés pues la gran mayoría de la población es en ellos población alfabetizada; en cambio, a esos países les interesarán más las diferencias en el nivel educativo de sus habitantes, o sea, que procurarán recoger principalmente los datos relativos al tipo de enseñanza que han recibido y al número de años de escuela.

Alfabetismo

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 escrito en cualquier idioma". También propuso que los datos correspondientes se refirieran a la población de más de quince años. A partir de esa edad, sugirió que los grupos de edad y sexo fueran: de 15 a 19 años; de 20 a 24 años, grupos con intervalos de 10 años, de los 25 a los 64 años, y de 65 años o más.

Procedimientos empleados para obtener datos sobre alfabetismo

En diferentes países, los censos han recogido datos sobre analfabetismo utilizando tres procedimientos:

De acuerdo con el primer procedimiento se pregunta a la persona si sabe leer, exclusivamente. El procedimiento es criticable en cuanto no satisface la definición de persona alfabeta (que es la que sabe leer y escribir). El segundo procedimiento consiste en preguntar si la persona sabe leer y escribir, pero tampoco basta porque muchas personas dicen que saben hacerlo porque saben escribir su nombre y esto no se ajusta tampoco a la definición propuesta. El tercer procedimiento pregunta, por separado —y de ser posible comprueba, en la práctica mediante una pequeña prueba— 1) si la persona sabe leer, y 2) si sabe escribir.

El último de los procedimientos mencionados tiene la ventaja de que permite constituir una categoría de "semi-alfabetos" con quienes saben leer, pero no escribir.

Tabulaciones

La tabulación de los datos sobre alfabetismo de la población tiene muchas variedades. Pueden contarse, en un extremo, países que no distinguen sino entre la población "menor de" y la "mayor de" determinada edad, mientras en el otro extremo, se recoge el alfabetismo de la población de los 7 años a los 119 años (edad máxima observada en el país correspondiente en el momento del censo) de año en año.

Nivel educativo

La comisión propuso que se indicara el nivel educativo alcanzado por cada uno de los pobladores censados, en términos completos de instrucción cursada. Simultáneamente, pidió indicar el grupo de edad al que pertenecen los interrogados (menores de 25 años, de entre 25 y 64 por intervalos de clase de diez en diez años, de mayores de 65 años).

En algunos censos se menciona no sólo el número de años de escuela cursados y aprobados, sino también el tipo de instrucción (universitaria, técnica, comercial, por correspondencia, etcétera).

Aunque en muchas ocasiones las cédulas censales piden muchos informes de los censados, en las tabulaciones suele aparecer sólo una parte mínima de la información correspondiente, en parte por la complejidad de los cuadros a que podrían dar lugar estos datos. En el caso de México, por ejemplo, en el censo de 1940, la información se redujo a la educación primaria, aunque las cédulas recogieron informes sobre la educación superior alcanzada.

Asistencia escolar

En relación con la comparabilidad internacional, los datos de asistencia escolar plantean varios problemas ya que algunos países preguntan acerca de las personas que asisten a la escuela en el momento del censo; en otras, la asistencia o inasistencia se refiere a un periodo determinado (del tantos del mes tal al tantos del mes tal); en algunos, hay amplitud en cuanto al intervalo de edades a que se refiere la pregunta (¿cuántos mayores de tantos años y menores de tantos otros asisten o no a la escuela en este momento? o ¿cuántos que se encuentran en esas condiciones de edad han asistido entre el día tantos y el día tantos?) En algunos países hay referencias concretas al tipo de institución al que se asiste (México, Polonia y Suecia hacen una categoría separada de las personas que asisten a instituciones como las academias comerciales y las escuelas de idiomas); en algunos otros se señala el nivel escolar al que se asiste y hay algunos (como Venezuela) que interrogan sobre las causas de la falta de asistencia a la escuela de personas en edad escolar. Las tabulaciones, también en este aspecto, sufren recortes considerables.

La clasificación de los asistentes o faltantes a las escuelas por grupos de edad, aparece en casi todos los censos, excepto en el mexicano de 1940, que, con ello, limita la utilidad de sus datos.

Población económicamente activa

Los datos relativos a la población económicamente activa interesan, principalmente, desde los ángulos económico y sociológico. Económicamente importan para determinar: 1) la disponibilidad de mano de obra; 2) la distribución de la misma por ramas de actividad en la producción total y el ingreso nacional, y 3) el avance económico. Sociológicamente, dichos datos interesan para determinar: 1) la constitución de los diferentes estratos económico-sociales; 2) la proporción de los asalariados frente a los no asalariados, y 3) la proporción de trabajadores manuales frente a los trabajadores intelectuales.

Fuentes sobre población económicamente activa

Las principales fuentes de obtención de datos acerca de la población económicamente activa son: 1) los censos de población; 2) los registros de la seguridad social (de desempleo, de enfermedad, retiro, etcétera), y 3) ya sean *a*) los registros de conscripción o *b*) ciertos muestreos que se realizan simultáneamente con el censo (en ocasiones).

La necesidad de emplear 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 —a más de los generales— consignan datos acerca de: 1) las personas empleadas en las organizaciones caritativas; 2) los trabajadores familiares que no reciben paga; 3) los trabajadores domésticos; 4) los trabajadores agrícolas, y 5) los empleados del gobierno.

Definiciones

En general, se define como "población económicamente activa" a "la parte de población que suministra el trabajo para la producción de bienes y servicios económicos". La definición incluye a quienes realizan una "operación lucrativa" o sea tanto a quienes son remunerados en dinero como a quienes lo son en especie, y sea que la remuneración la reciban en forma directa o indirecta. Por ello la definición abarca también a empleadores, a quienes trabajan por cuenta propia y a trabajadores miembros de la familia que no reciben pago en metálico.

Para fines censales, debe incluirse tanto a quienes se encuentran realizando un trabajo en el momento del censo como a los desempleados.

Criterios para la identificación de la población económicamente activa

Diferentes reuniones internacionales (del Comité de Expertos en la Liga de las Naciones, y la Sexta Conferencia Internacional de Estadísticos Laborales, principalmente) adoptaron dos criterios distintos para la identificación de los miembros de la población económicamente activa: 1) el criterio de la "ocupación remunerada" (o *gainful occupation*), y 2) el criterio de la "fuerza de trabajo" (*labour force*).

El criterio "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, tabulándose en seguida los datos que caen dentro del concepto "ocupación remunerada" que se delineó anteriormente. Este criterio tiene la ventaja de que permite caracterizar a las poblaciones desde el punto de vista socio-económico; tiene —en cambio— la desventaja de que carece de objetividad y precisión suficientes en cuanto hay personas que no tienen una ocupación única y bien definida (amas de casa, estudiantes, etcétera).

Conforme al criterio "fuerza de trabajo" se pregunta a los individuos cuál es su ocupación en el momento del censo o cuál fue o ha sido la de un corto periodo inmediatamente anterior al censo. Este criterio tiene la ventaja de ser objetivo; tiene la desventaja de que los datos obtenidos por este medio están afectados seriamente por condiciones atípicas que pueden imperar en el momento de recoger los datos censales (esto es particularmente importante en países en donde, en cortos espacios de tiempo, se producen grandes variaciones ocupacionales).

De lo anterior se desprende que los dos criterios no se excluyen, sino se complementan y que —utilizados con cuidado y en conexión con otras fuentes informativas acerca de los cambios de ocupación en cortos periodos— pueden proporcionar indicaciones valiosas acerca de la estratificación y la movilidad ocupacional, económica y social.

1.2. DECISIÓN CENSO-MUESTREAL

¿Censo o muestreo? o ¿censo y muestreo?

Censar es hacer una enumeración completa de todas las unidades de un conjunto.

Muestrear es seleccionar parte de un agregado para que represente a todo ese agregado.

El muestreo de material homogéneo es altamente representativo y confiable; en cambio, conforme aumenta la heterogeneidad del material, aumenta el riesgo que se corre al muestrear.

Conforme el material es más heterogéneo se requiere que la muestra sea mayor a fin de que sea más confiable el resultado; de modo que la confiabilidad total se logra cuando la muestra se confunde con el agregado todo que se enumera; o sea, cuando en realidad, se censa y no se muestrea.

Es indispensable distinguir entre muestra y censo incompleto. Cuando se carece de una parte de la información censal por no haberse obtenido, se habla de "censo incompleto"; sólo cuando se eligen, de acuerdo con un cierto *sistema*, ciertas unidades representativas se está autorizado para hablar de "muestra".

Cuando en una investigación se requiere una cobertura muy amplia de una población, lo aconsejable es censar. Con ello se logra una precisión muy alta, pero, naturalmente —en función de limitaciones de tiempo, dinero y energía disponibles— es necesario reducir el número de aspectos que se pueden cubrir.

Cuando en una investigación lo que se requiere es, en cambio, un gran detalle, hay que sacrificar, hasta cierto punto, la amplitud de la cobertura y en esas ocasiones es aconsejable el muestreo.

El grado en que se sacrifiquen amplitud y detalle tiene que determinarse en cada caso concreto, de acuerdo con las necesidades teóricas y prácticas de la investigación precisa que tenga que realizarse (entre las que cuenta, también, la velocidad con que se desean obtener los datos correspondientes).

Complementación de censo y muestreo

En muchas ocasiones, lo más conveniente es que una enumeración completa de la población, universo o agregado se complemente con uno o varios muestreos. En realidad, se puede pensar en toda una batería censo-muestreal que vaya del censo (que cubre el máximo de aspectos) a la muestra micrométrica y aun al estudio de casos (que cubre un mínimo de casos, pero con un máximo de detalle).

El muestreo

Investigación muestreal

El muestreo es una técnica estadística que toma una parte de un colectivo (universo de cosas o población de personas) en forma tal que las medidas estadísticas o ("estimadores") de esa porción o muestra puedan servir para estimar las medidas estadísticas correspondientes del colectivo, dentro de ciertos márgenes de seguridad.

Como puede verse, el muestreo implica las siguientes operaciones, interrelacionadas, pero distintas: 1) la toma de la muestra o porción representativa; 2) la medida de los valores estadísticos de la muestra, y 3) la inferencia de los valores poblacionales (estimados) a partir de los muestreales. Estos aspectos corresponden propiamente a: 1) la técnica constitutiva de la muestra misma; 2) la técnica de cálculo estadístico, en general, y 3) uno de los aspectos de la estadística inferencial, que, a su vez, se basa en la teoría de las distribuciones.

En efecto, los estimadores (o estadísticas de la muestra) se distribuyen en formas características (regidas por fórmulas y curvas típicas). Gracias al tipo de distribución al que corresponden, se puede precisar cuál es la probabilidad de que el valor calculado para la muestra sea igual al valor correspondiente de la población (en este sentido, este último es puramente un valor "estimado" y no uno calculado directamente).

Es por esto por lo que el muestreo enfatiza el aspecto metodológico de cómo se ha de tomar la muestra. Si la toma de la muestra (en realidad, la determinación del volumen que ha de tener y la selección de las unidades que deberán constituirla) no obedece a un proceso consciente y voluntario desde el ángulo matemático, no se conocerá la forma (también matemática, analítica o geométrica) en que se distribuirán —junto con sus propios estimadores— todos los estimadores análogos de las muestras parecidas que se podrían tomar de la población.

Éste será el aspecto que trataremos exclusivamente en este apartado —conforme a nuestro designio analítico inicial— dejando para más adelante la fundamentación de la inferencia estadística y la descripción de la técnica inferencial misma.

Conceptos fundamentales

Del concepto mismo de muestreo se desprenden algunos conceptos subordinados. Conviene subrayar algunas características de éstos, y agregar algunos más.

En un muestreo hay *una* población, pero hay *múltiples* muestras *posibles*. En la práctica no se obtienen todas las posibles, pues con ello se multiplicaría el trabajo estadístico y se encarecería la labor investigatoria. Como de lo que se trata es de simplificar el trabajo y abaratarlo (dentro de lo posible y deseable) se obtiene por lo general en la práctica *una* sola muestra *real*, mientras que en la experimentación de designio teórico, hay que tomar siempre más de una.

La población tiene ciertos valores estadísticos que la caracterizan. Cada uno de ellos, o es único, o forma con otros —interrelacionados— un conjunto único; así, por ejemplo: hay una media aritmética para un colectivo, y hay, para el mismo, un único conjunto de cuatro cuartilas. Cada muestra, a su vez, tiene su estadística propia o su propio conjunto de estadísticas interrelacionadas. Todas las muestras tienen, juntas, un conjunto de valores estadísticos correspondientes a cada valor individual de la población; así hay un conjunto de medias aritméticas procedentes de las diferentes muestras, mientras hay una media aritmética única de la población; hay un conjunto de cuartetas de cuartilas de las muestras, que corresponden a una sola cuarteta de cuartilas en la población.

Cada conjunto de estadísticas (todas las medias aritméticas de todas las muestras; todas las desviaciones medias de todas ellas, etcétera) tienen sus formas específicas de distribución. Esto significa que, dentro del conjunto, los valores pequeños de esas estadísticas tienen una probabilidad de aparición distinta de la que corresponde a los valores medios, y diferente de la que es propia de los valores altos. Así, las medias estadísticas del conjunto de muestras de una población (elegidas conforme se dirá más adelante) se distribuyen normalmente, o sea, que la probabilidad de que aparezca una de ellas corresponde a la dada por un tipo de curva a la que se conoce como "curva normal".

Tamaño y número de las muestras

Cada muestra posible de un colectivo está formada por una o varias unidades; por uno o varios de los individuos de dicho colectivo, universo o población. Un problema muestreal —que no es, con todo, el más importante— consiste en determinar cuántas unidades debe contener la muestra; cuántas conviene que tenga.

En principio, una muestra puede tener desde una hasta tantas unidades como sean las que constituyan la población. En forma correspondiente —también en principio— el número de muestras posibles es varias veces superior al número de unidades del colectivo. En caso extremo, puede haber tantas muestras como individuos tenga el colectivo (muestras de una unidad); en otro extremo, puede haber una sola muestra (formada por todas las unidades del colectivo). Pero hay otros extremos que, por otro lado, superan el mismo número de individuos del colectivo.

En el intermedio, hay muestras de dos, de tres, de cuatro, de ene unidades. El número de muestras de dos unidades cada una, que pueden formarse a partir de un colectivo de seis unidades es, por ejemplo, de 15. Este número está dado por las combinaciones de 6 objetos tomados de 2 en 2. En general el número de muestras posibles de un colectivo de N objetos o personas, tomados de n en n unidades muestreales, estará dado por las combinaciones de N en n, o sea por $\binom{N}{n}$.

Muestreo con y sin reemplazo

La forma anterior de constituir la muestra implica que si una unidad (a) ha salido en una muestra, no puede volver a figurar en ella. Al procedimiento se le designa como "muestreo sin reemplazo". Cuando una vez que sale una unidad muestreal en una muestra, puede volver a figurar en otra, se habla de "muestreo con reemplazo".

El muestreo sin reemplazo es más natural y preciso; pero da lugar a que aparezcan complicaciones matemáticas. Esto se evita, en la práctica, mediante el muestreo con reemplazo, menos natural y preciso, pero que simplifica el andamiaje matemático.

El problema principal del muestreo

Más importante que los anteriores es un problema muestreal que se relaciona al del tamaño de la muestra y es el de cómo elegir las unidades de la muestra real.

La principal solución al problema de cómo muestrear, es genérica: lo que se debe evitar es que el procedimiento sea arbitrario y parcial. No debe ser arbitrario, sino estar regulado por consideraciones matemáticas, y no debe ser parcial, sino imparcial para todas las unidades del colectivo.

El procedimiento no debe ser arbitrario porque, en la arbitrariedad, en la inconsciencia, en la falta de conocimiento del método preciso por el que se haya podido obtener una muestra no se puede saber cómo se distribuyen las estadísticas del conjunto de muestras posibles y también es imposible determinar qué probabilidad tiene cada una de ellas de representar la medida correspondiente de la población.

Tampoco debe ser parcial el procedimiento, porque si prefiere ciertas unidades del conjunto para que formen parte de la muestra, y deja preteridas a otras posibles unidades muestreales, la muestra no representará a la población toda, sino sólo a la parte de la población que haya sido *preferida* por el investigador. Así, habrá en los resultados no sólo elementos de hecho, sino valoraciones, regidas —éstas— por la idiosincrasia del investigador individual o por la ideología del grupo que investigue. La realidad será vista, pero a través de unos anteojos deformantes. Cuando esto ocurre, se dice que la elección muestreal fue prejuiciada.

A fin de evitar lo azaroso (que consideramos semánticamente distinto de lo aleatorio, en estadística) y lo prejuiciado, se recurre a una regulación matemática: el muestreo debe estar sujeto a probabilidades claramente determinables (debe ser aleatorio), pues toda la estadística basa sus resultados en fundamentos probabilitarios. Simultáneamente, al dar a cada unidad de la población igual oportunidad que a las otras para constituir la muestra, se asegura la imparcialidad, la falta de prejuicios en favor o en contra de ciertas porciones de la población.

Estratificación

La precisión de las estimaciones que se hagan a partir de una muestra se puede aumentar si el colectivo se divide en varios subcolectivos y de cada subcolección se toma una muestra aleatoria. A las subcolecciones en que se divide un colectivo se les conoce también como estratos, y al procedimiento de división, estratificación.

Mediante una estratificación conveniente se excluyen las posibilidades (muestras) extremas, que aparecen en un muestreo simple, o no estratificado.

Fracciones de estrato constantes y variables

Los estratos en que se divida un colectivo pueden ser iguales o distintos por el número de unidades que los formen. Si se atendiera a un criterio puramente estadístico se podría pensar en la conveniencia de formar estratos iguales; pero, cuando hay que considerar, sobre todo, el punto de vista social, conviene pensar en estratos de diferente tamaño.

Independientemente del tamaño de los estratos, y en forma más importante, hay que reconocer dos posibilidades distintas: la de que de cada estrato se tome siempre una misma fracción (una fracción muestreal común a todos los estratos del colectivo) y la de que, para cada estrato, se tome una fracción distinta.

Cuando se muestrean los estratos con fracción variable, esa variabilidad puede ser arbitraria o regulada. Como ya se ha dicho en otra ocasión, la arbitrariedad impide conocer la forma en que las decisiones correspondientes afectan los resultados, y la sujeción a una regla —en cambio— permite conocer la forma en que las mismas los favorecen o los dañan.

En general, si hay una fracción muestreal promedio, se dice que se "sobremuestrean" aquellos estratos de los que se toma una fracción mayor (mayor número de unidades muestreales) y que se "submuestrean" aquellos otros de los que se toma menor fracción muestreal (menor número de unidades).

Para saber qué estratos deben sobremuestrearse y cuáles submuestrearse, se necesita considerar qué tan variables son los datos dentro de cada estrato. Un estrato de variabilidad mayor que la variabilidad media se debe sobremuestrear; uno de variabilidad menor que la variabilidad media de los estratos, se debe submuestrear.

En efecto, la estratificación con fracción variable *puede ser* un instrumento que ayude a incrementar la precisión; *pero es, también* "un instrumento de doble filo, que puede volverse en contra nuestra si concentramos el muestreo en los estratos menos variables".

Puede decirse, en efecto, que la estratificación con fracción constante siempre aumenta la precisión (debido a que estratifica y elimina fuentes extremas de variabilidad). La estratificación con fracción variable (en cambio) aumenta o disminuye la precisión: la aumenta si las fracciones muestreales mayores se toman de los estratos más variables y las menores de las que lo son menos; la disminuye si las fracciones muestreales mayores se toman de los estratos menos variables y las menores de los que varían más.

Estratificación óptima

La óptima estratificación se logra cuando los estratos tienen medias aritméticas tan distintas como se pueda, y desviaciones cuadráticas medias tan pequeñas como sea posible. O sea, que debe haber máxima variabilidad *entre* los estratos y variabilidad mínima *dentro* de cada estrato.

El contraste de *entre* y *dentro* —debe decirse de paso— es de extraordinaria importancia en estadística (particularmente en relación con el "análisis de la variancia", como se verá después) y en el campo sociológico de las comparaciones intersocietarias lo es, también, de modo muy relevante.

Racimos muestreales

Mientras la estratificación divide a un colectivo en subcolecciones y toma muestras de *dentro* de cada subcolección, la formación de racimos divide a un colectivo en subcolecciones, pero decide, mediante muestreo, cuáles de *entre* esas colecciones se habrán de estudiar, para que se las tome como representativas del universo o colección.

En la estratificación se toman las unidades muestreales, en forma aleatoria, *de todos* los estratos; en el muestreo por racimos, se toman en forma aleatoria, unos (o *algunos*) racimos sí y otros no.

Estratificación y enracimamiento, o conglomeración

En la estratificación se gana precisión; en el muestreo por racimos se pierde precisión. La estratificación tiene la dificultad de que los estratos convenientes o no se encuentran siempre en forma inmediata en la realidad u oponen dificultades para su estudio; el muestreo por racimos —en cambio— procede sobre la base de *agrupamientos que ya ha formado la realidad*, que presentan contigüedad y facilidad práctica de estudio.

Los racimos naturales pueden quedar ejemplificados por los hogares de un vecindario. Sin embargo, como los miembros de un mismo hogar tienden a parecerse, estos racimos no satisfacen aquella prescripción del muestreo por racimos, según la cual, para incrementar la precisión de los resultados obtenidos por este tipo de muestreo se debe *aumentar la variabilidad dentro de cada racimo.* Como es fácil comprender ésta es una prescripción completamente opuesta, a la que rige el muestreo estratificado, pues para que en éste aumente la precisión, se debe *disminuir la variabilidad dentro de cada estrato.*

Muestreo en una o en varias etapas

La precisión obtenida de las muestras puede aumentar si en vez de realizarse el muestreo en una sola etapa, se realiza en varias. Una primera etapa se puede cubrir si se saca una muestra del conjunto de entidades de un país; una segunda, se puede llenar si se muestrea el conjunto de los distritos de cada una de las entidades de la primera; una tercera, al muestrear el conjunto de ayuntamientos de cada distrito de la segunda; una cuarta, si se hace sobre las casas de esos ayuntamientos, y una quinta, si se sortea a los individuos que viven en las casas elegidas en la etapa anterior.

Con este tipo de muestreo, se gana en flexibilidad, pero se pierde en sencillez. En efecto, en cada una de las etapas del muestreo multietápico, se puede emplear una técnica muestreal distinta (simple, estratificada, por racimos; con fracciones muestreales constantes o con fracciones muestreales variables) según convenga a la naturaleza peculiar de la colección de unidades que se muestree en cada etapa.

Tiene que ser producto de experiencia, reflexión y prudencia el elegir el conjunto de técnicas más apropiado y armonioso para este tipo de muestreo que, cuando se realiza correctamente, puede llegar a tener los caracteres de una obra de arte. En efecto, hay una combinación óptima de técnicas muestreales para cada colectivo y, entre los conjuntos prácticamente infinitos que pueden obtenerse de combinar todas las técnicas disponibles (no menos de cinco), las múltiples etapas posibles, y las formas también múltiples de estratificar y formar racimos de un mismo colectivo (especialmente si es complejo, según ocurre con los colectivos de unidades sociales), hay una conjunción óptima que requiere enorme talento y paciencia para su constitución.

1.3. RECOLECCIÓN DE LOS DATOS ESTADÍSTICOS

La recolección directa de datos estadísticos

El cuestionario, la cédula y el cuadro

El cuestionario, la cédula y el cuadro constituyen variantes de un mismo implemento de investigación del que se valen las ciencias sociales para hacer comparables las observaciones de diferentes investigadores individuales o de un mismo investigador que los aplica en diferentes ocasiones.

Estos instrumentos buscan, así, fundamentalmente: i) descargar la memoria del investigador al constituirse en una *guía escrita* para sus preguntas y un *esqueleto* en el que consignar sus respuestas; ii) la recolección de los mismos datos o rasgos en todos los casos; iii) la comparabilidad de las observaciones de diferentes investigadores por: *a*) separación analítica de diversos aspectos de la realidad estudiada, a los que se considera como elementales, y *b*) especificación previa de las unidades en que se han de consignar las respuestas respectivas, o *c*) el ámbito dentro del que se consideran válidos ciertos conceptos contenidos en las preguntas o en las respuestas esperadas.

El cuestionario e interrogatorio —como su nombre lo indica— es un conjunto de "cuestiones", "interrogantes" o preguntas; pero, para servir técnicamente a la investigación social tiene que reunir ciertas características entre las que las más importantes son las de ser: 1) unitario, o sea, girar en torno de un problema o situación, o de problemas y situaciones conexos; 2) ordenado, con fines prácticos, a fin de evitar lagunas, duplicaciones y traslapamientos indeseables (en tanto que preserva los que pueden servir con fines de comprobación o crítica ulterior); 3) de propósito sistemático, pues a más de conexión y orden, el cuestionario debe reflejar el sistema hipotético del que procede y el sistema teórico al que pretende llegar; 4) práctico, en cuanto debe de considerar las limitaciones de la situación propia de la pesquisa, las de preparación de los investigados y las de su capacidad de comprensión.

Cada uno de estos aspectos debe arrojar luz sobre todos los anteriores, y el cuestionario debe ser el resultado de la conjuntación armónica de todos ellos.

El cuestionario, la cédula y el cuadro no son sino variantes de un mismo implemento técnico, según hemos dicho. La diferencia que existe entre la entrevista, por una parte, y el cuestionario, la cédula y el cuadro, por otra, pueden entenderse mejor a la luz de las reflexiones de un lingüista como Terracini. El paso de la una a los otros es el de lo personal a lo impersonal; un aspecto más del estilo moderno, concentrado (propio de los telegramas, los pasaportes, los telegramas). En cada paso que se da en esta escala técnica "se lleva más y más hasta el extremo el esfuerzo por fichar, clasificar, planificar a los hombres y sus acciones".

En la construcción de un cuadro, hay que tener presente cuál es el objeto central para el cual se construye y cuáles los elementos indispensables, necesarios o simplemente auxiliares para el conocimiento de ese objeto. Esto permite eliminar del cuestionario, la cédula o el cuadro cuanto pudiera resultar simplemente interesante pero superfluo para el conocimiento, del objeto. A fin de construirlo adecuadamente, se debe hacer, antes, un cuadro "imaginario" o "fantasma" y el investigador debe —en seguida— tratar de llenarlo, o de hacer que trate de llenarlo un colaborador; en cuanto él o su colaborador tropiecen con alguna dificultad para preguntar o para contestar, deberán marcar el cuadro fantasma con lápiz rojo, a fin de modificar ulteriormente el punto o suprimirlo según convenga.

Después, conviene hacer una experiencia de prueba, utilizando a una porción pequeña de la población por estudiar, para apreciar las dificultades objetivas y las reacciones subjetivas de los investigados. Éstas se relacionan, generalmente, con: i) la insuficiente capacidad anticipatoria de los investigadores en lo que se refiere a las situaciones que pueden encontrar (en cuanto muchas de éstas son inesperadas); ii) a la incapacidad para redactar las preguntas en forma suficientemente clara para los investigados (en particular, a la incapacidad para adecuar el tono general del cuestionario, la cédula o el cuadro al nivel de instrucción de la población a la que se investiga), y iii) a la incapacidad para prever las reacciones desfavorables de los investigados, para suscitarlas tan poco como se pueda, mediante un planteamiento adecuado de las preguntas correspondientes).

Dentro de esta tónica, al redactar el cuestionario hay que recordar que no se está redactando una pieza literaria, sino se está elaborando un instrumento que debe ser funcionalmente útil a la investigación. Se debe recordar también que las preguntas se deben adecuar a los modos expresivos de los investigados, pero sin descender a tales niveles de expresión que hagan que resulte artificioso para el propio investigador o que provoque reacciones de burla de parte de los investigados al ver que éste trata de imitar —infructuosamente o sin autenticidad— sus modos expresivos. La presentación física, el tamaño y forma de los cuadros también deben cuidarse ya que, en muchas ocasiones, todos esos caracteres influyen en los resultados.

En síntesis, las fases por las que atraviesa la construcción de un cuestionario, de una cédula o de un cuadro son:

1º Formación de una lista de preguntas.

2º Constitución de una serie de cuadros-fantasmas.

3º 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, y
c) una representación física adecuada.

4º Realización de algunas experiencias de prueba.

5º Elaboración del cuadro definitivo.

6º Instrucciones para los recolectores de datos.

Al llenar cualquiera de estos instrumentos de trabajo, se debe de buscar que las respuestas sean inteligibles y que su registro sea legible. Para ello, se debe de aclarar el sentido de la respuesta cuando ésta sea ambigua (consignando, en el margen del cuadro, la pregunta auxiliar utilizada por el investigador para resolver la ambigüedad), y hay que evitar, en el momento de escribir, las abreviaturas y los signos de difícil o equívoca interpretación (a menos que se haya previsto la posibilidad de utilizarlos y se haya codificado su significado).

En ciertos casos se puede pensar en utilizar la taquigrafía, pero ésta, si bien acelera la recolección sobre el terreno, impone al recolector el trabajo de traducción de sus registros, por lo cual, en caso de usarla debe ejercer una gran vigilancia sobre dicho procedimiento.

Asimismo se deben de llenar con una línea aquellos renglones del cuestionario, la cédula o el cuadro para los que no se obtuvo respuesta. En caso de ser posible, se debe anotar también, con lápiz rojo (o, en general, de color distinto de aquel en que se anotan las respuestas directas del investigado) las circunstancias en que se produjo la laguna correspondiente. Con esto se evita que llegue a pensarse que un espacio dejó de llenarse por descuido del recopilador, y se posibilita una ulterior interpretación de las faltas de respuesta, las cuales —en veces— son tan sintomáticas o ilustrativas como las respuestas mismas.

Mientras el cuestionario conserva en buena parte los rasgos intersubjetivos (de presencia física del investigador), que caracterizan más particularmente a la entrevista, la cédula— y más aún el cuadro —tiende a perder este elemento. Lo intersubjetivo sigue existiendo, por supuesto, pero en forma difumada: el investigador raras veces es, en tales casos, una persona física; por lo general, es una persona moral, una institución social. Hay por supuesto, situaciones de transición: corresponden a aquellos casos en que no existe propiamente un investigador que interrogue, pero sí una persona que proporciona las formas (cuestionarios y más frecuentemente cédulas) en las que los investigados deben consignar sus informes. La situación responde bien a la investigación que procede de organismos gubernativos, burocráticos.

La administración de todo lo que no es entrevista o cuestionario se hace generalmente en forma personal (como la burocrática, ya indicada) o por correo (las formas se envían para ser llenadas y se devuelven llenas). Otra situación de tránsito se produce en la investigación a través del radio o la televisión, en la que hay un eco o residuo personal del investigador y las respuestas se reciben por correo.

La investigación personal permite que las preguntas sean más complejas y permite, también, deshacer las ambigüedades tanto de la pregunta como de la respuesta. La investigación que no es de este tipo lleva, interiormente construidos, los factores de su propia ambigüedad, que, a menos que el instrumento se haya construido con extremada cautela y pericia, acaba por afectar los resultados mismos.

En general puede decirse que los renglones de cuestionario, cédula y cuadro deben ser: i) pocos; ii) sencillos; iii) de preferencia sobre hechos objetivos o sobre cosas acerca de las que pueda responderse mediante opciones (sí, no; alto, mediano, bajo) o mediante un número.

En este punto, quizás sea prudente llamar la atención hacia el hecho de que si bien la estadística cuenta con un riquísimo arsenal de técnicas para el manejo de datos numéricos cardinales, la propensión de los investigados es a proporcionar datos mucho más burdos que, generalmente, no se expresan en forma de medidas precisas, sino en términos de órdenes de magnitud, de series ordenadas por "rango". Esto significa que deben de tratar de desarrollarse al máximo esas técnicas estadísticas ordenatorias; que inicialmente debe pensarse en utilizarlas casi exclusivamente, y que, sólo en términos de refinamiento y cuando se cuenta con datos que no se han recogido directamente de informantes no calificados, sino cuando se dispone de aquellos que han recolectado los informantes calificados, habituados a los procesos de medición, deben hacerse intervenir las otras técnicas estadísticas más finas.

Recolección de datos para construir escalas

Una escala sociométrica es un instrumento usado en el mensuramiento social para categorizar o jerarquizar: 1) aspectos del medio social o cultural o 2) conductas sociales, actitudes y opiniones. Estas escalas tratan de hacer comparables al máximo, sobre una base de rigor y objetividad extremados, dos o más instituciones, comunidades o sociedades, o de determinar grados de separación de la conducta de los miembros de una sociedad con respecto a las normas grupales. En el fondo, una escala, como la mayoría de las otras formas de investigación, se basa en una forma o modalidad particular del cuestionario. La modalidad propia de éste, en la construcción de las escalas depende de que: 1) las preguntas se hacen en dos etapas y no en una sola; 2) cubren no sólo a los investigados mismos, sino distinguen entre un grupo de informantes calificados (jueces) y otro de informantes no calificados; 3) las preguntas se hacen en forma disfrazada y extremadamente simple y dan lugar 4) a respuestas que concretan en una simple ordenación de ciertos aspectos que se someten a la consideración de los investigados.

En la construcción de una escala sociométrica se necesita:

1. Determinar qué se va a medir.
2. Seleccionar los criterios y elementos en que se basará la medición.
3. Hacer una escala sencilla.
4. Valorar y calificar cada elemento de la escala.
5. Someter la escala a la población por investigar.

La recolección de los datos gravita principalmente en torno de los puntos tercero y quinto.

Una escala sociométrica debe ser generalmente aplicable al agregado que se ha de trabajar y conservar a lo largo de su aplicación, su seguridad y validez. De ahí que haya de buscarse trabajar con colectivos tan homogéneos como sea posible y que se recomiende también que, en caso de tener un colectivo heterogéneo, se le divida en dos o más porciones homogénas y se elaboren tantas escalas distintas como porciones homogéneas haya. Esto, a más de contribuir a la seguridad y validez de la escala, coopera a la economía de la investigación pues, en cada escala, se reducen a un mínimo los renglones inaplicables.

Uno de los problemas principales de la elaboración de una escala consiste en adoptar un criterio para seleccionar los renglones o factores que hay que considerar dentro de la misma. Para ello, se recurre a personas que conozcan el terreno por medir. A éstos se les llama "jueces". A ellos se les somete una lista de elementos que, sobre una base deductiva se hayan descubierto como componentes del problema por estudiar, a fin de que, con base en su conocimiento de la realidad, así como de sus valoraciones del problema, incluyan o excluyan y ponderen diferencialmente los que han de intervenir en la escala.

Cuando se trata de problemas complejos (como el estudio integral de una comunidad) se tiene que recurrir a dos clases de jueces: los "jueces generales" que especifican qué grandes grupos de factores (entre los sanitarios, económicos, políticos, culturales, etcétera) han de considerarse para la medición de una actitud, de una opinión, de unas conductas, en tanto que los ("jueces especiales") médicos, economistas, jurisperitos, politólogos, antropólogos) determinan, en cada uno de esos aspectos, qué es lo que es necesario incluir en la escala.

Una vez que los jueces generales y especiales han seleccionado los aspectos que se han de considerar en la escala, se procede a calificar cada aspecto (esto, por corresponder a una sección diferente, a un tipo de operación distinta, no lo trataremos en este punto). Como resultado, se obtiene una serie de "indicadores" de la actitud, opinión o conducta, y una serie de calificaciones para cada indicador.

El siguiente paso recolector de datos consiste en aplicar la escala a la población investigada. De esta aplicación se obtiene una serie de respuestas de los individuos. Un individuo A puede mostrar, por ejemplo, la existencia del indicador 1, y no la del 2... Otro, la del 2 y no la del 1; otro más la del 1 y la del 2; otro más ni la del 1 ni la del 2. Cada uno de estos casos representa un distinto grado de intensidad de la actitud, opinión o conducta que se investiga. Dado que a cada uno de estos indicadores corresponde una calificación en la escala, la calificación total se obtiene al sumar las calificaciones de todos los aspectos o indicadores presentes. Así, el primer individuo tendrá por calificación 4 si 1 vale 4 y 2 vale 6. El segundo, en ese mismo supuesto, tendrá como calificación 6, el tercero 10 y el cuarto 0.

La recolección indirecta, de fuentes secundarias

Sobre la información estadística en México

Las estadísticas mexicanas se clasifican de acuerdo con tres criterios:

1. Por su amplitud.
2. Por su periodicidad.
3. Por su materia.

De acuerdo con su amplitud, se pueden distinguir:

1.1. Censo.
1.2. Muestreo.

Según su periodicidad, se puede hablar de:

2.1. Estadísticas aperiódicas.
2.2. Estadísticas periódicas.

Estas últimas se clasifican en:

2.21. Estadísticas de periodo largo.
2.22. Estadísticas de periodo corto o estadísticas permanentes.

De acuerdo con sus materia, las estadísticas son:

3.1. Económicas y
3.2. Sociales.

Censo y muestreo

Un censo es una enumeración directa y simultánea de todas las unidades sometidas a recuento.

Una muestra es una enumeración de una parte de las unidades sometibles a recuento, que representa el total, en forma matemáticamente definida, que permite inferir las características del todo, a partir de las características de esa parte suya.

Estadísticas periódicas y estadísticas permanentes

"Estadísticas periódicas" es el nombre que se da a aquellas que se elaboran periódicamente con intervalos largos, de cinco o diez años. Las "estadísticas permanentes" son también estadísticas periódicas pero cuyo periodo es muy corto (pues son, generalmente, anuales o mensuales).

Estadísticas económicas y sociales

Las estadísticas son, de acuerdo con su materia, ya económicas, ya sociales. Las económicas se clasifican por ramas de actividad, en agrícolas, mineras, industriales, bancarias, de transporte... Las sociales describen la vida familiar, los rasgos culturales, la ocupación, el ingreso de los mexicanos.

En México, los servicios estadísticos están encargados a una Dirección General de Estadística, dependiente de la Secretaría de Industria y Comercio, que está asesorada por un Consejo Consultivo, y que mantiene relaciones con Secretarías como las de Agricultura y Ganadería, la de Salubridad y Asistencia, la de Comunicaciones y Transportes y la de Obras Públicas, con Petróleos Mexicanos, los Ferrocarriles Nacionales y las Comisiones Bancarias de Valores, de Seguros, y con el Banco de México.

La dirección está organizada en Departamentos (Censal, de Estadísticas Económicas, de Estadísticas Económicas Básicas, de Estadística Social, Técnico y de Agrupamientos Mecánicos y Cómputo Electrónico).

Los censos

Los Censos Nacionales son de la exclusiva competencia de la Dirección Nacional de Estadística y comprenden:

1.11. Los censos de población.

1.12. Los censos agrícola-ganadero y ejidal.

1.13. Los económicos (industrial, comercial, de servicios y de transportes).

Las estadísticas permanentes

Las estadísticas permanentes recopiladas y elaboradas por la Dirección General de Estadística son:

2.221. demográficas,

2.222. educativas,

2.223. económicas,

2.224. monetarias y bancarias,

2.225. agrícolas.

Los censos de población

Los censos de población se realizan cada diez años en año terminado en cero. En 1960 se realizó el octavo y en 1970 el noveno. La información que de ellos se obtiene se presenta en:

Un resumen general a nivel nacional, por entidades.

Varios volúmenes a nivel de entidad por municipios (algunas entidades por el gran número de sus municipios están comprendidas en dos volúmenes del censo).

Ese resumen y esos volúmenes proporcionan informes sobre:

1. Edad.
2. Sexo.
3. Instrucción.
4. Estado civil.
5. Ocupación.
6. Características de la habitación:
   6.1. material de construcción,
   6.2. número de cuartos,
   6.3. número de ocupantes,
   6.4. servicios.

Los censos agrícola-ganadero y ejidal

Los censos agrícola-ganadero y ejidal también se realizan con periodicidad de diez años, en los años en que se realiza el censo de población, y como éste se presentan en un resumen nacional y en volúmenes por entidades. Se refieren a:

1. Propietarios de los predios.
2. Clasificación de las tierras.
3. Maquinarias, implementos y vehículos.
4. Personal ocupado.
5. Productos y producción.

Estadísticas permanentes

Entre las estadísticas permanentes, las económicas se refieren a:

1. Comercio exterior.
2. Producción industrial.
3. Producción minero-metalúrgica.
4. Producción de energía.
5. Producción de petróleo y sus derivados.
6. Finanzas públicas nacionales, estatales y municipales.
7. Transportes y comunicaciones.
8. Banca y moneda.

Algunos de estos datos se publican en los anuarios de los órganos respectivos; otros en boletines como los de las comisiones de valores y de seguros.

Estadísticas permanentes de materia social

Entre las estadísticas sociales permanentes se cuentan:

1. Las demográficas.
2. Las educativas.

Las demográficas tienen como fuente las oficinas del registro civil. La Dirección General de Estadística presenta cifras mensuales y tabulaciones anuales nacionales, estatales y municipales, de nacimientos, matrimonios y defunciones, por sexo, edad de la madre de los recién nacidos, edades de los contrayentes y edad del fallecido.

Esas estadísticas se complementan con otras, del movimiento social de la población, como son las de migración y turismo.

Las estadísticas educativas cubren desde los jardines de niños hasta la educación superior y, muy recientemente, se ha hecho cubrir la distribución de la población escolar por grados, sexos y edad.

Otras estadísticas de carácter social son las que se refieren a los aspectos laborales y judiciales de la actividad mexicana.

1.4. ELABORACIÓN DE LOS DATOS ESTADÍSTICOS

Codificación y análisis de contenido

La codificación es un paso indispensable en la investigación. Lo es, particularmente, porque las respuestas a un cuestionario, a una cédula, a un cuadro están dadas en forma textual —con palabras y expresiones del lenguaje diario— y el manejo de las mismas tiene que hacerse —en el caso de la estadística— en forma cuantitativa.

La codificación es, fundamentalmente, un proceso por el cual las respuestas a un interrogatorio o los datos de un documento, se expresan en forma tan sencilla como es posible, a fin de revelar, así, su contenido esencial; a fin de mostrar las semejanzas y diferencias básicas entre los individuos que han respondido o entre los casos descritos por los documentos que se han examinado.

La codificación puede considerarse como la forma más simple, inmediata y —en todo caso— como el cabo terminal de lo que, más ampliamente, constituye el "análisis de contenido". En efecto, en tanto que la desnuda codificación representa —en la mayoría de las ocasiones— un procedimiento de traducción simple de ciertas expresiones literales a símbolos numéricos, el análisis de contenido implica, en forma más compleja, que ha de tomarse un texto —particularmente, el que resulta de una entrevista, el que se obtiene de consultar un archivo, el que se enfrenta al examinar un periódico— y descoyuntarlo en sus partes significativas, a fin de poder manipularlo estadísticamente.

El análisis de contenido es un proceso complejo y delicado; por ello mismo, no puede dejarse sujeto sólo a la intuición del investigador; en efecto, puede y debe beneficiarse de las aportaciones lógico-matemáticas y de ciertas técnicas y concepciones lingüísticas como el análisis estructural o la glosemática de Hjelmslev. Con todo, ese análisis de contenido, puesto al servicio de las ciencias sociales, no puede ni debe agotarse en la combinación de lógica y lingüística, entre otras cosas, debido a que, en toda expresión inciden actitudes e ideologías que requieren especiales formas de observación y análisis.

Como es fácil comprender, el análisis de contenido es útil, muy especialmente, cuando la técnica de recolección de datos ha sido la entrevista o cuando los datos se recogen de una fuente secundaria en la que es posible prever que no ha habido una guía estricta de investigación. En cambio, la codificación pura y simple —de aplicación inmediata— puede servir, sobre todo, cuando el medio de recolección de datos ha sido el cuestionario (o alguna otra de las técnicas emparentadas con él).

En sentido metodológico más amplio se puede y debe concebir el análisis de contenido como un articulador que vincula dos técnicas aparentemente dispares, juzgadas por algunos como opuestas entre sí y que —en realidad— son complementarias: la técnica de casos y la técnica estadística. En efecto, de esas dos, una representa un enfoque sintético e individualizador; la otra un enfoque analítico colectivizador, pero de una a otra hay vías de tránsito como el análisis de contenido, que descoyunta las partes de un caso y permite relacionar cada una de ellas con las del mismo carácter, que proceden de otros casos, y como la tipificación social que, de las porciones disjuntas que aporta la elaboración estadística de las diferentes partes de un todo, forma, por conjuntación, un caso hipotético, un tipo ideal, construido a partir de la realidad.

La codificación misma —conforme señala Moses— se realiza en dos etapas: en la primera se establece el marco codificador; en la segunda, las respuestas individuales se ubican y distribuyen dentro de ese marco.

En muchas ocasiones, el marco de la más sencilla de las codificaciones está constituido por una alternativa de respuesta (positiva y negativa) y por las categorías que complementan la categoría principal de "quienes respondieron (en una u otra forma") con la de "quienes no pudieron o no quisieron responder" y con la de "aquellos a quienes no les era aplicable la pregunta". De este modo, típicamente, un marco codificador presenta las categorías siguientes:

Respondieron { Afirmativamente
Negativamente

No respondieron

La pregunta no les era aplicable.

A fin de facilitar la ulterior codificación, es frecuente que los cuestionarios no se hagan abiertos, sino cerrados; o sea, que es frecuente que en el cuestionario mismo se incluyan, frente a cada pregunta, las alternativas de respuesta esperadas.

En tales casos, con fines de aplicación expedita del propio cuestionario, conviene que en la forma impresa aparezca, en primer término, la categoría "no aplicable" pues eso elimina todas las restantes y remite al interrogador a la pregunta inmediata siguiente; que, en segundo término, se consigne la categoría de quienes "no respondieron", que también permite ese paso inmediato y que sólo en último término figuren quienes "respondieron" con sus diversas respuestas.

En la elaboración, el proceso debe ser —en cierto modo— inverso: interesa más conocer el número de "inaplicables", pues éste arroja luz sobre el cuestionario mismo, y el número de "rechazos" que es significativo en cuanto al conocimiento y las actitudes de quienes respondieron, en relación con la pregunta, debiendo considerarse, en último término, las respuestas mismas. De otro lado, la proporción de "inaplicables" y de "rechazados" dentro del total de casos examinados, y de éstos dentro del total de los examinables, da idea de la representatividad inicial de los resultados; del grado de reducción de la misma por encogimiento de la muestra planeada inicialmente, e indica, en aquellos casos en que este encogimiento es muy grande, la conveniencia de tratar una pregunta poco aplicable o con muchos rechazos, o simultáneamente poco aplicable y muy rechazada, en forma distinta a las demás preguntas. En caso extremo, puede indicar, también, la utilidad de realizar una investigación más particularizada, detenida y profunda en el sector cubierto por la pregunta afectada por la inaplicabilidad y el rechazo. Ésta, en su momento, deberá redactarse en otra forma y complementarse con otras preguntas que exploren, particularme, las actitudes de la población, y el origen probable de las mismas.

En la codificación suelen aparecer preguntas que no han sido contestadas mediante una simple opción entre un sí y un no, sino que implican una selección de una entre varias categorías. El investigador —al hacer el cuestionario— debe tener cuidado de que esas categorías sean, para el aspecto de que se trata, exhaustivas y mutuamente exclusivas. Deben ser exhaustivas en el sentido en el que su suma lógica debe cubrir todo el aspecto de que se trata, exhaustivas y mutuamente exclusivas. Deben ser que no debe haber entre ellas traslapamientos inadvertidos, pues si éstos existen, deben haber sido introducidos en forma consciente, voluntaria, con un fin determinado, y habiendo previsto las consecuencias de su introducción y las dificultades de manipulación que crean.

Sin embargo, en vista de las características de la investigación estadística, ya sea en la planificación del cuestionario o ya en la codificación de las respuestas al mismo, debe preferirse el exceso de análisis a su defecto. En general, es más fácil subsumir en una más amplia, categorías más estrechas que tratar de desentrañar qué categorías más estrechas pueden haber formado parte de la más amplia.

Como en otros casos, se trata de buscar un equilibrio adecuado entre el excesivo refinamiento y su falta total. El excesivo refinamiento analítico de las categorías y el establecimiento de subcategorías y subcategorías a la enésima potencia alarga el cuestionario y dificulta y hace más oneroso el mensuramiento social. La falta de refinamiento dificulta y hace que acabe por carecer de significado y de poder discriminador, el marco codificador, y convierte en un mazacote de difícil explotación, el conjunto de respuestas obtenidas.

El marco codificador es, por lo general, más fácil de establecer, cuando se manejan cuestionarios que cuando se manejan resultados de entrevista. En efecto, en el caso de los cuestionarios las respuestas están, en mayor o menor grado, precodificadas y los resultados sólo se "editan" o sujetan a una crítica que, en uno de sus aspectos, tiene que dar por resultado el refinamiento del marco codificador inicialmente empleado. En efecto, al lado de las categorías previstas suelen aparecer *otras* respuestas (y, en realidad *otros tipos* de respuesta). Esto impone —en muchos casos— que el marco inicial (el prefigurado en el cuestionario) se enriquezca o se depure gracias a la inclusión de esas categorías adicionales o gracias al uso que de las mismas se haga en calidad de agua fuerte, que pone a prueba la bondad del marco codificador prefigurado en el cuestionario. El establecimiento del marco codificador es más difícil cuando se manejan resultados de entrevista, contenidos documentales, etcétera, pues la preocupación principal del investigador debe consistir en formar un marco codificador adecuado que ni le resulte holgado ni le venga estrecho.

Para la constitución de ese marco codificador es necesario: agrupar todas las expresiones que se refieren a un solo aspecto o rubro analítico; clasificar esas respuestas, reuniendo, dentro del gran grupo —en subgrupos— las que, a pesar de su apariencia (de su estilo, de su fraseología, de su vocabulario, etcétera) son iguales entre sí, y separando cada subgrupo de los subgrupos constituidos por otras expresiones que, a pesar de su igualdad aparente, de forma, difieren de otras por su contenido mismo o por el que les da su ubicación dentro del contexto. Constituidas las categorías internas de un renglón analítico, conviene darles un rubro común: una expresión breve, fiel a todas ellas, para —en seguida— listar los rubros realmente obtenidos y complementarlos (en proceso deductivo) con los que potencialmente se habrían podido obtener. A la luz de esa nueva lista (marco codificador completo) hay que reexaminar las respuestas, y estudiar si no hay alguna o algunas de ellas que deban salir de un subgrupo y entrar a otro de los ya constituidos o de los que justamente acaben de agregarse.

Una vez formado el marco codificador —sea que se trate de codificación simple o de análisis de contenido— se procede a colocar en él las respuestas individuales. Para la manipulación más fácil o para la operación mecánica de estas respuestas, gracias a la universalidad y brevedad del signo numérico, la codificación de las respuestas de un renglón analítico de investigación culmina con el establecimiento de una relación biunívoca entre una categoría y un número que la representa.

De ahí parten todos los refinamientos procesales de la codificación, ya se trate del uso de simples tarjetas, de tarjetas perforadas a mano y manejadas mediante punzones —para su clasificación— o de tarjetas perforadas mecánicamente y manejadas mediante impulsos eléctricos. Lo que se vierte en esas tarjetas son expresiones literales que se han convertido en categorías numeradas y que, en último término, suelen convertirse en perforaciones que identifican cada categoría mediante la ubicación de la perforación misma en un espacio determinado (el de la tarjeta perforada).

Ordenación de los datos estadísticos

Para analizar una colección de datos o "agregado" estadístico, es útil ordenar esos datos. Un agregado sujeto a orden se conoce como "serie" o "distribución". La idea de "serie" es más amplia, más vaga, pues indica simple sucesión; mientras que la de "distribución" —más restringida y precisa— mienta ya, desde el principio, cierta idea de forma —manera en que se distribuyen los datos— la cual, o se incorpora en una fórmula matemática o en una gráfica.

Algunos textos hablan, en este sentido, de "escalas". Así, Freeman se refiere a "escalas nominales, ordinales e intervalares". Nos parece que el término escala es demasiado preciso (implica ascensos y descensos) y no es aplicable, en sentido estricto, a las nominales. Éstas son, en efecto, meros conjuntos de categorías, identificados por nombres y no por números y que, en este sentido, sólo en forma convencional o con criterio extraestadístico se pueden colocar como los escalones constitutivos de una escala. De ahí que prefiramos hablar, genéricamente, de "series" y que aceptemos, en cambio —por útil— la clasificación que Freeman hace de las que él llama escalas. Según esto, para nuestros propósitos, hay:

series nominales,
series ordinales, y
series cardinales (las escalas intervalares de Freeman).

O sea —en términos más simples— que hay series de nombres, series de números de orden, o series de números naturales.

Las series nominales son conjuntos ordenados de categoría. Las series ordinales arreglan sus datos según crezcan o decrezcan, pero atendiendo sólo al hecho de que determinada categoría del conjunto es "mayor que", "igual a" o "menor que" otra, sin que importe *qué tanto* mayor o menor sea. Las series cardinales, en cambio, consideran no sólo el hecho de que una categoría sea mayor, igual o menor que otra, para colocarla adelante, con o en seguida de ella, sino que estudia qué tantas veces mayor o menor que ella es.

Una serie puede *distribuirse,* además, de acuerdo con diferentes criterios, los principales son: 1) por ubicación de las categorías; 2) por fechamiento o ubicación temporal de las mismas, y 3) por referencia de las categorías a ejes, promedios, etcétera. Así, se habla de "distribuciones espaciales o geográficas", "distribuciones dinámicas o cronológicas" (también llamadas "diacrónicas") y "distribuciones estáticas o sincrónicas".

Series sincrónicas

Las series sincrónicas o estáticas pueden ser simples o de clases.

La serie simple se forma: 1) consignando todos los valores que la constituyen comenzando por el menor (o por el mayor); 2) tomando todos y cada uno de los valores que figuran en el agregado estadístico conforme a un orden creciente (o decreciente); 3) hasta concluir con el mayor (o el menor) de los datos del conjunto, y 4) *sin evitar repeticiones.*

La serie de clases está constituida por un conjunto de "casilleros" estadísticos; o sea, que está formada por ámbitos numéricos, delimitados por fronteras o límites también numéricos. Lo más frecuente es que éstos sean: un límite inferior y uno superior. Sin embargo, puede pensarse en casos de distribuciones bidimensionales (que atiendan simultáneamente a dos criterios) y en esos casos se puede hablar, en realidad, de clases que tienen cuatro límites: uno superior, uno inferior, uno lateral inferior, y otro lateral superior.

Distribuciones sincrónicas

Las distribuciones surgen, fundamentalmente, de las series, cuando a cada categoría de la serie se le asocia un dato numérico. El dato asociado es —por excelencia— el que responde a la pregunta "¿cuántos?". ¿Cuántos casos de esta categoría existen en el agregado? "Tantos". ¿Cuántos de esta otra? "Tantos otros". Ese dato asociado; ese "tantos", ese "tantos otros" se conoce como "frecuencia".

Según esto, se puede hablar de "distribuciones simples de frecuencias" o "distribuciones simples", y de "distribuciones de clases y frecuencias" o "distribuciones de clases".

Ordenación por formación de una serie simple

Para formar una serie simple se puede seguir la siguiente rutina: 1) cuéntese el número de casos que hay en el conjunto; 2) escríbanse, en columna, los números naturales, del *1* al que indique el número de casos del conjunto; 3) recórranse todos los datos del conjunto hasta encontrar aquel cuya magnitud sea mínima (o máxima si la serie ha de ser decreciente); 4) consígnese dicho valor frente al *1* de la primera columna y póngase una señal (√) frente al dato que se ha pasado del conjunto o agregado a la serie; 5) examínese de nuevo el conjunto de datos sin ordenar hasta descubrir aquel dato que —en magnitud— sigue inmediatamente en magnitud al mínimo (o al máximo), consígnesele en la serie y márquesele en el conjunto; 6) prosígase en la misma forma, sin evitar repetir en la serie los datos que aparezcan repetidos en el conjunto, hasta que todos los datos del conjunto sin ordenar aparezcan marcados con la señal y se haya llenado el espacio final de la segunda columna (correspondiente al último "número progresivo"). Estas dos comprobaciones (marcas en todos los datos del conjunto de datos sin ordenar, y cobertura del último espacio disponible frente a la columna de números progresivos) sirve para mostrar que ningún caso ha quedado fuera de la serie.

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

1º Número de datos: 20.

2º Columna de número progresivo:

Nº
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

3º Mínimo: 2.

4º Consignación en una segunda columna

						N	x
						1	2
4	4	5	4	3		2	
5	5	4	3	2		3	
6	6	6	6	2√		...	
7	7	7	7	7		20	

5º Señalar en los datos sin ordenar los que ya se hayan registrado en la columna x (estado de la tabulación hasta N = 10)

					N	x
4√	4√	5	4√	3√	1	2
5√	5√	4√	3√	2√	2	2
6	6	6	6	2√	3	3
7	7	7	7	7	4	3
					5	4
					6	4
					7	4
					8	4
					9	5
					10	5
					...	...
					20	—

Serie resultante:

N	x
1	2
2	2
3	3
4	3
5	4
6	4
7	4
8	4
9	5

Series cronológicas, diacrónicas o dinámicas

Serie diacrónica, dinámica o cronológica es 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 "cronológica", "diacrónica" o "dinámica").

En ocasiones, con fines de precisión, se distingue entre la serie empírica y la serie teórica (distinción que también puede hacerse respecto de las series sincrónicas o estáticas). La serie dinámica empírica ordena las magnitudes *observadas* de un fenómeno según la dimensión tiempo. La serie dinámica *teórica* ordena cronológicamente las magnitudes calculadas del fenómeno, obtenidas sobre la base de considerar que dichas variaciones de magnitud se encuentran ligadas al transcurso del tiempo.

Decir que entre los diversos valores del fenómeno estudiado mediante una serie dinámica, y las unidades de tiempo transcurridas hay una vinculación, equivale a afirmar, matemáticamente, que el fenómeno *está en función* (o "es función") del tiempo. O sea, que la magnitud del fenómeno dependerá del número de unidades de tiempo transcurridas a partir de determinado momento.

En las series cronológicas se manejan por lo menos dos variables (o, si se prefiere, se trabaja con dos tipos distintos de variable): una es el tiempo, y la otra (o las otras) son las variables no temporales. Aunque filosóficamente se puede justificar quien diga que el tiempo depende de los acontecimientos que en él ocurran, es mucho más fácil de justificar —y mucho más práctico— considerar que son los acontecimientos que varían de magnitud en el transcurso del tiempo los que dependen del tiempo mismo. De ahí que se considere al tiempo como la variable independiente, y a las otras como las dependientes.

El estudio estadístico de las series cronológicas: 1) primero es analítico; 2) después, es sintético, y 3) finalmente, trata de hacer pronósticos.

El análisis de las series cronológicas trata: 1.1) de determinar si el fenómeno crece en forma continua, y 1.2) en qué sentido y con qué ritmo crece o decrece; 1.3) si crece y decrece alternativamente, y 1.4) con qué periodicidad y amplitud crece y decrece alternativamente, así como si 1.5) está sujeto a variaciones de carácter aleatorio.

En forma no menos importante, el análisis trata de determinar en qué forma (matemática) se relacionan esos crecimientos, variaciones periódicas y oscilaciones aleatorias. En efecto, sólo si se conocen: 1) los resultados del análisis, y 2) las operaciones analíticas por las que se obtuvieron, es posible lograr la síntesis correspondiente.

La síntesis de una serie cronológica: 2.1) invierte los procedimientos de análisis para combinar entre sí los elementos resultantes del mismo, y 2.2) comprueba —mediante el cálculo de "errores probables" y "zonas de estimación"— el grado de adecuación entre los datos obtenidos mediante cálculo y los datos observados (que sirvieron de punto de partida al análisis y a la síntesis en que se funda dicho cálculo).

Una vez que se ha logrado un análisis aceptable, el pronóstico se puede hacer: 1) determinando el valor que corresponda al momento que interese, en relación con el punto de partida de la escala temporal que se haya utilizado, y 3.2) utilizando, para poner en marcha, el mecanismo sintético obtenido mediante las operaciones previas. En términos matemáticos esto representa: 3.1) dar un valor al momento que interesa, y 3.2) sustituirlo en la serie de ecuaciones o relaciones matemáticas que constituyen la síntesis correspondiente, a fin de obtener el correlativo valor teórico.

El pronóstico puede ser: en sentido estricto o en sentido amplio. En sentido estricto el pronóstico trata de anticipar cuál será el valor que alcanzará la serie en un momento dado del futuro, *si las fuerzas que han obrado para producirla siguen actuando sin perturbación alguna* (condición análoga a la económica de "siempre y cuando permanezcan constantes las restantes condiciones"). En sentido lato el pronóstico busca una de dos cosas: 1) o la determinación del valor de un fenómeno, correspondiente a un punto intermedio de la serie, para el que se carezca de observaciones (así, si los censos se realizan cada diez años, es probable que se tengan los valores de 1900, 1910, 1920, 1930, pero que no se tenga el valor observado de 1915, por ser un año intermedio), o que, a esa "prognosis" estadística se haga en sentido retrospectivo, para conocer con ciertos visos de verosimilitud el valor de un fenómeno para una fecha pasada, para la que no se cuente con registro estadístico. En este caso se realiza un "pronóstico hacia atrás", si se permite la paradoja.

Cuando se hacen esas determinaciones hacia atrás se habla de "extrapolación" de la serie; cuando se determinan valores intermedios se habla de "interpolación". Este último es, también, el término con el que se designa la representación de una serie dinámica empírica mediante una curva teórica.

Condensación por determinación de frecuencias

Para formar una distribución simple de frecuencias (o "distribución simple") se puede seguir esta rutina: 1) fórmese una *serie* sencilla, pero evitando repeticiones, y 2) asóciese a cada valor de la serie una frecuencia (o sea, el número de veces que se repite en el conjunto).

En forma más detallada, la rutina puede ser como sigue: 1) examínese el conjunto de datos sin ordenar, a fin de localizar el valor mínimo (o el máximo, en series decrecientes); 2) consígnese dicho valor en el primer lugar de la primera columna que constituirá la serie; 3) examínese el conjunto de datos para determinar la diferencia mínima entre la magnitud de un dato y la del inmediato siguiente (si son salarios, la diferencia entre los dos inmediatos siguientes puede ser de un peso, de cincuenta o de veinticinco centavos); 4) agréguese al valor mínimo (o réstese del máximo) una, dos, tres o más veces esa diferencia, anotando los resultados en esa columna hasta llegar al valor máximo (mínimo si la serie es decreciente); 5) examínese la serie original, siguiendo sistemáticamente sus datos del primero al último, marcando cada uno con una señal (√) y anotando frente al valor correspondiente de la serie: 5.1) una línea vertical (|), la primera vez que aparezca el dato; 5.2) una horizontal que forme ángulo con ella, la segunda vez (L); 5.3) otra vertical la tercera vez (⊔); 5.4) otra horizontal que complete el cuadrado (□), la cuarta; 5.5) una diagonal de cuadrado la quinta (⊠); 5.6), y repitiendo el procedimiento en caso de aparecer una sexta, séptima, octava . . . enésima vez (cuidando de separar los cuadrados que sirven para llevar la cuenta, para que la cercanía no haga que se confundan unidades contiguas) 6). Tras el examen sistemático del conjunto de datos sin ordenar, todos deberán aparecer marcados. El recuento de los cuadros cruzados (que valen 5), de los no cruzados (que valen 4) y de las barras sueltas (que valen 1 cada una) debe dar por resultado el total de datos del conjunto 7) las frecuencias que corresponden a los datos de la serie se obtendrán al sumar todos los cuadrados o barras que aparezcan delante de cada dato. Cuando frente a un dato no haya barras o cuadrados la frecuencia será cero.

Datos sin ordenar:

4	4	5	4	3
5	5	4	3	2
6	6	6	6	2
7	7	7	7	7

1º Mínimo: 2.

2º Consignado en primer término en la columna de las x_i:

x_i
2

3º Diferencia entre la magnitud de un dato y la del siguiente:

$$3 - 2 = 1.$$

4º Mínimo + 1, 2, 3, veces la diferencia mínima:

x_i
2
3
4
5
6
7 (valor máximo).

4√	4√	5√	4√	3√
5√	5√	4√	3√	2√
6√	6√	6√	6√	2√
7√	7√	7√	7√	7√

5º Recuento mediante cuadretes y señal de los ya contados frente a los datos.

2
3
4
5
6
7

6º Cuenta de cuadretes = 20
debe ser igual a total de datos. = 20

7º Recuento de los cuadretes correspondientes a cada valor de x_i (frecuencia — F).

x_i	f
2	2
3	2
4	4
5	3
6	4
7	5

Tabulación resultante,

Condensación por clasificación

Los datos de un agregado estadístico se pueden presentar en forma condensada si se les clasifica, o sea, si se forman, con ellos, "clases" estadísticas. Una clase estadística es un casillero numérico, que tiene límites también numéricos (uno inferior y uno superior, por lo menos). Esos casilleros o "clases" debidamente ordenados, forman una "serie de clases". Dentro de cada "clase" se colocan todos aquellos valores del agregado que caben en ella, y así se constituye una "distribución de clases".

La serie de clases

La formación de la pura serie de clases implica: 1) elegir un intervalo de clase, y 2) elegir los límites de clase (o, más precisamente, fijarlos).

Se conoce como "intervalo de clase" el número de unidades que separa a los límites superiores de los límites inferiores de las clases.

El límite superior de una clase indica el valor máximo que puede tener un dato para estar comprendido en ella; el límite inferior, el valor mínimo que ha de tener ese dato para su inclusión.

La elección del intervalo de clase debe atender tanto consideraciones de carácter sustantivo (que dependen del carácter social de los datos) como otras de carácter adjetivo (que se refieren a los aspectos puramente matemáticos de la clasificación). Estas consideraciones, a menudo no son convergentes, sino que compiten entre sí o se oponen; de ahí que deba ser producto de la prudencia y experiencia del investigador lograr un delicado equilibrio de todas ellas a fin de obtener el mejor intervalo asequible.

La consideración fundamental de carácter sustantivo que debe atenderse en la elección del intervalo de clase se refiere a que éste debe ser tal que, en lo posible, se ciña a las divisiones naturales de los datos. Así, si en términos generales se considera que una población puede dividirse en tres grandes grupos generacionales (de jóvenes, hasta los 30 años; de adultos, hasta los 60 y de ancianos hasta los 90) un intervalo de 30 años, o varios intervalos de los que 30 sea múltiplo (intervalos de 15, de 10 o de 5 años) pueden satisfacer este requerimiento.

Desde el punto de vista puramente estadístico, la elección del intervalo debe basarse en consideraciones como las siguientes (que se mencionan, más o menos, en orden de importancia): 1) debe relacionarse con la amplitud del agregado (ha de ser grande si la diferencia entre máximo y mínimo es grande, pequeño en caso contrario); 2) debe ser más pequeño conforme sea mayor la precisión que se busque (con lo cual aumenta el número de las clases); 3) debe ser constante para todas las clases (para facilitar los cálculos); 4) debe ser mayor conforme sea mayor la facilidad de operación que se busque, y 5) debe ser, en lo posible, un entero y, también dentro de lo posible, permitir que las "marcas de clase" (véase adelante) sean números enteros.

Fijación de los límites de clase

Elegido el intervalo de clase, el primer límite inferior (que determina la posición de todos los restantes límites inferiores y superiores) se fija a modo de contribuir, con el intervalo elegido, a que se satisfagan algunos de los requisitos mencionados anteriormente.

Una buena línea de conducta consiste en satisfacer los requisitos primordiales (del 1 al 4), mediante la elección de un buen intervalo y satisfacer el secundario (5) mediante una buena selección de límites. Sin embargo, si esa fijación de límites compite con el requisito (1), por ejemplo, se debe sacrificar la satisfacción del requisito (5). Lo mismo debe hacerse si compite, particularmente, con el requisito sustativo de que las clases estadísticas coincidan, en sus límites, con las delimitaciones sociológico-naturales de los grupos que pretenden representar.

Presentación de la serie de clases

Una vez que se haya fijado el primer límite inferior de la serie de clases, los límites inferiores siguientes se obtendrán a partir del primero, sumándole a éste el valor del intervalo, 1, 2, 3, . . .n veces hasta obtener un límite inferior igual al valor máximo del agregado (este último límite ya no será necesario). Una vez hecho este cálculo la serie se presenta en dos columnas: los límites inferiores se consignan en una columna; frente a los límites inferiores, en otra columna, se consignan los límites superiores.

Formas alternativas de presentación

Las clases se pueden identificar: 1) usando límites, o 2) usando valores representativos o "marcas de clase".

Cuando se usan los límites inferiores y superiores para identificar las clases: 1.1) se pueden hacer que coincida el límite superior de una clase con el límite inferior de la que le sigue, o 1.2) puede hacerse que difieran.

Si el límite inferior de una clase (la sexta clase, por ejemplo), coincide con el límite superior de la clase anterior (la quinta), el conjunto de las clases cubre toda la distribución sin solución de continuidad; pero, en cambio, no hay modo de adscribir unívocamente a una clase un dato que se encuentre justamente en el límite (uno cuyo valor sea igual al límite superior de la quinta y también igual al límite inferior de la sexta). El inconveniente se puede subsanar si se adopta, adicionalmente, una regla que diga que en los casos en que un dato se encuentra en el límite entre dos clases, debe de colocársele, siempre, en la clase inferior (o siempre en la superior).

Cuando se establece una diferencia entre el límite superior de una clase y el inferior de la siguiente, la diferencia puede hacerse: 1.21) por exclusión del límite superior, o 1.22) por exclusión del límite inferior. Estas exclusiones se hacen en dos formas alternativas: *a*) o se escribe un guión del límite excluido, o *b*) se señala "de *más* de tal límite inferior a tal límite superior" o "de tal límite inferior a *menos* de tal límite superior".

Las menciones anteriores son largas y relativamente imprecisas. Para mayor precisión, las exclusiones también pueden hacerse si se anotan no los límites en sentido estricto, sino el valor del límite inferior excluido más un décimo del intervalo, o el límite superior que se excluye menos un décimo de dicho intervalo. Este último procedimiento facilita la adscripción de los valores del agregado, pero afecta los cálculos (aunque, a veces, en mínima proporción).

La mejor solución, en este sentido, consiste en utilizar unos límites que no se presten a equívoco, en la adscripción de valores, y usar unos límites cuyo valor resulte afectado al mínimo, en el cálculo.

EJEMPLO DE LAS FORMAS ALTERNATIVAS DE PRESENTACIÓN DE LAS CLASES DE UNA SERIE

Diferentes maneras de presentar las primeras clases de la serie de intervalos quinquenales de edad, de acuerdo con la pregunta del censador mexicano en 1960.

El censador preguntó la edad "en años cumplidos" para los mayores de un año; "en meses cumplidos", para los menores de un año. Aquí se considera a los niños menores de un año subsumidos en la clase de quienes tenían "de cero a cinco años" (en realidad, a quienes tenían "desde unos minutos de haber nacido hasta quienes tenían unos minutos menos de los cinco años de su edad".

Primera presentación:

De 0 a más de 4 sin llegar a 5
De 5 a más de 9 sin llegar a 10
De 10 a más de 14 sin llegar a 15
. . . .
De 84 a más de 89 sin llegar a 90

Segunda presentación:

De 0 a 4 años 11 meses 364 días 23 horas
De 5 a 9 años 11 meses 364 días 23 horas
. . . .
De 84 a 89 años 11 meses 364 días 23 horas

Tercera presentación:

De 0 a 4.99 años
De 5 a 9.99 años
. . . .
De 85 a 89.99 años

Cuarta presentación:

0 — 4 +
5 — 9 +
....
85 — 89 +

Quinta presentación:

$0 \leqslant$ edad < 5 "edad igual o mayor que 0 pero menor que 5" (en realidad, en esta clase $0 <$ edad < 5, pues la edad $= 0$ es la del no nacido aún)
$5 \leqslant$ edad < 10
....
$85 \leqslant$ edad < 90

Si las respuestas hubieran sido del tipo: "de tantos años entrados a tantos más uno", la presentación posible sería:

de (0 cumplidos entrados a 1) a (4 cumplidos entrados a 5).

Esa presentación, en forma sintética, corresponde a la tabulación siguiente:

0 a 5 —
5 a 10 —
....
85 a 90 —

Series continuas y discontinuas

A fin de que el estudiante no se desoriente con la gran variedad de formas de presentación de las clases, conviene considerar que los límites inferiores son fijos, y que en lo único en que hay que pensar es en las alternativas de presentación de los límites superiores.

En este sentido, conviene —también— establecer una diferencia entre las series continuas y las discontinuas. Las series continuas son aquellas en que el tránsito de un valor a otro se hace por pequeños incrementos (como ocurre con años, meses, semanas, días, horas, minutos, segundos). Las series discontinuas, en cambio, son aquellas en las que los valores cambian por "saltos" (como ocurre con el número de miembros de una familia, que pueden ser 2, 3, 4, pero no 2.5 o 3.33).

Cuando las series son discontinuas puede convenir, como práctica general, hacer que el límite superior de cada clase sea igual al límite inferior de las clases siguientes, *menos una unidad.*

Cuando las series son continuas, puede convenir que los límites superiores de las clases sean iguales a los límites inferiores de las siguientes, *pero con la mención de que no están incluidos en la clase.* Esta mención genérica permite que cualquier dato del agregado se incluya en la clase misma si difiere, *aunque sea en una cantidad pequeñísima* del límite superior, y que se excluya de dicha clase sólo si es igual a dicho límite.

Presentación mediante las marcas de clase

Por "marcas de clase" se conocen ciertos valores representativos de las clases estadísticas. Generalmente se trata de los puntos medios. Sin embargo, si en una distribución la mayoría de los valores se concentra cerca de los límites inferiores (o cerca de los superiores), los límites mismos pueden funcionar como marcas de clase.

En general, se deben buscar como marcas de clase los valores más representativos de la distribución, aunque la rutina haya impuesto el uso de los puntos medios.

Cálculo de los puntos medios

Los puntos medios pueden calcularse en dos formas (que no siempre dan resultados coincidentes): 1) o se suman los límites y se divide su suma entre dos, tras de lo cual se agrega ½, o 2) se representan todos los valores posibles a lo largo de una línea y se elige el que ocupa la posición central.

Distribución de clases

Para formar la distribución de clases, se procede en forma parecida a como se procedió con la distribución simple. De lo que se trata ahora es, fundamentalmente, de determinar las frecuencias que deben corresponder a las diferentes clases, así como antes se determinaron las frecuencias correspondientes a los diferentes datos.

La rutina es como sigue: 1) hay que examinar sistemáticamente el agregado, del primero al último dato, para ver a qué clase corresponde cada uno de ellos; 2) marcar con (√) cada uno de los datos del agregado que se haya adscrito a una clase determinada; 3) marcar frente a la clase correspondiente una línea vertical si es el primer dato que le corresponde; una horizontal que forme ángulo, si es el segundo; otra vertical, si es el tercero; otra horizontal que forme cuadrado, si es el cuarto, y una diagonal, si es el quinto ($\boxtimes$) tal como se procedió en la determinación de las frecuencias de la distribución simple; 4) comprobar que todos los datos del agregado han sido marcados y que el total es igual al total de cuadrados y barras que aparezcan en la tabulación de trabajo. Una vez hecho esto; 5) se determinarán las frecuencias que correspondan a cada clase, contando los cuadrados cruzados y no cruzados y las barras que figuren delante de la misma, y 6) se copiará en limpio la tabulación.

La presentación final de la distribución de clases constará de: 1) dos columnas interrelacionadas, en donde figurarán, unos al lado de los otros, los límites inferiores y superiores de las clases, y 2) una columna que contendrá las frecuencias.

EJEMPLO DE FORMACIÓN DE UNA SERIE DE FRECUENCIAS ACUMULATIVAS

Localidades de México, en 1960, según el número de sus habitantes

CLASES *Localidades de tantos a tantos habitantes*	FRECUENCIAS *Número de localidades de esa clase*	FRECUENCIAS ACUMULADAS *Progresiva*	*Regresiva*
Sin habitantes o censadas con otras	56 100	56 100	145 712
de 1 a 99	51 555	107 655	89 612
de 100 a 499	27 098	134 753	38 057
de 500 a 999	6 156	140 909	10 959
de 1 000 a 2 499	3 342	144 251	4 803
de 2 500 a 4 999	865	145 116	1 461
de 5 000 a 9 999	340	145 456	596
de 10 000 a 19 999	146	145 602	256
de 20 000 a 29 999	38	145 640	110
de 30 000 a 39 999	20	145 660	72
de 40 000 a 49 999	11	145 671	52
de 50 000 a 74 999	15	145 686	41
de 75 000 a 99 999	9	145 695	26
de 100 000 a 249 999	12	145 707	17
de 250 000 a 499 999	2	145 709	5
de 500 000 a 3 000 000	3	145 712	3
	145 712		

COMPROBACIÓN

LECTURAS DE LA SERIE PROGRESIVA Y DE LA REGRESIVA DE FRECUENCIAS ACUMULADAS

hay 56 100 localidades sin habitantes o censadas con otras.
107 655 localidades con menos de 100 habitantes.
134 753 localidades con menos de 500 habitantes.
140 009 localidades con menos de 1 000 habitantes.
. . . .
hay 145 712 localidades con menos de 3 millones de habitantes.

hay 145 712 localidades con más de 0 habitantes.
89 612 localidades con más de 1 o con 1 habitante.
39 057 localidades con más de 100 o con 100 habitantes.
. . . .
hay 3 localidades con más de 500 000 o con 500 000 habitantes.

Consignación tabular

Un cuadro estadístico está constituido por los siguientes elementos principales: 1) un número; 2) un título; 3) cabezas de columna; 4) rubros de renglón; 5) columnas; 6) renglones; 7) líneas intercolumnares; 8) espacios interrenglonares; 9) fuentes, y 10) notas.

El número marca el orden que le corresponde al cuadro en el conjunto de cuadros de una investigación, y está destinado a permitir una referencia rápida al mismo. Puede tratarse de una numeración simple que abarque todos los cuadros de la investigación (especialmente si ésta es breve) o puede ser que la numeración corresponda a la de los capítulos. Así, por ejemplo, en *An Advanced Theory of Statistics*, de Kendall y Stuart, todos los cuadros del capítulo quinto están marcados por el entero 5 seguido de un punto decimal; detrás del punto decimal la numeración marca el orden de los cuadros dentro del conjunto de los que aparecen en dicho capítulo; así 5.3 indica que se trata del tercer cuadro del capítulo quinto. La sistematización puede llevarse más allá cuando, dentro de cada capítulo hay subdivisiones marcadas por una numeración de ese tipo. Kendall y Stuart emplean esa subdivisión de sus capítulos (así, hay un párrafo 5.7 que indica que es el séptimo del capítulo quinto), pero no hacen que se refleje dicha numeración en la de los cuadros (5.74 sería el cuadro cuarto del párrafo séptimo, dentro del capítulo quinto).

El título debe ser tan específico como se pueda (como debe ocurrir en el caso de cualquier título que se refiera a un asunto científico). Las principales especificaciones deben ser con respecto a cuál es el fenómeno de que se trata, cuál es su circunscripción espacial (respuesta a la pregunta ¿dónde?) y cuál la temporal (periodo que abarca, o respuesta a la pregunta ¿cuándo?). Cuando el cuadro es muy sencillo, en el título mismo pueden figurar las unidades (millones de habitantes, pesos mexicanos, etcétera); sin embargo, este caso es raro, y conviene consignar las unidades en las cabezas de columna o los rubros de renglón.

Las cabezas de columna son aquellas casillas que están destinadas a contener las clases en que se haya dividido el fenómeno observado (inmigrantes, emigrantes; población ocupada en las actividades primarias, en las secundarias, en las terciarias, etcétera) y las unidades correspondientes. Estas cabezas de columna son las que figuran en primer lugar, de arriba a abajo, en el cuadro. La primera de ellas, sirve para indicar el criterio de la clasificación de los rubros del cuadro (país de procedencia o destino de inmigrantes y emigrantes, divisiones político-administrativas o "entidades" en las que se practican determinadas actividades, etcétera).

En el caso de los cuadros de trabajo (distintos en algunos aspectos de los cuadros expositivos del informe final) las cabezas de columna generalmente se llenan con ciertos símbolos (d, para desviación, d^2 para cuadrado de las desviaciones, etcétera), pues en dichos cuadros de trabajo las columnas que encabezan esas casillas indican, generalmente, ciertas etapas del cálculo (en la columna encabezada por d se consignan las desviaciones, en la d^2 se registran los valores que se obtienen de elevar al cuadrado los valores de la columna anterior, etcétera).

Los rubros de renglón cumplen para las hileras o renglones de casillas una función análoga a la que desempeñan los encabezados para las columnas. Generalmente estos rubros corresponden a divisiones por países, divisiones administrativas (Estados y territorios) dentro de un mismo país, clases de actividad, u otras análogas.

En este aspecto, puede observarse que, en la práctica —incluso en la de quienes elaboran las publicaciones de los órganos centrales de estadística de algunos países— se desatienden ciertas precauciones y con ello, se dificulta, más tarde, la tarea del investigador. Así, es indispensable que en esas publicaciones (así como en cualquier informe estadístico-social que se les parezca) se establezcan, desde el principio, y en forma expresa, una serie de categorías de primer orden, que puedan ser comunes a todas las secciones de un anuario o de cualquier otra publicación estadística parecida (años, entidades, actividades) y que se presenten siempre en la misma forma (siempre como rubros o siempre como cabezas y no unas veces como rubros y otras como cabezas).

Por otra parte cuando se trata de actividades o entidades, por ejemplo, se puede optar por uno de dos criterios: o el ordenatorio o el sistemático. En el *Anuario Estadístico de los Estados Unidos Mexicanos* para 1960, por ejemplo, se emplean ambos: se trata de clasificar las actividades (con un criterio económico) y también se intenta a veces clasificar las entidades (por zonas geoeconómicas de características análogas) mientras que, en otras ocasiones, se listen esas entidades en forma puramente alfabética. Y, aunque es claro que la sistematización es actividad científicamente superior a la pura ordenación, como los criterios de sistematización pueden ser muy diversos (según el investigador, el tema de la investigación, etcétera) es preferible que, en fuentes primarias (y en informes sobre investigaciones que se asimilen a ellas) los rubros *se ordenen* simplemente; que las entidades se consignen por orden alfabético, por ejemplo.

Además, en tales casos, deben evitarse las supresiones de rubros (cuando a un rubro corresponda un cero, debe figurar el rubro y el cero, y cuando se desconozca o no se haya recogido el dato para dicho rubro, deberá seguir figurando el rubro con la mención "se ignora" frente al mismo).

Lo anterior, que es vigente para la tabulación de datos que corresponden a fuentes primarias o para los cuadros iniciales de una investigación, no tiene que seguirse al pie de la letra en el caso de aquellos otros cuadros de la misma que presenten algún principio de elaboración, pues buena parte de esa elaboración está constituida por la agrupación o clasificación de los rubros con criterio no simplemente ordenatorio, sino sistemático, ya que se elaboran de acuerdo con el plan de investigación.

En un cuadro, las columnas están constituidas por series de casillas ordenadas verticalmente. Todas las casillas de una misma columna corresponden a una de las clases constitutivas del cuadro y cubren —en cambio— toda la variedad de divisiones territoriales, de actividad o de cualquier otro tipo, de los diferentes rubros.

Los renglones están constituidos por conjuntos de casillas dispuestas horizontalmente. Todas las casillas de un mismo renglón se refieren a una misma división territorial, ocupacional, etcétera y cubren toda la variedad de manifestaciones del fenómeno para dicha división territorial, ocupacional o de otro tipo.

Cuando hay cierto número de columnas que pueden formar un grupo, distinto de otro grupo de columnas, en el mismo cuadro, se las puede unir, poniendo a cada grupo un nuevo encabezado, común a todas ellas. Lo mismo puede ocurrir en el caso de los renglones. Así, si bien el editor de un Anuario Estadístico debe pensar en ordenar las entidades alfabéticamente, un investigador puede comenzar reordenando los datos del anuario mediante la formación de grupos como los siguientes: "Entidades del Pacífico Norte", "Entidades del Norte", "Entidades del Golfo", "Entidades del Centro"; o bien "Entidades Socio-económicamente Excedentes", "Entidades Normales Socio-económicamente", "Entidades Socio-económicamente Deficientes", etcétera.

Las líneas intercolumnares sirven para separar los datos de una columna de los contenidos en otra columna. Generalmente son sencillas; pero, cuando las columnas se han reagrupado, formando clases más amplias, entre cada grupo de columnas y el siguiente se hace figurar una línea vertical doble (en realidad dos paralelas verticales) para conseguir la separación. También conviene señalar que, en el momento de la impresión, la práctica aceptada consiste: 1) en eliminar siempre estas líneas en los cuadros que tienen sólo dos columnas (en realidad sólo rubros y una columna); en usarlas generalmente en los cuadros que tienen más de dos columnas, y en eliminarlas —en caso de que lo imponga la economía— aun en los cuadros que tienen varias columnas, reemplazándolas por espacios blancos suficientemente grandes para evitar confusiones. Puede decirse, en suma, que es deseable usar esas líneas, pero que es permisible eliminarlas.

Los espacios entre los renglones cumplen función parecida a las líneas intercolumnares. Sin embargo, a diferencia de lo que ocurre con las columnas, no se marca ordinariamente mediante una línea la separación entre dos renglones contiguos y sí se deja un espacio en blanco cuando se trata de separar a un grupo de rubros (y los renglones que les corresponden) de otro grupo de rubros y renglones.

Cuando no se usan líneas para separar los renglones, y hay necesidad de consignar totales al pie de las columnas, una línea horizontal separa a dichos totales del cuerpo de la tabla o cuadro. Cuando se usan líneas para separar grupos de renglones, la separación entre el cuerpo del cuadro y la línea de totales se hace mediante una línea doble (o, mejor, mediante un par de paralelas horizontales).

ELEMENTOS PRINCIPALES DE UN CUADRO ESTADÍSTICO TÍPICO

TÍTULO: Fenómeno ———— Lugar ———— Fecha o Periodo ————

CATEGORÍAS		COLUMNA 1 Encabezado Unidades		COLUMNA 2 Encabezado Unidades
		Subcolumna 1.1	Subcolumna 1.2	Subcolumna única 2.1
RENGLÓN 1 Rubro Unidades de este rubro	Subrenglón 1.1 Subrenglón 1.2 Subrenglón 1.3	CASILLAS 1.1		CASILLAS 1.2
RENGLÓN 2 Rubro Unidades de este rubro	Subrenglón 2.1 Subrenglón 2.2	CASILLAS 2.1		CASILLAS 2.2
RENGLÓN 3 Rubro Unidades de este rubro	Subrenglón único 3.1	CASILLAS 3.1		CASILLAS 3.2

TOTALES

FUENTE: Publicación ———————— Cuadro ————————.

Nota: Aquí se han representado las divisiones principales con línea continua y las secundarias con línea distintinua; en caso de necesitarse, se pueden agrupar columnas o renglones separando cada grupo de los otros con líneas dobles.

La fuente de obtención de los datos es de gran importancia, y generalmente se hace figurar al pie del cuadro. Ocasionalmente se coloca, entre paréntesis, inmediatamente después del título.

Las notas sirven para hacer aclaraciones con respecto a los datos (señalar que un dato corresponde no a un recuento efectivo, sino a una cifra calculada, a una simple estimación, o a una apreciación hecha por determinada institución o entidad; que se ha empleado un criterio de obtención distinto del que rige el resto del cuadro, etcétera). Como los datos de estos cuadros son numéricos, las "llamadas" a la nota no se suelen hacer mediante números como cuando se indican las llamadas en un trozo de texto, sino mediante signos como el asterisco y otros semejantes.

Representación gráfica en estadística

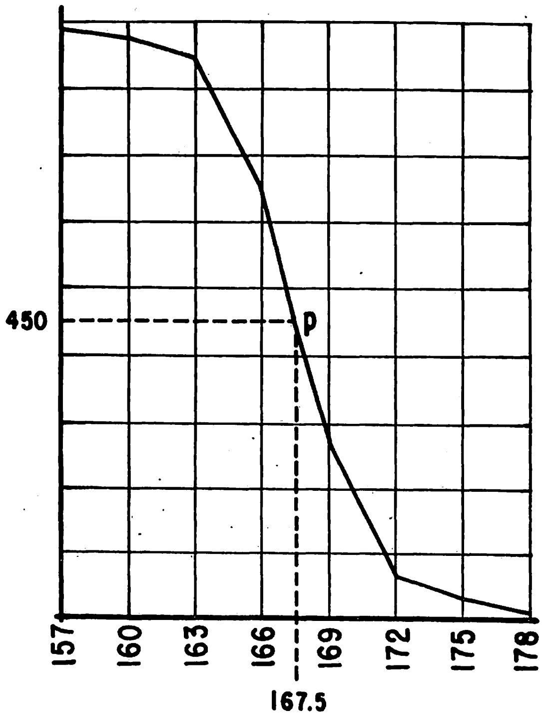
La representación gráfica en estadística tiene gran importancia. Su interés es doble en cuanto: 1) sirve como instrumento de trabajo y 2) sirve como medio de ilustrar ciertos resultados obtenidos mediante la elaboración estadística.

Como instrumento de trabajo, la representación gráfica de los fenómenos estadísticos sirve para que el estadígrafo logre una primera aproximación, de conjunto, al fenómeno que debe de estudiar en cuanto es más fácil darse cuenta de la magnitud del mismo y sus variaciones si se le representa por medio de una línea o un conjunto de líneas que si se manejan largas series de datos.

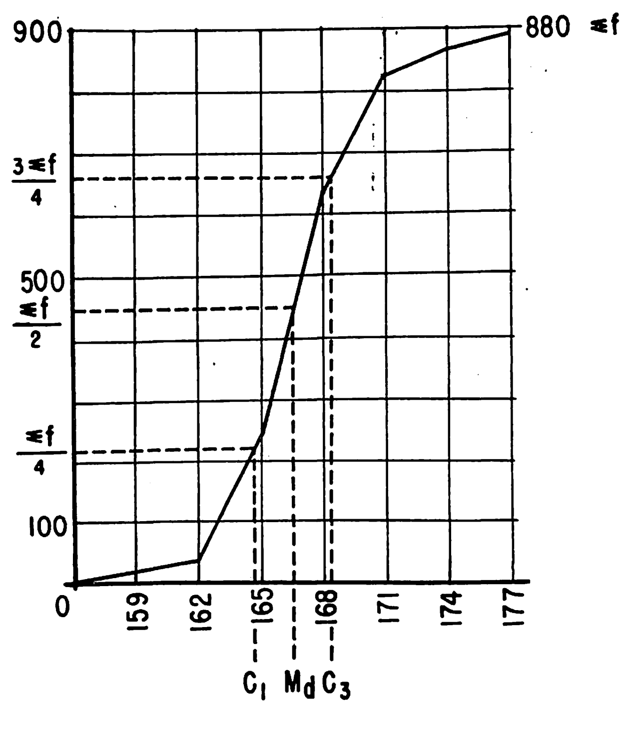
También sirve como instrumento de trabajo porque, en muchas ocasiones es posible obtener mediante un procedimiento gráfico más o menos rápido, resultados suficientemente aceptables que en otras condiciones tendrían que obtenerse mediante un largo proceso numérico.

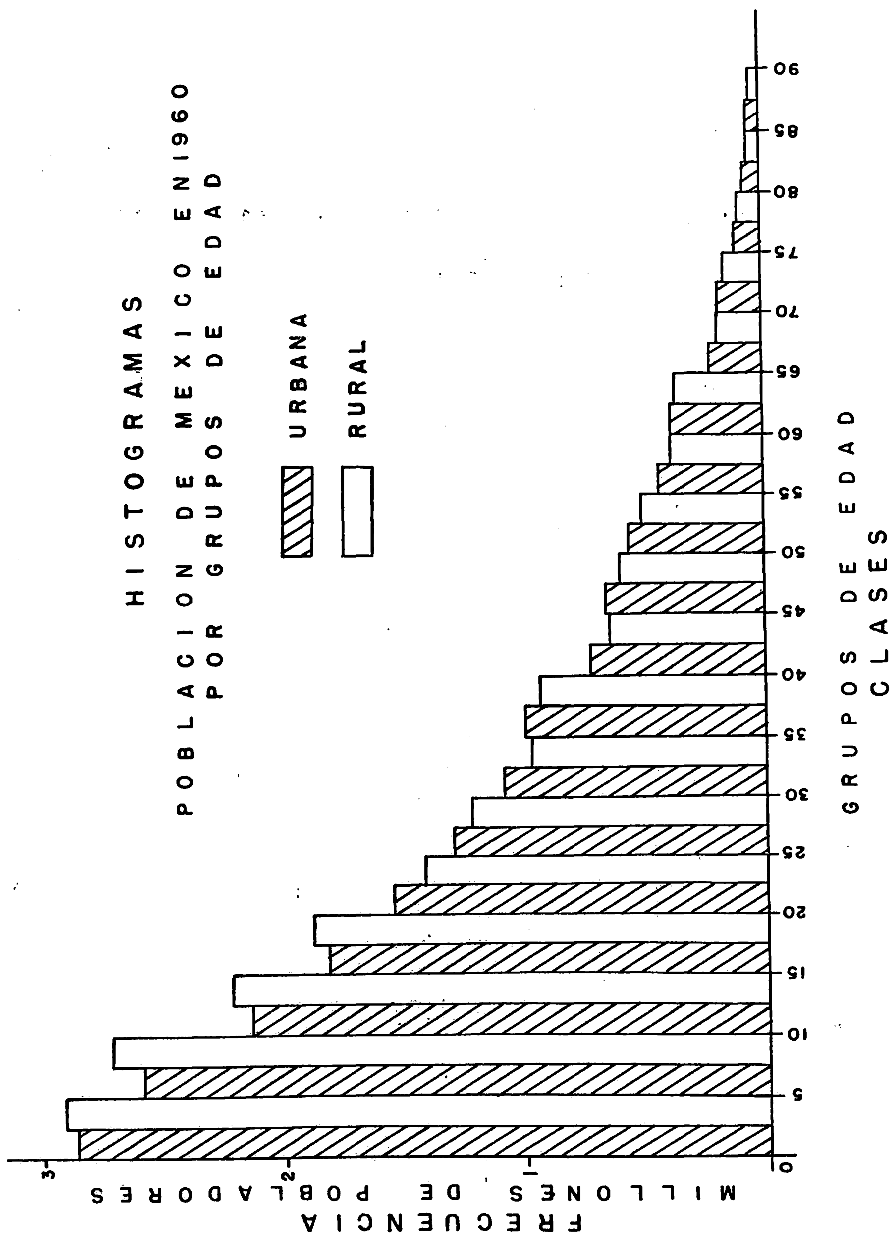
Como medio ilustrativo, las gráficas tienen un importante papel en estadística en cuanto permiten que el consumidor apresurado capte los caracteres más sobresalientes del fenómeno tal y como éstos han sido puestos de manifiesto por la elaboración estadigráfica.

OJIVA-DETERMINACION DE LA MEDIANA



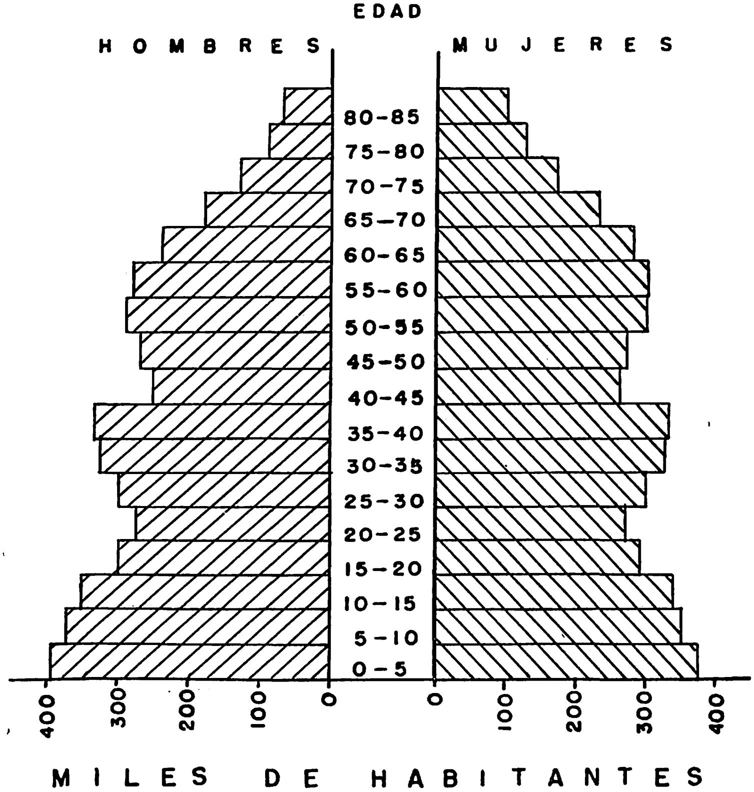
OJIVA-DETERMINACION DE LA MEDIANA Y LAS CUARTILAS



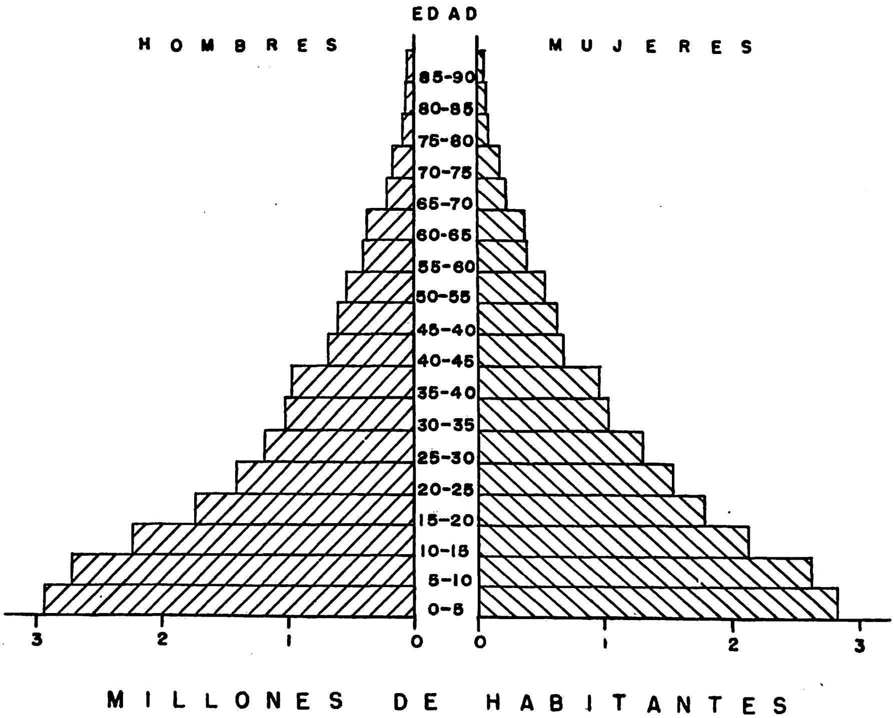

HISTOGRAMAS
POBLACION DE MEXICO EN 1960
POR GRUPOS DE EDAD
URBANA
RURAL
GRUPOS DE EDAD
CLASES
FRECUENCIA
MILLONES DE POBLADORES
0
1
2
3
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90

EJEMPLO DE DOBLE HISTOGRAMA

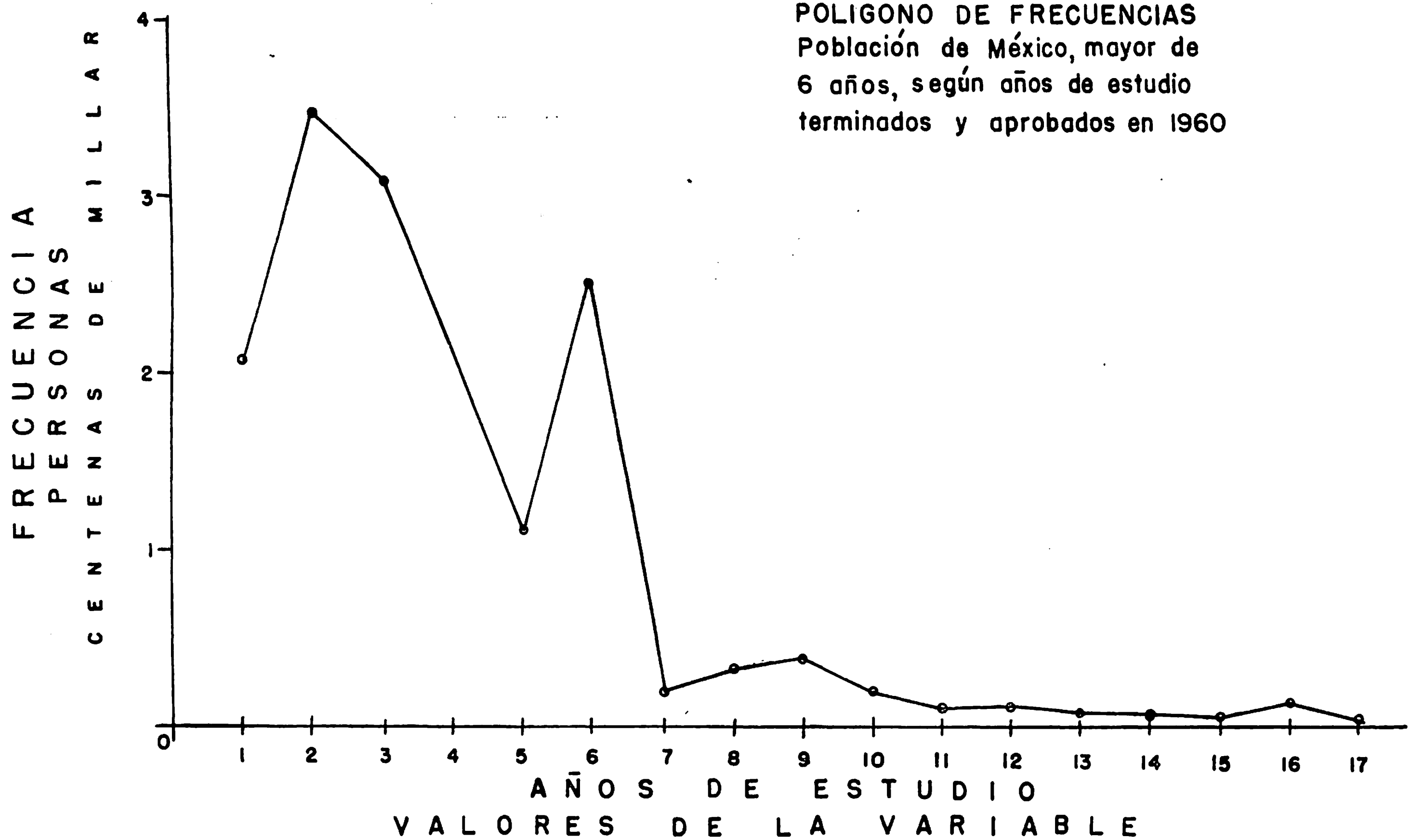
Pirámide de edades de la población Belga

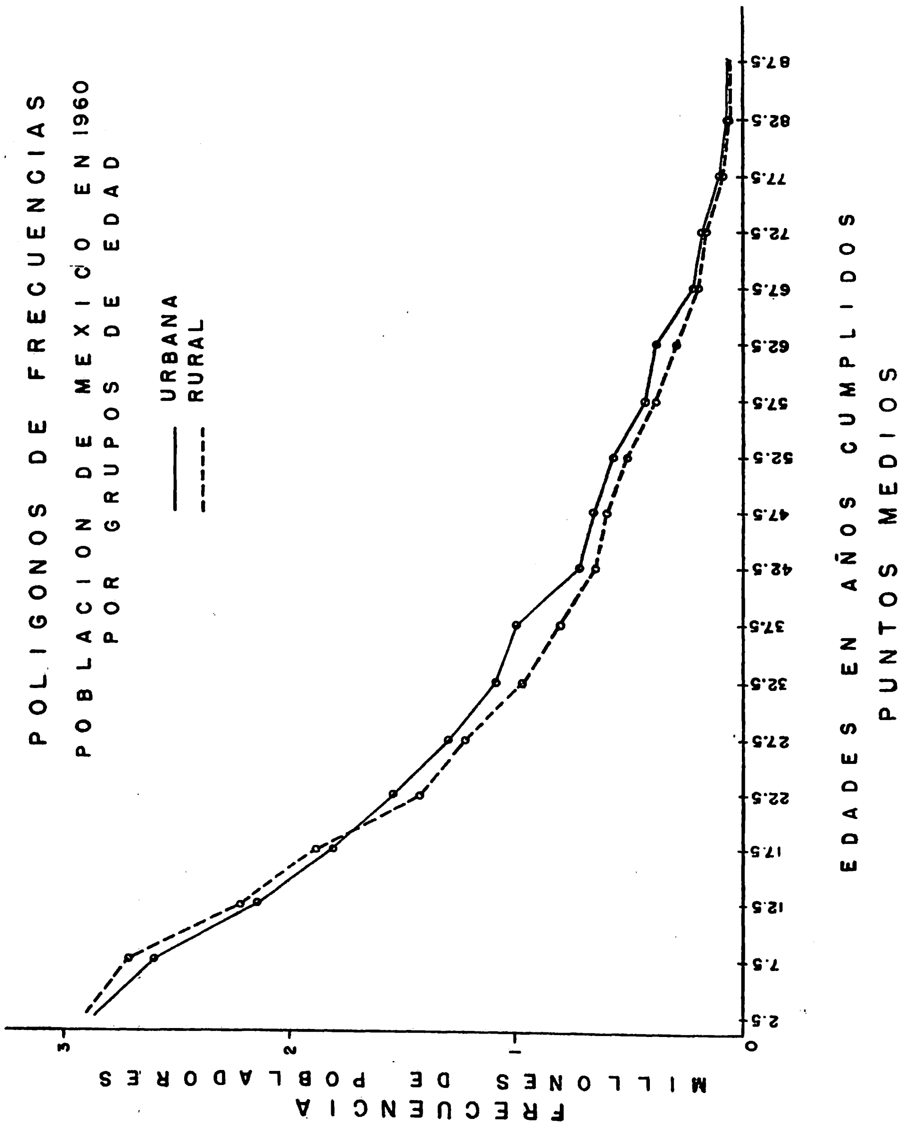


EJEMPLO DE DOBLE HISTOGRAMA
Pirámide de edades de la población mexicana



POLIGONO DE FRECUENCIAS
Población de México, mayor de 6 años, según años de estudio terminados y aprobados en 1960




POLIGONOS DE FRECUENCIAS
POBLACION DE MEXICO EN 1960
POR GRUPOS DE EDAD
URBANA
RURAL
EDADES EN AÑOS CUMPLIDOS
PUNTOS MEDIOS
2.5
7.5
12.5
17.5
22.5
27.5
32.5
37.5
42.5
47.5
52.5
57.5
62.5
67.5
72.5
77.5
82.5
87.5
0
1
2
3
FRECUENCIA
MILLONES DE POBLADORES

EJEMPLO DE REPRESENTACIÓN Y DE COMPARACIÓN GRAFICA DE DOS DISTRIBUCIONES ESTADÍSTICAS, MEDIANTE HISTOGRAMAS Y MEDIANTE POLÍGONOS DE FRECUENCIAS

DATOS

Distribución de la población urbana y rural de México, en 1960, por grupos de edad

CATEGORÍAS *Edades en años cumplidos*	PUNTOS MEDIOS *Edad media de cada clase*	FRECUENCIAS *Pobladores urbanos de esas edades*	 *Pobladores rurales de esas edades*
De 0 a 4	2.5	2 866 675	2 910 072
5 a 9	7.5	2 593 200	2 723 844
10 a 14	12.5	2 143 959	2 214 358
15 a 19	17.5	1 781 359	1 753 906
20 a 24	22.5	1 530 341	1 416 731
25 a 29	27.5	1 284 551	1 220 341
30 a 34	32.5	1 082 441	969 194
35 a 39	37.5	999 922	920 758
40 a 44	42.5	718 669	642 655
45 a 49	47.5	651 632	581 976
50 a 54	52.5	562 009	501 350
55 a 59	57.5	423 223	376 676
60 a 64	62.5	377 415	367 295
65 a 69	67.5	218 206	195 958
70 a 74	72.5	173 195	160 176
75 a 79	77.5	100 913	86 860
80 a 84	82.5	70 662	57 676
85 a 89	87.5	67 283	64 106

N.B. Por tratarse de "años cumplidos", los intervalos representan, en realidad: de 0 a 4 años o más, sin llegar a 5; o sea, de 0 a 5 años, y el punto medio corresponde a 2.5.

Las gráficas muestran dos histogramas y dos polígonos de frecuencias que correspon den a cada una de las dos distribuciones de la población (urbana la una, rural la otra).

EJEMPLO DE REPRESENTACIÓN DE SERIES ESTADÍSTICAS MEDIANTE MAPAS

DATOS

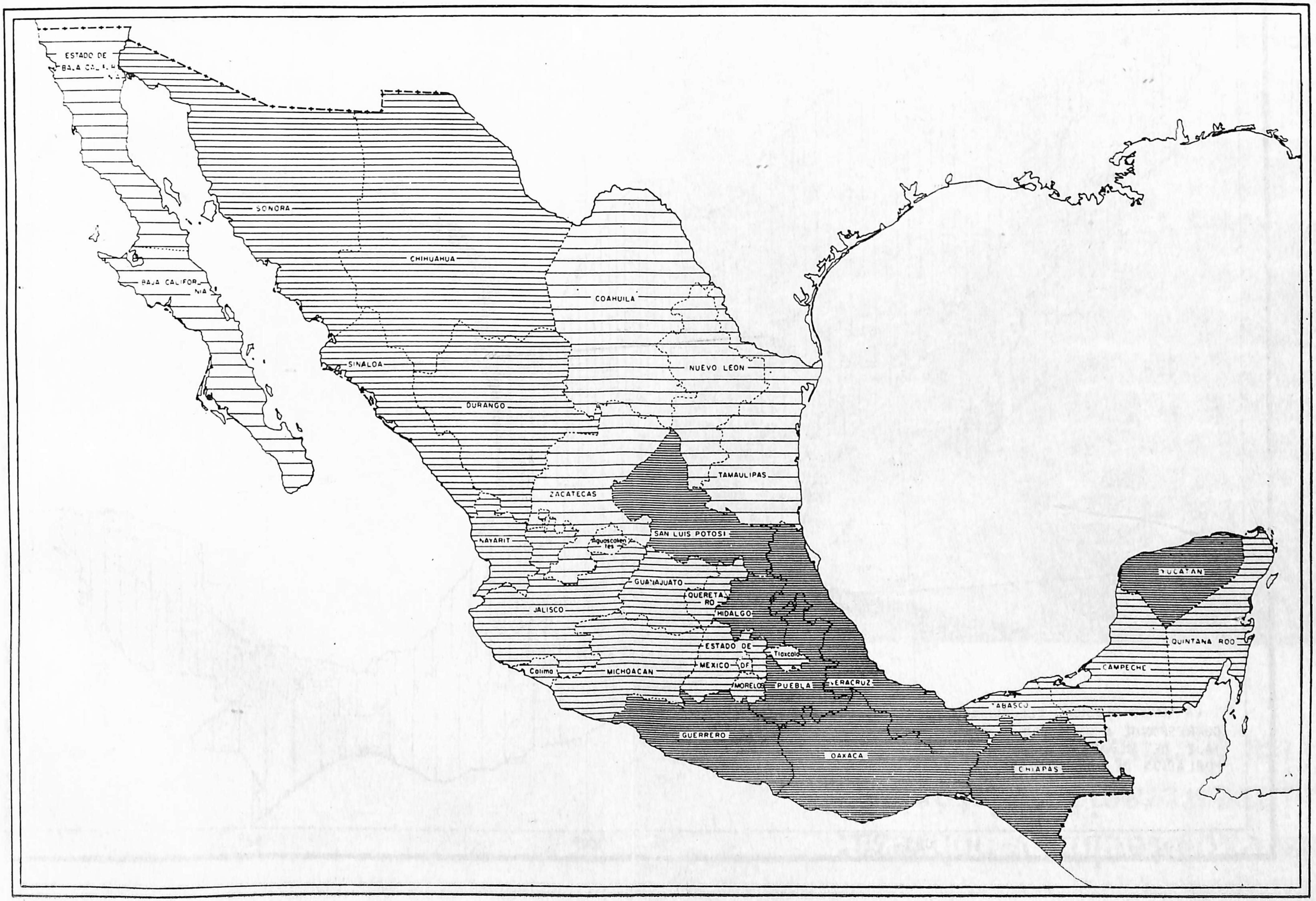
Distribución porcentual de la población hispanoparlante de México y distribución porcentual de la población alfabetizada del país, en 1960

CATEGORÍAS *Entidades federativas*	PORCIENTOS *de hispanoparlantes en la población total*	*de alfabetizados en la población total*
AGUASCALIENTES	99.21	72.93
BAJA CALIFORNIA N.	95.25	81.12
BAJA CALIFORNIA S.	99.18	79.51
CAMPECHE	74.04	68.10
COAHUILA	98.69	80.40
COLIMA	98.97	68.61
CHIAPAS	61.90	39.30
CHIHUAHUA	93.76	74.92
DISTRITO FEDERAL	97.94	83.42
DURANGO	98.70	75.16
GUANAJUATO	99.45	51.10
GUERRERO	79.82	37.19
HIDALGO	71.81	44.05
JALISCO	99.25	65.17
MÉXICO	89.03	57.40
MICHOACÁN	96.11	50.95
MORELOS	96.82	60.84
NAYARIT	96.91	65.94
NUEVO LEÓN	98.77	80.70
OAXACA	53.12	40.88
PUEBLA	82.16	50.23
QUERÉTARO	95.79	42.90
QUINTANA ROO	37.94	64.48
SAN LUIS POTOSÍ	86.25	53.34
SINALOA	99.00	66.03
SONORA	95.60	76.18
TABASCO	94.14	61.68
TAMAULIPAS	97.88	77.34
TLAXCALA	93.63	61.51
VERACRUZ	86.23	54.75
YUCATÁN	43.56	65.73
ZACATECAS	99.67	63.37

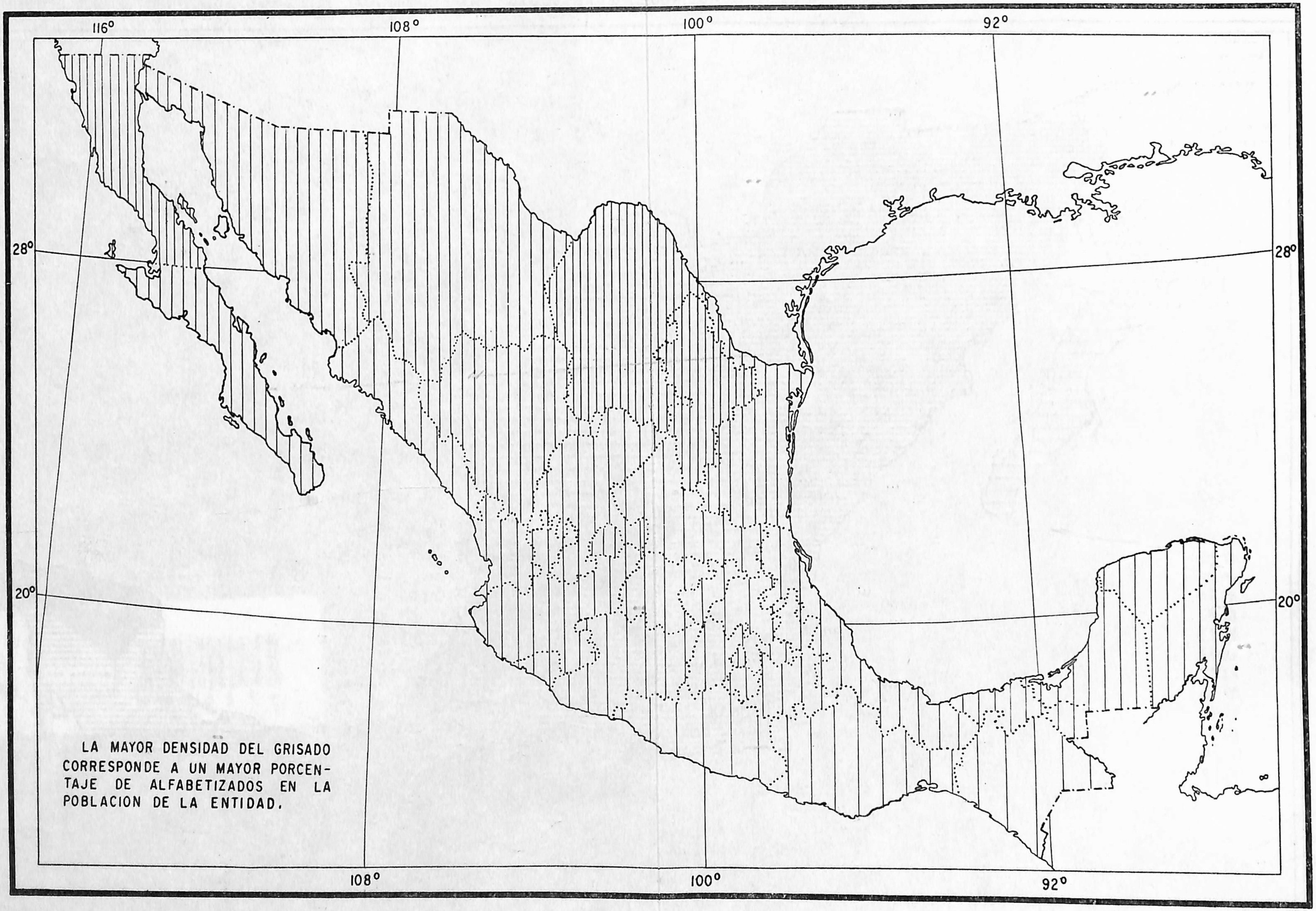
De estos datos se han hecho dos representacioses cartográficas: una (la de los hispanoparlantes) en papel opaco; la otra (la de los alfabetizados) en papel transparente.

ESTADOS UNIDOS MEXICANOS

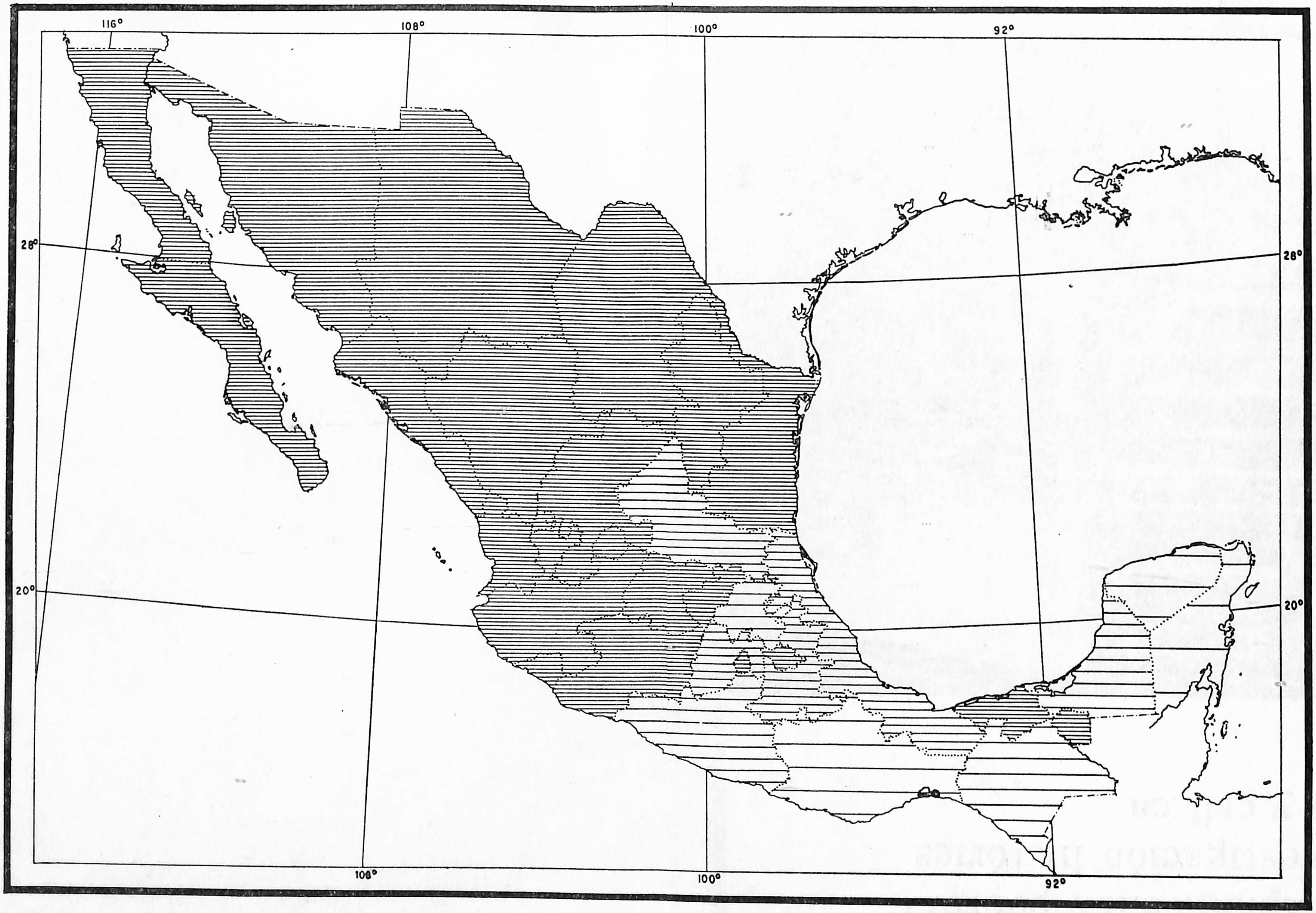
MONOLINGÜISMO ABSOLUTO



ALFABETIZADOS DE LA POBLACION DE MEXICO EN 1960



HISPANOPARLANTES DE MEXICO EN 1960



dib. J. CALONICO

El mapa de hispanoparlantes se ha grisado mediante paralelas horizontales.
El mapa de alfabetizados se ha grisado mediante paralelas verticales.
Las equivalencias entre los porcientos y la densidad de los grisados es la siguiente:

De 90 a 100%

De 80 a 90

De 70 a 80

De 60 a 70

De 50 a 60

De 40 a 50

De 30 a 40

Esto significa que si al superponer los mapas se forman cuadrados con los grisados horizontales y los verticales, existe concordancia entre la posición de la entidad según monolingües hispanoparlantes y la posición de la entidad de acuerdo con la proporción de alfabetizados en ella; que si se forman rectángulos, la posición de la entidad de acuerdo con una de estas proporciones es inferior a la que ocupa de acuerdo con la otra. La observación de estas discrepancias y del sentido de las mismas, permitirá continuar la investigación, gracias a esta ayuda proporcionada por la representación cartográfica de los correspondientes datos estadísticos.

EJEMPLO DE MAPA ESTADÍSTICO

Monolingüismo absoluto por entidades federativas en México, en 1960

Para elaborar este mapa, se utilizaron los datos consignados en el resumen general del VIII Censo General de Población de México, que consigna el total de quienes hablaban sólo una lengua indígena, para cada una de las entidades del país.

Los datos se ordenaron en forma decreciente, dando el número 1 a la entidad que tenía en 1960 el máximo de monolingües indígenas y terminando con el lugar número 32 asignado a la entidad que en esa fecha tenía el mínimo de monolingües indígenas.

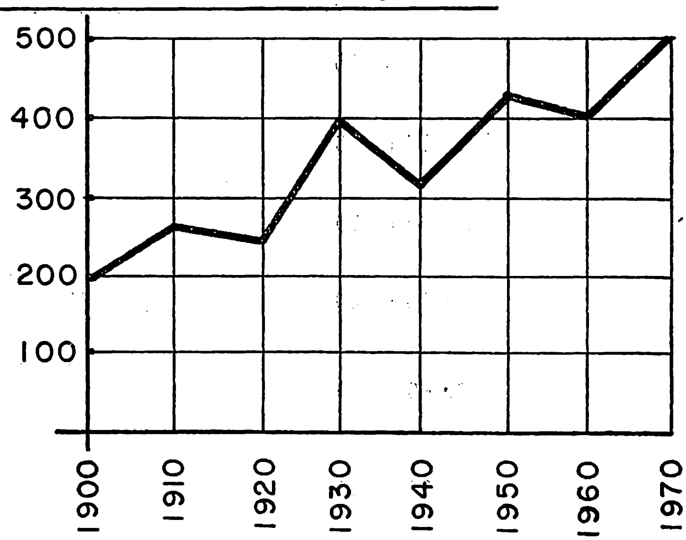
En seguida, se determinó cuáles eran las entidades que ocupaban la primera cuarta parte, cuáles las que ocupaban las dos siguientes cuartas partes (o centro) y cuáles las que ocupaban la cuarta parte final de la ordenación correspondiente.

Finalmente, se asignaron: un grisado más denso (de líneas paralelas más próximas entre sí) a las entidades que tenían —en estos términos— excesivo número de monolingües indígenas; uno intermedio a las que ocupaban —en esto— sitios que podían considerarse normales dentro de la situación sociolingüística de México en 1960, y uno menos denso a las que, en esto ocupaban posiciones de deficiencia.

Esto explica los tres grisados distintos que aparecen en el mapa, que se presenta a modo de ejemplo de cómo ciertos datos y resultados estadísticos pueden presentarse cartográficamente y dar así una visión plástica de una situación o problema social.

EJEMPLO DE REPRESENTACIÓN DE SERIE CRONOLÓGICA
Pobladores de Kalabá entre 1900 y 1970

Años	Pobl.
1900	200
1910	250
1920	240
1930	400
1940	320
1950	410
1960	400
1970	500

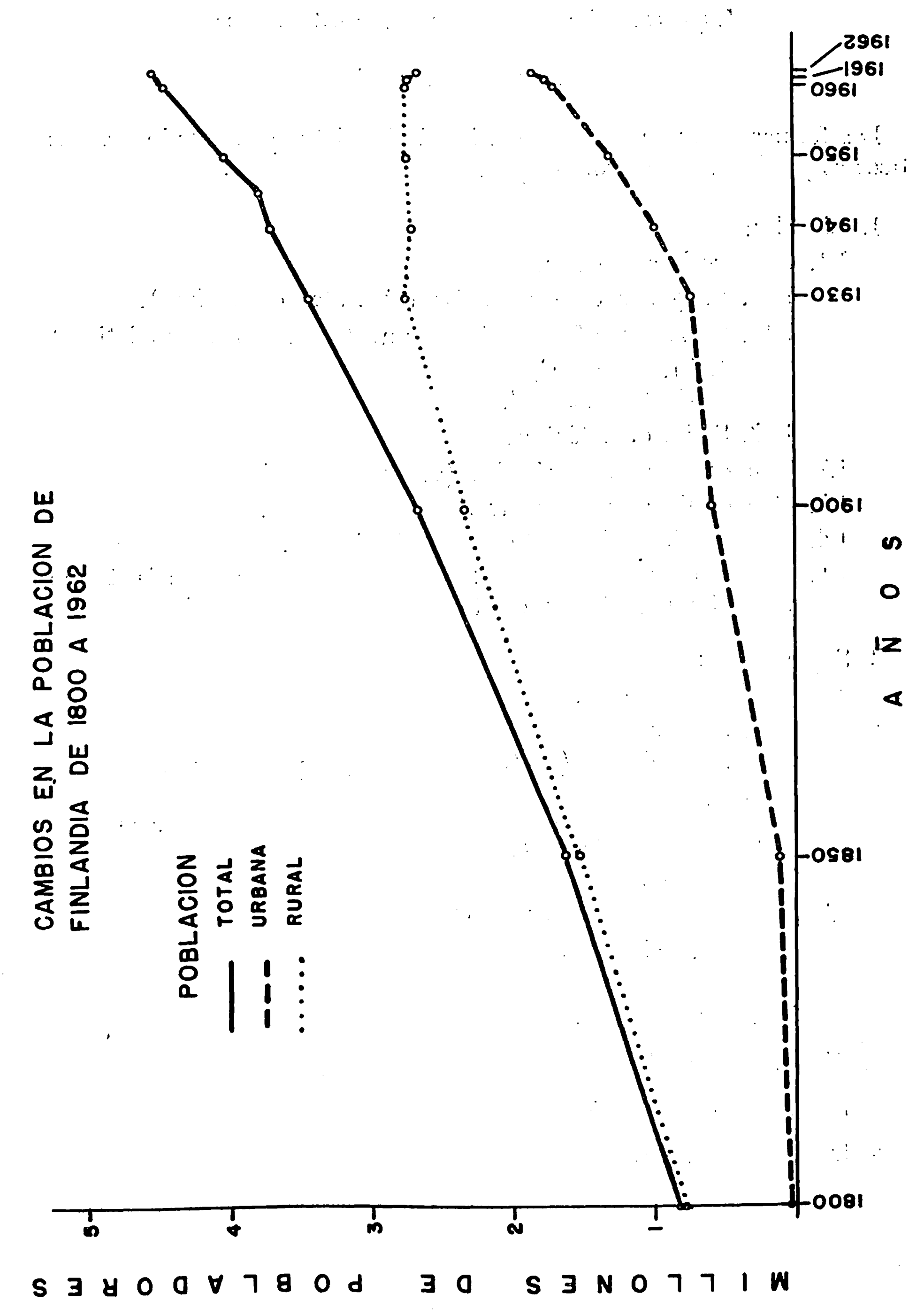


EJEMPLO DE REPRESENTACIÓN DE SERIES CRONOLÓGICAS, UNA SERIE RESULTANTE Y SUS DOS SERIES COMPONENTES

Cambios en la población total, y en las poblaciones urbana y rural de Finlandia, de 1800 a 1962

Años	*Población total*	*Población urbana*	*Población rural*
1800	832 700	46 600	786 100
1850	1 637 000	101 300	1 523 000
1900	2 655 900	333 300	2 322 600
1930	3 462 700	715 000	2 747 700
1940	3 695 600	991 700	2 703 900
1945	3 778 900	no hay dato	no hay dato
1950	4 029 800	1 302 400	2 727 400
1960	4 452 800	1 708 000	2 740 600
1961	no hay dato	1 753 300	2 736 700
1962	4 523 000	1 863 375	2 659 690

FUENTE: Finlandia en perspectiva.


CAMBIOS EN LA POBLACION DE
FINLANDIA DE 1800 A 1962
POBLACION
TOTAL
URBANA
RURAL
MILLONES DE POBLADORES
1
2
3
4
5
AÑOS
1800
1850
1900
1930
1940
1950
1960
1961
1962

Representación gráfica de datos estadísticos

Formas de representación

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

1. Lineales:
   1.1. Con escala aritmética.
      1.11. Polígonos de frecuencias (absolutas o relativas).
      1.12. Ojivas (con frecuencias acumulativas absolutas o relativas).
         1.121. "más de",
         1.122. "menos de".
      1.13. Series cronológicas.
   1.2. Con escala semilogarítmica (series cronológicas).
   1.3. Con escala logarítmica (series cronológicas).
   1.4. Con rayado probabilístico (distribuciones).
   1.5. Flechas de grosor variable (para expresar la importancia de movimientos migratorios, del tráfico citadino, etcétera).

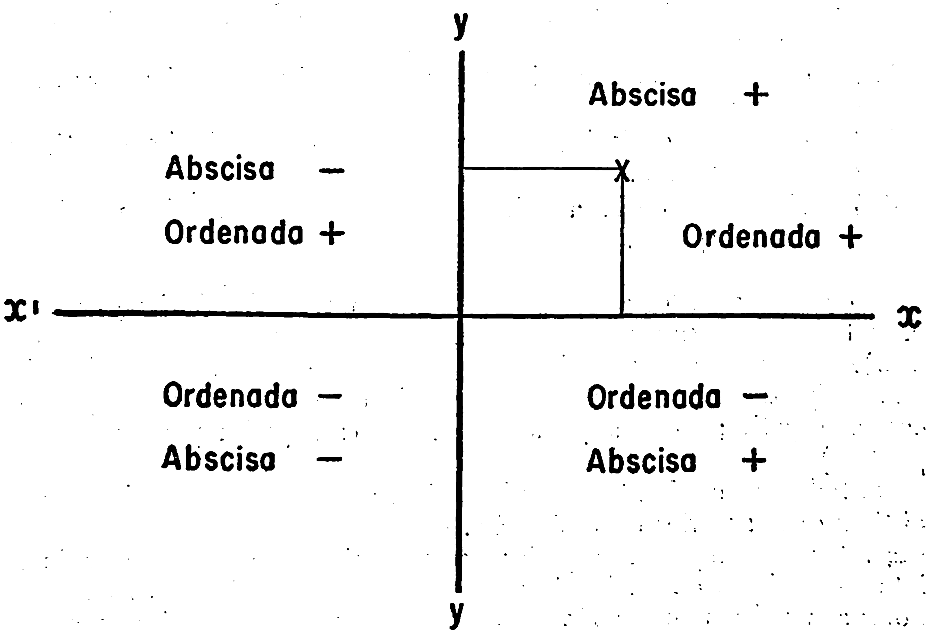
2. Superficiales:
   2.1. Gráficas de barras.
      2.11. Simples.
      2.12. Subdivididas (en forma absoluta o porcentual).
      2.13. Agrupadas.
      2.14. Pareadas (distribuciones por edad y sexo, etcétera).
      2.15. De desviaciones (o pérdidas y ganancias).
      2.16. Deslizantes.
   2.2. Histogramas y gráficas de escalones.
   2.3. Diagramas circulares.
      2.31. En pastel.
      2.32. En mariposa.

3. Sólidos o de volúmenes.

4. Diagramas de dispersión.

5. Mapas estadísticos.

SISTEMA DE COORDENADAS RECTANGULARES



La gran mayoría de las presentaciones gráficas se basa en el uso de un sistema de coordenadas cartesianas o rectangulares. Dicho sistema está constituido por dos rectas (una vertical y una horizontal) que se cortan en ángulo recto. A la vertical se le designa generalmente como "eje de las yes"; a la horizontal, se le llama "eje de las equis". Sus extremos se designan YY' (para la vertical) y XX' (horizontal). El punto en el que las rectas se cortan se llama "origen" y se le designa por 0.

Para localizar un punto P en el plano, se trazan desde él, dos perpendiculares (una a cada uno de los ejes), y se miden sobre ellas las distancias (dos) del punto a los ejes. Esas distancias se llaman "coordenadas" del punto. A la distancia del punto al eje horizontal se le llama "ordenada", a la del punto al eje vertical se le llama "abscisa".

Por convención, se considera que:

1. Las ordenadas que se miden hacia arriba del eje horizontal son positivas.
2. Las ordenadas que se miden hacia abajo de dicho eje 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:

1. En el cuadrante superior de la derecha, todos los puntos tienen abscisa positiva y ordenada positiva.
2. En el superior de la izquierda, los puntos tienen abscisa negativa y ordenada positiva.
3. En el inferior de la izquierda, los puntos tienen abscisa negativa y ordenada negativa.
4. En el inferior de la derecha, los puntos tienen abscisa positiva y ordenada negativa.

Es decir, que en los cuadrantes primero y tercero, la abscisa y la ordenada tienen el mismo signo (positivo en el primero, negativo en el tercero). En los otros dos cuadrantes (segundo y cuarto), las coordenadas tienen signos contrarios (en el segundo, la ordenada es positiva y en el cuarto, negativa; en el segundo, la abscisa es negativa y en el cuarto, positiva).

De todos los cuadrantes, el más utilizado en las representaciones estadísticas es el superior de la derecha. Ocasionalmente se usan los restantes; así, por ejemplo, el superior de la izquierda se emplea cuando se interpolan líneas de tendencia utilizando el año mediano; los dos inferiores se pueden usar cuando se representan variaciones cíclicas, pérdidas y ganancias, etcétera.

Para facilitar la localización de los puntos en el plano, se gradúan los dos ejes. Para ello, se toman sobre ellos espacios iguales que representen una unidad que se elija. Así, para localizar un punto cuyas coordenadas se conocen, se toman dichas coordenadas sobre cada uno de los 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: 1) se contarán m unidades a partir del origen, sobre el eje horizontal, y en ese punto se levantará una perpendicular; 2) se contarán n unidades a partir del origen sobre el eje vertical y en ese punto se levantará una perpendicular (o sea, una paralela al eje horizontal) hasta cortar la perpendicular trazada en primer lugar. En donde se corten esas perpendiculares estará el punto.

El sentido en el que se han de contar las unidades de la abscisa y las de la ordenada correspondientes dependerá de su signo. Así, si la abscisa es positiva, las unidades se contarán hacia la derecha del origen; si 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 (aunque en sentido lato se siga diciendo, también en este caso, que "se levantará" la perpendicular).

Se facilita aún más la localización de los puntos cuando se trabaja con papel cuadriculado o milimétrico; en tal caso no se necesitan trazar las perpendiculares correspondientes pues basta con contar los espacios 1) de la abscisa sobre el eje horizontal, y 2) a partir del punto de llegada y hacia arriba (o hacia abajo, según el signo), los espacios que corresponden a la ordenada.

Polígono de frecuencias

Para trazar un polígono de frecuencias, mediante el uso de las coordenadas rectangulares: 1) fórmese, con los datos, una distribución de frecuencias; 2) tómense como abscisas: 2.1) los valores de los datos, o 2.2) los puntos medios o marcas de clase, y en tales puntos, levántense las perpendiculares correspondientes; 3) tómese como ordenadas las frecuencias correspondientes 4). En los puntos determinados por esas coordenadas se encuentran los "vértices" del polígono 5). Únanse los vértices por medio de líneas rectas; 6) la línea quebrada resultante es el "polígono de frecuencias" representativo de la distribución.

Lectura de un polígono de frecuencias

Para poder leer un polígono de frecuencias: 1) se toma cada uno de los vértices del mismo (o cualquier punto que se encuentre sobre alguna de las rectas que los unen); 2) se mide, sobre la perpendicular (bajada del punto al eje) la distancia al eje horizontal; 3) se inicia la lectura diciendo "Hay tantos casos" (aquí la distancia medida) "para los cuales la magnitud del fenómeno tal es de..."; 4) se lee sobre el eje horizontal la distancia entre el pie de la perpendicular y el origen; 5) se completa la oración anterior diciendo "...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 un valor de x (determinado): 1) se buscará el valor de x sobre el eje horizontal; 2) en este punto se levantará una perpendicular, y 3) se medirá la distancia entre el pie de la perpendicular y la intersección de la perpendicular con el polígono de frecuencias; 4) esa medida indica cuántos casos tienen la magnitud x del fenómeno que se estudia.

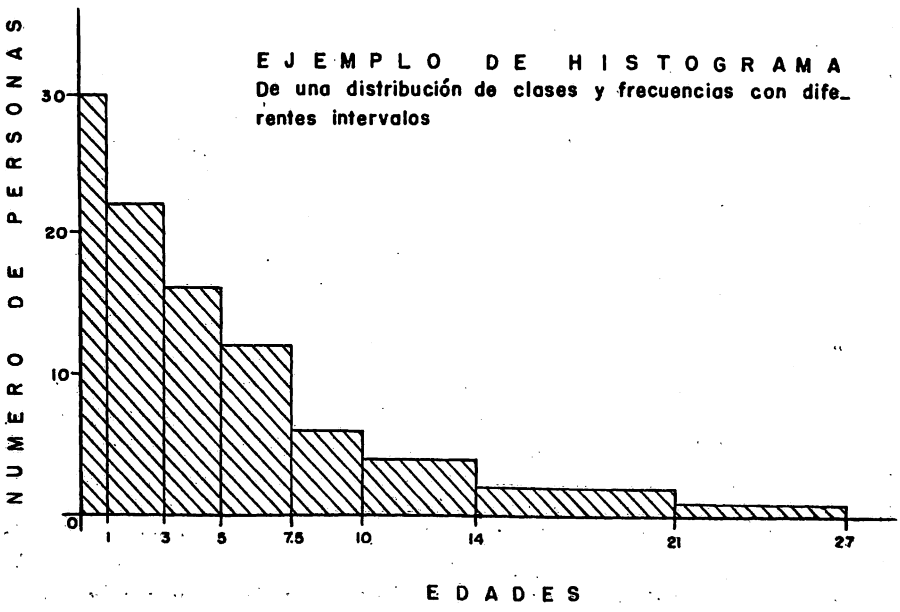
El trazo de un polígono de frecuencias no sólo sirve para "retratar" la forma de la distribución o para tener una idea de cuál puede ser su perfil teórico (una vez eliminados los errores accidentales y de otro tipo) sino que también permite tener una idea de cuál puede ser la frecuencia con que se presenten magnitudes intermedias del fenómeno que la agrupación en una serie de clases subsume bajo una misma denominación de magnitud (el punto medio o la marca de clase) o de aquellas magnitudes intermedias del fenómeno de las que no pudieron obtenerse las frecuencias en forma directa, mediante observación.

Los polígonos de frecuencias también sirven para comparar dos o más distribuciones. Esto es posible, especialmente, cuando se trata de polígonos que representan frecuencias relativas. Una frecuencia relativa se obtiene dividiendo la frecuencia absoluta correspondiente entre la suma de todas las frecuencias: $f_r = f / S f$. Para todos los fines prácticos, a una frecuencia relativa, en estadística, también se le puede considerar como una probabilidad de que ocurra la magnitud correspondiente. Así, si f_r corresponde a la magnitud r del fenómeno, se dice que hay "f_r posibilidades o probabilidades de que el fenómeno tenga la magnitud r".

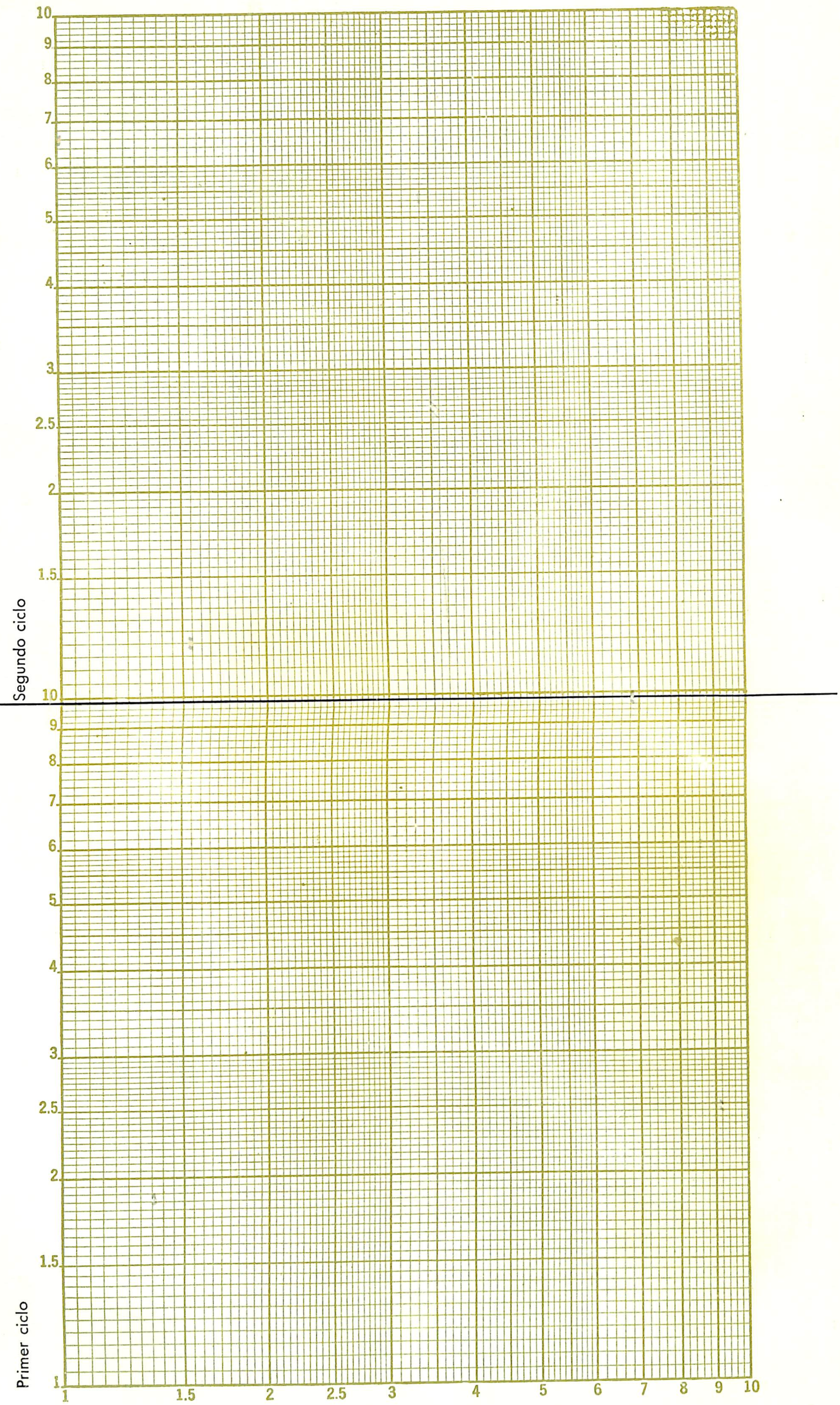
La representación de un polígono de frecuencias relativas sigue la misma rutina que la de un polígono de frecuencias absolutas. La lectura de este tipo de gráfica es análogo: se tratará de que: "Hay tantas probabilidades (aquí la frecuencia relativa u ordenada) de que la magnitud del fenómeno sea tal (aquí la abscisa)."

Cuando la frecuencia relativa se multiplica por cien, se obtiene una frecuencia porcentual. El polígono de frecuencias que se trace 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 formación 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 con trazo más grueso la paralela al eje horizontal que corresponda a la unidad (en el caso de los relativos) o al cien por ciento (en el caso de los tantos por ciento).



PAPEL LOGARÍTMICO (DE DOS CICLOS) PARA LA REPRESENTACIÓN ESTADÍSTICA



Nótese que el eje horizontal comienza en 1 no en 0 como en el rayado natural

EJEMPLO DE CÁLCULO DE LAS ALTURAS DEL HISTOGRAMA QUE REPRESENTA A UNA DISTRIBUCIÓN CUANDO LOS INTERVALOS DE CLASE NO SON IGUALES

Integración territorial mexicana por grupos de habitantes, en 1960

CATEGORÍAS *Localidades con este número de habitantes*	FRECUENCIAS *Número de localidades con esos habitantes*	PORCIENTO *respecto del total de localidades*	INTERVALO *al que corresponden esos porcientos*	
1 a 99	51 555	57.53	100	
100 a 499	27 098	30.24	400	
500 a 999	6 156	6.87	400	
1 000 a 2 499	3 342	3.73	1 500	
2 500 a 4 999	865	0.97	2 500	
5 000 a 9 999	340	0.38	5 000	
10 000 a 19 999	146	0.16	10 000	INTERVALO
20 000 a 29 999	38	0.04	10 000	ELEGIDO
30 000 a 39 999	20	0.02	10 000	PARA
40 000 a 49 999	11	0.01	10 000	ESTANDARIZAR
50 000 a 74 999	15	0.02	25 000	
75 000 a 99 999	9	0.01	25 000	
100 000 a 249 999	12	0.01	150 000	
250 000 a 499 999	2	0.01 —	250 000	
500 000 a 3 000 000	3	0.01 —	2 500 000	
Total de localidades	89 612			

RELACIÓN *entre el intervalo y el intervalo estándar*		ESTANDARIZACIÓN *Multiplicación de los porcientos por esa relación*
10 000/100	= 100	5 753.00
10 000/400	= 25	756.00
10 000/400	= 25	171.75
10 000/1 500	= 6.7	24.618
10 000/2 500	= 4	3.880
10 000/5 000	= 2	0.760
10 000/10 000	= 1	0.160
10 000/10 000	= 1	0.040
10 000/10 000	= 1	0.020
10 000/10 000	= 1	0.010
10 000/25 000	= 0.4	0.008
10 000/25 000	= 0.4	0.004
10 000/150 000	= 0.67	0.007
10 000/250 000	= 0.04	0.0004 —
10 000/2 500 000	= 0.004	0.00004 —

EJEMPLO DE LECTURA DE UN HISTOGRAMA CUYOS INTERVALOS DE CLASE SON DIFERENTES

Integración territorial mexicana
por grupos de habitantes, en 1960

ANCHURAS *Fracciones de un intervalo estándar de 10 000*		ALTURAS	ÁREA *Multiplica ancho por alto*
0.01 =	100	5753.000	57.53%
0.04 =	400	756.000	30.24%
0.04 =	400	171.750	6.87%
0.15 =	1 500	24.618	3.69%
0.25 =	2 500	3.880	0.97%
0.50 =	5 000	0.760	0.38%
1.00 =	10 000	0.160	0.16%
1.00 =	10 000	0.040	0.04%
1.00 =	10 000	0.020	0.02%
1.00 =	10 000	0.010	0.01%
2.50 =	25 000	0.008	0.02%
2.50 =	25 000	0.004	0.01%
15.00 =	150 000	0.007	0.01%

LÉASE: *En el intervalo que principia en . . .*	*y que abarca . . . unidades* (incluido límite)	*se encuentra el siguiente porciento de las entidades*
1	100	57.53
100	400	30.24
500	400	6.87
1 000	1 500	3.69
2 500	2 500	0.97
5 000	5 000	0.38
10 000	10 000	0.16
20 000	10 000	0.04
30 000	10 000	0.02
40 000	10 000	0.01
50 000	25 000	0.02
75 000	25 000	0.01
100 000	150 000	0.01
250 000	250 000	0.01—

Ojivas

Las ojivas son representaciones esiformes, sigmoides o de forma de s, que corresponden a las distribuciones acumulativas. Las distribuciones acumulativas, en vez de informar acerca de cuántos casos tiene una magnitud determinada, informan sobre cuántos casos tienen magnitudes iguales o inferiores a cierta magnitud (para lo cual se suman todas las frecuencias de las clases anteriores a aquella de la que se trate) o cuántos tienen una magnitud igual o superior a cierta magnitud (en cuyo caso, las que se suman son las frecuencias de las clases que siguen a aquella de cuya magnitud se trate). Hay, en forma correlativa, dos tipos de ojiva: 1) la ojiva "menos de", y 2) la ojiva "más de", para cada distribución. La primera, indica gráficamente cuántos casos son menores que cierto valor; la segunda, cuántos son mayores que dicho valor.

Para trazar las ojivas: 1) fórmese una distribución acumulativa; 1.1) sumando las frecuencias en forma progresiva, a partir de la correspondiente a los valores o clases inferiores de la variable y continuando hacia los superiores, si se trata de la ojiva "menos de"; 1.2) sumando las frecuencias en esa misma forma progresiva, pero *en sentido inverso* (de los valores o clases superiores a los inferiores), si se trata de la "más de"; 2) tómense los valores de la variable más o menos un medio (en series sencillas) o los límites de clase (en series de clases); 2.1) tomando los valores más un medio, o los límites superiores, en el caso de la ojiva "menos de", y 2.2) tomando los valores de la variable menos un medio, o los límites inferiores en 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, que son los "vértices" de la ojiva, y 4) únanse los vértices para obtener finalmente dicha 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 del punto al eje horizontal la distancia al eje, leyendo "Hay tantos casos (aquí la distancia medida) para los cuales la magnitud del fenómeno tal 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 tal es más de tanto".

Como en el caso de los polígonos de frecuencias, en el caso de las ojivas también se pueden utilizar relativos o tantos por ciento, en calidad de frecuencias. Las lecturas serán: "Hay tantas probabilidades de que la magnitud del fenómeno tal sea de menos (más de) tanto"; "De cada cien casos, hay tantos para los cuales la magnitud del fenómeno es menos de (más de) tanto."

Recuentos clasificatorios

Ocasionalmente, puede haber necesidad de hacer el recuento de los casos que correspondan a un género y, simultáneamente, de los que correspondan a sus especies (a fin de evitar, con ello, una duplicación de esfuerzos). Para concretar, puede pensarse en un recuento del vocabulario de varios periódicos de un país (*El Universal, Novedades, Excelsior, La Prensa,* etcétera) y en una simultánea determinación del vocabulario usado por la sección informativa, por la de artículos y por el conjunto de los editoriales de todos los periódicos del país. En ese caso, conviene diferenciar los palotes del recuento, empleando un color diferente para cada una de las secciones (rojo, para informaciones; verde, para artículos; negro, para editoriales, por ejemplo). En estas condiciones, en el momento del recuento, si se trata de determinar el número de voces empleadas por cada periódico, podrán contarse todos los palotes que le correspondan, sin tener en consideración el color de cada uno; si se trata de determinar el número de voces empleadas por las secciones informativas de todos los periódicos (o de uno de ellos en particular) habrá que contar sólo los palotes rojos, sólo los verdes, sólo los negros, etcétera.

Cuando se desea combinar una serie de criterios clasificatorios, a más de la distinción de los palotes por su color, se pueden emplear —en el recuento— series distintas de palotes (horizontales, verticales, inclinados) para cada una de las clases que se constituyan con el segundo criterio de clasificación. Según esto, el número de palotes (sin especificar color u orientación) dará la frecuencia genérica; el número de palotes de cada color dará las frecuencias específicas según un primer criterio de clasificación; el número de palotes de una misma inclinación dará las frecuencias específicas según un segundo criterio de clasificación, y el número de los palotes que tengan cierto color y determinada inclinación dará la frecuencia correspondiente a una clasificación conjunta, de acuerdo con ambos criterios.

Una distribución simple (o distribución simple de frecuencia), una vez formada, se consigna, en forma tabular, en dos columnas: 1) una contiene los valores (no repetidos) a los que corresponde una frecuencia distinta de cero (o sea, los que aparecen efectivamente en el conjunto), y 2) otra, los números (no ya los cuadrados o barras del recuento) correspondientes a la frecuencia o número de veces que aparece cada dato.

Manipulación matemática de los datos

Operaciones predominantemente matemáticas

Simbología estadística

La simbología estadística es, fundamentalmente, la misma simbología matemática. Está constituida por un conjunto de símbolos que pueden agruparse en las siguientes categorías: 1) símbolos que carecen de valor como magnitudes; 2) símbolos que indican orden, y 3) símbolos que representan magnitudes.

Los símbolos que no representan magnitudes pueden ser: 1.1) signos de operación; 1.2) operadores, y 1.3) signos de relación.

Los signos de operación son: $+$, $-$, $\times$, $\div$ / exponente, $\sqrt{}$

$+$ indica: cuéntense a partir de la última unidad de la magnitud colocada antes del signo, tantas unidades como indique la cifra colocada después de él; es decir "súmese".

$-$ significa: "cuéntense a partir de la última unidad de la cifra colocada antes del signo, pero *en sentido inverso*, tantas unidades como indique la cifra colocada después de él; es decir "réstese de... tanto, tanto".

$\times$ indica: súmese la cifra colocada antes del signo tantas veces consigo misma como indica la cifra colocada después de él; o sea, "multiplíquese esto por esto".

$\div$ equivale a: determínese cuántas veces hay que restar de la cifra colocada delante del signo la colocada después de él; 1) hasta que se obtenga cero, o 2) hasta que se obtenga una cifra menor que la colocada después del signo; es decir "divídase esto entre esto".

/ exponente quiere decir: multiplíquese la cifra colocada bajo el exponente tantas veces por sí misma como indique el exponente o cifra colocada arriba de ella; es decir: "elévese la cifra a la potencia tal".

$\sqrt[a]{b}$ significa: búsquese el número que tomando como factor tantas veces como indica a produce b.

Los operadores son signos condensados o taquigráficos de operación. Los que se emplean más en estadística son: las deltas minúscula y mayúscula (δ, Δ), la ese mayúscula latina o griega (S o Σ, sigma) y una ese latina alargada ($\int$).

Δ Delta mayúscula vale tanto como decir "increméntese o auméntese el valor siguiente en una cierta cantidad".

δ_x y delta minúscula se usa en ocasiones para indicar "tómese la derivada de lo siguiente (y)" o bien "tómese el límite de la relación entre el incremento de la cantidad colocada delante (y) y el incremento de la cantidad colocada como subíndice (x) de la delta, cuando el incremento de esta última cantidad (Δ_x) tiende hacia cero"; es decir, "tómese la derivada de y con respecto a x".

S o Σ (sigma mayúscula) equivale al imperativo "súmese todo lo que sigue" o bien "tómense sumandos del tipo del que sigue al operador, sigma mayúscula", o bien "súmense todos los *términos* del tipo siguiente".

$\int_v u$ (ese latina alargada) indica: "realícese con la cantidad colocada delante (u), la operación inversa de la derivación, tomando como variable de referencia el subíndice", o sea: "intégrese a (u) con respecto a v". En el fondo, se trata de un proceso de suma, en el que cada sumando es muy próximo del siguiente.

De los principales símbolos de relación, hay que considerar:

$$=, \neq, <, >.$$

$=$ indica: la cantidad colocada delante del signo es igual a la colocada detrás del signo. O bien: si se realizan las operaciones que quedan indicadas adelante del signo se obtendrán los mismos resultados que se obtendrían al realizar las operaciones indicadas después de él.

$\neq$ es la negación del anterior. Se trata de un signo genérico que afirma que la cantidad o el resultado de las operaciones que anteceden al signo difieren de la cantidad o del resultado de las operaciones que lo subsiguen. Los signos que especifican este tipo de relación, son: $<$ y $>$ ("menor que" y "mayor que").

$<$ señala que la cantidad colocada delante del signo no sólo es distinta de la colocada detrás de él, sino que es menor que ella.

$>$ precisa que la cantidad colocada antes del signo no sólo difiere de la colocada después, sino que es mayor que ella.

Estos signos de relación suelen combinarse: 1) para precisar relaciones más complejas, o 2) para indicar límites o campos de variación.

1.31. Símbolos indicadores de relaciones complejas. Resultan de la combinación del signo igual con uno de los precedentes:

$\leqslant$ o $\leqq$ representa lo mismo que $<$ o $=$ y quiere decir que la cifra o el resultado de las operaciones que preceden al signo son menores o iguales, pero *no mayores* que la cifra o el resultado de las operaciones que lo suceden. En realidad podría sustituirse por el signo *no mayor que* (que no se usa mucho).

$\geqslant$ representa tanto como $>$ o $=$ y quiere decir que lo que precede al signo es igual o mayor, pero *no menor* que lo que subsigue. Se podría sustituir por un signo *no menor que* (que también se usa poco).

Con un cierto rigor, el orden y la lectura de los signos complejos de relación debería ser:

$\leqslant$ menor que o igual a ... Nótese la paralela *interior* del "igual".

$\geqslant$ igual a o mayor que ... Nótese de nuevo la paralela *exterior*.

1.32. Delimitación de campos de variación. En estos casos, la variable se coloca entre los dos signos, y los límites entre los que varía a uno y otro lado de ellos. Ese campo de variación se conoce también, estadísticamente, como "amplitud" y a los límites, como "máximo" y "mínimo".

$10 < a < 100$ indica que *a* es mayor que 10 y menor que 100; o sea que la amplitud es de 10 a 100, pero el mínimo y el máximo no están determinados pues no se indica que *a* pueda ser igual a 10 o igual a 100.

Si el lector tiene dificultad para leer la expresión compleja, puede descomponerla en las dos simples siguientes:

$10 < a$ que quiere decir: "10 es menor que a", o sea "a es mayor que 10", pero $a < 100$: "a es menor que 100".

Esta forma de expresar el campo de oscilación *excluye* los límites; si se desean incluir éstos, a más de usar el doble símbolo de relación hay que usar el doble símbolo complejo.

$10 \leqslant a \leqslant 100$, que significa que *a* puede ser igual o mayor que 10 pero menor o igual a 100. En este caso, el mínimo es 10 (*a* no puede ser menor que 10) y el máximo 100 (*a* puede ser igual a 100, incluso).

2. Los símbolos de valor distintivo ordinal están representados generalmente por apóstrofos, por letras latinas o por números romanos. Así:

x, x', x'', x''' indican la primera, la segunda, la tercera equis, y se leen "equis", "equis prima", "equis biprima", etcétera.

x, x^s, x^t, x^u también indican ese orden (con el inconveniente de que las letras se pueden tomar como exponentes).

x, x^i, x^{ii},
x^{iii}, x^{iv}, en forma parecida.

3. Los símbolos que indican magnitudes pueden corresponder a: 3.1) constantes, y 3.2) variables.

Las constantes tienen un valor único, o un conjunto único de valores. Se representan, generalmente, por las mayúsculas o minúsculas de los albafetos latino y griego (aunque no se excluyan otros signos). Como generalmente conviene distinguir entre cantidades conocidas y cantidades desconocidas (datos e incógnitas, respectivamente) se usan las primeras letras para los datos y las últimas para las incógnitas. Cuando en un desarrollo conviene establecer alguna relación entre unas constantes originales y otras que de ellas puedan resultar, conviene emplear: primero, las minúsculas latinas (a, b, c...) después, sus mayúsculas, en forma correspondiente (A, B C...); después, las minúsculas griegas y, tras ellas, las mayúsculas griegas (α, β, γ...) (Α, Β Γ...) Sin embargo, en los últimos casos se presenta una dificultad —conocida del lingüista— ya que si las correspondencias se establecen de acuerdo con el puro orden, γ, por ejemplo, cuyo sonido es de g, corresponde a c, cuyo sonido, en español es de k o de s. Por otra parte, muchas de las mayúsculas griegas (alfa y beta mayúsculas, por ejemplo) se confunden con las correspondientes latinas y esto puede dar lugar a errores, por lo que conviene ser cauto al establecer esas correspondencias. Se debe precisar, en efecto, cuál es la convención que usa (correspondencia de sonido o de grafía) y cómo se deshace la anfibiología (A' para alfa mayúscula griega, por ejemplo frente a A para la mayúscula latina correspondiente).

Las constantes pueden ser: 3.11) constantes absolutas, y 3.12) constantes relativas.

Las constantes absolutas son aquellas que tienen un solo valor sea cual fuere el contexto en que se encuentren; en términos análogos a los que usa el lingüista, podría llamárseles "constantes libres de contexto".

π (pi minúscula) vale siempre 3.1416...

e (e minúscula) vale siempre 2.7182...

Las constantes relativas son aquellas que tienen un solo valor dentro de determinado contexto o en determinada ecuación, pero que alcanzan valores diferentes para diferentes contextos o ecuaciones (y por ello puede designárseles como "constantes ligadas por el contexto"). Así, en la ecuación de la recta ($y = a + b\,x$) "a" valdrá 3 *para la recta* que corte el eje de las yes en el punto 3; valdrá —8, para la recta que corte a dicho eje ocho unidades por debajo del origen: "a" será *constante*, siempre que se utilice esta ecuación de la recta, *para todas las rectas* posibles (lo sea, para el sistema rectilíneo) pero, dentro del conjunto de rectas, variará de una recta, concreta, particular, a otra.

Cuando se manejan sistemas de curvas (y no curvas aisladas) conviene enfatizar la relatividad de ciertas constantes que intervienen en las ecuaciones correspondientes afectándolas de subíndices, en forma parecida a como, según veremos dentro de un momento, se afectan las variables (si $y = a + bx$ representa la ecuación de una recta, en general y tenemos que representar 3 rectas distintas, las podemos expresar por:

$y_1 = a_1 + b_1\,x_1$; $y_2 = a_2 + b_2\,x_2$; $y_3 = a_3 + b_3\,x_3$).

Las variables son símbolos a los que corresponden valores múltiples. El signo representativo de una variable está en lugar de todo un género o conjunto de valores, sin que represente a ninguno de ellos en particular. En general, para indicar una variable no especificada, se suele afectar el símbolo correspondiente de un subíndice genérico. Ese subíndice genérico suele ser, convencionalmente, i minúscula; así x_i representa todos los valores posibles de la variable x; y_i, todos los posibles de la variable y, etcétera. Otro subíndice que también se emplea, cuando se quiere distinguir el dominio de variación de una variable del ya especificado por el subíndice i, es el subíndice j: u_j representa todos los valores posibles de la variable u. En castellano y otras lenguas que pluralizan agregando s conviene —según nuestro parecer— emplear el subíndice s (minúscula o mayúscula) en cuanto esto permite leer muchas veces como si se tratara de un plural: "Súmense las x_s", se puede leer directamente, "súmense las equis".

Así, si x representa la variable "edad" para un conjunto de tres niños de 14, 11 y 10 años, x_s (o x_i) 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 cambiará el subíndice genérico s por 1: así x_1 representa 14 años; para especificar que se trata de la edad del segundo, s se sustituirá por 2: x_2, representará 11 años, y x_3 representará los 10 años del tercer niño.

Denominación de los elementos de una operación

Las cifras unidas entre sí por el signo + se denominan "sumandos"; la cifra que precede al signo — se denomina "minuendo" y la que lo sucede, "sustraendo". Pero, en general, toda cifra o expresión unida por los signos + o — se llama "término".

Las cifras o las expresiones unidas entre sí por el signo x se denominan "multiplicando" (la que lo precede) y, "multiplicador" (la que lo subsigue); pero, en general, se conocen como "factores".

A la cifra o expresión que precede al signo ÷ se le denomina "dividendo", y a la que lo subsigue, "divisor". Como la división se representa también por medio de fracciones, a la cifra o expresión colocada encima de la raya del quebrado se le denomina "numerador" o dividendo, y a la colocada debajo, "denominador" o divisor.

Como en el caso de la suma y la resta, toda expresión en la que intervienen los signos x y ÷ puede considerarse como un factor (si interviene ÷ se habla de un factor fraccionario).

A la cifra que indica cuántas veces hay que tomar otra como factor común o igual, se le llama "exponente". A la que indica que hay que buscar cuál es el número que multiplicado por sí mismo tantas cuantas veces indica esa cifra, produce cierto valor, se le llama "índice del radical". En términos generales, a exponentes e índices de radical se les llama exponentes (considerándose como un exponente fraccionario el índice del radical). Así, un índice de radical 2 equivale a un exponente 1/2; un índice 3 a un exponente fraccionario 1/3, etcétera.

Simbología muestreal

Cuando se trabaja con muestras se presenta frecuentemente la necesidad de distinguir las medidas estadísticas que corresponden a una o varias muestras y las que corresponden a la población de la que forman parte (las "estadísticas" y los "parámetros"). En tales casos se recurre a la distinción siguiente:

Los parámetros o medidas características de la población se representan mediante letras griegas; las estadísticas o medidas que caracterizan a las muestras se representan mediante letras latinas.

Así, mientras

ξ_1 representa la media aritmética de la población,

$\bar{x}_{1\text{-}1}$ representa la media aritmética de una primera muestra,

$\bar{x}_{1\text{-}n}$ representa la media aritmética de la enésima muestra,

$\bar{x}_{m\text{-}n}$ representa la media de orden m de la enésima muestra,

ξ_m representa la media de orden m de la población.

En forma parecida, en tanto que σ representa la desviación media cuadrática de la población, s_n representa la desviación de la enésima muestra que se haya obtenido de ella; β_1, la medida pearsoniana de asimetría de la población y b_{1n} la medida pearsoniana de asimetría de su enésima muestra, así como β_2 representa la medida de curtosis de la población y b_{2n} la medida de curtosis de la enésima muestra de la población.

Signos de agrupación

Cuando se manejan cifras o expresiones simples no hay problemas de representación: cada una aparece en su sitio, antes o después de los signos de operación correspondientes. Cuando se manejan expresiones complejas (en las que queda indicado que hay que realizar ciertas operaciones antes de ejecutar otras) se utilizan ciertos signos de agrupación. Los signos de agrupación que más se usan son los paréntesis y sus afines, los corchetes o las llaves: ( ), [ ], { }, pero también se suele recurrir a otros procedimientos como el que consiste en subrayar o suprarrayar (cubrir con una línea) toda una expresión compleja.

En un primer nivel de complejidad conviene usar sólo paréntesis curvos: $(8 + 5) + (3 + 2)$ indica que queremos que se sumen primero 8 y 5; que después se sumen $3 + 2$ y después se sumen las sumas correspondientes (13 y 5) para obtener el resultado final. En forma parecida $(8 \times 5) + (7 \times 4)$ indica que queremos multiplicar primero 8 por 5, multiplicar después 7 por 4 y sumar finalmente los productos (40 más 28). Asimismo $(3 + 2) \times (5 + 7)$ indica que hay que sumar 3 y 2, sumar después 5 y 7 y multiplicar finalmente esas sumas (5 por 12). O sea, que, primero se deben ejecutar las operaciones de *dentro* de los paréntesis y luego las que se indican *entre* los paréntesis.

En un segundo nivel de complicación (si lo hay) conviene usar paréntesis rectangulares para abarcar expresiones en las que ya figuran paréntesis curvos: $[(8 + 5) + (3 + 2)] \times [(3 + 2) \times (5 + 7)]$ indica que primero hay que realizar la suma de los resultados de los dos primeros paréntesis curvos (como se indicó arriba), en seguida el producto de los resultados de los dos últimos y, finalmente, el producto de la suma de los primeros por el producto de los últimos. O sea, de nuevo, primero operaciones *dentro* de paréntesis rectangulares y después operaciones *entre* paréntesis rectangulares.

Si hay un tercer nivel de complicación, pueden emplearse llaves para abarcar las expresiones en paréntesis rectangulares. La regla será: primero las operaciones de dentro de las llaves y luego las indicadas entre las llaves.

En este tercer nivel de complicación, la operación total sería: 1) realizar las operaciones de dentro de los paréntesis curvos; 2) realizar las operaciones de dentro de los paréntesis rectangulares (que son operaciones entre paréntesis curvos); 3) realizar las operaciones de dentro de las llaves (que lo son entre paréntesis rectangulares), y 4) realizar las operaciones entre las llaves.

Cuando hay cifras o expresiones que no están en paréntesis se considera como si estuviesen encerradas en un paréntesis del orden correspondiente. $7\ (2 + 3)$ indica: $(7) \times 2 + 3)$; $5\ [(8 + 2) + (5 - 3)]$ vale tanto como si 5 estuviese en un paréntesis rectangular multiplicando al otro paréntesis rectangular.

Potencias de los números

Cuadrado de un número es el resultado que se obtiene de tomar al número dos veces como factor. Se expresa mediante un pequeño 2 colocado arriba del número cuyo cuadrado se indica. Elevar a 3 al cuadrado equivale a tomar 3 dos veces como factor, o multiplicar a 3 por sí mismo una vez: 3×3. El resultado (9) es el cuadrado de 3, que se representa por 3^2. O sea que: $3^2 = 9$.

Cubo de un número es el resultado de tomarlo tres veces como factor. En forma parecida a la anterior, el cubo de 5 es $5 \times 5 \times 5$. El resultado, 125 se expresa como 5^3 (cinco al cubo) o sea que $5^3 = 125$.

En forma inversa, si se tiene un número multiplicado por sí mismo una vez, el resultado se puede expresar escribiendo el número y afectándolo de un exponente dos. Así, 3×3 puede expresarse como 3^2; $7 \times 7 \times 7 \times 7 \times 7$ como 7^4 (siete a la cuarta potencia).

Elevación de una potencia a otra potencia

Supongamos que se trata de encontrar el cuadrado del cuadrado de 3. El cuadrado de 3 se representa por 3^2; el cuadrado de este cuadrado puede representarse encerrando toda la expresión en un paréntesis (3^2) y elevando, en seguida, al cuadrado dicho paréntesis: $(3^2)^2$.

Si aplicamos nuestras reglas sobre el uso de paréntesis tendremos que en este caso, habrá que realizar primero la operación de *dentro* del paréntesis ($3^2 = 9$) y después ejecutar con el resultado (9) la operación indicada *fuera* del paréntesis: 9^2. Como 9^2 equivale a multiplicar a 9 por sí mismo una vez, se tendrá: $9 \times 9 = 81$. O sea, que $(3^2)^2 = 81$. Esto mismo se puede evidenciar por otro camino.

Como elevar al cuadrado equivale a tomar dos veces como factor la cantidad afectada por el exponente, para obtener $(3^2)^2$ tomaremos dos factores iguales a la cantidad de dentro del paréntesis $3^2 \times 3^2$. Pero cada uno de estos 3^2 es igual a 3×3; así, los dos factores $3^2 \times 3^2$ pueden escribirse como $3 \times 3 \times 3 \times 3 = 81$.

Por lo dicho en el último párrafo, puede verse que $(3^2)^2$ se puede representar por $3 \times 3 \times 3 \times 3$; pero, como, por otra parte, $3 \times 3 \times 3 \times 3$ es el producto de cuatro factores iguales a 3, este resultado se puede expresar como 3^4. O sea que $(3^2)^2 = 3^4$.

Si se examina cuál es la relación que existe entre el exponente 4 del resultado y los exponentes 2 y 2 del cuadrado del cuadrado, podremos ver que $4 = 2 \times 2$, de donde podemos decir que:

Para elevar al cuadrado $(\quad)^2$ 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: $(a^2)^2 = a^{2 \times 2} = a^4$.

En general, aunque no lo probemos aquí, el cuadrado de una cantidad previamente elevada al cuadrado es igual a la cuarta potencia de dicha cantidad. Y podríamos mostrar igualmente que el cuadrado $(\quad)^2$ de un cubo (4^3) será la sexta potencia ($2 \times 3 = 6$) de la cantidad (4): 4^6. El cuadrado de una cuarta potencia $(\quad)^2$ de (7^4) será la octava potencia ($2 \times 4 = 8$) o sea 7^8. Y en general:

Para elevar al cuadrado una cantidad previamente elevada a otra potencia, bastará con duplicar el exponente de dicha cantidad: $(a^n)^2 = a^{2n}$.

En forma parecida, se puede afirmar que:

Para elevar a una potencia m $(\quad)^m$, una cantidad (a) previamente elevada a una potencia n (a^n), será necesario elevar la cantidad (a) a la potencia que se obtiene de multiplicar m por n (elevarla a la mn). Así $(a^n)^m = a^{nm}$.

Esto se aplica en el caso de exponentes positivos como los estudiados, en el de los negativos (que convierten a la expresión exponenciada en una fracción) y en el de los fraccionarios (que represéntan raíces de la expresión que exponencian).

Raíces de los números

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; ése número será su raíz cuadrada exacta; así, de 25, la raíz cuadrado es 5; de 81, 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 contiene dicho número; la raíz de este cuadrado perfecto, seguida de una parte decimal, será la raíz cuadrada del número. El cálculo se hace en la siguiente forma:

1º Sepárense periodos 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 (o sea, si 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 varias de las siguientes situaciones:

a) Que el último periodo entero tenga 2 cifras o que tenga solamente una. Lo segundo no constituye problema, pues se trabaja igual que si se tratara de un periodo completo.

b) Que el último periodo decimal tenga 2 cifras o una. En este último caso, se completa el último periodo 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 periodo entero (primero del número en total) está constituido por una sola cifra (6), pero esto no molesta, pues se puede buscar cuál es el máximo cuadrado perfecto contenido en 6 ($4 = 2 \times 2$) en la misma forma en que se busca el contenido en 30 ($25 = 5 \times 5$), por ejemplo. El último periodo 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 periodo (¿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). Anótese ese cuadrado debajo de él y réstesele, anotándose la raíz cuadrada del mismo (2) al lado del número cuya raíz se busca:

$$\begin{array}{l|l} \sqrt{6/32./53/20} & 2 \\ \ \underline{4} & \text{—} \\ \ 2 & \end{array}$$

3º Bájese al lado de la resta el segundo periodo.

$$\begin{array}{l|l} \sqrt{6/32./53/20} & 2 \\ \ \underline{4} & \text{—} \\ \ 2\ 32 & \\ & \end{array}$$

4º Divídase la cantidad entre 10 veces el duplo de la raíz.

a) La "cantidad" es la formada por la resta del periodo anterior y por el periodo recién bajado (232).

b) El "duplo de la raíz" se obtiene multiplicando por 2 la raíz hallada (2 en el ejemplo) y anotando el resultado (4) en una línea horizontal, paralela a la de la raíz.

5º Anótese el entero del cociente anterior (5, entero que resulta de dividir $232 \div 40$) en la raíz y en el duplo.

$$\begin{array}{l|l}
\sqrt{6/32./52/20} & 25 \\
4 & \text{——} \\
\text{—} & \\
2\ 32 & 45 \\
\cline{2-2}
 &
\end{array}$$

Multiplíquese ese entero (5) por la cifra (45) que figura ahora en la línea del duplo ($5 \times 45 = 225$).

Regístrese ese producto (225) debajo de la cantidad con la que se opera (232).

$$\begin{array}{l|l}
\sqrt{6/32./52/20} & 25 \\
4 & \text{——} \\
\text{—} & \\
232 & \\
225 & 45 \\
\text{——} & \text{———} \\
 &
\end{array}$$

Réstense esas dos cantidades

$$\begin{array}{l|l}
\sqrt{6/32./52/20} & 25 \\
4 & \text{——} \\
\text{—} & \\
232 & \\
225 & 45 \\
\text{——} & \text{———} \\
7 &
\end{array}$$

6º Bájese el siguiente periodo (52) pero si éste ya está después del punto decimal, póngase un punto decimal en la raíz (25.).

7º Continúese la operación como si se tratara del (cuarto paso) hasta haber bajado el último periodo.

Equivalencia, igualdades y ecuaciones

El signo igual (=) sirve para establecer equivalencias, igualdades y ecuaciones. Estas formas de relación son —básicamente— las mismas, pero a veces conviene establecer —con fines prácticos— alguna diferenciación entre ellas.

Una equivalencia expresa que las magnitudes colocadas antes y después del signo igual valen lo mismo, aunque estén expresadas en diferentes unidades: así, 1 dólar = 12.5 pesos, expresa que 1 dólar estadunidense (en el momento en que se escribe esto) es igual a 12.50 pesos mexicanos. O sea, que hay una relación de equivalencia entre estas cantidades.

Las igualdades y las ecuaciones expresan, en lo fundamental, que la serie de operaciones que se ordena ejecutar antes de escribir el signo "igual" deben producir el mismo resultado que la serie de operaciones que se indica hay que realizar después del signo. Entre igualdades y ecuaciones hay una diferencia que consiste en que mientras la igualdad expresa una equivalencia de resultados sólo en términos conocidos (no hay magnitudes desconocidas), la ecuación expresa esa equivalencia de resultados en términos de cantidades conocidas y cantidades desconocidas.

Elementos de una ecuación

A las cantidades conocidas que intervienen en una ecuación se les llama "datos"; a las desconocidas, "incógnitas".

Una ecuación equivale fundamentalmente a una pregunta sobre cuál deberá ser el valor de la cantidad o de las cantidades desconocidas capaz de hacer que la expresión deje de ser ecuación para convertirse en igualdad.

Así: $8 \times 5 = 30 + x$ es una ecuación que equivale a la pregunta: "¿Qué cantidad (x) es necesario agregar (+) a 30 para obtener el mismo resultado (40) que se obtiene al multiplicar 8 por 5?". Cuando encontramos que 10 es el valor que responde a esa pregunta, decimos que "hemos encontrado el valor de la incógnita". Si ponemos ese valor en lugar de x, tenemos $8 \times 5 = 30 + 10$. Ésta ya es una igualdad, porque el resultado de multiplicar 8 por 5 es igual al que se obtiene al sumar 30 por 10.

Los dos elementos básicos de una igualdad o de una ecuación son sus dos miembros. "Miembro" de una ecuación o una igualdad es cada una de las expresiones, simples o complejas, que figuran antes y después del signo igual; a la que aparece antes se le llama primer miembro y a la que aparece después de dicho signo, segundo miembro. En el ejemplo anterior, 8×5 constituye el primer miembro y $30 + 10$ el segundo miembro de la igualdad; 8×5 el primer miembro de la ecuación y $30 + x$ el segundo miembro de dicha ecuación.

En el caso de las ecuaciones, se puede distinguir a los miembros por su número de orden (primer miembro, segundo miembro) pero conviene más hacerlo mediante la expresión "miembro de la incógnita" y "miembro en que no figura la incógnita". Aunque el "miembro de la incógnita" puede serlo el primero o el segundo, conviene, por lo general, que se considere como tal al primero.

La solución de una ecuación consiste, —precisamente— en hacer que la incógnita quede sola en un solo miembro, y que todos los datos queden en el otro. El proceso para lograr esto se conoce como "resolución de la ecuación", y conduce, como paso final, a lo que se denomina "despejar a la incógnita".

Tipos de ecuación

Hay diferentes tipos de ecuación, pero, básicamente, se puede distinguir entre: 1) ecuaciones en las que la incógnita aparece elevada a un solo exponente, y 2) ecuaciones en las que la incógnita aparece elevada a diferentes exponentes. El primer tipo de ecuaciones es el de las que requieren soluciones más simples; las segundas, en cambio, imponen soluciones que no sólo son más complejas, sino que son también más particulares; que están más próximas del "truco" matemático.

Resolución de ecuaciones en las que la incógnita aparece elevada a un solo exponente

Cuando en una ecuación la incógnita sólo está elevada a un exponente, el proceso genérico consiste en pasar al miembro que no es de la incógnita (o dejar en él, si ya están ahí) todos los términos, factores, exponentes, asociados con la incógnita. En forma analítica: 1) se deja en el miembro de la incógnita (convencionalmente, el primero) todo lo que la contiene; 2) se pasa del miembro que no es de la incógnita (del segundo) al de la incógnita (primero) todo lo que, estando en ese miembro, contiene a la incógnita; 3) se pasa del miembro de la incógnita (primero) al que no es de la incógnita, cuando no la contiene, y 4) se deja en el miembro que no es de la incógnita (segundo) lo que ya está en él y no contiene a la incógnita.

El orden en que se pasan las diferentes expresiones matemáticas de un miembro a otro es, a su vez, el siguiente: 1) primero, se cambian a su lugar correspondiente *los términos* (o sean las expresiones separadas por los signos más o menos) según contengan o no contengan a la incógnita; 2) se trasladan al miembro que les corresponda, a los *factores*, y 3) se cambian de miembros los *exponentes.*

Antes de iniciar los cambios de miembro, se puede hacer, en cada uno de ellos, la reducción de términos semejantes. De este modo, en las ecuaciones de este tipo quedan, en el primer miembro, un término en la potencia respectiva de la incógnita y un término independiente y, en el segundo, otro término de esa misma potencia de la incógnita y otro término independiente. O sea, que el proceso se inicia con una expresión del tipo $ax^n + b = cx^n + d$.

Con una expresión como la anterior, el primer paso consiste en cambiar al primer miembro (miembro de la incógnita) cx^n que figura en el segundo: $ax^n + b - cx^n = d$; el segundo, en cambiar b al segundo miembro (miembro que no es de la incógnita): $ax^n - cx^n = d - b$. El paso siguiente es expresar el primer miembro como el producto de x^n por el factor $a - c$: $(a - c)\, x^n = d - b$. Como se han pasado términos, se pueden pasar de un miembro a otros factores: el factor $(a - c)$ de x^n puede pasar del miembro de la incógnita al que no es de la incógnita como divisor: $x^n = \dfrac{d - b}{a - c}$. Finalmente, el exponente n que figura en el miembro de la incógnita puede pasar al miembro que no es de la incóg nita como índice de una raíz. Así se tiene la solución final.

$$x = \sqrt[n]{\frac{d - b}{a - c}}$$

Resolución de ecuaciones en las que figuran varias potencias enteras de la incógnita

En lo que se refiere a la resolución de ecuaciones en las que figuran varias potencias enteras de la incógnita, sólo hablaremos de los casos más sencillos (y de uso estadístico más frecuente).

Como en el caso anterior, el paso previo debe consistir en la reducción de términos semejantes en cada uno de los miembros de la ecuación. Una vez hecho esto: 1) se pasan términos de un miembro a otro; 2) se reduce a la unidad el coeficiente de la potencia máxima; 3) se resuelve la ecuación resultante, según el orden de la ecuación.

La reducción del coeficiente de la potencia máxima a la unidad (o sea, el segundo paso) se logra dividiendo todos los coeficientes de la incógnita y los términos independientes entre el coeficiente de la potencia máxima de la incógnita.

Así, si tras la reducción de términos y tras haber concentrado en un miembro todos los que contenían a la incógnita y en el otro todos los que no la contenían, se obtuvo:

$$ax^n + bx^{n-1} + cx^{n-2} = d$$

el coeficiente de x^n se reduce a la unidad si tanto los coeficientes de las ponencias de x (b, c) como el término independiente, se dividen entre a:

$$x^n + \frac{b}{a}x^{n-1} + \frac{c}{a}x^{n-2} = \frac{d}{a}$$

La resolución de acuerdo con el orden de la ecuación implica, a veces, la utilización de métodos particulares. Sin embargo, se puede plantear un método genérico cuando se trata de una ecuación de segundo grado (y, hasta cierto punto, cuando se opera con una de tercer grado). Ese método consiste, fundamentalmente, en: 4) reducir a cero el coeficiente de la potencia inmediatamente inferior a la máxima (primera potencia de la incógnita en la ecuación de segundo grado; segunda potencia, en la de tercero).

Solución de una ecuación de segundo grado por el método de los polinomios equivalentes

Supondremos que ya se han dado los pasos previos y que la ecuación aparece en la forma:

$$a_2x^2 + a_1\,x = -\,a_0$$

Para resolver la ecuación, procederemos como sigue: 1) reduciremos a la unidad el coeficiente del cuadrado de la incógnita; 2) anularemos el coeficiente de la primera potencia, transformando nuestro polinomio original en un polinomio equivalente en el que la nueva incógnita difiera de la anterior en una constante; 3) resolveremos la ecuación cuadrática simple resultante (por el procedimiento delineado anteriormente).

Para reducir a la unidad el coeficiente a^2 de x^2 en la ecuación $a_2x^2 + a_1x = -\,a_0$, dividiremos todos los coeficientes (a_2, a_1, a_0) entre a_2. Si los cocientes de esta división los representamos por 1 (a_2 dividido entre a_2) y por las mayúsculas correspondientes (A_1, A_0), tendremos como ecuación resultante: $x^2 + A_1x = -\,A_0$.

Para anular el coeficiente de la primera potencia de la incógnita, comenzaremos: 1) por transformar la ecuación en una equivalente que sea de su mismo grado; cuya incógnita difiera de la suya en D, y 2) en seguida, anularemos el coeficiente de la primera potencia de la incógnita.

Al sustituir x (equis minúscula) por X (equis mayúscula) más D ($X + D$), se obtiene: $X^2 + 2DX + D^2 + A_1X + A_1D = -\,A_0$. Al pasar al segundo miembro todos los términos que no contienen a la incógnita, se tiene: $X^2 + 2DX + A_1X = -\,D^2 - A_1D - A_0$. O bien: $X^2 + (2D + A_1)\,X = -\,D^2 - A_1D - A_0$.

El coeficiente de la primera potencia de X puede anularse haciendo $2D + A_1 = 0$, o sea, $D = -\,A_1/2$. Al hacerlo, se obtiene: $X^2 = -\,D^2 - A_1D - A_0$.

Para resolver esta ecuación, bastará con: 1) sustituir el valor de D, y 2) extraer la raíz cuadrada del resultado.

Al sustituir el valor de D, se tiene: $X^2 = \frac{-\,A^2_1}{4} + \frac{A^2_1}{2} - A_0$. Esto equivale a: $X^2 = \frac{A^2_1 - 4\,A_0}{4}$.

Para obtener el valor de X (paso 3, resolutorio de la ecuación) bastará extraer la raíz cuadrada del segundo miembro:

$$X = \frac{\sqrt{A^2_1 - 4\,A_0}}{2}$$

Para obtener el valor de x, bastará recordar que x (equis minúscula) es igual a X (mayúscula) más D; que X es igual a la expresión recién encontrada y que D es igual a $-A_1/2$. Esto da, como fórmula final:

$$x = \frac{A_1 \pm \sqrt{A^2_1 - 4A_0}}{2}$$

Si se quiere expresar todo esto en términos de los coeficientes originales, bastará sustituir las Aes mayúsculas por sus equivalentes en términos de las aes minúsculas. Así, puede escribirse:

$$x = \frac{a_1 \pm \sqrt{a^2_1 - 4a_2a_0}}{2a_2}$$

Por supuesto, en la solución de una ecuación en la que los coeficientes son números, esta fórmula se puede aplicar directamente, pero también se puede seguir el procedimiento. Lo usual, lo rutinario, es aplicar la fórmula. Lo aconsejable (desde un punto de vista de la formación mental del estudiante, en sentido metodológico) es seguir todo el proceso. Este procedimiento se puede delinear como sigue:

Procedimiento de resolución de la ecuación completa de segundo grado:

1º Redúzcase el coeficiente de x cuadrada a la unidad por división de todos los coeficientes (incluso el término independiente o coeficiente de x a la cero) entre el coeficiente que afecta a x cuadrada,

2º Obténgase una ecuación equivalente, del mismo grado, en X (que difiera de x en D) y anúlese en la ecuación resultante el coeficiente de X.

3º Obténgase, de la ecuación resultante, el valor de X.

4º A partir del valor de X y del de D, obténgase el de x.

Discusión de la ecuación de segundo grado

Una ecuación completa de segundo grado se puede representar por:

$$a_0 + a_1 x + a_2 x^2 = 0$$

La fórmula para obtener las raíces de esta ecuación es:

$$x_{1,2} = \frac{-a_1 \pm \sqrt{a^2_1 - 4a_0a_2}}{2a_2}$$

A la cantidad subradical de esta expresión se le da el nombre de "discriminante". Según las variaciones que sufra el discriminante, se obtendrán diferentes tipos de valor para las raíces de la ecuación. Así,

1. Si $a^2_1 - 4\,a_0a_2 = 0$, o sea, si $a^2_1 = 4\,a_2a_0$, las raíces serán: $x_1 = \frac{-a^1}{2a_2}$ y $x_2 = \frac{-a_1}{2a_2}$ puesto que, en la fórmula original habrá desaparecido el discriminante (aditivo para la primera raíz, substractivo para la segunda).

2. Si $(a^2_1 - 4\,a_0a_2)$ es mayor que cero, o sea, si su valor es positivo; o bien, si a^2_1 mayor que $4\,a_0a_2$, las raíces estarán formadas por dos fracciones, cuyo denominador será $2a_2$, y cuyos numeradores serán dos binomios, formados por una cantidad real $(-a_1)$ y por la raíz cuadrada de otra cantidad real (el subradical). O sea, que ambos numeradores y ambas fracciones serán reales (en cuanto formados por dos términos igualmente reales).

Por otra parte, en cuanto el segundo término del binomio numerador debe agregarse a $-a_1$ para obtener la primera raíz, y substraerse de dicha cantidad para obtener la segunda, dichas raíces serán desiguales (a diferencia del caso anterior, en que las dos raíces eran reales, pero iguales).

3. Si $a^2_1 - 4\,a_0a_2$ es menor que cero, o sea, si su valor es negativo; o bien, si a^2_1 es menor que $4\,a_0a_2$, las raíces estarán formadas por una fracción cuyo denominador será $2a_2$ y cuyos numeradores serán binomios constituidos por una parte real $(-a_1)$ y por la raíz cuadrada de una cantidad negativa (el subradical) o sea, por una segunda parte, imaginaria. En suma, que los numeradores no serán números reales sino números complejos. Las dos raíces serán diferentes; más concretamente, serán dos complejos conjugados por tener iguales sus partes reales y ser simétricas (o estar afectadas por signos opuestos, más y menos) sus partes imaginarias.

Aproximación de esta discusión a las necesidades estadísticas.

Las condiciones de aparición de los diferentes tipos de raíz de la ecuación de segundo grado se pueden expresar en forma ligeramente diferente. Esta otra forma de expresión es la más usual en estadística: registra el criterio discriminatorio en forma fraccionaria. De este modo:

$$\text{Si } (a^2_1)\ /\ (4a_0a_2) = 1$$

esto significa que numerador y denominador son iguales; que la resta del numerador menos el denominador vale cero, o bien, que la ecuación de segundo grado tendrá, para este *criterio unitario, dos raíces reales e iguales.*

Si $(a^2_1)\ /\ (4a_0a_2)$ es mayor que 1, eso significa que el numerador es mayor que el denominador; que la resta del numerador menos el denominador es positiva, o bien, que la ecuación de segundo grado tendrá, para este *criterio fraccionario, dos raíces reales y desiguales.*

Si (a^2_1) / $(4a_0a_2)$ es menor que 1, eso significa que el numerador es menor que el denominador; que la resta de numerador y denominador es negativa, o bien, que la ecuación de segundo grado tendrá, para este *criterio fraccionario, dos raíces complejas, conjugadas.*

Procedimiento para la solución de la ecuación completa de tercer grado

La ecuación completa de tercer grado es de la forma:

$$a_3x^3 + a_2x^2 + a_1x = - a_0.$$

Para resolverla, procederemos sobre líneas muy parecidas a las que se siguieron en la solución de la ecuación completa de segundo grado; o sea: 1) se reduce a la unidad el coeficiente de la tercera potencia de la incógnita; 2) se anula el coeficiente de la segunda potencia de la incógnita tal y como aparece en una ecuación equivalente en la que la nueva incógnita difiere de la primera en cierta cantidad, y 3) se resuelve la ecuación resultante, mediante un artificio propio de las ecuaciones de tercer grado.

La reducción del coeficiente del cubo de equis minúscula se logra, como en el caso anterior, mediante división de todos los coeficientes entre el coeficiente de equis al cubo. La formación del polinomio equivalente, se realiza al sustituir a equis minúscula (x) por equis mayúscula más "de mayúscula" (D): $(x = X + D)$.

La solución propia de la ecuación de tercer grado consiste en lo siguiente: 1) obtener las raíces cúbicas de una incógnita auxiliar (ξ), y 2) sustituir el valor de esa incógnita auxiliar en la incógnita originaria.

Se forma una ecuación cuadrática con una incógnita auxiliar que designaremos por la letra griega ksi (ξ); o sea, que en esta ecuación habrá un término en ksi cuadrada, otro en ksi a la primera y uno independiente de ksi. En esa ecuación, los coeficientes serán: 1.1) la unidad para el cuadrado de ksi; 1.2) el término independiente de la ecuación en equis mayúscula (alfa cero) para la primera potencia de ksi; 1.3) el cubo de un tercio del coeficiente de equis (o sea alfa uno sobre 3, elevado todo al cubo) para la potencia cero (o término independiente) de esta ecuación cuadrática auxiliar. Al resolver esta ecuación auxiliar, se obtiene para ksi un valor, cuyas tres raíces cúbicas se extraen.

Obtenido el valor de las raíces cúbicas de ksi: 2.1) se suman, por pares, para obtener el valor de equis mayúscula; 2.2) se sustituye el valor de X (equis mayúscula) en el de x (equis minúscula), y así queda resuelta la ecuación correspondiente.

Resolución de un sistema de ecuaciones

Cuando en una ecuación existe una sola incógnita se puede obtener de modo inequívoco el valor de ésta que satisface a la ecuación. Cuando en *una* sola ecuación existen *dos* incógnitas, no se puede precisar cuál es *el* valor de cada una, pues hay muchas combinaciones de valores que satisfacen a la ecuación. Así, si preguntamos ¿qué par de enteros, sumados, dan 4?, podremos representar la pregunta por la ecuación $x + y = 4$, y encontrar que "x" puede ser igual a 0 y "y" igual a 4 $(0 + 4 = 4)$, o "x" igual a 1 y "y" igual a 3 $(1 + 3 = 4)$ o "x" igual a 2 y "y" igual a 2 $(2 + 2 = 4)$ y así sucesivamente. Pero, 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. Eso puede verse por la tabulación siguiente:

Cuando x es igual a	y es igual a	x + y es igual a	2x es igual a	3y es igual a	y 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, se puede ver 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 4, no todos 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 "x" igual a 3 y "y" igual a 1 cumple *simultáneamente*, la condición de que siendo enteros, sumados den 4, y la otra, de que el duplo del primero y el triple del segundo, sumados den 9. En forma parecida, se podría hacer otra tabulación de valores que satisfacieran la segunda condición; se vería entonces que, a pesar de ello, sólo un par satisfaría *simultáneamente* la primera. La solución coincidiría con la que hemos encontrado aquí.

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 dos 'ecuaciones' (o sea, expresiones en que figuran cantidades desconocidas y conocidas) 'simultáneas' (en cuanto los valores de las incógnitas deben satisfacer a ambas) con dos 'incógnitas' (los dos valores desconocidos x y y que aparecen en ambas)".

Como el método de tanteo representado por las tabulaciones anteriores no podría seguirse en todos los casos, por embarazoso (piénsese, por ejemplo, en el problema de buscar diez sumandos que produzcan 245 y cuyos duplos sumados den 2 600) se recurre a otros procedimientos. De ellos, uno de los más generales es el de los determinantes (que será el único al que nos referiremos).

Resolución de un sistema de dos ecuaciones con dos incógnitas por medio de determinantes

Un sistema de dos ecuaciones con dos incógnitas se puede resolver mediante el uso de "determinantes" de segundo orden.

¿Qué es un determinante?

Un determinante de segundo orden es: 1) un binomio (expresión algebraica constituida por dos términos) cada uno de cuyos términos está constituido por 1.1 y 1.2) dos productos de dos factores, y que están separados (los productos) por, 2) el signo menos. Es decir, se trata de dos productos, ab y cd que forman el binomio ab-cd:

$$\underbrace{ab}_{1.1} \quad \underbrace{-}_{2} \quad \underbrace{cd}_{1.2}$$

Convencionalmente, esta expresión se escribe colocando las cuatro cantidades que en ella intervienen (a, b, c, d) en los extremos de las dos diagonales de un cuadrado (a y b en la diagional principal y c y d en la diagonal secundaria, encerrando todo entre dos líneas verticales:

$$\begin{vmatrix} a & d \\ c & b \end{vmatrix}$$

Según esto, el valor de un determinante de segundo orden se calcula formando los productos de las cuatro cantidades, en cruz, y restando del producto ab el producto cd.

Ejemplo: Si se trata de obtener el valor de

$$\begin{vmatrix} 5 & 3 \\ 2 & 4 \end{vmatrix}$$

multiplicaremos 5 por 4 (20), y a este producto le restaremos 6 (producto de 2 por 3). O sea, que el resultado final será $20 - 6 = 14$.

Uso de los determinantes de segundo orden en la resolución de un sistema de dos ecuaciones con dos incógnitas

El procedimiento de los "determinantes" consiste en formar sendas fracciones para las dos incógnitas. El numerador y el denominador de cada fracción son determinantes de segundo orden. El denominador de ambas (común a las dos) es un mismo determinante, al que se le llama "eliminante del sistema".

Sea el sistema:

$$3a + 5b = 42$$
$$7a + 9b = 82$$

Eliminante del sistema

El eliminante del sistema se obtiene si se sacan los coeficientes de las incógnitas (coeficientes de a: 3 y 7; coeficientes de b: 5 y 9) y se colocan en dos columnas y dos renglones en el orden en que aparecen en las ecuaciones. En el ejemplo, el resultado es:

$$\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix} = 3 \times 9 - 5 \times 7 = 27 - 35 = 8$$

Éste es el eliminante en el ejemplo.

El eliminante, de acuerdo con su nombre, nos permite eliminar una de las incógnitas, o si, se prefiere, nos permite encontrar directamente los valores de ambas, mediante fracciones como las siguientes:

$$a = \frac{\text{numerador I}}{\text{eliminante}} \text{ y } b = \frac{\text{numerador II}}{\text{eliminante}}$$

Numeradores de las incógnitas

Los numeradores (I y II) de estas fracciones, se encuentran a partir del eliminante; en él, los coeficientes de la incógnita cuyo valor se trata de determinar, se sustituyen por los términos independientes (42, 82) tomados en el orden en que aparecen en el sistema de ecuaciones.

Así, en el caso de a, el numerador se forma poniendo 42 y 82 en vez de 3 y 7; en el caso de b, el numerador se forma poniendo 42 y 82 en vez de 5 y 9:

$$\text{I} = \begin{vmatrix} 42 & 5 \\ 82 & 9 \end{vmatrix} \qquad \text{II} = \begin{vmatrix} 3 & 42 \\ 7 & 82 \end{vmatrix}$$

El valor de estos dos determinantes es $42 \times 9 - 82 \times 5 = 32$ y $3 \times 82 - 7 \times 42 = 48$. En consecuencia, al dividir cada uno de los valores, de los numeradores I y II entre el eliminante (— 8), se obtiene 4 como valor de a y 6 como valor de b.

Procedimiento

Según esto, el procedimiento para resolver un sistema de dos ecuaciones con dos incógnitas, por el método de los determinantes, se puede resumir como sigue: 1) se forma un determinante de segundo grado con los coeficientes de las incógnitas (eliminante del sistema); 2.1) se sustituyen los coeficientes de la primera incógnita, que figuran en el eliminante, por los términos independientes (los que no contienen a la incógnita; 2.2) se divide el determinante así formado entre el eliminante, y, con ello, se obtiene el valor de la primera incógnita; 2.3) se sustituyen, en el eliminante, los coeficientes de la segunda incógnita por los términos independientes, y 3.2) se divide el valor del determinante así formado entre el del eliminante, para obtener el valor de la segunda incógnita.

Simplificación

El procedimiento se puede simplificar mediante la eliminación de los pasos 2.2 y 3.2 que corresponden a la determinación directa de la segunda incógnita. Para ello, se sustituye el valor de la primera incógnita en una de las ecuaciones, y se resuelve para la segunda incógnita. Esta simplificación tiene la desventaja de que hace que los valores obtenidos no sean independientes entre sí sino dependan uno de otro; de ese modo, si hay un error en la obtención de la primera incógnita éste repercutirá en la obtención de la segunda; en cambio cuando las incógnitas se despejan por separado, un error de cálculo de la primera no repercute en la segunda.

Resolución de un sistema de tres ecuaciones con tres incógnitas

La resolución de un sistema de tres ecuaciones con tres incógnitas sigue los mismos lineamientos que se siguieron en la solución de un sistema de dos ecuaciones con dos incógnitas, pero, en este caso, se trabaja con eliminantes de tercer orden. Según esto: 1) se forma el eliminante del sistema, tomando los tres renglones y las tres columnas de coeficientes de las tres incógnitas, en su orden; 2) para cada incógnita, se forma otro determinante de tercer orden, sustituyendo en el eliminante sus coeficientes por los términos independientes, y 3) se obtiene el valor de cada una de ellas dividiendo el valor del determinante así formado entre el del eliminante.

Procedimiento de los cofactores, para el cálculo de los determinantes

En la solución de sistemas de ecuaciones de tres o más incógnitas, se utilizan, en forma correspondiente, —como ya dijimos— determinantes de orden superior al segundo. Para encontrar el valor de estos determinantes se utiliza el método llamado de los menores o de los cofactores. Para explicar este procedimiento, comenzaremos por definir qué es un menor y qué un cofactor.

Menor

Menor de un determinante (de orden n) es el determinante de orden inmediato inferior (n-1) que se obtiene al eliminar una columna y un renglón del determinante primitivo.

De este modo, un menor (hay nueve en total) de un determinante de tercer orden, será el determinante de segundo orden (n-1) que se obtenga de eliminar una de las tres columnas (cualquiera, la primera, la segunda o la tercera) y uno de los tres renglones (el que sea).

Si tenemos el determinante de tercer orden:

$$\begin{vmatrix} 5 & 1 & 8 \\ 4 & 6 & 3 \\ 9 & 2 & 7 \end{vmatrix}$$

y en él suprimimos la primera columna (5, 4, 9) y el primer renglón (5, 1, 8) obtendremos el menor que corresponde al primer elemento (de la primera columna y del primer renglón):

$$\begin{vmatrix} 6 & 3 \\ 2 & 7 \end{vmatrix}$$

En la misma forma se puede obtener el menor del elemento central (6) eliminando la columna central (1, 6, 2) y el renglón central (4, 6, 3), con lo que se tendrá un determinante de segundo orden, que tendrá como primer renglón 5, 8 y por segundo 9, 7.

Cofactor

Cofactor de un determinante de orden n es el menor correspondiente, multiplicado por — 1 (menos uno) elevado a una potencia igual a la suma del número de orden de la columna y el número de orden del renglón suprimido.

También puede decirse que el "cofactor" de un determinante es el "menor" correspondiente o el simétrico de dicho menor (el menor con signo contrario) según que el número de orden de columna y renglón sea número par o impar.

En el ejemplo anterior, como en el primer menor, se suprimió la columna 1 y el renglón 1, habrá que elevar (menos 1) a la segunda potencia (1 + 1), lo cual produce más 1, que multiplicado por el menor correspondiente, lo deja intacto. O sea, que, en este caso, el cofactor es igual al menor, y del mismo signo.

En cambio si se obtiene el menor de la primera columna y el segundo renglón, la suma da $(1 + 2 = 3)$; hay que elevar menos uno a una potencia impar, que da como resultado un número negativo; o sea, que en este caso el cofactor es igual a su menor, pero de signo contrario.

Definido lo que se entiende por menor y por cofactor, puede delinearse el procedimiento que sigue para el cálculo de un determinante: 1) se elige una columna (o un renglón) del determinante; 2) se calculan los cofactores de los elementos de esa columna (o renglón); 3) se multiplica cada elemento de la columna elegida (o del renglón) por su cofactor, y 4) se suman algebraicamente los resultados.

Si se elige 1) la primera columna, el procedimiento puede mecanizarse como sigue; 2) se calculan los menores de esa columna; 3.1) se multiplica el primer elemento por su menor; 3.2) el segundo por el simétrico de su menor (el menor con signo cambiado), y 3.3) el tercer elemento por su menor; 4) sumando, al final, los resultados.

Condensación de los datos

Condensación sumatoria

Lectura y significación de las sumatorias

Uno de los operadores (o signos que indican operación) que se emplean más en estadística es el operador sigma mayúscula (Σ) que indica que deben sumarse todos los términos del género o tipo del colocado delante de él. A la expresión afectada por dicho operador se le llama "sumatoria". De este modo:

Σx_i indica: "súmense todos los valores que puedan representarse mediante la sustitución de i por los números naturales, en la expresión x_i", o bien, "súmese todas las equis", o, finalmente, "súmese $x_1 + x_2 + x_3 + x_4 + \ldots x_n$".

Si x_i representa las estaturas de los individuos de una población

Σx_i indica: "súmense las estaturas de todos los individuos de la población".

Limitación del dominio del operador

En ocasiones la suma no se extiende a todo el universo o población; en tales ocasiones se especifica cuáles han de ser los límites entre los cuales se ha de aplicar el sumador (u operador sigma). Para ello, se coloca debajo del operador sigma mayúscula (Σ) el número de orden del valor de la variable que ha de servir como primer sumando, y arriba del operador, el número de orden del valor que ha de servir como último término de la suma. Así:

$\sum_{i=3}^{5} x_i$ significa, en el ejemplo anterior: "súmese las estaturas de los individuos marcados con los números de orden 3, 4, y 5".

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

Reglas de operación del sumador

En el uso del operador sigma mayúscula se deben considerar 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.

Regla sobre las constantes afectadas por el sumador

Cuando el operador sigma mayúscula afecta a una constante, la sumatoria correspondiente se puede sustituir por el producto de la constante multiplicada por el número de veces que aparezca dicha constante.

La regla anterior se explica porque sumar una serie de cantidades constantes, iguales, equivale a multiplicar la constante por el número de veces que aparece como término de la suma. Sumar $5 + 5 + 5$ (tres veces) equivale a multiplicar 5×3. O sea, si empleamos el sumador:

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

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

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

Regla sobre los coeficientes de expresiones sumatorias

Si los diferentes valores de una variable/ están multiplicados por una misma constante (coeficiente)/ y sus productos se suman/, dicha suma/ será igual al resultado que se obtendría si/ primero, se sumaran los valores de la variable, y/ después/, se multiplicara la suma por la constante/.

	Diferentes valores de la variable	*Esos mismos valores multiplicados por la constante $a = 2$*
	x_i	ax_i
	$x_1 = 14$	$ax_1 = 2 \times 14 = 28$
	$x_2 = 11$	$ax_2 = 2 \times 11 = 22$
	$x_3 = 10$	$ax_3 = 2 \times 10 = 20$
	$\Sigma x_i = 35$	$\Sigma ax_i = 70$
	Suma de los valores de la variable	Suma de los productos de la variable por su coeficiente

Es fácil comprobar que: $\Sigma\, ax_i = 70 = 2 \times 35 = 2\, \Sigma\, x_i = a\, \Sigma\, x_i$

Que es lo que establece la regla. En forma práctica, la regla se puede expresar diciendo que: todo factor constante puede entrar o salir del sumador (Σ) sin modificarse ni alterar la sumatoria.

Precaución en el uso del sumador aplicado al producto de variables.

Valores de la primer variable	*Valores de la segunda*	*Producto de ambas*
x_i	f_i	$x_i f_i$
$x_1 = 1.40$	$f_1 = 14$	$x_1 f_1 = 19.60$
$x_2 = 1.20$	$f_2 = 11$	$x_2 f_2 = 13.20$
$x_3 = 1.10$	$f_3 = 10$	$x_3 f_3 = 11.00$
$\Sigma x_i = 3.70$	$\Sigma f_i = 35$	$\Sigma x_i f_i = 33.80$
Suma de los valores de la primer variable	Suma de los valores de la segunda	Suma de los productos de la primer variable por la segunda
Producto de las sumas de cada variable hechas separadamente		
Σx_i por $\Sigma f_i = 3.70 \times 35 = 129.50$		

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

Como puede verse 129.50 (producto que se obtiene de multiplicar las sumas) no es igual a 33.80 (producto que se obtiene de sumar los productos). En forma condensada se puede decir que: la suma de los productos de dos o más variables *no es igual* al producto de las sumas de esas variables:

$$\Sigma\ x_i f_i \neq \Sigma\ x_i \text{ por } \Sigma\ f_i$$

¿Cuál de los dos resultados es correcto? Cualquiera de ellos puede serlo, según la operación que se requiera. Hay ocasiones en que se necesita la suma de los productos de las variables; hay otras en que se utiliza el producto de las sumas de las variables. En cada caso, hay que tomar en cuenta el *orden* en que deben realizarse las operaciones.

Si se necesita la suma de los productos: 1) primero se obtienen los productos, y 2) después, se suman.

Si se necesita el producto de las sumas: 1) primero se obtienen las sumas, y 2) después, se multiplican esas sumas.

Un análisis lingüístico puede ayudar a recordar el orden, pues primero se realiza una operación múltiple (*varios* productos *o varias* sumas) y después una unificadora (*una* suma o *un* producto).

Precaución genérica al usar el sumador en operaciones en que intervienen dos o más operaciones con variables.

Si los diferentes valores de una variable (las x_i) se someten a cierta operación (se multiplican por los de otra variable, se elevan a potencias, se logaritman o exponencian) y después se suman los resultados, el valor final obtenido *no es igual* al resultado que se obtendría sumando primero los diferentes valores de la o de las variables y sometiendo después a esa suma o a esas sumas a la operación de que se trate (multiplicándolas, elevando la suma a la ponencia de que se trate, logaritmando esa suma o exponenciándola). En esos casos, hay que estar plenamente seguro de que se sabe cuál es el resultado que se quiere obtener, para realizar las diferentes operaciones que haya que realizar *en el orden debido,* pues esto es vital para el acierto en los resultados.

Definición de derivada

La derivada de una función es el límite de la relación entre el incremento de esa función y el incremento correspondiente de la variable, cuando el incremento de la variable tiende hacia cero.

Lo anterior puede expresarse, en forma analítica y simbólica, del modo siguiente:

Derivada	D
de una función	f ( )
de la variable x	x
es el límite	lim
de la relación	——
	(línea de fracción)
entre el incremento (Δ) de la función f(x)	Δ f(x)
y el incremento de la variable	Δ x
cuando el incremento de la variable Δ x tiende a cero	$\Delta x \rightarrow 0$

Reunidos todos estos elementos en una sola fórmula, se tiene:

$$D\,f\,(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta f\,(x)}{\Delta x}$$

Representaciones convencionales de la derivada

Tres son las notaciones más comúnmente empleadas para representar una derivada. Se deben a Cauchy, a Lagrange y a Leibinitz. Éstas son:

$D\,f\,(x)$ o $\delta\,f\,(x)$	notación de Cauchy,
$f'\,(x)$	notación de Lagrange,
$d\,f\,(x)/dx$	notación de Leibinitz.

Etapas para el cálculo directo de una derivada

La definición misma de derivada —o su fórmula— indican suficientemente cuáles tienen que ser los pasos que hay que dar para calcular en forma directa una derivada. Éstos son:

1. Tomar la variable x e incrementarla sumándole al valor inicial (x) el incremento (Δ x). $\quad x + \Delta x$
2. Determinar el incremento correspondiente de la función, para lo cual:
   2.1. Se restará de la función incrementada $\quad f\,(x + \Delta x)$
   que se obtiene al sustituir x por $x + \Delta x$ y ejecutar las operaciones indicadas) ...

2.2. ... la función no incrementada. $f(x)$

$f(x + \Delta x) - f(x) = \Delta f(x)$

3. Dividir el incremento de la función $\Delta f(x)$ entre el incremento de la variable Δx. $\Delta f(x) / \Delta x$
4. Hacer que tienda a cero el incremento de la variable. $\Delta x \to 0$
5. Determinar hacia qué límite tiende, entonces, el cociente de los incrementos. $\Delta f(x) / \Delta x \to$ ¿?
6. El resultado (la respuesta) es la derivada de la función.

Derivada de una función afectada por constantes cuando se conoce la derivada de la función desnuda

Si se conoce cuál es la derivada de una función $f(x)$, es posible determinar fácilmente cuál es la derivada: de $f(x)$ adicionada de una constante; de $f(x)$ multiplicada por una constante; de $f(x)$ elevada a una constante. Para encontrar las reglas correspondientes conviene, sin embargo, comenzar por explorar (mediante aplicación directa del procedimiento anterior, cuál es la derivada de una suma de funciones).

Derivada de la suma de dos funciones

Si representamos por u y por v dos funciones de x, distintas entre sí y su suma la representamos por y (que, asimismo será función de x):

$$y = u + v$$

Para encontrar la derivada de y con respecto a x (variable común a las funciones-sumando u y v, y a la función-suma y) seguiremos los pasos del procedimiento indicado anteriormente.

1. Al incrementar la variable x, la función u, tendrá un incremento Δu que produce $u + \Delta u$
   la función v, tendrá un incremento Δv, que produce $v + \Delta v$
   y la función y resultará incrementada en Δy $y + \Delta y$
2. El valor del incremento de la función-suma resultará ser, así, lo que se obtenga de restar de la función incrementada $(u + \Delta u) + (v + \Delta v) = y + \Delta y$ la función no incrementada $u + v = y$ lo cual produce: y

$$\Delta u + \Delta v = \Delta y$$

3. Al dividir el incremento de la función Δy entre el incremento de la variable Δx se ejecuta una operación que equivale a dividir el equivalente de Δy (igual a $\Delta u + \Delta v$) entre el incremento de x (Δx)

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u + \Delta v}{\Delta x}$$

o bien:

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta x}$$

4. Al hacer que x tienda hacia cero, en cada una de estas relaciones, se obtiene un límite para cada una de ellas:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} + \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x}$$

5. Cuál es el límite de la relación del incremento de y y el incremento de x cuando el incremento de x tiende a cero?
De acuerdo con la expresión anterior:
El límite del incremento de la función-suma (y) entre el incremento de la variable (x), cuando este último incremento tiende hacia cero es igual: a la suma de los dos límites siguientes:
a) el límite de la relación entre el incremento de la función que constituye el primer sumando (u) y el incremento de la variable (x) cuando el incremento de la variable tiende hacia cero, y
b) el límite de la relación entre el incremento de la función que constituye el segundo sumando (v) y el incremento de la variable (x) cuando el incremento de la variable tiende hacia cero.
Los dos límites que forman el resultado son, de acuerdo con la definición:
a) la derivada del primer sumando (u) respecto de la variable x, y $D_x(u)$
b) la derivada del segundo sumando (v) respecto de esa variable común, (x) $D_x(v)$

Según esto:

6. La derivada de la suma de las funciones (y) respecto a x es igual $D_x(y)$
a la suma de las derivadas de los sumandos u y v:

$$D_x y = D_x u + D_x v$$

Si pasamos de un miembro a otro una de las derivadas de los sumandos, obtendremos:

$$D_x y - D_x u = D_x v \text{ (corolario)}$$

Así, puede afirmarse que:

La derivada de la suma de dos (o más funciones) es igual a la suma de las derivadas de los sumandos y

(Corolario) La derivada de la diferencia de dos funciones es igual a la diferencia entre las derivadas de esas funciones.

Derivada de una función modificada por constantes

Si una función es el resultado de la adición de otra función y una constante, la regla sobre la derivada de una suma se aplica, si se la modifica convenientemente. En efecto, la derivada de la suma de una función y una constante es igual a la suma de las derivadas de la función y de la constante. Sin embargo, la derivada de una constante es nula (porque cualquier incremento de la variable no puede afectar a la constante que, consecuentemente tiene un incremento nulo) y, consecuentemente, la derivada de la suma de la función y la constante es igual a la derivada de la pura función:

Si una función es el resultado de multiplicar una función distinta de ella por un factor constante, ésta se puede considerar como una función-suma resultante de adicionar tantas funciones-sumando iguales, como indique el factor constante. De acuerdo con esto, la derivada de la función-suma será igual a la suma de tantas derivadas de la función-sumando común, como indique el factor constante. O sea, finalmente, que la derivada de la función resultante será igual a la derivada de la función, multiplicada por el factor constante.

Derivada de una función de función

Si y es función de v, y v es función de x, se dice que y es función de función de x.

Es posible demostrar que:

$$D_x y = D_v y \cdot D_x v$$

O sea que: la derivada de una función de función (y) con respecto a la variable independiente (x) es igual a: la derivada de la función de función con respecto a la variable intermediaria (v) / por la derivada de la variable o función intermediaria (v) con respecto a la variable independiente (x).

Derivada de una función inversa

Si y es función de x, x es función inversa de y.

Puede demostrarse que:

$$D_x y = D_y x)^{-1}$$

Lo cual puede expresarse como que: la derivada de una función con respecto a la variable independiente es igual al recíproco de la derivada de la variable independiente con respecto a la función.

El número e como un límite

El número e, base de los logaritmos naturales se define como el límite de la potencia de un binomio en el que uno de los términos es la unidad y el otro es el recíproco del exponente, cuando el segundo término tiende a 0:

$$e = \lim_{x \to 0} (1 + x)^{x^{-1}}$$

Derivada de un logaritmo natural

Para obtener esta derivada aplicaremos la regla o procedimiento general de derivación. Si:

$$y = L\,x$$

la variable incrementada es $x + \Delta x$; el incremento (por evaluar) de la función es Δy, y la función incrementada es:

$$y + \Delta y = L\,(x + \Delta x)$$

Si de la función incrementada se resta la no incrementada $-y = Lx$ se obtiene $\Delta y = L\ (x + \Delta x) - Lx$

La diferencia de logaritmos del segundo miembro puede expresarse como el logaritmo de un cociente:

$$\Delta y = L\ \frac{x + \Delta x}{x}$$

Los cocientes de incrementos obtenidos mediante la división de ambos miembros entre el incremento de la variable son:

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x} L \frac{x + \Delta x}{\Delta x} = \frac{1}{\Delta x} L (1 + \frac{\Delta x}{x})$$

Si se multiplica y divide el segundo miembro por x:

$$\frac{\Delta y}{\Delta x} = \frac{1}{x} \frac{x}{\Delta x} L \left(1 + \frac{\Delta x}{x}\right) = \frac{1}{x} L \left(1 + \frac{\Delta x}{x}\right)^{\frac{x}{\Delta x}}$$

Al tomar límites en el primero y en el segundo miembro, se obtiene:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{1}{x} \lim_{\Delta x \to 0} L \left(1 + \frac{x}{x}\right)^{\frac{x}{\Delta x}}$$

El límite del primer miembro es, por definición, la derivada de y con respecto a x. En cuanto al límite del paréntesis potenciado es el límite de la potencia de un binomio cuyo primer término es la unidad, y cuyo segundo término es el recíproco de la potencia, cuando el segundo término del binomio tiende a 0, y esto, según hemos indicado antes, es conocido como el número e. Según esto, la expresión anterior se convierte en:

$$D_x y = \frac{1}{x} L e = \frac{1}{x}$$

puesto que $L e = 1$ (el logaritmo de la base es siempre la unidad, ya que logaritmo de un número es el exponente (1 en el caso) al que hay que elevar a la base para obtener el número).

Derivada del logaritmo natural de una función

Si:

$$y = L\ f(x)$$

Al tomar la derivada de y con respecto a la función de x se tiene:

$$D_{f(x)} y = \frac{1}{f(x)}$$

por aplicación del principio anterior.

Como se trata de una función de función, la derivada de la función de función (y) con respecto a la variable independiente es igual al producto de la derivada de la función de función con respecto a la variable intermediaria/ f(x)/ por la derivada de la función intermediaria con respecto a la variable independiente (x).

$$D_x y = D_{f(x)} y \; D_x f(x)$$

Como en este caso, el primero de estos factores ha resultado igual al recíproco de la función de x, la substitución producirá:

$$D_x y = \frac{1}{f(x)} D_x f(x) = \frac{D\,f(x)}{f(x)}$$

Es decir que: la derivada del logaritmo de una función con respecto a la variable independiente es igual a la derivada de la función (sin logaritmizar) con respecto a dicha variable, dividida entre la función misma.

Derivada de una exponencial

Si:

$$y = a^v$$

Al tomar logaritmos: $L\,y = v\,L\,a$

Despejando a v: $v = L\,y\,/La$

Derivando respecto a y: $D_y v = 1/La\; D_y\; Ly = 1/La \cdot 1/y$

Como la función inversa, derivada con respecto a la directa, es igual al recíproco de la derivada de la directa respecto a la inversa:

$$D_v\,y = (D_y\,v)^{-1} = \left[(y\,La)^{-1}\right]^{-1} = y\;La$$

Si en esta expresión se substituye y por su valor inicial a^v, se obtiene:

$$D_v\,y = a^v L\,a$$

Es decir que: la derivada de una constante elevada a un exponente variable es igual a esa misma función exponencial multiplicada por el logaritmo natural de la constante.

En caso de que la constante sea e, base de los logaritmos naturales:

$$y = e^v$$

$$D_v y = e^v\,L\,e$$

Pero como el logaritmo natural de e (base de los logaritmos naturales) es 1:

$$D_v y = e^v$$

o sea, que en este caso, función y derivada son iguales.

Derivada de una exponencial, en términos muy generales

Si:

$$y = u^v$$

y tanto u como v son funciones de una misma variable x;

$$D_x y = vu^{v-1} D_x u + u^v L\, u\, D_x v$$

Es decir que: la derivada de una función elevada a otra función es igual a la suma: 1º de la derivada de la exponencial obtenida considerando el exponente como constante, y 2º de la derivada de la exponencial considerando la función exponenciada como si fuera constante.

Integración

Integrar es la operación inversa de derivar. Integrar una función (llamada "integrando") equivale a encontrar aquella otra función (llamada "integral") de la cual esa función es primera derivada. O sea, que el integrando es la primera derivada de la integral.

La integración se expresa tradicionalmente mediante el operador que es, en realidad, una ese latina alargada. Como el operador Σ (sigma mayúscula), el operador $\int$ (ese alongada) expresa una forma especial de suma. En efecto, gracias a los esfuerzos de Lebesgue y de Stieltjes, la integración y la sumación se consideran como especies de un proceso genérico de operación. El operador "sigma mayúscula" especifica ese proceso en aquellos casos en que se opera con distribuciones discontinuas (sumación); el operador "ese alongada" especifica ese proceso cuando se opera con distribuciones continuas.

Integración de derivadas

En cuanto la integración es la operación inversa de la derivación, los efectos del integrador y del derivador que se aplican simultáneamente a una misma función se anulan. De este modo:

$$\int_x D_x f(x) = f(x)$$

$$D_x \int_x f(x) = f(x)$$

En muchos casos, la derivación no es expresa sino tácita; o sea —en ese caso— que el operador "ese alargada" está afectando a una expresión que ha sido resultado de una derivación o a la que se le puede considerar como tal. En tal caso, si se sabe cuál es la expresión originaria de la que es resultado esta otra expresión, también se produce la anulación de efectos; es decir, que se obtiene como resultado la función primitiva de la que el integrando es primera derivada.

$$\int_x n\, x^{n-1}$$

Es fácil reconocer que el integrando es la derivada de la enésima potencia de x o sea, que esta expresión se puede escribir, también:

$$\int_x D_x\, x^n$$

Como los efectos de dos operadores inversos que se aplican simultáneamente a una misma función, se anulan, el resultado de la integración de la derivada de x a la n es:

$$x^n$$

O sea que, la integral primitiva (o sea, la de n por x a la n-1 es igual a x a la n:

$$\int_x n\, x^{n-1} = x^n$$

Integral de una suma de funciones

Si buscamos la integral de $D_x u + D_x v$, o sea:

$$\int (D_x u + D_x v)$$

Esto equivale a preguntarse lo siguiente:

¿Cuál es la función de la que $(D_x u + D_x v)$ es primera derivada? O bien, equivale a decir: si $D_x u + D_x v$ es la primera derivada de una función, ¿cuál es esa función?". Según vimos, en el estudio de las derivadas:

$$D_x u + D_x v = D_x y$$

en donde:

$$y = u + v$$

o sea que:

$$D_x u + D_x v = D_x\, (u + v)$$

Es decir, que $D_x u + D_x v$ es la primera derivada de $u + v$.

Como tomar la integral de $D_xu + D_xv$ equivale a preguntar: "¿de qué función es esta suma de derivadas la primera derivada?", la respuesta es $u + v$, porque $D_xu + D_xv$ es la primera derivada de $u + v$. O sea que:

$$\int_x (D_xu + D_xv) = u + v$$

pero

$$u = \int_x D_xu$$

$$v = \int_x D_xv$$

Si se sustituyen estos valores en la expresión original se puede escribir:

$$\int_x (D_xu + D_xv) = \int_x D_xu + \int_x D_xv$$

Esta fórmula se puede expresar como sigue: la integral de una suma de funciones es igual a la suma de las integrales de las funciones sumandos.

Si se le quiere dar mayor generalidad a la expresión, hay que pasar por encima de la mención explícita de las derivadas (que introdujimos a causa de nuestro proceso deductivo). Así, si llamamos w y z a dos funciones de x, podremos escribir:

$$\int_x (w + z) = \int_x w + \int_x z$$

Corolario: la integral de una resta de funciones es igual a la integral del minuendo menos la integral del substraendo.

Integración de una función precedida de un coeficiente

Si se trata de encontrar la integral de aD_xu, o sea:

$$\int_x (a\,D_xu)$$

Bastará recordar que esto equivale a preguntar: ¿cuál es la función de la que aD_xu es derivada? Como (en la revisión de las derivadas) establecimos que aD_xu era la primera derivada de au, podemos escribir como respuesta, au.

Como $u = \int_x D_x u$, lo anterior, se puede escribir como sigue:

$$a\,D_xu = a\int_x D_xu$$

Es decir, que: la integral de una función afectada por un coeficiente (o factor constante) es igual: a la integral de la función no afectada por él/ multiplicada (la integral) / por el coeficiente.

Lo anterior equivale a decir —también— que: un coeficiente o factor constante puede entrar o salir del integrador, sin modificar la integral.

A la expresión previa se le puede dar una formulación más general. Para ello, hay que referirla al caso en el que las derivadas no son explícitas. En ese caso, se tiene:

$$\int_x a\, f(x) = a \int_x f(x)$$

Aquí, en el primer miembro, a está dentro, y en el segundo, fuera del integrador y, sin embargo, hay igualdad.

Integración por partes

Si en la revisión correspondiente a las derivadas hemos establecido que:

$$D_x y = u\, D_x v + v\, D_x u$$

cuando $\quad y = uv$

Podemos, en la primera expresión, despejar al primer término del segundo miembro:

$$u\, D_x v = D_x y - v\, D_x u$$

Como y es igual a uv; esta expresión equivale a la siguiente:

$$u\, D_x v = D_x uv - v D_x u$$

Como se trata de una suma de funciones, podemos integrarla de acuerdo con la regla correspondiente:

$$\int u\, D_x v = \int D_x uv - \int v\, D_x u$$

De estos términos, se puede evaluar inmediatamente el primero del segundo miembro. En él, la función uv está afectada simultáneamente por los dos operadores inversos que anulan mutuamente sus efectos; o sea, que al ejecutar esa operación, resulta:

$$\int u\, D_x v = uv - \int v D_x u$$

A ésta se le conoce como la fórmula de la "integración por partes" y dice que:

La integral de una función por la derivada de otra función, es igual: al producto de dichas funciones / menos la integral de la segunda función por la derivada de la primera (con respecto a la variable común).

Naturalmente, también es cierto que:

$$\int v\, D_x u = uv - \int u D_x v$$

Se trata, fundamentalmente, de la misma fórmula.

En la formulación previa, en la parte final, también se puede decir en general: "menos la integral de una de las dos funciones por la derivada de la otra" (se da por supuesto que la derivación debe hacerse con respecto a la variable común).

En forma muy general, si la derivada no es explícita, se puede establecer que:

$$\int_x w\, z = w \int_x z - \int_x \left(\int_x z\right) D_x w$$

Esta nueva forma de expresar esas relaciones se podría traducir como sigue:

La integral del producto de dos funciones / es igual a: / la primera por la integral de la segunda / menos el integral del producto de la integral de la segunda por la derivada de la primera.

De nuevo, en esta expresión (como es indiferente que a una de las funciones se la tome por primera y a la otra por segunda o que se proceda a la inversa), se puede decir: "menos el integral / del producto de la integral de una de ellas / por la derivada de la otra".

Integración de una función elevada a un exponente constante

En la integración de una potencia obtenida mediante la elevación de una función a un exponente constante se puede poner de manifiesto que la integración (como proceso complejo de operación) es la inversa del proceso (asimismo, complejo) de derivación.

Si se busca la integral de la expresión nv^{n-1};

$$\int_x nv^{n-1} D_x v$$

en realidad se está preguntando lo siguiente:

¿Cuál es la función de la que ésta es primera derivada?

Como hemos visto, en la revisión de derivadas, el integrando es derivada de v^n; o sea, que la respuesta es:

$$\int_x nv^{n-1} D_x v = v^n$$

O sea, que, para obtener el integrando:

1º Se tomó el exponente n como coeficiente,

2º Se disminuyó en una unidad (— 1) el exponente (o sea, se dividió entre v), y

3º Se tomó la derivada de la función respecto de la variable.

En cambio para obtener la integral se sigue, *en sentido inverso*, la serie de operaciones que se realizaron para obtener la derivada, *y se invierte el sentido de dichas operaciones*, pues:

1º (Inverso del 3º anterior), se tomará la integral de la función no potenciada sujeta a derivación: $\int_x D_x v = v$.

2º (Inverso del 2º anterior), se aumentará en 1 el exponente (o sea, que se multiplicará por v).

3º (Inverso del 1º anterior), se tomará el exponente incrementado como divisor.

De este modo, en términos generales:

$$\int_x [f(x)]^n f'(x) = \frac{[f(x)]^{n+1}}{n+1}$$

Se puede recordar aquí que el 1 del exponente representa a f (x) en cuanto resultante de la integración de f′ (x), o sea, de la integración de la primera derivada de f (x).

Como caso particular, se puede considerar el de:

$$\int_x \frac{f'(x)}{[f(x)]^2} = [f(x)]^{-1}$$

puesto que, como hemos dicho, la derivada del recíproco de una función es igual a la derivada de la función entre el cuadrado de dicha función.

Integral del recíproco de una función

Si se considera que:

$$D_x L f(x) = \frac{D_x f(x)}{f(x)}$$

o sea:

$$= [f(x)]^{-1} D_x f(x)$$

al tomar las integrales, se obtiene:

$$\int_x D_x L f(x) = \int [f(x)]^{-1} D_x f(x)$$

y, como en el primer miembro, integrador y derivador anulan mutuamente sus efectos:

$$L f(x) = \int [f(x)]^{-1} D_x f(x)$$

O sea, que: la integral del recíproco de una función, multiplicado por la derivada de dicha función es igual al logaritmo de la función.

Algunos principios útiles para la integración

En una integración debe de distinguirse el tratamiento de los factores variables del tratamiento de los factores constantes. No todo lo que se puede hacer con los constantes se puede realizar con los variables.

Un factor constante puede: 1) *dentro del integrando*, pasar de la porción diferencial a la no diferencial, o de la no diferencial a la diferencial, y 2) puede salir o entrar al integrador, todo ello, sin modificarse ni modificar la integral.

Un factor variable puede: 1) pasar, *dentro del integrando*, de la porción diferencial a la no diferencial, o 2) pasar de la no diferencial a la diferencial, sin modificarse en ambos casos. En cambio, un factor *variable no* puede —a menos que se modifique, o de que modifique a la integral— entrar o salir del integrador.

O sea, que los factores constantes tienen plena libertad de movimiento: dentro del integrando y dentro de la integral, mientras que los factores variables sólo se puede mover dentro del integrando.

Existencia o inexistencia de un factor constante de integración y determinación del mismo (en caso de existir).

En muchas ocasiones, el único escollo que dificulta una integración (que impide se realice por simple inspección y uso de las fórmulas típicas) consiste en que la porción diferencial es menor que su derivada o sea, que bastaría con multiplicar la porción diferencial por un factor constante adecuado (o, en caso necesario, dividirlo por una constante apropiada) puede obtener la forma integrable por simple inspección. Esto se puede hacer, con base en los principios anteriores; pero, para que la integral no se altere, habrá que dividir la integral resultante entre ese factor (o, en el caso contrario, multiplicarla por el que sirvió de divisor de la porción diferencial).

A ese factor constante, por el que hay que multiplicar o dividir un integrando para hacerlo integrable, se le llama "constante de integración". Esa constante "de integración" no siempre se puede determinar a simple vista.

Para saber si realmente existe un factor de integración de determinado integrando, se sigue este procedimiento: 1) se deriva la parte no-diferencial del integrando; 2) se multiplica dicha derivada por A (factor constante *por determinar*); 3) se iguala el producto a la porción diferencial del integrando; 4) se igualan los coeficientes de potencias iguales de la variable que aparece en ambos miembros y, en cada caso, se despeja el valor de A. Si A es igual en todos los casos, el valor de A es el factor constante de integración. Si A es diferente en los distintos casos, *no hay factor de integración.* Una vez determinado el factor de integración, se procede como se ha indicado anteriormente; es decir: 1) se multiplica (o divide) por él el integrando, y 2) se divide (o multiplica por él) la integral resultante.

Integración de fracciones racionales (o de ciertos cocientes)

Una fracción racional es aquella en la que tanto numerador como denominador son polinomios enteros o funciones racionales enteras de la variable. Es decir, que en una fracción racional no aparecen exponentes fraccionarios.

En una fracción racional, el grado del numerador puede ser mayor o menor que el del denominador. Si es mayor, habrá un cociente y un residuo: si es menor, sólo habrá residuo.

Si en una fracción racional, el grado del numerador es mayor que el del denominador, al dividir el uno entre el otro se obtiene un cociente entero y un residuo fraccionario, y en este residuo fraccionario, el grado del numerador es inferior al del denominador.

Para integrar la fracción racional originaria, será necesario: 1) integrar el cociente, entero, y 2) integrar el residuo, fraccionario. Para hacer esto último, se procede en la forma que se delínea más adelante.

Si el grado del numerador, en una fracción racional, es menor que el grado del denominador, no se obtendrá cociente entero. La integración de este segundo caso se reducirá al problema de integrar el residuo del caso anterior.

Integración de fracciones racionales en las que el grado del numerador es menor que el grado del denominador

La integración de una fracción en la que el grado del numerador es menor que el del denominador, se facilita si se descompone la fracción total en fracciones parciales. Para lograrlo, se necesita que: 1) el producto de los denominadores de las fracciones parciales reproduzca el denominador de la fracción total, y 2) los numeradores de las fracciones parciales se determinan de tal modo que, hecha la reducción a un común denominador, la suma de los numeradores reproduzca el numerador de la fracción originaria.

Se pueden presentar dos casos principales: 1) que la descomposición produzca sólo factores reales, y 2) que produzca factores imaginarios. En el primer caso, se puede tratar: 1.1) de factores de primer grado o, 1.2) de factores de grado superior al primero.

Con vistas a las aplicaciones estadísticas más inmediatas, sólo daremos indicaciones para la integración de los casos (1.1) y (2).

Integración de fracciones a las que se puede descomponer en factores de primer grado

Dentro de este caso pueden considerarse —aun— dos subcasos: (1.11) aquél en que el denominador se puede descomponer en varios factores, todos iguales, de primer grado.

En el primer subcaso, el factor típico de la descomposición del denominador se puede representar por $(x - a)$. Si llamamos A al numerador (cuyo valor está por determinar) de la fracción parcial correspondiente, la fracción $A/(x + a)$ es la forma típica de la fracción resultante de esta descomposición.

Para encontrar el valor de los numeradores (A) de las fracciones-sumando: 1) hay que multiplicar cada numerador por todos los factores resultantes de la descomposición del denominador, exceptuando el suyo; 2) sumar los resultados obtenidos e igualar la suma al numerador de la fracción originaria; 3) igualar los coeficientes de potencias iguales de la variable en los dos miembros de la igualdad resultante, y 4) despejar los valores de los numeradores, en dichas expresiones. Obtenidos los valores de los numeradores en las fracciones parciales; 5) habrá que integrar las fracciones parciales resultantes, y 6) sumar dichas integrales.

En el segundo caso, si el factor repetido es $(x + a)$, hay que formar las fracciones tomando como denominadores la primera, la segunda, la tercera, la enésima potencia del factor $(x + a)$ y deberá haber tantas fracciones como veces aparezca repetido el factor. Si n representa el número de veces que se repite el factor en el denominador originario $(x + a)$ será el denominador máximo. El proceso de integración es —a partir de aquí— análogo al anterior.

Integración de fracciones cuyo denominador contiene factores imaginarios

Cuando el denominador contiene factores imaginarios, éstos deben aparecer pareados (en forma de complejos conjugados) a fin de que el resultado sea real. Si el factor imaginario es de la forma $x + a_0 + ai$ su conjugado será $x + a - a\,i$. El producto de estos complejos conjugados produce $(x + a^2 + a)$. Las fracciones parciales correspondientes serán de la forma $(A_1 x + A_0)/[(x + a)^2 + b^2]$.

Relievamiento de características por descripción estadística

Cálculo de promedios

Medias

Media estadística es el valor representativo de una distribución que se obtiene si: 1) se sujeta a la variable a cierta operación; 2) se suman los valores obtenidos; 3) se divide la suma entre el número de datos, y 4) se somete al cociente a la operación inversa de aquella a que se sujetó la variable.

Según la operación que se realice con los diversos valores de la variable (y la operación inversa a la que se sujete el cociente que resulta de dividir la suma de los resultados entre el efectivo) la media recibirá el calificativo de: 1) aritmética; 2) cuadrática; 3) cúbica; 4) armónica, y 5) geométrica.

La media aritmética

La media aritmética se obtiene si: 1) se dejan intactos los valores de la variable; 2) se suman esos valores; 3) se divide la suma entre el número de datos, y 4) se deja intacto el resultado. Las medias cuadrática y cúbica se obtienen en forma parecida: 1) operando con los cuadrados y cubos respectivamente y 2) extrayendo las raíces cuadradas y cúbicas de los resultados respectivos.

La media armónica

La media armónica se obtiene: 1) si se obtienen los recíprocos de los valores de la variable; 2) se suman dichos recíprocos; 3) se divide la suma de los recíprocos entre el efectivo, y 4) se busca el recíproco del cociente.

La media geométrica

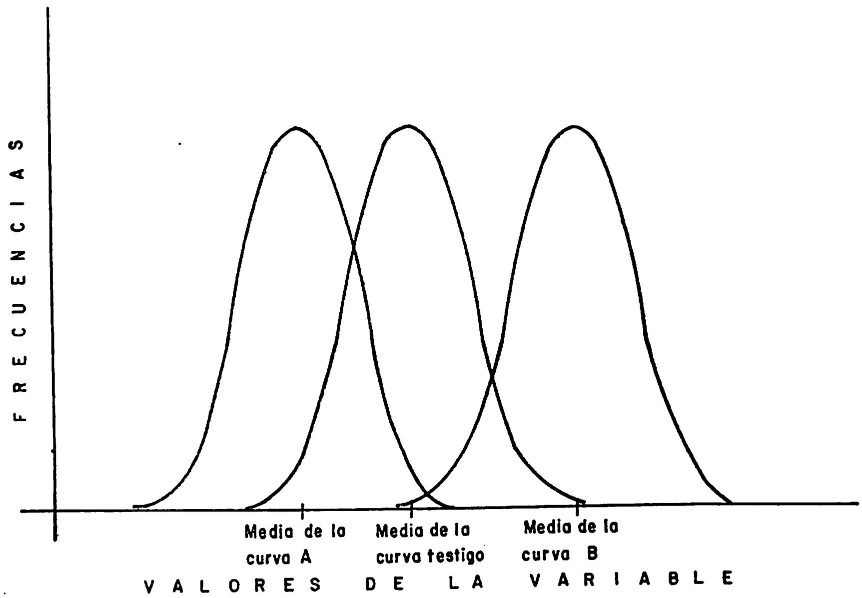
La media geométrica (también llamada logarítmica) se obtiene: 1) si se buscan los logaritmos de los valores de la variable; 2) se suman dichos logaritmos; 3) se divide la suma de los logaritmos entre el número de datos, y 4) se busca el antilogaritmo del cociente.

En el caso de la media geométrica, conviene advertir que si los logaritmos que se toman son decimales, el antilogaritmo del cociente que se busque al final deberá de ser —también— de base decimal; que si se usan logaritmos naturales de los valores de la variable el antilogaritmo del cociente deberá ser —también— antilogaritmo natural (de base e).

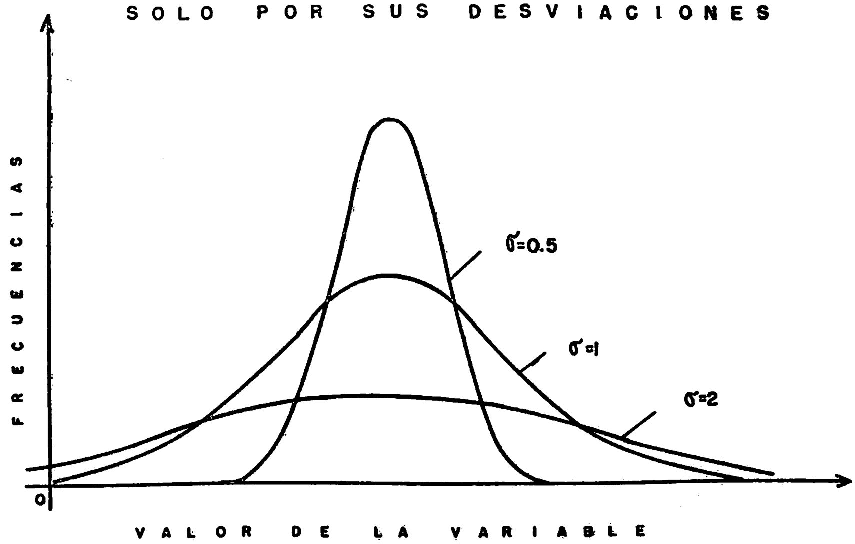
Simbolización de las medias

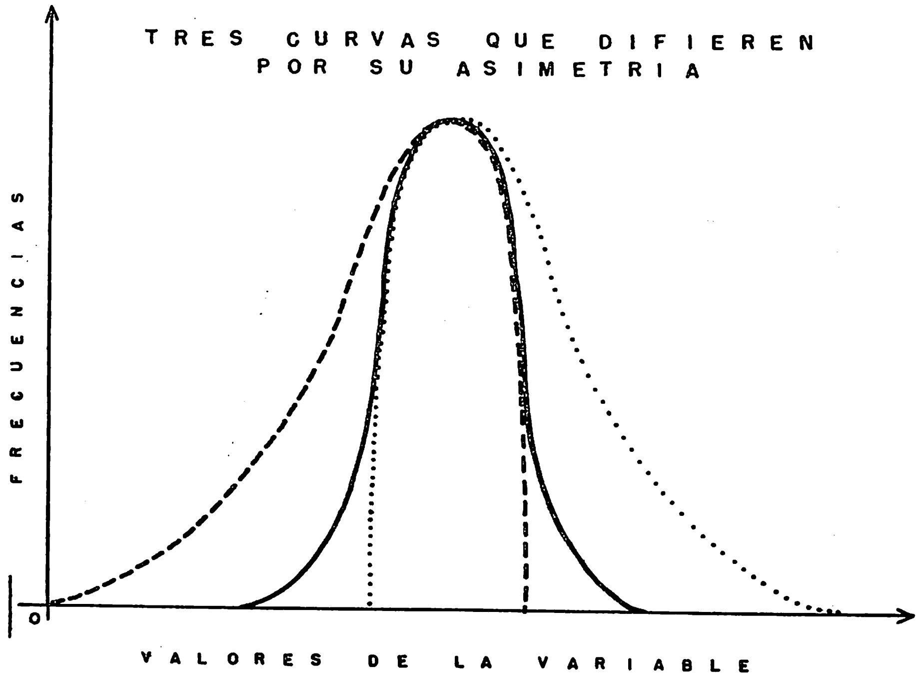
Las medias estadísticas se simbolizan colocando encima de la literal que representa la variable una raya horizontal. Como la media es un valor único y no múltiple (constante y no variable, para una distribución dada), el subíndice con que solemos afectar a la variable (generalmente el subíndice i o el j) no tiene razón de ser. En cambio, como hay varias medias estadísticas para cada distribución, conviene particularizar cada una. La solución más sencilla consiste en colocar como subíndice: a para la media aritmética, q para la cuadrática, c para la cúbica, a o h para la armónica (o harmónica) y g para la geométrica. Ésta era la práctica que aconsejábamos en *Técnicas estadísticas para investigadores sociales*; pero, como criticaba en forma muy prudente el profesor Francisco Hernández

TRES CURVAS QUE DIFIEREN SOLO POR SUS MEDIAS



TRES CURVAS QUE DIFIEREN SOLO POR SUS DESVIACIONES





Villarreal, de la Escuela Nacional de Ciencias Políticas y Sociales de México (a través de un alumno común), esos subíndices tenían significado por sí mismos para un hablante del español (ya que, en su mayoría corresponden a las iniciales de los adjetivos calificativos de "media") y, en cambio, podían significar poco —fuera del dominio convencional— para los hablantes de otros idiomas. De ahí que en este texto recomendemos que se usen los siguientes subíndices para la simbolización de las medias estadísticas: "1" para la aritmética, ya que en ella las potencias y raíces que intervienen son de *primer* grado; "2" para la cuadrática, pues en ella intervienen el cuadrado y la raíz cuadrada o de *segundo* grado; "3" para la cúbica, por razones análogas; "—1" para la armónica, pues en ella intervienen los recíprocos o potencias *negativas de primer* grado y "L" para la geométrica por intervenir en ella logaritmos y antilogaritmos.

La media geométrica obtenida sin logaritmos

En cuanto las operaciones con logaritmos tienen sus equivalentes en otras operaciones realizadas con los números naturales correspondientes, siendo la suma de logaritmos equivalente al producto de los números a los que corresponden, y el cociente a la raíz, la media geométrica puede calcularse si: 1) se dejan intactos los valores de la variable; 2) se multiplican esos valores entre sí; 3) se extrae del producto una raíz cuyo índice esté dado por el efectivo de la distribución (raíz enésima), y 4) se deja intacto el resultado.

Propiedad de la media aritmética

Cuando a los valores de una variable se les suma una constante, la media aritmética de los nuevos valores resulta igual a la media aritmética de los antiguos, pero aumentada en la constante que se sumó a los valores de la variable.

A base de la pura definición operativa de media aritmética, se puede demostrar la verdad de esta afirmación. En efecto, en este caso se tiene: 1) cada valor de la variable aumentado en la constante; 2) la suma de los valores de la variable, aumentada (dicha suma) en tantas veces la constante como valores tenga la variable (o sea, en la constante multiplicada por el efectivo); 3) el cociente de la suma de los valores de la variable y el producto de la constante por el efectivo entre el efectivo; este cociente produce: por un lado, 3.1) la suma de los valores de la variable entre el efectivo, o sea la vieja media aritmética, y, por otro lado, 3.2) la constante (porque ésta aparecía multiplicada por el efectivo, entre el que hubo de dividirse después). Si este resutlado se deja intacto, esta propiedad se puede expresar como se dijo anteriormente. Si lo que se quiere es obtener la vieja media aritmética a partir de esta nueva media artimética (aumentada en la constante) habrá que dar el cuarto paso (inverso del primero, como siempre), el cual consiste en restar de la nueva media aritmética la constante, para obtener la vieja media aritmética.

La última parte del razonamiento conducirá, en su momento, a la simplificación del cálculo de la media aritmética.

Sin embargo, como el razonamiento previo puede parecer demasiado abstracto y verborreico, el mismo se presenta, en seguida, en forma más accesible para el lector acostumbrado a pensar en términos artiméticos (conforme a una filosofía pedagógica de la estadística que hemos tenido ocasión de exponer en otro sitio).

Significación de la propiedad de la media aritmética

Mostraremos la significación de esta propiedad: 1) primero, en lenguaje aritmético, para 2) traducirlo después al lenguaje algebraico, y 3) demostrar el principio —finalmente— en términos ya puramente algebraicos.

Lenguaje aritmético

Sean los siete casos siguientes, 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.) les 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) menos la cantidad constante (3) que restamos de cada uno de los valores a partir de los cuales calculamos la media.

Y lo mismo que decimos de un valor restado de los de la variable, podemos decirlo, *mutatis mutandis,* de cualquier otro valor, también constante, que se le sume a cada uno de sus valores.

Traducción algebraica

Si se tiene un rasgo de magnitud x_i, 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 ($x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$ igual a $\Sigma\, x_i$) y dividiendo la suma entre el número de datos (7, en este caso, o N, en general). O sea, que la media aritmética será $\Sigma\, x_i \,/\, N$.

Si de cada uno de los valores de la variable ($x_1, x_2 \ldots$) 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: $(x_1 - x') + (x_2 - x') + (x_3 - x') + (x_4 - x') + (x_5 - x') + (x_6 - x') + (x_7 - x')$ o sea que tendremos, por un lado, todas las equis sumadas ($x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = \Sigma\, x_i$) y por otro, siete veces $-x'$ (o sea, $-7x'$, o, en general, $-Nx'$). Es decir, que el resultado de los valores de la nueva serie es: $\Sigma\, x_i - N\, x'$. Como el tercer paso del cálculo de la media consiste en dividir esta suma entre el efectivo (N), la nueva media resultará ser: $(x_i - N\, x')$ entre N, o sea: $\Sigma\, x_i \,/\, N$ menos $(Nx') \,/\, N$. Pero, $\Sigma\, x_i \,/\, N$ es igual a la antigua media aritmética $\bar{x}$ y $(N\, x') \,/\, N$ produce x'. Esto significa que la media de los nuevos valores $\bar{x}_I$ es igual a la media de los antiguos $\bar{x}$ menos la constante que se restó de cada valor de la variable (x').

Pensamiento algebraico

La fórmula de la media aritmética para una variable x_i está dada del modo siguiente:

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

Si de las x_i se resta una constante x', la nueva media aritmética (x_I) será:

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

Si se tiene en cuenta que la suma de una serie de sumas (o restas, como en este caso) es igual a la suma (o resta) de las sumas de los sumandos (o de los minuendos, por una parte, y de los substraendos, por otra), se tendrá:

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

Pero $\Sigma x'$ es la suma de una serie de constantes que es igual al producto de la constante por el número de veces que aparece (o sea N); por lo cual, lo anterior se convierte en:

$$= \Sigma x_i - Nx'$$

Al dividir todo el primer miembro entre N, se obtendrá la nueva media aritmética $(\bar{x}_I)$. Al dividir el segundo entre N se obtendrá una diferencia de fracciones, como se indica en seguida:

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

Es fácil reconocer, aquí, en el segundo miembro, que la primera fracción es la antigua media aritmética $(\bar{x})$ y que, como en la segunda figura N en el numerador y en el denominador, se puede escribir, en vez de ella, x':

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

Esto era lo que se quería demostrar.

Aplicación del principio al cálculo de la media aritmética de una serie

Cuando los datos de una serie son muy complicados (números muy grandes) se puede simplificar el cálculo de la media aritmética si se aplica el principio que se acaba de demostrar. Para el efecto, se debe proceder como sigue: 1) réstese de cada dato una cantidad fija; 2) calcúlese la media aritmética de los residuos de dichas restas, y 3) agréguese a la media aritmética obtenida la cantidad fija restada al principio; ésta es la media aritmética de la serie.

A la cantidad fija que se resta se la llama "media adivinada", porque, con el fin de simplificar al máximo las cantidades con las que se ha de trabajar, se procura adivinar cuál será, más o menos, el valor de la media. Conforme la constante que se reste esté más próxima de ella, los residuos con que se tenga que operar serán menores. De hecho, su suma será nula, si la "adivinación" fue tan acertada que permitió encontrar la verdadera media aritmética.

En ocasiones, el principio se puede aplicar también si no se resta, sino se agrega una constante. Así, por ejemplo, si los valores de una serie son 2.25, 4.25, 7.25, 8.25 puede convenirnos agregar a cada uno de ellos, 0.75 porque en esta forma obtenemos enteros (aunque, como es obvio también los obstendríamos, en el caso concreto, restando de continuo 0.25). Tras agregar la cantidad pertinente (0.75) procederíamos a obtener la media aritmética de los nuevos valores (enteros, más fáciles de manejar) y, después de obtenida, le restaríamos (como antes le agregamos) la constante 0.75.

Amplitud, oscilación o campo de variabilidad

Una de las características más sencillas y más fáciles de computar en una distribución estadística es la que se conoce como amplitud, oscilación o campo de variabilidad. Operativamente se le define como la diferencia entre el máximo absoluto y el mínimo absoluto de la distribución.

Cuando el mínimo es una cantidad negativa (cosa que ocurre cuando se recogen, por ejemplo, ciertas calificaciones a algunas pruebas de sociabilidad entre los asociales o antisociales) es necesario recordar que esa diferencia entre el máximo y el mínimo se debe entender en términos de diferencia o resta algebraica y que, para realizarla hay que cambiar el signo del sustraendo. Como el mínimo es una cantidad negativa (en estos casos) al cambiar de signo se convierte en positiva y, por lo tanto, hay que sumar aritméticamente su valor al máximo (en caso de que éste sea positivo). Si el máximo también es negativo, cosa que puede ocurrir si, en el ejemplo, el grupo está constituido exclusivamente por antisociales, habrá que sumar a este número negativo el que, siendo originalmente negativo se convirtió en positivo.

Si representamos por Amp. la amplitud, por m el mínimo de la distribución y por M el máximo de la misma, su fórmula será:

$$\text{Amp.} = M - m.$$

Los casos que se presentan en el cálculo son:

1. Máximo y mínimo positivos: Réstense aritméticamente.

2. Máximo y mínimo negativos: Réstense algebraicamente, o sea: cámbiese signo al mínimo (de − a +) réstese del mayor el menor y póngasele el signo del mayor.
3. Máximo positivo y mínimo negativo: Réstense algebraicamente, o sea: cámbiese signo al mínimo (de − a +) súmese al mayor el menor y aféctese todo de signo +.
4. Máximo negativo y mínimo positivo: Réstense algebraicamente, o sea: cámbiese signo al mínimo (de + a −) réstese del mayor el menor y póngasele el signo del mayor.

Medidas de variabilidad

Desviaciones

Desviación estadística es la diferencia (positiva o negativa) de un valor determinado con respecto a un valor constante o central que se toma como término de comparación.

Las desviaciones son medidas individuales (puesto que corresponde *una a* cada *uno* de los casos o valores individuales) de variación (ya que indican qué tanto varía *cada dato respecto de un valor* fijo).

Dispersión

Frente a la desviación, que es un carácter de cada uno de los datos de una distribución, la *dispersión es una característica de la distribución* considerada *en su conjunto,* como una unidad.

La dispersión de una distribución indica el mayor o menor grado de variabilidad de los datos con respecto a un punto fijo que es, generalmente, un promedio y, más particularmente, la media aritmética de la distribución. Consecuentemente, la dispersión indica si los datos de la distribución se *concentran* mucho en torno del promedio o si, por el contrario, se encuentran muy *dispersos* con respecto a él.

Las medidas de dispersión indican hasta qué punto es representativo de toda la distribución, el promedio respecto del cual se toman. Si una medida de dispersión indica que la distribución se concentra en torno del promedio, esto significa que el promedio se puede tomar como representativo de la distribución y que, al hacerlo, se corre poco riesgo. En cambio, si la medida indica que la distribución es dispersa con respecto al promedio, aún se puede tomar a éste como representativo de la distribución, pero hacerlo representa un riesgo mayor (en cuanto sólo representa bien a un grupo reducido de valores, y no a la mayoría o a la totalidad).

Como la dispersión es una medida propia de una distribución (es decir, es una característica de un conjunto o colección), y cumple para ésta la misma función que las desviaciones desempeñan para los casos individuales, está dentro de las líneas del pensamiento estadístico el definirla como un promedio (cualquiera) de desviaciones con respecto al centro de dispersión que se haya elegido.

Dos medidas de dispersión

Aquí nos ocuparemos sólo de dos medidas de dispersión. Una de ellas (la "desviación media aritmética") nos servirá para mostrar, en sus términos más simples, la forma en que para medir la dispersión se recurre a los promedios. La otra (la "desviación cuadrática") introduce una mayor complicación, pero se la debe conocer porque es la más usual (aquella a la que, en muchas ocasiones, se alude, excluyendo así a todas las demás medidas de dispersión).

La desviación media aritmética es la media aritmética de las desviaciones de los datos (generalmente con respecto a la medida aritmética de la distribución).

Esto significa que, para calcular esta medida de dispersión: 1) se calcula la media aritmética de los datos; 2) se resta de cada dato el resultado para obtener las desviaciones; 3.1) se suman las desviaciones así obtenidas y 3.2) se divide la suma entre el efectivo (o número de datos).

Si se considera que la suma de las desviaciones con respecto a la media aritmética es siempre cero, la desviación media aritmética (que habría que calcular a partir de ellas) como ya se dijo, sería siempre nula, y no serviría para medir nada. A partir de ella no podríamos conocer la mayor o menor concentración (o la dispersión menor o mayor) de los datos de una distribución respecto a su media, y menos aún comparar la mayor o menor concentración (o la dispersión menor o mayor) de los datos de varias distribuciones con respecto a sus correspondientes medias. A fin de evitar este inconveniente, se toman las desviaciones, pero *sin su signo*; o sea, que se trabaja con los valores absolutos de las desviaciones.

De acuerdo con esto, la desviación media aritmética es la media aritmética de los valores *absolutos* de las desviaciones de los datos con respecto a la media aritmética de la serie.

En efecto, en cuanto se eliminan los signos de las desviaciones (o sea, cuando se toman los valores absolutos de éstas) con respecto a la media aritmética, la suma de los valores positivos no anula ya la suma de los negativos y —gracias a esto— la desviación media aritmética sirve ya para los propósitos para los que fue creada: 1) la comparación de dispersiones—concentraciones de las series, y 2) la estimación de la representatividad o falta de representatividad de los promedios.

Si por d_i representamos las desviaciones de los diferentes datos y aplicamos la práctica simbológica de representación de las medias, * la des-

* (Colocar una raya horizontal sobre la literal de la variable, desprovista del subíndice i, marcando el número de orden del promedio con un número 1, para la media aritmética.)

viación media aritmética quedará representada por $\bar{d}_1$ (que se lee "d sub-uno testada o d testada sub-uno).

En símbolos, $\bar{d}_1$ es igual a $(\Sigma\, d_1)/N$

Desviación media cuadrática

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

La desviación media cuadrática responde —por otra vía— al inconveniente que tenía la anulación resultante de la suma de las desviaciones positivas y negativas con respecto a la media aritmética. En efecto, en la media cuadrática se usan cuadrados (en este caso, cuadrados de desviaciones, o sea, cuadrados de cantidades que unas veces son positivas y otras negativas), y los cuadrados son siempre positivos. O sea, que, por otro camino, se ha eliminado la posibilidad de que las desviaciones se anulen unas con otras. Como en toda media cuadrática, la extracción se colocará en el mismo nivel que los valores originarios (se asegura que será un valor de primero y no de segundo grado).

Conforme a las definiciones de desviación cuadrática media y de media cuadrática, el cálculo de esta medida de dispersión se hará en la forma siguiente: 1) se obtendrá la media aritmética; 2) se restará de cada valor de la variable dicha media; 3.1) se elevarán al cuadrado los resultados de dichas restas; 3.2) se sumarán esos cuadrados; 3.3) se dividirá la suma entre el efectivo, y 3.4) se extraerá la raíz cuadrada del cociente; ésta será la desviación cuadrática media o desviación media cuadrática, también conocida como "desviación estándar".

De acuerdo con nuestra simbología, como se trata de una desviación, representaremos esta medida por d; como es una desviación promedio, la d aparecerá "testada" ($\bar{d}$), y como el promedio es la media cuadrática (en la que intervienen la potencia y la raíz *segundas*) afectaremos la literal con el subíndice 2: $\bar{d}_2$.

En símbolos, $\bar{d}_2$ es igual a la raíz cuadrada (exponente ½) de

$$\left[\Sigma\,(d_1{}^2)\right] / N$$

Como la desviación media cuadrática es una medida de dispersión de uso frecuentísimo, merece que se la designe no sólo con los símbolos genéticos y las especificaciones correspondientes (des con unos subíndices peculiares) como cuando se da a un individuo común y corriente el apellido paterno y el propio nombre, sino con un símbolo propio (como cuando al ya consagrado se le designa con un nombre único e inseparable: "El Greco", por ejemplo) y así, la desviación media cuadrática se designa con la letra griega σ (sigma minúscula).

Aunque σ (sigma minúscula) no es signo que —por su forma— se confunda con Σ (sigma mayúscula) sí es fácil confundir una con otra si en sus nombres no se especifica que se trata de la mayúscula o de la minúscula. Esto, que en la escritura griega podría conducir a una falta contra la ortografía, en estadística conduciría a un error por la alusión que se haría a funciones completamente distintas. En efecto, la minúscula (σ) designa convencionalmente —según acabamos de ver— una *magnitud* (la desviación media cuadrática); en cambio, la mayúscula (Σ) no designa magnitud alguna, sino que es un operador: ordena que se ejecute una operación (una suma).

Propiedad de la desviación media cuadrática útil para simplificar su cálculo

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

Hay que aclarar que, cuando se trata de la "desviación media cuadrática de la nueva distribución" nos referimos a la que se calcula con respecto a la *media de los nuevos valores.*

Para una serie cuyos valores estén representados por x_i, la media aritmética será $\bar{x}$ (suprimimos el subíndice, en tanto no hay que hacer distinción entre la media aritmética y otras medias) y la desviación media cuadrática $\bar{d}_2$. $\bar{d}_2$ será, como ya dijimos, la raíz cuadrada (exponente ½) del cociente que resulta de dividir entre N la suma de las desviaciones de las x_i con respecto a x, o sea:

$$d_2 = \left[\frac{\Sigma \ (x_i - x)^2}{N} \right]^{1/2}$$

Si a cada x_i le restamos una constante x', los nuevos valores de la serie serán $(x_i - x')$.

Estos nuevos valores tendrán una nueva media aritmética conforme a la propiedad que ya se asentó en el lugar correspondiente. Esa nueva media será igual a la antigua ($\bar{x}$) menos la constante restada (x'); o sea, que la nueva media será $x - x'$.

De acuerdo con esto, las nuevas desviaciones serán iguales a los nuevos valores de la variable menos la nueva media: $(x_i - x') - (\bar{x} - x')$. Si quitamos paréntesis: $x_i - x' - \bar{x} + x' = x_i - \bar{x}$ (ya que las x' se anulan una a otra). O sea que, en el fondo, las desviaciones quedan incambiadas. Esto es natural, pues se han tomado desviaciones con respecto *a las medias respectivas.*

Si las desviaciones no cambian, la nueva desviación media cuadrática (σ_1) no tiene por qué cambiar pues, para obtenerla, se elevan esas desviaciones al cuadrado, se dividen entre N (una constante) se suman y se les extrae la raíz cuadrada (como en el caso de la σ originaria). Se hacen operaciones iguales con cantidades iguales y se obtienen, por lo mismo, resultados iguales. O sea que: $\sigma_I = \sigma$.

La variancia o variación estadística

Una medida íntimamente relacionada con la desviación cuadrática media que también se usa para medir la dispersión-concentración de las series o distribuciones es la medida de variación estadística o "variancia".

La variancia es el cuadrado de la desviación media cuadrática

Esto significa que la variancia es un resultado que se obtiene en el cómputo de la desviación media cuadrática; es el resultado que se obtiene durante dicho cómputo cuando no se ha llegado a extraer, aún, la raíz cuadrada final.

Según esto, la variancia se calcula: 1) obteniendo la media aritmética de la serie; 2) restándola de cada uno de los valores de la misma serie; 3.1) elevando al cuadrado los resultados (o sean las desviaciones); 3.2) sumando dichos cuadrados, y 3.3) dividiendo la suma entre el efectivo.

En este procedimiento de cálculo de la variancia se configura el patrón de cálculo de todo promedio: 1) se tomaron ciertos valores (los cuadrados de las desviaciones del paso 3.1), y se sometieron a una operación (se dejaron intactos); 2) se sumaron; 3) se dividieron entre el efectivo (en el paso 3.3), y 4) se ejecutó con el cociente la operación inversa a la realizada con los datos (es decir, se le dejó intacto porque intactos se tomaron los cuadrados de las desviaciones). O sea, que la variancia es una media aritmética (en la media aritmética se dejan intactos los datos y el cociente) de los cuadrados de las desviaciones (que fueron los datos, en este caso).

Según esto, la variancia o variación estadística está dada por la media aritmética de las segundas potencias de las desviaciones (con respecto a la media aritmética de la serie, en la mayoría de los casos).

Momentos respecto de la media aritmética

La variancia es un caso particular de lo que se conoce como momento con respecto a la media aritmética.

La variancia es una media aritmética de segundas potencias de desviaciones. Pero hay otras medias aritméticas de potencias de desviaciones. A estas "medias aritméticas de potencias de desviaciones" se les llama "momentos". La variancia, por serlo de las segundas potencias, se llama "segundo momento"; pero hay también terceros, cuartos y enésimos momentos. Un tercer momento será la media aritmética de las terceras potencias; el cuarto momento, la media de las cuartas, y el enésimo momento el de las enésimas *de las desviaciones.*

La asimetría

Para caracterizar a una serie (o a una distribución) en sus términos más simples se recurre a un promedio y a una medida de variabilidad. Sin embargo, hay ocasiones en que se buscan caracterizaciones más precisas; en tales casos, se recurre por lo menos a otras dos características conocidas como la asimetría y la curtosis.

Las medidas de asimetría-simetría tratan de señalar: 1) si los casos con valores superiores a la media aritmética (o a otro promedio elegido) tienen correspondencia exacta con otros valores de la distribución, inferiores a la misma (en cuyo caso se dice que hay simetría), o 2) si esto no ocurre (en cuyo caso hay asimetría).

El grado de asimetría de una distribución se puede medir si se utiliza el tercer momento. La medida de asimetría es ese tercer momento. O sea, que la asimetría se mide mediante la media aritmética de las terceras potencias de las desviaciones. Si este tipo de momento lo representamos por la letra griega (μ, mu) y especificamos el tercero mediante un subíndice 3 (μ_3) la medida de asimetría estará dada por $\mu_3 = [\Sigma\, d_i^3]/N$.

Si la distribución es simétrica, para cada valor positivo de d_i (x_i mayor que $\bar{x}$ en la fórmula $x_i - \bar{x}$) habrá un valor igual pero negativo de d_i (x_i menor que $\bar{x}$). El cubo de cada desviación positiva será positivo, de valor igual, pero de signo diferente 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 $+d_i$ se anulará con la correspondiente $-d^3_i$. Si dicha suma es nula, la fracción que tiene por numerador esa suma lo será también, y el tercer momento (al que equivale esa fracción) también será nulo. De este modo, cuando esta medida es igual a cero, la distribución es simétrica, y cuando esa medida es distinta de cero, la distribución es asimétrica.

La curtosis

La curtosis es una medida del aplanamiento o picudez de las distribuciones (su significado sociológico preciso puede verse en nuestro libro sobre *La matemática, la estadística y las ciencias sociales*). Esta característica se mide mediante el cálculo del cuarto momento, o sea, a través de la media de las cuartas potencias de las desviaciones. Como toda potencia de orden par (4) es positiva (sea negativa o positiva la desviación correspondiente), esta medida no suscita los problemas de la asimetría. No hay, por lo mismo, un valor central "cero" que marque la distinción entre lo medio, lo bajo y lo alto, como ocurre en la asimetría en que: 1) cero marca lo simétrico; 2) una asimetría negativa marca lo bajo, y 3) una positiva, lo alto. Dependerá de la teoría de las distribuciones el descubrir cuál es el valor medio de la curtosis. A partir de ese valor, se podrá establecer, o 1) que una distribución tiene curtosis media (es *meso-cúrtica* o normal); o que 2) tiene gran curtosis (*es lepto-cúrtica* o picuda), o que 3) tiene pequeña curtosis (es *plati-cúrtica o aplanada*).

Reducción a términos comparables

Una de las palabras de orden de la estadística es "comparabilidad". El estadístico se esfuerza siempre por hacer comparables, desde el ángulo numérico, todos aquellos colectivos que, en otras condiciones no lo serían.

Se puede mencionar varias formas de hacer comparables los datos; pero, casi todas se reducen —fundamentalmente— a la obtención de relativos (o sea, cocientes) de los valores originales (valores absolutos) con respecto a cierto valor o a cierta serie de valores que se toman como un término de comparación. Una variante de los relativos son los porcientos o porcentajes; éstos no son sino los relativos que aparecen multiplicados por 100 (lo cual se hace, generalmente, para evitar los valores fraccionarios que produce el cálculo de relativos). En términos más amplios y en esta misma vía de simplificación de los resultados, se calculan no sólo porcentajes, porcientos o tantos por ciento, sino tantos por mil, tantos por diez mil, o por un millón (y, en general tantos por 10 elevado a la potencia n).

El cálculo de relativos plantea algunos pequeños problemas matemáticos, y ofrece diversas variantes estadísticas. En primer término, nos ocuparemos de esos problemas; en segundo, hablaremos de esas variantes.

Problemas en el cálculo de relativos

El cálculo de relativos plantea dos problemas principales. Por una parte, se necesita conocer cuál es el tipo de datos que pueden relacionarse y cuál el de los que no se pueden relacionar estadísticamente. Por otra, se necesita conocer en qué forma pueden y deben, y en qué formas no se pueden ni se deben tratar los resultados de la relativización.

Tipos de relativos

Allen reconoce tres tipos de relativos: 1) el de los que relacionan a la parte con el todo; 2) el de los que relacionan una cifra con otra similar que procede de otro lugar, tiempo o sociedad, y 3) el de los que relacionan cifras de tipo diferente, pero asociado. Como ejemplo del primer tipo de relativos se puede mencionar la relación o el relativo que resulta de dividir el número de solteros entre el total de pobladores (solteros o no); como ejemplo del segundo tipo, el relativo resultante de dividir el precio actual de la leche entre el que tuvo hace diez años; como ejemplo del tercero (según el propio Allen), la relación entre el número de muertes que se producen *en un periodo dado* entre el número de habitantes existentes *en una fecha dada.*

Estas dos últimas son cantidades asociadas, similares, pero no del mismo tipo, pues una se refiere a un periodo y la otra a una fecha; pues una habla de muertos y la otra de vivos.

El problema de las unidades

En el cálculo de los relativos es necesario asegurar que los absolutos que se relacionen estén expresados en las mismas unidades o en unidades que puedan reducirse unas a las otras. En este sentido conviene, frecuentemente, ejecutar por separado las operaciones dimensionales, a fin de saber cuáles son las unidades resultantes de la división.

Hay ocasiones en las que las cantidades que hay que relacionar estadísticamente son del mismo tipo, pero están medidas en unidades diferentes (en centímetros unas, en pulgadas otras). En tales casos, *conviene* convertir unas unidades a las otras. Pero *no es indispensable* que esa conversión sea previa al cálculo propiamente dicho. Antes de relacionar las cantidades, se pueden colocar los factores de conversión (1 cm. = 0.39 in) en el lugar que les corresponde dentro de las fórmulas o relaciones matemáticas con que se opera. Una vez hecho esto, se puede trabajar con ellos como se trabaja con los números.

Ejemplo: Si se trata de relacionar 85 centímetros con 13 pulgadas, y se recuerda que 0.39 es el equivalente del centímetro en pulgadas, se puede escribir: (85 cm. x 0.39 in/cm.) /13 in. = 33.15 in /13 in = 2.55. Pero también se puede escribir, por una parte, la relación numérica y por otra, la dimensional; o sea por una (85/13) x 0.39 y, por la otra, un producto de dos relaciones dimensionales: la primera (cm / in) procedente de las cantidades relacionadas; la segunda, resultante de relacionar las dimensiones de sus unidades (in/cm) (equivalente de los centímetros en términos de pulgadas).

En una u otra forma, se obtiene el mismo resultado numérico y —como antes— una dimensión cero como propia de ese resultado.

Aproximaciones y redondeamientos

Las cifras obtenidas de los recuentos estadísticos, no son, por lo general, exactas; son sólo aproximadas. De ahí que sea presuntuoso hacer creer que los resultados estadísticos que se obtienen de ellas tienen gran exactitud y precisión. En efecto, no hay resultado que pueda ser más preciso de lo que son estos datos originales de los que se obtiene.

EJEMPLO DE CALCULO DE LOS PORCIENTOS

Distribución porcentual de la fuerza de trabajo
por ramas de actividad, en México, en 1950

CATEGORÍAS *Actividades*	FRECUENCIAS *Fuerza de trabajo en cada actividad*	CÁLCULO	RESULTADO *Porciento de la fuerza de trabajo de esa actividad, respecto de la fuerza total de trabajo*
Agricultura	4 823 901	$\frac{4\,823\,901}{8\,272\,093} \times 100 =$	58.31
Industrias extractivas	97 143	$\frac{97\,143}{8\,272\,093} \times 100 =$	1.17
Industria de transformación	972 542	$\frac{972\,542}{8\,272\,093} \times 100 =$	11.76
Industria de la construcción	224 512	$\frac{224\,512}{8\,272\,093} \times 100 =$	2.71
Electricidad y gas	24 966	$\frac{24\,966}{8\,272\,093} \times 100 =$	0.30
Comercio	684 092	$\frac{684\,092}{8\,272\,093} \times 100 =$	8.27
Transportes	210 592	$\frac{210\,592}{8\,272\,093} \times 100 =$	2.55
Servicios	879 379	$\frac{879\,379}{8\,272\,093} \times 100 =$	10.63
Insuficientemente especificadas	354 966	$\frac{354\,966}{8\,272\,093} \times 100 =$	4.29
TOTAL Fuerza total de trabajo	8 272 093	$\frac{8\,272\,093}{8\,272\,093} \times 100 =$	99.99%

En general —como asienta Allen— es generalmente seguro dar a entender que un cociente es correcto sólo hasta llegar a una cifra significativa, cuyo orden sea inmediatamente anterior al de la última cifra del menos exacto de los dos valores que intervienen en el cociente (trátese del dividendo o del divisor). La regla es aplicable, precisamente, en el cálculo de los relativos.

En forma inversa, si se quiere obtener una exactitud determinada, se necesita partir de datos que "incluyan —en el cómputo— una cifra significativa más de las que se requieran en los resultados".

Algunos usos estadísticos de los relativos

Los relativos se usan, en estadística, principalmente: 1) para reducir a términos comparables dos o más distribuciones de desviaciones que han sido clasificadas de acuerdo con diferentes intervalos; 2) para reducir las medidas n-dimensionales a medidas de dimensión cero (o sea, a números abstractos); 3) para determinar probabilidades o frecuencias relativas; 4) para calcular números-índice, y 5) para normalizar ciertas distribuciones.

Comparabilidad de las distribuciones de clase, con independencia del intervalo

Supongamos dos distribuciones de clases que tienen, cada una, un intervalo diferente. Si calculamos las desviaciones con respecto a un promedio determinado, y queremos comparar la distribución de esa serie de desviaciones (o sea, la forma en que —en cada caso— ciertas frecuencias se asocian con esas desviaciones) tropezaremos con una dificultad, pues existen dos fuentes simultáneas de variación: 1) la de las frecuencias mismas, y 2) la de las diferentes formas de clasificación (o sea la debida a la diferente magnitud de los intervalos).

Para facilitar la comparación, lo que se hace, en tales casos, es dividir cada desviación entre el intervalo común de la serie de clases a la que corresponde; así, si la primera serie tuvo las desviaciones: —15, —10, —5, 0, 5, 10 y las frecuencias asociadas fueron: 7, 5, 3, 1, 2, 4, y la segunda tuvo las desviaciones: —9, —6, —3, 0, 3, 6 y las frecuencias asociadas fueron también: 7, 5, 3, 1, 2, 4, podemos dividir las primeras desviaciones entre 5, intervalo de clase de la primera serie, y las segundas desviaciones entre 3, intervalo de clase de la segunda, y obtener, en ambos casos: —3, —2, —1, 0, 1, 2. Como las frecuencias son las mismas, y las desviaciones (en unidades del intervalo) son iguales, puede concluirse que las dos distribuciones, sin ser iguales, son de la misma forma o tipo.

Reducciones de medidas n-dimensionales a otras de dimensión cero

Las distribuciones estadísticas varían entre sí —muy principalmente— por la diferente concentración de sus valores en torno de un promedio central. Pero, aún antes, la posibilidad de comparar dos o más curvas en razón de su mayor o menor concentración (menor o mayor dispersión) está afectada por la diferencia de magnitud que el mismo promedio tiene para cada una de esas curvas. A fin de tener magnitudes estrictamente comparables de las medidas de variabilidad, hay que expresar éstas como una relación entre la medida de variabilidad misma (generalmente la desviación media cuadrática) y el promedio (la media aritmética, por lo general). El relativo correspondiente (generalmente expresado como porciento) se conoce como "coeficiente de variabilidad" y es, por tanto, igual a 100 $[d_2 / x]$ o 100 $[\sigma / \bar{x}]$. Esta medida indica qué tan variable es la distribución con respecto a la magnitud de su media aritmética y, por tanto, permite comparar dos o más distribuciones, en lo que se refiere a su concentración-dispersión, con independencia de la magnitud de sus medias (o *como si* sus medias fuesen iguales).

En la misma forma en que se busca asegurar la comparabilidad de las distribuciones por encima de las diferencias de magnitud de las medias, se logra la comparabilidad de las distribuciones en materia de asimetría y con independencia de la magnitud de la variación estadística.

Antes, conviene que exploremos cómo se puede relativizar la otra medida de variación estadística (la variancia). Ésta tiene mayor parecido con la medida de asimetría que el que tiene con esta última la desviación media cuadrática.

Si al relativo $\sigma / \bar{x}$ lo elevamos al cuadrado, la desviación cuadrática media se convertirá en la variancia o segundo momento, y la media aritmética aparecerá elevada al cuadrado: $\mu_2 / (\bar{x})^2$. O sea, que tendremos una nueva medida de variabilidad relativa (como el coeficiente de variación) que será igual al cuadrado del coeficiente de variabilidad. Para obtener este último, sólo se necesita extraer la raíz cuadrada del segundo momento en unidades del cuadrado de la media.

La medida pearsoniana de asimetría

Sobre la base del cálculo de una medida de variabilidad igual a $[\mu_2 / (\bar{x})^2]^{1/2}$ podemos diseñar una medida de asimetría que se base no en el segundo, sino en el tercer momento (μ_3) y que se relacione no con la media, sino con la medida inmediata anterior (con σ, o con μ_2). Si relacionamos a μ_3 con (σ), tendremos que recordar que, en esta última, los valores son de dimensión 1 (pues a las desviaciones se les eleva al cuadrado, pero después se les extrae raíz cuadrada, o sea que se elevan a

$2 \times 1/2 = 1$), mientras que μ_3 tiene tercera dimensión (pues las desviaciones se elevan en él a 3 y no se les extrae ninguna raíz). Como la relación debe plantearse entre valores *de igual dimensión*, para que los resultados tengan dimensión cero, la medida correspondiente se obtendrá de dividir μ_3 entre σ al cubo. Así, al dividir μ_3 (de tercera dimensión) entre σ^3 se obtiene el resultado apetecido; de orden cero, de primera dimensión (pues se trata de un número abstracto).

Si se trabaja con μ_3 y μ_2, como el primero se coloca en el nivel 3 y el segundo en el 2, la solución consiste en elevar a μ_3 a la potencia 2 (se eleva el nivel 6) y en elevar a μ_2 a la potencia 3 (con lo que también se eleva al nivel 6). Al dividir, se obtiene el resultado apetecido en cuanto a orden y dimensión: un número absoluto que mide la asimetría de las distribuciones *independientemente* de la mayor o menor concentración-dispersión de los datos.

La medida pearsoniana de curtosis

Pearson define una medida similar de curtosis, que es independiente de la concentración-dispersión de las distribuciones. Se obtiene a partir del cuarto momento. Nuevamente, es un relativo calculado respecto de σ o de μ_2. En el caso de emplear σ como término de relativización estadística, habrá que elevarla a la cuarta potencia a fin de compensar la elevación a la cuarta de los valores de los que procede μ_4. Si se emplea μ_2 para esa relativización habrá que elevarla simplemente al cuadrado, ya que la segunda potencia de los valores de los que procede, multiplicado por 2 da el 4 (compensador del cuarto nivel en que se coloca μ_4).

Si se sigue el proceso lógico impuesto por los desarrollos previos, es fácil comprender que más que esta medida pearsoniana de curtosis podría interesar otra, también relativa, en la que se considera no la dispersión sino el carácter inmediatamente previo (la asimetría). En estas condiciones, la medida relativa que propondríamos estaría dada, nuevamente, por la relación matemática entre el cuarto momento y el tercero. Para compensar las elevaciones al cuarto y al tercer nivel, multiplicaríamos el primero por 3, y el segundo por cuatro, y la medida estaría dada por μ_4^3 / μ_3^4.

Determinación de probabilidades o frecuencia relativa

Dos o más distribuciones estadísticas pueden *parecer* distintas por el hecho de que sus factores (número de casos) sean diferentes; pero pueden *ser* en realidad, iguales, dentro de su diferencia de escala. A fin de hacer comparables dos distribuciones de efectivos diferentes, con independencia de sus escalas, se pueden reducir las frecuencias (absolutas) a frecuencias relativas (conocidas también como probabilidades).

Para convertir una frecuencia absoluta en una relativa, se divide dicha frecuencia entre el efectivo de la distribución (o sea, entre la suma de las frecuencias).

En estas condiciones, la suma de las frecuencias *relativas* o probabilidades, resulta igual a la unidad (como puede comprobarse fácilmente si se considera que como cada una está dividida entre N, y que su suma, que es precisamente el efectivo N, al dividirse por sí misma, se reduce a 1).

La escala sigmática

Otra forma de reducir a términos comparables dos o más distribuciones distintas consiste en expresar cada una de ellas en términos de la desviación media cuadrática. Para esta reducción de escalas, se usan —generalmente— las desviaciones con respecto a la media aritmética.

En estas condiciones, la escala sigmática de una serie se obtiene: 1) calculando su media aritmética; 2) restándola de las marcas de clase para obtener desviaciones con respecto a la media aritmética; 3) calculando la desviación cuadrática media, y 4) dividiendo cada desviación entre la desviación cuadrática media. Así, se obtiene una serie de valores asociados a un conjunto de frecuencias, que se puede comparar con otra serie de valores análogos, asociados a otro conjunto de frecuencias (que procede de otra distribución). La comparación entre ambas distribuciones (nuevas) se establece, ahora, con independencia de sus distintas variabilidades.

Estandarización de tasas por comparación con una distribución normal

Números-índice

Para comparar los valores alcanzados por una variable (o por un conjunto de variables) en dos fechas, en dos lapsos de tiempo, en dos lugares o en dos instituciones sociales diferentes, se recurre a lo que se conoce como números-índice.

Los números-índice son, fundamentalmente, relativos o promedios de relativos. La relación se establece entre cada uno de los valores que adquiere el fenómeno en una fecha o institución y el que tuvo en otra fecha, otro lugar u otra institución que se considera como "base".

Si se trata de una sola variable x, y ésta alcanza un valor x_1 en un año, en un lugar, en una institución y un valor x_0 en otro año, otro lugar o institución considerada como base, el número-índice más simple es el relativo de x_1 dividido entre x_0 (x_1/x_0).

Cuando no hay una sola variable, sino un conjunto de variables u_1, v_1, w_1 en un lugar, año o institución, y esas variables alcanzaron los valores u_0, v_0, w_0 en otro lugar, año o institución, se pueden obtener varios números-índice aislados, dividiendo u_1 entre u_0, v_1 entre v_0, w_1 entre w_0. El problema que se plantea, en seguida, es el de cómo conjuntar estos diversos valores. Antes, se ofrece también la alternativa de conjuntar todos los valores correspondientes al año, lugar o institución de estudio, y relacionarlos estadísticamente (dividiendo el resultado) entre lo que se obtenga de conjuntar los valores correspondientes al año, lugar o institución de base.

Según esto, podemos sumar $u_1 + v_1 + w_1$ y dividir la suma entre la suma de $u_0 + v_0 + w_0$; o podremos dividir u_1 / u_0, v_1 / v_0, w_1 / w_0 y sumar los cocientes: $(u_1 / u_0) + (v_1 / v_0) + (w_1 / w_0)$.

Si representamos a los valores del año de estudio por p_1 y a los del año base por p_0, podemos decir que las alternativas de número-índice son: $\Sigma p_1 / \Sigma p_0$ o $\Sigma (p_1 / p_0)$. Al primero se le puede designar como "índice relativo de sumas", y al segundo como "índice suma de relativos".

Los dos índices esquematizados antes pueden parecer iguales; pero son, en realidad, distintos. El relativo de sumas mide las variaciones del conjunto de todas las variables de un año, lugar o institución, frente a otro. La suma de relativos mide el conjunto de las variaciones de cada variable de un lugar o institución, a otro.

Pero, los números-índice deben ser independientes del número de variables que intervengan; de ahí que los resultados anteriores deban dividirse entre el efectivo o número de variables correspondientes.

Es fácil ver que, especialmente cuando se divide la suma de los relativos $(\Sigma p_1 / p_0)$ entre N, se obtiene una media aritmética. Esto, a su vez, sugiere fácilmente, la posibilidad de emplear otras medias distintas para el cálculo de los números-índice. En forma destacada se puede usar o una media geométrica o una media armónica.

En el caso de la media armónica, se tomarán: 1) los recíprocos de los relativos $(p_1 / p_0)^{-1}$ * en seguida; 2) se sumarán esos recíprocos; 3) se dividirá la suma entre el efectivo N (número de variables consideradas), y 4) se tomará el recíproco del resultado, o sea: $N / \Sigma (p_0 / p_1)$, media armónica de los relativos. Éste será el número-índice buscado.

En el caso de la media geométrica o logarítmica: 1) se tomarán los logaritmos de los relativos (logaritmo de p_1 / p_0) (o sea, que se restará del logaritmo p_1 el logaritmo p_0); 2) se sumarán los resultados; 3) se dividirá la suma entre el efectivo, y se tomará; 4) el antilogaritmo del cociente. Ése será el número índice buscado.

Ajuste previo de series cronológicas

Los relativos también son útiles en el ajuste previo de las series cronológicas. Éste se realiza en dos sentidos: 1) como ajuste deflactario, y 2) como ajuste calendárico.

* O, puesto que el recíproco de una fracción es otra fracción que tiene por numerador el denominador de la primera (p_0) y por denominador el numerador de la primera (p_1), (p_0/p_1).

En efecto, hay muchas ocasiones en que los cambios que se producen a través del tiempo aparecen deformados debido a que —por ejemplo— se encuentran expresados en términos monetarios y, en el lapso correspondiente, se ha modificado el valor adquisitivo de la moneda; en tales ocasiones procede un ajuste previo de la serie cronológica que se conoce como ajuste deflactario.

Por otro lado, en el estudio de las variaciones estacionales que se producen en un año y, más particularmente, en las que se producen en un fenómeno de uno a otro mes, influye el hecho de que nuestro calendario actual cuenta con meses de diferente número de días; por eso se impone lo que se conoce como ajuste calendárico. Ambos ajustes se realizan por relativización. En el primero, la relativización se hace dividiendo entre un "número-índice de precios"; en el segundo, entre un "número-índice de días en el mes".

Ajuste deflactario

Para ajustar deflactariamente los datos de una serie, se usan 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 otro que se toma como base de comparación.

Afirmar que a un año dado le corresponde un índice de 1 o de 100 equivale a indicar que el sistema de precios de ese año es (en conjunto) el mismo 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 parecida, 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.

En el primer supuesto, o sea, en el de un año con índice de 1 o 100%, el ajuste deflactario tendrá que dejar sin alteración el dato registrado originalmente. En el caso de un índice inferior a 1 o a 100%, el hecho se interpretará en el sentido de que, de no haber sido por la variación conjunta del sistema general de precios, el valor o magnitud del fenómeno, para esa fecha, hubiese sido mayor que el valor registrado, y, de que por ello, la función del ajuste deflactario debe consistir en elevar la cifra registrada. En el tercer supuesto (o sea, cuando el índice es superior a 1 o a 100%) se deberá pensar que, de no haber sido por la variación conjunta del sistema general de precios, el valor o magnitud del fenómeno para esa fecha hubiese sido menor que la registrada, y el ajuste deflactario deberá dar dicha cifra, menor que la registrada.

Matemáticamente esos resultados se pueden obtener si se expresan los datos originales como relativos o porcientos del número-índice correspondiente, o sea, si se les divide 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 se puede concretar en la siguiente forma: valor deflactado = (valor no deflactado) / número-índice:

Ajuste calendárico

El segundo de los ajustes previos, importantes en el estudio de las series cronológicas, es el ajuste calendárico. Mediante él, se eliminan los efectos que —sobre las series— tienen las diferencias de los periodos de tiempo marcados por el calendario civil. Éstos, que, de acuerdo con su denominación común ("mes") podrían parecer equivalentes, en realidad, no lo son.

Así, el ajuste calendárico debe comenzar por considerar si el año es o no es bisiesto. En seguida debe considerar si dos periodos del año, a pesar de su común denominación de "meses", comprenden o no un mismo número de días. Finalmente, debe examinar, si cada uno de dichos meses tienen o no un mismo número de días *laborables* y laborados. *

Aquí sólo nos referiremos al más general y sencillo de los ajustes calendáricos: el que considera el carácter, bisiesto o no, del año, y la diferencia en el número de días de los distintos meses del año.

El ajuste calendárico se puede hacer de acuerdo con dos procedimientos: el primero es previo a la recolección de datos; el segundo es simultáneo a su elaboración. El primero consiste en elaborar un calendario dividido en unidades más regulares que las del actual (consistentes, por ejemplo, en 13 periodos de 28 días cada uno) y con uno o dos días que se quedan sin registrar en los años no bisiestos y en los bisiestos. El segundo es el que aquí nos interesa, en cuanto ejemplifica el uso de los relativos en estadística.

El procedimiento de ajuste aplicable a los datos ya registrados responde a las mismas necesidades y consideraciones que el ajuste deflactario. El razonamiento es como sigue:

1º Si, prescindiendo de consideraciones prácticas inmediatas, hubiésemos de pensar en un año dividido en 12 periodos que fueran exactamente iguales por su número de días, ¿cuántos días corresponderían a cada uno de esos periodos? La respuesta, en el caso de un año no-bisiesto sería: 365/12 = 30.41667 días. En el caso de uno bisiesto, sería: 366/12 = 30.5 días.

2º Si tomamos como base la duración de este mes teórico o ideal, y calculamos sobre él un número-índice de 1 o 100%, un mes que tuviera más días, tendría un número-índice superior a 1 o 100%, y uno que tuviera menos días, uno inferior a 1 o 100%.

* Debido al número de domingos, de sábados (para la "semana inglesa"), de fiestas fijas y móviles, en el mes, etcétera.

3º De acuerdo con lo anterior, a ninguno de los meses en que está dividido el año actual (de 28, 29, 30 y 31) le corresponde un índice de 1 o 100%. A enero, marzo, mayo, julio, agosto, octubre y diciembre les corresponde un número-índice superior a 1 (por tener 31 días). A abril, junio, septiembre y noviembre, uno inferior a 1 (por tener 30 días), y a febrero uno inferior al de los meses anteriormente mencionados (por tener 28 o 29 días según sea el año bisiesto o no).

En forma parecida a como hicimos en el ajuste deflactario, en este caso el ajuste se hará por división del número de días del mes entre el número-índice correspondiente al mes de que se trate; o sea que: los datos ajustados calendáricamente son iguales a los no ajustados divididos entre el número-índice de días en el mes, calculado con respecto al mes promedio del año correspondiente.

Los números-índices para el ajuste resultan ser:

	Año no bisiesto	*Año bisiesto*
Número medio de días por mes	30.41667 = 1	30.5 = 1
Índice para los meses de 31 días	31/30.41667 = 1.017177	31/30.5 = 1.016393
Índice para los meses de 30 días	30/30.41667 = 0.986304	30/30.5 = 0.983606
Índice para el mes de febrero	28/30.41667 = 1.920547	29/30.5 = 0.950819

Conceptos y operaciones indispensables para la inferencia estadística

Noción de probabilidad

La estadística se basa en la teoría de la probabilidad. Lo probable ubica un acontecimiento entre la incertidumbre total y la plena certeza. Las escalas en las que se ubica cada acontecimiento, de acuerdo con su mayor o menor probabilidad, van generalmente de menos infinito a más infinito, o de cero a uno.

La probabilidad como frecuencia relativa

La probabilidad se suele expresar como frecuencia relativa. Esto es posible sólo si se considera que cada acontecimiento es igualmente probable, en comparación con los demás. La definición correspondiente afirma lo siguiente:

Si hay n posibilidades: 1) igualmente probables; 2) mutuamente exclusivas, y 3) que agotan todos los acontecimientos posibles, y de esas probabilidades m favorecen al acontecimiento A, la relación matemática entre m y n (m/n) es la probabilidad de A.

La relación m/n representa la frecuencia con que puede aparecer A en comparación con el total de frecuencias; o sean: los casos en los que se producen: o el acontecimiento A o los acontecimientos complementarios de A (designados en forma genérica como los acontecimientos que no son A, o los acontecimientos no-A).

Modalidades de la probabilidad

La probabilidad de un acontecimiento se puede considerar en sí misma, sin hacer referencia a otras probabilidades de otros acontecimientos; pero también se la puede considerar como probabilidad de que ese acontecimiento se produzca junto con otro, o bien de que uno de esos acontecimientos ocurra cuando o con la condición de que ocurra otro. A cada una de estas modalidades de considerar la probabilidad se le conoce: 1) como probabilidad simple; 2) como probabilidad conjunta, y 3) como probabilidad condicionada.

Simbología

Las probabilidades se suelen representar mediante una letra pe mayúscula (P) que se hace preceder a un paréntesis en el que se encierran las letras que representan a los acontecimientos cuya probabilidad se estudia. Cuando se trata de representar la probabilidad conjunta de dos acontecimientos, las letras que les corresponden se incluyen en el paréntesis, separadas entre sí por una coma. Cuando se trata de representar la probabilidad condicionada, las letras que representan a los acontecimientos respectivos se incluyen en el paréntesis, colocando en primer término el acontecimiento condicionado y en seguida el condicionante (la condición), separando a uno del otro por medio de una línea vertical. Así, se tiene:

$P(A)$	"probabilidad de que ocurra A" (probabilidad simple).
$P(A,B)$	"probabilidad de que ocurran A y B" (probabilidad conjunta).
$P(A\|B)$	"probabilidad de que ocurra A si B ocurre" (probabilidad condicionada).

La probabilidad total es unitaria

Puesto que se ha definido la probabilidad de un acontecimiento A como la relación entre la frecuencia con que se produce o puede producirse (m) y la frecuencia con la que se producen tanto ese acontecimiento como los demás (igualmente probables, mutuamente exclusivos y exhaustivos), si se suman todas las probabilidades de los acontecimientos parciales se obtendrá una fracción cuyo denominador será el común a las probabilidades de cada acontecimiento (o sea, el total de frecuencias n) y cuyo numerador estará dado por la suma de las frecuencias de cada acontecimiento. O sea, que el resultado será una fracción que tiene por denominador a n, y por numerador una serie de términos que, sumados producen n. En último término, que la suma de todas las probabilidades produce como resultado uno. Es esto lo que significa decir que la probabilidad total es unitaria.

De otro modo: si A, B y C son tres acontecimientos que agotan todas las posibilidades, son mutuamente exclusivos e igualmente probables, y si A aparece 3 veces, B aparece 5 y C, 2, esto significa que el total de frecuencias es 10 $(3 + 5 + 2)$, que la probabilidad de A es $3/10$, la de B es $5/10$ y la de C es $2/10$. Si se suman esas probabilidades se obtiene, como probabilidad total: $(3/10) + (5/10) + (2/10) = 10/10 = 1$.

Probabilidad conjunta y probabilidad condicionada

La probabilidad conjunta de dos acontecimientos (A, B) depende de dos factores: 1) la probabilidad condicionada que hay de que ocurra uno de los acontecimientos *si* ocurre el otro, y 2) la probabilidad simple que hay de que ocurra ese otro, condicionante suyo. O sea, que la probabilidad conjunta es igual al producto de la probabilidad condicionada, por la probabilidad simple del condicionante.

$$P(A,B) = P(A|B)\,P(B).$$

En la forma en que hemos expresado la relación entre probabilidad condicionada y probabilidad conjunta se cubren dos y no una sola de las posibilidades de expresar una probabilidad conjunta de dos acontecimientos, pues también es cierto que:

$$P(A,B) = P(B|A)\,P(A).$$

Disyunción de probabilidades

La probabilidad que hay de que ocurran A o B separadamente (y no en forma conjunta) es igual a la probabilidad que hay de que ocurra A de por sí, más la que hay de que ocurra B de por sí, menos la probabilidad de que ocurran ambas.

Expectativa de una función

Una variable tiene cierta "probabilidad de ocurrir". De una función se dice que "tiene cierto valor esperado o expectativa". A esa "expectativa de la función" se la define como el valor probable de esa función a través de una serie de ensayos. Más concretamente, la expectativa o valor esperado de una función es su valor promedio.

Si representamos por x_i una variable y damos a i todos los valores que puede asumir dentro de un dominio determinado (de 1 a r, por ejemplo) se obtienen los valores $x_1, x_2 \ldots x_r$ de dicha variable. Por otro lado, si la probabilidad que tienen de ocurrir dichos valores se representan por $P(x_1)$, $P(x_2) \ldots P(x_r)$ la expectativa de x_i (que se representa por $E(x_i)$) estará dada por la suma de los productos de cada valor de x_i por su probabilidad $P(x_i)$:

$$E(x_i) = \sum_{i=1}^{r} x_i P(x_i)$$

O sea, que la expectativa es igual a una suma de los productos; los productos formados al multiplicar cada valor de la variable (o de la función) por su probabilidad; todo dentro del dominio o ámbito de variabilidad (de $i = 1$ a $i = r$) en este caso.

Funciones de frecuencia

Las probabilidades se suelen representar, en forma sintética, mediante funciones. Como las probabilidades son asimilables a las *frecuencias relativas* de un acontecimiento, a las funciones que se refieren a las probabilidades de un acontecimiento se las designa con el nombre de "funciones *de frecuencia*".

De este modo, la probabilidad que hay de que x esté entre dos valores límite (a y b) está dado: *a*) en el caso de las distribuciones discontinuas por la suma de los productos de los valores que x puede adquirir (x_i's) entre esos límites ($\sum_{i=1}^{r}$), y *b*) en el caso de las distribuciones continuas, por la integral ($\int$) entre esos límites ($\int^b_a$) de la función de x con respecto a x (dx).

$$\int_a^b p(x)\, dx$$

Nosotros preferiríamos representar esto por:

$$I_{x(a,b)}\, p(x)$$

Si f (x) es una función de x y p (x) es una función de probabilidad, la expectativa estará dada por la sumatoria o la integral de sus productos:

$$E[f(x)] = \int_{-\infty}^{\infty} f(x)\, p(x)\, dx$$

esto es otra forma de expresar lo que ya habíamos dicho previamente.

Función de distribución

Para salvar la diferencia de formas entre las distribuciones continuas y las discontinuas se usa la "función de distribución".

$$F(x) = \int_{-\infty}^{\infty} p(x)\, dx$$

Especificación de las distribuciones

Para pasar de la expresión genérica de una distribución a la específica (o sea, para especificar una distribución concreta) en la práctica (en las aplicaciones) se utilizan los "momentos".

Los momentos son las expectativas de las diferentes potencias de las desviaciones de la variable, respecto a un valor fijo (llamado "centro de momentos").

Los momentos que aparecen en forma más inmediata son los que toman como centro de momentos al origen de la distribución. Son más comunes los que toman como centro la media aritmética de la distribución. Pero, cuando, además, se requiere obtener medidas que sean independientes de la escala o de las unidades originales de medida, esos momentos se dividen entre las potencias del segundo momento con respecto a la media aritmética que correspondan al orden del momento de que se trate.

Función momentógena o generadora de momentos

La expansión en serie de la función exponencial de xt (o sea de e, base de los logaritmos naturales, elevada al producto de x por t) es:

$$e^{xt} = 1 + xt + \frac{x^2t^2}{2!} + \frac{x^3t^3}{3!} + \cdots$$

La expectativa de esta función se obtiene si a cada término de la misma se le antepone el sumador. Al anteponer el sumador a 1, se obtiene 1 (pues es como si se sumaran las potencias de orden cero de x); al anteponerlo al término en x, se obtiene la suma de las x's por t, o sea, el primer momento de x, por t; al hacerlo con el término en x^2 se obtiene la suma de los cuadrados de x por el cuadrado de t entre factorial 2, o sea, el segundo momento de x por el cuadrado de t entre factorial 2. O sea, que si sustituimos en la anterior, podremos escribir:

$$E[e^{xt}] = 1 + ut + u_2 \frac{t^2}{2!} + u_3 \frac{t_3}{3!}$$

O sea, que los coeficientes de las potencias de t son los momentos de primero, segundo, tercer grado, etcétera. De ahí que a la exponencial e elevada a xt se le llame "función generadora de momentos" o "función momentógena".

Funciones generadoras de sumas de variables

Entre la función momentógena de una suma de variables y las funciones momentógenas de esas variables hay una relación definida. La hay también entre la función cumulantógena de la suma y las funciones cumulantógenas de los sumandos.

La función momentógena de una suma de variables es igual al producto de las funciones momentógenas de esas variables.

La función cumulantógena de una suma de variables es igual a la suma de las funciones cumulantógenas de esas variables.

Estos resultados eran de esperarse ya que los cumulantes son, en nivel logarítmico, lo que los momentos en el nivel no logarítmico, pues, como se sabe, al pasar del nivel no logarítmico al logarítmico, los productos se convierten en sumas.

Función característica

Como hay ciertos casos, excepcionales, en los que la función momentógena no existe o no da lugar a una serie convergente, se deja que t adquiera valores imaginarios. Se usa, entonces, la función e elevada a ixt (en donde i representa la unidad imaginaria, igual a la raíz cuadrada de —1). A la expectativa de esta nueva función se le llama "función característica".

Función característica de la suma de variables

Como la función característica no es sino una variante de la momentógena (en el sentido de que se establece en el mismo nivel no logarítmico que ésta) es fácil anticipar el resultado siguiente:

La función característica de la suma de dos variables es igual al producto de las funciones características de dichas variables (siempre y cuando estas variables sean independientes entre sí).

Distribución de medidas estadísticas derivadas

La distribución de estadísticas derivadas es de particular interés para la teoría del muestreo. Muy particularmente, interesa determinar cuál es la distribución de las medias aritméticas de las muestras de una distribución conocida, o cuál es la distribución de sus desviaciones medias.

Hay varios métodos para derivar la distribución de determinada medida estadística a partir de una distribución-matriz, primitiva. Tres son los métodos principales: 1) el de sustitución de variable; 2) el procedimiento de sumas, y 3) el procedimiento probabilitario.

1º El de sustitución de variable, se emplea si se conocen:

1. La forma de la distribución, y
2. La forma matemática en que se relacionan la variable nueva y la vieja.

En ese caso, la distribución derivada se puede obtener de la primitiva por sustitución. Esa sustitución debe abarcar, en caso necesario: integrador e integrando, o sea: los límites y el referente de la integración, por una parte, y el integrando, por otra (si se usa nuestra terminología) o (si se prefiere la tradicional) los límites de integración y tanto la porción no diferencial como la diferencial del integrando.

2º El procedimiento de sumas, se utiliza si se conoce la distribución de cada una de las funciones de una serie y se quiere conocer la distribución de la suma de esas funciones. En tal caso, se usan las propiedades ya establecidas de las funciones momentógenas, cumulantógenas y características de las distribuciones de sumas de variables.

3º El procedimiento probabilitario, se emplea si se conoce una combinación apropiada de funciones probabilitarias (la condicionada y la condicionante, por ejemplo) y se desea encontrar otra función probabilitaria relacionada con ellas (la conjunta, en el ejemplo).

Estos tres procedimientos para encontrar las distribuciones de estadísticas derivadas, como indica Quenouille, a quien seguimos en esto, no introducen ningún concepto nuevo en estadística pero, en la práctica pueden requerir operaciones matemáticas un tanto complicadas.

La media del producto de dos estadísticas independientes entre sí

Mediante los procedimientos ya delineados, se puede saber cuál es la relación que liga a la media aritmética del producto de dos variables con las medias de dichas variables. Esa indagación llega a la conclusión de que:

La media del producto de dos variables es igual al producto de las medias de dichas variables.

Variancia del producto de dos medidas estadísticas independientes

En forma parecida, se puede llegar a establecer que:

La variancia del producto de dos variables es igual a: 1) el producto de las variancias de las variables factores, más 2) dos términos, formados por los productos de la variancia de cada factor por el cuadrado de la media aritmética del otro:

$$u_{2x}\, u_{2y} + (u_{2x}\, \bar{x}^2{}_{1y} + u_{2y}\, \bar{x}^2{}_{1x})$$

Función cumulantógena de una suma de gran número de variables

Cuando las variables que se suman no son dos sino muchas, la función cumulantógena de la suma sigue siendo igual a la suma de funciones cumulantógenas de las variables, pero, en esa suma, los términos cuyo grado es superior al segundo van siendo relativamente despreciables.

Esto mismo se puede expresar diciendo que cuando el número de variables que se suman crece indefinidamente, la función cumulantógena de su suma tiende a convertirse en una simple función cuadrática, pues los términos superiores al del segundo grado se pueden despreciar.

La cuadrática más general, cumulantógena, es:

$$K(t) = \bar{x}_1 t + \frac{u_2\, t^2}{2}$$

La momentógena da:

$$u_2 = \sigma^2 \; ; \; u_3 = 0 \; ; \; u_4 = 3\, \sigma^4$$

La función característica es:

$$C(t) = e^{iut - \frac{\sigma^2 t^2}{2}}$$

La función de frecuencia es:

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-\bar{x})^2}{2\sigma}}$$

Ésta es la llamada "distribución normal".

El hecho de que la función cumulantógena de la suma de un gran número de variables tienda como límite a una cuadrática se conoce como teorema central del límite.

La estadística tiene que ver, principalmente, con el problema inductivo de hacer afirmaciones sobre los colectivos a partir de sus muestras. Pero, es por deducción como se llegan a determinar: 1) los procedimientos de estimación; 2) las pruebas de significación, y 3) la aceptabilidad de las hipótesis.

Estimación

Una medida estadística que se ha obtenido de una muestra apunta hacia la medida correspondiente del colectivo de la que se extrajo esa muestra. Cuando a una medida muestreal se la usa para inferir el valor de la medida del colectivo correspondiente, se la designa con el nombre de "estimador", y a la medida cuyo valor se estima se la puede llamar "estimada".

Un estimador coincide con la estimada sólo en casos excepcionales pero, es deseable que conforme aumente el número de observaciones, el estimador tienda a la estimada (que sea "consistente"); que la expectativa de los estimadores iguale al valor de la estimada (que sea "insesgado" el estimador); que el estimador se aproxime a la estimada tanto como se pueda (que sea "eficiente") y que agote toda la información disponible en la muestra (que sea "suficiente").

Estas diferentes propiedades de un estimador están interrelacionadas; sin embargo, no todas ellas son reversibles; así, los estimadores insesgados y suficientes son los más eficientes; los eficientes son consistentes; pero, las afirmaciones inversas de las anteriores no son —en cambio— verdaderas, y un estimador insesgado puede carecer de todas las otras propiedades.

Pruebas de hipótesis

El más simple de los procedimientos de decisión estadística consiste en determinar si una hipótesis es aceptable o no.

Generalmente, con respecto a un hecho o a un acontecimiento, son varias las hipótesis que se pueden postular. Esto es particularmente cierto en relación con la asociación o falta de asociación de dos o más variables. La hipótesis que afirma que no existe asociación entre las variables se suele designar como "hipótesis nula" o "hipótesis cero"; las hipótesis que postulan la existencia de asociación (y de diferentes modalidades de asociación) entre las variables se designan con el nombre de "hipótesis alternativas".

En las pruebas de hipótesis, la estadística lo único que hace es refinar —mediante la expresión numérica— un procedimiento científico que se ha practicado siempre y, de acuerdo con el cual una hipótesis o ley se mantiene mientras no hay una observación que, por lo extremada, permita dudar de ella e imponga su sustitución por otra más amplia que abarque o explique ese extremo. O sea, que una hipótesis se sigue aceptando mientras sigue siendo "pequeña" la probabilidad de que ocurran esos casos extremos.

Qué tan pequeña deba ser esa probabilidad es algo que se determina a través de los llamados niveles de significación; así, una probabilidad de 0.05 se considera como un nivel "sugestivo o indicativo"; uno de 0.01 se considera como "significativo" y uno de 0.001 como "altamente significativo.

El nivel que se acepte para rechazar la hipótesis nula depende de las consecuencias que puedan tener: por una parte, el rechazo incorrecto; por otra, la incorrecta aceptación. O sea, que el nivel tiene que establecerse mediante una transacción entre la cautela y la audacia del investigador.

Límites de confianza

Las observaciones que se hacen en una muestra son razonablemente representativas de un parámetro desconocido dentro de ciertos límites. A estos límites se les llama "límites de confianza". Son ellos los que demarcan cuál es la amplitud dentro de la cual los valores obtenidos se pueden considerar como razonablemente representativos: el grado en que los resultados muestreales son representativos del universo o población.

Los límites de confianza dependen del nivel de significación que se haya aceptado; de ahí que cuando se mencionen unos límites se deba señalar —asociado a ellos— el coeficiente de confianza que permitió fijarlos. En efecto, conforme aumenta ese coeficiente, la amplitud o distancia entre los límites es mayor; es menor —en forma correspondiente— la exigencia que se ejerce sobre los valores, para reconocerlos como aceptables.

Los errores en las pruebas de hipótesis

Al docimar (o someter a prueba una hipótesis), se pueden cometer dos tipos de error: 1) el error de rechazar una hipótesis, a pesar de ser verdadera, y 2) el error de aceptar una hipótesis cuando es falsa. Correlativamente, los aciertos posibles son: 1) el de aceptar una hipótesis verdadera, y 2) el de rechazar una hipótesis falsa.

Selección de pruebas

A cada hipótesis se le puede someter a varias pruebas; de ahí que se necesite de un criterio para seleccionar de entre todas las pruebas de hipótesis posibles, la que resulta más conveniente en cada caso. Para elegir esa prueba, hay que determinar cuál es aquella prueba, 1) que, para una magnitud dada (admitida, fija), de uno de los dos tipos de error, hace que 2) sea mínimo el otro tipo de error; o sea:

1. Que para una magnitud fija determinada del error de rechazo de la hipótesis correcta, hay que buscar cuál es la prueba que hace mínimo el error de admitir la hipótesis incorrecta.
2. Que si se admite una magnitud dada para el error de aceptar la hipótesis incorrecta habrá que buscar cuál es la prueba que minimice el error de rechazar la hipótesis correcta.

A la probabilidad que una prueba tiene de rechazar correctamente la hipótesis nula se le llama su "poder" o su "potencia".

Verosimilitud

Dentro de la docimacia de hipótesis, el problema principal consiste en descubrir un método que permita encontrar aquellas pruebas que siendo de máxima potencia, sean —además—, en forma uniforme, de potencia máxima.

Para esto hay que considerar la frecuencia relativa con que un conjunto dado de observaciones puede surgir bajo diferentes hipótesis. Esto equivale a relacionar entre sí probabilidades condicionadas.

P(A'H) es la probabilidad de A en caso de ser cierta la hipótesis H,

P(A'H') es la probabilidad de A en caso de ser cierta la hipótesis H'.

Antes se consideró la relación entre probabilidades condicionadas que correspondían a diferentes acontecimientos (A, C) bajo la misma condición (B) o sea P(A'B) y P(C'B). Ahora se considera una relación entre probabilidades condicionadas de un mismo acontecimiento (A) bajo condiciones diferentes (H en un caso, H' en el otro).
Dentro de este contexto,

P(A'H) se designa como la verosimilitud de A bajo la hipótesis H.

P('A'H) se designa como la verosimilitud de A bajo la hipótesis H' y, a

P(A'H) / P(A'H') se le da el nombre de "razón entre verosimilitudes", "ratio verosimilitud" o "verosimilitud relativa".

La verosimilitud absoluta es numéricamente igual a la probabilidad; pero, conceptualmente, corresponde a un sistema diferente al de las probabilidades propiamente dichas.

La elección de las pruebas de significación

La búsqueda de la prueba de significación de máxima potencia se puede hacer mediante la comparación matemática de las verosimilitudes; pero, la ratioverosimilitud sólo refleja la plausibilidad de las hipótesis cuando éstas son simples.

Cuando una de las hipótesis es compleja, por lo general no se pueden encontrar pruebas que siendo de máximo poder, lo sean en forma uniforme. Así, hay casos en los que mientras unas son monotónicamente crecientes, otras son monotónicamente decrecientes, con lo que no hay medida estadística que actúe como indicador uniforme de la aceptabilidad relativa de las dos hipótesis, y no hay prueba que sea, uniformemente, de poder máximo.

2. ALGUNAS TÉCNICAS ESTADÍSTICAS DE APLICACIÓN SOCIAL INMEDIATA

2.1. ALGUNAS TÉCNICAS DE OBTENCIÓN DE PROMEDIOS

El modo

En el caso de las llamadas series nominales (o sea, aquellas en las que las categorías están dadas por nombres), el "modo" es el promedio más adecuado y más fácil de calcular (o, para ser más precisos, de determinar).

El modo es aquella categoría de la serie que se presenta con mayor frecuencia.

Se suele decir, también, que modo es la categoría de la serie a la que corresponde la frecuencia máxima, y a dicha frecuencia se le designa, en forma correspondiente, como "frecuencia modal".

Para determinar el modo de una serie se ve cuál, de entre todos los valores de la distribución que corresponden a las diferentes categorías de la serie, es el mayor. Ese valor es la frecuencia modal. Frente a dicha frecuencia modal aparece la categoría que, en estas condiciones, ha de considerarse como el modo de la distribución.

En el ejemplo siguiente, de entre los valores de la población correspondiente a cada una de las entidades federativas de México, en 1960, el mayor es 4 870 876 que es la frecuencia modal de la distribución. Esos 4 870 876 habitantes corresponden al Distrito Federal que es el modo de la distribución.

Mediana

El cálculo de la mediana en series de clases y frecuencias se puede hacer conforme al siguiente procedimiento:

1. Súmense todas las frecuencias y divídanse entre 2.
2. 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.
3. 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.
4. Divídase dicha diferencia entre la frecuencia de la clase en la que está comprendida la mitad de la suma de las frecuencias.
5. Multiplíquese el cociente por el valor del intervalo.
6. Agréguese el producto al límite inferior de la clase que contiene a la mitad de la suma de las frecuencias.

EJEMPLO DE DETERMINACIÓN DEL MODO. SERIES NOMINALES

Determinación del modo de las entidades federativas de México, en 1960, de acuerdo con su población total en esa fecha

CATEGORÍA *Nombre de la entidad*	FRECUENCIA *Población de esa entidad según la fuente*	FRECUENCIA MÁXIMA	MODO O CATEGORÍA DE MÁXIMA FRECUENCIA
AGUASCALIENTES	243 363		
BAJA CALIFORNIA NORTE	520 165		
BAJA CALIFORNIA SUR	81 594		
CAMPECHE	168 219		
COAHUILA	907 734		
COLIMA	164 450		
CHIAPAS	1 210 870		
CHIHUAHUA	1 226 793		
DISTRITO FEDERAL	4 870 876	4 870 876	DISTRITO FEDERAL
DURANGO	760 836		
GUANAJUATO	1 735 490		
GUERRERO	1 186 716		
HIDALGO	994 598		
JALISCO	2 443 261		
MÉXICO	1 897 851		
MICHOACÁN	1 851 876		
MORELOS	386 264		
NAYARIT	389 929		
NUEVO LEÓN	1 078 848		
OAXACA	1 727 266		
PUEBLA	1 973 837		
QUERÉTARO	355 045		
QUINTANA ROO	50 169		
SAN LUIS POTOSÍ	1 048 297		
SINALOA	838 404		
SONORA	783 378		
TABASCO	496 340		
TAMAULIPAS	1 024 182		
TLAXCALA	346 699		
VERACRUZ	2 727 899		
YUCATÁN	614 049		
ZACATECAS	817 831		

FUENTE: *VIII Censo General de Población, 1960. Resumen general.* Estados Unidos Mexicanos. Secretaría de Industria y Comercio. Dirección General de Estadística. México, D. F., 1962.

EJEMPLO DE DETERMINACIÓN DE LA MEDIANA EN UNA SERIE SENCILLA

Determinación de la mediana de las entidades federativas de México, en 1960, de acuerdo con el porciento de alfabetas en la población

CATEGORÍAS *Entidades federativas*	VARIABLE *Porciento de alfabetos*	ORDENACIÓN *Decreciente de porcientos*
AGUASCALIENTES	72.93	10
BAJA CALIFORNIA NORTE	81.12	2
BAJA CALIFORNIA SUR	79.51	5
CAMPECHE	68.10	12
COAHUILA	80.40	3
COLIMA	68.61	11
CHIAPAS	39.30	31
CHIHUAHUA	74.92	9
DISTRITO FEDERAL	83.42	1
DURANGO	75.16	8
GUANAJUATO	51.10	25
GUERRERO	37.19	32
HIDALGO	44.05	28
JALISCO	65.17	16
MÉXICO	57.40	22
MICHOACÁN	50.95	26
MORELOS	60.84	21
NAYARIT	65.94	15
NUEVO LEÓN	80.70	4
OAXACA	40.88	30
PUEBLA	50.23	27
QUERÉTARO	42.90	29
QUINTANA ROO	64.48	17
SAN LUIS POTOSÍ	53.34	24
SINALOA	66.03	13
SONORA	76.18	7
TABASCO	61.68	19
TAMAULIPAS	77.34	6
TLAXCALA	61.51	20
VERACRUZ	54.75	23
YUCATÁN	65.73	15
ZACATECAS	63.37	18

LUGARES CENTRALES: NÚMEROS DE ORDEN 32/2 y (32/2) + 1 = 16avo y 17avo

LUGAR CENTRAL: (16 + 17)/2 = 16.5

VALORES A LOS QUE CORRESPONDEN ESOS LUGARES: 16avo: 65.17%; 17avo: 64.48; 16.5: (65.17 + 64.48)/2 = 64.83% MEDIANA.

EJEMPLO DE CALCULO DE LA MEDIANA EN UNA SERIE DE FRECUENCIAS

Cálculo de la edad mediana de los menores de un año de México, en 1960

CATEGORÍAS *Edades de los menores de un año, en meses*	FRECUENCIAS *Niños de esa edad*	FRECUENCIAS ACUMULADAS	LOCALIZACIÓN DEL SEMI-EFECTIVO
0	58 096	58 096	
1	127 762	185 858	
2	114 240	300 098	
3	114 077	414 175	
4	97 215	511 390	
			572 093
5	91 508	602 898	
6	130 934	733 832	
7	87 920	821 752	
8	107 140	928 892	
9	84 875	1 013 767	
10	74 870	1 088 637	
11	55 550	1 144 187	

SUMA DE FRECUENCIAS 1 144 187

MITAD DE LA SUMA 572 093

ESTA MITAD DEL EFECTIVO ES MAYOR QUE 511 390
QUE CORRESPONDE A 4 MESES
Y MENOR QUE 602 898
QUE CORRESPONDE A 5 MESES

O SEA, QUE LA MEDIANA ES MAYOR QUE 4 meses
Y MENOR QUE 5 meses

DETERMINACIÓN DE QUÉ TANTO MAYOR QUE 4 Y MENOR QUE 5 MESES ES LA MEDIANA:

(5-4) MESES ES A (602 898-511 390) NIÑOS EN ESAS EDADES, COMO (MEDIANA - 4) ES A (572 093-511 390) NIÑOS EN ESAS EDADES.

1 : 91 508 :: (Md-4): 60 703

$$(\text{Md-4}) = \frac{1 \times 60\,703}{91\,508} = 0.6633$$

$$\text{Md} = 0.6633 + 4 = 4.6633.$$

N.B. Obsérvese que 4 se tomó como si fuera el límite inferior de su clase en vista de que el censador preguntó por la edad "en meses *cumplidos*", o sea que las edades de la tabla deben interpretarse como "de n meses cumplidos/o más/sin llegar a n + 1".

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:

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

El procedimiento consiste en:

1. Determinar cuál es la mitad de la distribución.
2. Determinar cuál es la frecuencia acumulada que contiene a dicha mitad de la distribución.
3. Restar de la frecuencia acumulativa que contiene a la mitad de la distribución, dicha mitad.
4. Dividir dicha diferencia entre la frecuencia no acumulada, de la clase que contiene a la mitad de la distribución.
5. Multiplicar dicho cociente por el intervalo.
6. 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.

EJEMPLO DE CALCULO DE LA MEDIANA EN SERIES DE CLASES Y FRECUENCIAS

Cálculo de la mediana de las edades de los habitantes de México, en 1960

CATEGORÍAS *Edades en años*	FRECUENCIAS *Pobladores de esas edades*	FRECUENCIAS ACUMULADAS	LOCALIZACIÓN DEL SEMI-EFECTIVO
de 0 a 4	5 776 747	5 776 747	
de 5 a 9	5 317 044	11 093 791	
de 10 a 14	4 358 316	15 452 107	
			17 404 793
de 15 a 19	3 535 265	18 987 372	
de 20 a 24	2 947 072	21 934 444	
de 25 a 29	2 504 892	24 439 336	
de 30 a 34	2 051 635	26 490 971	
de 35 a 39	1 920 680	28 411 651	
de 40 a 44	1 361 324	29 772 975	
de 45 a 49	1 233 608	31 006 583	
de 50 a 54	1 063 359	32 069 942	
de 55 a 59	799 899	32 869 841	
de 60 a 64	744 710	33 614 551	
de 65 a 69	414 164	34 028 715	
de 70 a 74	333 371	34 362 086	
de 75 a 79	187 773	34 549 859	
de 80 a 84	128 338	34 678 197	
de 85 a 89	131 389	34 809 586	
SUMA DE LAS FRECUENCIAS O EFECTIVO	34 809 586		

MITAD DEL EFECTIVO 17 404 793

EL SEMI-EFECTIVO 17 404 793 ES MAYOR QUE 15 452 107
Y MENOR QUE 18 987 372

O SEA, QUE LA MEDIANA ESTA ENTRE LAS CLASES DE 10-14 Y DE 15-19

LA MEDIANA ESTA ENTRE LAS EDADES DE QUIENES TIENEN 14 AÑOS O MENOS Y LAS DE QUIENES TIENEN 19 AÑOS O MENOS

LA MEDIANA ES MAYOR QUE 14 AÑOS
Y MENOR QUE 19 AÑOS

EJEMPLO DE CALCULO DE LA MEDIANA EN SERIES DE CLASES Y FRECUENCIAS (continuación)

Cálculo de la mediana de las edades de los habitantes de México, en 1960.

DETERMINACIÓN DE LA POSICIÓN EXACTA DE LA MEDIANA

De 15 a 19 años hay 3 535 265
De 15 a la mediana: 17 404 793 — 15 452 107 = 1 952 686
/(19+) — 15/ : 3 535 265 ::/Md — 15/: 1 952 686
5 : 3 535 265 :: Md — 15 : 1 952 686

$$\text{Md} - 15 = \frac{5 \times 1\,952\,686}{3\,535\,265} = \frac{9\,763\,430}{3\,535\,265}$$

$$\text{Md} - 15 = 2.76$$

$$\text{Md} \quad = 2.76 + 15 = 17.76 \text{ años}$$

COMPROBACIÓN:

De 15 a 19 años hay 3 535 265 personas
De 19+a la mediana: 18 987 372 — 17 404 793 = 1 582 579

/(19+) — 15/ : 3 535 265 :: (19+) — Md : 1 582 579
5 : 3 535 265 :: (19+) — Md : 1 582 579

$$(19+) - \text{Md} = \frac{5 \times 1\,582\,579}{3\,535\,265} = \frac{7\,912\,895}{3\,535\,265}$$

$$(19+) - \text{Md} = 2.23$$

$$-\text{Md} = 2.23 - (19+) = -16.77 + = -17.77$$

$$\text{Md} = 17.77 \text{ años.}$$

Medias

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

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

La media aritmética se obtiene si:

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

La media cuadrática se obtiene si:

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

La media cúbica se obtiene si:

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

La media armónica se obtiene si:

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

La media geométrica se obtiene si:

1. se buscan los logaritmos de los valores de la variable.
2. se suman dichos logaritmos,
3. se divide la suma de los logaritmos entre el número de datos, y
4. 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 decirse que:

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

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:

1. se sujeta a la variable a cierta operación, se multiplican los valores obtenidos por sus frecuencias correspondientes (ponderación),
2. se suman los productos así obtenidos,
3. se divide la suma entre el número de datos, y
4. se somete el cociente a la operación inversa de aquella a la que se sujetó a la variable.

OPERACIONES PARA LA OBTENCIÓN DE MEDIAS ESTADÍSTICAS

Media	*Operaciones que (1) deben ejecutarse con la variable*	*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".

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

$$\bar{x}_g = \text{antilogaritmo } \frac{\Sigma\,(\log m_i)\ f_i}{\Sigma\, f_i}$$

$$\bar{x}_g = \text{antilogaritmo } \frac{72.09730}{36} = 100.63$$

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	$m_i—m'$	$\frac{m_i—m'}{i} = d'$	$\frac{m_i—m'}{i} f_i$
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
						—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'),

$\frac{m_i—m'}{i}$ desviación de los puntos medios (m_i) con respecto a la media adivinada (m') en unidades del intervalo (i):

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

$$\bar{x}_a = 40 + \left[\frac{—119}{79} \right] 10$$

$\bar{x}_a = 40 + [(-1.5) \times 10] = 40 + (-15) = 40 - 15 = 25.$

CALCULO SIMPLIFICADO DE LA MEDIA ARITMETICA 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

$$\bar{x}_a = m' + \frac{\Sigma\ (m_i — m')\ f_i}{\Sigma\ f_i}$$

$$\bar{x}_a = 25 + \frac{—256}{96} = 25 - 2.67 = 22.33.$$

La fórmula de la media aritmética ponderada indica que:

1. hay que multiplicar cada uno de los valores de la variable por su frecuencia,
2. sumar los productos,
3. dividir la suma de esos productos entre la suma de las frecuencias, y
4. tomar ese cociente como media aritmética ponderada de la serie de frecuencias.

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, cuál es 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.

EJEMPLO DE CALCULO DE LA MEDIA ARITMÉTICA EN SERIES SENCILLAS

Cálculo de la población media de las entidades federativas de México, en 1960

Número	*Entidad*	*Población*
1	AGUASCALIENTES	243 363
2	BAJA CALIFORNIA NORTE	520 165
3	BAJA CALIFORNIA SUR	81 594
4	CAMPECHE	168 219
5	COAHUILA	907 734
6	COLIMA	164 450
7	CHIAPAS	1 210 870
8	CHIHUAHUA	1 226 793
9	DISTRITO FEDERAL	4 870 876
10	DURANGO	760 836
11	GUANAJUATO	1 735 490
12	GUERRERO	1 186 716
13	HIDALGO	994 598
14	JALISCO	2 443 261
15	MÉXICO	1 897 851
16	MICHOACAN	1 851 876
17	MORELOS	386 264
18	NAYARIT	389 929
19	NUEVO LEÓN	1 078 848
20	OAXACA	1 727 266
21	PUEBLA	1 973 837
22	QUERÉTARO	355 045
23	QUINTANA ROO	50 164
24	SAN LUIS POTOSÍ	1 048 297
25	SINALOA	838 404
26	SONORA	783 378
27	TABASCO	496 340
28	TAMAULIPAS	1 024 182
29	TLAXCALA	346 699
30	VERACRUZ	2 727 899
31	YUCATAN	614 049
32	ZACATECAS	817 831
	Efectivo	34 923 129
	Número de entidades	32

$$\frac{\text{Efectivo}}{\text{Número}} \qquad \frac{34\,923\,129 \text{ pobladores}}{32 \text{ entidades}} =$$

MEDIA ARITMÉTICA = 1 091 348 pobladores por entidad.

EJEMPLO DE CALCULO DE LA MEDIA ARITMÉTICA PONDERADA SERIES DE FRECUENCIAS

Cálculo de la media aritmética de las edades de los habitantes de México, que eran menores de un año en 1960

Edad en meses	*Pobladores de esa edad*	*Edad por número de pobladores de esa edad*
0	58 096	0
1	127 762	127 762
2	114 240	228 480
3	114 077	342 231
4	97 215	388 860
5	91 508	457 540
6	130 934	785 604
7	87 920	615 440
8	107 140	857 120
9	84 875	763 865
10	74 870	748 700
11	55 550	611 050
SUMAS	1 144 187	5 926 652
	Efectivo	Suma de edades ponderadas por sus frecuencias

MEDIA ARITMÉTICA $\dfrac{5\,926\,652 \text{ meses X niños}}{1\,144\,187 \text{ niños}} =$

MEDIA ARITMÉTICA 5.18 meses

EJEMPLO DE CÁLCULO DE LA MEDIA ARITMÉTICA PONDERADA SERIES DE CLASES Y FRECUENCIAS

Cálculo de la edad media de la población urbana de México, en 1960

CLASES *Edades en años*	PUNTOS MEDIOS *Edad media de la clase*	FRECUENCIAS *Pobladores entre esas edades*	PONDERACIÓN *Pobladores por edad media de la clase*
0 a 4 +	2.5	2 866 675	7 166 687.5
5 a 9 +	7.5	2 593 200	19 449 000.0
10 a 14 +	12.5	2 143 958	26 799 475.0
15 a 19 +	17.5	1 781 359	31 173 782.5
20 a 24 +	22.5	1 530 341	34 432 672.5
25 a 29 +	27.5	1 284 551	35 325 152.5
30 a 34 +	32.5	1 082 441	35 179 332.5
35 a 39 +	37.5	999 922	37 497 075.0
40 a 44 +	42.5	718 669	30 543 432.5
45 a 49 +	47.5	651 632	30 952 520.0
50 a 54 +	52.5	562 009	29 505 472.5
55 a 59 +	57.5	423 223	24 335 322.5
60 a 64 +	62.5	377 415	23 588 437.5
65 a 69 +	67.5	218 206	14 728 905.0
70 a 74 +	72.5	173 195	12 556 637.5
75 a 79 +	77.5	100 913	7 820 757.5
80 a 84 +	82.5	70 662	5 829 615.0
85 a 89 +	87.5	67 283	5 887 262.5
85 años o más, en la fuente			
SUMAS		17 645 654	412 771 540.0
		EFECTIVO	SUMA de las edades medias ponderadas por su frecuencia de clase
MEDIA ARITMÉTICA			$\frac{412\,771\,540.0}{17\,645\,654}$ = 23.39 años

EJEMPLO DE CALCULO DEL PRIMER MOMENTO RESPECTO DE UNA MEDIA ADIVINADA
SERIES DE CLASES Y FRECUENCIAS

Primer momento respecto a una media adivinada de 32.5 años, de la población urbana de México, en 1960

En unidades originarias

CLASES *Edades en años*	PUNTOS MEDIOS *Edad media de cada clase*	FRECUENCIAS *Pobladores entre esas edades*	DESVIACIÓN *Diferencia respecto de 32.5 años*	PONDERACIÓN *Pobladores por diferencia de edad respecto de 32.5*
0 a 4+	2.5	2 866 675	(2.5 — 32.5) = — 30	— 86 000 250
5 a 9+	7.5	2 593 200	(7.5 — 32.5) = — 25	— 64 830 000
10 a 14+	12.5	2 143 958	(12.5 — 32.5) = — 20	— 42 879 160
15 a 19+	17.5	1 781 359	(17.5 — 32.5) = — 15	— 26 720 385
20 a 24+	22.5	1 530 341	(22.5 — 32.5) = — 10	— 15 303 410
25 a 29+	27.5	1 284 551	(27.5 — 32.5) = — 5	— 6 422 755
30 a 34+	32.5	1 082 441	(32.5 — 32.5) = 0	0
35 a 39+	37.5	999 922	(37.5 — 32.5) = + 5	+ 4 999 610
40 a 44+	42.5	718 669	(42.5 — 32.5) = + 10	+ 7 186 690
45 a 49+	47.5	651 632	(47.5 — 32.5) = + 15	+ 9 774 480
50 a 54+	52.5	562 009	(52.5 — 32.5) = + 20	+ 11 240 180
55 a 59+	57.5	423 223	(57.5 — 32.5) = + 25	+ 10 580 575
60 a 64+	62.5	377 415	(62.5 — 32.5) = + 30	+ 11 322 450
65 a 69+	67.5	218 206	(67.5 — 32.5) = + 35	+ 7 637 210
70 a 74+	72.5	173 195	(72.5 — 32.5) = + 40	+ 6 927 800
75 a 79+	77.5	100 913	(77.5 — 32.5) = + 45	+ 4 541 085
80 a 84+	82.5	70 662	(82.5 — 32.5) = + 50	+ 3 533 100
85 a 89+	87.5	67 283	(87.5 — 32.5) = + 55	+ 3 700 565
				— 242 155 960
				+ 81 443 745
SUMAS		17 645 654		— 160 712 215

$$\text{PRIMER MOMENTO RESPECTO DE 32.5 AÑOS} = \frac{-160\,712\,215}{17\,645\,654}$$

$$\text{PRIMER MOMENTO RESPECTO DE 32.5 AÑOS} = -9.11 \text{ años}$$

O sea, que entre 32.5 años y la media verdadera (punto de equilibrio) hay una diferencia de 9.11 años. Esto permite anticipar que la media aritmética será igual a 32.5 años — 9.11 años = 23.39 años.

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 sustituir 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:

EJEMPLO DE CALCULO DIRECTO DE LA MEDIA ARITMÉTICA

Edad media de los inmigrantes menores de 60 años que entraron a México, en 1956

l_i	l_s	f	m	mf
0	5—	139	2.5	347.5
5	10—	129	7.5	967.5
10	15—	88	12.5	1 110.0
15	20—	421	17.5	7 367.5
20	25—	630	22.5	14 175.0
25	30—	521	27.5	14 327.5
30	35—	269	32.5	8 742.5
35	40—	207	37.5	7 762.5
40	45—	128	42.5	5 440.0
45	50—	109	47.5	5 177.5
50	55—	75	52.5	3 937.5
55	60—	46	57.5	2 645.0
	Σ	2 762.		72 000.0

$$\bar{x}_1 = \frac{\Sigma\ mf}{\Sigma\ f} = \frac{72\,000}{2\,762} = 26.068 \doteq 27 \text{ años.}$$

Cálculo de los momentos

En series sencillas:

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

1. restando de cada valor de la variable el valor de la media arbitraria,
2. multiplicando esos residuos (o "desviaciones") por la frecuencia correspondiente,
3. sumando dichos residuos (o "desviaciones"), y
4. dividiendo la suma entre el número de casos.

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

1. restando de cada valor de la variable el valor de la media arbitraria.
   (Ponderación); multiplicando esos residuos o "desviaciones" por la frecuencia correspondiente,
2. sumando dichos productos, y
3. dividiendo la suma entre la suma de las frecuencias.

El primer momento en series de clases y frecuencias se calcula:

1. 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,
2. sumando dichos productos, y
3. dividiendo dicha suma entre la suma de las frecuencias de clase.

EJEMPLO DE CALCULO SIMPLIFICADO DE LA MEDIA ARITMÉTICA MEDIANTE RESTA DE UNA CONSTANTE Y SUMA ULTERIOR DE LA MISMA

Edades de las mujeres inmigrantes menores de 60 años que entraron a México, en 1956

l_i	l_s	f	m	$d' = (m - 27.5)$	d'f
0	5 —	127	2.5	— 25	— 3 175
5	10 —	108	7.5	— 20	— 2 160
10	15 —	77	12.5	— 15	— 1 155
15	20 —	128	17.5	— 10	— 1 280
20	25 —	189	22.5	— 5	— 945
25	30 —	225	27.5	0	0
30	35 —	168	32.5	+ 5	+ 840
35	40 —	120	37.5	+ 10	+ 1 200
40	45 —	109	42.5	+ 15	+ 1 635
45	50 —	81	47.5	+ 20	+ 1 620
50	55 —	56	52.5	+ 25	+ 1 400
55	60 —	50	57.5	+ 30	+ 1 500
		1 438			— 8 715
					+ 8 195
					— 520

$$\bar{x}_a = x' + \frac{\Sigma\, d'f}{\Sigma\; f} = 27.5 + \left(\frac{-520}{1\,438}\right) =$$

$$= 27.5 - 0.36 = 27.14 \text{ años.}$$

EJEMPLO DE CÁLCULO SIMPLIFICADO DE LA MEDIA ARITMÉTICA MEDIANTE DIVISIÓN ENTRE UNA CONSTANTE Y ULTERIOR MULTIPLICACIÓN POR ELLA

Edad de las mujeres inmigrantes menores de 60 años que entraron a México, en 1957

l_i	l_s	f	m	m/5	(m/5)f
0	5 —	129	2.5	0.5	64.5
5	10 —	157	7.5	1.5	235.5
10	15 —	121	12.5	2.5	302.5
15	20 —	162	17.5	3.5	567.5
20	25 —	231	22.5	4.5	1 039.5
25	30 —	241	27.5	5.5	1 325.5
30	35 —	233	32.5	6.5	1 514.5
35	40 —	162	37.5	7.5	1 215.0
40	45 —	106	42.5	8.5	901.0
45	50 —	88	47.5	9.5	836.0
50	55 —	66	52.5	10.5	693.0
55	60 —	41	57.5	11.5	471.5
	Σ	1 737			9 166.0

$$\bar{x}_1 = 5\left[\frac{\sum \frac{m}{5}f}{\sum f}\right] = 5\left[\frac{9\,166.0}{1\,737}\right] = 5 \times 5.276 =$$

$$\doteq 26.38 \text{ años} \doteq 26.4.$$

EJEMPLO DE CÁLCULO DE LA MEDIA ARITMÉTICA MEDIANTE RESTA DE UNA CONSTANTE Y DIVISIÓN ENTRE OTRA Y ULTERIOR MULTIPLICACIÓN POR LA SEGUNDA Y SUMA DE LA PRIMERA

Edad de los hombres que inmigraron a México en 1957
y que eran menores de 60 años

l_i	l_s	f	m	m — 32.5	$\dfrac{m - 32.5}{5} = D'_i$	D'f
0	5 —	181	2.5	— 30	— 6	— 1 086
5	10 —	178	7.5	— 25	— 5	— 890
10	15 —	157	12.5	— 20	— 4	— 628
15	20 —	572	17.5	— 15	— 3	— 1 716
20	25 —	700	22.5	— 10	— 2	— 1 400
25	30 —	575	27.5	— 5	— 1	— 575
30	35 —	310	32.5	0	0	0
35	40 —	235	37.5	5	+ 1	+ 235
40	45 —	146	42.5	10	+ 2	+ 292
45	50 —	124	47.5	15	+ 3	+ 372
50	55 —	95	52.5	20	+ 4	+ 380
55	60 —	63	57.5	25	+ 5	+ 315
	Σ	3 336				— 6 295 + 1 594
						— 4 701

$$\bar{x}_1 = \left(\frac{\Sigma D'f}{\Sigma f}\right) i + x' = \left(\frac{-4\,701}{3\,336}\right) 5 + 32.5 =$$

$$= (-1.409) \times 5 + 32.5 =$$

$$= -7.05 + 32.5 = 25.45 \doteq 25.5 \text{ años.}$$

EJEMPLO DE CÁLCULO DEL PRIMER MOMENTO RESPECTO DE UNA MEDIA ADIVINADA

SERIE DE FRECUENCIAS

Cálculo del primer momento de la serie de edades de los habitantes de México que eran menores de un año en 1960, respecto de la edad de 6 meses (media adivinada)

VARIABLE *Edad en meses*	FRECUENCIA *Pobladores de esa edad*	DESVIACIÓN *Diferencia de la edad media*	PONDERACIÓN *Pobladores × diferencia de edad*
0	58 096	$(0-6)=-6$	− 348 576
1	127 762	$(1-6)=-5$	− 638 810
2	114 240	$(2-6)=-4$	− 456 960
3	114 077	$(3-6)=-3$	− 342 231
4	97 215	$(4-6)=-2$	− 194 430
5	91 508	$(5-6)=-1$	− 91 508
6	130 934	$(6-6)=\ 0$	0
7	87 920	$(7-6)=+1$	+ 87 920
8	107 140	$(8-6)=+2$	+ 214 280
9	84 875	$(9-6)=+3$	+ 254 625
10	74 870	$(10-6)=+4$	+ 299 480
11	55 550	$(11-6)=+5$	+ 277 750
SUMAS	1 144 187		− 2 072 515 + 1 134 055
			− 938 460

PRIMER MOMENTO

$$\frac{-\ 938\,460}{+\ 1\,144\,187} =$$

$$= -\ 0.82 \text{ meses.}$$

EJEMPLO DE CÁLCULO DEL PRIMER MOMENTO RESPECTO DE LA MEDIA ARITMÉTICA

Cálculo del primer momento de la serie de edades de los habitantes de México que eran menores de un año en 1960, respecto de la edad media aritmética de 5.18 meses calculada anteriormente

VARIABLE *Edad en meses*	FRECUENCIA *Pobladores de esa edad*	DESVIACIÓN *Diferencia entre su edad y la edad media del grupo*	PONDERACIÓN *Pobladores × diferencia de sus edades respecto a la media*
0	58 096	(0 − 5.18) = − 5.18	− 300 937.28
1	127 762	(1 − 5.18) = − 4.18	− 534 045.16
2	114 240	(2 − 5.18) = − 3.18	− 363 283.20
3	114 077	(3 − 5.18) = − 2.18	− 248 687.86
4	97 215	(4 − 5.18) = − 1.18	− 114 713.70
5	91 508	(5 − 5.18) = − 0.18	− 16 471.44
6	130 934	(6 − 5.18) = + 0.82	+ 107 365.88
7	87 920	(7 − 5.18) = + 1.82	+ 160 014.40
8	107 140	(8 − 5.18) = + 2.82	+ 302 134.80
9	84 875	(9 − 5.18) = + 3.82	+ 324 222.50
10	74 870	(10 − 5.18) = + 4.82	+ 360 873.40
11	55 550	(11 − 5.18) = + 5.82	+ 323 301.00
SUMAS	1 144 187		− 1 578 138.64 + 1 577 911.98
			+ 226.66

PRIMER MOMENTO RESPECTO

A LA MEDIA (5.18) $= \dfrac{+\ 226.66}{+\ 1\,144\,187} = +\ 0.00019.$

O sea prácticamente cero según ocurre siempre con el primer momento respecto de la media aritmética; aquí no se obtuvo 0 porque 5.18 es el valor aproximado y no el exacto de la media aritmética de la serie.

EJEMPLO DE CALCULO SIMPLIFICADO DEL PRIMER MOMENTO RESPECTO DE LA MEDIA ADIVINADA SERIES DE CLASES Y FRECUENCIAS

Primer momento respecto de una media adivinada de 32.5 años, de la población urbana de México, en 1960
En unidades del intervalo (5)

CLASES *Edades en años*	PUNTOS MEDIOS *Edad media de cada clase*	FRECUENCIAS *Pobladores entre esas edades*	DESVIACIÓN INTERVALAR *Diferencia a 32.5, entre 5*	PONDERACIÓN *Pobladores por diferencia de edades*
0 a 4 +	2.5	2 866 675	(2.5 — 32.5)/5 = — 6	— 17 200 050
5 a 9 +	7.5	2 593 200	— 25 /5 = — 5	— 12 966 000
10 a 14 +	2.5	2 143 958	— 20 /5 = — 4	— 8 575 832
15 a 19 +	17.5	1 781 359	— 15 /5 = — 3	— 5 344 077
20 a 24 +	22.5	1 530 341	— 10 /5 = — 2	— 3 060 682
25 a 29 +	27.5	1 284 551	— 5 /5 = — 1	— 1 284 551
30 a 34 +	32.5	1 082 441	0 /5 = 0	0
35 a 39 +	37.5	999 922	+ 5 /5 = + 1	+ 999 922
40 a 44 +	42.5	718 669	+ 10 /5 = + 2	+ 1 437 338
45 a 49 +	47.5	651 632	+ 15 /5 = + 3	+ 1 954 896
50 a 54 +	52.5	562 009	+ 20 /5 = + 4	+ 2 248 036
55 a 59 +	57.5	423 223	+ 25 /5 = + 5	+ 2 116 115
60 a 64 +	62.5	377 415	+ 30 /5 = + 6	+ 2 264 490
65 a 69 +	67.5	218 206	+ 35 /5 = + 7	+ 1 527 442
70 a 74 +	72.5	173 195	+ 40 /5 = + 8	+ 1 385 560
75 a 79 +	77.5	100 913	+ 45 /5 = + 9	+ 908 217
80 a 84 +	82.5	70 662	+ 50 /5 = 10	+ 706 620
85 a 89 +	87.5	67 283	(87.5 — 32.5)/5 = 11	+ 740 113
				— 48 431 192
				+ 16 288 749
SUMAS		17 645 654		— 32 142 443

PRIMER MOMENTO RESPECTO DE 32.5 AÑOS EN UNIDADES DEL INTERVALO DE CLASE (5 AÑOS) $= \dfrac{-32\,142\,443}{17\,645\,654} =$
$= -1.82$ intervalos

Si el primer momento en unidades del intervalo es de — 1.82 intervalos de 5 años cada uno, esto significa que, el primer momento en unidades originarias será de $-1.82 \times 5 =$
$= -9.10$ años aproximadamente, como lo mostró el cálculo directo (— 9.11 años).

EJEMPLO DE CÁLCULO SIMPLIFICADO DE LA MEDIA ARITMÉTICA

SERIES DE CLASES Y FRECUENCIAS

Cálculo de la edad media de la población rural de México en 1960, utilizando una media adivinada de 27.5

CLASES *Edades en años*	PUNTOS MEDIOS *Edad media de la clase*	FRECUENCIA *Pobladores entre esas edades*	DESVIACIONES *Número de clases que dista de la clase en que 27.5 es punto medio*	PONDERACIONES *Pobladores por ese número de clases*
0 a 4 +	2.5	2 910 072	− 5	− 14 550 360
5 9 +	7.5	2 724 844	− 4	− 10 895 376
10 14 +	12.5	2 214 358	− 3	− 6 643 074
15 19 +	17.5	1 753 906	− 2	− 3 507 812
20 24 +	22.5	1 416 731	− 1	− 1 416 731
25 29 +	27.5	1 220 341	0	0
30 34 +	32.5	969 194	+ 1	+ 969 194
35 39 +	37.5	920 758	+ 2	+ 1 841 516
40 44 +	42.5	642 655	+ 3	+ 1 927 965
45 49 +	47.5	581 976	+ 4	+ 2 327 904
50 54 +	52.5	501 350	+ 5	+ 2 506 750
55 59 +	57.5	376 676	+ 6	+ 2 260 056
60 64 +	62.5	376 295	+ 7	+ 2 571 065
65 69 +	67.5	195 958	+ 8	+ 1 567 664
70 74 +	72.5	160 176	+ 9	+ 1 441 584
75 79 +	77.5	86 860	+ 10	+ 868 600
80 84 +	82.5	57 676	+ 11	+ 634 436
85 89 +	87.5	64 106	+ 12	+ 769 272
			Cálculo en la tirilla anexa.	− 37 013 353 + 19 686 006
SUMAS		17 163 932		− 17 327 347

PRIMER MOMENTO RESPECTO DE LA MEDIA ADIVINADA (27.5) = $\dfrac{-17\,327\,347}{17\,163\,932} =$

$= -1.009$

PRIMER MOMENTO POR EL INTERVALO (5) $= -1.009 \times 5 =$
$= -5.045 \doteq -5.05$

MEDIA ARITMÉTICA = MEDIA ADIVINADA + PRIMER MOMENTO × INTERVALO:

$= 27.5 + (-5.05) =$
$= 22.45$ años

EJEMPLO DE CALCULO SIMPLIFICADO DE LA MEDIA ARITMÉTICA

SERIES DE CLASES Y FRECUENCIAS. TIRILLA ANEXA

Cálculo de las desviaciones o número de clases que dista la clase correspondiente de aquella otra clase en la que la media adivinada (27.5) es punto medio

PUNTOS MEDIOS	CÁLCULO DE LA DESVIACIÓN	DESVIACIÓN
0 a 4+/2	(PUNTO MEDIO/MEDIA ADIVINADA)/I	
2.5	(2.5 — 27.5) / 5	— 5
7.5	(7.5 — 27.5) / 5	— 4
12.5	(12.5 — 27.5) / 5	— 3
17.5	(17.5 — 27.5) / 5	— 2
22.5	(22.5 — 27.5) / 5	— 1
27.5	(27.5 — 27.5) / 5	0
32.5	(32.5 — 27.5) / 5	+ 1
37.5	(37.5 — 27.5) / 5	+ 2
42.5	(42.5 — 27.5) / 5	+ 3
47.5	(47.5 — 27.5) / 5	+ 4
52.5	(52.5 — 27.5) / 5	+ 5
57.5	(57.5 — 27.5) / 5	+ 6
62.5	(62.5 — 27.5) / 5	+ 7
67.5	(67.5 — 27.5) / 5	+ 8
72.5	(72.5 — 27.5) / 5	+ 9
77.5	(77.5 — 27.5) / 5	+10
82.5	(82.5 — 27.5) / 5	+11
87.5	(87.5 — 27.5) / 5	+12

Cálculo simplificado de la media aritmética

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

1. se determina o elige una media arbitraria o adivinada (m') entre los puntos medios (uno de los m_i),
2. 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 $\dfrac{m_i - m'}{i}$; 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),

3. se multiplican las desviaciones de cada clase por las frecuencias correspondientes y se divide su suma entre la suma de las frecuencias,
4. se multiplica el cociente ("primer momento con respecto a la media arbitraria en unidades del intervalo") por el valor del intervalo,
5. se suma al producto anterior la media elegida, adivinada o arbitraria.

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

1. Se elige uno de los puntos medios de las clases como media arbitraria (media arbitraria m').
2. Frente a la clase de la media arbitraria se coloca cero, hacia arriba —1, —2, —3, etcétera, y hacia abajo 1,2,3... etcétera, 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).
3. Se multiplica cada una de las desviaciones de clase por la correspondiente frecuencia de clase.
4. Se suman los productos de las frecuencias por las desviaciones.
5. Se divide la suma de los productos de las frecuencias por las desviaciones entre la suma de las frecuencias (primer momento).
6. Se multiplica el primer momento por el valor del intervalo.
7. Se suma al producto del primer momento por el intervalo el valor de la media arbitraria.

2.2. ALGUNAS TÉCNICAS PARA LA OBTENCIÓN DE MEDIDAS DE DISPERSIÓN

La razón de variación

En las series nominales, el promedio más empleado es el modo. Éste —como todos los promedios— es más o menos representativo de toda la serie. Para saber qué tan representativo es, se calcula —en éste, como en el caso de otros promedios como la mediana o las medias— una medida de dispersión. En el caso de las series nominales, la medida de dispersión es la "razón de variación".

La razón de variación es una diferencia de relativos (o de porcientos). Mide en qué tanto se aleja de 1 (o de 100%) la frecuencia modal relativa.

La frecuencia modal relativa se obtiene dividiendo la frecuencia modal absoluta entre el efectivo (o suma de las frecuencias) de la distribución y es, por tanto, una probabilidad. La frecuencia modal relativa indica cuál es la probabilidad que hay de acertar si se considera al modo como categoría representativa de toda la serie de la que forma parte.

Por su parte, 1 (o 100%) representa la probabilidad total de acertar o fallar al considerar al modo como representantivo de toda la serie. O sea, que la diferencia entre 1 y la frecuencia modal relativa representa la probabilidad que hay de fallar si se considera al modo como categoría representativa de toda la serie.

En suma, para calcular la razón de variación:

1. Se toma la frecuencia máxima ("frecuencia modal").
2. Se la divide entre el efectivo (suma de frecuencias).
3. El cociente obtenido se resta de 1.

EJEMPLO DE CÁLCULO DE LA RAZÓN DE VARIACIÓN
SERIES NOMINALES

Cálculo de la razón de variación de la población total de las entidades federativas de México, en 1960

Modo: Distrito Federal (4 870 876 habitantes)

CATEGORÍA *Nombre de la entidad*	FRECUENCIA *Población de esa entidad*
AGUASCALIENTES	243 363
BAJA CALIFORNIA NORTE	520 165
BAJA CALIFORNIA SUR	81 594
CAMPECHE	168 219
COAHUILA	907 734
COLIMA	164 450
CHIAPAS	1 210 870
CHIHUAHUA	1 226 793
DISTRITO FEDERAL	4 870 876
DURANGO	760 836
GUANAJUATO	1 735 490
GUERRERO	1 186 716
HIDALGO	994 598
JALISCO	2 443 261
MÉXICO	1 897 851
MICHOACÁN	1 851 876
MORELOS	386 264
NAYARIT	389 929
NUEVO LEÓN	1 078 848
OAXACA	1 727 266
PUEBLA	1 973 837
QUERÉTARO	355 045
QUINTANA ROO	50 169
SAN LUIS POTOSÍ	1 048 297
SINALOA	838 404
SONORA	783 378
TABASCO	496 340
TAMAULIPAS	1 024 182
TLAXCALA	346 699
VERACRUZ	2 727 899
YUCATÁN	614 049
ZACATECAS	817 831
SUMA	34 923 129

CÁLCULO DE LA RAZÓN DE VARIACIÓN

FRECUENCIA MODAL $=$
$= 4\,870\,876$

FRECUENCIA MODAL RELATIVA $=$

$$= \frac{\text{FRECUENCIA MODAL}}{\text{EFECTIVO}} =$$

$$= \frac{4\,870\,876}{34\,923\,129} =$$

$= 0.1395 = 13.95\%$

RAZÓN DE VARIACIÓN $=$

$=$ UNO $-$ FRECUENCIA MODAL RELATIVA $=$

$= 1 - 0.1395 =$

$= 0.8605 = 86.05\%$

Promedios laterales o
promedios de graduación

En tanto que al modo, a la mediana y a las diversas medias se les conoce como promedios centrales por ocupar el centro de las distribuciones (cuando éstas son simétricas) o encontrarse cerca del centro (cuando no lo son), hay otros promedios que se disponen a uno y otro lado de los promedios centrales a los que, por ese motivo, se les conoce como promedios laterales.

Los promedios laterales sirven para graduar las distribuciones y también para apreciar —por resta de los más próximos a la media— la desviación-promedio.

A los promedios de graduación se les conoce genéricamente como "cuantilas" 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 la 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:

$$\left.\begin{array}{l}\text{1. Zona de deficiencia,}\\ \text{2. Zona de normalidad, y}\\ \text{3. Zona de excedencia,}\end{array}\right\}\text{ de un fenómeno.}$$

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.

Desviación media cuadrática

Para obtener el valor de la desviación media cuadrática de una serie basta con:

1. 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 ... hacia arriba, 1, 2 ... hacia abajo) la media aritmética de dichas desviaciones.
2. Elevar al cuadrado esos residuos.
3. Multiplicar esos cuadrados por las frecuencias correspondientes.
4. Sumar esos productos.
5. Dividir dicha suma entre la suma de las frecuencias.
6. Extraer la raíz cuadrada del cociente.
7. Multiplicar la raíz cuadrada por el intervalo.

Cálculo de la desviación σ en términos de los dos primeros momentos

1. Se calcula el primer momento de la distribución:
   - 1.1. multiplicando las desviaciones con respecto a la media arbitraria en unidades de intervalo (—n ... —2, —1, 0, 1, 2, ... n) por las frecuencias correspondientes,
   - 1.2. se suman esos productos,
   - 1.3. se dividen entre la suma de las frecuencias.
2. Se eleva al cuadrado el primer momento.
3. Se calcula el segundo momento:
   - 3.1. elevando al cuadrado las desviaciones ... (—n, ... —2, —1, 0, 1, 2 ... n),
   - 3.2 multiplicando dichos cuadrados por las frecuencias,
   - 3.3. sumando dichos productos,
   - 3.4. dividiendo la suma entre la suma de las frecuencias.
4. Se resta del segundo momento el cuadrado del primer momento.
5. Se extrae la raíz cuadrada de la diferencia.
6. Se multiplica la raíz cuadrada por el intervalo.

EJEMPLO DE CALCULO DE LA DESVIACIÓN MEDIA CUADRATICA SERIES DE CLASES Y FRECUENCIAS

Cálculo de la desviación media cuadrática de la población rural de México en 1960, respecto de su media aritmética (aproximada) de 22.5 años

La media aritmética exacta fue de 22.45 años según cálculo anterior

PUNTOS MEDIOS *Edad media de cada clase*	FRECUENCIA *Pobladores de esa edad*	DESVIACIONES *Distancia de la edad media de cada clase respecto de la media de la serie*	CUADRADOS DE DESVIACIONES	PONDERACIÓN *Pobladores por cuadrados de desviaciones*
5.2	2 910 072	(2.1 — 22.5) = − 20	400	1 164 028 800
7.5	2 723 844	(7.5 — 22.5) = − 15	225	612 864 900
12.5	2 214 358	(12.5 — 22.5) = − 10	100	221 435 800
17.5	1 753 906	(17.5 — 22.5) = − 5	25	43 847 650
22.5	1 416 731	(22.5 — 22.5) = 0	0	0
27.5	1 220 341	(27.5 — 22.5) = + 5	25	30 508 525
32.5	969 194	(32.5 — 22.5) = + 10	100	96 919 400
37.5	920 758	(37.5 — 22.5) = + 15	225	207 170 550
42.5	642 655	(42.5 — 22.5) = + 20	400	257 062 000

47.5	581 976	(47.5 — 22.5) = + 25	625	363 735 000
52.5	501 350	(52.5 — 22.5) = + 30	900	451 215 000
57.5	376 676	(57.5 — 22.5) = + 35	1 225	461 428 100
62.5	367 295	(62.5 — 22.5) = + 40	1 600	587 672 000
67.5	195 958	(67.5 — 22.5) = + 45	2 025	396 814 950
72.5	160 176	(72.5 — 22.5) = + 50	2 500	400 440 000
77.5	86 860	(77.5 — 22.5) = + 55	3 025	262 751 500
82.5	57 676	(82.5 — 22.5) = + 60	3 600	207 633 600
87.5	64 106	(87.5 — 22.5) = + 65	4 225	270 847 850
	17 163 932			6 036 375 625

SEGUNDO MOMENTO RESPECTO DE LA MEDIA $\dfrac{6\,036\,375\,625}{17\,163\,932} =$

$= 351.6895$

DESVIACIÓN MEDIA CUADRÁTICA $= \sqrt{351.6895} =$

$= 18.75$ años

O sea, que los datos de la serie se desvían de la media de 22.5 años, en promedio, 19 años aproximadamente.

EJEMPLO DE CALCULO SIMPLIFICADO DE LA DESVIACIÓN MEDIA CUADRATICA

SERIES DE CLASES Y FRECUENCIAS

Cálculo de la desviación media cuadrática de la población rural de México en 1960, respecto de su media aritmética (aproximada) de 22.5 años

En unidades del intervalo

PUNTOS MEDIOS *Edad media de cada clase*	FRECUENCIA *Pobladores en esa edad media*	DESVIACIONES EN UNIDADES DEL INTERVALO (5)	CUADRADOS DE LAS DESVIACIONES	PONDERACIÓN *Pobladores × las desviaciones cuadrados de*
2.5	2 910 072	− 20/5 = − 4	16	46 561 152
7.5	2 723 844	− 15/5 = − 3	9	24 514 596
12.5	2 214 358	− 10/5 = − 2	4	8 857 432
17.5	1 753 906	− 5/5 = − 1	1	1 753 906
22.5	1 416 731	0/5 = 0	0	0
27.5	1 220 341	+ 5/5 = + 1	1	1 220 341
32.5	969 194	+ 10/5 = + 2	4	3 876 776
37.5	920 758	+ 15/5 = + 3	9	8 286 822
42.5	642 655	+ 20/5 = + 4	16	10 282 480
47.5	581 976	+ 25/5 = + 5	25	14 549 400
52.5	501 350	+ 30/5 = + 6	36	18 048 600
57.5	376 676	+ 35/5 = + 7	49	18 457 124

62.5	367 295	$+40/5 = +8$	64	23 506 880
67.5	195 958	$+45/5 = +9$	81	15 872 598
72.5	160 176	$+50/5 = +10$	100	16 017 600
77.5	86 860	$+55/5 = +11$	121	10 510 060
82.5	57 676	$+60/5 = +12$	144	8 305 344
87.5	64 106	$+65/5 = +13$	169	10 833 914
SUMAS	17 163 932			241 455 025

SEGUNDO MOMENTO RESPECTO DE LA MEDIA EN UNIDADES DEL INTERVALO DE CLASES (5)

$$\frac{241\,455\,025}{17\,163\,932} = $$

$$= 14.0676$$

DESVIACIÓN MEDIA CUADRATICA EN UNIDADES DEL INTERVALO $= \sqrt{14.0676} = 3.75$

Para obtener la desviación media cuadrática en unidades originarias bastará con multiplicar esta desviación en unidades del intervalo (3.75) por el intervalo (5): $3.75 \times 5 = 18.75$ que fue el valor obtenido por el procedimiento anterior, más bromoso.

EJEMPLO DE CALCULO ULTRASIMPLIFICADO DE LA DESVIACIÓN MEDIA CUADRATICA

SERIES DE CLASES Y FRECUENCIAS

Cálculo de la desviación media cuadrática de la población rural de México en 1960, mediante el cálculo de los dos primeros momentos, respecto de una media adivinada (de 27.5 años). El primer momento, calculado anteriormente es = 5.05. Aquí se calcula el segundo momento

PUNTOS MEDIOS *Edad media de la clase*	FRECUENCIAS *Pobladores en esa edad media*	DESVIACIONES *Respecto a la media adivinada (27.5) en unidades del intervalo (5)*	CUADRADOS DE ESAS DESVIACIONES	PONDERACIÓN *Pobladores × cuadrados de desviaciones*
2.5	2 910 072	—5	25	72 751 800
7.5	2 723 844	—4	16	43 581 504
12.5	2 214 358	—3	9	19 929 222
17.5	1 753 906	—2	4	7 015 624
22.5	1 416 731	—1	1	1 416 731
27.5	1 220 341	0	0	0
32.5	969 194	1	1	969 194
37.5	920 758	2	4	3 683 032
42.5	642 655	3	9	5 783 895
47.5	581 976	4	16	9 311 616
52.5	501 350	5	25	12 533 750
57.5	376 676	6	36	13 560 336
62.5	367 676	7	49	17 997 455
67.5	195 958	8	64	12 541 312
72.5	160 176	9	81	12 974 256
77.5	86 860	10	100	8 686 000
82.5	57 676	11	121	6 978 796
87.5	64 106	12	144	9 231 264
SUMAS	17 163 932			258 945 787

SEGUNDO MOMENTO
RESPECTO DE LA
MEDIA ADIVINADA (27)
EN UNIDADES DEL INTERVALO (5)

$$\frac{258\,945\,787}{17\,163\,932} = 15.0866$$

PRIMER MOMENTO
RESPECTO DE LA
MEDIA ADIVINADA (27)
EN UNIDADES DEL INTERVALO (5) —1.009

CUADRADO DEL PRIMER MOMENTO $(-1.009)^2 = +1.0181$

SEGUNDO MOMENTO RESPECTO DE LA MEDIA *ARITMÉTICA*:

SEGUNDO MOMENTO RESPECTO DE LA MEDIA *ARBITRARIA*,
menos
CUADRADO DEL PRIMER MOMENTO RESPECTO A LA *ARBITRARIA*:

$15.0866 - 1.0181 = 14.0685$

DESVIACIÓN MEDIA CUADRÁTICA *EN UNIDADES DEL INTERVALO*:

$$\sqrt{14.0685} = 3.75$$

Puesto que esta desviación está en unidades del intervalo, y el intervalo es de 5 años, para obtener la desviación cuadrática media en unidades originarias habrá que multiplicar esa desviación por 5 : 3.75 × 5 años = 18.75 años, desviación cuadrática media de la serie, como mostró el cálculo directo y comprobó el cálculo simplificado.

2.3. ALGUNAS TÉCNICAS PARA LA OBTENCIÓN DE ÍNDICES DE ASIMETRÍA Y DE CURTOSIS

Cálculo de un momento de orden r

Los procedimientos para calcular un momento de orden r son los siguientes:

A) Momentos con respecto a la media aritmética.

1. Calcúlese la media aritmética.
2. 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).
3. 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, multiplíquese esas potencias de las desviaciones por las frecuencias correspondientes.

4. Súmense los valores así obtenidos.
5. 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.

1. 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).
2. Réstese esta media elegida de los datos o de los puntos medios (según el tipo de serie).
3. Continúese el procedimiento como para A.

Asimetría

Con el fin de calcular la asimetría de una distribución habrá que:

1. Calcular la desviación cúbica media con respecto a la media aritmética (o bien la raíz cúbica del tercer momento con respecto a dicha media), y
2. Dividir dicha desviación cúbica media entre la desviación cuadrática media.

EJEMPLO DE CALCULO DEL MOMENTO DE TERCER ORDEN RESPECTO DE LA MEDIA ARITMÉTICA

Cálculo del momento de tercer orden de la población rural de México en 1960, respecto de su media aritmética (aproximada) de 22.5 años

PUNTOS MEDIOS *Edad media de cada clase*	FRECUENCIA *Pobladores en esa edad*	DESVIACIONES	CUBOS DE DESVIACIONES	PONDERACIÓN
2.5	2 910 072	− 20	− 8 000	− 23 280 576 000
7.5	2 723 844	− 15	− 3 375	− 9 192 973 500
12.5	2 214 358	− 10	− 1 000	− 2 214 358 000
17.5	1 753 906	− 5	− 125	− 219 238 250
22.5	1 416 731	0	0	0
27.5	1 220 341	+ 5	+ 125	+ 152 542 625
32.5	969 194	+ 10	+ 1 000	+ 969 194 000
37.5	920 758	+ 15	+ 3 375	+ 3 107 558 250
42.5	642 655	+ 20	+ 8 000	+ 5 141 240 000
47.5	581 976	+ 25	+ 15 625	+ 9 093 375 000
52.5	501 350	+ 30	+ 27 000	+ 13 536 450 000
57.5	376 676	+ 35	+ 42 875	+ 16 149 983 500
62.5	367 295	+ 40	+ 64 000	+ 23 506 880 000
67.5	195 958	+ 45	+ 91 125	+ 17 856 672 750
72.5	160 176	+ 50	+ 125 000	+ 20 022 000 000
77.5	86 860	+ 55	+ 166 375	+ 14 451 332 500
82.5	57 676	+ 60	+ 216 000	+ 12 458 016 000
87.5	64 106	+ 65	+ 274 625	+ 17 605 110 250
SUMAS	17 163 932			− 34 907 145 750
				+ 154 050 354 875
				+ 119 143 209 125

TERCER MOMENTO RESPECTO DE LA MEDIA EN UNIDADES ORIGINARIAS

$$\frac{+119\,143\,209\,125}{17\,163\,932} = 6\,941.48.$$

EJEMPLO DE CALCULO SIMPLIFICADO DEL MOMENTO DE TERCER ORDEN CON RESPECTO A LA MEDIA ARITMÉTICA

Cálculo del momento de tercer orden de la población rural de México en 1960, respecto de su media aritmética (aproximada) de 22.5 años

En unidades del intervalo (5)

PUNTOS MEDIOS *Edad media de clase*	FRECUENCIAS *Pobladores en esa edad*	DESVIACIONES EN UNIDADES DE INTERVALO *(entre 5)*	CUBOS DE DESVIACIONES	PONDERACIÓN
2.5	2 910 072	(− 20 ÷ 5 =) − 4	− 64	− 186 244 608
7.5	2 723 844	(− 15 ÷ 5 =) − 3	− 27	− 73 543 788
12.5	2 214 358	(− 10 ÷ 5 =) − 2	− 8	− 17 714 864
17.5	1 753 906	(− 5 ÷ 5 =) − 1	− 1	− 1 753 906
22.5	1 416 731	(0 ÷ 5 =) 0	0	0
27.5	1 220 341	(+ 5 ÷ 5 =) + 1	+ 1	+ 1 220 341
32.5	969 194	(+ 10 ÷ 5 =) + 2	+ 8	+ 7 753 552
37.5	920 758	(+ 15 ÷ 5 =) + 3	+ 27	+ 24 860 466
42.5	642 655	(+ 20 ÷ 5 =) + 4	+ 64	+ 41 129 920
47.5	581 976	(+ 25 ÷ 5 =) + 5	+ 125	+ 72 747 000
52.5	501 350	(+ 30 ÷ 5 =) + 6	+ 216	+ 108 291 600

57.5	376 676	(+ 35 ÷ 5 =)+ 7	+ 343	+ 129 199 868
62.5	367 295	(+ 40 ÷ 5 =)+ 8	+ 512	+ 188 055 040
67.5	195 958	(+ 45 ÷ 5 =)+ 9	+ 729	+ 142 853 382
72.5	160 176	(+ 50 ÷ 5 =)+ 10	+1 000	+ 160 176 000
77.5	86 860	(+ 55 ÷ 5 =)+ 11	+1 331	+ 115 610 660
82.5	57 676	(+ 60 ÷ 5 =)+ 12	+1 728	+ 99 664 128
87.5	64 106	(+ 65 ÷ 5 =)+ 13	+2 197	+ 140 840 882
				− 279 257 166
				+ 1 232 402 839
SUMAS	17 163 932			+ 953 145 673

TERCER MOMENTO RESPECTO DE LA MEDIA EN UNIDADES DEL INTERVALO

$$\frac{+\ 953\,145\,673}{17\,163\,932} = \; \doteq \; 55.53$$

Para obtener el tercer momento con respecto a la media en unidades originarias hay que multiplicar este valor por el cubo (125) del intervalo (5). Así, se obtiene: 55.53×125 años $= 6\,941.25$ como tercer momento respecto de la media en unidades originarias.

EJEMPLO DE CALCULO DEL ÍNDICE DE ASIMETRÍA

Cálculo de la asimetría de la población rural de México, en 1960

TERCER MOMENTO
RESPECTO A LA MEDIA ARITMÉTICA
EN UNIDADES ORIGINARIAS (ya calculado) = 6 941

DESVIACIÓN *CÚBICA* MEDIA = $\sqrt[3]{6\,941} \doteq 19.07$

DESVIACIÓN *CUADRATICA* MEDIA (ya calculada) = 18.75 ("SIGMA")

ÍNDICE DE ASIMETRÍA = DESVIACIÓN CÚBICA MEDIA *SIGMATICA*

$$= \frac{\text{DESVIACIÓN CÚBICA MEDIA}}{\text{SIGMA}} = \frac{19.07}{18.75}$$

$$\doteq 1.017$$

EJEMPLO DE CALCULO DEL ÍNDICE PEARSONIANO DE ASIMETRÍA

Cálculo de la asimetría de la población rural de México, en 1960

SEGUNDO MOMENTO
RESPECTO A LA MEDIA ARITMÉTICA
EN UNIDADES ORIGINARIAS (ya calculado) $\doteq$ 352

TERCER MOMENTO
RESPECTO A LA MEDIA ARITMÉTICA
EN UNIDADES ORIGINARIAS (ya calculado) $\doteq$ 6 941

ÍNDICE PEARSONIANO DE ASIMETRÍA (BETA MINÚSCULA SUBUNO) =

$$\frac{\text{CUADRADO DEL TERCER MOMENTO}}{\text{CUBO DEL SEGUNDO MOMENTO}} =$$

$$= \frac{(6\,941)^2}{(\ 352)^3} = \frac{48\,177\,481}{43\,614\,208} = 1.105.$$

O bien, para calcular el índice pearsoniano de asimetría se tendrá que:

1. Calcular el segundo momento respecto a la media aritmética y elevarlo al cubo.
2. Calcular el tercer momento respecto a la media aritmética y elevarlo al cuadrado.
3. Dividir el resultado obtenido en (2) entre el obtenido en (1).
   A ese resultado se le conoce como el índice pearsoniano de asimetría, y se le representa por beta minúscula subíndice 1 (β_1).

Curtosis, achatamiento o picudez

Para determinar el grado de curtosis de una distribución, o bien se calcula, 1º el cuarto momento con respecto a la media aritmética, y 2º se le divide entre la cuarta potencia de la desviación cuadrática media (o entre la segunda potencia del segundo momento), o:

1. Se calcula la desviación media bicuadrática de la distribución (igual a la raíz cuarta del cuarto momento), y
2. Se le divide entre la desviación media cuadrática de la distribución.

2.4. ALGUNAS TÉCNICAS PARA OBTENER LAS MEDIDAS NECESARIAS PARA LA DESCRIPCIÓN ESTADÍSTICA DE UNA DISTRIBUCIÓN

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:

1. Calcúlense los momentos con respecto a la media arbitraria en unidades del intervalo, para lo cual:
   1.1. Obténganse las desviaciones con respecto a una media elegida, en unidades del intervalo:
   1.11. determinando los puntos medios de los intervalos,
   1.12. eligiendo uno de dichos puntos medios como media arbitraria,

1.13. 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.)

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

1.21. 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),

1.22. 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,

1.23. 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,

1.24. multiplicando los productos de la columna anterior por las desviaciones, a fin de obtener los productos de las cuartas potencias de las desviaciones por las frecuencias, anotando los resultados en otra columna.

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

1.31. 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,

1.32, 1.33, 1.34. 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.

2. Se calcularán las potencias crecientes del primer momento con respecto a la media arbitraria en unidades del intervalo, A, B y C, elevando al cuadrado, al cubo y a la cuarta potencia el cociente obtenido en a del inciso anterior.

3. 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 calculado en 1º y 2º (ver fórmulas de relaciones entre unos momentos y otros en Apéndice 1).

4. 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:
   4.1. Elévese el valor del intervalo a la segunda, a la tercera y a la cuarta potencia.
   4.2. Multiplíquese:
   4.21. el cuadrado del intervalo por el segundo momento, con respecto a la media aritmética en unidades del intervalo,
   4.22. la tercera potencia del intervalo por el tercer momento,
   4.23. 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.
5. Se aplicará la corrección por agrupamiento en caso necesario.
6. La media aritmética se calculará haciendo las substituciones correspondientes en la fórmula:

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

en donde μ_1' representa él primer momento con respecto a la media arbitraria en unidades del intervalo.

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

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

8. 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).

9. 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).

EJEMPLO DE CALCULO CONJUNTO DE LA MEDIA ARITMÉTICA, LA DESVIACIÓN MEDIA CUADRÁTICA Y LAS MEDIDAS DE ASIMETRÍA Y DE CURTOSIS

SERIE DE CLASES Y FRECUENCIAS

Número de asegurados que, en un grupo, están expuestos a enfermar durante el año próximo

CATEGORÍA *edades*	FRECUENCIA *Expuestos de esa edad*	PUNTO MEDIO	DESVIACIÓN	PONDERACIÓN DE LAS POTENCIAS DE LAS DESVIACIONES POR LAS FRECUENCIAS			
	f_i	m_i	d'_i	$d'_i f_i$	$d'^2_i f_i$	$d'^3_i f_i$	$d'^4_i f_i$
14.5 a 19.5	35	17	— 4	— 140	560	— 2 240	8 960
19.5 24.5	145	22	— 3	— 435	1 305	— 3 915	11 745
24.5 29.5	160	27	— 2	— 320	640	— 1 280	2 560
29.5 34.5	150	32	— 1	— 150	150	— 150	150
34.5 39.5	125	37	0	0	0	0	0
39.5 44.5	100	42	1	100	100	100	100
44.5 49.5	90	47	2	180	360	720	1 440
49.5 54.5	75	52	3	225	675	2 025	6 075
54.5 59.5	50	57	4	200	800	3 200	12 800
59.5 64.5	35	62	5	175	875	4 375	21 875
64.5 69.5	25	67	6	150	900	5 400	32 400
69.5 74.5	15	72	7	105	735	5 145	36 015

74.5	79.5	5	77	8	40	320	2 560	20 480
79.5	84.5	3	82	9	27	243	2 187	19 683
84.5	89.5	1	87	10	10	100	1 000	10 000
					− 1 045 + 1 212		− 7 585 + 26 712	
SUMAS		1 014			+ 167	7 763	+ 19 127	184 283
Divididas entre	1 014:				+ 0.1646	7.655	+ 18.862	181.738
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: $\mu'_1 = 0.1646$

Segunda: $\mu'^2_1 = 0.0271$

Tercera: $\mu'^3_1 = 0.0045$

Cuarta: $\mu'^4_1 = 0.0007$

MOMENTOS CON RESPECTO A LA MEDIA ARITMÉTICA

Primero: $\mu_1 = 0$

Segundo:
$$\begin{aligned}\mu_2 &= \mu'_2 - \mu'^2_1 = \\ &= 7.6558 - 0.0271 = \\ &= 7.6287\end{aligned}$$

Tercero:
$$\begin{aligned}\mu_3 &= \mu'_3 - 3\,\mu'_2\mu'_1 + 2\,\mu'^3_1 = \\ &= 18.8629 - 3(7.6558)(0.1646) + 2(0.0045) = \\ &= 18.8629 - 3.7804 + 0.0090 = \\ &= 15.0915\end{aligned}$$

Cuarto:
$$\begin{aligned}\mu_4 &= \mu'_4 - 4\,\mu'_3\mu'_1 + 6\,\mu_2\mu'^2_1 - 3\,\mu'^4_1 \\ &= 181.7386 - 4(15.0915)(0.1646) + 6(7.6287)(0.0271) - \\ &- 3(0.0007) \\ &= 181.7386 - 9.9362 + 1.2404 - 0.0021 = \\ &= 173.0407\end{aligned}$$

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:

Primero: $u_1 = 0.1646 \times 5 = 0.8230$

Segundo: $u_2 = 7.6287 \times 25 = 190.7175$

Tercero: $u_3 = 15.0915 \times 125 = 1\,886.4375$

Cuarto: $u_4 = 173.0407 \times 625 = 108\,150.4375$

OBTENCIÓN DE LA MEDIA ARITMÉTICA

$$\bar{x}_1 = m' + \mu_1 = 37 + 0.8230 = 37.823$$

EJEMPLO DE CÁLCULO CONJUNTO... (conclusión)

OBTENCIÓN DE LA DESVIACIÓN MEDIA CUADRÁTICA

$$\sigma = \sqrt{\mu_2} = \sqrt{190.7175} = 13.81$$

OBTENCIÓN DEL ÍNDICE PEARSONIANO DE ASIMETRÍA

$$\beta_1 = \frac{\mu^2{}_3}{\mu^3{}_2} = \frac{1\,886.4375^2}{190.7175^3} = \frac{3\,558\,646.4414}{6\,936\,999.0577} = 0.5118$$

OBTENCIÓN DEL ÍNDICE PEARSONIANO DE CURTOSIS

$$\beta_2 = \frac{u_4}{u_2{}^2} = \frac{108\,150.4375}{190.7175^2} = \frac{108\,150.4375}{36\,373.1648} = 2.97.$$

2.5. ALGUNAS TÉCNICAS DE INTERPOLACIÓN DE CURVAS TÍPICAS A LAS DISTRIBUCIONES EMPÍRICAS

Las curvas simétricas centrales

De acuerdo con una norma prudente, que nos impone partir de lo conocido para llegar a lo desconocido, tomaremos como punto de partida las distribuciones simétricas ya conocidas. O sea, que usaremos la distribución binomial (discontinua) que tiende hacia la distribución normal (continua).

La distribución binomial puede expresarse en la forma siguiente:

$$f(S) = \binom{M}{S}\, p^{S}\, (1\text{-}p)^{M\text{-}S}$$

En esta fórmula: f(S) representa la frecuencia con la que aparece el carácter S (número de soles obtenidos, por ejemplo, al tirar al aire un cierto número de monedas); M representa el número de individuos del conjunto (número de monedas que se arrojan al aire); p, la probabilidad que hay de que salga un sol en una moneda.

Por consiguiente, 1-p representa la probabilidad (complementaria) que hay de que *no* salga el carácter considerado (la probabilidad de que salga un águila). M-S señala el número (complementario respecto a M) de individuos que, en el conjunto, *no* mostraron el carácter S (número de monedas que, tras haber caído, muestran águilas y no soles en su cara visible.

En el estudio de la distribución binomial se muestra que p y 1-p pueden ser diferentes entre sí. Esto es lo que ocurre cuando se trata de determinar la probabilidad que hay de obtener un 6 al lanzar un dado (p = 1/6) y la —complementaria— que hay de obtener un número que no sea 6 al lanzar dicho dado ($1\text{-}p = 5/6$). Sin embargo, para partir del caso más sencillo, consideraremos una distribución binomial en la que p y 1-p son iguales (según ocurre en el caso de las probabilidades que hay de obtener un sol o un águila con una moneda, ya que, en tal caso, $p = 1/2$ y 1-p ($= 1 - 1/2$) también es igual a 1/2.

En el caso de una binomial como ésta, la expresión anterior se convierte en:

$$f(S) = \binom{M}{S} p^{S} p^{M\text{-}S} = \binom{M}{S} p^{M}$$

La expresión que acabamos de escribir da la frecuencia con la que aparecen S soles con M monedas. La expresión que da la frecuencia con la que aparecerán $S + 1$ soles, si se utiliza el mismo número (M) de monedas es parecida. Si restamos esta última expresión de la previa, obtendremos el incremento que sufren las frecuencias, cuando la variable (que ha pasado de S a $S + 1$) se incrementa en una unidad.

En forma parecida, si sumamos las dos expresiones y dividimos la suma resultante entre dos, obtendremos la frecuencia correspondiente a un valor intermedio.

Si dividimos el incremento o expresión primera entre la frecuencia intermedia, obtendremos la pendiente, o sea:

$$\frac{\Delta f(S)}{f(S + .5)} = \frac{2p^{M}}{p^{M}} \frac{\left[\binom{M}{S+1} - \binom{M}{S}\right]}{\left[\binom{M}{S+1} + \binom{M}{S}\right]}$$

Mediante una serie de sustituciones se llega a escribir una expresión en la que pueden reconocerse: en el numerador, un término variable y la suma de dos términos constantes en el denominador, dos términos constantes. Si representamos por x la expresión variable, y por a y por b las sumas de constantes, la expresión anterior se puede reducir a:

$$\frac{x + a}{b}$$

Por lo que se refiere al primer miembro de la igualdad, éste se puede considerar, en el límite, como la derivada de la función con respecto a la variable, dividida (la derivada) entre la función misma; o sea:

$$\frac{D_x y}{y}$$

Consiguientemente, toda la expresión puede escribirse:

$$\frac{D_x y}{y} = \frac{x + a}{b}$$

Esta expresión corresponde a la forma diferencial hacia la que tiende la binomial cuando de discontinua pasa a continua; o sea, cuando en vez de crecer "por saltos", crece por incrementos infinitamente pequeños.

A fin de mostrar el proceso que lleva de la forma diferencial de la ecuación a al integral, tomaremos las integrales de ambos miembros de la ecuación (con respecto a x):

$$\int_x \frac{D_x y}{y} = \int_x \frac{x + a}{b}$$

La integral del primer miembro es igual al logaritmo de la función (o sea, que todo el primer miembro se reduce a Ly). En cuanto al segundo miembro, es la integral de dos fracciones, de las que una tiene por numerador a x y por denominador (común) a b, mientras la otra tiene por numerador a a, y por denominador a b. La integración de estas fracciones nos permite escribir:

$$Ly = \frac{x^2 + 2ax}{2b} + C$$

Al tomar antilogaritmos de ambos miembros, obtenemos:

$$y = e^{\left(\frac{x^2 + 2ax}{2b} + C\right)}$$

En el segundo miembro figura e (base de los logaritmos naturales) elevado a una fracción más C (constante de integración). O sea, que dicho miembro se puede descomponer en dos factores exponenciales: de ellos, el primero es e elevado a la fracción; el segundo es e elevado a C. Si, además, a e elevado a C lo representamos por Y', la expresión anterior se puede escribir como:

$$y = y' e^{\frac{x^2 + 2ax}{2b}}$$

Ésta es la forma integral de la llamada curva normal. En ella existen: una variable (x) y dos parámetros (a y b) cuyos valores hay que determinar. El valor de y' no nos interesa por el momento.

Para la determinación del valor de los parámetros recurriremos a los momentos. Si utilizamos desviaciones con respecto al origen, operaremos con los datos mismos x, y si suponemos que la serie es de frecuencias tendremos que multiplicar sus potencias respectivas por la frecuencia, que está representada —en este caso— por y:

$$v_n = \frac{\Sigma x^n y}{N}$$

Ésta es la fórmula del enésimo momento de la serie discontinua. Pero, si en vez de ser discontinua es continua la serie, habrá que transformar la sumatoria en integral (sustituir el operador sigma mayúscula por el operador ese alargada). Según esto, la fórmula del enésimo momento, para una distribución continua como la que nos interesa, resulta ser:

$$v_n = \int_x x^n y$$

Si en la expresión diferencial de la curva continua quitamos denominadores (pasando b que está como denominador en el segundo al primer miembro, como factor, y pasando igualmente a y que figuraba como denominador del primer miembro al segundo, como factor) tendremos:

$$b D_x y = y (x + a)$$

Si se parte de esta expresión concreta, se puede obtener el enésimo momento de la distrbuición. Para ello, a ambos miembros se les deberá multiplicar por la enésima potencia de la variable independiente (x^n).

Para obtener el enésimo momento, se necesitará —además— integrar ambos miembros con respeto a x (o sea, que habrá que anteponerles el operador $\int$). Así, se obtendrá:

$$\int_x b D_x y = \int_x y (x + a)$$

Así, se obtiene:

$$v_{n+1} = a v_n - b n v_{n-1}$$

Esta expresión genérica resultante nos indica que, en esta curva típica: si 1) se multiplica por el parámetro a, el enésimo momento; 2) se multiplica el momento anterior a él por b y por n, y 3) se resta del primer resultado el segundo, se obtendrá el momento siguiente (el de orden n + 1).

Si —por una parte— tenemos una distribución y sus momentos ya calculados y —por otra— tenemos la expresión genérica que puede corresponderle, podemos determinar la curva específica que se adecúe mejor a la distribución, al determinar (primero) los valores de a y de b de la distribución empírica y sustituirlos (después) en la expresión teórica general.

Para calcular los parámetros a y b hay que recordar que en cuanto representan dos incógnitas necesitaremos dos ecuaciones simultáneas que las contengan. A su vez, para tener esas ecuaciones, habremos de especificar la expresión genérica (que da la relación entre los momentos de la curva) dando a n los valores sucesivos 0 y 1. Así, obtendremos:

$$
\begin{aligned}
av_0 + b(0)v_{-1} &= -v_1 \\
av_1 + b(1)v_0 &= -v_2
\end{aligned}
$$

Al resolver este sistema resultante, se obtienen como fórmulas para los parámetros a y b las siguientes:

$$
\begin{aligned}
a &= -v_1 \\
b &= -v_2 + v_1^2
\end{aligned}
$$

Conocidas la expresión genérica y las fórmulas de los parámetros de esa expresión en términos de los momentos, se puede interpolar una curva normal a una serie cuyos dos primeros momentos con respecto al origen (v_1 y v_2) pueden calcularse por cómputos ya conocidos.

La retención de la fórmula es más fácil si se recuerda que $+v_1$ (primer momento con respecto al origen) no es otra cosa que la media aritmética y que $v_2 - v_1^2$ (simétrico de b) es la fórmula para la desviación cuadrática media en términos de los momentos, elevada al cuadrado.

De acuerdo con esto, la expresión para la curva normal puede escribirse como:

$$
y = y' e^{-\frac{\delta^2}{2}}
$$

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:

1. Un valor constante (el de la ordenada máxima).
2. 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).

1. Para calcular el valor de la ordenada máxima:

1.1. Se obtiene el efectivo de la distribución sumando todas las frecuencias.

1.2. 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).

1.3. 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.

2. Para obtener el valor del factor exponencial e:

$$-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)$$

2.1. Se resta de cada dato (/ de cada punto medio) la media aritmética de la distribución.

2.2. Se divide dicha desviación entre la desviación cuadrática media de la distribución.

2.3 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).

Precaución en el uso de la tabla de ordenadas

Las tablas de ordenadas de la curva normal suelen presentar dos formas alternativas pues: o listan los valores de las ordenadas de la curva normal, expresadas en fracciones de la ordenada máxima, o listan los valores que se obtienen al dividir los anteriores entre 2.506628 (o sea entre la raíz cuadrada de dos pi).

Al primer tipo de tabla corresponde la del Apéndice V de Frederick Cecil Mills, *Métodos estadísticos aplicados a la economía y a los negocios.* Al segundo tipo corresponde la tabulación de Charles Hodgman, *Mathematical Tables.* Este segundo tipo de tabla hace que resulte redundante el uso del factor 2.506628 del paso 1.2.

EJEMPLO DE
INTERPOLACIÓN DE UNA CURVA NORMAL
MEDIANTE LA APLICACIÓN DIRECTA DE SUS FÓRMULAS

Datos de la serie por interpolar		Cálculo de μ_2				Cálculo de Y_0
M	f	M f	d	d^2	d^2 f	$Y_0 =$
60	9	540	—90	8 100	72 900	$= \frac{N i}{\sqrt{2 \pi \mu_2}} =$
81	20	1 620	—69	4 761	95 220	
102	33	3 366	—48	2 304	76 032	
123	47	5 781	—27	729	34 263	$= \frac{300 \times 21}{2.5 \times 46}$
144	55	7 920	— 6	36	1 980	
165	52	8 580	15	225	11 700	
186	39	7 254	36	1 296	50 544	
207	26	5 382	57	3 249	84 474	
229	12	2 748	79	6 241	74 892	
249	5	1 445	99	9 801	49 005	
270	2	540	120	44 400	88 800	
	300	45 176				$Y_0 = 55$

$\boxed{\bar{x}_a = 150.39}$

EJEMPLO DE
INTERPOLACIÓN DE UNA CURVA NORMAL
MEDIANTE LA APLICACIÓN DIRECTA DE SUS FÓRMULAS

$$Y = Y_0 \, e^{-\sigma^2/2} = 55 \, e^{-\sigma^2/2}$$

Cálculo de las yes teóricas

σ	σ^2	$-\sigma^2/2$	$e^{\sigma^2/2}$	$Y_0 \, e^{-\sigma^2/2}$
—1.96	3.8416	—1.92	0.146607	8
—1.50	2.2500	—1.13	0.323033	17
—1.04	1.0816	—0.54	0.582748	32
—0.58	0.3369	—0.17	0.843665	46
—0.13	0.0169	—0.08	0.990050	54
0.33	0.1089	—0.05	0.951229	52
0.79	0.6241	—0.31	0.733447	40
1.24	1.5376	—0.87	0.418952	23
1.72	2.9584	—1.43	0.239309	13
2.15	4.6225	—2.31	0.099261	5
2.61	6.8121	—3.41	0.033041	1
			TABLAS	

Se recomienda el uso de Charles D. Hodgman, *Mathematical Tables.* Chemical Rubber Publishing Co. Cleveland, Ohio. La tabla de la función exponencial, que aquí se utiliza, figura en las páginas 151-7.

Sistemas de curvas asimétricas
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:

$$y_t = \sum f_i \frac{i}{\sigma} \left[\varphi_0 (\delta_1) - \frac{\sqrt{\beta_1}}{6} \varphi_3 (\delta_i) + \frac{\beta_2 - 3}{24} \varphi_4 (\delta_i) \right]$$

En esta fórmula:

f_1 representa las frecuencias,

i representa el intervalo es la desviación media cuadrática

φ (δ_1) representa los valores listados como "ordenadas" en la tabla (Hodgman, *Mathematical Tables,* pp. 209-13 *"ordinate"*)

φ_3 (δ_1) la tercera derivada de esa función, cuyos valores se listan en la tabla (*third derivative*)

φ_4 (δ_1) su cuarta derivada (*fourth derivative*)

β_1 el índice pearsoniano de asimetría

β_2 el índice pearsoniano de curtosis.

El ajuste de una curva de Gram-Charlier se hará conforme al siguiente procedimiento:

1. 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).

1.1.-A. restando de cada punto medio la media aritmética, y
1.2.-A. dividiendo los resultados entre la desviación media cuadrática o.

B. (Indirecto).

1.1.-B. 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,
1.2.-B. obteniendo el cociente del intervalo entre la desviación cuadrática media, y
1.3.-B. multiplicando por dicho cociente los residuos obtenidos en a.

2. 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.
3. 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.
4. 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 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.

5. Se suman los valores:

5.1. de las ordenadas,

5.2. de los productos de las terceras derivadas (φ_3) por su coeficiente,

$$\left(\frac{\sqrt{\beta_1}}{6} \right) \text{ y}$$

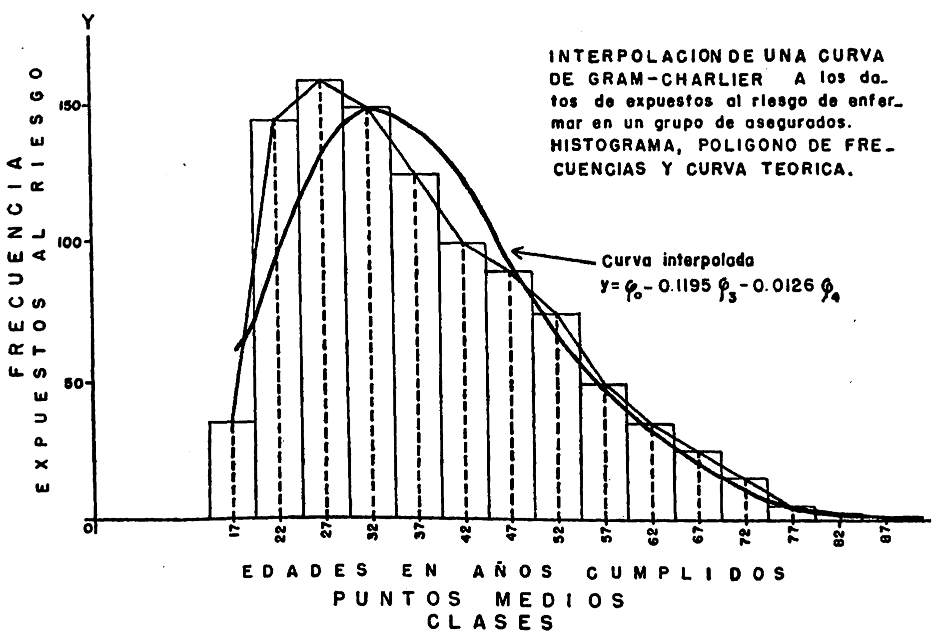
5.3. de los productos de las cuartas derivadas (φ_4) por su coeficiente

$$\left(\frac{\beta_2 - 3}{24} \right)$$

teniendo cuidado de hacer esa suma algebraicamente, o sea, teniendo en cuenta los signos.

6. 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.

7. Multiplíquese el producto obtenido en 6º, por los valores de las sumas algebraicas obtenidas en 5°. Los resultados, contenidos en una última columna, serán los valores teóricos de ajuste en la curva de Gram-Charlier correspondiente.



EJEMPLO DE PERECUACIÓN DE UNA CURVA DE GRAM-CHARLIER

DISTRIBUCIÓN DE CLASES Y FRECUENCIAS

Número de asegurados que —en un grupo— están expuestos a enfermar durante el año próximo

PUNTOS MEDIOS m_i	FRECUENCIA f_i	DESVIACIÓN $m_i - x_1$	DESVIACIÓN SIGMÁTICA $\frac{m_i - x_1}{\sigma}$
17	35	-20.823	-1.53
22	145	-15.823	-1.17
27	160	-10.823	-0.80
32	150	-5.823	-0.43
37	125	-0.823	-0.06
42	100	4.177	$+0.31$
47	90	9.177	$+0.68$
52	75	$+14.177$	$+1.04$
57	50	$+19.177$	$+1.41$
62	35	$+24.177$	$+1.78$
67	25	$+29.177$	$+2.15$
72	15	$+34.177$	$+2.52$
77	5	$+39.177$	$+2.88$
82	3	$+44.177$	$+3.25$
87	1	$+49.177$	$+3.62$

Valores obtenidos anteriormente:

$\bar{x}_1 = 37.823$

$\sigma = 13.81$

$\beta_1 = 0.5118$

$\beta_2 = 2.97$

EJEMPLO DE PERECUACIÓN DE UNA CURVA DE GRAM-CHARLIER

DISTRIBUCIÓN DE CLASES Y FRECUENCIAS (CONTINUACIÓN)

Número de asegurados que —en un grupo— están expuestos a enfermar durante el año próximo (cálculo de las φ's)

VALOR TABULADO DE φ_0	VALOR TABULADO DE φ_3	SIGNO	$-167\sqrt{\beta_1}\varphi_3$ $-0.1195\sqrt{\beta_1}\varphi_3$	φ_4	$0.42(\beta_2-3)\varphi_4$ $-0.0126\varphi_4$
0.1238	+0.1238	—	+0.0149	—0.6888	+0.0087
0.2012	+0.3840	—	+0.0459	—0.6720	+0.0085
0.2897	+0.5469	—	+0.0654	—0.1247	+0.0016
0.3637	+0.4403	—	+0.0526	+0.7001	—0.0088
0.3982	+0.0716	—	+0.0085	+1.1894	—0.0150
0.3802	+0.3423	+	—0.0409	+0.9250	—0.0117
0.3166	+0.5463	+	—0.0652	+0.1391	—0.0018
0.2323	+0.4635	+	—0.0553	—0.5389	+0.0068
0.1476	+0.2107	+	—0.0252	—0.7347	+0.0093
0.0818	—0.0245	—	+0.0029	—0.4887	+0.0062
0.0396	—0.1380	—	+0.0165	—0.1332	+0.0017
0.0167	—0.1408	—	+0.0168	+0.0871	—0.0011
0.0063	—0.0962	—	+0.0115	+0.1384	—0.0017
0.0020	—0.0499	—	+0.0060	+0.1039	—0.0013
0.0006	—0.0208	—	+0.0025	+0.0547	—0.0007

Valores que fue necesario calcular para continuar estos desarrollos:

$$\sqrt{\beta_1} = \sqrt{0.5118} = 0.7154$$

$$0.167\sqrt{\beta_1} = 0.167 \times 0.7154 = 0.1195$$

— — — — — — — — — — —

$$\beta_2-3 = 2.97-3 = -0.03$$

$$0.42(\beta_2-3) = 0.42(-0.03) = -0.0126$$

EJEMPLO DE PERECUACIÓN DE UNA CURVA DE GRAM-CHARLIER

DISTRIBUCIÓN DE CLASES Y FRECUENCIAS (CONTINUACIÓN II)

Número de asegurados que —en un grupo— están expuestos a enfermar durante el año próximo (cálculo de la ordenada y las frecuencias teóricas)

SUMAS $\varphi_0 - 0.167\sqrt{\beta_1}\varphi_3 - 0.42(\beta_2 - 3)\varphi_4$ $\varphi_0 - 0.1195\ \ \varphi_3 - 0.0126\ \ \ \varphi_4$	FRECUENCIAS TEÓRICAS Sumas por ordenada Sumas por 367.13
0.1474	54.11
0.2556	93.80
0.3567	130.95
0.4075	149.61
0.3917	143.80
0.3510	128.86
0.2532	92.96
0.1838	67.48
0.1317	48.35
0.0909	33.37
0.0578	21.22
0.0324	11.90
0.0161	5.91
0.0067	2.46
0.0024	0.88

Cálculo de la ordenada, necesaria para el cálculo de los valores teóricos:

$$\text{Ordenada} = \frac{\Sigma f_i^1}{\sigma} = \frac{1\,014 \times 5}{13.81} = \frac{5\,070}{13.81} = 367.13$$

El sistema asimétrico

El recorrido que hemos hecho por terreno parcialmente conocido, nos permite ver cuál es el método que hay que seguir para obtener fórmulas que sirvan para la descripción estadística de series que ya no sean simétricas (como la curva normal) sino asimétricas.

Un recorrido paralelo al que hemos seguido antes consiste en:

1. Partir de la expresión de una curva discontinua apropiada.
2. Determinar la relación que existe entre el incremento unitario y la frecuencia media del intervalo unitario de esa curva.
3. Determinar, en la expresión resultante, cuáles son los términos constantes y cuáles los variables, y representarlos en la forma más simple.
4. Pasar de la expresión diferencial a la expresión integral de la curva correspondiente, y
5. Determinar la relación entre los parámetros de la expresión obtenida y los momentos de la distribución.

En el caso anterior, nuestro punto de partida estuvo en una binomial simétrica, en la que p era igual a l-p. En este caso, en que no nos interesa obtener una curva continua simétrica, sino un sistema de curvas continuas asimétricas, tomaremos como punto de partida una binomial asimétrica (o sea, una en la que p sea diferente de l-p).

En la fórmula correspondiente, se puede suponer que M es el número de dados empleados en una jugada, y que S es el número de los que dieron un 6, mientras que M-S es el número de los que dieron una puntuación distinta de 6.

Si mediante esta fórmula se calcula el incremento de las frecuencias para una unidad del incremento de la variable, se calcula —en seguida— la frecuencia media del intervalo, y se las relaciona, podrá obtenerse, en el segundo miembro, una expresión fraccionaria. En el numerador de ésta figurarán variables y constantes, y en su denominador (a diferencia de lo que ocurrió en el caso anterior, en el que sólo figuraron constantes) figuran también variables y constantes.

De acuerdo con esto si, como en casos previos, designamos por x a la variable, y representamos con a, b y c a las constantes, podremos escribir:

$$\frac{D_x y}{y} = \frac{x + a}{bx + c}$$

En este caso, hemos igualado el segundo miembro de la expresión original a $(D_x y)/y$, por razones análogas a las que nos movieron a hacerlo en el caso anterior. La expresión es la fórmula diferencial del sistema de curvas asimétricas.

Cuando se utilizan desviaciones con respecto a la media aritmética, la expresión diferencial es:

$$\frac{D_d y}{y} = \frac{d + a'}{bd' + c'}$$

Si, para simplificar aún más la expresión, en vez de tomar desviaciones respecto a la media en unidades originarias (d), tomamos desviaciones con respecto a la media en unidades sigmáticas (δ) los parámetros también variarán:

$$\frac{D_\delta y}{y} = \frac{\delta + \alpha}{\beta_1 \delta + \beta o}$$

Esta expresión —a su vez— es más fácil de integrar si sustituimos a δ por D— β_0 / β_1. Después de hecha la sustitución y de realizadas las operaciones, la expresión puede escribirse como sigue si a β_0 / β_1 se le representa por A:

$$\frac{D\ y}{y} = \frac{1}{\beta_1} \left(1 + \frac{A}{D}\right)$$

Ésta es la fórmula diferencial del sistema de ecuaciones que se estudia.

Para obtener la forma integral del sistema, bastará con anteponer el integrador ($\int$) a los dos miembros, refiriéndolo: a x, en el primero ($\int_x$) y a D, en el segundo ($\int_D$). Como en el segundo miembro aparece un factor constante, el mismo puede quedar fuera del integrador, que afectará —por tanto— sólo a la suma de dentro del paréntesis:

$$\int_x \frac{D_x y}{y} = \frac{1}{\beta_1} \int_D \left(1 + \frac{A}{D}\right)$$

Así, se obtiene:

$$y = Y \left(e^{D} D^{A}\right)^{1/\beta_1}$$

Para obtener la segunda expresión genérica, o sea, la fórmula para Y integramos la anterior. Y puede quedar fuera del integrador ya que como factor constante puede entrar o salir del integrando sin alterarse o alterar la expresión:

$$\int_0^\alpha y = Y \int_0^\alpha e^{D/\beta_1} D^{A/\beta_1}$$

El factor integral del segundo miembro es muy característico: se le conoce como la función gamma mayúscula (Γ) del exponente de la variable (D), o sea: A/β_1 adicionado (dicho exponente) de una unidad $[(A/\beta_1) + 1]$.

Según esto, toda la expresión previa se transforma en la que sigue si —además— para facilitar la escritura, representamos el resultado inmediatamente anterior (exponente de D, + 1) por R; y si —después— despejamos de la ecuación resultante a Y:

$$Y = \frac{N}{(-\beta_1)^R \ \Gamma\,(R)}$$

Para especificar las expresiones genéricas habrá que encontrar —como en el primer caso— la equivalencia de los parámetros en términos de los momentos.

En el caso anterior, usamos momentos respecto al origen porque nuestra expresión diferencial contenía x ("desviaciones con respecto al origen"). Como en este caso tratamos con desviaciones sigmáticas, nuestros momentos tendrán que serlo con respecto a la media aritmética en unidades sigmáticas.

Los siguientes desarrollos muestran los resultados obtenidos al sustituir n por 0, 1, y 2, efectuar las reducciones y despejar los valores de los tres parámetros α, β_0 y β_1.

En último término, en esos desarrollos, hemos sustituido esos valores en las fórmulas de los parámetros complejos: D, A y R.

Con ayuda de las dos fórmulas genéricas y de estas fórmulas de los parámetros, es posible interpolar una curva asimétrica de este sistema a una distribución empírica.

$$\begin{array}{llllll}
\alpha\gamma & - \ \beta_0\gamma_{n-1} & - & (n-1)\ \beta_1\gamma & = & -\gamma_{n+1} \\
\alpha\gamma_0 & - \ 0\ \beta_0\gamma_{-1} & - & 1\ \beta_1\gamma_0 & = & -\gamma_1 \\
\alpha\gamma_1 & - \ \beta_0\gamma_0 & - & 2\ \beta_1\gamma_1 & = & -\gamma_2 \\
\alpha\gamma_2 & - \ 2\ \beta_0\gamma_1 & - & 3\ \beta_1\gamma_2 & = & -\gamma_3 \\
\hline
\alpha & - \ 0 & - & \beta_1 & = & 0 \\
0 & - \ \beta_0 & - & 0 & = & -1 \\
\alpha & - \ 0 & - & -3\ \beta_1 & = & -\gamma_3 \\
\hline
\alpha & = -\beta_1 & & & & \\
 & \beta_0 & = & -\ 1 & & \\
 & & & \beta_1 & = & \dfrac{\gamma_3}{2} \\
\hline
\end{array}$$

Así, los parámetros complejos resultan ser:

$$\boxed{D = \delta - \frac{1}{\beta_1}}$$

$$\boxed{A = -\beta - \frac{1}{\beta_1}}$$

$$\boxed{R = \frac{1}{\beta_1^2}}$$

Interpolación de curvas asimétricas

Procedimiento detallado

1. Calcular el tercer momento con respecto a la media en unidades sigmáticas. Para ello habrá que:
   1.1. Calcular la media aritmética.
   1.2. Restarla de cada dato, para obtener desviaciones.
   1.3. Elevar al cuadrado las desviaciones obtenidas, sumar los cuadrados, dividir la suma entre el efectivo, y extraer la raíz cuadrada del cociente, para obtener la desviación cuadrática media.
   1.4. Dividir cada diferencia o desviación entre la desviación cuadrática media, para obtener desviaciones sigmáticas.
   1.5. Elevar al cubo los cocientes, sumar los cubos, y dividir entre el efectivo su suma.
2. Dividir el tercer momento sigmático entre 2, y cambiar el signo del cociente para obtener β_1.
3. Tomar el recíproco de beta-uno y restarlo de δ (delta) para obtener las Des (mayúsculas).
4. Tomar β_1, cambiarle signo y sumarle el recíproco de β_1 para tener A.
5. Tomar el recíproco de β_1 y dividirlo entre β_1 para obtener el recíproco del cuadrado, o sea R.

6. Obtener el valor de Y:
   6.1. buscando, en tablas, la función gamma de R,
   6.2. elevando β_1 a la R,
   6.3. multiplicando entre sí los dos resultados anteriores,
   6.4. dividiendo el efectivo entre el producto así obtenido.
7. Sustituir el valor de los diferentes parámetros en la fórmula genérica del sistema de curvas que nos ocupa.
8. Dar valores a δ en esa ecuación (ya especificada mediante la sustitución de los valores de los parámetros) para obtener los valores teóricos de las frecuencias.

Tabla de la función gamma (Γ)

n	Γ(n)	n	Γ(n)	n	Γ(n)	n	Γ(n)
1.00	1.0000	1.25	.9064	1.50	.8862	1.75	.9191
1.01	.9943	1.26	.9044	1.51	.8865	1.76	.9214
1.02	.9888	1.27	.9025	1.52	.8870	1.77	.9238
1.03	.9835	1.28	.9007	1.53	.8876	1.78	.9262
1.04	.9784	1.29	.8990	1.54	.8882	1.79	.9288
1.05	.9735	1.30	.8975	1.55	.8889	1.80	.9314
1.06	.9687	1.31	.8960	1.56	.8896	1.81	.9341
1.07	.9641	1.32	.8946	1.57	.8905	1.82	.9369
1.08	.9597	1.33	.8934	1.58	.8914	1.83	.9397
1.09	.9555	1.34	.8922	1.59	.8924	1.84	.9426
1.10	.9513	1.35	.9811	1.60	.8935	1.85	.9456
1.11	.9474	1.36	.8902	1.61	.8947	1.86	.9487
1.12	.9436	1.37	.8893	1.62	.8959	1.87	.9518
1.13	.9399	1.38	.8885	1.63	.8972	1.88	.9551
1.14	.9364	1.39	.8879	1.64	.8986	1.89	.9584
1.15	.9330	1.40	.8873	1.65	.9001	1.90	.9618
1.16	.9298	1.41	.8868	1.66	.9017	1.91	.9652
1.17	.9267	1.42	.8864	1.67	.9033	1.92	.9688
1.18	.9237	1.43	.8860	1.68	.9050	1.93	.9724
1.19	.9209	1.44	.8858	1.69	.9068	1.94	.9761
1.20	.9182	1.45	.8857	1.70	.9086	1.95	.9799
1.21	.9156	1.46	.8856	1.71	.9106	1.96	.9837
1.22	.9131	1.47	.8856	1.72	.9126	1.97	.9877
1.23	.9107	1.48	.8857	1.73	.9147	1.98	.9917
1.24	.9085	1.49	.8859	1.74	.9168	1.99	.9958

EJEMPLO DE INTERPOLACIÓN DE UNA CURVA ASIMÉTRICA

Valores de μ_2 y de μ_3 de la serie por interpolar

$$\mu_2 = 1.41787 \qquad \mu_3 = 3.606622$$

$$\gamma_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{3.606622}{1.731200} \doteq 2.08331$$

$$\beta_1 = \frac{-\gamma_3}{2} = \frac{-2.08331}{2} = -1.0417$$

$$\frac{1}{\beta_1} = \frac{1}{-1.0417} = -0.959969$$

$$D = \delta - \frac{1}{\beta_1} = \delta + 0.959969$$

$$A = -\beta_1 - \frac{1}{\beta_1} = 1.0417 - 0.959969 = 0.081731$$

$$R = \frac{1}{\beta^2} = \frac{-0.959969}{-1.0417} = 0.92157$$

Cálculo de Y

$$Y = \frac{N}{(-\beta)^R\,\Gamma(R)} = \frac{251}{1.0417^{0.92157}\,\Gamma(0.92157)}$$

$$\Gamma(R-1) = \Gamma(R)/(R-1)$$

$$\Gamma(0.92157) = \frac{\Gamma(1.92157)}{0.92157} = 1.05302$$

$1.04^{0.92}$

$$\log 1.04 = 0.01703$$
$$x\ 0.92 = 0.01567$$
$$\text{antilogaritmo de} \quad 0.01567 \doteq 1.04$$

$$Y = \frac{251}{1.04 \times 1.05} = \frac{251}{1.09}$$

$$Y = 230.2$$

EJEMPLO DE INTERPOLACIÓN DE UNA CURVA ASIMÉTRICA

(continuación)

Substitución en la fórmula genérica.

$$Y = 192 \left[e^{\delta+0.96} (\delta+0.96)^{0.08} \right]^{-0.96}$$

Transformación de los datos en desviaciones sigmáticas + K

Datos de la serie		Cálculo de las desviaciones		Transformación
X	f	d	δ	δ + 0.96
1	44	—1.3346612	—1.11	—0.15
2	135	—0.3336612	—0.28	0.68
3	45	0.6653388	0.55	1.51
4	12	1.6653388	1.39	2.35
5	8	2.6653388	2.22	3.18
6	3	3.6653388	3.05	4.01
7	1	4.6653388	3.88	4.84
8	3	5.6653388	4.72	5.68

Fórmula genérica escrita en forma logarítmica

$$\log Y = \log 192 - 0.96 \left[(\delta + 0.96) \log e + 0.08 \log (\delta + 0.96) \right]$$

Cálculo de las yes teórica

I D log e	II log D	III 0.08 log D	IV 0.96 por (I + III)	(V) log 192 menos 0.96 (I+III)	Y = Antilog (V)
0.2953172	—1.83251	—0.0143992	0.269681280	2.01462	103
0.6557779	0.17898	0.0143184	0.643292448	1.64001	44
1.0205815	0.37107	0.0296856	1.008256416	1.27504	19
1.3810422	0.50245	0.0401944	1.364387136	0.91891	8
1.7415029	0.60314	0.0482512	1.718163936	0.56514	4
2.1019636	0.68485	0.0547880	2.071481536	0.21282	2
2.4667672	0.75435	0.0603480	2.426030592	—0.14373	1
				—1.75627	

2.6. ALGUNAS TÉCNICAS PARA EL ANÁLISIS Y LA SÍNTESIS DE LAS SERIES DINÁMICAS

Concepto y obtención de las sumas móviles

Si se tiene una serie temporal, se conoce como "suma móvil", la que se obtiene de sumar un cierto número de valores sucesivos: "suma móvil de dos en dos" es la que se obtiene al sumar el primero de los valores de la variable con el segundo; de sumar por separado el "segundo de esos valores con el tercero", etcétera, hasta terminar con la suma del penúltimo y el último que constituye la última suma móvil de dos en dos. Si la suma móvil es "de tres en tres" se obtendrá de sumar, por un lado, el primero, el segundo y el tercer valor de la variable; en seguida, el segundo, el tercero y el cuarto, para terminar con la suma del antepenúltimo, el penúltimo y el último, que constituirá la última suma móvil de tres en tres.

En general, una suma móvil de n en n se obtiene si se suman, primero, los términos del primero al de orden ene; en seguida, por su lado, los valores del segundo al enésimo más uno (o sea, al que sigue al enésimo); después, por otra parte, los valores del tercero al que está dos lugares después del enésimo . . . para terminar con la suma de los ene términos que anteceden al último, que son los que constituyen la última "suma móvil de n en n".

Concepto y obtención de las medias móviles

En forma parecida a como hemos dado la noción de suma móvil, podemos dar la de media móvil. En efecto, cuando se divide cada una de las sumas móviles, obtenidas por el procedimiento anterior, entre el número de datos que constituyen cada suma móvil, se obtiene una media móvil. Así, la media móvil de ene en ene datos se obtiene sumando los valores de la serie, de ene en ene, comenzando por los ene primeros y terminando por los ene últimos, y dividiendo cada suma entre ene.

Utilidad de las medias móviles

El cálculo de las medias móviles sirve para descubrir la línea de tendencia de las series dinámicas empíricas. Cuando se superpone su representación gráfica a la de la serie empírica, se observa que las medias móviles *suavizan* el trazo de la curva.

EJEMPLO DE CALCULO DE SUMAS MÓVILES

Sumas móviles de 12 en 12 meses de los depósitos del sistema bancario mexicano
Se ejemplifica sólo con algunos meses de 1960 a 1961

Años	*Mes*	*Depósitos en millones de pesos*	*Sumas móviles de 12 en 12 meses*
1960	E	10 077	
			No hay suma móvil de 12 en 12; sólo hay 7 valores disponibles: E—J
	F	10 332	
			No hay suma móvil de 12 en 12; sólo hay 8 valores disponibles: E—A
	M	10 216	
			No hay suma móvil de 12 en 12; sólo hay 9 valores disponibles: E—S
	A	10 381	
			No hay suma móvil de 12 en 12; sólo hay 10 valores disponibles: E—O
	M	10 552	
			No hay suma móvil de 12 en 12; sólo hay 11 valores disponibles: E—N
	J	10 358	
			10 077 + 10 332 + 10 216 + 10 381 + 10 552 + 10 358 + 10 418 + 10 753 + 10 773 + 10 663 + 10 934 + 10 722 = 126 179
	J	10 418	
			10 332 + 10 216 + 10 381 + 10 552 + 10 358 + 10 418 + 10 753 + 10 773 + 10 663 + 10 934 + 10 722 + 10 495 = 126 597
	A	10 753	
			10 216 + 10 381 + 10 552 + 10 358 + 10 418 + 10 753 + 10 773 + 10 663 + 10 934 + 10 722 + 10 495 + 10 561 = 126 826
	S	10 773	
			126 826 — 10 216 + 10 601 = 127 211
	O	10 663	
			127 211 — 10 381 + 11 353 = 128 183
	N	10 934	
			No hay suma móvil de 12 en 12; sólo hay 11 valores disponibles: J—A
	D	10 722	
			No hay suma móvil de 12 en 12; sólo hay 10 valores disponibles: A—A
1961	E	10 495	
			No hay suma móvil de 12 en 12; sólo hay 9 valores disponibles: S—A
	F	10 561	
			No hay suma móvil de 12 en 12; sólo hay 8 valores disponibles: O—A
	M	10 601	
			No hay suma móvil de 12 en 12; sólo hay 7 valores disponibles: N—A
	A	11 353	
	M		

El número de sumas y de medias móviles de una serie es siempre menor que el número de datos

Al calcular las sumas móviles y las medias móviles correspondientes se obtienen siempre menos sumas o medias que datos. En el caso de las medias móviles de 3 en 3 hay 2 sumas móviles menos que datos; en el de las medias móviles de 5 en 5, hay 4 sumas y medias móviles menos que datos. Cuando se calculan medias de 2 en 2 datos hay una suma y una media móviles menos que datos; si se calculan de 4 en 4 hay 3 sumas y 3 medias móviles menos que datos.

Correspondencia entre las medias móviles y los datos

Cuando se calculan sumas y medias móviles de ene en ene, y ene es un número impar, es posible establecer un parangón entre la suma o la media móvil resultante y el dato mediano, de entre la serie de datos que sirvieron para obtener la suma o media. Así, si se calculan medias móviles de 3 en 3, la primera media móvil será equiparable al dato mediano de la serie constituida por el primero, el segundo y el tercero; o sea, que se la puede equiparar con el segundo dato; a la segunda media móvil de 3 en 3, se la puede equiparar con el dato mediano de la serie constituida por el segundo, el tercero y el cuarto, o sea, con el tercero; a la tercera media móvil se la podrá equiparar con el cuarto, etcétera.

En forma parecida, si la media móvil es de 5 en 5, a la primera se la puede equiparar con el dato mediano de la serie constituida por el primero, el segundo, el tercero, el cuarto y el quinto, o sea, con el quinto; a la segunda media móvil con el sexto, y así sucesivamente.

Así, resulta claro que, en el primer caso, los datos primero y último no tienen media móvil con que equipararse; que, en el segundo caso, los datos primero y segundo, por una parte y último y penúltimo por otra no tienen medias móviles con las que ponerse en parangón, tal como se dijo antes.

Pero, cuando el número de datos de los que se calcula cada media móvil es par, no sólo los extremos de la serie observada carecen de media móvil con que parangonarse, puesto que los mismos datos centrales no tienen media móvil con que equipararse. Así, por ejemplo, si se calculan medias móviles de 2 en 2, la primera media móvil obtenida no corresponde ni al primer dato ni al segundo, sino a un punto intermedio entre ambos; la segunda no corresponde ni al segundo ni al tercero sino a otro punto intermedio entre los ocupados por éstos, y así, sucesivamente.

De ahí que cuando se quieren emplear medias móviles de pares (o, en general de números pares) de datos, se imponga la necesidad adicional de una operación más, conocida como "centrar".

EJEMPLO DE CALCULO DE LAS MEDIAS MÓVILES Y DE LA OPERACIÓN DE CENTRAR

Medias móviles de 12 en 12 meses y medias móviles de 12 en 12 centradas para el 7º $(12/2 + 1)$ mes, de los depósitos del sistema bancario mexicano

Años	Mes	*Depósitos en millones de pesos*	*Sumas móviles de 12 en 12 meses*	*Medias móviles de 12 en 12 meses*	*Medias móviles centradas para el 7º dato*	*Mes*
1960	E	10 077				
	F	10 332				
	M	10 216				
	A	10 381				
	M	10 552				
	J	10 358				
			126 179	$126\,179 \div 12 = 10\,515$		
	J	10 418				
			126 597	$126\,597 \div 12 = 10\,550$	$\dfrac{10\,515 + 10\,550}{2} = 10\,533$	J
	A	10 753				
			126 826	$126\,826 \div 12 = 10\,569$	$\dfrac{10\,550 + 10\,569}{2} = 10\,560$	A
	S	10 773				
			127 211	$127\,211 \div 12 = 10\,601$	10 585	S
	O	10 663				
			128 183	$128\,183 \div 12 = 10\,682$	10 641	O
	N	10 934				
	D	10 722				
1961	E	10 495				
	F	10 561				
	M	10 601				
	A	11 353				

Cómo se centran las medias móviles

La operación de centrar medias móviles de números pares de datos consiste, fundamentalmente, en tomar una nueva serie de medias móviles a partir de las medias móviles de primer orden. De este modo, las dos primeras medias móviles que quedaban colocadas entre el primero y el segundo, y entre el segundo y el tercer datos producen una nueva media móvil "centrada" en el segundo y comparable, por ello mismo, con ese dato.

Como es fácil comprender, la operación de centrar las medias móviles sólo tiene razón de ser cuando ene es par ya que cuando es impar, la operación de centrar se convertiría, en realidad, en una operación descentradora.

Perecuación de tendencias

Si bien el uso de las sumas y medias móviles permite tener una visión de la tendencia que sigue determinado fenómeno al través del tiempo, su empleo no permite —en cambio— la obtención de una expresión analítica (de una fórmula) que pueda considerarse como ley matemática del fenómeno.

Es por esto por lo que, para obtener una expresión matemática de la tendencia, se perecúan líneas o fórmulas típicas a las series empíricas. Entre los métodos de perecuación destaca el de los mínimos cuadrados.

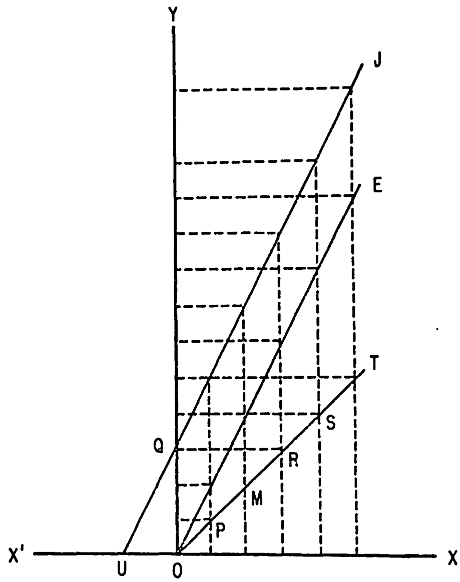
Método de los mínimos cuadrados para el cálculo de la tendencia

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.

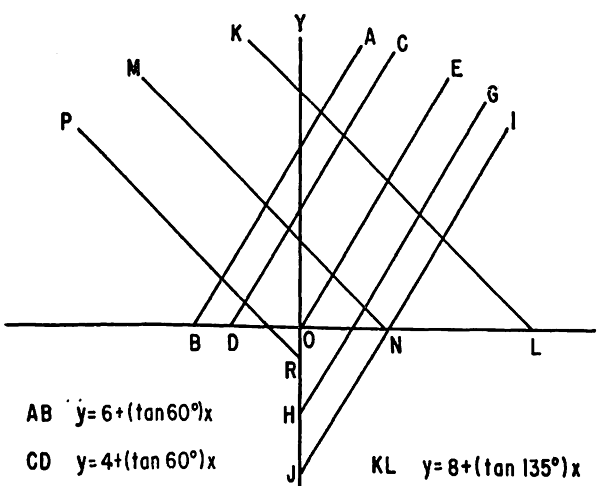
El procedimiento general de interpolación puede resumirse en las siguientes etapas:

1. Representación gráfica de la serie dinámica empírica con el objeto de poder determinar el tipo de curva por interpolar.

RECTAS
representación de ecuaciones de primer grado



RECTAS Y ECUACIONES DE PRIMER GRADO



AB $y = 6 + (\tan 60°)x$

CD $y = 4 + (\tan 60°)x$

EO $y = (\tan 60°)x$

GH $y = -3 + (\tan 60°)x$

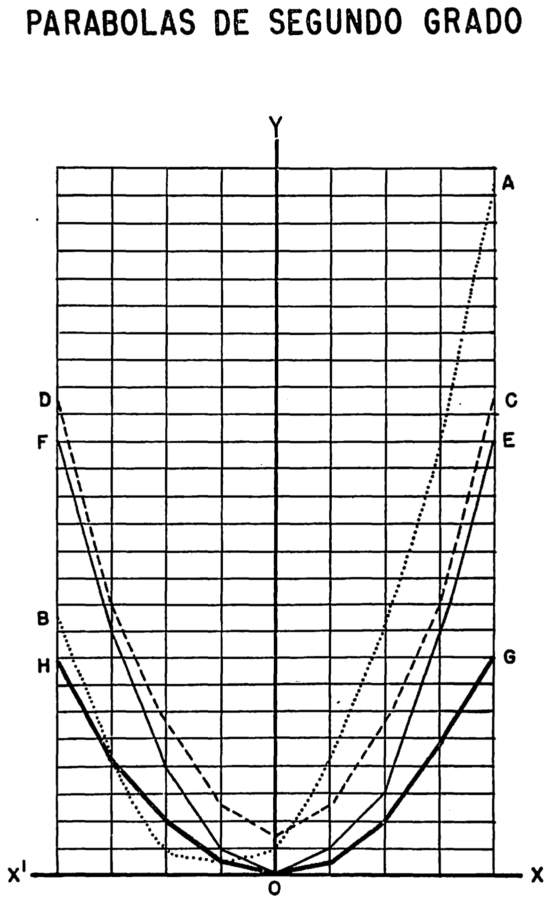
IJ $y = -5 + (\tan 60°)x$

KL $y = 8 + (\tan 135°)x$

$y = 8 - (\tan 45°)x$

MN $y = 3 - (\tan 45°)x$

PR $y = -1 - (\tan 45°)x$



2. Elección —por inspección ocular— de una de las curvas siguientes como representativa de la serie dinámica empírica.
   2.1. Recta.
   2.2. Parábola de 2º grado.
   2.3. Parábola de 3er. grado.
   2.4. Hipérbola.
   2.5. Logarítmica.
3. Consignación de la fórmula general de la curva elegida:
   3.1. Recta:

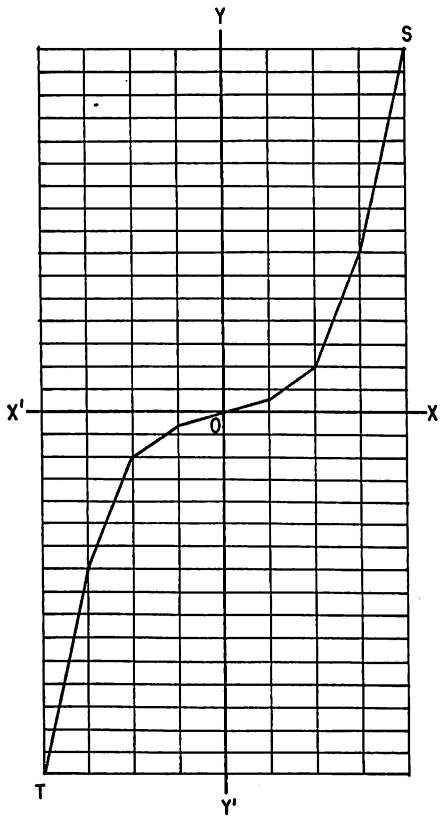
$$y = a + bx.$$

   3.2. Parábola de 2º grado;

$$y = a + cx + bx^2.$$

PARABOLA DE TERCER GRADO

SOT $y = x^3$



3.3. Parábola de 3er. grado:

$$y = a + cx + dx^2 + bx^3.$$

3.4. Hipérbola (equilátera):

$$y = a + bx^{-1}.$$

3.5. Logarítmica:

$$y = a + b \text{ Log. } x.$$

3.6. Exponencial:

$$y = ab^x \text{ o}$$

$$\text{Log. } y = \text{Log. } a + x \text{ Log. } b.$$

4. Obtención de tantas ecuaciones de interpolación como parámetros tenga la fórmula general de la curva elegida.

4.01. Recta:
2 parámetros (a, b),
2 ecuaciones.

4.02. Parábola de 2º grado:
3 parámetros (a, c, b),
3 ecuaciones.

4.03. Parábola de 3er. grado:
4 parámetros (a, c, d, b),
4 ecuaciones.

4.04. Hipérbola (equilátera):
2 parámetros (a, b),
2 ecuaciones.

4.05. Logarítmica:
2 parámetros (a, b),
2 ecuaciones.

4.06. Exponencial:
2 parámetros (a, b),
2 ecuaciones.

Forma de las ecuaciones de interpolación:

4.1. Para obtener la primera de interpolación, agréguese un operador Σ a cada uno de los términos de la ecuación general:

4.11. Recta:

$$\Sigma = Na + \Sigma(x)b.$$

4.12. Parábola de 2º grado:

$$\Sigma y = Na + \Sigma(x)c + \Sigma(x^2)b.$$

4.13. Parábola de 3er. grado:

$$\Sigma y = Na + \Sigma(x)c + \Sigma(x^2)d + \Sigma(x^3)b.$$

4.14. Hipérbola:

$$\Sigma y = Na + \Sigma(x^{-1})b.$$

4.15. Logarítmica:

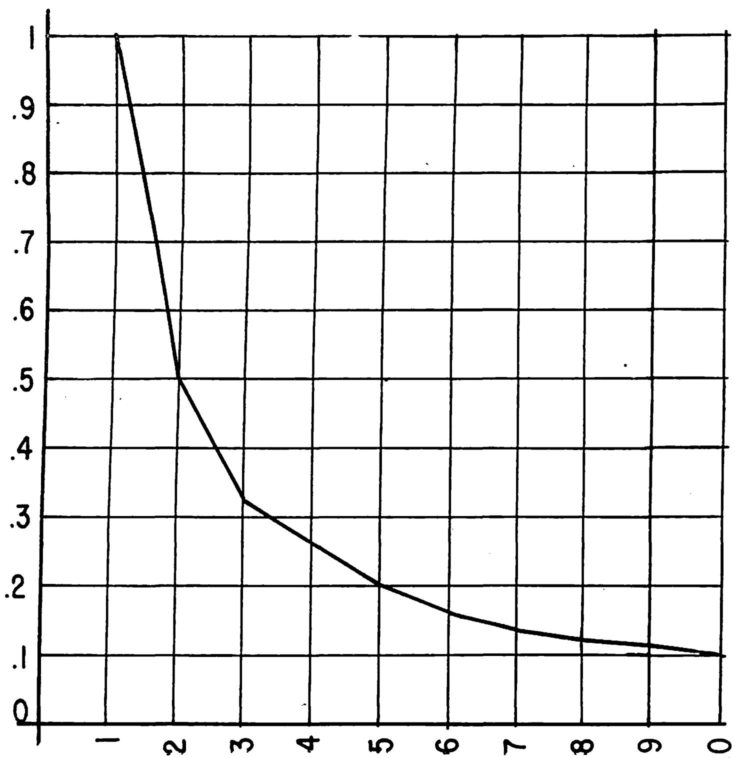
$$\Sigma y = Na + \Sigma(\log x)b.$$

4.1. Exponencial:

$$\Sigma \log y = N \log A + (\Sigma x)\log b.$$

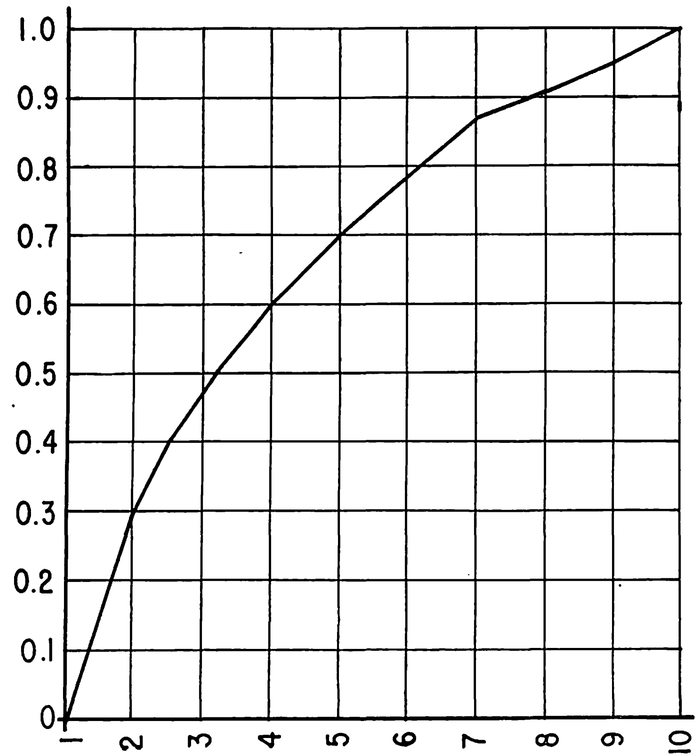
HIPERBOLA

$y = x^{-1}$

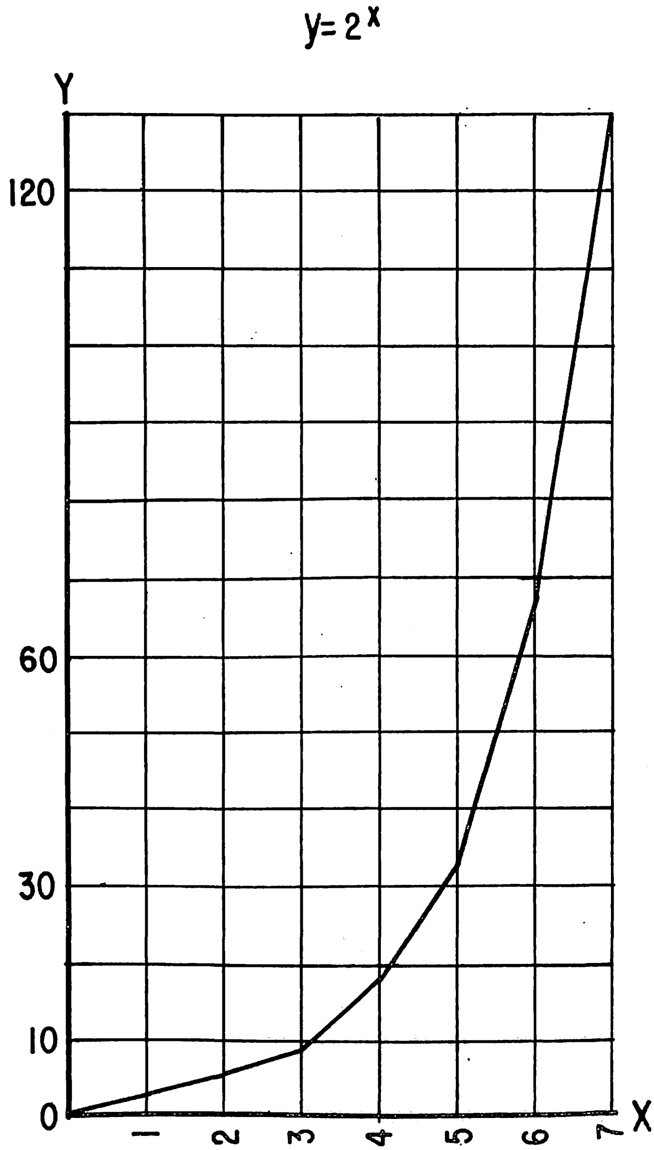


LOGARITMICA

$y = \log x$



EXPONENCIAL



4.2. 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:

4.21. Recta:

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)b.$$

4.22. Parábola de 2º grado:

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^4)b.$$

4.23. Parábola de 3er. grado:

$$\Sigma(yx^3) = \Sigma(x^3)a + \Sigma(x^6)b.$$

4.24. Hipérbola:

$$\Sigma(yx^{-1}) = \Sigma(x^{-1})a + (x^{-2})b.$$

4.25. Logarítmica:

$$\Sigma(y \log x) = \Sigma(\log^2 x)b.$$

4.26. Exponencial:

$$\Sigma(x \log y) = \Sigma x \log a + \Sigma x^2 \log b.$$

4.3. 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:

4.32. Parábola de 2º grado:

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)b.$$

4.33. Parábola de 3er. grado:

$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)d + \Sigma(x^4)b.$$

4.4. 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:

4.43. Parábola de 3er. grado:

$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^3)c + \Sigma(x^4)d + \Sigma(x^5)b.$$

5. Fórmese con las ecuaciones anteriores un sistema de tantas ecuaciones como parámetros:

5.1. Recta:

$$\Sigma(y) = Na \quad + \Sigma(x)b$$
$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)b.$$

5.2. Parábola de 2º grado:

$$\Sigma(y) = Na \quad + \Sigma(x)c + \Sigma(x^2)b$$
$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)b$$
$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^3)c + \Sigma(x^4)b.$$

5.3. Parábola de 3er. grado:

$$\Sigma(y) = Na \quad + \Sigma(x)c + \Sigma(x^2)d + \Sigma(x^3)b$$
$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)c + \Sigma(x^3)d + \Sigma(x^4)b$$
$$\Sigma(yx^2) = \Sigma(x^2)a + \Sigma(x^3)c + \Sigma(x^4)d + \Sigma(x^5)b$$
$$\Sigma(yx^3) = \Sigma(x^3)a + \Sigma(x^4)c + \Sigma(x^5)d + \Sigma(x^6)b.$$

5.4. Hipérbola:

$$\Sigma(y) = Na \quad + \Sigma(x^{-1})b$$
$$\Sigma(yx^{-1}) = \Sigma(x^{-1})a + \Sigma(x^{-2})b.$$

5.5. Logarítmica:

$$\Sigma(y) = Na \quad + (\log x)b$$
$$\Sigma(y \log x) = \Sigma(\log x)a + \Sigma(\log^2 x)b.$$

6. Resuélvase el sistema de ecuaciones así formado (por ejemplo, por medio de determinantes).

7. Substitúyanse los valores de los parámetros así obtenidos en la ecuación general de la curva correspondiente.

EJEMPLO DE INTERPOLACIÓN DE UNA RECTA

Población económicamente activa en Kalabá

Años	*x*	*Población 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
	Σx	Σy	$\Sigma(yx)$	$\Sigma(x^2)$

Ecuación general:

$$y = a + bx$$

Ecuaciones de interpolación:

$$\Sigma y = Na + \Sigma(x)b$$
$$\Sigma(yx) = \Sigma(x)a + \Sigma(x^2)b$$

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:

$$a = \frac{\text{Determinante para a}}{\text{Eliminante}} = \frac{4\,136}{825} = 5.01$$

Valor de b:

$$b = \frac{\text{Determinante para b}}{\text{Eliminante}} = \frac{163}{825} = 0.1975$$

Ley del fenómeno:

$$y = 5.01 + 0.1975x$$

EJEMPLO DE INTERPOLACIÓN SIMPLIFICADA DE UNA RECTA

Población económicamente activa en Kalabá

Años	*x'*	*Población 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	
	$\Sigma(x')$	$\Sigma(y)$	$\Sigma(x'y)$	$\Sigma(x'^2)$

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:

$$\Sigma y = Na' + \Sigma(x')b$$
$$\Sigma(yx') = \Sigma(x')a + \Sigma(x'^2)b$$

Substitución de valores, previa simplificación:

$$\Sigma y = Na'$$
$$\Sigma(yx') = \Sigma(x'^2)b$$

Substitución:

$$54 = 9a'$$
$$11.8 = 60b$$

Resolución de las dos ecuaciones independientes:

$$a' = \frac{54}{9} = 6 \qquad b = \frac{11.8}{60} = .196$$

Ley del fenómeno:

$$y = 6 + .196x'$$

EJEMPLO DE INTERPOLACIÓN DE UNA RECTA
SIMPLIFICACIÓN DEL AÑO MEDIANO (I)

Población económicamente activa en Kalabá

Años	*x"*	*Población 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:

$$\Sigma y = Na'' + \Sigma(x'')b$$
$$\Sigma(yx'') = \Sigma(x'')a + \Sigma(x''^2)b$$

Ecuaciones simplificadas:

$$\Sigma y = Na''$$
$$\Sigma(yx'') = \Sigma(x''^2)b$$

Substitución de valores:

$$61 = 10a''$$
$$16.30 = 82.50b$$

Resolución de las dos ecuaciones independientes:

$$a'' = \frac{61}{10} = 6.1 \qquad b = \frac{16.30}{82.50} = .197$$

Ley del fenómeno:

$$y = 6.1 + .197x''$$

EJEMPLO DE INTERPOLACIÓN DE UNA RECTA SIMPLIFICACIÓN DEL AÑO MEDIANO (II)

Población económicamente activa en Kalabá

Años	*x''*	*Población 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	
	$\Sigma(x'')$	$\Sigma(y)$	$\Sigma(yx'')$	$\Sigma(x''^2)$

Mediana de los años, con respecto a la cual se tomaron las equis como desviaciones en medios años o semestres: 1 950.5

Ecuación general:

$$y = a'' + b'x''$$

Ecuaciones de interpolación:

$$\Sigma(y) = Na'' + \Sigma(x'')b'$$
$$\Sigma(yx'') = \Sigma(x'')a + \Sigma(x''^2)b'$$

Ecuaciones simplificadas:

$$\Sigma(y) = Na''$$
$$\Sigma(yx'') = \Sigma(x''^2)b'$$

Substitución de valores:

$$61 = 10a''$$
$$32.6 = 330b'$$

Resolución de las dos ecuaciones independientes:

$$a'' = \frac{61}{10} = 6.1$$

$$b' = \frac{32.6}{330} = .0987$$

Ley del fenómeno:

$$y = 6.1 + .0987x''$$
x'' (como desviación semestral)

o $$y = 6.1 + .1974\,x''$$
x'' (como desviación anual)

EJEMPLO DE INTERPOLACIÓN DE UNA PARABOLA

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
	$\Sigma(x'')$	$\Sigma(y)$	$\Sigma(x''^2)$	$\Sigma(yx''^2)$	$\Sigma(x''^4)$	$\Sigma(yx'')$

Ecuación general:

$$y = a'' + c''x'' + b'x''^2$$

Ecuaciones de Interpolación:

$$\Sigma(y) = Na'' + \Sigma(x'')c'' + \Sigma(x''^2)b'$$

$$\Sigma(yx''^2) = \Sigma(x'')a'' + \Sigma(x''^2)c'' + \Sigma(x''^3)b'$$

$$\Sigma(yx''^2) = \Sigma(x''^2)a'' + \Sigma(x''^3)c'' + \Sigma(x''^4)b'$$

Simplificación:

$$\Sigma(y) = Na'' + \Sigma(x''^2)b'$$

$$\Sigma(yx'') = \Sigma(x'')c''$$

$$\Sigma(yx''^2) = \Sigma(x''^2)a'' + \Sigma(x''^4)b'$$

* 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.*

Substitución de valores en el sistema formado por las ecuaciones primera y última, y ulterior substitución en la ecuación independiente:

$$\begin{aligned} 940 &= 10\,a'' + 330\,b' \\ 35\,220 &= 330\,a'' + 19\,338\,b' \end{aligned}$$

Resolución del sistema mediante determinantes de 2º orden:

Eliminante:

$$\begin{vmatrix} 10 & 330 \\ 330 & 19\,338 \end{vmatrix} \; 193\,380 - 108\,900 = 84\,480$$

Determinante para a:

$$\begin{vmatrix} 940 & 330 \\ 35\,220 & 19\,338 \end{vmatrix} = 18\,177\,720 - 11\,622\,600 = \\ = 6\,555\,120$$

Valor de a:

$$a = \frac{\text{Determinante para a}}{\text{Eliminante}} = \frac{6\,555\,120}{84\,480} = 77.59$$

Determinante para b:

$$\begin{vmatrix} 10 & 940 \\ 330 & 35\,220 \end{vmatrix} = 352\,200 - 310\,200 = 42\,000$$

Valor de b:

$$b = \frac{\text{Determinante para b}}{\text{Eliminante}} = \frac{42\,000}{84\,480} = .497$$

Para encontrar el valor de c, utilícese la ecuación independiente:

$$\Sigma yx'' = \Sigma(x''^2)c$$

Substitución de valores:

$$150 = 330c$$

$$c = \frac{150}{330} = .45$$

Ley del fenómeno:

$$y = 77.59 + .45''x + .497x''^2$$

EJEMPLO DE INTERPOLACIÓN DE UNA PARABOLA DE SEGUNDO GRADO (SIMPLIFICACIÓN DEL AÑO MEDIANO)

Crecimiento de la población de Kalabá

Años	*x'*	*Población 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:

$$\begin{aligned} \Sigma(y) &= Na' + \Sigma(x')c' + \Sigma(x'^2)b' \\ \Sigma(yx') &= \Sigma(x')a' + \Sigma(x'^2)c' + \Sigma(x'^3)b' \\ \Sigma(yx'^2) &= \Sigma(x'^2)a' + \Sigma(x'^3)c' + \Sigma(x'^4)b' \end{aligned}$$

Simplificación:

$$\begin{aligned} \Sigma(y) &= Na' + \Sigma(x'^2)b' \\ \Sigma(yx') &= \Sigma(x'^2)c' \\ \Sigma(yx'^2) &= \Sigma(x'^2)a' + \Sigma\,(x'^4)b' \end{aligned}$$

Substitución de valores en las ecuaciones primera y tercera que forman sistema, y después en la independiente:

$$\begin{aligned} 94 &= 9a' + 60b' \\ 684 &= 60a' + 708b' \end{aligned}$$

Resolución del sistema:

Eliminante:

$$\begin{vmatrix} 9 & 60 \\ 60 & 708 \end{vmatrix} = 6\,372 - 3\,600 = 2\,772$$

Determinante de a:

$$\begin{vmatrix} 94 & 60 \\ 684 & 708 \end{vmatrix} = 66\,552 - 41\,040 = 25\,512$$

Valor de a:

$$a = \frac{\text{Determinante para a}}{\text{Eliminante}} = \frac{25\,512}{2\,772} = 9.2$$

Determinante para b:

$$\begin{vmatrix} 60 & 684 \\ 9 & 94 \end{vmatrix} = 6\,156 - 5\,940 = 516$$

Valor de b:

$$b' = \frac{\text{Determinante para b}'}{\text{Eliminante}} = \frac{516}{2\,772} = .0861$$

Para encontrar el valor de c, utilícese la ecuación independiente:

$$\Sigma(yx') = \Sigma(x'^2)c$$

$$163 = 60c$$

$$c = \frac{163}{60} = 2.71$$

Ley del fenómeno:

$$y = 9.2 + 2.71\,x + .0861\,x^2$$

EJEMPLO DE INTERPOLACIÓN DE (LA RAMA POSITIVA DE) UNA HIPÉRBOLA EQUILÁTERA

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}	*Elaboración* 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	2.927	2.0680	1.546
	$\Sigma(y)$	$\Sigma(x^{-1})$	$\Sigma(yx^{-1})$	$\Sigma(x^{-1})^2$

Ecuación general:

$$y = a + bx^{-1}$$

Ecuaciones de interpolación:

$$\Sigma(y) = Na + \Sigma(x^{-1})b$$
$$\Sigma(yx^{-1}) = \Sigma(x^{-1})a + \Sigma(x^{-1})^2b$$

Substitución de valores:

$$2.0680 = 2.927a + 1.546b$$
$$4.91 = 10a + 2.927b$$

Resolución del sistema de ecuaciones por medio de determinantes:
Eliminante:

$$\begin{vmatrix} 10 & 2.927 \\ 2.927 & 1.546 \end{vmatrix} = 15.46 - 8.567329 = 6.892671$$

Determinante para a:

$$\begin{vmatrix} 4.91 & 2.927 \\ 2.0680 & 1.546 \end{vmatrix} = 7.59086 - 6.0319616 = 1.5688984$$

Valor de a:

$$a = \frac{\text{Determinante para a}}{\text{Eliminante}} = \frac{1.56889884}{6.892671} = .22$$

Determinante para b:

$$\begin{vmatrix} 10 & 4.91 \\ 2.927 & 2.0608 \end{vmatrix} = 20.608 - 14.37157 = 6.23643$$

Valor de b:

$$b = \frac{\text{Determinante para b}}{\text{Eliminante}} = \frac{6.23643}{6.892671} = 0.94$$

Ley del fenómeno:

$$y = .22 + 0.94\,x^{-1}$$

EJEMPLO DE INTERPOLACIÓN DE UNA EXPONENCIAL

Crecimiento de la población de la República Mexicana
(1895 - 1950)

Años	*x*	*Población 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
	Σx		$\Sigma(\log y)$	$\Sigma(x \log y)$	$\Sigma(x^2)$

Ecuación general:

$$y = ab^x$$

Transformación por logaritmos:

$$\log y = \log a + x \log b$$

Ecuaciones de interpolación:

$$\Sigma \log y = N \log a + \Sigma(x) \log b$$
$$\Sigma(x \log y) = \Sigma(x) \log a + \Sigma(x^2) \log b$$

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).

Hay algunas series (particularmente algunas de carácter demográfico) a las que resulta difícil interpolarles una curva adecuada por el procedimiento de los mínimos cuadrados que hemos delineado en el estudio de la tendencia secular. En estos casos se emplea un procedimiento llamado de "agrupación" que vamos a aplicar, primero, al caso muy sencillo de la recta y, en seguida a los de la logarítmica, la antilogarítmica o exponencial y sus derivaciones las curvas de Gomperz y de Pearl-Reed o logística.

Cálculo de la tendencia rectilínea por agrupamiento:

1. Divídanse las equis y las yes en dos grupos.
2. Obténganse:
   2.1. la suma del primer grupo de equis (S_{1x}), y
   2.2. la suma del segundo grupo de equis (S_{2x}).
3. Réstese de la segunda suma la primera (D_x).
4. Obténganse:
   4.1. la suma del primer grupo de yes (S_{1y}), y
   4.2. la suma del segundo grupo de yes (S_{2y}).
5. Réstese de la segunda suma la primera (D_y).
6. Divídase la diferencia de las yes (D_y) entre la diferencia de las equis (D_x). El resultado es el parámetro b de la tendencia rectilínea, o sea, el ritmo de cambio del fenómeno.
7. Para obtener el parámetro a, o intersección de la recta con el eje vertical:
   7.1. multiplíquese b por la primer suma de las equis,
   7.2. réstese el producto de la primer suma de las yes,
   7.3. duplíquese la diferencia, y
   7.4. divídase el duplo entre el total de datos.

Las fórmulas correspondientes son:

$$b = \frac{D_y}{D_x}$$

$$a = \frac{2(S_{1y} - b\, S_{1x})}{N}$$

OBTENCIÓN DE LAS FÓRMULAS PARA EL CALCULO DE LA TENDENCIA RECTILÍNEA POR AGRUPAMIENTO

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	$y_6 = a + 6b$

$$S_{1y} = y_1 + y_2 + y_3 = 3a + 6b; \qquad S_{1x} = 1 + 2 + 3 = 6$$

$$S_{2y} = y_4 + y_5 + y_6 = 3a + 15b; \qquad S_{2x} = 4 + 5 + 6 = 15$$

$$D_y = S_{2y} - \\ = 3a + 15b - (3a + 6b) = \\ = 9b \qquad D_x = S_{2x} - S_{1x} = \\ = 15 - 6 = \\ = 9$$

$$\frac{D_y}{D_x} = \frac{9b}{9} = b$$

$$S_{1y} = 3a + 6b$$

$$= \frac{N}{2} a + S_{1x} b$$

$$a = \frac{2}{N}(S_{1y} - S_{1x} b)$$

$$y = a + b x$$

EJEMPLO PEDAGÓGICO DE CÁLCULO DE UNA TENDENCIA RECTILÍNEA POR AGRUPAMIENTO

	x	y
	1	8
	2	13
	3	18
	4	23
	5	28
N =	6	33

$$S_{1y} = 8 + 13 + 18 = 39 \qquad S_{1x} = 1 + 2 + 3 = 6$$

$$S_{2y} = 23 + 28 + 33 = 84 \qquad S_{2x} = 4 + 5 + 6 = 15$$

$$D_y = 84 - 39 = 45 \qquad D_x = 15 - 6 = 9$$

$$\frac{D_y}{D_x} = \frac{45}{9} = 5 = b$$

$$a = \frac{2}{N}(S_{1y} - bS_{1x}) = \frac{2}{N}(39 - 6 \times 5)$$

$$= \frac{2}{6}(39 - 30) = 3$$

$$y = a + b x$$
$$y = 3 + 5 x$$

Cálculo de la tendencia logarítmica por agrupamiento

Paso previo. Obténganse los logaritmos de las equis o unidades de tiempo.

Pasos centrales. Trabájese con esos logaritmos de las equis como si fueran las equis, y calcúlese la tendencia rectilínea por agrupamiento.

Paso posterior. Substitúyanse los valores de los parámetros a y b, así obtenidos, en la ecuación general de la logarítmica:

$$y = a + b \log x.$$

Cálculo de la tendencia exponencial por agrupamiento

Paso previo. Obténganse los logaritmos de las yes, o sean las magnitudes variables del fenómeno en el tiempo.

Pasos centrales. Trabájese con los logaritmos de las yes como si fueran yes, y calcúlese la tendencia rectilínea por agrupamiento.

Paso posterior. Búsquense los antilogaritmos de los valores encontrados para los parámetros y substitúyanse dichos valores en la ecuación general de la exponencial:

$$y = ab^x$$

Cálculo de la exponencial modificada, mediante agrupamiento

Se conoce como exponencial modificada aquella en la que a la expresión ordinaria de la tendencia exponencial que aquí representaremos por $y = bc^x$ se agrega un término que se conoce como "elemento de corrección", y que representaremos por a:

$$y = bc^x + a$$

El cálculo de esta tendencia se hace en dos etapas:

1. Cálculo de c.
2. Cálculo de a y de b.

1. Para calcular el parámetro c, tómese un número de datos que sea múltiplo de tres, y dense a las equis valores que vayan de 0 en adelante. En seguida:
   1.1. Agrúpense los datos en tres grupos iguales.
   1.2. Súmense los valores de cada grupo para obtener las tres sumas: s_1, s_2 y s_3.

1.3. Réstense por pares las sumas más inmediatas: s_2 de s_1 para obtener una primera diferencia d_1, s_3 de s_2 para obtener una segunda diferencia d_2.

1.4. Divídase la segunda diferencia entre la primera para obtener un cociente q.

1.5. Extráigase del cociente q una raíz igual al número de datos comprendidos en cada uno de los tres grupos formados (i. e. extráigase la raíz cuadrada si hay seis datos en total; cúbica si hay nueve, cuarta si son doce). El resultado es c.

2. Para calcular a y b:
Paso previo. Calcúlense las diferentes potencias de c, elevándolo a 0, 1, 2 ... N.
Paso central. Tómense los valores de esas potencias como si fueran equis, y calcúlese una tendencia rectilínea por agrupamiento.
Paso posterior. Los valores de los parámetros obtenidos mediante esa interpolación, deberán substituirse junto con el valor de c en la ecuación de la exponencial o geométrica modificada:

$$y = a + bc^x$$

Cálculo de la curva de Gomperz por agrupamiento

La ecuación de la curva de Gomperz es:

$$y = ab^{c^x}$$

Esta ecuación se convierte en forma logarítmica, en:

$$\log y = \log a + c^x \log b$$

En vista de que logaritmo y ocupa el lugar de las yes en la exponencial modificada; que logaritmo de a ocupa el de a y logaritmo de b el de b, el procedimiento de cálculo consiste en:

Paso previo. Obtener los logaritmos de las yes.
Paso central. Tomar los logaritmos de las yes como si fueran yes, e interpolar una geométrica modificada.
Paso posterior. Substituir los valores de los parámetros encontrados, en la forma logarítmica de la ecuación de Gomperz.
Paso alternativo. Buscar los antilogaritmos de los parámetros así encontrados y substituirlos en lugar de a y de b en la forma no logarítmica de la ecuación de Gomperz:

$$y = ab^{c^x}$$

La curva logística o de Pearl-Reed tiene por ecuación:

$$y^{-1} = a + bc^{x}$$

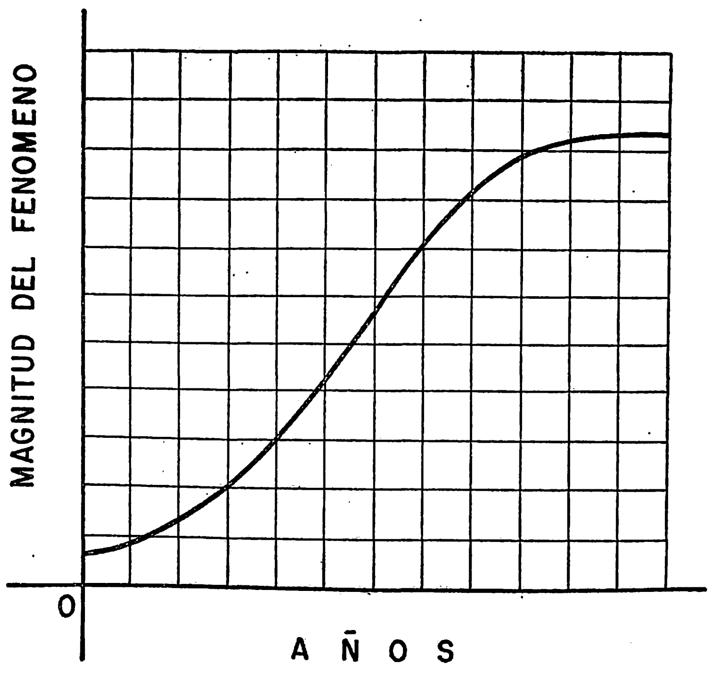
En esta expresión, los recíprocos de las yes ocupan el lugar que corresponde a las yes en la geométrica modificada. Según esto:

Paso previo. Obténganse los recíprocos de las yes.

Paso central. Tómense los recíprocos de las yes como si fuesen yes, e interpólese una exponencial modificada.

Paso ulterior. Substitúyanse los valores de los parámetros a, b, y c en la ecuación general de la logística (dada más arriba).

FORMA (ESIFORME) DE LA LOGISTICA



EJEMPLO PEDAGÓGICO DE ADAPTACIÓN DE UNA EXPONENCIAL POR AGRUPAMIENTO

x	y	log y Y
1	10	1.0000
2	20	1.3010
3	40	1.6021
4	80	1.9031
5	160	2.2041
6	320	2.5051

$S_{1Y} = 3.9031;$ $\qquad S_{1x} = 6$

$S_{2Y} = 6.6123;$ $\qquad S_{1y} = 15$

$D_y = 2.7092 \qquad D_x = 9$

$$\frac{D_y}{D_x} = \frac{2.7092}{9} = 0.3010 = C$$

$$B = \frac{2}{N}(S_{1Y} - CS_{1X}) =$$

$$= \frac{2}{N}(3.9031 - 6 \times 0.3010) =$$

$$= \frac{2}{6}(3.9031 - 1.8060) =$$

$$= \frac{1}{3}(2.0971) = 0.6990$$

$$Y = B + Cx = 0.6990 + 0.3010\,x$$

$$\log y = \log b + \log c \times x$$

$$b = \text{antilog } 0.6990 = 5$$

$$c = \text{antilog } 0.3010 = 2$$

$$\log y = \log 5 + (\log 2)\,x$$

$$y = 5 \times 2^x$$

EJEMPLO PEDAGÓGICO DE CÁLCULO DE UNA LOGARÍTMICA POR EL MÉTODO DE AGRUPACIÓN

x	y	X = Log. x
1	20 000	0.0000
2	29 030	0.3010
3	34 313	0.4771
4	38 063	0.6021
5	40 970	0.6990
6	43 346	0.7782

$S_{1y} = 83\,343; \qquad S_{1x} = 0.7781$

$S_{2y} = 122\,379; \qquad S_{2x} = 2.0793$

$D_y = 39\,036 \qquad D_x = 1.3012$

$$\frac{D_y}{D_x} = \frac{39\,036}{1.3012} = 30\,000 = b$$

$$a = \frac{2}{N}(S_{1y} - b\,S_{1x}) =$$

$$= \frac{2}{6}(83\,343 - 30\,000 \times 0.7781) =$$

$$= \frac{1}{3}(83\,343 - 23\,343) =$$

$$= \frac{60\,000}{3} = 20\,000$$

$$y = a + b \log X =$$

$$y = a + b \log X =$$

$$y = 20\,000 + 30\,000 \text{ Log. } x$$

EJEMPLO DE ADAPTACIÓN DE UNA EXPONENCIAL MODIFICADA POR EL MÉTODO DE AGRUPACIÓN

x	y	Cálculo del parámetro c
1	13	$s_1 = 36$
2	23	$d_1 = 126 - 36 = 90$
3	43	$s_2 = 126$
4	83	$q = 360/90 = 4; \sqrt[2]{4} = 2 = c$
5	163	$s_3 = 486$; $d_2 = 486 - 126 = 360$
6	323	

Cálculo de los parámetros a y b

x	$2^x = X$	y
1	2	13
2	4	23
3	8	43
4	16	83
5	32	163
6	64	323

$S_{1y} = 79$; $S_{1x} = 2 + 4 + 8 = 14$

$S_{2y} = 569$; $S_{2x} = 16 + 32 + 64 = 112$

$D_y = 490$; $D_x = 98$

$$\frac{D_y}{D_x} = \frac{490}{98} = 5 = b$$

$$a = \frac{2}{N}(S_{1y} - bS_{1x}) = \frac{2}{6}(79 - 5 \times 14) =$$

$$= \frac{79 - 70}{3} = \frac{9}{3} = 3 = a$$

$y = a + bc^x$

$y = 3 + 5 \times 2^x$

EJEMPLO PEDAGÓGICO DE CÁLCULO DE UNA LOGÍSTICA POR EL MÉTODO DE AGRUPACIÓN

x	y	$y^{-1} = Y$
1	0.625	1.6
2	0.455	2.2
3	0.294	3.4
4	0.172	5.8
5	0.094	10.6
6	0.050	20.2

$s_1 = 3.8$

$s_2 = 9.2$

$s_3 = 30.8$

$d_1 = 5.4$

$d_2 = 21.6$

$$q = \frac{21.6}{5.4} = 4$$

$$c = \sqrt[2]{q} = \sqrt[2]{4} = 2$$

(raíz cuadrada porque en cada uno de los tres grupos de datos hay 2 datos)

x	$2^x = X$	$y^{-1} = Y$
1	2	1.6
2	4	2.2
3	8	3.4
4	16	5.8
5	32	10.6
6	64	20.2

$S_{1Y} = 7.2$; $S_{1X} = 14$

$S_{2Y} = 36.6$; $S_{2X} = 112$

$D_Y = 29.4$ $\quad D_X = 98$

$$\frac{D_Y}{D_X} = \frac{29.4}{98} = 0.3 = b$$

$$a = \frac{2}{N}(S_{1Y} - b\,S_{1X}) =$$

$$= \frac{2}{6}(7.2 - 0.3 \times 14) =$$

$$= \frac{1}{3}(7.2 - 4.2) = \frac{1}{3} \times 3 = 1$$

$$y^{-1} = a + bc^x = 1 + 0.3\ (2)^x$$

Variaciones periódicas

En tanto que el movimiento secular presenta una tendencia más o menos constante a aumentar o disminuir, y mientras que las fluctuaciones irregulares representan altibajos que no muestran periodicidad alguna, hay otros movimientos de ascenso y descenso que acusan una marcada periodicidad. De ellos, unos repiten sus altas y sus bajas dentro del año ("variaciones estacionales") mientras que otros cubren ciclos mayores o menores que el ciclo anual ("variaciones cíclicas").

Variaciones estacionales

Muchos fenómenos, a más de la tendencia general a aumentar o disminuir en el curso de los años muestran —combinadamente— una tendencia al alza en determinadas *estaciones* [1] y a la baja en otras. Esas fluctuaciones del fenómeno en el curso de las diversas estaciones del año se conocen como: "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 a modo de que el promedio de ese promedio de relativos sea igual a 1 (o a 100, en caso de haber trabajado con porcientos).

Los índices así calculados expresan en qué tanto (por ciento) se alejan proporcionalmente (en más o menos) los valores reales de los valores que cabría esperar en caso de que el fenómeno no estuviera sujeto a variaciones estacionales.

Según esto, los cuatro pasos esenciales para el cálculo de los índices de variación estacional son:

1. Determinar el valor central que servirá de base para calcular los relativos (media, mediana, modo) o los valores normales (magnitudes calculadas mediante tendencia secular, medias móviles, etcétera) que puedan servir de base a ese cálculo.

2. Calcular los relativos de cada valor observado respecto del central o del normal correspondiente, dividiendo cada dato entre la base elegida.

3. Promediar todos los relativos correspondientes a cada una de las estaciones en los diversos años (mediante cálculo de la media o la determinación de la mediana).

[1] *Estación* designa aquí: o las estaciones propiamente dichas o los meses del año, o los trimestres, o los períodos (cualesquiera que sea el tipo de ellos) que convenga hacer en el curso del año, o de acuerdo con los cuales se registren los datos correspondientes.

4. Ajustar los promedios, a fin de que su suma dividida por el número de estaciones del año (12 si son meses, 4 si son trimestres) resulte igual a 1.

Dentro de este proceso general que se sigue siempre en el cálculo de los índices de variación estacional, hay las cuatro variantes siguientes:

1. Procedimiento de las medias mensuales.
2. Procedimiento de los relativos de la tendencia.
3. Procedimiento de relativos de las medias móviles.
4. Procedimiento de los eslabones relativos.

Variaciones estacionales por el procedimiento de las medias mensuales

Éste es uno de los procedimientos más sencillos y rápidos, y consiste, fundamentalmente, en:

1. Elegir como base de comparación un valor central (en particular, la media). Para ello:
   1.1. Se calcula la media de cada estación, sumando los valores observados y dividiéndolos entre el número de años en que se han hecho las observaciones.
   1.2. Se suman las medias de todas las estaciones y se divide el resultado entre el número de estaciones que hay en el año, para obtener la media general.
2. Calcular los relativos de los valores observados, con respecto a ese valor central (o media) dividiendo cada uno de los valores observados entre la media general.
3. Promediar todos los relativos correspondientes a una misma estación, sumándolos y dividiéndolos entre el número de años (o determinar el valor mediano de dichos relativos).
4. Ajustar los promedios de cada uno de los meses. Para ello:
   4.1. Se suman dichos promedios estacionales y se divide entre el número de estaciones del año.
   4.2. Se ve si este promedio es:
      a) igual a 1, en cuyo caso, no necesita ajuste,
      b) mayor que 1, caso en que el exceso deberá de dividirse entre el número de estaciones y el cociente, restarse de cada uno de los promedios estacionales; éstos ya eran índices estacionales, pero aún no estaban ajustados.
      c) menor que 1, caso en que el defecto habrá de dividirse entre el número de estaciones y el cociente resultante, sumarse a cada uno de los promedios mensuales a fin de obtener índices estacionales ajustados.

Variaciones estacionales por el procedimiento de relativos de la tendencia

Éste es uno de los procedimientos que encajan mejor dentro del marco de estudio de las series temporales. Fundamentalmente, consiste en:

1. Elegir como término de comparación una serie de valores normales (los valores calculados mediante la tendencia). Para hacerlo:
   1.1. Se obtienen los promedios de cada año, ya sea calculando su media aritmética o determinando su mediana.
   1.2. Con los promedios de cada año, se forma una serie cronológica, a la que se le interpola la curva correspondiente.
   1.3. Se determinan los valores mensuales de la tendencia, substituyendo la variable dependiente (en la fórmula) para la tendencia por $1/n$, $2/n \ldots n/n = 1 \ldots 2 \; 3 \ldots 4 \ldots N$.
2. Calcular los relativos de los valores observados con respecto a los valores normales que les correspondan, dividiendo el valor observado en una estación específica entre el valor teórico de la tendencia, calculado para esa precisa estación del año.
3. Promediar todos los relativos correspondientes a una misma estación en los diferentes años, ya sea calculando su media o determinando su mediana.
4. Ajustar los promedios de relativos de cada estación.
   Para hacerlo:
   4.1. Se calcula el promedio (media aritmética o mediana) de dichos relativos mensuales promediados, o sea, el promedio mensual general.
   4.2. Se divide cada promedio estacional de relativos (que es ya un índice estacional no ajustado) entre el promedio general. Con esto se obtiene un índice ajustado de variación estacional.

Variaciones estacionales por el procedimiento de las medias móviles

En este procedimiento:

1. Se eligen ciertos valores normales (medias móviles de ene en ene datos si ene es el número de estaciones, y, en caso necesario, se las centra para la estación de orden $[(n/2) + 1]$.
2. Se calculan los relativos de los valores observados con respecto a las medias móviles centradas que les correspondan, dividiendo los valores observados entre esas medias móviles centradas.
3. Se promedian los relativos de los valores observados respecto a sus medias móviles que correspondan a una misma estación (se suman

los relativos de todos los meses de julio, por ejemplo, y la suma se divide entre el número de relativos que se hayan sumado o sea el número de años menos 1 en virtud de la pérdida de medias móviles equiparables).

4. 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, para obtener los correspondientes índices ajustados de variación estacional.

Hay quienes consideran (según es el caso de Rigleman y Frisbee) que los índices de variación estacional calculados por este procedimiento de las medias móviles son más representativos que los que se obtienen por otros procedimeintos en vista de que las medias móviles siguen bastante de cerca la curva cíclica.

EJEMPLO DE CALCULO DE VARIACIONES ESTACIONALES
PROCEDIMIENTO DE LAS MEDIAS MENSUALES

Mes	*1950*	*1960*	*1970*	*Media mensual*	*Corrección*	*Media corregida*	*Indice*
Enero	8 190	8 370	8 550	8 370	——	8 370	8 370/9 114
Febrero	8 280	8 460	8 640	8 460	—15	8 445	0.93
Marzo	8 640	8 820	9 000	8 820	—30	9 790	0.96
Abril	8 490	8 670	8 850	8 670	—45	8 625	0.95
Mayo	9 010	9 190	9 370	9 190	—60	9 130	1.00
Junio	9 130	9 310	9 490	9 310	—75	9 235	1.01
Julio	9 110	9 290	9 470	9 290	—90	9 200	1.08
Agosto	9 750	9 930	10 110	9 930	—105	9 825	1.11
Septiembre	10 050	10 230	10 410	10 230	—120	10 110	1.15
Octubre	10 470	10 650	10 830	10 650	—135	10 515	1.01
Noviembre	9 180	9 360	9 540	9 360	—150	9 210	9 210/9 114
Diciembre	7 900	8 080	8 260	8 080	—165	7 915	
Suma:	108 200	110 360	112 520			109 370	
Media anual:	9 017	9 197	9 377			9 114	
Tendencia:	180		180				
Mensual:	180/12 = 15						

EJEMPLO DE CÁLCULO DE LOS ÍNDICES DE VARIACIÓN ESTACIONAL COMO RELATIVOS DE LA TENDENCIA

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

$y \;= a + bx$

$\Sigma y \;= Na \qquad 495 = 3a \qquad a = 165$

$\Sigma yx = Sx^2b \qquad 805 = 2b \qquad b = \;40$

Ritmo anual de crecimiento:

$$b = 40$$

Ritmo trimestral:

$$b/4 = 40/4 = 10$$

Ritmo semitrimestral:

$$b/4/2 = 10/2 = 5$$

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	

Variaciones estacionales por el procedimiento de los eslabones relativos

En el proceso de los eslabones relativos o de los relativos eslabonados, el proceso general de cálculo de las variaciones estacionales, se convierte en un proceso complejo; en él, 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: 1) la eliminación de la discrepancia entre los valores obtenidos para una misma estación en un año y en el que le sigue, dentro de la cadena de relativos, y 2) la expresión de cada elemento de dicha cadena como relativo de la media de la cadena (ajustada ya) para eliminar la discrepancia previa.

De ahí que, en lo que sigue, usemos dos ordenaciones de las etapas del procedimiento de los eslabones relativos: una muestra la sucesión de las diferentes fases del procedimiento con independencia de cualquier otra consideración (y en ella usamos la notación ordinaria, con cardinales arábicos); la otra señala la relación de esas diferentes fases específicas con las etapas del proceso general de cálculo de las variaciones estacionales (por lo que usaremos los romanos I, II, III, IV en correspondencia con las cuatro etapas del proceso general).

Según esto, el procedimiento de los eslabones relativos puede delinearse como sigue:

1 (I). Elíjanse como base de comparación los valores correspondientes al mes que preceda a aquel del que se va a calcular el eslabón relativo.

N. B. Diferencia de éste con los otros procedimientos, pues ni se elige un valor central ni una serie de valores normales calculados, sino una serie de valores empíricamente observados.

2 (II). Calcúlense los relativos de cada valor observado entre el anterior: el correspondiente a febrero entre el de enero; el correspondiente a marzo entre el de febrero... el correspondiente a enero del segundo año entre el que corresponda a diciembre del primer año, para obtener los "eslabones". Los "eslabones" representan cuánto es el valor de cada estación (mes en el ejemplo) con respecto al valor de la estación (o mes) precedente.

3 (II). Promédiense todos los eslabones correspondientes a una misma estación, para obtener "medianas estacionales de eslabones" que indiquen cuánto en *promedio*, representa el valor de cada estación con respecto al del anterior, en la serie de años estudiada.

4 (I). Elíjase la mediana de eslabones de la primera estación del año como base fija de comparación de las demás medianas estacionales.

5 (II). Calcúlense los relativos de cada promedio mensual con respecto a la base fija (mediana de la primera estación del año) teniendo en cuenta que:

5.1. El relativo de la primera estación con respecto a sí mismo es igual a uno ($x_e/x_e = 1$).

5.2. El relativo de la segunda estación con respecto a la primera (x_f/x_e) es la mediana misma de los eslabones de la segunda estación con respecto a la primera (x_f/x_e).

5.3. El relativo de la tercera estación con respecto a la primera, es igual a la mediana de los eslabones de la tercera con respecto a la segunda (x_m/x_f) multiplicada por el relativo de la segunda con respecto a la primera (x_f/x_e). En efecto, el producto de estos relativos produce (x_m/x_f por x_f/x_e) $= x_m/x_e$.

5.n. En general, el relativo de un mes con respecto al mes primero, o de una estación respecto de la primera estación del año será igual a la mediana de los eslabones de la estación respecto de la estación precedente, multiplicada por el relativo de la estación precedente con respecto a enero.

5. (n + 1). El relativo de la primera estación con respecto a la primera estación (enero con respecto a enero) será igual a la mediana de los eslabones relativos de la primera estación respecto de la última del año anterior (enero respecto de diciembre anterior: x_e/x_d) multiplicada por el relativo de la estación precedente (la última del año: diciembre si se trata de meses) con respecto a la primera (x_d/x_e).

N. B. Ese producto debería dar 1 (xe/x_d por $x_d/x_e = 1$) pero esto ocurre excepcionalmente en la realidad. Esto era de esperar pues, por ejemplo, en el caso de los meses, el "diciembre" con respecto al cual se tomaron los eslabones y respecto del que está dada la mediana de los eslabones de enero es el diciembre del año anterior, mientras que el relativo de diciembre con respecto a enero contiene como numerador el diciembre del año mismo: o sea, que si se llama x_{d-1} al del año previo, y x_d al del mismo año, se tendrá: x_e/x_{d-1} por $x_d/x_e \neq 1$ e igual, en cambio, a x_{d-1}/x_d). Esta diferencia con respecto a 1 es la que impone el que se haga el ajuste.

6 (IV). Ajuste en dos etapas:

6.1. Reducción del segundo relativo de la primera estación al valor 1 del primer relativo de esa estación, para lo cual se considera la discrepancia entre ambos como debida a la tendencia (que se dividirá entre el número de estaciones de un año para restar o agregar 1/n al valor de la segunda estación, 2/n al de la tercera, y n/n al segundo relativo de la primera estación).

7 (II). Reducción de los valores anteriores (que dan cada valor mensual con respecto a la primera estación) a relativos del promedio de los mismos.

7.1. Se encontrará la mediana o se calculará la media aritmética de los valores obtenidos, y

7.2. Se dividirá cada uno de esos valores entre ese promedio, obteniéndose así los índices ajustados de variación estacional, por el método de los eslabones relativos.

EJEMPLO PEDAGÓGICO DE CÁLCULO DE LAS VARIACIONES ESTACIONALES POR EL PROCEDIMIENTO DE LOS ESLABONES RELATIVOS

ESTACIONES *Trimestrales*	DATOS TRIMESTRALES 1950	1951	1952	ESLABONES 1950	1951	1952	MEDIANAS	CADENA
I	88	120	152	—	0.96	0.94	0.95	1.000
II	144	192	240	1.64	1.60	1.58	1.60	1.600
III	143	186	230	0.99	0.97	0.98	0.97	1.552
IV	125	162	198	0.88	0.87	0.86	0.87	1.350
								1.283
			Primera mediana trimestral repetida				0.95	

ESTACIONES *Trimestrales*	AJUSTE	ÍNDICE
I	0.93	77.5
II	1.46	121.6
III	1.34	111.6
IV	1.07	89.1
	1.00	
Suma de los cuatro primeros eslabones ajustados de la cadena	4.80	
Media	$\frac{4.80}{4} = 1.20$	

Análisis de una serie temporal

Para realizar el análisis de una serie temporal, se sigue este procedimiento:

1. Se determina la tendencia de la serie.
2. Con base en la tendencia, se calculan los *valores teóricos* de la serie.
3. Se determinan los índices de variación estacional.
4. Se multiplican los valores teóricos de la tendencia por los índices de variación estacional, para obtener los llamados *valores normales*.
5. Se dividen los valores reales entre los valores normales, para expresar los valores reales como porcientos de los normales.
6. Se resta 100 de los resultados así obtenidos, para determinar así los *índices de variación cíclica* que son desviaciones porcentuales.

Un análisis como éste descompone una serie temporal en los tres elementos siguientes:

1. Una línea y una ecuación de tendencia.
2. Unos porcientos de variación estacional.
3. Unas desviaciones porcentuales de carácter más o menos cíclico (si hay auténtico ciclo, a esas desviaciones se les podrá interpolar una senoidal u otra línea de tipo parecido al de ésta).

La serie empírica está constituida por los valores teóricos de la tendencia modulados por las variaciones estacionales y los ciclos, y afectados por las fluctuaciones.

Si de la serie se eliminan los valores de la tendencia (por resta o división, según convenga) queda un complejo de variación estacional, ciclo y fluctuación.

Si de la serie se eliminan variaciones estacionales y ciclo, queda la tendencia afectada por las fluctuaciones.

Como el análisis se puede hacer eliminando los componentes por substracción o por división, en el momento complementario de síntesis la serie debe reconstituirse precisamente a base de las operaciones inversas: mediante suma si se restó; mediante multiplicación si se dividió.

EJEMPLO DE PERECUACIÓN DE UNA EXPRESIÓN DE PRIMER GRADO A UNA SERIE CRONOLÓGICA

Entrada de nacionales y extranjeros, de México, entre 1946 y 1959

TIEMPO *Años*	CÓDIGO *Número de orden del año en el periodo*	MAGNITUD *Personas que salieron de México en ese año*	MAGNITUD POR CÓDIGO	MAGNITUDES AL CUADRADO
1946	1	319 576	319 576	1
1947	2	308 180	616 360	4
1948	3	315 035	945 105	9
1949	4	374 760	1 499 040	16
1950	5	456 356	2 281 780	25
1951	6	504 267	3 025 602	36
1952	7	524 878	3 674 146	49
1953	8	504 067	4 032 536	64
1954	9	593 898	5 345 082	81
1955	10	636 944	6 369 440	100
1956	11	711 399	7 825 389	121
1957	12	756 792	9 081 504	144
1958	13	795 778	10 345 114	169
1959	14	856 519	11 991 266	196
Σ	105	7 658 449	67 351 940	1 015
	Σx	Σy	Σxy	Σx^2

COMPROBACIÓN DE LAS SUMAS DE NÚMEROS NATURALES y DE CUADRADOS DE ESOS NÚMEROS;

$$\Sigma x = N\frac{(N+1)}{2} = 14 \times \frac{15}{2} = 7 \times 15 = 105$$

$$\Sigma x^2 = N\frac{(N+1)}{2}\frac{(2N+1)}{3} = 14 \times \frac{15}{2} \times \frac{29}{3} = 1\,015$$

ECUACIÓN GENERAL

$$y = a + bx$$

ECUACIONES DE INTERPOLACIÓN

$$\Sigma y = aN + b\Sigma x$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

$$7\,658\,449 = 14a + 105b$$
$$67\,351\,940 = 105a + 1\,015b$$

$$-[114\,876\,735 = 210a + 1\,575b]$$
$$134\,703\,880 = 210a + 2\,030b$$

$$19\,827\,145 = 455b$$
$$\frac{19\,827\,145}{455} = b =$$

$b = 43\,576.14$ personas por año

$$7\,658\,449 = 14a + 105 \times 43\,576.142$$
$$= 14a + 4\,575\,494.910$$
$$-4\,575\,494.91 + 7\,658\,449 = 14a$$
$$+3\,082\,954 = 14a$$
$$\frac{3\,082\,954}{14} =$$

$a = 220\,211$ personas

ECUACIÓN GENERAL ESPECIFICADA

$$y = 220\,211 + 43\,576.14x$$

O sea que en el periodo 1946 a 1959, en México, a partir de una salida teórica de nacionales y extranjeros de 220 211 personas, las salidas aumentaron a razón de 43 576 personas adicionales por año.

Síntesis de una serie temporal

La síntesis de una serie temporal se realiza, principalmente, cuando se quiere conocer —tan aproximadamente como sea posible— el valor que tuvo, ha tenido o tendrá el fenómeno en un momento dado del pasado remoto, del pasado reciente o del futuro más o menos próximo. Para realizar la síntesis se combinan los tres elementos principales de toda serie temporal (tendencia, variación estacional y variación cíclica) del modo siguiente:

1. Se determina cuántos años dista la fecha que interesa, del origen temporal de la serie con que se trabaja (más concretamente, de la parte de la serie que sirvió para precisar la tendencia).
2. Se precisa cuál ha sido la fecha en que se recogieron los datos de la serie real, en el curso de cada año (por ejemplo, si los datos corresponden al 1º de julio de cada año) y se determina si el dato que interesa corresponde o no a esa fecha. En caso de no corresponder a ella, se determinará cuántos meses dista la fecha que interesa de la fecha central en que se recogieron los datos, y se agrega tantas veces $1/12 = 0.0933$ como meses haya entre una y otra, al número de años determinado en el 1er. paso.
3. Se substituye esa suma en el lugar de x dentro de la ecuación de la tendencia, para obtener el valor teórico.
4. Se determina a qué estación (mes, trimestre, etcétera) corresponde la fecha y cuál es el índice de variación que corresponde a esa estación.
5. Se multiplica el valor teórico obtenido en el 3er. paso por el índice de variación estacional encontrado en el 4º paso, con lo que se obtienen valores normales.
6. Se determina a qué fase del ciclo corresponde la fecha con la que se trabaja y se precisa cuál es la desviación porcentual que corresponde a esa fase.
7. Se le suma a la desviación cíclica porcentual, 100, y
8. Se multiplica el resultado obtenido en el 5º paso por la suma obtenida en el 7º

Ejemplo. Calcúlese para enero de 1980 la población de un lugar que en 1970 (10 años) tenía, en julio (10 años menos 6 meses $= 10 - 6 \times .0933 = 9.44$ años), una población de 50 000 ($=$ a) que crecía a razón de 150 por año (b) (de donde $y = 50\,000 + 150 \times 9.44 = 51\,416$) que en los meses de enero tenía una variación estacional de 92% ($51\,416 \times 0.92 = 47\,303$) y que tenía como variación cíclica $+ 5\%$ ($1 + 0.05 = 1.05$) Resultado: $47\,303 \times 1.05 = 49\,668$.

EJEMPLO DE CÁLCULO DE LOS VALORES TEÓRICOS DE UNA TENDENCIA DE PRIMER GRADO

Entrada de nacionales y extranjeros, a México, en el periodo comprendido de 1946 a 1959, según la tendencia:

$$y = 220\,211 + 43\,576\,x$$

Años	*Código* x	*Incremento del año respecto del principio del periodo* 43 576 x	*Valor teórico de la tendencia para ese año = valor inicial + incremento hasta ese año:* 220 211 + 43 576 x
	0	0	220 211
1946	1	43 576	263 787
1947	2	87 152	307 363
1948	3	130 728	350 939
1949	4	174 304	394 515
1950	5	217 880	438 091
1951	6	261 456	481 667
1952	7	305 032	525 243
1953	8	348 608	568 819
1954	9	392 184	612 395
1955	10	435 760	655 971
1956	11	479 336	699 547
1957	12	522 912	743 123
1958	13	566 488	786 699
1959	14	610 064	830 275

2.7. ALGUNAS TÉCNICAS PARA LA DETERMINACIÓN DEL GRADO DE ASOCIACIÓN ENTRE LAS SERIES ESTADÍSTICAS

Índice de predecibilidad de Guttman

El grado de asociación entre dos series nominales puede determinarse —como otros índices de correlación— si se comparan entre sí dos medidas de dispersión: la de la distribución univariada y la de la distribución bivariada, constituidas por una de las series y por la conjunción de ambas. En este caso, la comparación se establece entre las razones de variación de la serie nominal univariada (a la que quiera considerarse como variable dependiente) y de la bivariada constituida por ella misma y por la otra serie (a la que se considerará como variable independiente).

Supongamos que, de acuerdo con la primera serie (considerada en aislamiento):

1. Tienen el carácter que sirve de criterio clasificador, a individuos (por ejemplo: a individuos tienen instrucción universitaria), y
2. Carecen del carácter que sirve para la clasificación, b individuos (b individuos no tienen instrucción universitaria).

En este caso, el total de individuos es a + b.

Si, además, damos por supuesto que b es mayor que a, la razón de variación del fenómeno será: $1 - \frac{b}{a+b} = \frac{a+b-b}{a+b}$. O sea, que la razón de variación del fenómeno es: $\frac{a}{a+b}$.

En seguida, supondremos que clasificamos a esos mismos a + b individuos, simultáneamente, de acuerdo con el primer criterio "instrucción universitaria" y de acuerdo con un segundo criterio como el de si "pertenecen o no a la clase media".

Supongamos que, de acuerdo con esta doble clasificación de los a + b individuos del grupo, se tiene la situación siguiente:

hay c individuos que tienen instrucción universitaria y que pertenecen a la clase media,
hay d individuos que no tienen instrucción universitaria pero pertenecen a la clase media,
hay e individuos que tienen instrucción universitaria pero no pertenecen a la clase media,
hay f individuos que no tienen instrucción universitaria y no pertenecen a la clase media.

Esta situación puede esquematizarse en un cuadro de doble entrada, también llamado cuadro de contingencia, como el que sigue:

		FENÓMENO 2 *Pertenecen a la clase media*	
		TIENEN EL CARÁCTER *Son de clase media*	NO TIENE EL CARÁCTER *No son de clase media*
FENÓMENO 1 Instrucción universitaria	TIENEN EL CARÁCTER *Tienen instrucción universitaria*	c	e
	NO TIENEN EL CARÁCTER *No tienen instrucción universitaria*	d	f

El grupo total está formado por los $a + b$ individuos del principio; o sea, por $c + d + e + f$ individuos según esta forma conjunta de clasificarlos.

Si suponemos que al observar los valores del cuadro descubrimos que c es mayor que e, esto significa que la razón de variación de la distribución constituida por el primer renglón será: $1 - \frac{c}{c+e} = \frac{e}{c+e}$.

En forma parecida, si f es mayor que d, la razón de variación del segundo renglón será: $1 - \frac{f}{d+f} = \frac{d}{d+f}$.

En cambio, la razón de variación total será: $\frac{e+d}{c+d+e+f}$.

Ahora es cuando se puede establecer una comparación entre las dos razones de variación: $\frac{a}{a+b}$ y $\frac{e+d}{c+d+e+f}$.

Al restar una de otra esas razones de variación, se obtiene:

$$\frac{a}{a+b} - \frac{e+d}{c+d+e+f}$$

Esta diferencia representa la que existe entre el error cometido al dar como representativo el modo y el error que se comete al tomar la segunda serie como base para predecir los valores de la primera.

Como $a + b = c + d + e + f$, se puede escribir:

$$\frac{a - (e+d)}{a+b}$$

Si queremos obtener el error relativo, dividiremos este resultado entre la razón de variación de la serie univariada: $(a/(a + b)$:

$$\frac{\frac{a-(e+d)}{a+b}}{\frac{a}{a+b}}$$

Como el denominador de la fracción denominadora y el de la fracción numeradora son iguales, se reducen a la unidad. O sea, que el error relativo se reduce a:

$$\frac{a - (e+d)}{a}$$

EJEMPLO PEDAGÓGICO DE CÁLCULO DEL ÍNDICE DE PREDECIBILIDAD DE GUTTMAN

Ejemplo tomado de Linton C. Freeman:

Elementary Applied Statistics for Students in Behavioral Science, pero resuelto conforme a un procedimiento distinto

		RESPUESTA A UN PROBLEMA LÓGICO		SUMAS
		Pasaron	*No pasaron*	*Total de examinados*
ESTUDIOS PREVIOS DE MATEMÁTICAS	Los tenían	22	3	25
	No los tenían	8	17	25
SUMAS	Total de alumnos con o sin antecedentes matemáticos	30	20	50

a = clase no modal de la dependiente "respuesta al problema lógico" — 20

e = clase no modal de quienes tenían antecedentes matemáticos — 3

d = clase no modal de quienes no tenían antecedentes matemáticos — 8

$$\text{Índice: } \frac{a-(e+d)}{a} = \frac{20-(3+8)}{20} = \frac{20-11}{20} = \frac{9}{20} = 0.45$$

Este error relativo da el grado de asociación entre las series: a es, en él, la frecuencia no modal de la distribución univariada; e y d son las frecuencias no modales de las distribuciones filares * del fenómeno considerado como dependiente.

Expresado en términos y con fórmulas un tanto distintas, Friedman presenta este índice en su texto de estadística y le atribuye el nombre de "índice de predecibilidad de Guttman".

* El término "filar" se refiere aquí a las filas o hileras del cuadro, en forma parecida a como el término "columnar" hace referencia a las columnas del mismo.

EJEMPLO DE DETERMINACIÓN DEL GRADO DE ASOCIACIÓN ENTRE DOS SERIES NOMINALES

Asociación entre el sexo del individuo y su actividad económica, en México, en 1960

		ACTIVIDAD ECONÓMICA *Activos*	*Inactivos*	SUMAS *Activos e inactivos*
SEXO	Hombres	9 296 723	8 118 597	17 415 320
	Mujeres	2 035 293	15 472 516	17 507 809
SUMAS	Hombres y Mujeres	11 332 016	23 591 113	34 923 129

a = frecuencia no modal de la dependiente "actividad económica" — 11 332 016

e = frecuencia no modal de la primera filar de la independiente: "hombres" — 8 118 597

d = frecuencia no modal de la segunda filar de la independiente: "mujeres" — 2 035 293

$$\text{Índice} = \frac{a-(c+d)}{a} = \frac{11\,332\,016-(8\,118\,597+2\,035\,293)}{11\,332\,016} =$$

$$= \frac{11\,332\,016-10\,153\,890}{11\,332\,016} =$$

$$= \frac{1\,178\,126}{11\,332\,016}$$

$$= 0.1039 = 10.39\%$$

EJEMPLO DE CALCULO DEL ÍNDICE DE CORRELACIÓN (DE SPEARMAN) SERIES ORDINALES

Correlación entre número de municipios y total de habitantes de las entidades federativas de México, en 1960

CATEGORÍAS *Entidades federativas*	VARIABLES 1ª *Municipios*	2ª *Habitantes*	ORDENACIONES DE ACUERDO CON *la 1ª y la 2ª* VARIABLES		DESVIACIONES	CUADRADOS DE DESVIACIONES
AGUASCALIENTES	8	243 363	28.5	28	+ 0.5	.25
BAJA CALIFORNIA N.	4	520 165	31.5	22	+ 9.5	90.25
BAJA CALIFORNIA S.	7	81 594	30	31	– 1.0	1.00
CAMPECHE	8	168 219	28.5	29	– 0.5	.25
COAHUILA	38	907 734	19.5	16	+ 3.5	12.25
COLIMA	9	164 450	27	30	– 3.0	9.00
CHIAPAS	111	1 210 870	6	10	– 4.0	16.00
CHIHUAHUA	66	1 226 793	12	9	+ 3.0	9.00
DISTRITO FEDERAL	13	4 870 876	26	1	+ 25.0	625.00
DURANGO	38	760 836	19.5	20	– 0.5	0.25
GUANAJUATO	46	1 735 490	16	7	+ 9.0	81.00
GUERRERO	75	1 186 716	10	11	– 1.0	1.00
HIDALGO	82	994 598	9	15	– 6.0	36.00
JALISCO	124	2 443 261	4	3	+ 1.0	1.00
MÉXICO	119	1 897 851	5	5	0.0	0.00
MICHOACAN	110	1 851 876	7	6	+ 1.0	1.00

MORELOS	32	386 264	21	25	— 4.0	16.00
NAYARIT	19	389 929	22	24	— 2.0	4.00
NUEVO LEÓN	52	1 078 848	15	12	+ 3.0	9.00
OAXACA	571	1 727 266	1	8	— 7.0	49.00
PUEBLA	222	1 973 837	2	4	— 2.—	4.00
QUERÉTARO	18	355 045	23	26	— 3.0	9.00
QUINTANA ROO	4	50 169	31.5	32	— 0.5	0.25
SAN LUIS POTOSÍ	54	1 048 297	13	13	0.0	0.00
SINALOA	16	838 404	25	17	+ 8.0	64.00
SONORA	72	783 378	11	19	— 8.0	64.00
TABASCO	17	496 340	24	23	+ 1.0	1.00
TAMAULIPAS	41	1 024 182	18	14	+ 4.0	16.00
TLAXCALA	44	346 699	17	27	— 10.0	100.00
VERACRUZ	198	2 727 899	3	2	+ 1.0	1.00
YUCATÁN	106	614 049	8	21	— 13.0	169.00
ZACATECAS	53	817 831	14	18	— 4.0	16.00
			528	528	0.0	1 405.50
(32 × 33) / 2 =			COMPROBACIONES			SUMA DE CUADRADOS DE DESVIACIONES

SUMA DE CUADRADOS DE LAS DESVIACIONES DEL ORDEN DE LAS ENTIDADES SEGÚN NÚMERO DE MUNICIPIOS RESPECTO DE SU ORDEN SEGÚN NÚMERO DE HABITANTES: 1 405.50

EJEMPLO DE CALCULO DEL ÍNDICE DE CORRELACIÓN (DE SPEARMAN) SERIES ORDINALES (continuación)

Correlación entre número de municipios y total de habitantes de las entidades federativas de México, en 1960

SUMA DE CUADRADOS DE LAS DESVIACIONES DEL *ORDEN* DE LAS ENTIDADES SEGÚN EL NÚMERO DE MUNICIPIOS RESPECTO DE SU *ORDEN* SEGÚN NÚMERO DE HABITANTES = 1 405.50

SEIS VECES ESA SUMA = 1 405.50 × 6 = 8 433.00

NÚMERO DE ENTIDADES = 32

CUBO DEL NÚMERO DE ENTIDADES = 32^3 = 32 768

CUBO DEL NÚMERO MENOS EL NÚMERO = 32 728 − 32 = 32 736

COCIENTE DE LOS DOS RESULTADOS; $\frac{8\,433}{32\,736} = 0.257$

DIFERENCIA ENTRE UNO Y EL COCIENTE ANTERIOR: 1 − 0.257 = 0.743

ÍNDICE DE CORRELACIÓN DE SPEARMAN: 0.743 = 74.3%

N. B. En este ejemplo, se ha aplicado el procedimiento de Spearman en su forma original, sin considerar que hay "rangos ligados" (Aguascalientes y Campeche, de acuerdo con la primer variable tienen números de orden o rango 28.5, por ejemplo). Se ha procedido así, sin usar la modificación del procedimiento impuesto por la existencia de dichos "rangos ligados" en función de dos consideraciones: 1ª que, como dice Kendall, los resultados obtenidos cuando el número de ligas no es muy grande, son aproximadamente iguales, úsese uno u otro procedimiento, y 2ª que el procedimiento más fino puede resultar excesivamente complicado para un nivel elemental de instrucción como es éste en el que nos movemos.

Distribuciones bivariadas

Una distribución es bivariada cuando frente a cada uno de los individuos del conjunto correspondiente se consignan dos magnitudes: una que mide un carácter y la otra que mide otro carácter de dicho individuo. Así, por ejemplo, si frente al señor A.M.X. se consigna que gana 500 pesos mensuales y tiene 9 años de instrucción; frente a B.N.Y. que gana 5 000 y tiene 12 años de instrucción, y frente a C.O.Z. que gana 10 000 y tiene 18 años de instrucción, 500, 5 000 y 10 000; 9, 12 y 18 constituyen una distribución bivariada.

Una distribución bivariada puede considerarse constituida —para ciertos propósitos— por dos distribuciones univariadas. En el ejemplo, la distribución bivariada 500-9; 5 000-12; 10 000-18 está constituida por las univariadas 500, 5 000, 10 000 y 9, 12, 18.

Cada distribución bivariada tiene sus medidas estadísticas propias (destacadamente, el índice de correlación), pero, cada una de las distribuciones univariadas componentes conserva sus estadísticas particulares (una media aritmética para cada una de las dos distribuciones; una desviación cuadrática media para cada una de ellas).

La construcción estadística y la parsimonia metodológica imponen que las nuevas medidas descriptivas de la distribución bivariada resultante se basen en las antiguas (las de las distribuciones univariadas componentes). Así, el índice de correlación se expresa en términos de desviaciones de cada una de las dos series componentes respecto a su media, expresadas en unidades sigmáticas de la correspondiente distribución univariada.

CUADRO DE DOBLE ENTRADA

para las distribuciones bivariadas

	CARACTER 2	COLUMNAS
	CARACTER 1	Clases del segundo carácter de la distribución bivariada
HILERAS	Clases del primer carácter de la distribución bivariada	Individuos que tienen simultáneamente los dos caracteres de la distribución bivariada.

La presentación de las series bivariadas puede hacerse utilizando dos columnas contiguas para registrar en cada una de ellas, frente a cada individuo, el valor de cada uno de sus dos caracteres:

INDIVIDUO	CARÁCTER 1	CARÁCTER 2
A.M.X.	$ 500	9 años
B.N.Y.	$ 5 000	12 años
C.O.Z.	$ 10,000	18 años

Sin embargo, ésta es una forma particular de presentación que no se puede usar cuando se trata de clases y no de individuos.

Cuadros de doble entrada

Cuando los datos de una distribución bivariada están clasificados (a lo largo de uno o de los dos caracteres considerados) se requiere un cuadro de doble entrada para su presentación.

Un cuadro de doble entrada consigna en forma ordinaria (o sea, verticalmente) la serie de clases del primer carácter (en calidad de rubros) y coloca —formando ángulo recto con él— la serie de clases del segundo (a manera de encabezados). Ver el cuadro.

Frecuencias de una distribución bivariada

Una vez que está formada la serie bivariada de clases, se examina a cada individuo para ver a qué clase pertenece según el primero y a cuál corresponde de acuerdo con el segundo carácter. En el cruce de la hilera y la columna correspondientes, se marca la existencia de ese individuo mediante un palote.

Cuando se han consignado todos los individuos de la serie por medio de palotes, se hace el recuento de los que hay en cada casilla (en forma parecida a como se procedió en la formación de series de frecuencias). El número resultante es la frecuencia con la que aparece conjuntamente la clase que indica el rubro de la hilera y la clase indicada por el encabezado de la columna correspondiente, en las distribución. A éstas se les llama "frecuencias conjuntas".

Frecuencias de las distribuciones univariadas componentes

Como se dijo antes, el cálculo de las estadísticas propias de la distribución bivariada resultante depende del de las que caracterizan las distribuciones univariadas componentes. Estas últimas, a su vez, dependen, para su cálculo, de las frecuencias de cada distribución univariada.

En un cuadro de doble entrada, para obtener la distribución univariada del primer carácter basta con sumar renglón por renglón todas las frecuencias conjuntas que en él aparezcan y consignar el resultado de cada suma al final de la hilera. Las clases que sirven de rubros, y la columna resultante de la consignación de esas sumas constituye la primera distribución univariada componente.

Para obtener la segunda distribución univariada, bastará con sumar columna por columna todas las frecuencias conjuntas que aparezcan en ellas y consignar la suma al pie. Las clases que sirven de encabezado y la hilera de esas sumas constituye la segunda distribución componente.

Cálculo de estadísticas de las distribuciones univariadas componentes

Para cualquier cálculo estadístico en el que intervengan las distribuciones univariadas, se procede como si no existiera el centro del cuadro con sus frecuencias conjuntas.

Si en las guías de procedimiento se dice, por ejemplo, "se consignan en una columna las desviaciones" hay que entender esto literalmente para la primera distribución univariada componente; pero hay que cambiar la orden por la de "se consignan en un *renglón* las desviaciones" si se trata de la segunda distribución univariada componente.

Ponderación de las distribuciones bivariadas

En cuanto en el cálculo de una medida estadística intervienen no una sino las dos variables de la distribución u otras magnitudes derivadas de ellas, la ponderación tiene que hacerse mediante las frecuencias conjuntas.

Generalmente, en estos casos, se trata de multiplicaciones; en ellas intervienen tres factores: uno que corresponde al primer carácter; otro que corresponde al segundo, y la frecuencia conjunta que liga a ambos.

En la práctica, lo que se hace es tomar cada frecuencia conjunta y multiplicarla por las magnitudes correspondientes que se consignan en su mismo renglón y en su misma columna.

Toda ponderación de este tipo requiere: o la consignación de los resultados en un ángulo de la casilla correspondiente, o de la elaboración de un cuadro adicional que reproduzca el que ya registra las frecuencias conjuntas y que, en la práctica o se traza al lado de éste o se elabora en papel transparente, superponiéndolo al principal y consignando en él los resultados.

Efectivo de la distribución bivariada

En toda distribución bivariada, el efectivo es igual a la suma de las frecuencias conjuntas. Esta suma es —además— idéntica a la suma de las frecuencias de cada distribución univariada componente.

EJEMPLO DE PRESENTACIÓN DE LAS DOS SERIES UNIVARIADAS COMPONENTES DE LA SERIE BIVARIADA

		INSTRUCCIÓN EN AÑOS TERMINADOS Y APROBADOS				
		1 a 6	7 a 9	10 a 13	14 a 18	19 a 21
I	0 a 500 —					
N	500 a 1 000 —					
G	1 000 a 1 500 —					
R	1 500 a 2 000 —					
E	2 000 a 2 500 —					
S	2 500 a 3 000 —					
O	3 000 a 3 500 —					
S	3 500 a 4 000 —					

EJEMPLO DE PRESENTACIÓN DE LA DISTRIBUCIÓN BIVARIADA

		INSTRUCCIÓN EN AÑOS TERMINADOS Y APROBADOS				
		1 a 6	7 a 9	10 a 13	14 a 18	19 a 21
I	0 a 500 —	6				
N	500 a 1 000 —	5	4			
G	1 000 a 1 500 —		5	7	3	
R	1 500 a 2 000 —			8	4	
E	2 000 a 2 500 —				2	9
S	2 500 a 3 000 —					5
O	3 000 a 3 500 —				1	6
S	3 500 a 4 000 —					3

6 es la frecuencia conjunta de 0 a 500 — y de 1 a 6

8 es la frecuencia conjunta de 1 500 a 2 000 y de 10 a 13

EJEMPLO DE CALCULO DE LAS FRECUENCIAS DE LAS DISTRIBUCIONES UNIVARIADAS COMPONENTES A PARTIR DE LAS FRECUENCIAS CONJUNTAS DE LA DISTRIBUCIÓN BIVARIADA

		INSTRUCCIÓN EN AÑOS TERMINADOS Y APROBADOS					
		1 a 6	7 a 9	10 a 13	14 a 18	19 a 21	f_1
I	0 a 500 —	6					6
N	500 a 1 000 —	5	4				5 + 4
G	1 000 a 1 500 —		5	7	3		5 + 3 + 7
R	1 500 a 2 000 —			8	4		8 + 4
E	2 000 a 2 500 —				2	9	2 + 9
S	2 500 a 3 000 —					5	5
O	3 000 a 3 500 —				1	6	1 + 6
S	3 500 a 4 000 —					3	3
	f_2	11	9	15	10	23	

LAS DISTRIBUCIONES UNIVARIADAS COMPONENTES SON:

I	0 a 500	F	6
N	500 a 1 000	R	11
G	1 000 a 1 500	E	15
R	1 500 a 2 500	C	12
E	2 000 a 2 500	U	11
S	2 500 a 3 000	E	5
O	3 000 a 3 500	N	7
S	3 500 a 4 000	C	3
		I	
		A	
S	3 500 a 4 000	S	3

Y:

INSTRUCCIÓN EN AÑOS TERMINADOS Y APROBADOS				
1 a 6	7 a 9	10 a 13	14 a 18	19 a 21
FRECUENCIAS				
11	9	15	10	23

Índice de correlación

En el estudio de las series bivariadas, importa conocer el grado de intensidad con que las variaciones de magnitud de uno de los caracteres de la serie sigue las variaciones de magnitud del otro carácter: el grado en el que, por ejemplo, se asocian en un grupo de individuos el hecho de ser pobre y el de ser ineducado. Esa intensidad se mide estadísticamente mediante un momento bivariado: el momento 11.

El momento 11 —como todo momento— es una media (o sea, fundamentalmente una suma de algo, entre el efectivo: Σ / N). Como en otros momentos, en éste también intervienen desviaciones, o sea diferencias entre los datos y un valor central tomado como eje de momentos. Por comodidad, conviene hablar desde el principio del momento 11 con respecto a las medias aritméticas de las distribuciones componentes de la distribución bivariada: o sea que las desviaciones que en él intervienen son desviaciones con respecto a las medias aritméticas $(x_1 - \bar{x}_1)$ y $(x_2 - \bar{x}_2)$ que en este caso, representan las desviaciones de los datos de la primera serie (x_1) respecto de su media aritmética $(\bar{x}_1)$ y las desviaciones de los datos de la segunda serie (x_2) respecto de su media aritmética $(\bar{x}_2)$.*

El hecho de que se trate del momento bivariado de orden 11 indica no que las desviaciones de una serie única hayan de elevarse a la potencia once, sino que hay que tomar las primeras potencias de una de las dos series componentes y las primeras potencias de la segunda de esas series.

Para conjuntar la desviaciones de ambas series, se recurre a la multiplicación.

O sea, que el momento de orden once es: la suma/ del producto/ de las desviaciones de la primera serie respecto de su media aritmética/ multiplicadas por/ las desviaciones de la segunda serie respecto de la suya/ dividido todo ese producto entre el efectivo de la distribución o número de datos:

$$\frac{\Sigma\,[(x_1 - \bar{x}_1)\ (x_2 - \bar{x}_2)]}{N}$$

Cuando la serie bivariada está formada por dos series univariadas, cuyos datos están agrupados en clases, es necesario ponderar por las frecuencias correspondientes; o sea, que hay que multiplicar cada desviación por su frecuencia de clase. Como hay una doble agrupación, cada frecuencia (que aquí designaremos como "frecuencia conjunta de las distribuciones 1

* Nótese que, en este caso, para simplificar la notación usamos el símbolo $\bar{x}_2$ para representar la media aritmética (o media primera de la segunda serie y no como convinimos en el estudio general de las medias, para indicar la media cuadrática o segunda media de las equis).

y 2, o f_{12}") tiene que multiplicarse tanto por la desviación de los datos de la primera serie con respecto a su media, como por la desviación de los datos de la segunda serie respecto de la suya. Es decir, que en el numerador aparecerá:

$$\Sigma [(x_1 - \bar{x}_1) \; f_{12} \; (x_2 - \bar{x}_2)]$$

Por su parte, el efectivo es tanto la suma de las f_1 como la suma de las f_2, como la suma de las f_{12}. O sea, que en el denominador debe figurar: Σf_{12}

El momento 11 resulta ser:

$$\frac{\Sigma [(x_1 - \bar{x}_1) \; f_{12} \; (x_2 - \bar{x}_2)]}{\Sigma f_{12}}$$

A fin de reducir esta medida de relación a términos abstractos y comparables de serie a serie hay que expresarla en unidades sigmáticas.

En este caso, como la serie es bivariada, hay dos series componentes, cada una de las cuales tiene su propia sigma o desviación cuadrática media. Para conjuntarlas, hay que multiplicar la desviación cuadrática media de la primera serie (σ_1) por la desviación cuadrática media de la segunda (σ_2).

O sea, que el momento de orden once en unidades sigmáticas es igual a:

$$\frac{\Sigma [(x_1 - \bar{x}_1) \; f_{12} \; (x_2 - \bar{x}_2)]}{\sigma_1 \; \Sigma \; f_{12} \; \sigma_2}$$

A esta medida se le llama "índice de correlación (rectilínea)" y se le representa por r.

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:

1. 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.

2. 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.

3. Calcúlese la media de la primer variable:
   3.1. Multiplicando cada valor o punto medio de dicha variable por la frecuencia correspondiente y anotando dichos productos en un renglón.
   3.2. Sumando todos los productos así obtenidos, y
   3.3. Dividiendo la suma entre el total de las frecuencias.

4. Calcúlese en forma análoga (usando las columnas) la media aritmética de la segunda variable.

5 y 6. Determínense en los renglones y en las columnas, respectivamente, las desviaciones con respecto a la media aritmética:
   5.1 y 6.1. 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
   5.2 y 6.2. Restando de cada valor o punto medio de la segunda variable la media aritmética correspondiente, anotando los resultados en una columna.

7. 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 registre el producto de estas tres 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.

8. Obténganse las desviaciones medias cuadráticas de las dos variables. Para ello:
   8.1. Elévense al cuadrado las desviaciones obtenidas en 5º y 6º, anotando los resultados en un renglón y una columna.
   8.2. Súmense los cuadrados de las desviaciones de cada una de las variables con respecto a sus correspondientes medias (por separado).
   8.3. Divídanse las sumas así obtenidas entre la suma de las frecuencias.
   8.4. Extráiganse las raíces cuadradas de dichos cocientes. Los resultados son las desviaciones medias cuadráticas de cada una de las variables.

9. 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.

10. Divídase el valor obtenido en 7º entre el valor obtenido en 9º El resultado es el índice de correlación r.

Decir que entre dos variables (x_1 y x_2) hay una relación rectilínea, equivale a afirmar:

1. Que la primera variable (x_1) es igual:
   al producto de la segunda (x_2) por una constante (b_{12}) más un término constante (a_{12}), y
2. Que, en forma conversa, la segunda variable (x_2) es igual a:
   el producto de la primera variable (x_1) por una constante (b_{21}) más un término constante (a_{21}).

Como es fácil ver b_{12} no es igual a b_{21} pero está relacionada con ella, y a_{12} tampoco es igual con a_{21}, pero con ella se relaciona.

En términos simbólicos, lo anterior equivale a establecer las dos ecuaciones siguientes:

$$x_1 = a_{12} + b_{12} x_2$$

$$x_2 = a_{21} + b_{21} x_1$$

A estas dos se les conoce como ecuaciones de regresión, de la primera variable en la segunda y de la segunda en la primera.

Las ecuaciones de regresión en términos de desviación

Si de ambos miembros de la primera ecuación de regresión restamos una constante la ecuación no se altera. Como la media aritmética de la primera variable (x_1) es constante, podemos restarla de ambos miembros, sin producir alteración:

$$x_1 - \bar{x}_1 = a_{12} - \bar{x}_1 + b_{12} x_2$$

Lo que ahora aparece en el primer miembro es, por definición, la desviación de las x_1 con respecto a su media aritmética, que podemos representar por d_1

$$x_1 - \bar{x}_1 = d_1$$

Lo que aparece en el segundo miembro de la ecuación que precede a ésta es la diferencia entre dos constantes a_{12} y $\bar{x}_1$ más un producto. La diferencia entre esas constantes puede representarse por otra constante:

$$a_{12} - \bar{x} = A_{12}$$

Si, además, para homogeneidad de la simbología, representamos al coeficiente de x_2 por la mayúscula y no por la minúscula:

$$b_{12} = B_{12}$$

podremos escribir la primera ecuación de regresión:

$$d_1 = A_{12} + B_{12}\, x_2$$

Si, basándonos en la misma propiedad, restamos de ambos miembros la media de las x_2 ($\bar{x}_2$) multiplicada por B_{12}, la ecuación no se altera pues se trata de la resta de un producto constante:

$$d_1 - B_{12}\,\bar{x}_2 = A_{12} + B_{12}\,x_2 - B_{12}\bar{x}_2$$

Al sacar, en el segundo miembro como factor común a B_{12}, se obtiene:

$$d_1 - B_{12}\,\bar{x}_2 = A_{12} + B_{12}\,(x_2 - \bar{x}_2)$$

En el segundo término del segundo miembro, aparecen ahora las desviaciones de la segunda variable, que se pueden representar por d_2

$$d_1 - B_{12}\,\bar{x}_2 = A_{12} + B_{12}\,d_2$$

Al pasar $B_{12}\,\bar{x}_2$ del primero al segundo miembro, con signo contrario:

$$d_1 = A_{12} - B_{12}\,\bar{x}_2 + B_{12}\,d_2$$

En el segundo miembro, los dos primeros términos son constantes y pueden representarse por una constante (A'_{12}):

$$d_1 = A'_{12} + B_{12}\,d_2$$

Ésta es la primera ecuación de regresión expresada en desviaciones de las variables respecto de sus medias.

La segunda ecuación de regresión se puede transformar en forma parecida, en:

$$d_2 = A'_{21} + B_{21}\,d_1$$

Esta expresión es simplificable, pues puede demostrarse que A'_{12} es nula, y que es nula, también A'_{21} · En efecto, si se aplica el sumador a ambos miembros de cualesquiera de las ecuaciones de regresión y se divide entre N, se obtiene:

$$\frac{\Sigma\, d_1}{N} = \frac{NA'_{12}}{N} + \frac{\Sigma\, B_{12}\,d_2}{N}$$

En esta expresión es fácil reconocer en el primer miembro, y en el segundo término del segundo miembro, primeros momentos de las desviaciones (d) respecto de la media aritmética. Esos primeros momentos, como se sabe, son nulos. Además de eso, en el primer término del segundo miembro A'_{12} aparece multiplicada y dividida por N, o sea, multiplicada por la unidad, con lo cual resulta:

$$0 = A'_{12} + 0$$

Es decir que A'_{12} es nula, y puede escribirse la ecuación de regresión:

$$d_1 = B_{12}\, d_2$$

En forma parecida, la segunda de regresión se reduce a:

$$d_2 = B_{21}\, d_1$$

Las ecuaciones de regresión en términos sigmáticos

Las ecuaciones anteriores se pueden expresar en unidades sigmáticas. Si en la primera de regresión dividimos ambos miembros por la desviación cuadrática media de la primera variable (σ_1):

$$\frac{d_1}{\sigma_1} = \frac{B_{12}\, d_2}{\sigma_1}$$

Si representamos la desviación sigmática del primer miembro por δ_1:

$$\delta_1 = \frac{B_{12}\, d_2}{\sigma_1}$$

Si, en el segundo miembro, multiplicamos y dividimos por la desviación cuadrática media de la segunda variable (σ_2):

$$\delta_1 = \frac{B_{12}\, \sigma_2\, d_2}{\sigma_1\, \sigma_2}$$

En esta expresión d_2/σ_2 representa la desviación de la segunda variable en unidades sigmáticas, o sea δ_2:

$$\delta_1 = \frac{B_{12}\, \sigma_2}{\sigma_1}\, \delta_2$$

En forma parecida, la segunda ecuación de regresión puede escribirse, en unidades sigmáticas:

$$\delta_2 = \frac{B_{21}\ \sigma_1}{\sigma_2}\ \delta_1$$

Los coeficientes beta de las ecuaciones de regresión

En las expresiones anteriores, en los segundos miembros, aparecen como coeficientes de las desviaciones en unidades sigmáticas dos cantidades constantes: $(B_{12}\,\sigma_2)/\sigma_1$ y $(B_{21}\,\sigma_1)/\sigma_2$ · Para mayor simplicidad, cada una de estas constantes se puede representar por un nuevo símbolo. El símbolo elegido convencionalmente ha sido la letra griega beta minúscula. De este modo:

$$\frac{B_{12}\ \sigma_2}{\sigma_1} \text{ se representa por } \beta_{12}, \text{ y}$$

$$\frac{B_{21}\ \sigma_1}{\sigma_2} \text{ se representa por } \beta_{21}.$$

O sea, que, en conjunto, a los dos coeficientes de las ecuaciones de regresión expresadas en términos de desviaciones en unidades sigmáticas se les conoce con el nombre de "coeficientes beta".

Estos coeficientes beta indican por qué cantidad hay que multiplicar las desviaciones sigmáticas de una de las variables que intervienen en la regresión, para obtener los valores teóricos de las desviaciones sigmáticas de la otra variable.

El índice de correlación rectilínea como un promedio

Los coeficientes beta se pueden relacionar con el índice de correlación rectilínea r (o r_{12}). El índice de correlación rectilínea es —en efecto— la media geométrica de los coeficientes beta. Según esto, un procedimiento de cálculo del índice de correlación consistirá en:

1. Reducir las dos variables que se relacionan a desviaciones respecto de sus medias, en unidades de sus respectivas desviaciones cuadráticas medias.
2. Interpolar a cada serie una ecuación de regresión y obtener así los dos coeficientes beta (de regresión de la primera en la segunda variable y de regresión de la segunda en la primera variable), y
3. Tomar la media geométrica de ambos coeficientes.

El resultado es el índice de correlación rectilínea.

EJEMPLO DE CALCULO DEL INDICE DE CORRELACIÓN A PARTIR DE LOS COEFICIENTES DE REGRESIÓN

Ejemplo pedagógico de correlación entre las variables x_1 y x_2, cuyos valores fueron los siguientes en los años comprendidos entre 1960 y 1970 (datos: $\bar{x}_1 = 31$; $\bar{x}_2 = 64$; $\sigma_1 = 6.7$; $\sigma_2 = 19.1$)

Años	x_1	x_2	d_1	d_2	δ_1	δ_2	$\delta_1\,\delta_2$	$\delta^2{}_2$	$\delta^2{}_1$
1960	21	41	− 10	− 23	− 1.5	− 1.2	1.80	1.44	2.25
1961	18	34	− 13	− 30	− 1.9	− 1.6	3.04	2.56	3.61
1962	23	38	− 8	− 26	− 1.2	− 1.4	1.68	1.96	1.44
1963	34	67	+ 3	+ 3	+ 0.4	+ 0.2	0.08	0.04	0.16
1964	36	68	+ 5	+ 4	+ 0.7	+ 0.2	0.14	0.04	0.49
1965	38	84	+ 7	+ 20	+ 1.0	+ 1.0	1.00	1.00	1.00
1966	38	76	+ 7	+ 12	+ 1.0	+ 0.6	0.60	0.36	1.00
1967	36	72	+ 5	+ 8	+ 0.7	+ 0.4	0.28	0.16	0.49
1968	32	99	+ 1	+ 35	+ 0.1	+ 1.8	0.18	3.24	0.01
1969	33	67	+ 2	+ 3	+ 0.3	+ 0.2	0.06	0.04	0.09
1970	32	58	+ 1	− 6	+ 0.1	− 0.3	0.03	0.09	0.01
SUMAS							8.89	10.93	10.55

Ecuaciones de regresión:

$$\delta_1 = \beta_{12}\,\delta_2 \qquad\qquad \delta_2 = \beta_{21}\,\delta_1$$

ECUACIONES DE INTERPOLACIÓN:

$$\Sigma\,\delta_1\,\delta_2 = \beta_{12}\,\Sigma\,\delta^2{}_2 \qquad\qquad \Sigma\,\delta_2\,\delta_1 = \beta_{21}\,\Sigma\,\delta^2{}_1$$

SUSTITUCIÓN:

$$8.89 = \beta_{12}\,(10.93) \qquad\qquad 8.89 = \beta_{21}\,(10.55)$$

DESPEJE:

$$\beta_{12} = \frac{8.89}{10.93} = 0.8134; \qquad\qquad \beta_{21} = \frac{8.89}{10.55} = 0.8118$$

OBTENCIÓN DE LA MEDIA GEOMÉTRICA DE LOS COEFICIENTES BETA

$$\sqrt{\beta_{12}\,\beta_{21}} = \sqrt{0.8134 \times 0.8118} = \sqrt{0.6603} = 0.81 = r_{12} = r_{21}$$

* Los valores de las equis, el cálculo de las des y los valores de las medias y las desviaciones se tomaron de Riggleman and Frisbee quienes aplicaron a su ejemplo la fórmula de Pearson-Bravais y no el procedimiento que aquí se sigue, obteniendo 0.82 como índice de correlación.

Cálculo del coeficiente de correlación mediante el cálculo previo de los coeficientes de regresión

1. 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 un cuadro de doble entrada.

2. 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 de aquélla; anótense los productos de estos tres factores en una columna y súmense para obtener el numerador (común) de los coeficientes de regresión.

3. Elévense al cuadrado —separadamente— 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.

4. Divídase el numerador común obtenido en 2º entre los denominadores obtenidos en 3º para obtener los valores de los coeficientes de regresión.

5. 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}).

Cálculo simplificado del índice de correlación

El procedimiento es el siguiente:

1. Consígnense las frecuencias conjuntas (f_{12}) en un cuadro de doble entrada cuyos encabezados columnares serán los valores (o los puntos medios) de la primera variable, y cuyos encabezados de las hileras serán los valores (o los puntos medios) de la segunda variable.

2. Súmense las frecuencias de cada columna para determinar las frecuencias de la primera variable, y las frecuencias de cada renglón para determinar las frecuencias de la segunda variable.

3. Calcúlese el momento 11 con respecto a las medias arbitrarias en unidades de los intervalos de las variables:

   3.1. Eligiendo de entre los puntos medios de cada una de las variables, sendos valores a los que se considerará como medias arbitrarias.

3.2. Colocando:

0 frente a cada una de las medias arbitrarias 1, 2, 3..., hacia arriba o hacia la derecha, +1, +2, +3..., +n hacia abajo o hacia la izquierda,

3.3. Multiplicando cada una de las frecuencias conjuntas (f_{12}) por las desviaciones calculadas en 3.2 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,

3.4. 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}).

4. 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:

4.1. Multiplíquense las desviaciones obtenidas en 3.2 por las frecuencias de la distribución univariada de la primera 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).

4.2. Súmense separadamente los valores de los productos obtenidos para la primera variable y, por otra parte, los productos obtenidos para la segunda.

4.3. Divídanse las dos sumas obtenidas en 4.2 entre la suma total de frecuencias a fin de obtener el primer momento de la primera 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}).

4.4. Multiplíquense los momentos obtenidos en C, y

5. Réstese el producto obtenido anteriormente, del momento 11 con respecto a las medias aritméticas en unidades de los intervalos, obtenidos en 3º, con el fin de obtener el valor del numerador para el índice de correlación ($M'_{11} = M'_{11} - M'_{10} M'_{01}$).

6. Determínense los valores de las desviaciones cuadráticas medias en unidades del intervalo para las dos variables. Para ello:

6.1. Calcúlense los segundos momentos de las respectivas distribuciones univariadas,

6.11. 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,

6.12. sumando separadamente los productos obtenidos en 6.11 para cada una de las variables, y

6.13. 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.

6.2. Calcúlense los cuadrados de los primeros momentos (M'_{10} y M'_{01}) con respecto a las medidas arbitrarias en unidades de los intervalos, elevando al cuadrado los valores obtenidos en 4.3.

6.3. Réstese de cada segundo momento, el cuadrado del primer momento correspondiente: de M'_{20} $(M'_{10})^2$; de M'_{02} $(M'_{01})^2$.

6.4. 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.

7. 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 ($\sigma_1 \sigma_2$).

8. Divídase el valor obtenido en 5º entre el valor obtenido en 7º a fin de obtener el valor del índice de correlación r_{12}.

La forma del índice de correlación resulta ser:

$$r = \frac{M'_{11} - M'_{01} M'_{10}}{N \sigma_{01} \sigma_{10}}$$

EJEMPLO DE CALCULO DEL ÍNDICE DE CORRELACIÓN EN SERIES DE FRECUENCIAS

Correlación entre el número de cuartos y el de ocupantes en la población urbana de Baja California Sur, en 1960

Ocupantes / Cuartos	1	2	3	4	5	6	f_x	d_x	f_xd_x	$f_xd_x^2$
1	181	157	217	226	293	166	1 240	—3	—3 720	11 160
2	52	154	189	215	183	211	1 004	—2	—2 008	4 016
3	25	72	79	107	135	86	504	—1	— 504	504
4	19	27	63	70	67	75	321	0	0	0
5	4	16	17	24	32	32	125	1	+ 125	125
6	3	6	7	13	12	16	57	2	+ 114	228
7	10	4	9	5	18	12	58	3	+ 174	522
f_y	294 +	436 +	581 +	660 +	740 +	598	= 3 309		—6 232	16 555
									+ 413	
d_y	—3	—2	—1	0	1	2			Σf_xd_x = —5 819	
f_yd_y	—882	—872	—581	0	740	1 196 +	—2 335 + 1 936 =		$\boxed{= \Sigma f_yd_y = — 399}$	
$f_yd_y^2$	2 646 +	1 744 +	581 +	0 +	740 +	2 392 =	+ 8 103			

EJEMPLO DE CALCULO DEL INDICE DE CORRELACIÓN EN SERIES DE FRECUENCIAS (CONTINUACIÓN)

Correlación entre el número de cuartos y el de ocupantes en la población urbana de Baja California Sur, en 1960.
(Continuación: cálculo del momento M_{11})

Ocupantes / Cuartos	1	2	3	4	5	6
1	181 X —3 X—3 = +1 629	157 X —3 X—2 = + 942	217 X —3 X—1 = + 651	226 X —3 X 0 = 0	293 X —3 X 1 = — 679	166 X —3 X 2 = — 996
2	52 X —2 X—3 = + 312	154 X —2 X—2 = + 616	189 X —2 X—1 = + 378	215 X —2 X 0 = 0	183 X —2 X 1 = — 366	211 X —2 X 2 = — 844
3	25 X —1 X—3 = + 75	72 X —1 X—2 = + 144	79 X —1 X—1 = + 79	107 X —1 X 0 = 0	135 X —1 X 1 = — 135	86 X —1 X 2 = — 172
4	19 X 0 X—3 = 0	27 X 0 X—2 = 0	63 X 0 X—1 = 0	70 X 0 X 0 = 0	67 X 0 X 1 = 0	75 X 0 X 2 = 0
5	4 X 1 X—3 = — 12	16 X 1 X—2 = — 32	17 X 1 X—1 = — 17	24 X 1 X 0 = 0	32 X 1 X 1 = + 32	32 X 1 X 2 = + 64
6	3 X 2 X—3 = — 18	6 X 2 X—2 = — 24	7 X 2 X—1 = — 14	13 X 2 X 0 = 0	12 X 2 X 1 = + 24	16 X 2 X 2 = + 64
7	10 X 3 X—3 = — 90	4 X 3 X—2 = — 24	9 X 3 X—1 = — 27	5 X 3 X 0 = 0	18 X 3 X 1 = + 554	12 X 3 X 2 = + 72
	+ 2 016 — 120	+ 1 702 — 80	+ 1 108 — 58		— 1 380 + 110	— 2 012 + 200
	+ 1 896 +	+ 1 622	+ 1 050 +	0 +	— 1 270 +	— 1 812
$\Sigma f_{xy} d_x d_y$		4 568		MENOS = 1 486	3 082	

Nota: X representa "por", no es una literal

EJEMPLO DE CALCULO DEL ÍNDICE DE CORRELACIÓN EN SERIES DE FRECUENCIAS

Correlación entre el número de cuartos y el de ocupantes en la población urbana de Baja California Sur, en 1960. (Continuación: Obtención de los momentos para calcular el índice.)

MOMENTO DE ORDEN 11 = $\sum f_{xy}d_xd_y$ dividida entre $\sum f_x$ (o $\sum f_y$) = 1 486 / 3 309 = 0.4491

MOMENTO DE ORDEN 01 = $\sum f_x d_x$ dividida entre $\sum f_x$ = — 5 819/ 3 309 = — 1.7585

MOMENTO DE ORDEN 10 = $\sum f_y d_y$ dividida entre $\sum f_y$ = — 399/ 3 309 = — 0.1206

MOMENTO 11 menos (MOMENTO 01 por MOMENTO 10) = 0.4491 — (— 1.7585) (— 0.1206) =

= 0.4491 — 0.2121 = + 0.2370

SIGMA 01 = RAÍZ DE LA DIFERENCIA DE MOMENTO DE ORDEN 02 menos MOMENTO 01 AL CUADRADO =

$$\sqrt{\left(\frac{16\,555}{3\,309}\right) - (-1.7585)^2} =$$

$$= \sqrt{1.9107}$$

SIGMA 10 = RAÍZ DE LA DIFERENCIA DE MOMENTO DE ORDEN 20 menos MOMENTO 02 AL CUADRADO =

$$\sqrt{\left(\frac{8\,103}{3\,309}\right) - (-0.1206)^2} =$$

$$= \sqrt{2.4343}$$

PRODUCTO DE LAS SIGMAS = $\sqrt{1.9107 \times 2.4343} = \sqrt{4.65121701} = 2.1566$

DIVISIÓN DE LA RESTA DE LOS MOMENTOS ENTRE EL PRODUCTO DE LAS SIGMAS = 0.2370/2.1566 =

= 0.10989 = 10.98%

ÍNDICE DE CORRELACIÓN entre número de cuartos y número de ocupantes en Baja California Sur, en 1960: 10.98%.

2.8. ALGUNAS TÉCNICAS DE LA ESTADÍSTICA DESCRIPTIVA QUE SIRVEN DE PUNTO DE PARTIDA A LA INFERENCIA ESTADÍSTICA

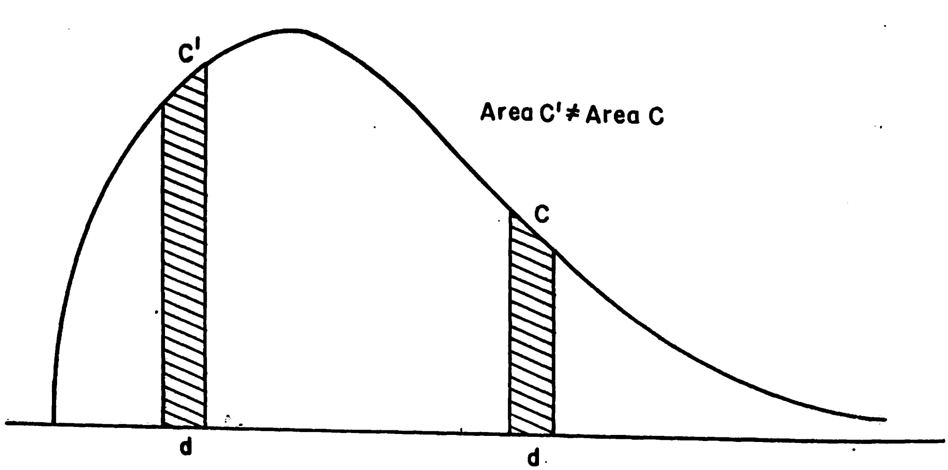
Áreas bajo distintas curvas

Las distribuciones estadísticas difieren unas de otras por características que se reflejan gráficamente. Unas distribuciones (y sus gráficas) son simétricas, otras son asimétricas; unas son picudas, otras son chatas. Según que sean asimétricas o simétricas, chatas o picudas, sus frecuencias se acumularán más o menos en una u otra parte de la curva, de modo que las áreas correspondientes a bases iguales serán —en la mayoría de los casos— distintas.

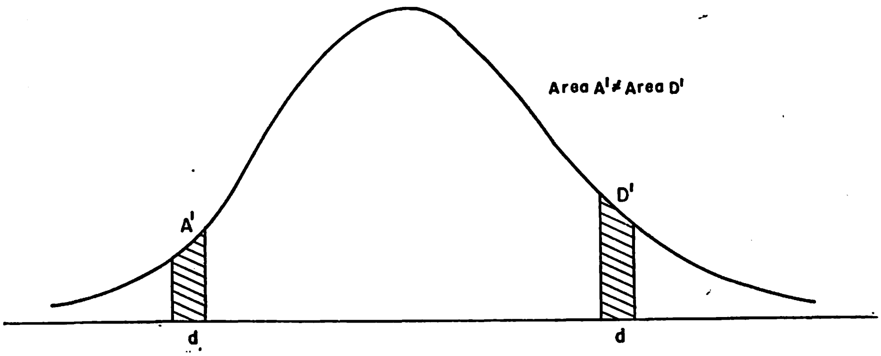
De ese modo: 1) en dos curvas de diferente grado de curtosis, achatamiento o picudez, a un mismo intervalo, situado en la misma zona de la curva, corresponden unas frecuencias o unas áreas que difieren de curva a curva; 2) en una misma curva asimétrica, a un mismo intervalo, corresponden —según su ubicación— frecuencias o áreas distintas, y 3) en una misma curva simétrica, pueden darse dos casos pues: a) si el intervalo es igual, pero se mide en zonas que equidistan del centro, las áreas o frecuencias serán iguales y b) si el intervalo es igual, pero se mide en zonas no equidistantes del centro, las frecuencias o áreas serán diferentes.

En las tres curvas que siguen, se tomó siempre un intervalo d. A pesar de ello: en el primer caso, el área C′ del principio de la curva es distinta del área C del final; en el segundo, aunque la curva es simétrica, por haber tomado primero en un sitio y después en otro el mismo intervalo, son diferentes las áreas respectivas A′ y D′, y, en el tercer caso, si bien las áreas A, B y C difieren entre sí por su diferente colocación a lo largo del eje horizontal las áreas B y D son iguales porque a distancia o intervalo d se tomó —para ellas dos— a igual distancia del eje de simetría de la curva.

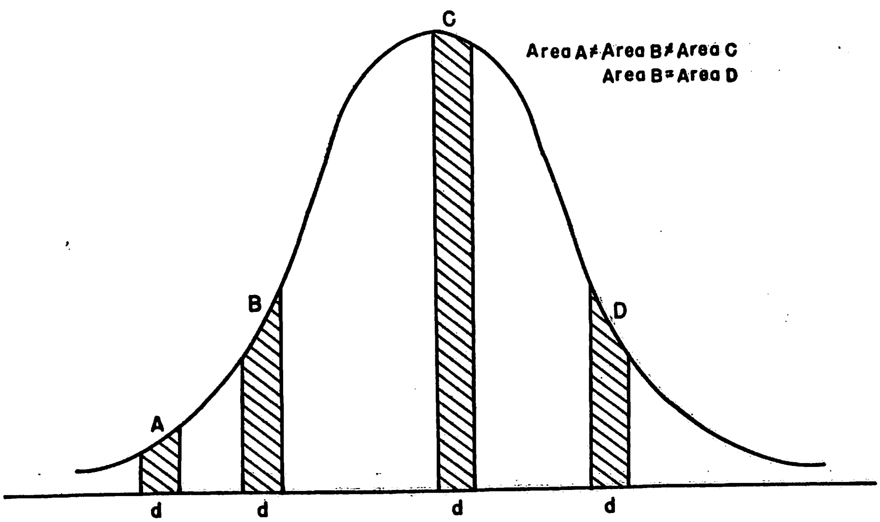
AREAS BAJO DISTINTAS CURVAS



AREAS BAJO DISTINTAS CURVAS



AREAS BAJO DISTINTAS CURVAS

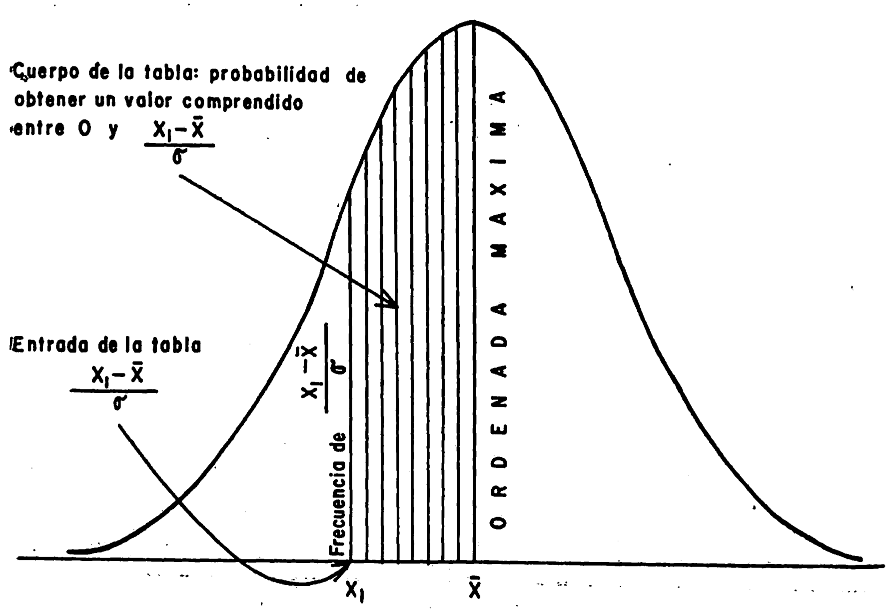


Áreas bajo la curva normal

A una distribución se la representa gráficamente tomando las magnitudes de la variable como abscisas y las frecuencias que le corresponden como ordenadas. En estas condiciones, el área limitada por los ejes coordenados y la curva, representa la suma de todas las frecuencias, o efectivo de la distribución.

El área bajo la curva corresponde, en el aspecto gráfico, al resultado que se obtiene, en el analítico, al integrar la función que asocia las frecuencias de la distribución con los distintos valores de la variable. O sea, que si se integra esa expresión entre más infinito y menos infinito, se obtiene el efectivo, el 100% de las frecuencias o la unidad (que corresponde a la probabilidad total de una distribución estadística).

AREAS BAJO LA CURVA NORMAL



Si en vez de integrar entre menos infinitivo y más infinito, se integra entre menos infinito y cero, se obtendrá sólo una parte de la distribución. En caso de que la distribución (y la curva que la represente) sean simétricas (según es el caso de la normal), se obtendrá la mitad del efectivo, el 50% de las frecuencias o 0.5 de la probabilidad total.

En forma parecida, si en vez de integrar entre menos y más infinito, se integra entre dos magnitudes cualquiera de la variable, se obtendrá un porciento determinado del efectivo o del área total bajo la curva. Sin embargo, aun cuando la distancia entre los límites de integración sean iguales en varios casos, serán distintos los porcientos que se obtengan: *a*) para distintas distribuciones o curvas, y *b*) para distintas porciones incluso de la misma curva, según que los límites se encuentren hacia el centro de ésta o cerca de sus extremos (llamados "colas") en caso de ser simétrica la distribución o curva, y según estén próximos de una u otra cola, en caso de ser asimétricas.

La integral de una distribución no siempre se tiene que hacer directamente. Hay muchas distribuciones típicas (comenzando por la normal) de las que se han tabulado las integrales comprendidas entre un valor suyo importante (la media, en el caso de la distribución o curva normal) y las diferentes magnitudes que puede alcanzar la variable a uno y otro lado de ese valor referencial. En el caso de la curva normal, las integrales correspondientes se encuentran en un cuadro llamado "tabla de áreas bajo la curva normal".

Como sería prácticamente imposible hacer una tabla que diera las áreas bajo la curva, de todas las distribuciones posibles de medidas estadísticas (desde las micrométricas hasta las macrométricas), la entrada a la tabla de áreas bajo la curva está dada no en unidades originarias, sino en unidades de la desviación cuadrática media (o unidades sigmáticas).

O sea, que para determinar qué área bajo la curva (o qué porciento de los casos de la distribución) se encuentra entre la media de la curva normal y un valor determinado de la variable, hay que expresar ese valor como desviación sigmática (o sea, que hay que: 1) restarle la media aritmética, y 2) a la diferencia dividirla entre la desviación cuadrática media, σ).

Frente a cada uno de los valores de la desviación sigmática se encuentra, en la tabla de áreas bajo la curva, un número (un decimal). Ese número indica qué porciento del área total de la curva está comprendido entre el eje de las equis, la ordenada (o perpendicular a ese eje levantada desde el lugar ocupado por la media), la porción de curva correspondiente, y la ordenada levantada en el punto que corresponde a esa desviación sigmática.

El valor dado por la tabla de áreas bajo la curva expresa, así, cuál es la probabilidad que hay de obtener un valor que —en términos sigmáticos— esté comprendido entre la media aritmética y el valor que sirvió de entrada a la tabla.

Así, por ejemplo, si como entrada se toma +1 (una desviación cuadrática media), en la tabla se encuentra 0.3413. Esto significa que hay un 34.13 por ciento de probabilidades de que un valor de la distribución se encuentre comprendido entre cero desviaciones sigmáticas (desviación de la media con respecto a sí misma) y una desviación sigmática.

Como la curva normal es simétrica, también es cierto que hay 34.13 por ciento de probabilidades de que el valor de una distribución normal se aparte de la media entre cero y menos una desviación sigmática.

En forma complementaria, también es posible resolver el problema inverso. Si queremos que un valor de una distribución normal tenga una probabilidad determinada de ocurrir, buscaremos en el cuerpo de la tabla esa probabilidad y, en seguida, leeremos en las entradas a qué valor corresponde. O sea, que para que un valor de la distribución tenga esa probabilidad, se necesitará que el mismo tenga una desviación sigmática comprendida entre 0 y el valor que se encontró a la entrada de la tabla.

Si, una vez hecha esa determinación, se quiere saber a qué valores originarios corresponde esa probabilidad bastará con realizar, en orden inverso, las operaciones inversas a las que se realizaron al principio. O sea, que se necesitará tomar ese valor de la entrada y 1) multiplicarlo por el valor de la desviación cuadrática media, y 2) sumar al producto la media aritmética. La probabilidad buscada se encontrará siempre que el valor esté comprendido entre el valor encontrado a la entrada de la tabla y el valor de la media aritmética.

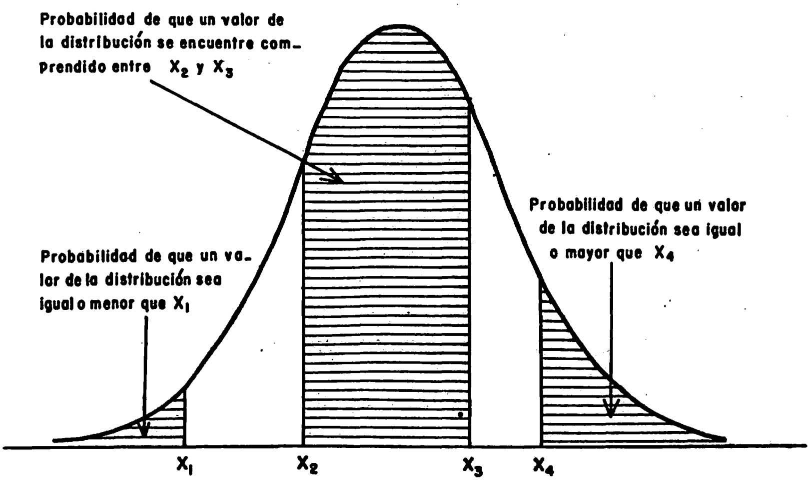
Puede ocurrir, sin embargo, que lo que interese no sea la probabilidad asociada a cierto intervalo central de la distribución, sino la que corresponde a un intervalo extremo. Puede ser que lo que interese sea determinar la probabilidad que hay de que ocurra, en una distribución normal, un valor que no sea mayor que cierto límite (por un extremo de la curva), o que lo que importe sea determinar la probabilidad de que aparezca un valor que no sea inferior a cierto límite (por el otro extremo). En esos casos y otros parecidos, se tiene que recordar que:

La distribución total cubre el 100% del efectivo.
Entre un extremo y la media hay 50% de ese efectivo.
Entre la media y el otro extremo hay otro 50%.
Entre la media y el límite hay p% de casos.

Según esto, entre el límite y el extremo hay [100 - (50 + p)] % de casos o (50-p)% de casos.

Todo lo anterior lo hemos referido a la distribución normal porque ésta es la más fácil de enteder y de manejar en nivel elemental; pero, lo mismo se puede hacer con otras distribuciones cuando se somete a prueba una hipótesis estadística. Esa prueba —en efecto— consiste, fundamentalmente, en determinar si el valor de una medida estadística

AREAS BAJO LA CURVA NORMAL



calculado de una muestra es muy probable o es poco probable que sea la medida estadística correspondiente (el parámetro) de la población total, o si dadas dos medidas estadísticas de dos distribuciones, la diferencia que existe entre ellas es probable que dependa del puro azar y de la forma de selección de las muestras, o es más probable que se deba a un factor significativo, a una fuerza capaz de ser identificada individualmente y no a la acción de todo ese conjunto de fuerzas difíciles de identificar, que se conoce como azar.

Sin embargo, cuando las distribuciones con las que se trabaja no son normales, las tabulaciones de probabilidades o áreas bajo la curva tienen que complicarse. No basta, en tales casos, con una entrada —como en el caso de la normal— sino que se requieren dos o más entradas en cuya intersección se encuentra la probabilidad correspondiente, en el cuerpo de la tabla.

Problemas de obtención de áreas bajo la curva normal

Primero. ¿Cuál es la probabilidad de que un valor de la distribución normal (expresado en unidades sigmáticas o convertido a ellas) sea menor o igual a cierto valor fijo x_1, si x_1 es menor que la media aritmética?

Solución: 1) búsquese en la tabla de áreas bajo la curva, la probabilidad de que el valor se encuentre entre x_1 y la media, y 2) réstese esa probabilidad de 0.5 (o 50%), probabilidad de que el valor sea inferior a la media.

Mutatis mutandis, el problema y la solución se aplican en tratándose de la probabilidad de un valor mayor a cierto valor fijo x_3, mayor que la media.

Segundo. ¿Cuál es la probabilidad de que un valor de la distribución sea mayor que un valor fijo x_1 y menor que otro x_2, ambos inferiores a la media aritmética?

Solución: 1) búsquese la probabilidad asociada a x_1, que suponemos el más alejado de la media; 2) búsquese la probabilidad que se asocia con x_2 el valor más próximo de la media, y 3) réstese de la primera probabilidad la segunda.

En forma parecida se plantea y resuelve el problema de la probabilidad asociada a un intervalo limitado por dos valores x_3 y x_4 superiores a la media.

Tercero. ¿Cuál es la probabilidad de que un valor de la distribución normal sea mayor que un valor fijo x_1, y menor que la media, si x_1 es inferior a la media?

Solución: 1) búsquese direcatmente esa probabilidad en la tabla de áreas bajo la curva, frente al valor de x_1.

En forma parecida se plantea y resuelve el problema de encontrar la probabilidad de que un valor de la distribución normal sea menor que un valor fijo x_3 y mayor que la media, si x_3 es superior a la media.

Cuarto. ¿Cuál es la probabilidad de que ocurra en una distribución normal un valor que sea mayor que un valor fijo x_1 (menor que la media) y que sea menor que otro valor fijo x_3 (mayor que la media)?

Solución: 1) búsquese la probabilidad asociada a x_1; 2) búsquese la que se asocia con x_3, y 3) súmense las probabilidades obtenidas.

Quinto. ¿Cuál es la probabilidad de que ocurra un valor de la distribución normal que sea menor que un valor fijo x_3 (mayor que la media)?

Solución: 1) búsquese la probabilidad asociada a x_3, y 2) súmesele 0.5 (o 50%) probabilidad de que ocurran los valores inferiores a la media.

De modo análogo se plantea y resuelve el problema de la probabilidad asociada a un valor mayor que un valor fijo x_1 (inferior a la media).

Sexto. ¿Cuál es la probabilidad de que un valor de la distribución normal sea mayor que un valor fijo x_3 (mayor que la media)?

Solución: 1) búsquese la probabilidad asociada a x_3, y 2) réstesela de 1 (o 100%) probabilidad total tanto de los valores inferiores como de los iguales o de los superiores a x_3.

Planteamiento y solución son análogos si el problema consiste en determinar la probabilidad de que el valor de una distribución normal sea menor que un valor fijo x_1, menor que la media aritmética.

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: 1) 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; 2) 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; 3) 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 (a 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: a) 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 b) de los límites superiores de las clases posteriores a las 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:

1. Calcular las desviaciones con respecto a la media aritmética en términos de la desviación cuadrática media:
   1.1. 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
   1.2. Dividiendo el resultado de dicha resta (desviación con respecto a la media aritmética) entre la desviación cuadrática media de la distribución.
2. Determinar la porción de área (% de frecuencia) comprendido entre las desviaciones con respecto a la media en unidades sigmáticas y la media.

2.1. Buscando en la entrada de la tabla de áreas bajo la curva normal el valor de las desviaciones,

2.2. Encontrando en el cuerpo de la tabla, los valores de área correspondientes.

3. Determinar la porción de área (% de frecuencias) correspondiente a cada valor (a cada clase).

 3.1. 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.

 3.2. 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,

 3.3. 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.

4. Cálculo de las frecuencias teóricas de la distribución no unitaria:

 4.1. Multiplicando los valores obtenidos (% de frecuencias) por el efectivo (o suma total de frecuencias de la distribución empírica que se estudia).

Importancia de las áreas

Las áreas de las curvas estadísticas son importantes porque la inferencia estadística se basa en ellas.

Para inferir, básicamente:

1º Se determina cuál es el tipo de curva que revela cómo se distribuye una medida estadística en el conjunto de muestras *posibles* de una población.

2º Se determina cuál es, en esa curva, el área o probabilidad que corresponde a la medida obtenida de la muestra *real.*

3º Se decide si esa área o probabilidad es suficiente para permitir que se considere a la medida muestral como representativa de la medida poblacional, o si es insuficiente para justificar esta inferencia.

INTERPOLACIÓN DE UNA CURVA NORMAL POR EL MÉTODO DE LAS ÁREAS

Distribución por edades de los pobladores de un caserío

Clases de edad		f_i	$l_i - x_a$	$L_i - x_a$	$l_i - x_a$	$L_i - x_a$	Áreas entre la x_a y l	Áreas entre la x_a y L	*Áreas en cada intervalo*	*Áreas de distribución normal unitaria* $\times \Sigma f_i$
0	10 —	10	—49.07		—2.53		.4943		.0160	20.8
10	20 —	45	—39.07		—2.02		.4783		.0451	58.6
20	30 —	125	—29.07		—1.50		.4332		.0967	125.7
30	40 —	210	—19.07		—0.98		.3365		.1557	202.4
40	50 —	280	— 9.07	.93	—0.47	0.05	.1808	.0199	.2007	260.9
50	60 —	295		10.93		0.56		.2123	.1924	250.1
60	70 —	185		20.93		1.08		.3599	.1476	191.9
70	80 —	105		30.93		1.60		.4452	.0853	110.9
80	90 —	40		40.93		2.11		.4826	.0374	48.6
90	100 —	5		50.93		2.63		.4957	.0131	17.0

Valores calculados previamente:

$x_a = 49.07$

$\sigma = 19.37$

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 cuáles excedencias con respecto a lo que debería esperarse que ocurriera normalmente.

APÉNDICE

Diferentes tipos de desviación y de momento, y sus relaciones

I. Clasificación según centro de referencia		DESVIACIONES — II. Clasificación según escala o unidades: 1. Unidades originarias	2. Unidades del intervalo	3. Unidades sigmáticas
1. Con respecto al origen	Símbolo	d_o	D_o	δ_o
	Fórmula	$= x_i - 0$	$= \dfrac{x_i - 0}{i}$	$= \dfrac{x_i - 0}{\sigma}$
2. Con respecto a una media arbitraria m′ o x′	Símbolo	d'_i	D'_i	δ'_i
	Fórmula	$= x_i - x'$	$= \dfrac{x_i - x'}{i}$	$= \dfrac{x_i - x'}{\sigma}$
3. Con respecto a la media primera o aritmética $\overline{x_1}$	Símbolo	d_i	D_i	δ_i
	Fórmula	$= x_i - \overline{x_1}$	$= \dfrac{x_i - \overline{x_1}}{i}$	$= \dfrac{x_i - \overline{x_1}}{\sigma}$

I.—Clasificación según centro de momentos		MOMENTOS — II.—Clasificación según escala o según unidades: 1.—En unidades originarias (1os. 2os. 3os. 4os.)	2.—En unidades de intervalo (1os. 2os. 3os. 4os.)	3.—En unidades sigmáticas (1os. 2os. 3os. 4os.)
1.—Respecto al origen	Símbolo	$\nu_1 \quad \nu_2 \quad \nu_3 \quad \nu_4$	$\nu_{1i} \quad \nu_{2i} \quad \nu_{3i} \quad \nu_{4i}$	$\nu_{1\sigma} \quad \nu_{2\sigma} \quad \nu_{3\sigma} \quad {}_{4\sigma}$
	Fórmulas	$\nu_r = \dfrac{\Sigma d^r_o f_i}{\Sigma f_i}$	$\nu_{ri} = \dfrac{\Sigma D^r_o f_i}{\Sigma f_i}$	${}_{r\sigma} = \dfrac{\Sigma \delta^r_i f_i}{\Sigma f_i}$
2.—Respecto a una media arbitraria	Símbolo	$\mu'_1 \quad \mu'_2 \quad \mu'_3 \quad \mu'$	$\mu_{1i} \quad \mu_{2i} \quad \mu_{3i} \quad \mu'_{4i}$	$\mu_{1\sigma} \quad \mu'_{2\sigma} \quad \mu_{3\sigma} \quad \mu'_{4\sigma}$
	Fórmulas	$\mu'_r = \dfrac{\Sigma d'^r_i f_i}{\Sigma f_i}$	${}'_{ri} = \dfrac{\Sigma D'^r_i f_i}{\Sigma f_i}$	$\mu'_{r\sigma} = \dfrac{\Sigma \delta'^r_i f_i}{\Sigma f_i}$
3.—Respecto a la media primera	Símbolo	$\mu_1 \quad \mu_2 \quad \mu_3 \quad \mu_4$	$\mu_{1i} \quad \mu_{2i} \quad \mu_{3i} \quad \mu_{4i}$	$\mu_{1\sigma} \quad \mu_{2\sigma} \quad \mu_{3\sigma} \quad \mu_{4\sigma}$
	Fórmulas	$\mu_r = \dfrac{\Sigma d^r_i f_i}{\Sigma f_i}$	$\mu_{ri} = \dfrac{\Sigma D^r_i f_i}{\Sigma f_i}$	$\mu_{r\sigma} = \dfrac{\Sigma \delta^r_o f_i}{\Sigma f_i}$

1. Relación entre momentos en unidades originarias y momentos en unidades del intervalo.

Momento de orden r en unidades originarias igual a:
Momento de orden r en unidades del intervalo por
la potencia r del intervalo

$$\mu_r = \mu'_{r1}\,\sigma^r$$

2. Relación entre momentos en unidades originarias y momentos en unidades sigmáticas.

Momento de orden r en unidades originarias igual a:
Momento de orden r en unidades sigmáticas por
la potencia r de sigma

$$\mu_r = \mu'_{r1}\, i^r$$

3. Relación entre momento con respecto a una medida arbitraria y momentos con respecto a la media aritmética:

Equivalencia del primer momento respecto a la media aritmética:

$$\mu_1 = 0$$

Equivalencia del segundo momento:

$$\mu_2 = \mu'_2 - \mu'^{\,2}_1$$

Equivalencia del tercer momento:

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

Equivalencia del cuarto momento:

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

3. TABLAS DE USO INDISPENSABLE EN ESTADÍSTICA ELEMENTAL

TABLAS DE LA CURVA NORMAL

Variable como desviación sigmática	*Ordenada*	*Tercera derivada*	*Cuarta derivada*	*Áreas bajo la curva*
0.0	0.3989	0.0000	1.1968	0.0000
0.1	0.3970	0.1187	1.1671	0.0398
0.2	0.3910	0.2315	1.0799	0.0793
0.3	0.3814	0.3330	0.9413	0.1179
0.4	0.3683	0.4184	0.7607	0.1554
0.5	0.3521	0.4841	0.5501	0.1915
0.6	0.3332	0.5278	0.3231	0.2258
0.7	0.3123	0.5486	0.0937	0.2580
0.8	0.2897	0.5469	— 0.1247	0.2881
0.9	0.2661	0.5245	— 0.3203	0.3159
1.0	0.2420	0.4839	— 0.4839	0.3413
1.1	0.2179	0.4290	— 0.6091	0.3643
1.2	0.1942	0.3635	— 0.6926	0.3849
1.3	0.1714	0.2918	— 0.7341	0.4032
1.4	0.1497	0.2180	— 0.7364	0.4192
1.5	0.1295	0.1457	— 0.7043	0.4332
1.6	0.1109	0.0781	— 0.6441	0.4452
1.7	0.0941	0.0176	— 0.5632	0.4554
1.8	0.0790	— 0.0341	— 0.4692	0.4641
1.9	0.0656	— 0.0761	— 0.3693	0.4713
2.0	0.0540	— 0.1080	— 0.2700	0.4773
2.1	0.0440	— 0.1302	— 0.1765	0.4821
2.2	0.0355	— 0.1436	— 0.0927	0.4861
2.3	0.0283	— 0.1492	— 0.0214	0.4893
2.4	0.0224	— 0.1483	0.0362	0.4918
2.5	0.0175	— 0.1424	0.0800	0.4938
2.6	0.0136	— 0.1328	0.1105	0.4953
2.7	0.0104	— 0.1207	0.1293	0.4965
2.8	0.0079	— 0.1073	0.1379	0.4974
2.9	0.0060	— 0.0934	0.1385	0.4981
3.0	0.0044	— 0.0798	0.1330	0.4987
3.1	0.0033	— 0.0669	0.1231	0.4990
3.2	0.0024	— 0.0552	0.1107	0.4993
3.3	0.0017	— 0.0449	0.0969	0.4995
3.4	0.0012	— 0.0359	0.0829	0.4997
3.5	0.0009	— 0.0283	0.0694	0.4998
3.6	0.0006	— 0.0219	0.0570	0.4998
3.7	0.0004	— 0.0168	0.0460	0.4999
3.8	0.0003	— 0.0127	0.0365	0.4999
3.9	0.0002	— 0.0095	0.0284	0.5000

TABLAS DE LA FUNCIÓN GAMMA

n	Gamma (n)	Gamma ($n+1$)	n	Gamma (n)	Gamma ($n+1$)
1.00	1.00000	1.00000	1.40	.88726	1.24216
1.01	.99433	1.00427	1.41	.88676	1.25033
1.02	.98884	1.00862	1.42	.88636	1.25863
1.03	.98335	1.01285	1.43	.88604	1.26704
1.04	.97844	1.01758	1.44	.88580	1.27555
1.05	.97350	1.02217	1.45	.88565	1.28419
1.06	.96874	1.02686	1.46	.88560	1.29298
1.07	.96415	1.03164	1.47	.88563	1.30188
1.08	.95973	1.03651	1.48	.88575	1.31091
1.09	.95546	1.04145	1.49	.88595	1.32007
1.10	.95135	1.04685	1.50	.88623	1.32935
1.11	.94739	1.05160	1.51	.88659	1.33875
1.12	.94359	1.05682	1.52	.88704	1.34830
1.13	.93993	1.06212	1.53	.88757	1.35798
1.14	.93642	1.06752	1.54	.88818	1.36780
1.15	.93304	1.07300	1.55	.88887	1.37777
1.16	.92980	1.07857	1.56	.88964	1.38784
1.17	.92670	1.08424	1.57	.89049	1.39807
1.18	.92373	1.09000	1.58	.89142	1.40844
1.19	.92088	1.09585	1.59	.89243	1.41896
1.20	.91817	1.10180	1.60	.89352	1.42963
1.21	.91558	1.10785	1.61	.89468	1.44043
1.22	.91311	1.11399	1.62	.89592	1.45139
1.23	.91075	1.12022	1.63	.89724	1.46250
1.24	.90852	1.12656	1.64	.89864	1.47377
1.25	.90640	1.13300	1.65	.90012	1.48520
1.26	.90440	1.13954	1.66	.90167	1.49677
1.27	.90250	1.14618	1.67	.90330	1.50851
1.28	.90072	1.15292	1.68	.90500	1.52040
1.29	.89904	1.15976	1.69	.90678	1.53246
1.30	.89747	1.16671	1.70	.90864	1.54469
1.31	.89600	1.17376	1.71	.91057	1.55707
1.32	.89464	1.18092	1.72	.91258	1.56964
1.33	.89338	1.18819	1.73	.91466	1.56236
1.34	.89222	1.19557	1.74	.91683	1.59528
1.35	.89115	1.20305	1.75	.91906	1.60835
1.36	.89018	1.21060	1.76	.92137	1.62161
1.37	.88931	1.21835	1.77	.92376	1.63505
1.38	.88854	1.22619	1.78	.92623	1.64869
1.39	.88785	1.23411	1.79	.92877	1.66250

TABLAS DE LA FUNCIÓN GAMMA

n	*Gamma* (n)	*Gamma* $(n+1)$	n	*Gamma* (n)	*Gamma* $(n+1)$
1.80	.93138	1.67648	1.90	.96177	1.82736
1.81	.93408	1.69068	1.91	.96523	1.84359
1.82	.93685	1.70507	1.92	.96878	1.86006
1.83	.93969	1.71963	1.93	.97240	1.87673
1.84	.94261	1.73440	1.94	.97610	1.89363
1.85	.94561	1.74938	1.95	.97988	1.91077
1.86	.94869	1.76456	1.96	.98374	1.92813
1.87	.95184	1.77994	1.97	.98768	1.94573
1.88	.95507	1.79553	1.98	.99171	1.96359
1.89	.95838	1.81134	1.99	.99581	1.98166

BIBLIOGRAFÍA

ACKOFF, Rusell I., *The Design of Social Research.* The University of Chicago Press, Chicago, 1953.

ALLEN, R. G. D., *Statistics for Economists.* Huntchinson University Library, London, first published, 1949, eleventh impression, 1961.

ARKIN, Herbert and RAYMOND R. Colton, *Statistical Methods.* College Outline Series, Barnes & Noble, Inc. New York, fourth edition, 1939, revised, 1950, 1955, thirteenth printing, 1955.

BLALOCK, Hubert M., *Estadística Social.* Primera edición en inglés, 1960, traducción de Carlos Gerhard, revisada por Ana Ma. Flores; primera edición en español, Fondo de Cultura Económica, México-Buenos Aires, 1966.

DAVID, F. N., A *Statistical Primer.* Charles Griffin & Company Limited, London, 1953.

FREEMAN, Linton C., *Elementary Applied Statistics: for Students in Behavioral Science.* John Wiley & Sons, Inc. New York, London, Sydney, 1965.

GARCÍA PÉREZ, Andrés, *Elementos de método estadístico.* Cuarta edición, Universidad Nacional Autónoma de México, México, 1966 (en el momento de escribirse estas líneas estaba en prensa la quinta edición).

GLEASON GALICIA, Rubén, *Las estadísticas y censos de México: su organización y estado actual.* Instituto de Investigaciones Sociales, México, 1968.

HOLGUÍN QUIÑONES, Fernando, *Estadística descriptiva.* Universidad Nacional Autónoma de México, México, 1970.

HODGMAN, Charles D., *Mathematical Tables from handbook of chemistry and physics.* Tenth edition, Chemical Rubber Publishing Co. Cleveland, Ohio, 1954.

KENDALL, Maurice G. and STUART, Alan, *The Advanced Theory of Statistics*. Three-volume edition, vol. 1: Distribution Theory; first published, 1958, vol. 2: Inference and Relationship; first published, 1961, vol. 3: Design and Analysis, and Time-Series; first published, 1966; Charles Griffin & Company Limited, London.

LUNDBERG, George A., *Técnica de la investigación social*. Traducción de José Miranda, Fondo de Cultura Económica, 1ª edición, México, 1949.

McCORMICK, Thomas Carson, *Técnica de la estadística social*. Primera edición en inglés, 1941. Traducción de Francisco Rived, revisada por Luis Alaminos; primera edición en español. Fondo de Cultura Económica, México-Buenos Aires, 1954.

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, 1950.

MOSER, C. A., *Survey Methods in Social Investigation*. William Heinemann Ltd., Melbourne, London, Toronto, first published, 1958, reprinted, 1959.

Naciones Unidas (Organización de las), *Informe sobre la definición y medición internacional del nivel de vida*. Nueva York, 1954.

QUENOUILLE, M. H., *The Fundamentals of Statistical Reasoning*, being number three of Griffin's Statistical Monographs & Courses edited by M. G. Kendall, Charles Griffin & Company Limited, London, 1965.

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*. McGraw-Hill Book Company, Inc. New York, Toronto, London, third edition, 1951.

STUART, Alan, *Basic Ideas of Scientific Sampling*; being number four of Griffin's statistical Monographs & Courses edited by M. G. Kendall, Charles Griffin & Company Limited, London, 1964.

United Nations, *Population Census Methods*. Population studies, Nº 4, Lake Success, New York, November 1949.

YATES, Frank, *Sampling Methods for Censuses and Surveys*. Third edition, second impression, Charles Griffin & Company Limited, London, 1965.

URIBE VILLEGAS, Óscar, *Curvas sociográficas*, Fundamento matemático y técnica de aplicación. Instituto de Investigaciones Sociales, México, 1969.

YOUNG, Pauline V., *Métodos científicos de investigación social*. Traducción por Ángela Müller Montiel, Instituto de Investigaciones Sociales de la Universidad Nacional Autónoma de México, México, 1960.

YULE, G. Undy and M. G. Kendall, *Introducción a la estadística matemática*. Traducción de la 13ª edición inglesa y prólogo por José Ros Jimeno, M. Aguilar, editor, Madrid, 1947.

EJEMPLOS

GRÁFICAS

ÍNDICE POR MATERIA

ÍNDICE

En la Imprenta Universitaria, bajo la dirección de Jorge Gurría Lacroix, se terminó la impresión de *Los elementos de la estadística social*, el día 15 de septiembre de 1971. La composición se paró en tipos Electra 11:12, 10:11, 9:10 y 8:9. Se tiraron 2 000 ejemplares.

IIS