Schaum

ESTADÍSTICA

SEGUNDA EDICIÓN

Murray R. Spiegel

- 975 problemas resueltos con soluciones completamente detalladas
- Más de 700 problemas suplementarios con solución
- Especial énfasis en la comprensión de los métodos de resolución de problemas prácticos
- Abarca los aspectos teóricos esenciales de la estadística

McGraw Hill

Utilizado por millones de estudiantes y recomendado por profesores de todo el mundo

ESTADISTICA

Segunda edición

ESTADISTICA

Segunda edición

MURRAY R. SPIEGEL
Hartford Graduate Center

Traducción
RAFAEL HERNANDEZ HEREDERO
Dpto. de Métodos Matemáticos de la Física
Universidad Complutense de Madrid

Revisión Técnica
LORENZO ABELLANAS RAPUN
Catedrático de Métodos Matemáticos de la Física
Universidad Complutense de Madrid

McGraw Hill

MADRID • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA
MEXICO • NUEVA YORK • PANAMA • SAN JUAN • SANTAFE DE BOGOTA • SANTIAGO • SAO PAULO
AUCKLAND • HAMBURGO • LONDRES • MILAN • MONTREAL • NUEVA DELHI
PARIS • SAN FRANCISCO • SIDNEY • SINGAPUR • ST. LOUIS • TOKIO • TORONTO

ESTADISTICA (Segunda edición)

Edificio Valrealty, 1.ª planta
Basauri, 17
28023 Aravaca (Madrid)

Traducido de la segunda edición en inglés de **STATISTICS**

ISBN: 0-07-060234-4
ISBN: 84-7615-562-X
Depósito legal: M. 522-1997

Fotocompuesto en MonoComp, S. A.

IMPRESO POR INDUSTRIAS GRAFICAS 3f S.A.

IMPRESO EN CHILE - PRINTED IN CHILE

Contenido

Prólogo

La Estadística o los métodos estadísticos, como se denomina a veces, está jugando un papel más y más importante en casi todas las facetas del comportamiento humano. Ocupada inicialmente en asuntos de Estado, y de ahí su nombre, la influencia de la Estadística se ha extendido ahora a la agricultura, biología, negocios, química, comunicaciones, economía, educación, electrónica, medicina, física, ciencias políticas, psicología, sociología y otros muchos campos de la ciencia y la ingeniería.

El propósito de este libro es presentar una introducción a los principios básicos de la Estadística, que serán de utilidad con independencia del campo de interés específico del lector. Se ha diseñado para ser usado como suplemento a un texto estándar o como libro de texto para un curso formal de Estadística. Será de considerable interés, asimismo, como libro de consulta, para todos aquellos que estén implicados en aplicar la Estadística a sus propios problemas de investigación.

Cada capítulo comienza con enunciados claros de las definiciones pertinentes, teoremas y principios, junto con otro material ilustrativo y descriptivo. Ello viene seguido de problemas resueltos y suplementarios que en muchos casos utilizan datos obtenidos en situaciones estadísticas reales. Los problemas resueltos sirven para ilustrar y ampliar la teoría, arrojan luz sobre los puntos sutiles, sin lo cual el estudiante se sentiría siempre sobre arenas movedizas, y proporcionan la oportunidad de repetir los principios básicos, vital para un aprendizaje eficaz. Numerosas demostraciones de fórmulas han quedado incluidas entre los problemas resueltos. El elevado número de problemas suplementarios con solución, completa la revisión del material expuesto en cada capítulo.

La única base matemática requerida para la comprensión del libro consiste en aritmética y rudimentos de álgebra. En el primer capítulo se presenta un repaso de los conceptos matemáticos usados posteriormente. Puede leerse al comienzo o guardarlo como referencia para cuando sea preciso.

La primera parte del libro trata el análisis de las distribuciones de frecuencia y las medidas asociadas de tendencia central, dispersión, sesgo (asimetría) y curtosis (aplastamiento). Lo cual conduce naturalmente a una discusión de teoría elemental de probabilidades y sus aplicaciones, que allana el camino para la teoría del muestreo. Se consideran en primer lugar las técnicas de grandes muestras, que involucran a la distribución normal, y aplicaciones a la estimación estadística y al contraste de hipótesis y significación. La teoría de pequeñas muestras, que emplea la distribución t de Student, la ji-cuadrado y la distribución F, aparece en un capítulo posterior, junto con sus aplicaciones. Otro capítulo sobre ajuste de curvas y el método de mínimos cuadrados lleva lógicamente a los temas de correlación y regresión en dos variables. La correlación parcial y múltiple, en más de dos variables, se estudia en un capítulo aparte. Luego siguen capítulos sobre el análisis de varianza y los métodos no paramétricos, nuevos en esta segunda edición. Dos capítulos finales tratan el análisis de series en el tiempo y los números índice, respectivamente.

Hemos incluido más material del que puede cubrirse en un curso habitual, con el fin de hacer el libro más flexible, ampliarlo y mejorarlo como libro de consulta y estimular el interés por otros temas. Al usar el libro es posible alterar el orden de muchos capítulos e incluso omitir algunos. Así, los Capítulos 13-15 y 18-19 en su casi totalidad pueden introducirse tras el Capítulo 5, si se desea estudiar correlación, regresión, series en el tiempo y números índice antes que la teoría de muestreo. Análogamente, el Capítulo 6 puede omitirse casi completo si no se quiere perder mucho tiempo en las probabilidades. En un primer curso, todo el Capítulo 15 pude ser omitido. Hemos elegido el orden que aparece porque existe la tendencia creciente, en los cursos modernos, de introducir la teoría del muestreo y la inferencia estadística lo antes posible.

Deseo agradecer a las diversas instituciones, tanto gubernamentales como privadas, por su cooperación al proporcionarme datos para las tablas. En el texto figuran las referencias oportunas a las fuentes consultadas. En particular, estoy agradecido al profesor Sir Ronald A. Fisher, F. R. S., Cambridge; doctor Frank Yates, F. R. S., Rothamsted, y Messrs. Oliver y Boyd Ltd., Edinburgh, por conceder autorización para utilizar los datos de la Tabla III de su libro *Statistical Tables for Biological, Agricultural, and Medical Research*. Quiero dar las gracias, asimismo, a Esther y Meyer Scher por su apoyo y al personal de McGraw-Hill por su colaboración.

MURRAY R. SPIEGEL

CAPITULO 1

Variables y gráficos

ESTADISTICA

La Estadística estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis.

En un sentido menos amplio, el término *estadística* se usa para denotar los propios datos, o números derivados de ellos, tales como los promedios. Así se habla de estadística de empleo, estadística de accidentes, etc.

POBLACION Y MUESTREO; ESTADISTICA INDUCTIVA Y DESCRIPTIVA

Al recoger datos relativos a las características de un grupo de individuos u objetos, sean alturas y pesos de estudiantes de una universidad o tuercas defectuosas producidas en una fábrica, suele ser imposible o nada práctico observar todo el grupo, en especial si es muy grande. En vez de examinar el grupo entero, llamado *población* o *universo*, se examina una pequeña parte del grupo, llamada *muestra*.

Una población puede ser *finita* o *infinita*. Por ejemplo, la población consistente en todas las tuercas producidas por una fábrica un cierto día es finita, mientras que la determinada por todos los posibles resultados (caras, cruces) de sucesivas tiradas de una moneda, es infinita.

Si una muestra es representativa de una población, es posible inferir importantes conclusiones sobre la población a partir del análisis de la muestra. La fase de la Estadística que trata con las condiciones bajo las cuales tal diferencia es válida se llama *estadística inductiva* o *inferencia estadística*. Ya que dicha inferencia no es del todo exacta, el lenguaje de las *probabilidades* aparecerá al establecer nuestras conclusiones.

La parte de la Estadística que sólo se ocupa de describir y analizar un grupo dado, sin sacar conclusiones sobre un grupo mayor, se llama *estadística descriptiva* o *deductiva*.

Antes de entrar en el estudio de la Estadística, recordemos algunas nociones matemáticas relevantes.

VARIABLES: DISCRETAS Y CONTINUAS

Una *variable* es un símbolo, tal como X, Y, H, x o B, que puede tomar un conjunto prefijado de valores, llamado *dominio* de esa variable. Si la variable puede tomar un solo valor, se llama *constante*.

Una variable que puede tomar cualquier valor entre dos valores dados se dice que es una *variable continua*; en caso contrario diremos que la *variable es discreta.*

EJEMPLO 1. El número N de hijos en una familia puede ser 0, 1, 2, 3, ... pero no 2.5 ó 3.842. Es una variable discreta.

EJEMPLO 2. La altura H de una persona, que puede ser 62 pulgadas (abreviatura «in»), 63.8 in o 65.8341 in, dependiendo de la precisión de la medida, es una variable continua.

Los datos que admiten descripción mediante una variable discreta o continua se denominan respectivamente *datos discretos* y *continuos*. El número de hijos en cada una de 1000 familias es un ejemplo de datos discretos, mientras que las alturas de 100 universitarios lo es de datos continuos. En general, las *mediciones* dan lugar a datos continuos, y las *enumeraciones o recuentos*, a datos discretos.

A veces conviene extender la noción de variable a entidades no numéricas; por ejemplo, el color C en un arco iris es una variable que puede tomar los «valores» rojo, anaranjado, amarillo, verde, azul, añil y violeta. Suele ser posible sustituir tales variables por entidades numéricas; por ejemplo, denotando el rojo como 1, el anaranjado como 2, etc.

REDONDEO DE DATOS

El resultado de redondear un número como 72.8 en unidades es 73, pues 72.8 está más próximo de 73 que de 72. Análogamente, 72.8146 se redondea en centésimas (o sea con dos decimales) a 72.81, porque 72.8146 está más cerca de 72.81 que de 72.82.

Al redondear 72.465 en centésimas nos hallamos ante un dilema, ya que está *equidistante* de 72.46 y de 72.47. Se adopta en tales casos la costumbre de redondear al *entero par* que preceda al 5. Así pues, 72.465 se redondea a 72.46, 183.575 se redondea a 183.58 y 116,500,000 se redondea en millones a 116,000,000. Esta estrategia es particularmente útil para minimizar los *errores de redondeo acumulados* cuando se efectúa un gran número de operaciones (véase Prob. 1.4).

NOTACION CIENTIFICA

Al escribir números, especialmente los que tienen muchos ceros antes o después del punto decimal, interesa emplear la notación científica mediante potencias de 10.

EJEMPLO 3. $10^1 = 10$, $10^2 = 10 \times 10 = 100$, $10^5 = 10 \times 10 \times 10 \times 10 \times 10 = 100{,}000$ y $10^8 = 100{,}000{,}000$.

EJEMPLO 4. $10^0 = 1$; $10^{-1} = .1$, o sea 0.1; $10^{-2} = .01$, o sea 0.01, y $10^{-5} = .00001$, o sea 0.00001.

EJEMPLO 5. $864{,}000{,}000 = 8.64 \times 10^8$, y $0.00003416 = 3.416 \times 10^{-5}$.

Nótese que al multiplicar un número por 10^8, por ejemplo, el punto decimal se mueve ocho posiciones *a la derecha*, y al multiplicar por 10^{-6} se mueve seis posiciones *a la izquierda*.

A menudo escribiremos 0.1253 en vez de .1253 para recalcar el hecho de que no se ha omitido accidentalmente un entero no nulo delante del punto decimal. Sin embargo, ese cero puede omitirse cuando no exista riesgo de confusión, por ejemplo, en tablas.

Con frecuencia usamos paréntesis o puntos para denotar el producto de dos o más números. Así pues, $(5)(3) = 5 \cdot 3 = 5 \times 3 = 15$, y $(10)(10)(10) = 10 \cdot 10 \cdot 10 = 10 \times 10 \times 10 = 1000$. Si se usan letras para representar los números, se suelen omitir los paréntesis y los puntos; por ejemplo, $ab = (a)(b) = a \cdot b = a \times b$.

La notación científica resulta útil en el cálculo, sobre todo para localizar puntos decimales. Se utilizan entonces las reglas

$$(10^p)(10^q) = 10^{p+q} \qquad \frac{10^p}{10^q} = 10^{p-q}$$

donde p y q son números arbitrarios.

En 10^p, p se llama *exponente* y 10 *base*.

EJEMPLO 6. $(10^3)(10^2) = 1000 \times 100 = 100{,}000 = 10^5$ es decir, 10^{3+2}

$$\frac{10^6}{10^4} = \frac{1{,}000{,}000}{10{,}000} = 100 = 10^2 \quad \text{es decir, } 10^{6-4}$$

EJEMPLO 7. $(4{,}000{,}000)(0.0000000002) = (4 \times 10^6)(2 \times 10^{-10}) = (4)(2)(10^6)(10^{-10}) = 8 \times 10^{6-10}$
$= 8 \times 10^{-4} = 0.0008$

EJEMPLO 8. $\dfrac{(0.006)(80{,}000)}{0.04} = \dfrac{(6 \times 10^{-3})(8 \times 10^4)}{4 \times 10^{-2}} = \dfrac{48 \times 10^1}{4 \times 10^{-2}} = \left(\dfrac{48}{4}\right) \times 10^{1-(-2)}$
$= 12 \times 10^3 = 12{,}000$

DIGITOS SIGNIFICATIVOS

Si una altura se anota con la mejor precisión posible como 65.4 in, eso significa que está entre 65.35 y 65.45. Los dígitos empleados, aparte de los ceros necesarios para localizar el punto decimal, se llaman *dígitos significativos* o *cifras significativas*, del número.

EJEMPLO 9. 65.4 tiene tres cifras significativas.

EJEMPLO 10. 4.5300 tiene cinco cifras significativas.

EJEMPLO 11. $.0018 = 0.0018 = 1.8 \times 10^{-3}$ tiene dos cifras significativas.

EJEMPLO 12. $.001800 = 0.001800 = 1.800 \times 10^{-3}$ tiene cuatro cifras significativas.

Los números asociados a enumeraciones, por contraposición a los obtenidos por mediciones, son exactos y tienen una cantidad ilimitada de cifras significativas. No obstante, en algunos de estos casos puede resultar difícil decidir qué cifras son significativas sin información adicional. Así, el número 186,000,000 puede tener 3, 4, ..., 9 cifras significativas. Si se sabe que tiene cinco, es mejor escribirlo como 186.00 millones o bien 1.8600×10^8.

CALCULOS

Al efectuar cálculos que impliquen productos, divisiones y raíces de números, el resultado final no puede tener más dígitos significativos que el ingrediente con menor cantidad de ellos (véase Problema 1.9).

EJEMPLO 13. $73.24 \times 4.52 = (73.24)(4.52) = 331.$

EJEMPLO 14. $1.648/0.023 = 72.$

EJEMPLO 15. $\sqrt{38.7} = 6.22.$

EJEMPLO 16. $(8.416)(50) = 420.8$ (si 50 es exacto).

Al hacer sumas y restas, el resultado final no puede tener más cifras significativas tras el punto decimal que el ingrediente con menor cantidad de ellas (véase Prob. 1.10).

EJEMPLO 17. $3.16 + 2.7 = 5.9.$

EJEMPLO 18. $83.42 - 72 = 11.$

EJEMPLO 19. $47.816 - 25 = 22.816$ (si 25 es exacto).

La regla precedente admite generalización (véase Prob. 1.11).

FUNCIONES

Si a cada valor posible de una variable X le corresponden uno o más valores de otra variable Y, decimos que Y es *función* de X y escribimos $Y = F(X)$ (léase «Y igual a F de X») para indicar esa dependencia funcional. Cabe utilizar en vez de F otras letras (G, ϕ, etc.).

La variable X se llama la *variable independiente* e Y la *variable dependiente.*

Si a cada valor de X le corresponde un solo valor de Y, se dice que Y es *función univaluada* de X; en caso contrario, se dice *multivaluada.*

EJEMPLO 20. La población total P de EE.UU. es función del tiempo t, y escribimos $P = F(t)$.

EJEMPLO 21. La longitud L de un muelle vertical es función del peso P que soporta. En símbolos, $L = G(P)$.

La dependencia funcional (o correspondencia) entre variables se anota a veces en una tabla. Sin embargo, puede también indicarse con una ecuación que conecta ambas variables, tal como $Y = 2X - 3$, de la que Y se determina a partir de X.

Si $Y = F(X)$, se suele denotar por $F(3)$ el «valor de Y cuando $X = 3$», por $F(10)$ el «valor de Y cuando $X = 10$», etc. Así que si $Y = F(X) = X^2$, entonces $F(3) = 3^2 = 9$ es el valor de Y para $X = 3$.

El concepto de función admite extensión a varias variables (véase Prob. 1.17).

COORDENADAS RECTANGULARES

Consideremos dos rectas perpendiculares $X'OX$ e $Y'OY$, llamadas *ejes* X e Y, respectivamente (véase Fig. 1.1), sobre las que se indican escalas apropiadas. Estas rectas dividen el plano que determinan, llamado *plano* XY, en cuatro regiones denotadas por I, II, III y IV, que llamaremos primero, segundo, tercero y cuarto *cuadrantes*, respectivamente.

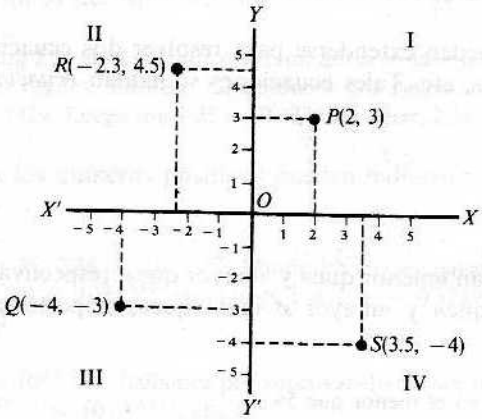


Figura 1.1.

El punto O se llama *origen* o *punto cero*. Dado un punto P, tracemos perpendiculares a los ejes X e Y desde P. Los valores de X, Y en los puntos donde tales perpendiculares cortan a los ejes se conocen como las *coordenadas rectangulares*, o simplemente *coordenadas* de P y se denotan (X, Y). La coordenada X se llama *abscisa*, y la Y *ordenada*, del punto. En la Figura 1.1 la abscisa del punto P es 2 y la ordenada es 3, de modo que las coordenadas de P son $(2, 3)$.

Recíprocamente, dadas las coordenadas de un punto, podemos localizar (marcar) el punto. Así, los puntos con coordenadas $(-4, -3)$, $(-2.3, 4.5)$ y $(3.5, -4)$ están representados en la Figura 1.1 por Q, R y S, respectivamente.

Construyendo un eje Z que pase por O y sea perpendicular al plano XY, podemos extender fácilmente las ideas anteriores. En tal caso, las coordenadas de un punto P se denotan (X, Y, Z).

GRAFICOS

Un *gráfico* es una representación de la relación entre variables. Muchos tipos de gráficos aparecen en Estadística, según la naturaleza de los datos involucrados y el propósito del gráfico. Entre ellos citemos los *gráficos de barras*, *circulares*, etc. Estos gráficos se refieren a veces como diagramas. Hablaremos, por tanto, de *diagramas* de barras, circulares, etc. (véanse Probs. 1.23, 1.24, 1.26 y 1.27).

ECUACIONES

Las ecuaciones son enunciados del tipo $A = B$, donde A se llama *miembro* (o *lado*) *izquierdo*, y B *miembro derecho*, de la ecuación. Siempre que se efectúe sobre ambos miembros de una ecuación una misma operación, se obtendrán ecuaciones equivalentes. Por tanto, se puede sumar, restar, multiplicar o dividir ambos lados de una ecuación por el mismo número y se llegará a una ecuación equivalente, con la única excepción de *la división por cero*, que *no está permitida.*

EJEMPLO 22. Dada la ecuación $2X + 3 = 9$, restemos 3 de ambos lados: $2X + 3 - 3 = 9 - 3$, o sea $2X = 6$. Dividimos ambos miembros por 2: $2X/2 = 6/2$, es decir $X = 3$. Este valor de X es una *solución* de la ecuación dada, como se ve sustituyendo X por 3, obteniéndose $2(3) + 3 = 9$ ó $9 = 9$, que es una *identidad.* Este proceso de hallar soluciones de una ecuación se llama *resolver* la ecuación.

Las ideas precedentes pueden extenderse para resolver dos ecuaciones en dos incógnitas, tres ecuaciones en tres incógnitas, etc. Tales ecuaciones se llaman *ecuaciones simultáneas* (véase Problema 1.30).

DESIGUALDADES

Los símbolos $<$ y $>$ significan «menor que» y «mayor que», respectivamente. Los símbolos $\leqslant$ y $\geqslant$ significan «menor o igual que» y «mayor o igual que», respectivamente. Son los *símbolos de desigualdad.*

EJEMPLO 23. $3 < 5$ se lee «3 es menor que 5».

EJEMPLO 24. $5 > 3$ se lee «5 es mayor que 3».

EJEMPLO 25. $X < 8$ se lee «X es menor que 8».

EJEMPLO 26. $X \geqslant 10$ se lee «X es mayor o igual que 10».

EJEMPLO 27. $4 < Y \leqslant 6$ se lee «4 es menor que Y, que es menor o igual que 6», o bien «Y está entre 4 y 6, excluyendo el 4, pero incluyendo el 6», o sea, «Y es mayor que 4, y menor o igual que 6».

Las relaciones que usan símbolos de desigualdad se llaman *desigualdades.* Igual que hablamos de miembros de una ecuación, hablaremos de *miembros* (o *lados*) *de una desigualdad.* De modo que en la desigualdad $4 < Y \leqslant 6$, los miembros son 4, Y y 6.

Una desigualdad válida permanece válida si:

1. Se suma o resta el mismo número de ambos lados

EJEMPLO 28. Como $15 > 12$, $15 + 3 > 12 + 3$ (es decir, $18 > 15$) y $15 - 3 > 12 - 3$ (es decir, $12 > 9$).

2. Se multiplica o divide cada lado por un mismo número *positivo.*

EJEMPLO 29. Como $15 > 12$, $(15)(3) > (12)(3)$ (es decir, $45 > 36$) y $15/3 > 12/3$ (es decir, $5 > 4$).

3. Se multiplica o divide cada lado por un mismo número *negativo* y se invierte el símbolo de desigualdad.

EJEMPLO 30. Como $15 > 12$, $(15)(-3) < (12)(-3)$ (es decir, $-45 < -36$) y $15/(-3) < 12/(-3)$ (es decir, $-5 < -4$).

LOGARITMOS

Todo número positivo N puede expresarse como potencia de 10; es decir, podemos encontrar p tal que $N = 10^p$. Se dice que p es el *logaritmo de N en base 10*, o el *logaritmo común o decimal* de N, y se escribe en breve $p = \log N$, o bien $p = \log_{10} N$. Por ejemplo, como $1000 = 10^3$, $\log 1000 = 3$. Del mismo modo, como $0.01 = 10^{-2}$, $\log 0.01 = -2$.

Cuando N está entre 1 y 10 (o sea, 10^0 y 10^1), $p = \log N$ es un número entre 0 y 1, y se puede hallar con la tabla de logaritmos del Apéndice VII.

EJEMPLO 31. Para hallar log 2.36 en el Apéndice VII, miramos en la columna de la izquierda, encabezada por N, hasta encontrar los dos dígitos iniciales, 23. Entonces nos desplazamos a la derecha a la columna encabezada por 6. Allí leemos 3729. Luego $\log 2.36 = 0.3729$ (es decir, $2.36 = 10^{0.3729}$).

Los logaritmos de todos los números positivos pueden hallarse a partir de los de los números comprendidos entre 1 y 10.

EJEMPLO 32. Del Ejemplo 31, $2.36 = 10^{0.3729}$. Multiplicando sucesivamente por 10, tenemos $23.6 = 10^{1.3729}$, $236 = 10^{2.3729}$, $2360 = 10^{3.3729}$, etc. Luego $\log 2.36 = 0.3729$, $\log 23.6 = 1.3729$, $\log 236 = 2.3729$, y $\log 2360 = 3.3729$.

EJEMPLO 33. Como $2.36 = 10^{0.03729}$, hallamos por sucesivas divisiones por 10 que $0.236 = 10^{0.3729-1} = 10^{-0.6271}$, $0.0236 = 10^{0.3729-2} = 10^{-1.6271}$, etc.

A menudo escribimos $0.3729 - 1$ como $9.3729 - 10$, o $\bar{1}.3729$; y $0.3729 - 2$ como $8.3729 - 10$, o $\bar{2}.3729$; etcétera. Con esa notación se tiene

$$\log 0.236 = 9.3729 - 10 = \bar{1}.3729 = -0.6271$$

$$\log 0.0236 = 8.3729 - 10 = \bar{2}.3729 = -1.6271$$

etcétera.

La parte decimal .3729 en todos esos logaritmos se llama *mantisa*. El resto, que antecede al punto decimal [o sea, 1, 2, 3, y $\bar{1}$ y $\bar{2}$ (o sea $9 - 10$, $8 - 10$, respectivamente)] se llama la *característica*.

Es sencillo demostrar las siguientes reglas:

1. Para un número mayor que 1 la característica es positiva y vale una unidad *menos* que el número de dígitos que preceden al punto decimal.

EJEMPLO 34. Las características de los logaritmos de 2360, 236, 23.6 y 2.36 son 3, 2, 1 y 0, y los logaritmos son 3.3729, 2.3729, 1.3729 y 0.3729.

2. Para un número menor que 1, la característica es negativa y vale uno *más* que el número de ceros que siguen al punto decimal.

EJEMPLO 35. Las características de los logaritmos de 0.236, 0.0236 y 0.00236 son -1, -2 y -3, y los logaritmos son $\bar{1}.3729$, $\bar{2}.3729$ y $\bar{3}.3729$, o sea $9.3729 - 10$, $8.3729 - 10$ y $7.3729 - 10$, respectivamente.

Si se precisan logaritmos de números de cuatro cifras (como 2.364 y 758.2) debe usarse *interpolación* (véase Prob. 1.36).

ANTILOGARITMOS

En la forma exponencial $2.36 = 10^{0.3729}$, el número 2.36 se llama el *antilogaritmo* de 0.3729, o sea antilog 0.3729. Es el número cuyo logaritmo es 0.3729. Se sigue que antilog $1.3729 = 23.6$, antilog $2.3729 = 236$, antilog $3.3729 = 2360$, antilog $9.3729 - 10 =$ antilog $\bar{1}.3729 = 0.236$ y antilog $8.3729 - 10 =$ antilog $\bar{2}.3729 = 0.0236$. El antilogaritmo de cualquier número se puede hallar con el Apéndice VII.

EJEMPLO 36. Para hallar antilog $8.6284 - 10$, miramos la mantisa .6284 dentro de la tabla. Como aparece en la fila del 42 y en la columna encabezada con 5, los dígitos requeridos son 425. Y ya que la característica es $8 - 10$, el número es 0.0425.

Análogamente, antilog $3.6284 = 4250$ y antilog $5.6284 = 425{,}000$.

Si no se encuentra la mantisa en el Apéndice VII, úsese interpolación (véase Prob. 1.37).

CALCULOS USANDO LOGARITMOS

Estos cálculos recurren a las siguientes propiedades:

$$\log MN = \log M + \log N$$

$$\log \frac{M}{N} = \log M - \log N$$

$$\log M^p = p \log M$$

Combinando esos resultados obtenemos, por ejemplo,

$$\log \frac{A^p B^q C^r}{D^s E^t} = p \log A + q \log B + r \log C - s \log D - t \log E$$

Véanse Problemas 1.38 al 1.45.

PROBLEMAS RESUELTOS

VARIABLES

1.1. Decir cuáles de estos datos son discretos y cuáles continuos:

(*a*) Número de acciones vendidas un día en la Bolsa de Valores.
(*b*) Temperaturas medidas en un observatorio cada media hora.
(*c*) Vida media de los tubos de televisión producidos por una fábrica.
(*d*) Ingresos anuales de los profesores de Enseñanza Media.
(*e*) Longitudes de 1000 tornillos producidos en una empresa.

Solución

(*a*) Discretos; (*b*) continuos; (*c*) continuos; (*d*) discretos; (*e*) continuos.

1.2. Dar el dominio de las siguientes variables y decir cuáles son continuas:

(*a*) Número G de galones (gal) de agua en una lavadora.
(*b*) Número B de libros en una estantería.
(*c*) Suma S de los puntos obtenidos al lanzar un par de dados.
(*d*) Diámetro D de una esfera.
(*e*) País P de Europa.

Solución

(*a*) *Dominio:* Cualquier valor entre 0 gal y la capacidad de la lavadora. *Variable:* Continua.
(*b*) *Dominio:* 0, 1, 2, 3, ... hasta el número total de libros que caben en la estantería. *Variable:* Discreta.
(*c*) *Dominio:* Los puntos de un dado pueden ser 1, 2, 3, 4, 5 ó 6. Luego la suma de dos dados puede ser 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ó 12, que es el dominio de S. *Variable:* Discreta.
(*d*) *Dominio:* Todos los valores positivos. *Variable:* Continua.
(*e*) *Dominio:* Francia, Italia, ..., etc., que pueden representarse numéricamente como 1, 2, ... *Variable:* Discreta.

REDONDEO DE DATOS

1.3. Redondear cada número con la precisión establecida:

(*a*)	48.6	unidades	(*f*)	143.95	decenas
(*b*)	136.5	unidades	(*g*)	368	centenas
(*c*)	2.484	centésimas	(*h*)	24,448	millares
(*d*)	0.0435	milésimas	(*i*)	5.56500	centésimas
(*e*)	4.50001	unidades	(*j*)	5.56501	centésimas

Solución

(*a*) 49; (*b*) 136; (*c*) 2.48; (*d*) 0.044; (*e*) 5; (*f*) 144.0; (*g*) 400; (*h*) 24,000; (*i*) 5.56; (*j*) 5.57.

1.4. Sumar los números 4.35, 8.65, 2.95, 12.45, 6.65, 7.55 y 9.75 (*a*) directamente, (*b*) redondeando en décimas con la regla del «entero par», y (*c*) idem eligiendo el entero dudoso más alto anterior al 5.

Solución

(*a*)		(*b*)		(*c*)	
	4.35		4.4		4.4
	8.65		8.6		8.7
	2.95		3.0		3.0
	12.45		12.4		12.5
	6.65		6.6		6.7
	7.55		7.6		7.6
	9.75		9.8		9.8
Total	52.35	Total	52.4	Total	52.7

Nótese que el método (*b*) es mejor que el (*c*) por cuanto minimiza la *acumulación de errores de redondeo*.

NOTACION CIENTIFICA Y DIGITOS SIGNIFICATIVOS

1.5. Expresar los siguientes números sin usar potencias de 10:

(*a*) 4.823×10^7 (*c*) 3.80×10^{-4} (*e*) 300×10^8
(*b*) 8.4×10^{-6} (*d*) 1.86×10^5 (*f*) $70{,}000 \times 10^{-10}$

Solución

(*a*) Movemos el punto decimal siete lugares a la derecha y obtenemos 48,230,000; (*b*) moviendo ahora seis posiciones a la izquierda queda 0.0000084; (*c*) 0.000380; (*d*) 186,000; (*e*) 30,000,000,000; (*f*) 0.0000070000.

1.6. ¿Cuántas cifras significativas hay en cada uno de estos números, supuesto que han sido redondeados correctamente?

(*a*) 149.8 in (*d*) 0.00280 m (*g*) 9 casas
(*b*) 149.80 in (*e*) 1.00280 m (*h*) 4.0×10^3 libras (lb)
(*c*) 0.0028 metros (m) (*f*) 9 gramos (g) (*i*) 7.58400×10^{-5} dinas

Solución

(*a*) cuatro; (*b*) cinco; (*c*) dos; (*d*) tres; (*e*) seis; (*f*) uno; (*g*) sin límite; (*h*) dos; (*i*) seis.

1.7. ¿Cuál es el máximo error en cada una de estas medidas, supuesto que se han anotado del modo más preciso posible?

(*a*) 73.854 in (*b*) 0.09800 pies cúbicos (ft^3) (*c*) 3.867×10^8 kilómetros (km)

Solución

(*a*) La medida debe estar entre 73.8535 a 73.8545 in; luego el máximo error es 0.0005 in. Hay 5 cifras significativas.
(*b*) El número de pies cúbicos está entre 0.097995 a 0.098005 pies cúbicos; luego el error máximo es 0.000005 pies cúbicos. Cuatro cifras significativas.
(*c*) El número real de kilómetros es mayor que 3.8665×10^8 pero menor que 3.8675×10^8; por tanto, el máximo error posible es 0.0005×10^8, o sea 50,000 km. Cuatro cifras significativas.

1.8. Escribir cada número en notación científica. Salvo mención expresa en contra, se suponen todas las cifras significativas.

(*a*) 24,380,000 (cuatro cifras significativas) (*c*) 7,300,000,000 (cinco cifras significativas)
(*b*) 0.000009851 (*d*) 0.00018400

Solución

(*a*) 2.438×10^7; (*b*) 9.851×10^{-6}; (*c*) 7.3000×10^9; (*d*) 1.8400×10^{-4}.

CALCULOS

1.9. Probar que el producto de 5.74 y 3.8, supuesto que tienen tres y dos cifras significativas, no puede lograrse con más de dos cifras significativas.

Solución

Primer método

$5.74 \times 3.8 = 21.812$, pero no todas las cifras de este producto son significativas. Para ver cuántas lo son, nótese que 5.74 puede ser cualquier número entre 5.735 y 5.745, mientras que 3.8 es cualquiera entre 3.75 y 3.85. Luego el menor valor posible del producto es $5.735 \times 3.75 = 21.50625$, y el mayor $5.745 \times 3.85 = 21.11825$.

Como el posible rango de valores es 21.50625 a 22.11825, es claro que sólo las dos primeras cifras del producto son cifras significativas, pudiendo escribir el resultado como 22. Observemos que 22 debe interpretarse como cualquier número entre 21.5 y 22.5.

Segundo método

Con las cifras dudosas en cursiva, el producto es:

$$\begin{array}{r} 5.7\mathit{4} \\ 3.\mathit{8} \\ \hline \mathit{4592} \\ 172\mathit{2} \\ \hline 2\mathit{1.812} \end{array}$$

No debemos conservar más de una cifra dudosa en el producto, que es en consecuencia 22 con dos cifras significativas. Es, por tanto, innecesario arrastrar más cifras significativas de las que figuren en el factor menos preciso; así, si 5.74 se redondea a 5.7, el producto es $5.7 \times 3.8 = 21.66 = 22$ con dos cifras significativas, de acuerdo con el resultado ya sabido.

Al calcular a mano, se ahorra trabajo no guardando más que una o dos cifras más allá de las que tenga el factor menos preciso, y redondeando al número adecuado de cifras significativas el resultado final. Con calculadoras que manejan muchos dígitos, debe tenerse cuidado en no creer que todas las obtenidas son cifras significativas.

1.10. Sumar 4.19355, 15.28, 5.9561, 12.3 y 8.472, suponiendo que todas son cifras significativas.

Solución

Pondremos en el cálculo (*a*) las cifras dudosas en cursiva. La respuesta final con sólo una cifra dudosa se presenta como 46.2.

(*a*)	(*b*)
4.1935*5*	4.19
15.2*8*	15.28
5.956*1*	5.96
12.*3*	12.3
8.47*2*	8.47
46.*20165*	46.20

Se ahorra esfuerzo guardando, como en (*b*), un decimal significativo más que en el número preciso. La respuesta final, redondeada a 46.2, coincide con el cálculo (*a*).

1.11. Calcular $475{,}000{,}000 + 12{,}684{,}000 - 1{,}372{,}410$ si esos números tienen 3,5 y 7 cifras significativas, respectivamente.

Solución

En (*a*) conservaremos todas las cifras y redondearemos el resultado final. En (*b*), usamos un método análogo al del Problema 1.10(*b*). En ambos casos, las cifras dudosas están en cursiva.

$$(a)\quad \begin{array}{r} 475{,}000{,}000 \\ +\ 12{,}684{,}000 \\ \hline 487{,}684{,}000 \end{array} \qquad \begin{array}{r} 487{,}684{,}000 \\ -\ 1{,}372{,}410 \\ \hline 486{,}311{,}590 \end{array} \qquad (b)\quad \begin{array}{r} 475{,}000{,}000 \\ +\ 12{,}700{,}000 \\ \hline 487{,}700{,}000 \end{array} \qquad \begin{array}{r} 487{,}700{,}00 \\ -\ 1{,}400{,}000 \\ \hline 486{,}300{,}000 \end{array}$$

El resultado final se redondea a 486,000,000; o mejor, para mostrar que hay 3 cifras significativas, escribirlo como 486 millones o 4.86×10^8.

1.12. Efectuar cada operación indicada.

(*a*) 48.0×943

(*b*) $8.35/98$

(*c*) $(28)(4193)(182)$

(*d*) $\dfrac{(526.7)(0.001280)}{0.000034921}$

(*e*) $\dfrac{(1.47562 - 1.47322)(4895.36)}{0.000159180}$

(*f*) Si los denominadores 5 y 6 son exactos, $\dfrac{(4.38)^2}{5} + \dfrac{(5.482)^2}{6}$

(*g*) $3.1416\sqrt{71.35}$

(*h*) $\sqrt{128.5 - 89.24}$

Solución

(*a*) $48.0 \times 943 = (48.0)(943) = 45{,}300$

(*b*) $8.35/98 = 0.085$

(*c*) $(28)(4193)(182) = (2.8 \times 10^1)(4.193 \times 10^3)(1.82 \times 10^2)$
$= (2.8)(4.193)(1.82) \times 10^{1+3+2} = 21 \times 10^6 = 2.1 \times 10^7$

Esto puede escribirse también como 21 millones para mostrar las dos cifras significativas.

(*d*) $$\frac{(526.7)(0.001280)}{0.000034921} = \frac{(5.267 \times 10^2)(1.280 \times 10^{-3})}{3.4921 \times 10^{-5}} = \frac{(5.267)(1.280)}{3.4921} \times \frac{(10^2)(10^{-3})}{10^{-5}}$$
$$= 1.931 \times \frac{10^{2-3}}{10^{-5}} = 1.931 \times \frac{10^{-1}}{10^{-5}}$$
$$= 1.931 \times 10^{-1+5} = 1.931 \times 10^4$$

Que cabe presentar como 19.31 miles mostrando las cuatro cifras significativas.

(*e*) $$\frac{(1.47562 - 1.47322)(4895.36)}{0.000159180} = \frac{(0.00240)(4895.36)}{0.000159180} = \frac{(2.40 \times 10^{-3})(4.89536 \times 10^3)}{1.59180 \times 10^{-4}}$$
$$= \frac{(2.40)(4.89536)}{1.59180} \times \frac{(10^{-3})(10^3)}{10^{-4}} = 7.38 \times \frac{10^0}{10^{-4}} = 7.38 \times 10^4$$

Esto puede expresarse como 73.8 miles, mostrando sus tres cifras significativas. Nótese que aunque habia seis cifras significativas en cada número inicial, algunas se han perdido al restar 1.47322 de 1.47562.

(*f*) Si los denominadores 5 y 6 son exactos, $\frac{(4.38)^2}{5} + \frac{(5.482)^2}{6} = 3.84 + 5.009 = 8.85$

(*g*) $3.1416\sqrt{71.35} = (3.1416)(8.447) = 26.54$

(*h*) $\sqrt{128.5 - 89.24} = \sqrt{39.3} = 6.27$

1.13. Evaluar lo que sigue, dado que $X = 3$, $Y = -5$, $A = 4$ y $B = -7$, donde todos los números son exactos:

(*a*) $2X - 3Y$

(*b*) $4Y - 8X + 28$

(*c*) $\frac{AX + BY}{BX - AY}$

(*d*) $X^2 - 3XY - 2Y^2$

(*e*) $2(X + 3Y) - 4(3X - 2Y)$

(*f*) $\frac{X^2 - Y^2}{A^2 - B^2 + 1}$

(*g*) $\sqrt{2X^2 - Y^2 - 3A^2 + 4B^2 + 3}$

(*h*) $\sqrt{\frac{6A^2}{X} + \frac{2B^2}{Y}}$

Solución

(*a*) $2X - 3Y = 2(3) - 3(-5) = 6 + 15 = 21$

(*b*) $4Y - 8X + 28 = 4(-5) - 8(3) + 28 = -20 - 24 + 28 = -16$

(*c*) $\frac{AX + BY}{BX - AY} = \frac{(4)(3) + (-7)(-5)}{(-7)(3) - (4)(-5)} = \frac{12 + 35}{-21 + 20} = \frac{47}{-1} = -47$

(*d*) $X^2 - 3XY - 2Y^2 = (3)^2 - 3(3)(-5) - 2(-5)^2 = 9 + 45 - 50 = 4$

(*e*) $2(X + 3Y) - 4(3X - 2Y) = 2[(3) + 3(-5)] - 4[3(3) - 2(-5)]$

$= 2(3 - 15) - 4(9 + 10) = 2(-12) - 4(19) = -24 - 76 = -100$

Otro método

$$2(X + 3Y) - 4(3X - 2Y) = 2X + 6Y - 12X + 8Y = -10X + 14Y = -10(3) + 14(-5)$$
$$= -30 - 70 = -100$$

(*f*) $\frac{X^2 - Y^2}{A^2 - B^2 + 1} = \frac{(3)^2 - (-5)^2}{(4)^2 - (-7)^2 + 1} = \frac{9 - 25}{16 - 49 + 1} = \frac{-16}{-32} = \frac{1}{2} = 0.5$

(*g*) $\sqrt{2X^2 - Y^2 - 3A^2 + 4B^2 + 3} = \sqrt{2(3)^2 - (-5)^2 - 3(4)^2 + 4(-7)^2 + 3}$

$= \sqrt{18 - 25 - 48 + 196 + 3} = \sqrt{144} = 12$

(*h*) $\sqrt{\frac{6A^2}{X} + \frac{2B^2}{Y}} = \sqrt{\frac{6(4)^2}{3} + \frac{2(-7)^2}{-5}} = \sqrt{\frac{96}{3} + \frac{98}{-5}} = \sqrt{12{,}4} = 3.52$ aproximadamente

FUNCIONES

1.14. La Tabla 1.1 muestra el número de bushels (bu) de trigo y maíz producidos en la cooperativa PQR durante los años 1975-1985. Con referencia a esa tabla, determinar el año o años durante los cuales: (*a*) la producción de trigo fue mínima, (*b*) la de maíz fue máxima, (*c*) se dio el mayor descenso en la producción de trigo, (*d*) decreció la producción de maíz respecto del año anterior y creció la de trigo, (*e*) se produjo idéntica cantidad de trigo y (*f*) la producción conjunta de trigo y maíz fue máxima.

Tabla 1.1

Año	Número de bushels de trigo	Número de bushels de maíz
1975	200	75
1976	185	90
1977	225	100
1978	250	85
1979	240	80
1980	195	100
1981	210	110
1982	225	105
1983	250	95
1984	230	110
1985	235	100

Solución

(*a*) 1976; (*b*) 1981 y 1984; (*c*) 1980; (*d*) 1978, 1982, 1983 y 1985; (*e*) 1977 y 1982, y 1978 y 1983; (*f*) 1983.

1.15. Sean W y C, respectivamente el número de bushels de trigo y maíz producidos en el año t en la cooperativa PQR del Problema 1.14. Es claro que W y C son ambas funciones de t, lo que podemos indicar como $W = F(t)$ y $C = G(t)$.

(*a*) Hallar W cuando $t = 1981$.
(*b*) Hallar C cuando $t = 1978$ y 1984.
(*c*) Hallar t cuando $W = 225$.
(*d*) Hallar $F(1979)$.
(*e*) Hallar $G(1983)$.
(*f*) Hallar C cuando $W = 210$.
(*g*) ¿Cuál es el dominio de la variable t?
(*h*) ¿Es W función univaluada de t?
(*i*) ¿Es t función de W? Si lo es, ¿es univaluada?
(*j*) ¿Es C función de W?
(*k*) ¿Qué variable es independiente, t o W?

Solución

(*a*) 210; (*b*) 85 y 110, respectivamente; (*c*) 1977 y 1982; (*d*) 240; (*e*) 95; (*f*) 110; (*g*) los años 1975, 1976, ..., 1985.

(*h*) Sí, pues a cada valor de t en su dominio le corresponde uno y sólo un valor de W.

(*i*) Sí, porque a cada valor de W podemos suponer que le corresponden uno o más valores de t, que pueden hallarse con la Tabla 1.1. Como puede haber más de un valor de t para cada valor de W (así ocurre con $W = 225$ y $t = 1977$ ó 1982), la función es multivaluada. Esta dependencia funcional de t en W se puede expresar como $t = H(W)$.

(*j*) Sí, pues a cada valor que puede tomar W le corresponden uno o más valores de C, como enseña la Tabla 1.1. Análogamente, W es función de C.

(*k*) Físicamente, suele pensarse en W como determinado por t, y no al revés. Así pues, físicamente t es la variable independiente y W la dependiente. Matemáticamente, sin embargo, cualquiera de las variables puede verse como independiente y la otra como dependiente. A la que se asignan diversos valores es la independiente; la que viene determinada como resultado es la dependiente.

1.16. Una variable Y queda determinada por la variable X mediante la ecuación $Y = 2X - 3$, donde 2 y 3 son exactos.

(*a*) Hallar Y cuando $X = 3, -2$ y 1.5.
(*b*) Poner en una tabla los valores de Y para $X = -2, -1, 0, 1, 2, 3$ y 4.
(*c*) Si denotamos la dependencia de Y en X por $Y = F(X)$, determinar $F(2.4)$ y $F(0.8)$.
(*d*) ¿Qué valor de X corresponde a $Y = 15$?
(*e*) ¿Puede expresarse X como función de Y?
(*f*) ¿Es Y función univaluada de X?
(*g*) ¿Es X función univaluada de Y?

Solución

(*a*) Cuando $X = 3$, $Y = 2X - 3 = 2(3) - 3 = 6 - 3 = 3$. Cuando $X = -2$, $Y = 2X - 3 = 2(-2) - 3 = -4 - 3 = -7$. Cuando $X = 1.5$, $Y = 2X - 3 = 2(1.5) - 3 = 3 - 3 = 0$.
(*b*) Los valores de Y, calculados como en (*a*), se indican en la Tabla 1.2. Nótese que pueden construirse otras tablas escogiendo otros valores de X. La relación $Y = 2X - 3$ es equivalente a la colección de todas las posibles tablas.

Tabla 1.2

X	-2	-1	0	1	2	3	4
Y	-7	-5	-3	-1	1	3	5

(*c*) $F(2.4) = 2(2.4) - 3 = 4.8 - 3 = 1.8$ y $F(0.8) = 2(0.8) - 3 = 1.6 - 3 = -1.4$.
(*d*) Sustituir $Y = 15$ en $Y = 2X - 3$. Se obtiene $15 = 2X - 3$, $2X = 18$ y $X = 9$.
(*e*) Sí. Como $Y = 2X - 3$, $Y + 3 = 2X$ y $X = \frac{1}{2}(Y + 3)$. Esto expresa X *explícitamente* como función de Y.
(*f*) Sí, porque para cada valor posible de X (hay infinitos) le corresponde un solo de Y.
(*g*) Sí, porque de la parte (*e*), $X = \frac{1}{2}(Y + 3)$, de modo que correspondiente a cada valor de Y hay uno y uno sólo de X.

1.17. Si $Z = 16 + 4X - 3Y$, hallar el valor de Z correspondiente a: (*a*) $X = 2$, $Y = 5$; (*b*) $X = -3$, $Y = -7$; (*c*) $X = -4$, $Y = 2$.

Solución

(*a*) $Z = 16 + 4(2) - 3(5) = 16 + 8 - 15 = 9$.
(*b*) $Z = 16 + 4(-3) - 3(-7) = 16 - 12 + 21 = 25$.
(*c*) $Z = 16 + 4(-4) - 3(2) = 16 - 16 - 6 = -6$.

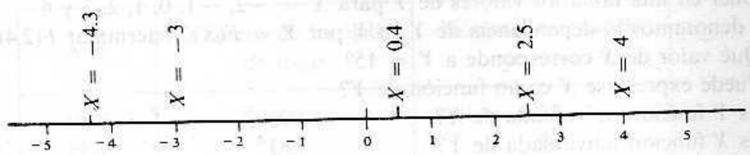
Dados valores de X e Y, les corresponde uno de Z. Podemos denotar esta dependencia de Z en X e Y como $Z = F(X, Y)$ (se lee «Z es función de X e Y»). $F(2.5)$ denota el valor de Z cuando $X = 2$ e $Y = 5$, que es 9; véase (*a*). De la misma manera, $F(-3, -7) = 25$ y $F(-4, 2) = -6$ por las partes (*b*) y (*c*), respectivamente.

Las variables X, Y se llaman *variables independientes*, y Z la *variable dependiente*.

GRAFICOS

1.18. Localizar en el eje X de un sistema coordenado los puntos correspondientes a: (*a*) $X = 4$, (*b*) $X = -3$, (*c*) $X = 2.5$, (*d*) $X = -4.3$ y (*e*) $X = 0.4$, suponiendo que esos valores son exactos.

Solución



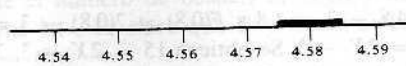
Cada valor exacto de X corresponde a un punto y sólo uno sobre el eje X. Recíprocamente, se demuestra en matemáticas más avanzadas que a cada punto del eje le corresponde un valor de X y sólo uno.

Así pues, teóricamente existe un punto asociado a $X = 22/7 = 3.142857142857...$, o al $X = \pi = 3.14159265358...$ En la práctica, naturalmente, no es factible su localización exacta, porque el lápiz hace una marca de cierta anchura y cubre una infinidad de puntos. El propio eje X tiene grosor. De modo que el diagrama adjunto es una representación física de la situación matemática.

1.19. Sea X el diámetro en centímetros (cm) de una bola. Si $X = 4.58$ con tres cifras significativas, ¿cómo debe representarse en el eje X?

Solución

La verdadera medida está entre 4.575 y 4.585 cm, luego hay que representarla por el segmento grueso de la figura adjunta.



1.20. Localizar en un sistema de coordenadas rectangulares los puntos de coordenadas: (*a*) (5, 2), (*b*) (2, 5), (*c*) (−3, 1), (*d*) (1, −3), (*e*) (3, −4), (*f*) (−2.5, −4.8), (*g*) (0, −2.5) y (*h*) (4, 0). Suponemos exactos todos esos números.

Solución

Véase Figura 1.2.

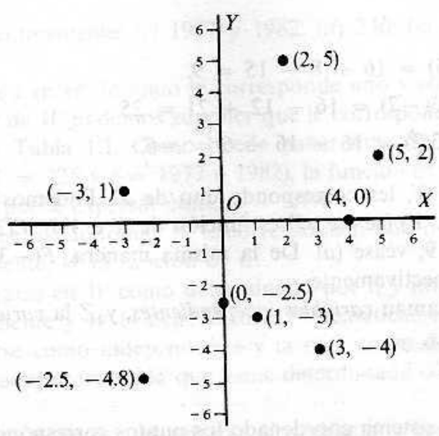


Figura 1.2.

1.21. Representar la ecuación $Y = 2X - 3$.

Solución

Tomando $X = -2, -1, 0, 1, 2, 3$ y 4, obtenemos que $Y = -7, -5, -3, -1, 1, 3$ y 5, respectivamente [véase Prob. 1.16(*b*)]. Luego los puntos vienen dados en el gráfico por $(-2, -7)$, $(-1, -5)$, $(0, -3)$, $(1, -1)$, $(2, 1)$, $(3, 3)$ y $(4, 5)$, que pueden verse representados en coordenadas rectangulares en la Figura 1.3. Todos ellos, así como los obtenidos a partir de otros valores de X, yacen en una recta que es la gráfica pedida.

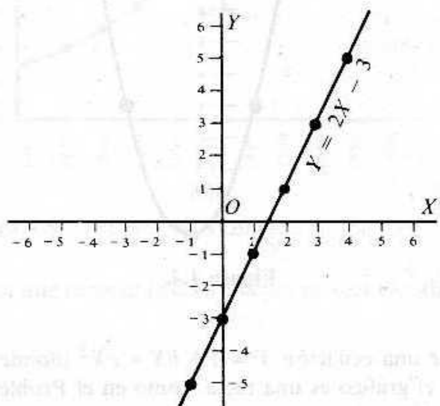


Figura 1.3.

Como la gráfica de $Y = 2X - 3$ es una línea recta, se dice que $F(X) = 2X - 3$ es una *función lineal*. En general, $F(X) = aX + b$ (con a, b constantes) es una función lineal cuya gráfica es una recta.

Nótese que sólo se necesitan dos puntos para hallar la gráfica de una función lineal, pues dos puntos determinan una recta.

1.22. Representar la ecuación $Y = X^2 - 2X - 8$.

Solución

La Tabla 1.3 muestra los valores de Y correspondientes a algunos valores de X; por ejemplo, cuando $X = -2$, $Y = (-2)^2 - 2(-2) - 8 = 4 + 4 - 8 = 0$. De esa tabla vemos que están sobre la gráfica los puntos $(-3, 7)$, $(-2, 0)$, $(-1, -5)$, $(0, -8)$, $(1, -9)$, $(2, -8)$, $(3, -5)$, $(4, 0)$ y $(5, 7)$. Estos puntos, y otros calculados mediante otros valores de X, están sobre la curva de la Figura 1.4, llamada parábola. La función $F(X) = X^2 - 2X - 8$ se llama una *función cuadrática*.

Tabla 1.3

X	-3	-2	-1	0	1	2	3	4	5
Y	7	0	-5	-8	-9	-8	-5	0	7

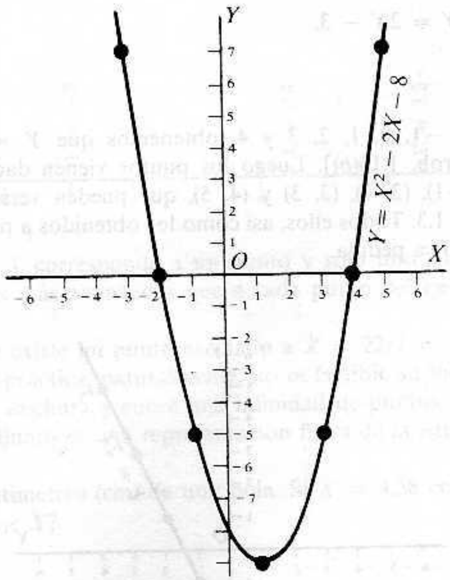


Figura 1.4.

En general, el gráfico de una ecuación $Y = a + bX + cX^2$ (donde a, b y c son constantes y $c \neq 0$) es una parábola. Si $c = 0$, el gráfico es una recta, como en el Problema 1.21.

1.23. La Tabla 1.4 muestra la población de EE.UU. (en millones) en los años 1860-1980. Representar esos datos.

Solución

Primer método

En la Figura 1.5, la población P es la variable dependiente y el tiempo t la variable independiente. Los puntos se localizan del modo habitual por las coordenadas leídas en la tabla, como (1880, 50.2). Se conectan los puntos sucesivos con trazos rectos, ya que no disponemos de información sobre P en los tiempos intermedios; de ahí que el gráfico se llame un *gráfico de trazos.*

Obsérvese que las unidades en los ejes son distintas, como al dibujar el gráfico de $Y = 2X - 3$. Ello es correcto, pues de hecho las dos variables son magnitudes completamente diferentes.

Asimismo, el cero se ha indicado en el eje vertical, pero (por razones obvias) no en el horizontal. Debe indicarse el cero siempre que sea posible, sobre todo en el eje vertical. Si no fuese posible por alguna razón, y si tal omisión pudiera provocar alguna conclusión errónea, es aconsejable advertirlo de algún modo, por ejemplo como en el Problema 1.26.

Tabla 1.4. Población de EE.UU., 1860-1980

Año	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980
Población (millones)	31.4	39.8	50.2	62.9	76.0	92.0	105.7	122.8	131.7	151.1	179.3	203.3	226.5

Fuente: U.S. Bureau of the Census.

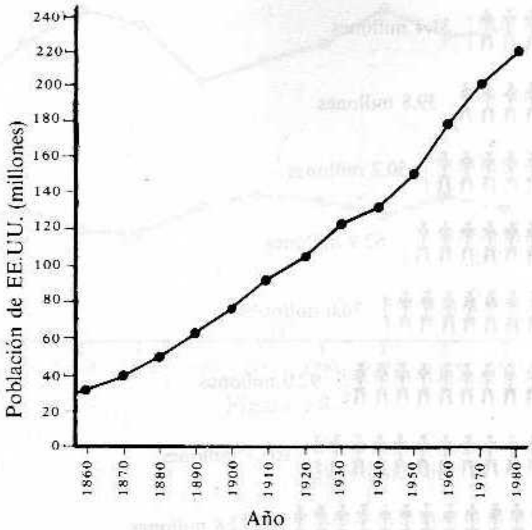


Figura 1.5. (*Fuente:* U.S. Bureau of the Census.)

Una tabla o una gráfica que recojan la distribución de una variable en función del tiempo, se llaman *series en el tiempo.*

Segundo método

La Figura 1.6 se llama un *gráfico o diagrama de barras.* La anchura de cada barra, todas idénticas, no tienen importancia en este caso y se escoge a capricho (siempre que las barras no se solapen).

Los números sobre las barras pueden omitirse. Si se mantienen, la escala vertical de la izquierda es innecesaria.

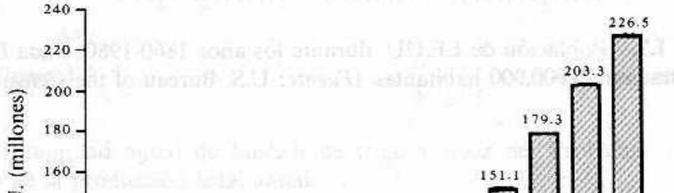


Figura 1.6. (*Fuente:* U.S. Bureau of the Census.)

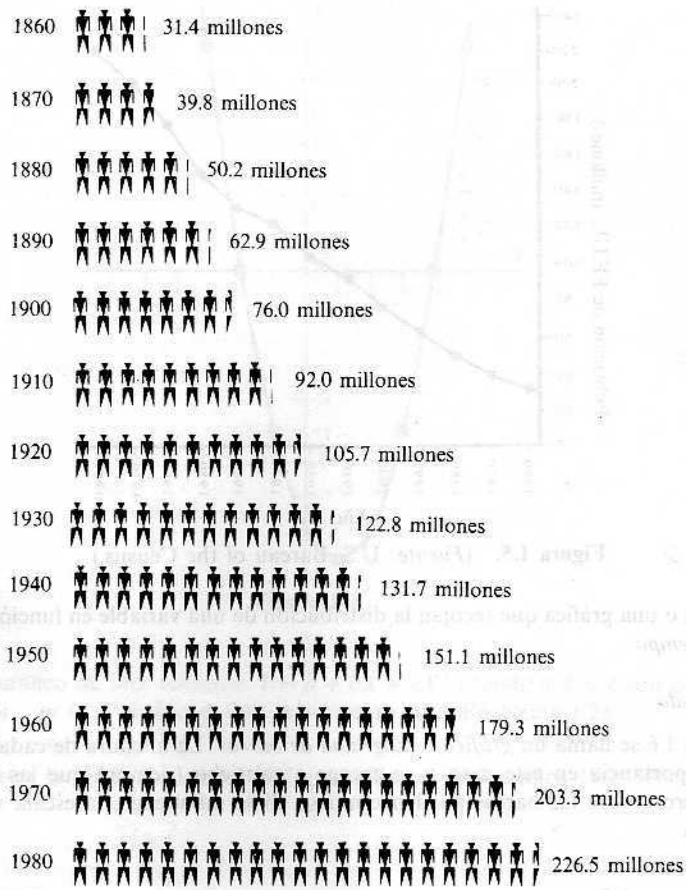


Figura 1.7. Población de EE.UU. durante los años 1860-1980. Cada figura representa 10,000,000 habitantes. (*Fuente:* U.S. Bureau of the Census.)

Tercer método

La Figura 1.7 es un *pictograma*, usado a menudo para representar datos en Estadística de una forma que sea nítida para el gran público. Muchos de ellos conllevan una buena dosis de ingenuidad y originalidad en el arte de la presentación de datos.

El número de la derecha de los monigotes puede omitirse. Incluso en ese caso, el lector podrá estimar la población en una franja de 5 millones.

1.24. Representar los datos del Problema 1.14 usando: (*a*) gráficos de trazos y (*b*) gráficos de barras.

Solución

(*a*) La Figura 1.8 muestra el gráfico de trazos.
(*b*) Véanse las Figuras 1.9 y 1.10. El gráfico de la Figura 1.10 se llama un *gráfico de barras en componentes*.

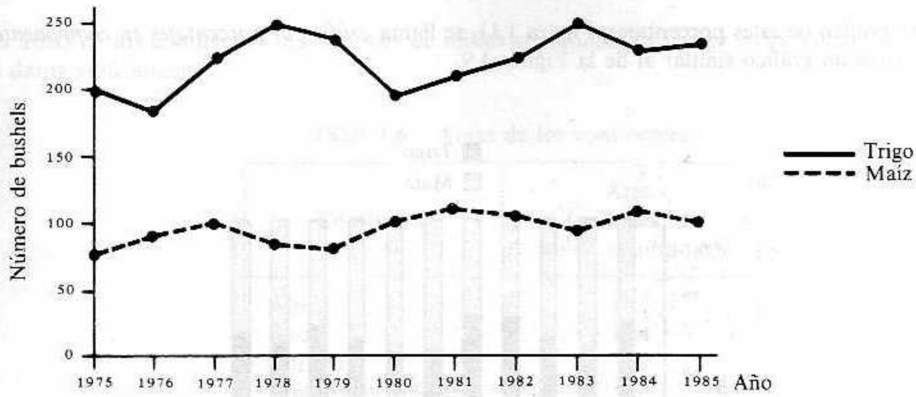


Figura 1.8.

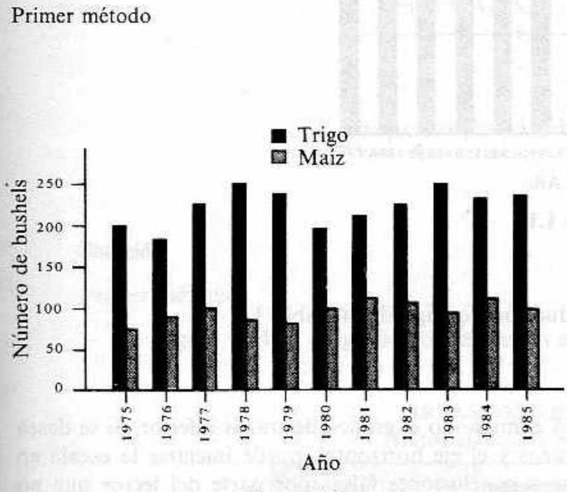


Figura 1.9.

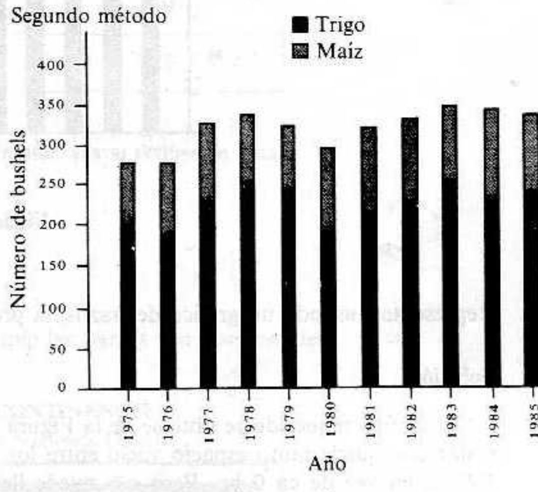


Figura 1.10.

1.25. (*a*) Expresar la cantidad anual de bushels de trigo y maíz del Problema 1.14 (Tabla 1.1) como porcentajes de la producción total anual.

(*b*) Representar los porcentajes obtenidos en la parte (*a*).

Solución

(*a*) En 1975 el porcentaje de trigo = 200/(200 + 75) = 72.7%, y el maíz 100% − 72.7% = 27.3%; etc. Los porcentajes se indican en la Tabla 1.5.

Tabla 1.5

Año	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Porcentaje de trigo	72.7	67.3	69.2	74.6	75.0	66.1	65.6	68.2	72.5	67.6	70.1
Porcentaje de maíz	27.3	32.7	30.8	25.4	25.0	33.9	34.4	31.8	27.5	32.4	29.9

(*b*) El gráfico de tales porcentajes, Figura 1.11, se llama *gráfico de porcentajes en componentes*. Puede usarse un gráfico similar al de la Figura 1.9.

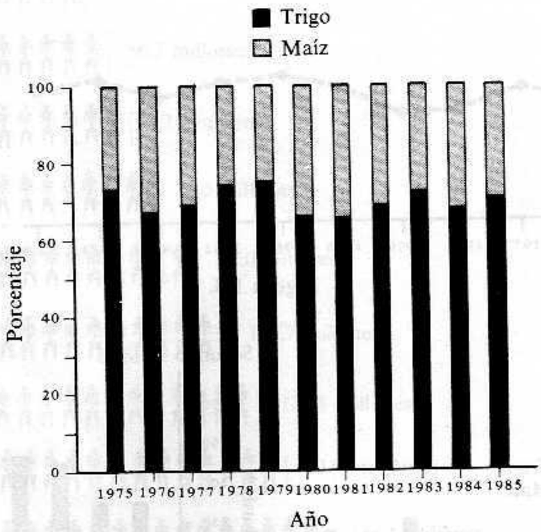


Figura 1.11.

1.26. Representar, usando un gráfico de trazos, la producción de trigo de la Tabla 1.1.

Solución

El gráfico requerido se obtiene de la Figura 1.8 eliminando el gráfico de trazos inferior. Si se desea evitar que quede tanto espacio vacío entre los trazos y el eje horizontal, puede iniciarse la escala en 150 bu en vez de en 0 bu. Pero eso puede llevar a conclusiones falsas por parte del lector que no advierta la omisión del cero. Para advertirle de ello, cabe construir el gráfico de la Figura 1.12.

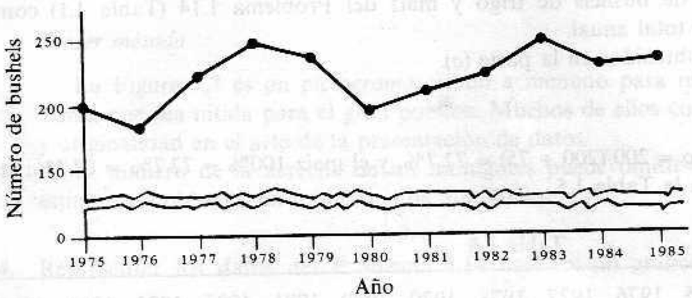


Figura 1.12.

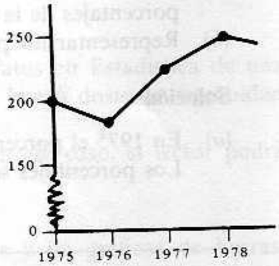


Figura 1.13.

Otro truco frecuente para llamar la atención sobre la supresión del cero es el uso de una línea en zigzag en uno de los ejes (Fig. 1.13).

1.27. Las áreas de los continentes (en millones de millas cuadradas) se recoge en la Tabla 1.6. Representar los datos gráficamente.

Tabla 1.6. Areas de los continentes

Continente	Area (millones de millas cuadradas)
Africa	11.7
Asia	10.4
Europa	1.9
América del Norte	9.4
Oceanía	3.3
América del Sur	6.9
Unión Soviética	7.9
Total	51.5

Fuente: Naciones Unidas.
Nota: Europa excluye Turquía, que se incluye en Asia.

Solución

Primer método

La Figura 1.14 es un gráfico de barras en el que las barras son horizontales.

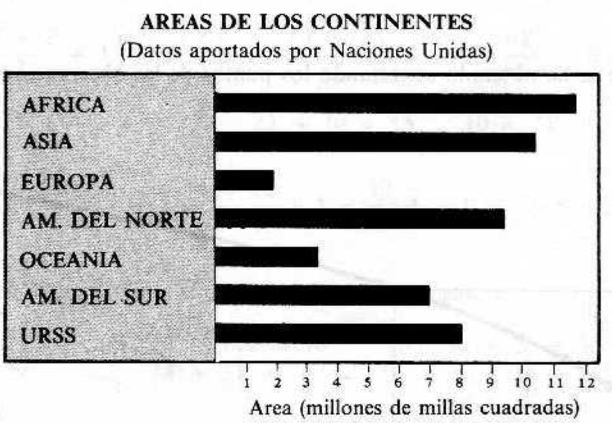


Figura 1.14.

La Figura 1.15 se llama un *diagrama circular*. Para construirlo, hacemos que el área total, 51.5 millones de millas cuadradas, corresponda a los 360° del círculo. Así, un millón corresponde a 360°/51.5. Se deduce que Africa, con 11.7 millones, ocupa un arco de 11.7(360°/51.5) = 82°, mientras Asia, Europa, Norteamérica, Oceanía, América del Sur y la URSS ocupan 73°, 13°, 66°, 23°, 48° y 55°, respectivamente.

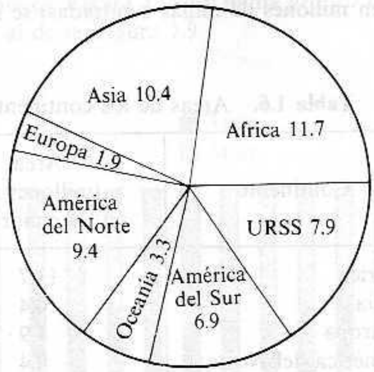


Figura 1.15. Areas de los continentes (en millones de millas cuadradas).

1.28. El tiempo T (en segundos) requerido para una oscilación completa de un péndulo simple de longitud L cm, se ve en la Tabla 1.7, que da las observaciones obtenidas en un laboratorio de Física.

(*a*) Representar gráficamente T como función de L.
(*b*) De la gráfica en (*a*), estimar T para un péndulo de 40 cm.

Tabla 1.7

L	10.1	16.2	22.2	33.8	42.0	53.4	66.7	74.5	86.6	100.0
T	0.64	0.81	0.95	1.17	1.30	1.47	1.65	1.74	1.87	2.01

Solución

(*a*) La Figura 1.16 se ha obtenido conectando los puntos de las observaciones con una curva suave.

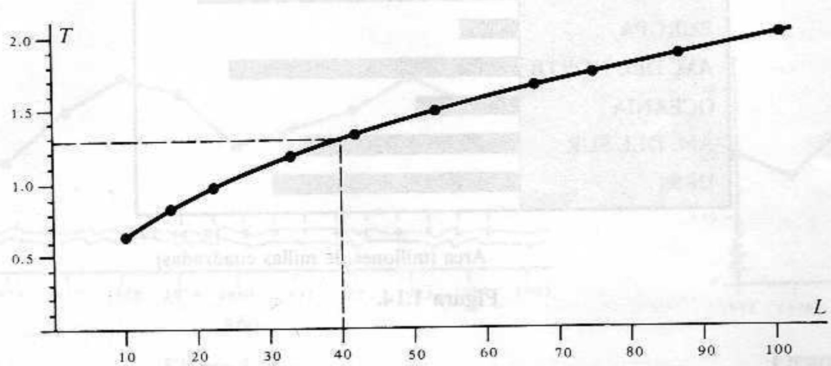


Figura 1.16.

(*b*) El valor estimado de T es 1.27 segundos.

ECUACIONES

1.29. Resolver las ecuaciones:

(*a*) $4a - 20 = 8$ (*c*) $18 - 5b = 3(b + 8) + 10$

(*b*) $3X + 4 = 24 - 2X$ (*d*) $\dfrac{Y+2}{3} + 1 = \dfrac{Y}{2}$

Solución

(*a*) Sumar 20 a cada lado: $4a - 20 + 20 = 8 + 20$, o sea $4a = 28$.
Dividir ambos lados por 4: $4a/4 = 28/4$ y $a = 7$.
Comprobación: $4(7) - 20 = 8$, $28 - 20 = 8$ y $8 = 8$.

(*b*) Restar 4 de ambos miembros: $3X + 4 - 4 = 24 - 2X - 4$, o sea $3X = 20 - 2X$.
Sumar $2X$ a ambos lados: $3X + 2X = 20 - 2X + 2X$, o sea $5X = 20$.
Dividir por 5: $5X/5 = 20/5$ y $X = 4$.
Comprobación: $3(4) + 4 = 24 - 2(4)$, $12 + 4 = 24 - 8$ y $16 = 16$.

Puede obtenerse el resultado mucho más fácilmente dándose cuenta de que cada término puede ser trasladado de un miembro de la ecuación al otro sin más que cambiarle el signo. Así, podemos hacer

$$3X + 4 = 24 - 2X \qquad 3X + 2X = 24 - 4 \qquad 5X = 20 \qquad X = 4$$

(*c*) $18 - 5b = 3b + 24 + 10$ y $18 - 5b = 3b + 34$.
Trasponiendo, $-5b - 3b = 34 - 18$, o sea $-8b = 16$.
Dividiendo por -8, $-8b/(-8) = 16/(-8)$ y $b = -2$.
Comprobación: $18 - 5(-2) = 3(-2 + 8) + 10$, $18 + 10 = 3(6) + 10$ y $28 = 28$.

(*d*) Multiplicamos primero ambos lados por 6, el común denominador.

$$6\left(\frac{Y+2}{3} + 1\right) = 6\left(\frac{Y}{2}\right) \qquad 6\left(\frac{Y+2}{3}\right) + 6(1) = \frac{6Y}{2} \qquad 2(Y+2) + 6 = 3Y$$

$$2Y + 4 + 6 = 3Y \qquad 2Y + 10 = 3Y \qquad 10 = 3Y - 2Y \qquad Y = 10$$

Comprobación: $\dfrac{10+2}{3} + 1 = \dfrac{10}{2}, \dfrac{12}{3} + 1 = \dfrac{10}{2}, 4 + 1 = 5$ y $5 = 5$.

1.30. Resolver cada uno de los conjuntos de ecuaciones simultáneas:

(*a*) $3a - 2b = 11$
$5a + 7b = 39$

(*b*) $5X + 14Y = 78$
$7X + 3Y = -7$

(*c*) $3a + 2b + 5c = 15$
$7a - 3b + 2c = 52$
$5a + b - 4c = 2$

Solución

(*a*) Multiplicar la primera ecuación por 7: $21a - 14b = 77$ (1)
Multiplicar la segunda ecuación por 2: $10a + 14b = 78$ (2)
Sumar: $31a = 155$
Dividir por 31: $a = 5$

Nótese que al multiplicar cada ecuación por un número apropiado, somos capaces de escribir dos ecuaciones equivalentes, (1) y (2), en las que los coeficientes de la incógnita b son iguales, de modo que al sumar se elimina b y hallamos a.

Sustituimos $a = 5$ en la primera ecuación: $3(5) - 2b = 11$, $-2b = -4$ y $b = 2$. Así pues, $a = 5$ y $b = 2$.

Comprobación: $3(5) - 2(2) = 11$, $15 - 4 = 11$ y $11 = 11$; $5(5) + 7(2) = 39$, $25 + 14 = 39$ y $39 = 39$.

(*b*)

Multiplicar la primera ecuación por 3:	$15X + 42Y = 234$	(3)
Multiplicar la segunda ecuación por -14:	$-98X - 42Y = 98$	(4)
Sumar:	$-83X = 332$	
Dividir por -83:	$X = -4$	

Sustituimos $X = -4$ en la primera ecuación: $5(-4) + 14Y = 78$, $14Y = 98$ e $Y = 7$. Luego $X = -4$ e $Y = 7$.

Comprobación: $5(-4) + 14(7) = 78$, $-20 + 98 = 78$ y $78 = 78$; $7(-4) + 3(7) = -7$, $-28 + 21 = -7$ y $-7 = -7$.

(*c*)

Multiplicar la segunda por 2:	$6a + 4b + 10c = 77$	
Repetir la tercera ecuación por -5:	$-35a + 15b - 10c = -260$	
Sumar:	$-29a + 19b = -230$	(5)

Multiplicar la segunda por 2:	$14a - 6b + 4c = 104$	
Repetir la tercera ecuación:	$5a + b - 4c = 2$	
Sumar:	$19a - 5b = 106$	(6)

Así hemos eliminado c y nos quedan dos ecuaciones, (5) y (6), para deducir a y b.

Multiplicar la ecuación (5) por 5:	$-145a + 95b = -1150$
Multiplicar la ecuación (6) por 19:	$361a - 95b = 2014$
Sumar:	$216a = 864$
Dividir por 216:	$a = 4$

Sustituyendo $a = 4$ en (5) o (6) vemos que $b = -6$.
Sustituyendo $a = 4$ y $b = -6$ en alguna de las ecuaciones dadas, se obtiene $c = 3$.
Así pues, $a = 4$, $b = -6$ y $c = 3$.
Comprobación: $3(4) + 2(-6) + 5(3) = 15$ y $15 = 15$; $7(4) - 3(-6) + 2(3) = 52$ y $52 = 52$; $5(4) + (-6) - 4(3) = 2$ y $2 = 2$.

DESIGUALDADES

1.31. Expresar en palabras el significado de:

(*a*) $N > 30$ (*b*) $X \leqslant 12$ (*c*) $0 < p \leqslant 1$ (*d*) $\mu - 2t < X < \mu + 2t$

Solución

(*a*) N es mayor que 30.
(*b*) X es menor o igual que 12.

(*c*) p es mayor que 0, pero menor o igual que 1.
(*d*) X es mayor que $\mu - 2t$, pero menor que $\mu + 2t$.

1.32. Traducir lo que sigue en símbolos:

(*a*) La variable X tiene valores entre 2 y 5 inclusive.
(*b*) La media aritmética $\bar{X}$ es mayor que 28.42, pero menor que 31.56.
(*c*) m es un número positivo menor o igual que 10.
(*d*) P es un número no negativo.

Solución

(*a*) $2 \leqslant X \leqslant 5$; (*b*) $28.42 < \bar{X} < 31.56$; (*c*) $0 < m \leqslant 10$; (*d*) $P \geqslant 0$.

1.33. Usando símbolos de desigualdad, poner 3.42, −0.6, −2.1, 1.45 y −3 en: (*a*) orden creciente y (*b*) orden decreciente.

Solución

(*a*) $-3 < -2.1 < -0.6 < 1.45 < 3.42$
(*b*) $3.42 > 1.45 > -0.6 > -2.1 > -3$

Nótese que al marcar los puntos en una recta, crecen de izquierda a derecha.

1.34. Escribir como desigualdades en X (o sea, despejar X):

(*a*) $2X < 6$
(*b*) $3X - 8 \geqslant 4$
(*c*) $6 - 4X < -2$
(*d*) $-3 < \dfrac{X-5}{2} < 3$
(*e*) $-1 \leqslant \dfrac{3-2X}{5} \leqslant 7$

Solución

(*a*) Dividiendo ambos lados por 2 resulta $X < 3$.
(*b*) Sumando 8 a ambos lados, $3X \geqslant 12$; dividiendo ambos lados por 3, $X \geqslant 4$.
(*c*) Sumando −6 queda $-4X < -8$; dividiendo por −4, $X > 2$. Hagamos constar que, como en las ecuaciones, podemos pasar un término al otro lado sin más que cambiarle el signo. Por la parte (*b*), por ejemplo, $3X \geqslant 8 + 4$.
(*d*) Multiplicar por 2, $-6 < X - 5 < 6$; sumando 5, $-1 < X < 11$.
(*e*) Multiplicando por 5, $-5 \leqslant 3 - 2X \leqslant 35$; sumando −3, $-8 \leqslant -2X \leqslant 32$; dividiendo por −2, $4 \geqslant X \geqslant -16$, es decir $-16 \leqslant X \leqslant 4$.

LOGARITMOS Y ANTILOGARITMOS

1.35. Determinar la característica de los logaritmos comunes (base 10) de los números:

(*a*) 57	(*d*) 35.63	(*g*) 186,000	(*j*) 0.0325
(*b*) 57.4	(*e*) 982.5	(*h*) 0.71	(*k*) 0.0071
(*c*) 5.63	(*f*) 7824	(*i*) 0.7314	(*l*) 0.0003

Solución

(*a*) 1; (*b*) 1; (*c*) 0; (*d*) 1; (*e*) 2; (*f*) 3; (*g*) 5; (*h*) 9−10; (*i*) 9−10; (*j*) 8−10; (*k*) 7−10; (*l*) 6−10.

1.36. Calcular los siguientes logaritmos:

(*a*) log 87.2	(*f*) log 0.382	(*k*) log 4.638	(*p*) log 0.2548
(*b*) log 37,300	(*g*) log 0.00159	(*l*) log 6.753	(*q*) log 0.04372
(*c*) log 753	(*h*) log 0.0753	(*m*) log 183.2	(*r*) log 0.009848
(*d*) log 9.21	(*i*) log 0.000827	(*n*) log 43.15	(*s*) log 0.0001788
(*e*) log 54.50	(*j*) log 0.0503	(*o*) log 876,400	

Solución

(*a*) Mantisa = .9405, y característica = 1; de modo que log 87.2 = 1.9405; (*b*) 4.5717; (*c*) 2.8768; (*d*) 0.9643; (*e*) 1.7364; (*f*) Mantisa = .5821, y característica = 9 − 10; por tanto log 0.382 = 9.5821 − 10; (*g*) 7.2014 − 10; (*h*) 8.8768 − 10; (*i*) 6.9175 − 10; (*j*) 8.7016 − 10; (*k*) La mantisa de log 4638 está a 0.8 de camino entre la de log 4630 y la de log 4640.

$$\begin{aligned} \text{Mantisa de log } 4640 &= .6665 \\ \text{Mantisa de log } 4630 &= \underline{.6656} \\ \text{Diferencia tabular} &= .0009 \end{aligned}$$

La mantisa de log 4.638 = .6656 + (0.8)(.0009) = .6663 con cuatro dígitos; luego log 4.638 = .6663. Este proceso se llama interpolación lineal. Si se desea, la tabla de partes proporcionales del Apéndice VII permite deducir la mantisa directamente (6656 + 7).

(*l*) 0.8295 (8293 + 2); (*m*) 2.2630 (2625 + 5); (*n*) 1.6350 (6345 + 5); (*o*) 5.9427 (9425 + 2); (*p*) 9.4062 − 10 (4048 + 14); (*q*) 8.6407 − 10 (6405 + 2); (*r*) 7.9933 − 10 (9930 + 3); (*s*) 6.2524 − − 10 (2504 + 20).

1.37. Calcular los siguientes antilogaritmos:

(*a*) antilog 1.9058	(*c*) antilog 7.8657 − 10	(*f*) antilog 2.6715
(*b*) antilog 3.8531	(*d*) antilog 9.8267 − 10	antilog 4.1853
antilog 2.1875	antilog $\bar{2}.3927$	antilog 0.9245
antilog 0.4997	antilog 7.7443 − 10	(*g*) antilog $\bar{1}.6089$
antilog 4.9360	(*e*) antilog 9.3842 − 10	antilog 8.8907 − 10
		antilog 1.2000

Solución

(*a*) En el Apéndice VII la mantisa .9058 corresponde al número 805. Como la característica es 1, el número debe tener dos cifras delante del punto decimal; por tanto, es 80.5 (es decir, antilog 1.9058 = 80.5).

(*b*) antilog 3.8531 = 7130, antilog 2.1875 = 154, antilog 0.4997 = 3.16 y antilog 4.9360 = 86,300.

(*c*) En el Apéndice VII la mantisa .8657 corresponde al número 734. Como la característica es 7 − 10, el número tiene dos ceros tras el punto decimal. En consecuencia, el número es 0.00734 (o sea, antilog 7.8657 − 10 = 0.00734). La tabla de partes proporcionales del Apéndice VII la daría también.

(*d*) antilog 9.8267 − 10 = 0.671, antilog $\bar{2}.3927$ = 0.0247 y antilog 7.7443 − 10 = 0.00555.

(*e*) Como la mantisa no aparece en la tabla, hay que usar interpolación:

$$\begin{aligned} \text{Mantisa de log } 2430 &= .3856 \\ \text{Mantisa de log } 2420 &= \underline{.3838} \\ \text{Diferencia tabular} &= .0018 \end{aligned} \qquad \begin{aligned} \text{Mantisa dada} &= .3842 \\ \text{Mantisa inferior más próxima} &= \underline{.3838} \\ \text{Diferencia} &= .0004 \end{aligned}$$

Luego 2420 + (4/18)(2430 − 2420) = 2422 con cuatro dígitos, y el número pedido es 0.2422.

(*f*) antilog 2.6715 = 469.3 (3/9 × 10 = 3 aproximadamente), antilog 4.1853 = 15,320 (6/28 × 10 = 2 aproximadamente), y antilog 0.9245 = 8.404 (2/5 × 10 = 4).

(*g*) antilog $\bar{1}.6089$ = 0.4064 (4/11 × 10 = 4 aproximadamente), antilog 8.8907 − 10 = 0.07775 (3/6 × 10 = 5) y antilog 1.2000 = 15.85 (13/27 × 10 = 5 aproximadamente).

CALCULOS USANDO LOGARITMOS

Calcular cada una de las cantidades que siguen, usando logaritmos.

1.38. $P = (3.81)(43.4)$.

Solución

$\log P = \log 3.81 + \log 43.4$:

$$\begin{array}{rl} \log 3.81 & = 0.5809 \\ (+)\log 43.4 & = 1.6375 \\ \hline \log P & = 2.2184 \end{array}$$

Por tanto, P = antilog 2.2184 = 165.3, o sea, 165 con tres dígitos significativos. Nótese el significado del cálculo en exponenciales:

$$(3.81)(43.4) = (10^{0.5809})(10^{1.6375}) = 10^{0.5809+1.6375} = 10^{2.2184} = 165.3$$

1.39. $P = (73.42)(0.004620)(0.5143)$.

Solución

$\log P = \log 73.42 + \log 0.004620 + \log 0.5143$:

$$\begin{array}{rl} \log 73.42 & = 1.8685 \\ (+)\log 0.004620 & = 7.6646 - 10 \\ (+)\log 0.5143 & = 9.7112 - 10 \\ \hline \log P & = 19.2416 - 20 = 9.2416 - 10 \end{array}$$

Luego $P = 0.1744$.

1.40. $P = \dfrac{(784.6)(0.0431)}{28.23}$.

Solución

$\log P = \log 784.6 + \log 0.0431 - \log 28.23$:

$$\begin{array}{rl} \log 784.6 & = 2.8947 \\ (+)\log 0.0431 & = 8.6345 - 10 \\ \hline & \ 11.5292 - 10 \\ (-)\log 28.23 & = 1.4507 \\ \hline \log P & = 10.0785 - 10 = 0.0785 \end{array}$$

Así pues, P = 1.198, o sea 1.20 con tres dígitos significativos. En términos de exponenciales:

$$\frac{(784.6)(0.0431)}{28.23} = \frac{(10^{2.8947})(10^{8.6345-10})}{10^{1.4507}} = 10^{2.8947+8.6345-10-1.4507} = 10^{0.0785} = 1.198$$

1.41. $P = (5.395)^8$

Solución

$$\log P = 8 \log 5.395 = 8(0.7320) = 5.8560 \text{ y } P = 717{,}800, \text{ o sea } 7.178 \times 10^5.$$

1.42. $P = \sqrt{387.2} = (387.2)^{1/2}$.

Solución

$$\log P = \tfrac{1}{2} \log 387.2 = \tfrac{1}{2}(2.5879) = 1.2940 \text{ y } P = 19.68.$$

1.43. $P = (0.08317)^{1/5}$.

Solución

$$\log P = \tfrac{1}{5} \log 0.08317 = \tfrac{1}{5}(8.9200 - 10) = \tfrac{1}{5}(48.9200 - 50) = 9.7840 - 10 \text{ y } P = 0.6081.$$

1.44. $P = \dfrac{\sqrt{0.003654}(18.37)^3}{(8.724)^4 \sqrt[4]{743.8}}$.

Solución

$$\log P = \tfrac{1}{2} \log 0.003654 + 3 \log 18.37 - (4 \log 8.724 + \tfrac{1}{4} \log 743.8):$$

Numerador N

$$\begin{aligned}
\tfrac{1}{2} \log 0.003654 &= \tfrac{1}{2}(7.5628 - 10) \\
&= \tfrac{1}{2}(17.5628 - 20) = 8.7814 - 10 \\
3 \log 18.37 &= 3(1.2641) = 3.7923 \\
\text{Sumar:} \quad \log N &= 12.5737 - 10 \\
(-) \log D &= 4.4806 \\
\log P &= 8.0931 - 10 \\
P &= 0.01239
\end{aligned}$$

Denominador D

$$\begin{aligned}
4 \log 8.724 &= 4(0.9407) = 3.7628 \\
\tfrac{1}{4} \log 743.6 &= \tfrac{1}{4}(2.8714) = 0.7178 \\
\text{Sumar:} \quad \log D &= 4.4806
\end{aligned}$$

1.45. $P = \sqrt{\dfrac{(874.3)(0.03816)(28.53)^3}{(1.754)^4(0.007352)}}$.

Solución

$$\log P = \tfrac{1}{2}[\log 874.3 + \log 0.03816 + 3 \log 28.53 - (4 \log 1.754 + \log 0.007352)]:$$

$$\begin{aligned}
\log 874.3 &= 2.9417 &&= 2.9417 \\
\log 0.03816 &= 8.5816 - 10 &&= 8.5816 - 10 \\
3 \log 28.53 &= 3(1.4553) &&= 4.3659 \\
\text{Sumar:} & && 15.8892 - 10 \qquad (1) \\
4 \log 1.754 &= 4(0.2440) &&= 0.9760 \\
\log 0.007352 & &&= 7.8664 - 10 \\
\text{Sumar:} & && 8.8424 - 10 \qquad (2)
\end{aligned}$$

De (1) y (2) tenemos que

$$\log P = \tfrac{1}{2}[(15.8892 - 10) - (8.8424 - 10)] = \tfrac{1}{2}(7.0468) = 3.5234 \text{ y } P = 3338$$

PROBLEMAS SUPLEMENTARIOS

VARIABLES

1.46. Decir cuáles de los que siguen representan datos discretos y cuáles continuos:

C (*a*) Centímetros de lluvia en una ciudad durante varios meses. c
C (*b*) Velocidad de un coche (km/h). c
D (*c*) Número de billetes de \$20 en circulación en EE.UU. en cada momento.
C (*d*) Volumen de negocio diario en la Bolsa de Tokio.
D (*e*) Número de estudiantes matriculados en una Universidad en varios años.

1.47. Dar el dominio de cada variable y decir si son discretas o continuas:

(*a*) Número W de bushels de trigo producidos por acre en un campo en varios años.
(*b*) Número N de miembros en una familia. D
(*c*) Estado civil de una persona. D
(*d*) Tiempo de vuelo T de un misil. C
(*e*) Número P de pétalos de una flor. D

REDONDEO DE DATOS, NOTACION CIENTIFICA Y DIGITOS SIGNIFICATIVOS

1.48. Redondear cada número con la precisión indicada:

(*a*) 3256 centenas.
(*b*) 5.781 decenas.
(*c*) 0.0045 milésimas.
(*d*) 46.7385 centésimas.
(*e*) 125.9995 dos cifras decimales.
(*f*) 3,502,378 millones.
(*g*) 148.475 unidades.
(*h*) 0.000098501 millonésimas.
(*i*) 2184.73 decenas.
(*j*) 43.87500 centésimas.

1.49. Expresar cada número sin usar potencias de 10:

(*a*) 132.5×10^4
(*b*) 418.72×10^{-5}
(*c*) 280×10^{-7}
(*d*) 7300×10^6
(*e*) 3.487×10^{-4}
(*f*) 0.0001850×10^5

1.50. ¿Cuántos dígitos significativos hay en estos números, supuesto que se dan con la mayor precisión posible?

(*a*) 2.54 cm
(*b*) 0.004500 yd
(*c*) 3,510,000 bu
(*d*) 3.51 millones bu
(*e*) 10.000100 pies
(*f*) 378 personas
(*g*) 378 oz
(*h*) 4.50×10^{-3} km
(*i*) 500.8×10^5 kg
(*j*) 100.00 mi

1.51. ¿Cuál es el error máximo en cada una de las medidas siguientes, supuesto que se dan con la mayor precisión posible? Decir en cada caso el número de dígitos significativos.

(*a*) 7.20 millones bu
(*b*) 0.00004835 cm
(*c*) 5280 pies
(*d*) 3.0×10^8 m
(*e*) 186,000 mi/seg
(*f*) 186 miles mi/seg

1.52. Escribir estos números en notación científica, supuesto que todos son dígitos significativos salvo mención expresa en contra.

(*a*) 0.000317
(*b*) 428,000,000 (cuatro cifras significativas)
(*c*) 21,600.00
(*d*) 0.000009810
(*e*) 732 miles
(*f*) 18.0 diezmilésimas

CALCULOS

1.53. Probar que: (*a*) el producto y (*b*) el cociente de 72.48 y 5.16, supuesto que tienen cuatro y tres dígitos significativos, respectivamente, no admiten más de tres dígitos significativos. Escribir los resultados con la mejor precisión posible.

1.54. Efectuar cada operación, suponiendo que los números se dan en la mayor precisión posible.

(*a*) 0.36×781.4

(*b*) $\dfrac{873.00}{4.881}$

(*c*) $5.78 \times 2700 \times 16.00$

(*d*) $\dfrac{0.00480 \times 2300}{0.2084}$

(*e*) $\sqrt{120 \times 0.5386 \times 0.4614}$ (120 exacto)

(*f*) $\dfrac{(416{,}000)(0.000187)}{\sqrt{73.84}}$

(*g*) $14.8641 + 4.48 - 8.168 + 0.36125$

(*h*) $4{,}173{,}00 - 170{,}264 + 1{,}820{,}470 - 78{,}320$ (los números son exactos en, respectivamente, 4, 6, 6 y 5 cifras significativas)

(*i*) $\sqrt{\dfrac{7(4.386)^2 - 3(6.47)^2}{6}}$ (3, 6 y 7 son exactos)

(*j*) $4.120\sqrt{\dfrac{3.1416[(9.483)^2 - (5.075)^2]}{0.0001980}}$

1.55. Evaluar lo que sigue, sabiendo que $U = -2$, $V = \frac{1}{2}$, $W = 3$, $X = -4$, $Y = 9$ y $Z = \frac{1}{6}$, donde todos los números son exactos.

(*a*) $4U + 6V - 2W$

(*b*) $\dfrac{XYZ}{UVW}$

(*c*) $\dfrac{2X - 3Y}{UW + XV}$

(*d*) $3(U - X)^2 + Y$

(*e*) $\sqrt{U^2 - 2UV + W}$

(*f*) $3X(4Y + 3Z) - 2Y(6X - 5Z) - 25$

(*g*) $\sqrt{\dfrac{(W-2)^2}{V} + \dfrac{(Y-5)^2}{Z}}$

(*h*) $\dfrac{X - 3}{\sqrt{(Y - 4)^2 + (U + 5)^2}}$

(*i*) $X^3 + 5X^2 - 6X - 8$

(*j*) $\dfrac{U - V}{\sqrt{U^2 + V^2}}[U^2V(W + X)]$

FUNCIONES, TABLAS Y GRAFICOS

1.56. Una variable Y queda determinada por otra X mediante $Y = 10 - 4X$.

(*a*) Hallar Y tal que $X = -3, -2, -1, 0, 1, 2, 3, 4$ y 5, y poner los resultados en una tabla.

(*b*) Hallar Y tal que $X = -2.4, -1.6, -0.8, 1.8, 2.7, 3.5$ y 4.6.

(*c*) Si denotamos la dependencia entre X e Y por $Y = F(X)$, calcular $F(2.8)$, $F(-5)$, $F(\sqrt{2})$ y $F(-\pi)$.

(*d*) ¿Qué valor de X corresponde a $Y = -2, 6, -10, 1.6, 16, 0$ y 10?

(*e*) Expresar X explícitamente como función de Y.

1.57. Si $Z = X^2 - Y^2$, calcular Z cuando: (*a*) $X = -2$, $Y = 3$, y (*b*) $X = 1$, $Y = 5$. (*c*) En la notación funcional $Z = F(X, Y)$, cuando $F(-3, -1)$.

1.58. Si $W = 3XZ - 4Y^2 + 2XY$, calcular W cuando: (*a*) $X = 1$, $Y = -2$, $Z = 4$, y (*b*) $X = -5$, $Y = -2$, $Z = 0$. (*c*) Con la notación funcional $W = F(X, Y, Z)$, calcular $F(3, 1, -2)$.

1.59. Localizar en un sistema de coordenadas rectangulares los puntos de coordenadas: (*a*) (3, 2), (*b*) (2, 3), (*c*) −4, 4), (*d*) (4, −4), (*e*) (−3, −2), (*f*) (−2, −3), (*g*) (−4.5, 3), (*h*) (−1.2, −2.4), (*i*) (0, −3) y (*j*) (1.8, 0).

1.60. Representar las ecuaciones: (*a*) $Y = 10 - 4X$ (véase Prob. 1.56), (*b*) $Y = 2X + 5$, (*c*) $Y = \frac{1}{3}(X - 6)$, (*d*) $2X + 3Y = 12$ y (*e*) $3X - 2Y = 6$.

1.61. Representar las ecuaciones: (*a*) $Y = 2X^2 + X - 10$ y (*b*) $Y = 6 - 3X - X^2$.

1.62. Representar $Y = X^3 - 4X^2 + 12X - 6$.

1.63. La Tabla 1.8 muestra el número de trabajadores, agrícolas o no, en EE.UU. durante 1840-1980. Representar los datos usando: (*a*) gráfico de trazos, (*b*) gráfico de barras y (*c*) gráfico de barras en componentes.

Tabla 1.8

Año	Trabajadores agrícolas (millones)	Trabajadores no agrícolas (millones)
1840	3.72	1.70
1860	6.20	4.33
1880	8.59	8.80
1900	10.90	18.17
1920	11.46	30.97
1940	9.22	43.75
1960	4.19	65.70
1980	2.33	103.76

Fuente: U.S. Bureau of the Census.

1.64. Con los datos de la Tabla 1.8, diseñar un pictograma que muestre la variación en el número de trabajadores: (*a*) agrícolas y (*b*) no agrícolas. ¿Puede diseñar otro que las muestre a la vez?

1.65. Con los datos de la Tabla 1.8, construir un gráfico que muestre el porcentaje de trabajadores: (*a*) agrícolas y (*b*) no agrícolas. ¿Puede diseñar otro que las muestre a la vez?

1.66. La Tabla 1.9 da la expectativa de vida de un niño nacido en EE.UU. durante 1920-1980. Llevar los datos a un gráfico.

Tabla 1.9

Año	Varones	Hembras
1920	53.6	54.6
1930	58.1	61.6
1940	60.8	65.2
1950	65.6	71.1
1960	66.6	73.1
1970	67.1	74.7
1980	70.0	77.4

Fuente: National Center for Health Statistics.

1.67. La Tabla 1.10 recoge las velocidades orbitales de los planetas del sistema solar. Representar esos datos.

Tabla 1.10

Planeta	Velocidad (m/seg)
Mercurio	29.7
Venus	21.8
Tierra	18.5
Marte	15.0
Júpiter	8.1
Saturno	6.0
Urano	4.2
Neptuno	3.4
Plutón	3.0

1.68. En la Tabla 1.11 se ven los números (en millones) de estudiantes en enseñanza elemental, media y superior («colleges») en EE.UU. Representar los datos usando: (*a*) gráficos de trazos, (*b*) gráficos de barras y (*c*) gráficos de barras en componentes.

Tabla 1.11

Año	1960	1965	1970	1975	1980
Elemental	32.4	35.5	37.1	33.8	30.6
Media	10.2	13.0	14.7	15.7	14.6
Superior	3.6	5.7	7.4	9.7	10.2

Fuente: U.S. Bureau of the Census.

1.69. Representar los datos de la Tabla 1.11 en un gráfico de porcentajes en componentes.

1.70. La Tabla 1.12 muestra el estado civil de hombres y mujeres (de más de 18 años) en EE.UU. en 1983. Representar los datos mediante: (*a*) dos gráficos circulares de igual diámetro y (*b*) un gráfico de diseño propio.

Tabla 1.12

Estado civil	Varones (% total)	Hembras (% total)
Soltero	25.1	18.4
Casado	66.7	61.3
Viudo	2.4	12.4
Divorciado	5.8	7.9

Fuente: U.S. Bureau of the Census.

1.71. En la Tabla 1.13 figuran las declaraciones de quiebra habidas en EE.UU. en 1975-1986. Representar los datos usando gráficos adecuados.

Tabla 1.13

Año	Total de declaraciones de quiebra
1975	11,432
1976	9,628
1977	7,919
1978	6,619
1979	7,564
1980	11,742
1981	16,794
1982	24,908
1983	31,334
1984	52,078
1985	57,252
1986	61,183

Fuente: Survey of Current Business.

1.72. La Tabla 1.14 recoge la relación entre divorcios y bodas en EE.UU. durante 1900-1980. Representar los datos en dos tipos de gráficos.

Tabla 1.14

Año	Relación entre divorcios y bodas
1900	0.079
1910	0.088
1920	0.134
1930	0.174
1940	0.165
1950	0.231
1960	0.258
1970	0.328
1980	0.491

Fuente: U.S. Department of Health and Human Services.

1.73. La Tabla 1.15 da, redondeados al millón, los paises más poblados en 1986. Representar los datos por dos métodos diferentes.

Tabla 1.15

País	Población (millones)
China	1038
India	768
U.R.S.S.	278
EE.UU.	239
Indonesia	173
Brasil	135
Japón	121

Fuente: Naciones Unidas.

1.74. Representar los datos de la Tabla 1.15 teniendo en cuenta que la población mundial era en 1986 de 4850 millones.

1.75. En la Tabla 1.16 se ven las áreas de los océanos en millones de millas cuadradas. Representar los datos usando: (*a*) un gráfico de barras y (*b*) un gráfico circular.

Tabla 1.16

Océano	Area (millones de millas cuadradas)
Pacífico	63.8
Atlántico	31.5
Indico	28.4
Antártico	7.6
Artico	4.8

Fuente: Naciones Unidas.

ECUACIONES

1.76. Resolver las ecuaciones:

(*a*) $16 - 5c = 36$

(*b*) $2Y - 6 = 4 - 3Y$

(*c*) $4(X - 3) - 11 = 15 - 2(X + 4)$

(*d*) $3(2U + 1) = 5(3 - U) + 3(U - 2)$

(*e*) $3[2(X + 1) - 4] = 10 - 5(4 - 2X)$

(*f*) $\frac{2}{5}(12 + Y) = 6 - \frac{1}{4}(9 - Y)$

1.77. Resolver las ecuaciones simultáneas:

(*a*) $2a + b = 10$
$7a - 3b = 9$

(*b*) $3a + 5b = 24$
$2a + 3b = 14$

(*c*) $8X - 3Y = 2$
$3X + 7Y = -9$

(*d*) $5A - 9B = -10$
$3A - 4B = 16$

(*e*) $2a + b - c = 2$
$3a - 4b + 2c = 4$
$4a + 3b - 5c = -8$

(*f*) $5X + 2Y + 3Z = -5$
$2X - 3Y - 6Z = 1$
$X + 5Y - 4Z = 22$

(*g*) $3U - 5V + 6W = 7$
$5U + 3V - 2W = -1$
$4U - 8V + 10W = 11$

1.78. (*a*) Representar las ecuaciones $5X+2Y=4$ y $7X - 3Y = 23$, usando el mismo sistema coordenado.
(*b*) Determinar, con tales gráficos, la solución simultánea de ambas ecuaciones.
(*c*) Repetir las partes (*a*) y (*b*) para las ecuaciones simultáneas (*a*)-(*d*) del Problema 1.77.

1.79. (*a*) Usar el gráfico del Problema 1.61(*a*) para resolver la ecuación $2X^2 + X - 10 = 0$. (*Ayuda:* Hallar los valores de X en que la parábola corta al eje X, es decir, donde $Y = 0$.)
(*b*) Por el método de la parte (*a*), resuélvase $3X^2 - 4X - 5 = 0$.

1.80. Las soluciones de la ecuación cuadrática $aX^2 + bX + c = 0$ vienen dadas por la fórmula cuadrática:

$$X = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Usarla para hallar las soluciones de: (*a*) $3X^2 - 4X - 5 = 0$, (*b*) $2X^2 + X - 10 = 0$, (*c*) $5X^2 + 10X = 7$ y (*d*) $X^2 + 8X + 25 = 0$.

DESIGUALDADES

1.81. Usando símbolos de desigualdad, poner los números -4.3, -6.15, 2.37, 1.52 y -1.5 en orden: (*a*) creciente y (*b*) decreciente.

1.82. Expresar con símbolos de desigualdad las afirmaciones siguientes:

(*a*) N está entre 30 y 50 inclusive.
(*b*) S no es menor que 7.
(*c*) X es mayor o igual que -4, pero menor que 3.
(*d*) P es a lo sumo 5.
(*e*) X sobrepasa a Y en al menos 2.

1.83. Resolver las desigualdades:

(*a*) $3X \geqslant 12$
(*b*) $4X < 5X - 3$
(*c*) $2N + 15 > 10 + 3N$
(*d*) $3 + 5(Y - 2) \leqslant 7 - 3(4 - Y)$
(*e*) $-3 \leqslant \frac{1}{5}(2X + 1) \leqslant 3$
(*f*) $0 < \frac{1}{2}(15 - 5N) \leqslant 12$
(*g*) $-2 \leqslant 3 + \frac{1}{2}(a - 12) < 8$

LOGARITMOS Y ANTILOGARITMOS

1.84. Hallar los logaritmos comunes de:

(*a*) 387
(*b*) 0.387
(*c*) 0.0792
(*d*) 14,630
(*e*) 0.6042
(*f*) 0.002795
(*g*) 476.3
(*h*) 1.007
(*i*) 7.146
(*j*) 71.46
(*k*) 0.00098
(*l*) 84,620,000

1.85. Hallar los antilogaritmos de:

(*a*) 3.5611
(*b*) 9.8293 − 10
(*c*) 1.7045
(*d*) 8.9266 − 10
(*e*) 2.4700
(*f*) 6.4700 − 10
(*g*) $\bar{2}.8003$
(*h*) 3.7072
(*i*) 0.0800
(*j*) 6.3841

1.86. Evaluar mediante logaritmos:

(*a*) (783.6)(1654)

(*b*) $\dfrac{21.7}{378.2}$

(*c*) $\dfrac{(0.04556)(624.1)}{(14.32)(0.003572)}$

(*d*) $(1.562)^{15}$

(*e*) $\dfrac{(0.3854)^4(12.48)^2}{(0.04382)^3}$

(*f*) $0.04182\sqrt{0.6758}$

(*g*) $\sqrt[3]{3728}$

(*h*) $\sqrt[5]{(21.63)(33.81)(47.53)(65.28)(87.47)}$

(*i*) $\sqrt{\dfrac{(48.79)(0.00574)^3}{(2.143)^5}}$

(*j*) $\dfrac{3.781}{0.01873}\sqrt{\dfrac{(43.25)(0.08743)}{(0.002356)(6.824)}}$

1.87. Representar: (*a*) $Y = \log X$ y (*b*) $Y = 10^X$ y discutir las analogías entre ambos gráficos.

1.88. Escribir sin usar logaritmos las ecuaciones: (*a*) $2 \log X - 3 \log Y = 2$ y (*b*) $\log Y + 2X = \log 3$.

1.89. Si $a^p = N$, donde a y p son positivos y $a \neq 1$, llamamos a p el *logaritmo de N en base a*, y escribimos $p = \log_a N$. Evaluar: (*a*) $\log_2 8$, (*b*) $\log_{25} 125$, (*c*) $\log_4 1/16$, (*d*) $\log_{1/2} 32$ y (*e*) $\log_5 1$.

1.90. Probar que $\log_e N = 2.303 \log_{10} N$, aproximadamente, donde $e = 2.71828...$ se llama *base natural* de logaritmos y donde $N > 0$.

1.91. Probar que $(\log_b a)(\log_a b) = 1$, donde $a > 0$, $b > 0$, $a \neq 1$ y $b \neq 1$.

CAPITULO 2

Distribuciones de frecuencias

FILAS DE DATOS

Una *fila de datos* consiste en datos recogidos que no han sido organizados numéricamente, por ejemplo, las alturas de 100 estudiantes por letra alfabética.

ORDENACIONES

Una *ordenación* es un conjunto de datos numéricos en orden creciente o decreciente. La diferencia entre el mayor y el menor se llama *rango* de ese conjunto de datos. Así, si la mayor altura de entre los 100 estudiantes era de 74 in y la menor de 60 in, el rango es $74 - 60 = 14$ in.

DISTRIBUCIONES DE FRECUENCIAS

Al resumir grandes colecciones de datos, es útil distribuirlos en *clases* o *categorías*, y determinar el número de individuos que pertenecen a cada clase, llamado *frecuencia de clase*. Una disposición tabular de los datos por clases junto con las correspondientes frecuencias de clase, se llama *distribución de frecuencias* (o *tabla de frecuencias*). La Tabla 2.1 es una distribución de frecuencias de alturas (con precisión de 1 pulgada) de 100 estudiantes varones de la Universidad XYZ.

Tabla 2.1. Alturas de 100 estudiantes varones de la Universidad XYZ

Altura (in)	Número de estudiantes
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
	Total 100

La primera clase (o categoría), por ejemplo, consta de las alturas entre 60 y 62 in, y se indica por el rango 60-62. Como hay 5 estudiantes en esta clase, la correspondiente frecuencia de clase es 5.

Los datos así organizados en clases como en la anterior distribución de frecuencias se llaman *datos agrupados*. Aunque el proceso de agrupamiento destruye en general detalles de los datos iniciales, es muy ventajosa la visión nítida obtenida y las relaciones evidentes que saca a la luz.

INTERVALOS DE CLASE Y LIMITES DE CLASE

El símbolo que define una clase, como el 60-62 en la Tabla 2.1, se llama un *intervalo de clase*. Los números extremos, 60 y 62, se llaman *límite inferior de clase* (60) y *límite superior de clase* (62). Con frecuencia se intercambian los términos clase e intervalo de clase, aunque el intervalo de clase es un símbolo para la clase.

Un intervalo de clase que, al menos en teoría, carece de límite superior o inferior indicado, se llama *intervalo de clase abierto*. Por ejemplo, refiriéndonos a edades de personas, la clase «65 años o más» es un intervalo de clase abierto.

FRONTERAS DE CLASE

Si se dan alturas con precisión de 1 pulgada, el intervalo de clase 60-62 incluye teóricamente todas las medidas desde 59.5000 a 62.5000. in. Estos números, indicados más brevemente por los números exactos 59.5 y 62.5, se llaman *fronteras de clase* o verdaderos límites de clase; el menor (59.5) es la *frontera inferior* y el mayor (62.5) la *frontera superior*.

En la práctica, las fronteras de clase se obtienen promediando el límite superior de una clase con el inferior de la siguiente.

A veces se usan las fronteras de clase como símbolos para la clase. Así, las clases de la primera columna de la Tabla 2.1 se pueden indicar por 59.5-62.5, 62.5-65.5, etc. Para evitar ambigüedad en tal notación, las fronteras no deben coincidir con valores realmente medidos. De modo que si una observación diera 62.5, no sería posible decidir si pertenece al intervalo de clase 59.5-62.5 o al 62.5-65.5.

TAMAÑO O ANCHURA DE UN INTERVALO DE CLASE

El *tamaño* o *anchura de un intervalo de clase* es la diferencia entre las fronteras de clase superior e inferior. Si todos los intervalos de clase de una distribución de frecuencias tienen la misma anchura, la denotaremos por c. En tal caso, c es igual a la diferencia entre dos límites inferiores (o superiores) de clases sucesivas. Para los datos de la Tabla 2.1, por ejemplo, la anchura del intervalo de clase es $c = 62.5 - 59.5 = 65.5 - 62.5 = 3$.

MARCA DE CLASE

La *marca de clase* es el punto medio del intervalo de clase y se obtiene promediando los límites inferior y superior de clase. Así que las marcas de clase del intervalo 60-62 es $(60 + 62)/2 = 61$. La marca de clase se denomina también *punto medio de la clase*.

A efectos de análisis subsiguientes, todas las observaciones pertenecientes a un mismo intervalo de clase se supone que coinciden con la marca de clase. De manera que todas las alturas en el intervalo de clase 60-62 in se considerarán de 61 in.

REGLAS GENERALES PARA FORMAR DISTRIBUCIONES DE FRECUENCIAS

1. Determinar el mayor y el menor de todos los datos, hallando así el rango (diferencia entre ambos). 35
2. Dividir el rango en un número adecuado de intervalos de clase del mismo tamaño. Si ello no es factible, usar intervalos de clase de distintos tamaños o intervalos de clase abiertos (véase Problema 2.12). Se suelen tomar entre 5 y 20 intervalos de clase, según los datos. Los intervalos de clase se eligen también de modo tal que las marcas de clase (o puntos medios) coincidan con datos realmente observados. Ello tiende a disminuir el llamado *error de agrupamiento* que se produce en análisis ulteriores. No obstante, las fronteras de clase no debieran coincidir con datos realmente observados.
3. Determinar el número de observaciones que caen dentro de cada intervalo de clase; esto es, hallar las frecuencias de clase. Esto se logra mejor con una *hoja de recuentos* (véase Prob. 2.8).

HISTOGRAMAS Y POLIGONOS DE FRECUENCIAS

Los histogramas y los polígonos de frecuencias son dos representaciones gráficas de las distribuciones de frecuencias.

1. Un *histograma* o *histograma de frecuencias*, consiste en un conjunto de rectángulos con: (*a*) bases en el eje X horizontal, centros en las marcas de clase y longitudes iguales a los tamaños de los intervalos de clase y (*b*) áreas proporcionales a las frecuencias de clase.

   Si los intervalos de clase tienen todos la misma anchura, las alturas de los rectángulos son proporcionales a las frecuencias de clase, y entonces es costumbre tomar las alturas iguales a las frecuencias de clase. En caso contrario, deben ajustarse las alturas (véase Problema 2.13).

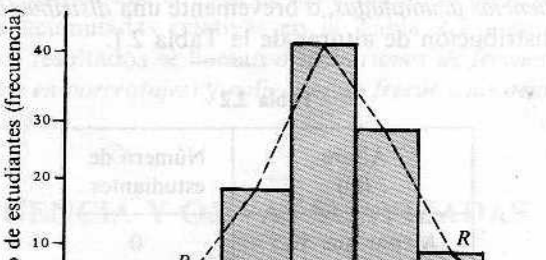


Figura 2.1.

frec. VARIABLE

2. Un *polígono de frecuencias* es un gráfico de trozos de la frecuencia de clase con relación a la marca de clase. Puede obtenerse conectando los puntos medios de las partes superiores de los rectángulos del histograma.

Histograma y polígono de frecuencias correspondientes a la distribución de frecuencias de alturas en la Tabla 2.1 se indican sobre los mismos ejes en la Figura 2.1. Suelen añadirse las longitudes PQ y RS a las marcas de clase extremas como asociadas a una frecuencia de clase cero. En tal caso, la suma de las áreas de los rectángulos del histograma es igual al área total limitada por el polígono de frecuencias y el eje X (véase Prob. 2.11).

DISTRIBUCIONES DE FRECUENCIAS RELATIVAS

La *frecuencia relativa de una clase* es su frecuencia dividida por la frecuencia total de todas las clases y se expresa generalmente como un porcentaje. Por ejemplo, la frecuencia relativa de la clase 66-68 en la Tabla 2.1 es $42/100 = 42\%$. La suma de las frecuencias relativas de todas las clases da obviamente 1, o sea 100 por 100.

Si se sustituyen las frecuencias de la Tabla 2.1 por las correspondientes frecuencias relativas, la tabla resultante se llama una *distribución de frecuencias relativas, distribución de porcentajes o tablas de frecuencias relativas.*

La representación gráfica de distribuciones de frecuencias relativas se puede obtener del histograma o del polígono de frecuencias sin más que cambiar la escala vertical de frecuencias a frecuencias relativas, manteniendo exactamente el mismo diagrama. Los gráficos resultantes se llaman *histogramas de frecuencias relativas* (o *histogramas de porcentajes*) y *polígonos de frecuencias relativas* (o *polígonos de porcentajes*), respectivamente.

DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS Y OJIVAS

La frecuencia total de todos los valores menores que la frontera de clase superior de un intervalo de clase dado se llama *frecuencia acumulada* hasta ese intervalo de clase inclusive. Por ejemplo, la frecuencia acumulada hasta el intervalo de clase 66-68 en la Tabla 2.1 es $5 + 18 + 42 = 65$, lo que significa que 65 estudiantes tienen alturas por debajo de 68.5 in.

Una tabla que presente tales frecuencias acumuladas se llama una *distribución de frecuencias acumuladas, tabla de frecuencias acumuladas*, o brevemente una *distribución acumulada*, y se muestra en la Tabla 2.2 para la distribución de alturas de la Tabla 2.1.

Tabla 2.2

Altura (in)	Número de estudiantes
Menor que 59.5	0
Menor que 62.5	5
Menor que 65.5	23
Menor que 68.5	65
Menor que 71.5	92
Menor que 74.5	100

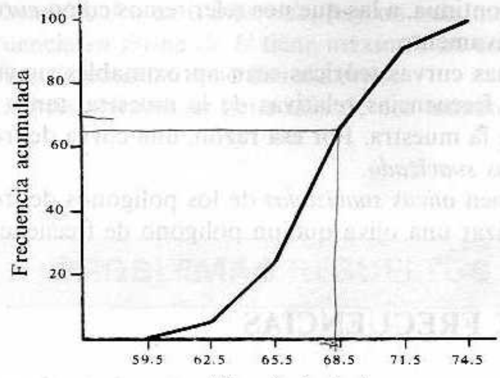


Figura 2.2.

Un gráfico que recoja las frecuencias acumuladas por debajo de cualquiera de las fronteras de clase superiores respecto de dicha frontera se llama un *polígono de frecuencias acumuladas* u *ojiva*, y se ilustra en la Figura 2.2 para las alturas de estudiantes de la Tabla 2.1.

A ciertos efectos, es deseable considerar una distribución de frecuencias acumuladas de todos los valores mayores o iguales que la frontera de clase inferior de cada intervalo de clase. Como eso hace considerar alturas de 59.5 in o más, de 62.5 in o más, etc., se le suele llamar una *distribución acumulada «o más»*, mientras que la antes considerada es una *distribución acumulada «menor que»*. Es fácil deducir una de otra (véase Prob. 2.15). Las correspondientes ojivas se conocen con los mismos apodos. Siempre que nos refiramos a distribuciones acumuladas u ojivas sin más, estaremos hablando del caso «menor que».

DISTRIBUCIONES DE FRECUENCIAS RELATIVAS Y OJIVAS DE PORCENTAJES

La *frecuencia acumulada relativa* o *frecuencia acumulada en porcentajes*, es la frecuencia acumulada dividida por la frecuencia total. Así, la frecuencia acumulada relativa de alturas menores que 68.5 in es $65/100 = 65\%$, lo que significa que el 65% de los estudiantes mide menos de 68.5 in.

Si se usan frecuencias acumuladas relativas en la Tabla 2.2 y en la Figura 2.2 en vez de frecuencias acumuladas, los resultados se llaman *distribuciones de frecuencias acumuladas relativas* (o *distribuciones acumuladas en porcentajes*) y *polígonos de frecuencias acumuladas relativas* (u *ojivas de porcentajes*), respectivamente.

CURVAS DE FRECUENCIA Y OJIVAS SUAVIZADAS

Los datos recogidos pueden considerarse usualmente como pertenecientes a una muestra de una población grande. Ya que son posibles muchas observaciones sobre esa población, es teóricamente posible (para datos continuos) escoger intervalos de clase muy pequeños y tener todavía números razonables de observaciones en cada clase. Así que cabe esperar que el polígono de frecuencias o el polígono de frecuencias relativas para una gran población tenga tantos pequeños segmentos que aparezca como casi una curva continua, a las que nos referiremos como *curva de frecuencias* o *curva de frecuencias relativas*, respectivamente.

Es sensato esperar que dichas curvas teóricas sean aproximables suavizando los polígonos de frecuencias o los polígonos de frecuencias relativas de la muestra, tanto mejor la aproximación cuanto mayor sea el tamaño de la muestra. Por esa razón, una curva de frecuencias se cita a veces como un *polígono de frecuencias suavizado*.

De forma análoga, se obtienen *ojivas suavizadas* de los polígonos de frecuencias acumuladas u ojivas. Suele ser más fácil suavizar una ojiva que un polígono de frecuencias (véase Prob. 2.18).

TIPOS DE CURVAS DE FRECUENCIAS

Las curvas de frecuencia que aparecen, en la práctica adoptan ciertas formas características, como ilustra la Figura 2.3.

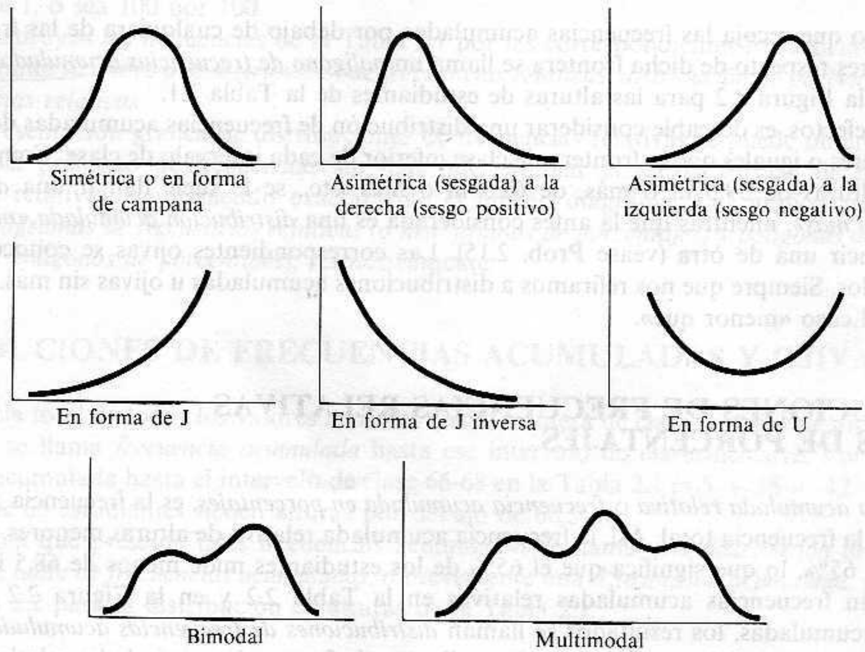


Figura 2.3.

1. Las curvas de frecuencias *simétricas* o *en forma de campana*, se caracterizan porque las observaciones equidistantes del máximo central tienen la misma frecuencia. Ejemplo importante es la curva normal.
2. En las curvas de frecuencia *poco asimétricas*, o *sesgadas*, la cola de la curva a un lado del máximo central es más larga que al otro lado. Si la cola mayor está a la derecha, la curva se dice *asimétrica a la derecha* o de *asimetría positiva*. En caso contrario, se dice *asimétrica a la izquierda* o de *asimetría negativa*.

3. En una curva *en forma de J* o de *J invertida,* hay un máximo en un extremo.
4. Una curva de frecuencia *en forma de U* tiene máximos en ambos extremos.
5. Una curva de frecuencia *bimodal* tiene dos máximos.
6. Una curva de frecuencia *multimodal* tiene más de dos máximos.

PROBLEMAS RESUELTOS

ORDENACIONES

2.1. (*a*) Disponer los números 17, 45, 38, 27, 6, 48, 11, 57, 34 y 22 en lista ordenada.
(*b*) Determinar el rango de esos números.

Solución

(*a*) En orden creciente: 6, 11, 17, 22, 27, 34, 38, 45, 48, 57. En orden decreciente: 57, 48, 45, 38, 34, 27, 22, 17, 11, 6.

(*b*) El menor es 6 y el mayor 57, luego el rango es $57 - 6 = 51$.

2.2. Las calificaciones finales en Matemáticas de 80 estudiantes figuran en la tabla adjunta.

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Hallar en esa tabla:

(*a*) La calificación más alta. 97
(*b*) La más baja. 53
(*c*) El rango. (44)
(*d*) Las cinco más altas. 97-96-94-95-95
(*e*) Las cinco más bajas. 53-59
(*f*) La décima de mayor a menor.
(*g*) El número de estudiantes con calificaciones de 75 o más.
(*h*) Idem por debajo de 85.
(*i*) El porcentaje de estudiantes con calificaciones mayores que 65 pero no superiores a 85.
(*j*) Las calificaciones que no aparecen.

Solución

Algunas de estas cuestiones son tan de detalle que se contestan mejor en una ordenación, lo cual se hace subdividiendo los datos en clases y colocando cada número de la tabla en su clase, como en la Tabla 2.3, llamada *tabla de entrada única*. Ordenando entonces los de cada clase, como en la Tabla 2.4 es fácil deducir las respuestas a las cuestiones planteadas.

(*a*) 97.
(*b*) 53.
(*c*) Rango = 97 − 53 = 44.
(*d*) Las cinco más altas son 97, 96, 95, 95 y 94.
(*e*) Las cinco más bajas son 53, 57, 59, 60 y 60.
(*f*) 88.

Tabla 2.3

50-54	53 1
55-59	59, 57 2
60-64	62, 60, 61, 62, 63, 60, 61, 60, 62, 62, 63 11
65-69	68, 68, 65, 66, 69, 68, 67, 65, 65, 67 10
70-74	73, 73, 71, 74, 72, 74, 71, 71, 73, 74, 73, 72 12
75-79	75, 76, 79, 75, 75, 78, 78, 75, 77, 78, 75, 79, 79, 78, 76, 75, 78, 76, 76, 75, 77 = 21
80-84	84, 82, 82, 83, 80, 81 6
85-89	88, 88, 85, 87, 89, 85, 88, 86, 85
90-94	90, 93, 93, 94
95-99	95, 96, 95, 97

Tabla 2.4

50-54	53
55-59	57, 59
60-64	60, 60, 60, 61, 61, 62, 62, 62, 62, 63, 63
65-69	65, 65, 65, 66, 67, 67, 68, 68, 68, 69
70-74	71, 71, 71, 72, 72, 73, 73, 73, 73, 74, 74, 74
75-79	75, 75, 75, 75, 75, 75, 75, 76, 76, 76, 76, 77, 77, 78, 78, 78, 78, 78, 79, 79, 79
80-84	80, 81, 82, 82, 83, 84
85-89	85, 85, 85, 86, 87, 88, 88, 88, 89
90-94	90, 93, 93, 94
95-99	95, 95, 96, 97

(*g*) 44 estudiantes.
(*h*) 63 estudiantes.
(*i*) El porcentaje 85 es 49/80 = 61.2%.
(*j*) No aparecen 0, 1, 2, 3, ..., 52, 54, 55, 56, 58, 64, 70, 91, 92, 98, 99 y 100.

DISTRIBUCIONES DE FRECUENCIA, HISTOGRAMAS Y POLIGONOS DE FRECUENCIAS

2.3. La Tabla 2.5 muestra una distribución de frecuencia de los salarios semanales de 65 empleados de la empresa P&R. Determinar de esa tabla:

(*a*) El límite inferior de la sexta clase.
(*b*) El límite superior de la cuarta clase.
(*c*) La marca de clase (o punto medio) de la tercera clase.
(*d*) Las fronteras de clase del quinto intervalo.

(*e*) La anchura del quinto intervalo de clase.
(*f*) La frecuencia de la tercera clase.
(*g*) La frecuencia relativa de la tercera clase.
(*h*) El intervalo de clase con máxima frecuencia, que se llama *intervalo de clase modal*. Su frecuencia es la *frecuencia de clase modal*.
(*i*) El porcentaje de empleados que cobran menos de $280.00 a la semana.
(*j*) El porcentaje de empleados que cobran menos de $300.00 pero al menos $260.00 por semana.
(K) Moda

Tabla 2.5

Salarios	Número de empleados
$250.00-$259.99	8
260.00-269.99	10
270.00-279.99	16
280.00-289.99	14
290.00-299.99	10
300.00-309.99	5
310.00-319.99	2
Total	65

10 - 100
16 - X

Solución

(*a*) $300.00.
(*b*) $289.99.
(*c*) La marca de clase de la tercera clase = $\frac{1}{2}$($270.00 + $279.99) = $274.995. A efectos prácticos se redondeará a $275.00.
(*d*) La frontera de clase inferior de la quinta clase = $\frac{1}{2}$($290.00 + $289.99) = $289.995. La superior = $\frac{1}{2}$($299.99 + $300.00) = $299.995.
(*e*) Anchura del quinto intervalo de clase = frontera superior de la quinta clase − frontera inferior de la quinta clase = $299.995 − $289.985 = $10.00. En este caso, todos los intervalos de clase son de la misma anchura: $10.00.
(*f*) 16.
(*g*) 16/65 = 0.246 = 24.6%.
(*h*) $270.00-$279.99.
(*i*) Número de empleados que ganan menos de $280 por semana = 16 + 10 + 8 = 34. Porcentaje de empleados que ganan menos de $280 por semana = 34/65 = 52.3%.
(*j*) Número de empleados que cobran menos de $300.00 pero al menos $260 por semana = 10 + + 14 + 16 + 10 = 50. Porcentaje de empleados que cobran menos de $300.00 pero al menos $260 por semana = 50/65 = 76.9%.
(K) $ 277.5

2.4. Si las marcas de clase en una distribución de frecuencias de pesos de estudiantes son 128, 137, 146, 155, 164, 173 y 182 libras (lb), hallar: (*a*) la anchura del intervalo de clase, (*b*) las fronteras de clase y (*c*) los límites de clase, suponiendo que los pesos se midieron con 1 libra de precisión.

Solución

(*a*) Anchura del intervalo de clase = diferencia común entre marcas de clase sucesivas = 137 − 128 = = 146 − 137 = etc. = 9 lb.

(*b*) Como los intervalos de clase son de igual anchura, las fronteras de clase están a mitad de camino entre las marcas de clase, luego son

$$\tfrac{1}{2}(128 + 137),\ \tfrac{1}{2}(137 + 146),\ \ldots,\ \tfrac{1}{2}(173 + 182) \quad \text{o sea} \quad 132.5,\ 141.5,\ 150.5,\ \ldots,\ 177.5 \text{ lb}$$

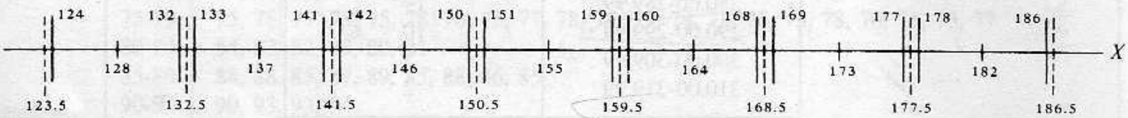
La primera frontera de clase es $132.5 - 9 = 123.5$ y la última $177.5 + 9 = 186.5$, ya que la anchura común de los intervalos de clase es 9 lb. Así pues, las fronteras de clase son

$$123.5,\ 132,\ 141.5,\ 150.5,\ 159.5,\ 168.5,\ 177.5,\ 186.5 \text{ lb}$$

(*c*) Como los límites de clase son enteros, los elegimos como los enteros más cercanos a las fronteras de clase, a saber, 123, 124, 132, 133, 141, 142, ... Luego la primera clase tiene límites 124-132, la siguiente 133-141, etc.

2.5. Representar gráficamente los resultados del Problema 2.4.

Solución



El gráfico se ve en el diagrama adjunto. Las marcas de clase 128, 137, 146, ..., 182 están localizadas en el eje X. Las fronteras de clase se indican por los segmentos verticales discontinuos, y los límites de clase por segmentos verticales sólidos.

2.6. La menor de 150 medidas es 5.18 in y la mayor 7.44 in. Determinar un conjunto apropiados de: (*a*) intervalos de clase, (*b*) fronteras de clase y (*c*) marcas de clase que puedan usarse para formar la distribución de frecuencias de esas medidas.

Solución

El rango es $7.44 - 5.18 = 2.26$ in. Para un mínimo de cinco intervalos de clase, la anchura de estos es $2.26/5 = 0.45$ aproximadamente; y para un máximo de 20 intervalos de clase la anchura es $2.26/20 = 0.11$ aproximadamente. Elecciones convenientes de la anchura de los intervalos de clase están entre 0.11 y 0.45, es decir, podrían ser 0.20, 0.30 ó 0.40.

(*a*) Las columnas I, II y III de la tabla adjunta muestran intervalos de clase de anchuras 0.20, 0.30 y 0.40, respectivamente.

a= 0.20 I	a=0.30 II	a=0.40 III
5.10-5.29	5.10-5.39	5.10-5.49
5.30-5.49	5.40-5.69	5.50-5.89
5.50-5.69	5.70-5.99	5.90-6.29
5.70-5.89	6.00-6.29	6.30-6.69
5.90-6.09	6.30-6.59	6.70-7.09
6.10-6.29	6.60-6.89	7.10-7.49
6.30-6.49	6.90-7.19	
6.50-6.69	7.20-7.49	
6.70-6.89		
6.90-7.09		
7.10-7.29		
7.30-7.49		

Nótese que el límite inferior de cada primera clase podría haber sido distinto de 5.10; por ejemplo, si en la columna I hubiéramos partido de 5.15 como límite inferior, el primer intervalo de clase hubiera sido 5.15-5.34.

(*b*) Las fronteras de clase correspondientes a las columnas I, II y III de la parte (*a*) vienen dadas, respectivamente, por

I 5.095-5.295, 5.295-5.495, 5.495-5.695, ..., 7.295-7.495
II 5.095-5.395, 5.395-5.695, 5.695-5.995, ..., 7.195-7.495
III 5.095-5.495, 5.495-5.895, 5.895-6.295, ..., 7.095-7.495

Obsérvese que tales fronteras de clase son correctas, pues no coinciden con medidas obtenidas.

(*c*) Las marcas de clase correspondientes a las columnas I, II y III de (*a*) son

I 5.195, 5.395, ..., 7.395 II 5.245, 5.545, ..., 7.345 III 5.295, 5.695, ..., 7.295

Estas marcas de clase tienen la desventaja de no coincidir con medidas observadas.

2.7. Al contestar el Problema 2.6(*a*), un estudiante escogió los intervalos de clase 5.10-5.40, 5.40-5.70, ..., 6.90-7.20 y 7.20-7.50. ¿Hay algo incorrecto en su elección?

Solución

Esos intervalos de clase se solapan en 5.40, 5.70, ..., 7.20. Luego una medida anotada como 5.40, por ejemplo, podría ser colocada en cualquiera de los dos primeros intervalos de clase. Algunos estadísticos justifican esta elección decidiendo asignar la mitad de los casos dudosos a una clase y la otra mitad a la otra.

La ambigüedad desaparece escribiendo los intervalos de clase como 5.10 hasta 5.40, 5.40 hasta 5.70, etc. En este caso, los límites de clase coinciden con las fronteras de clase, y las marcas de clase pueden coincidir con datos observados.

En general, es deseable evitar solapamientos de intervalos de clase si es posible y escogerlos de modo que las fronteras de clase no coincidan con los datos observados. Por ejemplo, los intervalos de clase del Problema 2.6 podían haberse escogido como 5.095-5.395, 5.395-5.695, etc., sin ambigüedad. Una desventaja de esta elección particular es que las marcas de clase no coinciden con los datos observados.

2.8. En la tabla que sigue se recogen los pesos de 40 estudiantes varones de una universidad, con precisión de 1 libra. Construir una distribución de frecuencias.

138	164	150	132	144	125	149	157
146	158	140	147	136	148	152	144
168	126	138	176	163	119	154	165
146	173	142	147	135	153	140	135
161	145	135	142	150	156	145	128

Solución

Los pesos extremos son 176 y 119 lb, luego el rango es $176-119=57$ lb. Si se usan 5 intervalos de clase, su anchura será $57/5=11$ aproximadamente; si se usan 20 intervalos de clase, será de $57/20=3$, aproximadamente.

Una colección razonable es 5 lb. Es conveniente, asimismo, elegir las marcas de clase como 120, 125, 130, 135, ..., lb. De modo que los intervalos de clase pueden tomarse como 118-122, 123-127, 128-132, ... Con tal elección, las fronteras de clase son 117.5, 122.5, 127.5, ..., que no coinciden con los datos observados.

$\frac{57}{x}=5 \rightarrow x=11,4 \cong 12$

Tabla 2.6

Peso (lb)	Recuento	Frecuencia
118-122	/	1
123-127	//	2
128-132	//	2
133-137	////	4
138-142	///// /	6
143-147	///// ///	8
148-152	/////	5
153-157	////	4
158-162	//	2
163-167	///	3
168-172	/	1
173-177	//	2
		Total 40

Tabla 2.7

Peso (lb)	Recuento	Frecuencia
118-126	///	3
127-135	/////	5
136-144	///// ////	9
145-153	///// ///// //	12
154-162	/////	5
163-171	////	4
172-180	//	2
		Total 40

La distribución de frecuencias requerida se ve en la Tabla 2.6. La columna central, llamada *hoja de recuentos*, se usa para tabular las frecuencias de clase y suele omitirse en la presentación final de la distribución de frecuencias. No es necesario hacer ordenación, aunque si se dispone de ella puede utilizarse para tabular las frecuencias.

Otro método

Naturalmente, existen otras distribuciones de frecuencias. La Tabla 2.7, por ejemplo, muestra una distribución de frecuencias con sólo 7 clases, en la que la anchura del intervalo de clase es 9 lb.

2.9. Construir: (*a*) un histograma y (*b*) un polígono de frecuencias para la distribución de pesos del Problema 2.8.

Solución

El histograma y el polígono de frecuencias para cada caso del Problema 2.8 vienen dados en las Figuras 2.4(*a*) y 2.4(*b*). Nótese que los centros de las bases de los rectángulos están localizados en las marcas de clase.

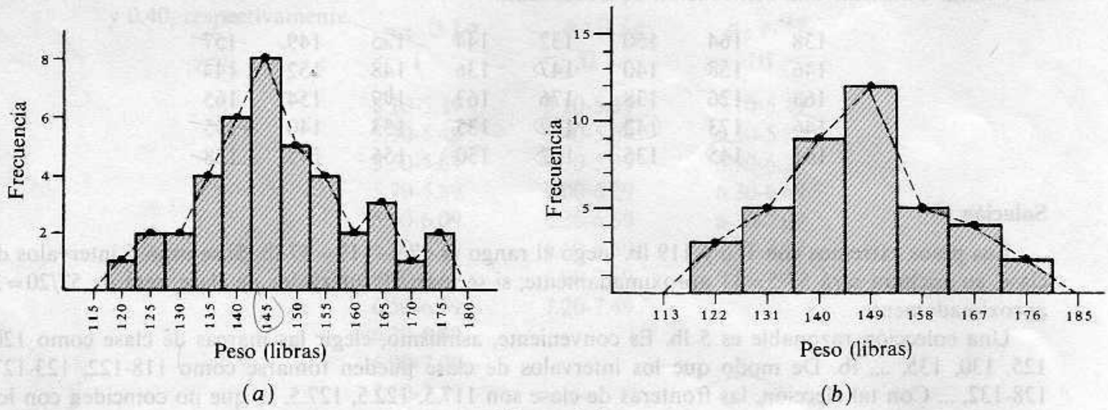


Figura 2.4.

2.10. Con los datos de la Tabla 2.5 del Problema 2.3, construir: (*a*) una distribución de frecuencias relativas, (*b*) un histograma, (*c*) un histograma de frecuencias relativas, (*d*) un polígono de frecuencias y (*e*) un polígono de frecuencias relativas.

Solución

(*a*) La distribución de frecuencias relativas de la Tabla 2.8 se obtiene de la distribución de frecuencias de la Tabla 2.5 dividiendo cada frecuencia de clase por la frecuencia total (65) y expresando el resultado como porcentaje.

(*b*) y (*c*) El histograma y el histograma de frecuencias relativas se muestran en la Figura 2.5. Nótese que para pasar de uno a otro sólo es necesario añadir al histograma una escala vertical con las frecuencias relativas, como se ve a la derecha en la Figura 2.5.

(*d*) y (*e*) El polígono de frecuencias y el polígono de frecuencias relativas se indican por la gráfica de trazos en la Figura 2.5. Así pues, para convertir un polígono de frecuencias en un polígono de frecuencias relativas, basta añadir una escala vertical con las frecuencias relativas.

Si sólo se desea un polígono de frecuencias relativas, la figura adjunta no contendría el histograma y el eje de las frecuencias relativas aparecería en la izquierda en lugar del eje de frecuencias.

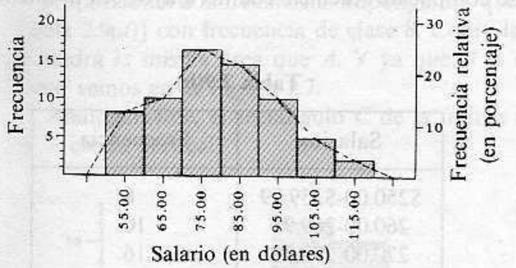


Figura 2.5.

Tabla 2.8

Salarios	Frecuencia relativa (como porcentaje)
\$250.00-\$259.99	12.3
260.00-269.99	15.4
270.00-279.99	24.6
280.00-289.99	21.5
290.00-299.99	15.4
300.00-309.99	7.7
310.00-319.99	3.1
Total	100.0

2.11. Probar que en un histograma el área total de los rectángulos es igual al área total limitada por el correspondiente polígono de frecuencias y el eje X.

Solución

Lo probaremos para el caso de un histograma con tres rectángulos (Fig. 2.6) y el polígono de frecuencias asociado, que se indica con trazo discontinuo.

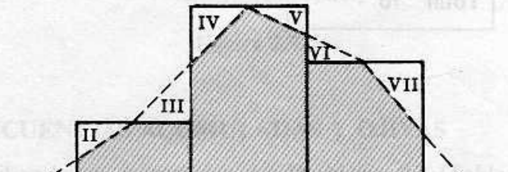


Figura 2.6.

Area total de los rectángulos = área sombreada + área II + área IV + área V + área VII
= área sombreada + área I + área III + área VI + área VIII
= área total acotada por el polígono de frecuencias y el eje X

Como área I = área II, entonces área III = área IV, área V = área VI y área VII = área VIII.

2.12. En la empresa P&R (Prob. 2.3), se ha contratado a cinco nuevos trabajadores con salarios semanales de \$285.34, \$316,83, \$335.78, \$356.21 y \$374.50. Construir una distribución de frecuencia de los salarios de los 70 trabajadores.

Solución

La Tabla 2.9 muestra posibles distribuciones de frecuencia.

En la Tabla 2.9(*a*) se ha usado un mismo tamaño de intervalos de clase \$10.00. Como consecuencia, hay demasiadas clases vacías y la información es más detallada en el extremo superior de la escala de salarios.

En la Tabla 2.9(*b*) las clases vacías y los detalles finos han sido evitados usando el intervalo de clase abierto «\$320.00 o más». Una desventaja sería que la tabla se haría menos cómoda al efectuar ciertos cálculos. Así, es imposible determinar la cantidad total pagada a la semana porque «\$320.00 o más» podría significar que hay individuos que cobran incluso \$1400.00 a la semana.

En la Tabla 2.9(*c*) se usa una anchura de intervalo de clase de \$20.00, con la desventaja de que se que ciertas operaciones matemáticas posteriores se complican. Además, cuanto mayor sea la anchura, mayor el error de agrupamiento.

Tabla 2.9(a)

Salarios	Frecuencia
\$250.00-\$259.99	8
260.00-269.99	10
270.00-279.99	16
280.00-289.99	15
290.00-299.99	10
300.00-309.99	5
310.00-319.99	3
320.00-329.99	0
330.00-339.99	1
340.00-349.99	0
350.00-359.99	1
360.00-369.99	0
370.00-379.99	1
	Total 70

Tabla 2.9(b)

Salarios	Frecuencia
\$250.00-\$259.99	8
260.00-269.99	10
270.00-279.99	16
280.00-289.99	15
290.00-299.99	10
300.00-309.99	5
310.00-319.99	3
320.00 en adelante	3
	Total 70

Tabla 2.9(c)

Salarios	Frecuencia
\$250.00-\$269.99	18
270.00-289.99	31
290.00-309.99	15
310.00-329.99	3
330.00-349.99	1
350.00-369.99	1
370.00-389.99	1
	Total 70

Tabla 2.9(d)

Salarios	Frecuencia	
\$250.00-\$259.99	8	8
260.00-269.99	10	18
270.00-279.99	16	34
280.00-289.99	15	49
290.00-299.99	10	59
300.00-319.99	8	67
320.00-379.99	3	70
	Total 70	

2.13. Construir un histograma para la distribución de frecuencias de la Tabla 2.9(*d*).

Solución

La Figura 2.7 muestra el histograma solicitado. Para construirlo usamos el hecho de que el área es proporcional a la frecuencia. Supongamos que el rectángulo *A* corresponde a la primera clase [véase Tabla 2.9(*d*)] con frecuencia de clase 8. Como la sexta clase tiene también frecuencia 8, su rectángulo *B* tendrá la misma área que *A*. Y ya que *B* es doble ancho que *A*, tendrá la mitad de su altura, tal como vemos en la Figura 2.7.

Análogamente, el rectángulo *C* de la última clase en la Tabla 2.9(*d*) tiene media unidad de altura en la escala vertical.

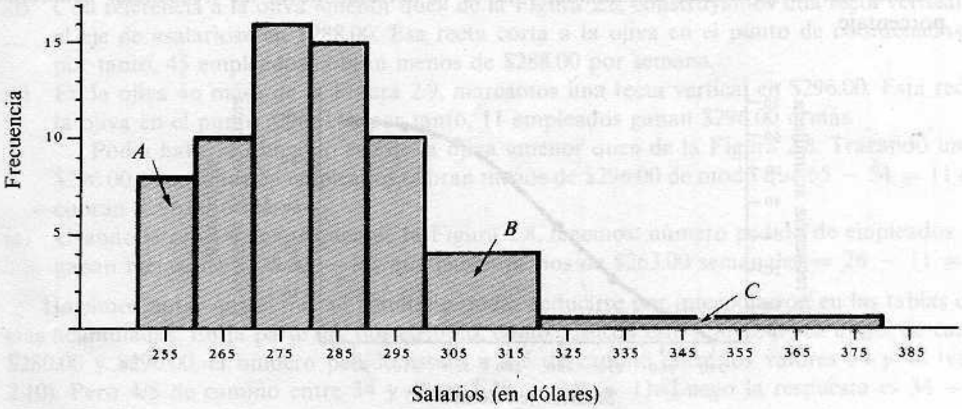


Figura 2.7.

DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS Y OJIVAS

2.14. Construir para la distribución de frecuencias del Problema 2.3 (Tabla 2.5): (*a*) una distribución de frecuencias acumuladas, (*b*) una distribución acumulada de porcentajes, (*c*) una ojiva y (*d*) una ojiva de porcentajes.

Solución

(*a*) y (*b*) La distribución de frecuencias acumuladas y la distribución acumulada en porcentajes (o distribución de frecuencias relativas acumuladas) se combinan en la Tabla 2.10.

Tabla 2.10

Salarios	Frecuencia acumulada	Porcentaje acumulativo de distribución
Menor que 250.00	0	0.0
Menor que 260.00	8	12.3
Menor que 270.00	18	27.7
Menor que 280.00	34	52.3
Menor que 290.00	48	73.8
Menor que 300.00	58	89.2
Menor que 310.00	63	96.9
Menor que 320.00	65	100.0

Nótese que cada entrada de la columna 2 se obtiene sumando entradas sucesivas de la columna 2 de la Tabla 2.5. Luego 18 = 8 + 10, 34 = 8 + 10 + 16, etc.

Cada entrada en la columna 3 se obtiene de la anterior dividiendo por 65, la frecuencia total, y expresando el resultado como porcentaje. Así, 34/65 = 52.3%. También podían haberse obtenido sumando entradas sucesivas de la columna 2 de la Tabla 2.8. Así, 27.7 = 12.3 + 15.4, 52.3 = 12.3 + 15.4 + 24.6, etc.

(*c*) y (*d*) La ojiva (o polígono de frecuencias acumuladas) y la ojiva de porcentajes (o polígono de frecuencias acumuladas relativas) se ven en la Figura 2.8. La escala vertical de la izquierda nos permite leer la frecuencia acumulada, y la de la derecha indica las frecuencias acumuladas en porcentaje.

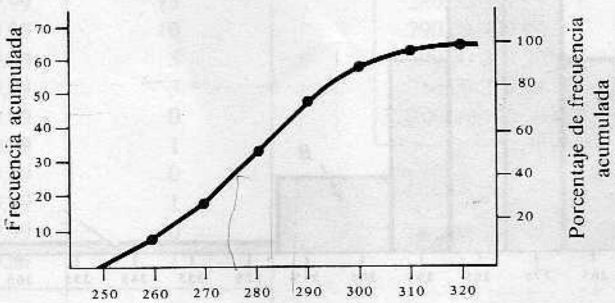


Figura 2.8.

Las anteriores suelen llamarse ojiva o distribución de frecuencias acumuladas «menor que», por la manera de acumular las frecuencias.

2.15. A partir de la distribución de frecuencias de la Tabla 2.5 del Problema 2.3, construir: (*a*) una distribución de frecuencias acumuladas «o más» y (*b*) una ojiva «o más».

fue. Ni!

Solución

(*a*) Cada entrada de la columna 2 en la Tabla 2.11 se obtiene sumando entradas sucesivas de la columna 2 de la Tabla 2.55, *comenzando por abajo;* asi pues, $7 = 2 + 5$, $17 = 2 + 5 + 10$, etc. Estas entradas pueden obtenerse también restando cada entrada en la columna 2 de la Tabla 2.10 de la frecuencia total, 65, es decir, $57 = 65 - 8$, $47 = 65 - 18$, etc.

(*b*) La Figura 2.9 muestra una ojiva «o más».

Tabla 2.11

Salarios	Frecuencia acumulada «o más»
\$250.00 o más	65
260.00 o más	57
270.00 o más	47
280.00 o más	31
290.00 o más	17
300.00 o más	7
310.00 o más	2
320.00 o más	0

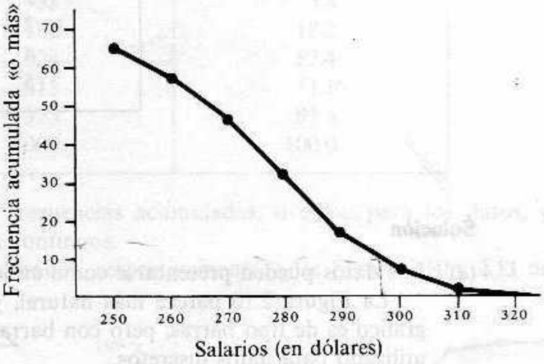


Figura 2.9.

2.16. De las ojivas en las Figuras 2.8 y 2.9 (de los Probs. 2.14 y 2.15, respectivamente), estimar el número de empleados que cobran por semana: (*a*) menos de \$288.00, (*b*) \$296.00 o más y (*c*) al menos \$263.00, pero menos de \$275.00.

Solución

(*a*) Con referencia a la ojiva «menor que» de la Figura 2.8, construyamos una recta vertical que corte al eje de «salarios» en \$288.00. Esa recta corta a la ojiva en el punto de coordenadas (288, 45); por tanto, 45 empleados cobran menos de \$288.00 por semana.

X Y

(*b*) En la ojiva «o más» de la Figura 2.9, marcamos una recta vertical en \$296.00. Esta recta corta a la ojiva en el punto (296, 11); por tanto, 11 empleados ganan \$296.00 o más.

Podía haberse obtenido eso de la ojiva «menor que» de la Figura 2.8. Trazando una recta en \$296.00, vemos que 54 empleados cobran menos de \$296.00 de modo que $65 - 54 = 11$ empleados cobran \$296.00 o más.

(*c*) Usando la ojiva «menor que» de la Figura 2.8, tenemos: número pedido de empleados = los que ganan menos de \$275.00 − los que ganan menos de \$263.00 semanales $= 26 - 11 = 15$.

Hagamos notar que el mismo resultado podía deducirse por interpolación en las tablas de frecuencias acumuladas. En la parte (*a*), por ejemplo, como \$288.00 está a 8/10, o sea a 4/5, de camino entre \$280.00 y \$290.00, el número pedido estará a 4/5 de camino entre los valores 34 y 48 (véase Tabla 2.10). Pero 4/5 de camino entre 34 y 48 es $\frac{4}{5}(48 - 34) = 11$. Luego la respuesta es $34 + 11 = 45$ empleados.

2.17. Se lanzan cinco monedas 1000 veces. El número de lanzamientos en los que han salido 0, 1, 2, 3, 4 y 5 caras se indican en la Tabla 2.12.

(*a*) Representar los datos de esa tabla.

(*b*) Construir una tabla que muestre los porcentajes de tiradas que han dado un número de caras menor que 0, 1, 2, 3, 4, 5 ó 6.

(*c*) Representar los datos de la tabla de la parte (*b*).

Tabla 2.12

Número de caras	Número de tiradas (frecuencia)
0	38
1	144
2	342
3	287
4	164
5	25
	Total 1000

Solución

(*a*) Los datos pueden presentarse como en las Figuras 2.10 ó 2.11.

La Figura 2.10 parece más natural, ya que el número de caras no puede ser 1.5 ó 3.2. Este gráfico es de tipo barras, pero con barras de anchura cero. Se llama *gráfico de varillas* y es muy utilizado para datos discretos.

La Figura 2.11 es un histograma de los datos. El área total del histograma es la frecuencia total, 1000, como debe ser. Al usar la representación en histograma o el correspondiente polígono de frecuencias, estamos tratando los datos como si fueran continuos. Luego veremos que tal perspectiva es útil. Recuérdese que ya hemos utilizado el histograma y los polígonos de frecuencias para datos discretos en el Problema 2.10.

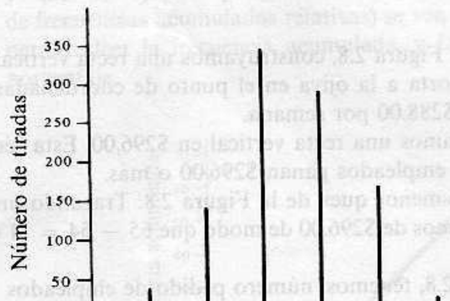


Figura 2.10.

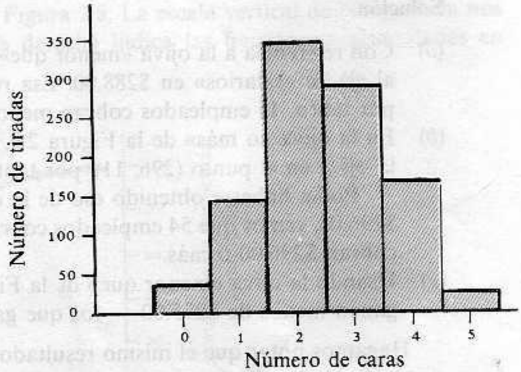


Figura 2.11.

(*b*) La Tabla 2.13 muestra simplemente una distribución de frecuencias acumuladas y una distribución de porcentajes acumulados del número de caras. Debe observarse que las entradas «menor que 1», «menor que 2», etc., podrían haberse sustituido por entradas «menor o igual que».

(*c*) El gráfico pedido puede presentarse como en la Figura 2.12 o como en la Figura 2.13.

La Figura 2.12 parece más natural para presentar datos discretos, pues el porcentaje de tiradas con menos de 2 caras ha de ser igual que para menos de 1.75, 1.56 ó 1.23 caras, de manera que debe verse el mismo porcentaje (18.2%) para esos valores (indicado por un segmento horizontal).

Tabla 2.13

Número de caras	Número de tiradas (frecuencia acumulada)	Porcentaje de número de tiradas (porcentaje de frecuencia acumulada)
Menor que 0	0	0.0
Menor que 1	38	3.8
Menor que 2	182	18.2
Menor que 3	524	52.4
Menor que 4	811	81.1
Menor que 5	975	97.5
Menor que 6	1000	100.0

La Figura 2.13 muestra el polígono de frecuencias acumuladas, u ojiva, para los datos, y esencialmente trata los datos como si fueran continuos.

Nótese que las Figuras 2.12 y 2.13 corresponden, respectivamente, a las Figura 2.10 y 2.11 de la parte (*a*).

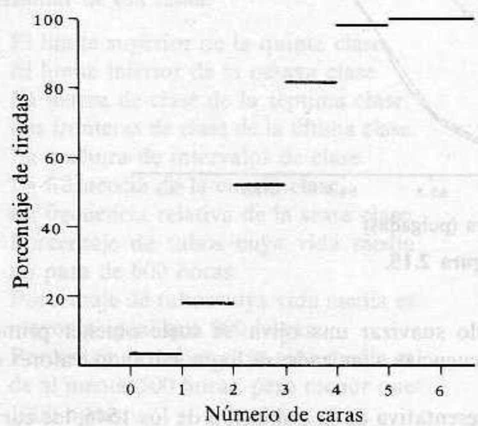


Figura 2.12.

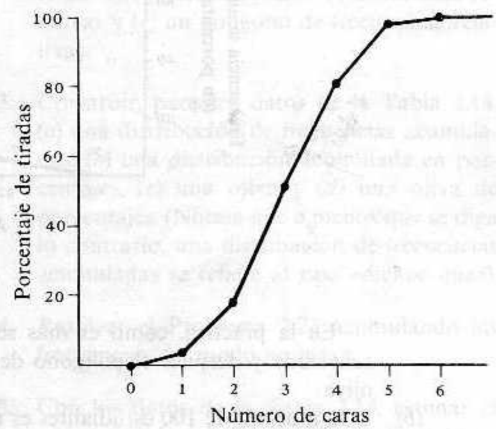


Figura 2.13.

CURVAS DE FRECUENCIA Y OJIVAS SUAVIZADAS

2.18. Los 100 estudiantes de la Universidad XYZ (Tabla 2.1) constituían en realidad una muestra de los 1546 estudiantes varones de esa universidad.

(*a*) De los datos de esa muestra, construir un polígono de frecuencias en porcentajes suavizado (curva de frecuencias) y una ojiva suavizada en porcentajes «menor que».

(*b*) De los resultados de una de las construcciones de la parte (*a*), estimar el número de estudiantes con alturas entre 65 y 70 in. ¿Qué hipótesis hay que hacer?

(*c*) ¿Cabe utilizar los resultados para estimar la proporción de varones en EE.UU. con alturas entre 65 y 70 in?

Solución

(*a*) En las Figuras 2.14 y 2.15 los gráficos discontinuos representan los polígonos de frecuencias y las ojivas, y se han obtenido de las Figuras 2.1 y 2.2, respectivamente. Las gráficas suavizadas (en trazo continuo) se obtienen aproximando los anteriores mediante curvas continuas.

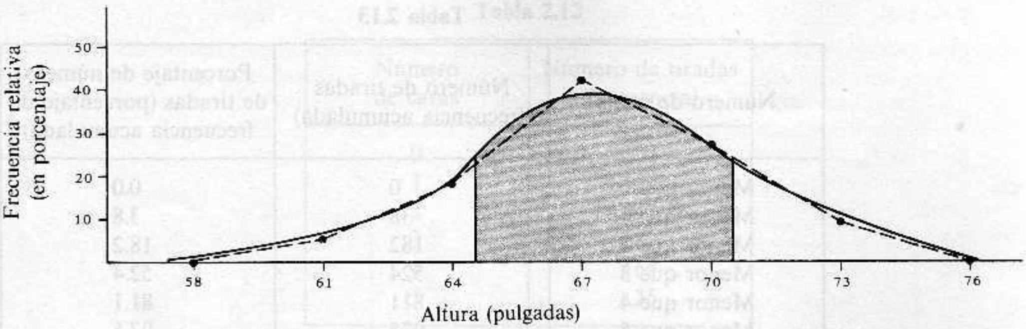


Figura 2.14.

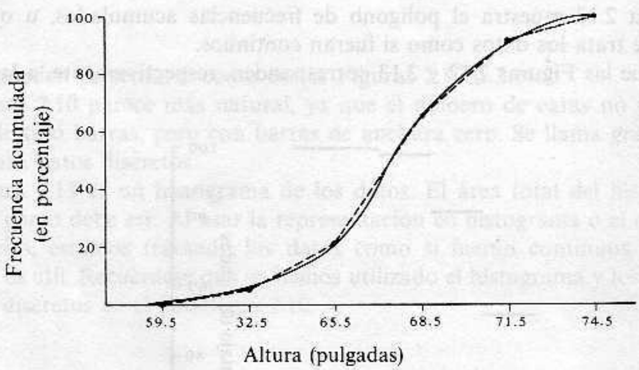


Figura 2.15.

En la práctica, como es más sencillo suavizar una ojiva, se suele obtener primero la ojiva suavizada y después el polígono de frecuencias suavizado se logra mirando valores en la citada ojiva.

(*b*) Si la muestra de 100 estudiantes es representativa de la población de los 1546, las curvas suavizadas de las Figuras 2.14 y 2.15 pueden considerarse como la curva de frecuencias en porcentajes y la ojiva de porcentajes de esa población. Esta hipótesis es correcta sólo si la muestra es aleatoria (o sea, si cada estudiante tiene la misma probabilidad de salir elegido en la muestra).

Como las alturas anotadas entre 65 y 70 in, con precisión de pulgada, en realidad representan alturas entre 64.5 y 70.5 in, el porcentaje de estudiantes en la población que tiene esas alturas se encuentra dividiendo el área sombreada de la Figura 2.14 por el área total acotada por la curva suavizada y el eje X.

Es más sencillo, no obstante, usar la Figura 2.15, de la que vemos que

$$\text{Porcentaje de estudiantes por debajo de 70.5 in} = 82\%$$
$$\text{Porcentaje de estudiantes por debajo de 64.5 in} = 18\%$$

luego el porcentaje con alturas entre 64.5 y 70.5 in $= 82\% - 18\% = 64\%$. Así pues, el número de estudiantes de esa universidad que miden entre 65 y 70 in es el 64% de $1546 = 989$.

Otra forma de decir eso es afirmar que la probabilidad de que una persona, elegida al azar de entre esas 1546, tenga altura comprendida entre 65 y 70 in, es 64%, 0,64 ó 64 de cada 100. A causa de la relación con las probabilidades (tratadas en el Capítulo 6), las curvas de frecuencia relativa se llaman *curvas de probabilidad* o *distribuciones de probabilidad*.

(*c*) Podríamos estimar la requerida proporción en un 64% (ahora con mucho más margen de error) sólo si estuviéramos convencidos de que la muestra de 100 estudiantes fuera realmente aleatoria vista desde la población masculina de EE.UU. Lo cual es improbable, porque algunos estudiantes no habrán alcanzado aún su altura tope y las generaciones jóvenes tienden a ser más altas que las anteriores, aparte de otros factores.

PROBLEMAS SUPLEMENTARIOS

2.19. (*a*) Ordenar los números 12, 56, 42, 21, 5, 18, 10, 3, 61, 34, 65 y 24 y (*b*) hallar su rango.

2.20. La Tabla 2.14 muestra una distribución de frecuencias de las vidas medias de 400 válvulas de radio probadas en la empresa L&M. Determinar de esa tabla:

(*a*) El límite superior de la quinta clase.
(*b*) El límite inferior de la octava clase.
(*c*) La marca de clase de la séptima clase.
(*d*) Las fronteras de clase de la última clase.
(*e*) La anchura de intervalos de clase.
(*f*) La frecuencia de la cuarta clase.
(*g*) La frecuencia relativa de la sexta clase.
(*h*) Porcentaje de tubos cuya vida media no pasa de 600 horas.
(*i*) Porcentaje de tubos cuya vida media es mayor o igual que 900 horas.
(*j*) Porcentaje de tubos cuya vida media es de al menos 500 horas, pero menor que 1000 horas.

Tabla 2.14

Vida media (horas)	Número de tubos	
1) 300-399	14	
2) 400-499	46	60
3) 500-599	58	118
4) 600-699	76	194
5) 700-799	68	262
6) 800-899	62	324
7) 900-999	48	372
8) 1000-1099	22	394
9) 1100-1199	6	400
	Total 400	

2.21. Construir: (*a*) un histograma y (*b*) un polígono de frecuencias correspondientes a la distribución de frecuencias de la Tabla 2.14.

2.22. Para los datos de la Tabla 2.14 (Prob. 2.20), construir: (*a*) una distribución de frecuencias relativas, (*b*) un histograma de frecuencias relativas y (*c*) un polígono de frecuencias relativas.

2.23. Construir, para los datos de la Tabla 2.14, (*a*) una distribución de frecuencias acumuladas, (*b*) una distribución acumulada en porcentajes, (*c*) una ojiva y (*d*) una ojiva de porcentajes. (Nótese que a menos que se diga lo contrario, una distribución de frecuencias acumuladas se refiere al tipo «menor que»).

2.24. Resolver el Problema 2.23 acumulando las frecuencias del modo «o más».

2.25. Con los datos de la Tabla 2.14, estimar el porcentaje de tubos con vida media: (*a*) menor que 560 horas, (*b*) 970 horas o más y (*c*) entre 620 y 890 horas.

14%

2.26. Los diámetros internos de los tubos fabricados por una empresa se miden con precisión de milésima de pulgada. Si las marcas de clase de una distribución de frecuencias de esos diámetros vienen dadas por 0.321, 0.324, 0.327, 0.330, 0.333 y 0.336, hallar: (*a*) la anchura del intervalo de clase, (*b*) las fronteras de clase y (*c*) los límites de clase.

2.27. La tabla adjunta muestra los diámetros en centímetros de una muestra de 60 bolas de cojinete manufacturadas por una fábrica. Construir una distribución de frecuencias con intervalos de clase apropiados.

1.738	1.729	1.743	1.740	1.736	1.741
1.735	1.731	1.726	1.737	1.728	1.737
1.736	1.735	1.724	1.733	1.742	1.736
1.739	1.735	1.745	1.736	1.742	1.740
1.728	1.738	1.725	1.733	1.734	1.732
1.733	1.730	1.732	1.730	1.739	1.734
1.738	1.739	1.727	1.735	1.735	1.732
1.735	1.727	1.734	1.732	1.736	1.741
1.736	1.744	1.732	1.737	1.731	1.746
1.735	1.735	1.729	1.734	1.730	1.740

2.28. Para los datos del Problema 2.27, construir: (*a*) un histograma, (*b*) un polígono de frecuencias, (*c*) una distribución de frecuencias relativas, (*d*) un histograma de frecuencias relativas, (*e*) un polígono de frecuencias relativas, (*f*) una distribución de frecuencias acumuladas, (*g*) una distribución acumulada en porcentajes, (*h*) una ojiva e (*i*) una ojiva de porcentajes.

2.29. Determinar, a partir de los resultados del Problema 2.28, el porcentaje de bolas con diámetros: (*a*) mayores que 1.732 cm, (*b*) no mayor que 1.736 cm y (*c*) entre 1.730 y 1.738 cm. Comparar los resultados con los obtenidos directamente de los datos del Problema 2.27.

2.30. Repetir el Problema 2.28 para los datos del Problema 2.20.

2.31. La Tabla 2.15 muestra la distribución de porcentajes de ventas totales para plantaciones de tipo familiar en EE.UU. en 1982. Usando esa tabla, responder las siguientes cuestiones:

(*a*) ¿Cuál es la anchura del segundo intervalo de clase? ¿Y del séptimo?

(*b*) Cuántos tamaños diferentes de intervalos de clase hay?

(*c*) ¿Cuántos intervalos de clase abiertos hay?

(*d*) ¿Cómo habría que escribir el primer intervalo de clase para que su anchura sea igual a la del segundo?

(*e*) ¿Cuál es la marca de clase del segundo intervalo de clase? ¿Y del séptimo?

(*f*) ¿Cuáles son las fronteras de clase del cuarto intervalo de clase?

(*g*) ¿Qué porcentaje de las plantaciones tuvo ventas de $20,000 o más? ¿Y por debajo de $10,000?

(*h*) ¿Qué porcentaje logró ventas de al menos $10,000, pero no mayores que $40,000?

(*i*) ¿Qué porcentaje tuvo ventas entre $15,000 y $25,000? ¿Qué hipótesis se han hecho en ese cálculo?

(*j*) ¿Por qué los porcentajes de la Tabla 2.15 no suman 100%?

Tabla 2.15

Ventas (dólares)	Explotaciones (%)
Menos de 2,500	25.9
2,500-4,999	13.2
5,000-9,999	13.0
10,000-19,999	11.7
20,000-39,999	11.0
40,000-99,999	14.4
100,000-249,999	8.5
250,000-499,999	1.8
500,000 o más	0.6

2.32. (*a*) ¿Por qué es imposible construir un histograma de porcentajes o un polígono de frecuencias para la distribución de la Tabla 2.15?

(*b*) ¿Cómo modificaría la distribución para que pudieran construirse ambos?

(*c*) Llevar a cabo la modificación y la construcción.

2.33. El número total de plantaciones en la distribución de la Tabla 2.15 es 1,945,000. A partir de ese dato, determinar el número de plantaciones con ventas: (*a*) superiores a $40,000, (*b*) menores que $40,000 y (*c*) entre $30,000 y $50,000.

2.34. (*a*) Construir un polígono de frecuencias en porcentajes suavizado y una ojiva en porcentajes suavizada para los datos de la Tabla 2.14.

(*b*) Estimar con ellos la probabilidad de que un tubo se deteriore antes de 600 horas.

(*c*) Discutir el riesgo del fabricante al garantizar los tubos por 425 horas. Idem con 875 horas.

(*d*) Si el fabricante ofrece una garantía de 90 días para la devolución del importe de un tubo, ¿cuál es la probabilidad de que devuelva el importe, supuesto que el tubo esté en uso 4 horas diarias? ¿Y con 8 horas diarias?

2.35. (*a*) Lanzar 4 monedas 50 veces y anotar el número de caras en cada ocasión.

(*b*) Construir una distribución de frecuencias que indique el número de veces que se han obtenido 0, 1, 2, 3 y 4 caras.

(*c*) Construir una distribución de porcentajes correspondiente a la parte (*b*).

(*d*) Comparar el porcentaje obtenido en (*c*) con los teóricos 6,25%, 25%, 37.5%, 25% y 6.25% (proporcional a 1, 4, 6, 4 y 1) deducidos por las leyes de las probabilidades.

(*e*) Representar las distribuciones de las partes (*b*) y (*c*).

(*f*) Construir una ojiva de porcentajes para los datos.

2.36. Repetir el problema anterior con otros 50 lanzamientos y véase si el experimento está más de acuerdo con lo esperado teóricamente. Si no, dar posibles razones para tales discrepancias.

CAPITULO 3

Media, mediana, moda y otras medidas de tendencia central

NOTACION DE INDICES

Denotemos por X_j (léase «X sub j») cualquiera de los N *valores* $X_1, X_2, X_3, \ldots, X_N$ que toma una variable X. La letra j en X_j, que puede valer 1, 2, 3, ..., N se llama *subíndice*. Es claro que podíamos haber empleado cualquier otra letra en vez de j, por ejemplo, i, k, p, q o s.

NOTACION DE SUMA

El símbolo $\sum_{j=1}^{N} X_j$ denotará la suma de todos los X_j desde $j = 1$ a $j = N$; por definición,

$$\sum_{j=1}^{N} X_j = X_1 + X_2 + X_3 + \cdots + X_N$$

Cuando no ocasione confusión, denotaremos esa suma simplemente por $\sum X$, $\sum X_j$ o $\sum_j X_j$. El símbolo $\sum$ es la letra griega *sigma* mayúscula, que denota suma.

EJEMPLO 1. $\displaystyle\sum_{j=1}^{N} X_j Y_j = X_1 Y_1 + X_2 Y_2 + X_3 Y_3 + \cdots + X_N Y_N$.

EJEMPLO 2. $\displaystyle\sum_{j=1}^{N} aX_j = aX_1 + aX_2 + \cdots + zX_N = a(X_1 + X_2 + \cdots + X_N) = a\sum_{j=1}^{N} X_j$, donde a es una constante. Más sencillamente, $\sum aX = a \sum X$.

EJEMPLO 3. Si a, b, c son constantes, entonces $\sum (aX + bY - cZ) = a \sum X + b \sum Y - c \sum Z$. Véase Problema 3.3.

PROMEDIOS O MEDIDAS DE TENDENCIA CENTRAL

Un *promedio* es un valor típico o representativo de un conjunto de datos. Como tales valores suelen situarse hacia el centro del conjunto de datos ordenados por magnitud, los promedios se conocen como *medidas de tendencia central*.

Se definen varios tipos, siendo los más comunes la *media aritmética*, la *mediana*, la *moda*, la *media geométrica* y la *media armónica*. Cada una tiene ventajas y desventajas, según los datos y el objetivo perseguido.

LA MEDIA ARITMETICA

La *media aritmética*, o simplemente *media*, de un conjunto de N números $X_1, X_2, X_3, ..., X_N$ se denota por $\bar{X}$ (léase «X barra») y se define por

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{j=1}^{N} X_j}{N} = \frac{\sum X}{N} \tag{1}$$

EJEMPLO 4. La media aritmética de los números 8, 3, 5, 12 y 10 es

$$\bar{X} = \frac{8 + 3 + 5 + 12 + 10}{5} = \frac{38}{5} = 7.6$$

Si los números $X_1, X_2, ..., X_K$ ocurren $f_1, f_2, ..., f_k$ veces, respectivamente (o sea, con frecuencias $f_1, f_2, ..., f_k$), la media aritmética es

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + \cdots + f_KX_K}{f_1 + f_2 + \cdots + f_K} = \frac{\sum_{j=1}^{K} f_jX_j}{\sum_{j=1}^{K} f_j} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N} \tag{2}$$

donde $N = \sum f$ es la *frecuencia total* (o sea, el número total de casos).

EJEMPLO 5. Si 5, 8, 6 y 2 ocurren con frecuencias 3, 2, 4 y 1, respectivamente, su media aritmética es

$$\bar{X} = \frac{(3)(5) + (2)(8) + (4)(6) + (1)(2)}{3 + 2 + 4 + 1} = \frac{15 + 16 + 24 + 2}{10} = 5.7$$

LA MEDIA ARITMETICA PONDERADA

A veces asociamos con los números $X_1, X_2, ..., X_K$ ciertos *factores peso* (o *pesos*) $w_1, w_2, ..., w_K$, dependientes de la relevancia asignada a cada número. En tal caso,

$$\bar{X} = \frac{w_1X_1 + w_2X_2 + \cdots + w_KX_k}{w_1 + w_2 + \cdots + w_K} = \frac{\sum wX}{\sum w} \tag{3}$$

se llama la *media aritmética ponderada* con pesos $f_1, f_2, ..., f_K$.

EJEMPLO 6. Si el examen final de un curso cuenta tres veces más que una evaluación parcial, y un estudiante tiene calificación 85 en el examen final y 70 y 90 en los dos parciales, la calificación media es

$$\bar{X} = \frac{(1)(70) + (1)(90) + (3)(85)}{1 + 1 + 3} = \frac{415}{5} = 83$$

PROPIEDADES DE LA MEDIA ARITMETICA

1. La suma algebraica de las desviaciones de un conjunto de números respecto de su media aritmética es cero.

EJEMPLO 7. Las desviaciones de los números 8, 3, 5, 12 y 10 respecto de su media aritmética 7.6 son $8 - 7.6$, $3 - 7.6$, $5 - 7.6$, $12 - 7.6$ y $10 - 7.6$, o sea 0.4, -4.6, -2.6, 4.4 y 2.4, con suma algebraica $0.4 - 4.6 - 2.6 + 4.4 + 2.4 = 0$.

2. La suma de los cuadrados de las desviaciones de un conjunto de números X_j respecto de un cierto número a es mínima si y sólo si $a = \bar{X}$ (véase Prob. 4.27).
3. Si f_1 números tienen media m_1, f_2 números tiene media m_2, ..., f_K números tienen media m_K, entonces la media de todos los números es

$$\bar{X} = \frac{f_1 m_1 + f_2 m_2 + \cdots + f_K m_K}{f_1 + f_2 + \cdots + f_K} \tag{4}$$

es decir, una media aritmética ponderada de todas las medias (véase Prob. 3.12).

4. Si A es una *media aritmética supuesta o conjeturada* (que puede ser cualquier número) y si $d_j = X_j - A$ son las desviaciones de X_j respecto de A, las ecuaciones (1) y (2) se convierten, respectivamente, en

$$\bar{X} = A + \frac{\sum_{j=1}^{N} d_j}{N} = A + \frac{\sum d}{N} \tag{5}$$

$$\bar{X} = A + \frac{\sum_{j=1}^{K} f_j d_j}{\sum_{j=1}^{K} f_j} = A + \frac{\sum fd}{N} \tag{6}$$

donde $N = \sum_{j=1}^{K} f_j = \sum f$. Nótese que las fórmulas (5) y (6) se resumen en $\bar{X} = A + \bar{d}$ (véase Prob. 3.18).

CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS

Cuando los datos se presentan en una distribución de frecuencias, todos los valores que caen dentro de un intervalo de clase dado se consideran iguales a la marca de clase, o punto medio, del intervalo. Las fórmulas (2) y (6) son válidas para tales datos agrupados si interpretamos X_j como la marca de clase, f_j como su correspondiente frecuencia de clase, A como cualquier marca de clase conjeturada y $d_j = X_j - A$ como las desviaciones de X_j respecto de A.

Los cálculos con (2) y (6) se llaman *métodos largos y cortos*, respectivamente (véanse Probs. 3.15 y 3.20).

Si todos los intervalos de clase tienen idéntica anchura c, las desviaciones $d_j = X_j - A$ pueden expresarse como cu_j, donde u_j pueden ser 0, ± 1, ± 2, ± 3, ..., y la fórmula (6) se convierte en

$$\bar{X} = A + \left(\frac{\sum_{j=1}^{K} f_j u_j}{N}\right) = A + \left(\frac{\sum fu}{N}\right)c \tag{7}$$

que es equivalente a la ecuación $\bar{X} = A + c\bar{u}$ (véase Prob. 3.21). Esto se conoce como *método de compilación* para calcular la media. Es un método muy breve y debe usarse siempre para datos agrupados con intervalos de clase de anchuras iguales (véanse Probs. 3.22 y 3.23). Nótese que en el método de compilación los valores de la variable X *se transforman* en los valores de la variable u de acuerdo con $X = A + cu$.

LA MEDIANA

La *mediana* de un conjunto de números ordenados en magnitud es o el valor central o la media de los dos valores centrales.

EJEMPLO 8. El conjunto de números 3, 4, 4, 5, 6, 8, 8, 8 y 10 tiene mediana 6.

EJEMPLO 9. El conjunto de números 5, 5, 7, 9, 11, 12, 15 y 18 tiene mediana $\frac{1}{2}(9 + 11) = 10$.

Para datos agrupados, la mediana obtenida por interpolación viene dada por

$$\text{Mediana} = L_1 + \left(\frac{\frac{N}{2} - (\sum f)_1}{f_{\text{mediana}}}\right)c \tag{8}$$

=> P50

donde:

L_1 = frontera inferior de la clase de la mediana.
N = número de datos (frecuencia total).
$(\sum f)_1$ = suma de frecuencias de las clases inferiores a la de la mediana.
f_{mediana} = frecuencia de la clase de la mediana.
c = anchura del intervalo de clase de la mediana.

Geométricamente la mediana es el valor de X (abscisa) que corresponde a la recta vertical que divide un histograma en dos partes de igual área. Ese valor de X se suele denotar por $\tilde{X}$.

LA MODA

La *moda* de un conjunto de números es el valor que ocurre con mayor frecuencia; es decir, el valor más frecuente. La moda puede no existir, e incluso no ser única en caso de existir.

EJEMPLO 10. El conjunto 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12 y 18 tiene moda 9.

EJEMPLO 11. El conjunto 3, 5, 8, 10, 12, 15 y 16 no tiene moda.

EJEMPLO 12. El conjunto 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 y 9 tiene dos modas, 4 y 7, y se llama *bimodal.*

Una distribución con moda única se dice *unimodal.*

En el caso de datos agrupados donde se haya construido una curva de frecuencias para ajustar los datos, la moda será el valor (o valores) de X correspondiente al máximo (o máximos) de la curva. Ese valor de X se denota por $\hat{X}$.

La moda puede deducirse de una distribución de frecuencias o de un histograma a partir de la fórmula

$$\text{Moda} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)c \tag{9}$$

donde:

L_1 = frontera inferior de la clase modal (clase que contiene a la moda).
Δ_1 = exceso de la frecuencia modal sobre la de la clase inferior inmediata.
Δ_2 = exceso de la frecuencia modal sobre la clase superior inmediata.
c = anchura del intervalo de clase modal.

RELACION EMPIRICA ENTRE MEDIA, MEDIANA Y MODA

Para curvas de frecuencia unimodales que sean poco asimétricas tenemos la siguiente relación empírica

$$\text{Media} - \text{moda} = 3(\text{media} - \text{mediana}) \tag{10}$$

Las Figuras 3.1 y 3.2 muestran las posiciones relativas de la media, la mediana y la moda para curvas de frecuencia asimétricas a derecha e izquierda, respectivamente. Para curvas simétricas, los tres valores coinciden.

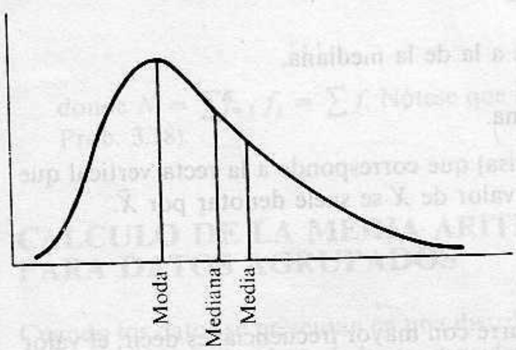


Figura 3.1.

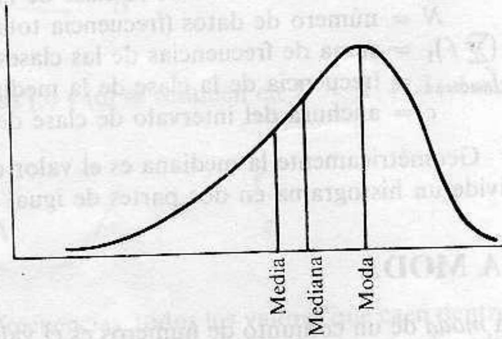


Figura 3.2.

LA MEDIA GEOMETRICA G

La *media geométrica* G de un conjunto de N números positivos $X_1, X_2, X_3, \ldots, X_N$ es la raíz N-ésima del producto de esos números:

$$G = \sqrt[N]{X_1 X_2 X_3 \cdots X_N} \tag{11}$$

EJEMPLO 13. La media geométrica de 2, 4 y 8 es $G = \sqrt[3]{(2)(4)(8)} = \sqrt[3]{64} = 4$.

Podemos calcular G por logaritmos (véase Prob. 3.35) o con una calculadora. Para la media geométrica de datos agrupados, véanse Problemas 3.36 y 3.91.

LA MEDIA ARMONICA H

La *media armónica* H de un conjunto de números $X_1, X_2, X_3, \ldots, X_N$ es el recíproco de la media aritmética de los recíprocos de esos números:

$$H = \frac{1}{\dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\frac{1}{X_j}} = \frac{N}{\displaystyle\sum \frac{1}{X}} \tag{12}$$

En la práctica es más fácil recordar que

$$\frac{1}{H} = \frac{\displaystyle\sum \frac{1}{X}}{N} = \frac{1}{N}\sum \frac{1}{X} \tag{13}$$

EJEMPLO 14. La media armónica de los números 2, 4 y 8 es

$$H = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = \frac{3}{\frac{7}{8}} = 3.43$$

Para la media armónica de datos agrupados, véanse Problemas 3.99 y 3.100.

RELACION ENTRE LAS MEDIAS ARITMETICA, GEOMETRICA Y ARMONICA

La media geométrica de una colección de números positivos $X_1, X_2, \ldots, X_N$ es menor o igual que su media aritmética, pero mayor o igual que su media armónica. En símbolos,

$$H \leqslant G \leqslant \bar{X} \tag{14}$$

La igualdad ocurre si y sólo si todos los números $X_1, X_2, \ldots, X_N$ son idénticos.

EJEMPLO 15. El conjunto 2, 4, 8 tiene media aritmética 4.67, media geométrica 4 y media armónica 3.43.

LA MEDIA CUADRATICA (MQ)

La *media cuadrática* (MQ) de un conjunto de números $X_1, X_2, \ldots, X_N$ se suele denotar por $\sqrt{\overline{X^2}}$ y se define como

$$\mathrm{MQ} = \sqrt{\overline{X^2}} = \sqrt{\frac{\sum_{j=1}^{N} X_j^2}{N}} = \sqrt{\frac{\sum X^2}{N}} \tag{15}$$

Este tipo de promedio se utiliza con frecuencia en las aplicaciones físicas.

EJEMPLO 16. La media cuadrática del conjunto 1, 3, 4, 5 y 7 es

$$\sqrt{\frac{1^2 + 3^2 + 4^2 + 5^2 + 7^2}{5}} = \sqrt{20} = 4.47$$

CUARTILES, DECILES Y PERCENTILES

Si un conjunto de datos está ordenado por magnitud, el valor central (o la media de los dos centrales) que divide al conjunto en dos mitades iguales, es la mediana. Extendiendo esa idea, podemos pensar en aquellos valores que dividen al conjunto en cuatro partes iguales. Esos valores, denotados Q_1, Q_2 y Q_3, se llaman primer, segundo y tercer *cuartiles*, respectivamente. El Q_2 coincide con la mediana.

Análogamente, los valores que dividen a los datos en 10 partes iguales se llaman *deciles*, y se denotan $D_1, D_2, \ldots, D_9$, mientras que los valores que los dividen en 100 partes iguales se llaman *percentiles*, denotados por $P_1, P_2, \ldots, P_{99}$. El 5.° decil y el 50.° percentil coinciden con la mediana. Los 25.° y 75.° percentiles coinciden con el primer y tercer cuartiles.

Colectivamente, cuartiles, deciles y percentiles se denominan *cuantiles*. Para su cálculo con datos agrupados, véanse Problemas 3.44 al 3.46.

PROBLEMAS RESUELTOS

Para Divide en 100
25 25 75 25

NOTACION DE SUMA

3.1. Escribir explícitos los términos en cada suma:

(a) $\sum_{j=1}^{6} X_j$ (c) $\sum_{j=1}^{N} a$ (e) $\sum_{j=1}^{3} (X_j - a)$

Q_1 Q_2 Q_3

(b) $\sum_{j=1}^{4} (Y_j - 3)^2$ (d) $\sum_{k=1}^{5} f_k X_k$

b) $(Y_1-3)^2+(Y_2-3)^2+(Y_3-3)^2+(Y_4-3)^2$

Solución

(a) $X_1 + X_2 + X_3 + X_4 + X_5 + X_6$

(b) $(Y_1 - 3)^2 + (Y_2 - 3)^2 + (Y_3 - 3)^2 + (Y_4 - 3)^2$

c)

(*c*) $a + a + a + \cdots + a = Na$

(*d*) $f_1X_1 + f_2X_2 + f_3X_3 + f_4X_4 + f_5X_5$

(*e*) $(X_1 - a) + (X_2 - a) + (X_3 - a) = X_1 + X_2 + X_3 - 3a$

3.2. Expresar cada suma en notación abreviada de suma:

(*a*) $X_1^2 + X_2^2 + X_3^2 + \cdots + X_{10}^2$ $\qquad \sum_{j=1}^{10} X_j^2$

(*b*) $(X_1 + Y_1) + (X_2 + Y_2) + \cdots + (X_8 + Y_8)$ $\qquad \sum_{j=1}^{8} X_j + Y_j$

(*c*) $f_1X_1^3 + f_2X_2^3 + \cdots + f_{20}X_{20}^3$ $\qquad \sum_{j=1}^{20} f_j X_j^3$

(*d*) $a_1b_1 + a_2b_2 + a_3b_3 + \cdots + a_Nb_N$ $\qquad \sum_{j=1}^{N} a_j b_j$

(*e*) $f_1X_1Y_1 + f_2X_2Y_2 + f_3X_3Y_3 + f_4X_4Y_4$ $\qquad \sum_{j=1}^{4} f_j X_j Y_j$

Solución

(*a*) $\sum_{j=1}^{10} X_j^2$ $\qquad$ (*c*) $\sum_{j=1}^{20} f_jX_j^3$ $\qquad$ (*e*) $\sum_{j=1}^{4} f_jX_jY_j$

(*b*) $\sum_{j=1}^{8} (X_j + Y_j)$ $\qquad$ (*d*) $\sum_{j=1}^{N} a_jb_j$

3.3. Probar que $\sum_{j=1}^{N} (aX_j + bY_j - cZ_j) = a\sum_{j=1}^{N} X_j + b\sum_{j=1}^{N} Y_j - c\sum_{j=1}^{N} Z_j$, donde a, b y c son constantes.

Solución

$$
\begin{aligned}
\sum_{j=1}^{N} (aX_j + bY_j - cZ_j) &= (aX_1 + bY_1 - cZ_1) + (aX_2 + bY_2 - cZ_2) + \cdots + (aX_N + bY_N - cZ_N) \\
&= (aX_1 + aX_2 + \cdots + aX_N) + (bY_1 + bY_2 + \cdots + bY_N) - (cZ_1 + cZ_2 + \cdots + cZ_N) \\
&= a(X_1 + X_2 + \cdots + X_N) + b(Y_1 + Y_2 + \cdots + Y_N) - c(Z_1 + Z_2 + \cdots + Z_N) \\
&= a\sum_{j=1}^{N} X_j + b\sum_{j=1}^{N} Y_j - c\sum_{j=1}^{N} Z_j
\end{aligned}
$$

o más abreviado, $\sum (aX + bY - cZ) = a\sum X + b\sum Y - c\sum Z$.

3.4. Dos variables X e Y toman los valores $X_1 = 2$, $X_2 = -5$, $X_3 = 4$, $X_4 = -8$ e $Y_1 = -3$, $Y_2 = -8$, $Y_3 = 10$, $Y_4 = 6$, respectivamente. Calcular: (*a*) $\sum X$, (*b*) $\sum Y$, (*c*) $\sum XY$, (*d*) $\sum X^2$, (*e*) $\sum Y^2$, (*f*) $(\sum X)(\sum Y)$, (*g*) $\sum XY^2$ y (*h*) $\sum (X + Y)(X - Y)$.

Solución

Nótese que en cada caso el subíndice j de X e Y ha sido omitido, y la $\sum$ se entiende como $\sum_{j=1}^{4}$. Así pues, $\sum X$, por ejemplo, es una abreviatura para $\sum_{j=1}^{4} X_j$.

(*a*) $\sum X = (2) + (-5) + (4) + (-8) = 2 - 5 + 4 - 8 = -7$

(*b*) $\sum Y = (-3) + (-8) + (10) + (6) = -3 - 8 + 10 + 6 = 5$

(*c*) $\sum XY = (2)(-3) + (-5)(-8) + (4)(10) + (-8)(6) = -6 + 40 + 40 - 48 = 26$

(*d*) $\sum X^2 = (2)^2 + (-5)^2 + (4)^2 + (-8)^2 = 4 + 25 + 16 + 64 = 109$

(*e*) $\sum Y^2 = (-3)^2 + (-8)^2 + (10)^2 + (6)^2 = 9 + 64 + 100 + 36 = 209$

(*f*) $(\sum X)(\sum Y) = (-7)(5) = -35$, usando las partes (*a*) y (*b*). Nótese que $(\sum X)(\sum Y) \neq \sum XY$.

(*g*) $\sum XY^2 = (2)(-3)^2 + (-5)(-8)^2 + (4)(10)^2 + (-8)(6)^2 = -190$

(*h*) $\sum (X + Y)(X - Y) = \sum (X^2 - Y^2) = \sum X^2 - \sum Y^2 = 109 - 209 = -100$, usando las partes (*d*) y (*e*).

3.5. Si $\sum_{j=1}^{6} X_j = -4$ y $\sum_{j=1}^{6} X_j^2 = 10$, calcular: (*a*) $\sum_{j=1}^{6} (2X_j + 3)$, (*b*) $\sum_{j=1}^{6} X_j(X_j - 1)$ y (*c*) $\sum_{j=1}^{6} (X_j - 5)^2$.

Solución

(*a*) $$\sum_{j=1}^{6} (2X_j + 3) = \sum_{j=1}^{6} 2X_j + \sum_{j=1}^{6} 3 = 2\sum_{j=1}^{6} X_j + (6)(3) = 2(-4) + 18 = 10$$

(*b*) $$\sum_{j=1}^{6} X_j(X_j - 1) = \sum_{j=1}^{6} (X_j^2 - X_j) = \sum_{j=1}^{6} X_j^2 - \sum_{j=1}^{6} X_j = 10 - (-4) = 14$$

(*c*) $$\sum_{j=1}^{6} (X_j - 5)^2 = \sum_{j=1}^{6} (X_j^2 - 10X_j + 25) = \sum_{j=1}^{6} X_j^2 - 10\sum_{j=1}^{6} X_j + 25(6) = 10 - 10(-4) + 25(6) = 200$$

Si se desea, puede omitirse el subíndice j y usar $\sum$ en lugar de $\sum_{j=1}^{6}$ siempre que se manejen con soltura estas abreviaturas.

LA MEDIA ARITMETICA

3.6. Las notas de un estudiante en seis exámenes fueron 84, 91, 72, 68, 87 y 78. Hallar la media aritmética.

Solución

$$\bar{X} = \frac{\sum X}{N} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = \frac{480}{6} = 80$$

A menudo se usa el término *promedio* como sinónimo de *media aritmética*. Estrictamente hablando, sin embargo, esto es incorrecto, porque hay otros promedios además de la media aritmética.

3.7. Diez medidas del diámetro de un cilindro fueron anotadas por un científico como 3.88, 4.09, 3.92, 3.97, 4.02, 3.95, 4.03, 3.92, 3.98 y 4.06 centímetros (cm). Hallar la media aritmética de tales medidas.

Solución

$$\bar{X} = \frac{\sum X}{N} = \frac{3.88 + 4.09 + 3.92 + 3.97 + 4.02 + 3.95 + 4.03 + 3.92 + 3.98 + 4.06}{10} = \frac{39.82}{10} = 3.98 \text{ cm}$$

3.8. Los salarios anuales de 4 individuos son \$15,000, \$16,000, \$16,500 y \$40,000.
(*a*) Hallar su media aritmética.
(*b*) ¿Puede decirse que ese promedio es *típico* de dichos salarios?

Solución

(*a*) Supuesto que todas las cifras eran significativas en los salarios anotados,

$$\bar{X} = \frac{\$15{,}000 + \$16{,}000 + \$16{,}500 + \$40{,}000}{4} = \frac{\$87{,}500}{4} = \$21{,}875$$

(*b*) La media \$21,875 no es ciertamente típica de esos salarios, y presentarla como un promedio sin más comentarios sería muy engañoso. Una gran desventaja de la media es que se ve muy afectada por valores extremos.

3.9. Hallar la media aritmética de los números 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5 y 4.

Solución

Primer método

$$\bar{X}=\frac{\sum X}{N}=\frac{5+3+6+5+4+5+2+8+6+5+4+8+3+4+5+4+8+2+5+4}{20}=\frac{96}{20}=4.8$$

Segundo método

Hay 6 cincos, 2 treses, 5 cuatros, 2 doses y 3 ochos. Luego

$$\bar{X}=\frac{\sum fX}{\sum f}=\frac{\sum fX}{N}=\frac{(6)(5)+(2)(3)+(2)(6)+(5)(4)+(2)(2)+(3)(8)}{6+2+2+5+2+3}=\frac{96}{20}=4.8$$

3.10. De entre 100 números, 20 son cuatros, 40 son cincos, 30 son seises y los restantes sietes. Hallar su media aritmética.

Solución

$$\bar{X}=\frac{\sum fX}{\sum f}=\frac{\sum fX}{N}=\frac{(20)(4)+(40)(5)+(30)(6)+(10)(7)}{100}=\frac{530}{100}=5.30$$

4 | 20
5 | 40
6 | 30
7 | 10
100

3.11. Las calificaciones finales de un estudiante en cuatro asignaturas fueron 82, 86, 90 y 70. Si los respectivos créditos otorgados a esos cursos son 3, 5, 3 y 1, determinar una calificación media apropiada.

Solución

Usamos una media aritmética ponderada, con pesos dados por los créditos otorgados. Así pues,

$$\bar{X}=\frac{\sum wX}{\sum w}=\frac{(3)(82)+(5)(86)+(3)(90)+(1)(70)}{3+5+3+1}=85$$

3.12. De los 80 empleados de una empresa, 60 cobran \$7,00 a la hora y el resto \$4,00 a la hora.

7 | 60
4 | 20
80

(*a*) Hallar cuánto cobran de media por hora.
(*b*) ¿Sería idéntica la respuesta si los 60 cobraran de media \$4,00 a la hora? Demuestre su respuesta.
(*c*) ¿Cree que la media es representativa?

Solución

(*a*)
$$\bar{X}=\frac{\sum fX}{N}=\frac{(60)(\$7{,}00)+(20)(\$4{,}00)}{60+20}=\frac{\$500.00}{80}=\$6.25$$

(*b*) Sí, el resultado es el mismo. Para verlo, supongamos que f_1 números tienen media m_1 y que f_2 números tienen media m_2. Debemos probar que la media de todos esos números es

$$\bar{X}=\frac{f_1m_1+f_2m_2}{f_1+f_2}$$

Sea M_1 la suma de los f_1 números y M_2 la de los otros f_2. Entonces, por definición de media aritmética,

$$m_1=\frac{M_1}{f_1}\qquad m_2=\frac{M_2}{f_2}$$

o sea $M_1 = f_1 m_1$ y $M_2 = f_2 m_2$. Cuando los $(f_1 + f_2)$ números suman $(M_1 + M_2)$, la media aritmética de todos los números es

$$\bar{X} = \frac{M_1 + M_2}{f_1 + f_2} = \frac{f_1 m_1 + f_2 m_2}{f_1 + f_2}$$

como habíamos anunciado. El resultado se generaliza con facilidad.

(c) Podemos decir que \$6.25 es representativo en el sentido de que la mayoría de los trabajadores cobra \$7.00 a la hora, que no difiere mucho de \$6.25. Hay que recordar que siempre que resumimos datos numéricos en un solo número (un promedio, por ejemplo), estamos abocados a cometer algun error. No obstante, el resultado no es tan engañoso como el del Problema 3.8.

Realmente, para pisar suelo firme, es preciso dar alguna estimación de la «dispersión» o «variación» de los datos respecto de la media (u otro promedio). Eso se llama *dispersión* de los datos. En el Capítulo 4 veremos diversas medidas de la dispersión.

3.13. Cuatro grupos de estudiantes, consistentes en 15, 20, 10 y 18 individuos, dieron pesos medios de 162, 148, 153 y 140 lb, respectivamente. Hallar el peso medio de todos esos estudiantes.

Solución

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{(15)(162) + (20)(148) + (10)(153) + (18)(140)}{15 + 20 + 10 + 18} = 150 \text{ lb}$$

3.14. Si los ingresos medios anuales de los trabajadores agrícolas y no agrícolas en EE.UU. son \$9000 y \$15,000, respectivamente, ¿la media anual de todos ellos sería $\frac{1}{2}$(\$9000 + \$15,000) = \$12,000?

Solución

Sería \$12,000 sólo si hubiera tantos trabajadores de un tipo como de otro. Para hallar la verdadera media sería necesario conocer los números relativos de trabajadores de cada tipo. Si, por ejemplo, hay uno agrícola por cada diez no agrícolas, la media será

$$\bar{X} = \frac{(1)(\$9000) + (11)(\$15{,}000)}{1 + 11} = \$14{,}500$$

Es una media aritmética ponderada.

3.15. Usar la distribución de frecuencias de alturas en la Tabla 2.1 para hallar la altura media de esos 100 estudiantes.

Solución

La Tabla 3.1 indica cómo se hace. Nótese que todos los estudiantes que tienen entre 60 y 62 in, o entre 63 y 65, etc., se consideran como de 61 in, 64 in, etc. El problema se reduce entonces a hallar la altura media de 100 estudiantes, de los cuales 5 miden 61 in, 18 miden 64 in, etc.

Los cálculos exigidos pueden ser tediosos, sobre todo para casos de números grandes y con muchas clases. Hay técnicas que acortan el trabajo; véanse, por ejemplo, los Problemas 3.20 y 3.22.

Tabla 3.1

Altura (in)	Marca de clase (X)	Frecuencia (f)	fX
60-62	61	5	305
63-65	64	18	1152
66-68	67	42	2814
69-71	70	27	1890
72-74	73	8	584
		$N=\sum f=100$	$\sum fX=6745$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N} = \frac{6745}{100} = 67.45 \text{ in}$$

PROPIEDADES DE LA MEDIA ARITMETICA

3.16. Probar que la suma de desviaciones de $X_1, X_2, \ldots, X_N$ respecto de su media $\bar{X}$ es cero.

Solución

Sean $d_1 = X_1 - \bar{X}$, $d_2 = X_2 - \bar{X}$, ..., $d_N = X_N - \bar{X}$ las desviaciones de $X_1, X_2, \ldots, X_N$ respecto de su media $\bar{X}$. Entonces

$$\text{Suma de desviaciones} = \sum d_j = \sum (X_j - \bar{X}) = \sum X_j - N\bar{X}$$

$$= \sum X_j - N\left(\frac{\sum X_j}{N}\right) = \sum X_j - \sum X_j = 0$$

donde hemos usado $\sum$ en vez de $\sum_{j=1}^{N}$. Hubiéramos podido omitir el subíndice j en X_j, supuesto que queda *sobreentendido*.

3.17. Si $Z_1 = X_1 + Y_1$, $Z_2 = X_2 + Y_2$, ..., $Z_N = X_N + Y_N$, probar que $\bar{Z} = \bar{X} + \bar{Y}$.

Solución

Por definición,

$$\bar{X} = \frac{\sum X}{N} \qquad \bar{Y} = \frac{\sum Y}{N} \qquad \bar{Z} = \frac{\sum Z}{N}$$

Luego

$$\bar{Z} = \frac{\sum Z}{N} = \frac{\sum (X + Y)}{N} = \frac{\sum X + \sum Y}{N} = \frac{\sum X}{N} + \frac{\sum Y}{N} = \bar{X} + \bar{Y}$$

donde los subíndices j en X, Y y Z han sido suprimidos, y donde $\sum$ significa $\sum_{j=1}^{N}$.

3.18. (*a*) Si N números $X_1, X_2, \ldots, X_N$ tienen desviaciones respecto de un número A dadas por $d_1 = X_1 - A$, $d_2 = X_2 - A$, ..., $d_N = X_N - A$, respectivamente, probar que

$$\bar{X} = A + \frac{\sum_{j=1}^{N} d_j}{N} = A + \frac{\sum d}{N}$$

(*b*) En el caso de que $X_1, X_2, ..., X_K$ tengan, respectivamente, frecuencias $f_1, f_2, ..., f_K$ y $d_1 = X_1 - A, ..., d_K = X_K - A$, probar que el resultado de la parte (*a*) queda sustituido por

$$\bar{X} = A + \frac{\sum_{j=1}^{K} f_j d_j}{\sum_{j=1}^{K} f_j} = A + \frac{\sum fd}{N} \quad \text{donde} \sum_{j=1}^{K} f_j = \sum f = N$$

Solución

(*a*) *Primer método:*

Como $d_j = X_j - A$ y $X_j = A + d_j$, se tiene

$$\bar{X} = \frac{\sum X_j}{N} = \frac{\sum (A + d_j)}{N} = \frac{\sum A + \sum d_j}{N} = \frac{NA + \sum d_j}{N} = A + \frac{\sum d_j}{N}$$

donde hemos usado $\sum$ en vez de $\sum_{j=1}^{N}$ por brevedad.

Segundo método:

Tenemos $d = X - A$, o sea $X = A + d$, omitiendo los subíndices en d y X. Luego por el Problema 3.17,

$$\bar{X} = \bar{A} + \bar{d} = A + \frac{\sum d}{N}$$

ya que la media de un conjunto de constantes iguales todas a A es A.

(*b*)

$$\bar{X} = \frac{\sum_{j=1}^{K} f_j X_j}{\sum_{j=1}^{K} f_j} = \frac{\sum f_j X_j}{N} = \frac{\sum f_j (A + d_j)}{N} = \frac{\sum A f_j + \sum f_j d_j}{N} = \frac{A \sum f_j + \sum f_j d_j}{N}$$

$$= \frac{AN + \sum f_j d_j}{N} = A + \frac{\sum f_j d_j}{N} = A + \frac{\sum fd}{N}$$

Hagamos notar que *formalmente* el resultado se obtiene de (*a*) sustituyendo d_j por $f_j d_j$ y sumando desde $j = 1$ hasta K en vez de hacerlo desde $j = 1$ hasta N. El resultado es equivalente a $\bar{X} = A + \bar{d}$, donde $\bar{d} = (\sum fd)/N$.

CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS

3.19. Usar el método del Problema 3.18(*a*) para hallar la media aritmética de los números 5, 8, 11, 9, 12, 6, 14 y 10, escogiendo como «media conjeturada» A los valores: (*a*) 9 y (*b*) 20.

Solución

(*a*) Las desviaciones de los números dados respecto de A son $-4, -1, 2, 0, 3, -3, 5$ y 1, y la suma de estas desviaciones es $\sum d = -4 - 1 + 2 + 0 + 3 - 3 + 5 + 1 = 3$. Por tanto

$$\bar{X} = A + \frac{\sum d}{N} = 9 + \frac{3}{8} = 9.375$$

(*b*) Las desviaciones respecto de 20 son $-15, -12, -9, -11, -8, -14, -6$ y -10, y $\sum d = -85$. Así pues,

$$\bar{X} = A + \frac{\sum d}{N} = 20 + \frac{(-85)}{8} = 9.375$$

3.20. Usar el método del Problema 3.18(*b*) para hallar la media aritmética de las alturas del Problema 3.15.

Solución

El método queda indicado en la Tabla 3.2. Tomamos como media conjeturada la marca de clase 67 (que tiene la máxima frecuencia), aunque podría usarse cualquier marca de clase. Observemos que los cálculos son más sencillos que los del Problema 3.15. Para abreviarlos aún más, podemos proceder como en el Problema 3.22, haciendo uso de que las desviaciones (columna 2 de la Tabla 3.2) son todas múltiplos enteros de la anchura del intervalo de clase.

Tabla 3.2

Marca de clase (X)	Desviación $d = X - A$	Frecuencia (f)	fd
61	-6	5	-30
64	-3	18	-54
$A \rightarrow 67$	0	42	0
70	3	27	81
73	6	8	48
		$N = \sum f = 100$	$\sum fd = 45$

$$\bar{X} = A + \frac{\sum fd}{N} = 67 + \frac{45}{100} = 67.45 \text{ in}$$

3.21. Sea $d_j = X_j - A$ las desviaciones de cada marca de clase en una distribución de frecuencias respecto de una marca de clase dada A. Probar que si todos los intervalos de clase tienen la misma anchura c, entonces: (*a*) las desviaciones son todas múltiplos de c (es decir, $d_j = cu_j$, donde $u_j = 0, \pm 1, \pm 2, \ldots$) y (*b*) la media aritmética es calculable mediante la fórmula

$$\bar{X} = A + \left(\frac{\sum fu}{N}\right)c$$

Solución

(*a*) El resultado se ilustra en la Tabla 3.2 del Problema 3.20, donde se ve que las desviaciones en la columna 2 son todas múltiplos de la anchura $c = 3$ in.

Para ver que el resultado es cierto en general, notemos que si $X_1, X_2, X_3, \ldots$ son sucesivas marcas de clase, su diferencia común será igual a c, de modo que $X_2 = X_1 + c$, $X_3 = X_1 + 2c$, y en general $X_j = X_1 + (j - 1)c$. Entonces, cualquier par de marcas de clase, digamos X_p y X_q, difieren en

$$X_p - X_q = [X_1 + (p - 1)c] - [X_1 + (q - 1)c] = (p - q)c$$

que es múltiplo de c.

(*b*) Por la parte (*a*), las desviaciones de todas las marcas de clase respecto de cualquiera de ellas son múltiplos de c (o sea, $d_j = cu_j$). Usando el Problema 3.18(*b*), tendremos

$$\bar{X} = A + \frac{\sum f_j d_j}{N} = A + \frac{\sum f_j(cu_j)}{N} = A + c\frac{\sum f_j u_j}{N} = A + \left(\frac{\sum fu}{N}\right)c$$

Nótese que esto es equivalente al resultado $\bar{X} = A + c\bar{u}$, que puede obtenerse de $\bar{X} = A + \bar{d}$ haciendo $d = cu$ y observando que $\bar{d} = c\bar{u}$ (véase Prob. 3.18).

3.22. Hacer uso del resultado del Problema 3.21(*b*) para hallar la altura media de los 100 estudiantes del Problema 3.20.

Solución

El método, resumido en la Tabla 3.3, se llama *método de compilación*, y debe utilizarse siempre que sea posible.

Tabla 3.3

	X	u	f	fu
	61	−2	5	−10
	64	−1	18	−18
$A \longrightarrow$	67	0	42	0
	70	1	27	27
	73	2	8	16
			$N = 100$	$\sum fu = 15$

$$\bar{X} = A + \left(\frac{\sum fu}{N}\right)c = 67 + \left(\frac{15}{100}\right)(3) = 67.45 \text{ in}$$

3.23. Calcular el salario semanal medio de los 65 empleados de la empresa P&R a partir de la distribución de frecuencias de la Tabla 2.5, usando: (*a*) el método largo y (*b*) el método de compilación.

Solución

Las Tablas 3.4 y 3.5 muestran las respectivas soluciones a (*a*) y (*b*).

Cabe suponer que se ha introducido error en esas tablas porque las marcas de clase verdaderas son \$254.995, \$264.995, etc., en lugar de \$255.00, \$265.00, etc. Si se usan en la Tabla 3.4 esas marcas de clase verdaderas en vez de las otras, $\bar{X}$ resulta ser \$279.76 en vez de \$279.77, y la diferencia es despreciable.

$$\bar{X} = \frac{\sum fX}{N} = \frac{\$18{,}185.00}{65} = \$279.77 \qquad \bar{X} = A + \left(\frac{\sum fu}{N}\right)c = \$275.000 + \frac{31}{65}(\$10.00) = \$279.77$$

Tabla 3.4

X	f	fX
\$255.00	8	\$2040.00
265.00	10	2650.00
275.00	16	4400.00
285.00	14	3990.00
295.00	10	2950.00
305.00	5	1525.00
315.00	2	630.00
	$N = 65$	$\sum fX = \$18{,}185.00$

Tabla 3.5

X	u	f	fu
\$255.00	−2	8	−16
265.00	−1	10	−10
$A \rightarrow$ 275.00	0	16	0
285.00	1	14	14
295.00	2	10	20
305.00	3	5	15
315.00	4	2	8
		$N = 65$	$\sum fu = 31$

3.24. Usando la Tabla 2.9(*d*), hallar el salario medio de los 70 trabajadores de la empresa P&R.

Solución

En este caso, los intervalos de clase no son de la misma anchura y hemos de recurrir al método largo, como muestra la Tabla 3.6.

Tabla 3.6

X	f	fX
\$255.00	8	\$2040.00
265.00	10	2650.00
275.00	16	4400.00
285.00	15	4275.00
295.00	10	2950.00
310.00	8	2480.00
350.00	3	1050.00
	$N = 70$	$\sum fX = \$19{,}845.00$

$$\bar{X} = \frac{\sum fX}{N} = \frac{\$19{,}845.00}{70} = \$283.50$$

LA MEDIANA

3.25. Las notas de un estudiante en seis exámenes han sido 84, 91, 72, 68, 87 y 78. Hallar la mediana de esas notas.

Solución

Las notas ordenadas son 68, 72, 78, 84, 87 y 91. Como hay un número par de ellas, hay dos valores centrales, 78 y 84, cuya media aritmética $\frac{1}{2}(78 + 84) = 81$ es la nota pedida. Comparar con el Problema 3.6, donde la media aritmética era 80.

3.26. Cinco oficinistas cobran \$4.52, \$5.96, \$5.28, \$11.20 y \$5.75 a la hora. Hallar: (*a*) la mediana y (*b*) la media de esas cantidades.

Solución

(*a*) Los salarios, en ordenación, son \$4.52, \$5.28, \$5.75, \$5.96 y \$11.20. Como hay un número impar de ellos, sólo hay un valor central, \$5.75, que es la mediana.

(*b*) La media aritmética es

$$\frac{\$4.52 + \$5.96 + \$5.28 + \$11.20 + \$5.75}{5} = \$6.54$$

Nótese que la mediana no se ve afectada por el valor extremo \$11.20, mientras que la media sí. En este caso, la mediana da mejor indicación del salario medio que la media.

3.27. Si (*a*) 85 y (*b*) 150 números se ordenan, ¿cómo calcularía la mediana de esos números?

Solución

(*a*) Como hay 85 números, y 85 es impar, el único valor central es el 43.°, y ese es la mediana. Deja 42 números a cada lado.

(*b*) Ahora 150 es par, y hay dos valores centrales, el 75.° y el 76.°. Dejan 74 números a cada lado. Su promedio es la mediana.

3.28. Del Problema 2.8, hallar la mediana de los pesos de esos 40 estudiantes, usando: (*a*) la distribución de frecuencias de la Tabla 2.7 (reproducida aquí como Tabla 3.7) y (*b*) los datos originales.

Solución

(*a*) *Primer método* (por interpolación)

Los pesos en la distribución de frecuencias de la Tabla 3.7 se suponen distribuidos continuamente. En tal caso, la mediana es aquel peso para el que la mitad de la frecuencia total (40/2 = 20) quede por encima y la mitad por debajo.

Tabla 3.7

Peso (lb)	Frecuencia	
118-126	3	3
127-135	5	8
136-144	9	17
145-153	12	29
154-162	5	34
163-171	4	38
172-180	2	40
	Total 40	

Ahora bien, la suma de las tres primeras frecuencias de clase es 3 + 5 + 9 = 17. Luego para llegar al 20 deseado tomamos 3 más de entre los 12 casos de la cuarta clase. Puesto que el cuarto intervalo de clase, 145-153, realmente corresponde a pesos desde 144.5 a 153.5, la mediana debe estar a 3/12 de camino entre 144.5 y 153.5; es decir, la mediana es

$$144.5 + \frac{3}{12}(153.5 - 144.5) = 144.5 + \frac{3}{12}(9) = 146.8 \text{ lb}$$

Segundo método (usando la fórmula)

Como la suma de las frecuencias de las tres y cuatro primeras clases son $3 + 5 + 9 = 17$ y $3 + 5 + 9 + 12 = 29$, respectivamente, es claro que la mediana cae en la cuarta clase, que es, por tanto, la clase de la mediana. Entonces

$$L_1 = \text{frontera de la clase inferior a la de la mediana} = 144.5$$
$$N = \text{Número de datos} = 40$$
$$(\sum f)_1 = \text{suma de las clases inferiores a la de la mediana} = 3 + 5 + 9 = 17$$
$$f_{\text{mediana}} = \text{frecuencia de la clase de la mediana} = 12$$
$$c = \text{tamaño del intervalo de la clase de la mediana} = 9$$

luego

$$\text{Mediana} = L_1 + \left(\frac{N/2 - (\sum f)_1}{f_{\text{mediana}}}\right)c = 144.5 + \left(\frac{40/2 - 17}{12}\right)(9) = 146.8 \text{ lb}$$

(*b*) Ordenados, los pesos originales eran

119, 125, 126, 128, 132, 135, 135, 135, 136, 138, 138, 140, 140, 142, 142, 144, 144, 145, 145, 146, 146, 147, 147, 148, 149, 150, 150, 152, 153, 154, 156, 157, 158, 161, 163, 164, 165, 168, 173, 176

La mediana es la media aritmética de los pesos 20.° y 21.° en esa ordenación, a saber, 146 lb.

3.29. Mostrar cómo se puede obtener el peso mediana en el Problema 3.28 de: (*a*) un histograma y (*b*) una ojiva de porcentajes.

Solución

(*a*) La Figura 3.3(*a*) muestra el histograma de los pesos del Problema 3.28. La mediana es la abscisa correspondiente a la recta LM, que divide el histograma en dos áreas iguales. Como en un histograma el área corresponde a la frecuencia, LM es tal que el área total a izquierda y a derecha es la mitad de la frecuencia total, o sea, 20. Así pues, las áreas $AMLD$ y $MBEL$ corresponden a frecuencias de 3 y 9. Entonces, $AM = \frac{3}{12}AB = \frac{3}{12}(9) = 2.25$, y la mediana es $144.5 + 2.25 = 146.75$, o sea 146.8 lb redondeada a la décima de libra. El valor aproximado puede adivinarse del gráfico.

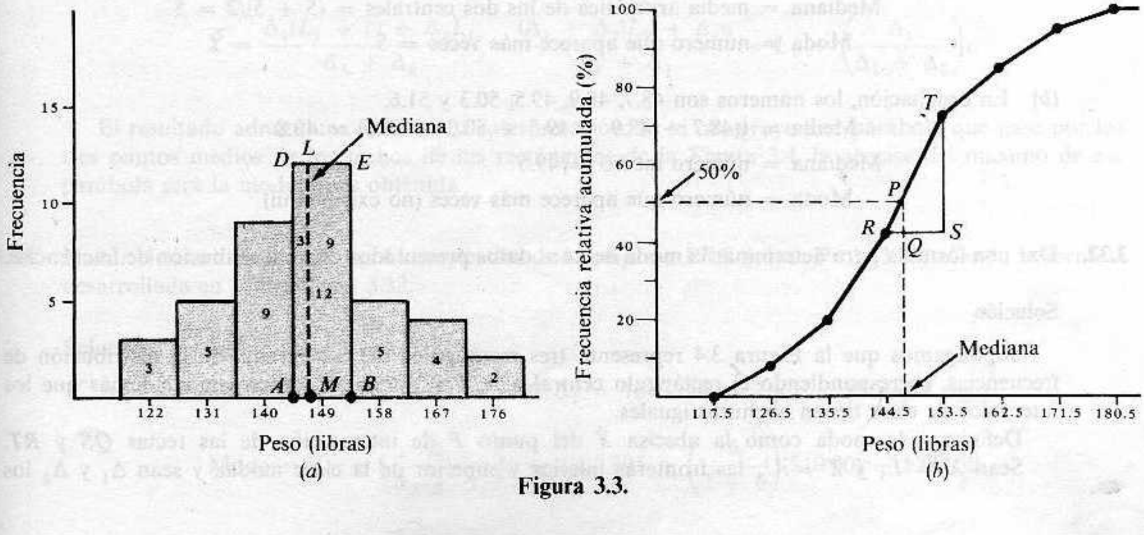


Figura 3.3.

(*b*) La Figura 3.3(*b*) muestra el polígono de frecuencia relativa acumulada (u ojiva de porcentajes) para los pesos del Problema 3.28. La mediana es la abscisa del punto P en esa ojiva, cuya ordenada es 50%. Para calcular ese valor, vemos de los triángulos semejantes PQR y RST que

$$\frac{RQ}{RS} = \frac{PQ}{ST} \quad \text{o sea} \quad \frac{RQ}{9} = \frac{50\% - 42\%}{72.5\% - 42.5\%} = \frac{1}{4} \quad \text{así que} \quad RQ = \frac{9}{4} = 2.25$$

Por tanto

$$\text{Mediana} = 144.5 + RQ = 144.5 + 2.25 = 146.75 \text{ lb}$$

o sea 146.8 lb, con precisión de décima de libra.. Este valor puede verse también aproximadamente en el gráfico.

3.30. Hallar la paga media de los 65 empleados de la empresa P&R (véase Prob. 2.3).

Solución

Aquí $N = 65$ y $N/2 = 32.5$. Como las sumas de las primeras dos y tres frecuencias de clase son $8 + 10 = 18$ y $8 + 10 + 16 = 34$, respectivamente, la clase de la mediana es la tercera. Usando la fórmula,

$$\text{Mediana} = L_1 + \left(\frac{N/2 - (\sum f)_1}{f_{\text{mediana}}}\right)c = \$269.995 + \left(\frac{32.5 - 18}{16}\right)(\$10.00) = \$279.06$$

LA MODA

3.31. Hallar la media, la mediana y la moda para los conjuntos: (*a*) 3, 5, 2, 6, 5, 9, 5, 2, 8, 6 y (*b*) 51.6, 48.7, 50.3, 49.5, 48.9.

Solución

(*a*) Ordenados, los números son 2, 2, 3, 5, 5, 5, 6, 6, 8 y 9.

$$\text{Media} = \tfrac{1}{10}(2 + 2 + 3 + 5 + 5 + 5 + 6 + 6 + 8 + 9) = 5.1$$
$$\text{Mediana} = \text{media aritmética de los dos centrales} = (5 + 5)/2 = 5$$
$$\text{Moda} = \text{número que aparece más veces} = 5$$

(*b*) En ordenación, los números son 48.7, 48.9, 49.5, 50.3 y 51.6.

$$\text{Media} = \tfrac{1}{5}(48.7 + 48.9 + 49.5 + 50.3 + 51.6) = 49.8$$
$$\text{Mediana} = \text{número medio} = 49.5$$
$$\text{Moda} = \text{número que aparece más veces (no existe aquí)}$$

3.32. Dar una fórmula para determinar la moda de unos datos presentados como distribución de frecuencias.

Solución

Supongamos que la Figura 3.4 representa tres rectángulos del histograma de la distribución de frecuencias, correspondiendo el rectángulo central a la clase modal. Y supongamos además que los intervalos de clase tienen anchuras iguales.

Definimos la moda como la abscisa $\hat{X}$ del punto P de intersección de las rectas QS y RT. Sean $X = L_1$ y $X = U_1$ las fronteras inferior y superior de la clase modal, y sean Δ_1 y Δ_2 los excesos de frecuencia de la clase modal sobre las de las clases adyacentes a izquierda y derecha, respectivamente.

De los triángulos semejantes PQR y PST, tenemos

$$\frac{EP}{RQ} = \frac{PF}{ST} \quad \text{o sea} \quad \frac{\hat{X} - L_1}{\Delta_1} = \frac{U_1 - \hat{X}}{\Delta_2}$$

Entonces

$$\Delta_2(\hat{X} - L_1) = \Delta_1(U_1 - \hat{X}) \qquad \Delta_2\hat{X} - \Delta_2 L_1 = \Delta_1 U_1 - \Delta_1\hat{X} \qquad (\Delta_1 + \Delta_2)\hat{X} = \Delta_1 U_1 + \Delta_2 L_1$$

o

$$\hat{X} = \frac{\Delta_1 U_1 + \Delta_2 L_1}{\Delta_1 + \Delta_2}$$

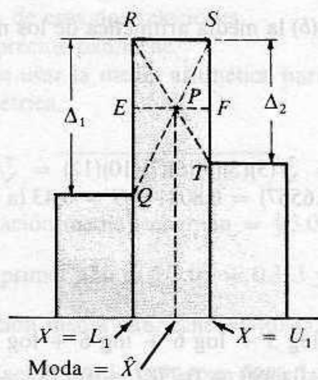


Figura 3.4.

Como $U_1 = L_1 + c$, donde c es la anchura de los intervalos de clase, eso se convierte en

$$\hat{X} = \frac{\Delta_1(L_1 + c) + \Delta_2 L_1}{\Delta_1 + \Delta_2} = \frac{(\Delta_1 + \Delta_2)L_1 + \Delta_1 c}{\Delta_1 + \Delta_2} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)c$$

El resultado admite una interesante interpretación: Si se construye una parábola que pase por los tres puntos medios de los techos de los rectángulos de la Figura 3.4, la abscisa del máximo de esa parábola será la moda antes obtenida.

3.33. Hallar el salario modal de los 65 empleados de la empresa P&R (véase Prob. 3.23) usando la fórmula desarrollada en el Problema 3.32.

Solución

Aquí $L_1 = \$269.995$, $\Delta_1 = 16 - 10 = 6$, $\Delta_2 = 16 - 14 = 2$ y $c = \$10.00$. Luego

$$\text{Moda} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)c = \$269.995 + \left(\frac{6}{2 + 6}\right)(\$10.00) = \$277.50$$

RELACION EMPIRICA ENTRE MEDIA, MEDIANA Y MODA

3.34. (*a*) Usar la fórmula empírica media − moda = 3(media − mediana) para hallar el salario modal de los 65 empleados de la empresa P&R.

(*b*) Comparar el resultado con el del Problema 3.33.

Solución

(*a*) De los Problemas 3.23 y 3.30 tenemos media = \$279.77 y mediana = \$279.06. Entonces

$$\text{Moda} = \text{media} - 3(\text{media} - \text{mediana}) = \$279.77 - 3(\$279.77 - \$279.06) = \$277.64$$

(*b*) Del Problema 3.33 vemos que el salario modal es \$277.50, así que está en buen acuerdo con el resultado empírico.

LA MEDIA GEOMETRICA

3.35. Hallar: (*a*) la media geométrica y (*b*) la media aritmética de los números 3, 5, 6, 6, 7, 10 y 12, supuestos exactos.

Solución

(*a*) La media geométrica $= G = \sqrt[7]{(3)(5)(6)(6)(7)(10)(12)} = \sqrt[7]{453{,}600}$. Usando logaritmos comunes, $\log G = \frac{1}{7}\log 453{,}600 = \frac{1}{7}(5.6567) = 0.8081$ y $G = 6.43$ (a la centésima). Alternativamente, puede usarse una calculadora.

Otro método

$$\begin{aligned}\log G &= \tfrac{1}{7}(\log 3 + \log 5 + \log 6 + \log 6 + \log 7 + \log 10 + \log 12)\\ &= \tfrac{1}{7}(0.4771 + 0.6990 + 0.7782 + 0.7782 + 0.8451 + 1.0000 + 1.0792)\\ &= 0.8081\\ \text{y} \qquad G &= 6.43\end{aligned}$$

(*b*) Media aritmética $= \bar{X} = \frac{1}{7}(3 + 5 + 6 + 6 + 7 + 10 + 12) = 7$. Esto ilustra el hecho de que la media geométrica de un conjunto de números distintos positivos es menor que la media aritmética.

3.36. Los números $X_1, X_2, ..., X_K$ ocurren con frecuencias $f_1, f_2, ..., f_K$, donde $f_1 + f_2 + \cdots + f_K = N$ es la frecuencia total.

(*a*) Hallar su media geométrica G.

(*b*) Deducir una expresión para $\log G$.

(*c*) ¿Cómo pueden usarse esos resultados para hallar la media geométrica de datos agrupados en una distribución de frecuencias?

Solución

(*a*)

$$G = \sqrt[N]{\underbrace{X_1X_1\cdots X_1}_{f_1 \text{ veces}}\ \underbrace{X_2X_2\cdots X_2}_{f_2 \text{ veces}}\cdots \underbrace{X_KX_K\cdots X_K}_{f_K \text{ veces}}} = \sqrt[N]{X_1^{f_1}X_2^{f_2}\cdots X_K^{f_K}}$$

donde $N = \sum f$. Esto se llama a veces la *media geométrica ponderada*.

(*b*)

$$\log G = \frac{1}{N}\log(X_1^{f_1}X_2^{f_2}\cdots X_K^{f_K}) = \frac{1}{N}(f_1\log X_1 + f_2\log X_2 + \cdots + f_K\log X_K)$$

$$= \frac{1}{N}\sum_{j=1}^{K} f_j \log X_j = \frac{\sum f \log X}{N}$$

donde suponemos que todos los números son positivos; de lo contrario, los logaritmos no estarían definidos.

Nótese que el logaritmo de la media geométrica de un conjunto de números es la media aritmética de los logaritmos de tales números.

(*c*) El resultado puede aplicarse para calcular la media geométrica de datos agrupados tomando X_1 X_2, ..., X_K como marca de clase y $f_1, f_2, ..., f_K$ como las correspondientes frecuencias de clase.

3.37. Mientras durante un año la relación entre el precio de la leche (un cuarto de galón) y el de la hogaza de pan era 3.00, al año siguiente pasó a ser 2.00.

(*a*) Hallar la media aritmética de esas dos relaciones.
(*b*) Idem para la relación de precios pan/leche.
(*c*) Discutir la conveniencia de usar la media aritmética para promediar relaciones.
(*d*) Idem para la media geométrica.

Solución

(*a*)

$$\text{Relación media leche/pan} = \tfrac{1}{2}(3.00 + 2.00) = 2.50$$

(*b*) La relación pan/leche del primer año es $1/3.00 = 0.333$ y para el segundo $1/2.00 = 0.500$. Luego

$$\text{Relación media pan/leche} = \tfrac{1}{2}(0.333 + 0.500) = 0.417$$

(*c*) Sería de esperar que la relación media leche/pan fuese la recíproca de la pan/leche, si la media es un promedio adecuado. Sin embargo, $1/0.417 = 2.40 \neq 2.50$. Eso demuestra que la media aritmética es un pobre promedio para manejar cocientes entre magnitudes.

(*d*)

$$\text{Media geométrica de las relaciones leche/pan} = \sqrt{(3.00)(2.00)} = \sqrt{6.00}$$

$$\text{Media geométrica de las relaciones pan/leche} = \sqrt{(0.333)(0.500)} = \sqrt{0.0167} = 1/\sqrt{6.00}$$

Como estos promedios son recíprocos, la conclusión es que la media geométrica es más adecuada que la media aritmética para promediar relaciones del tipo propuesto en este problema.

3.38. La población de bacterias en un cultivo creció de 1000 a 4000 en 3 días. ¿Cuál fue el crecimiento medio diario?

Solución

Ya que de 1000 a 4000 es un 300% de crecimiento, uno podría sospechar que el crecimiento medio diario es $300\%/3 = 100\%$. Sin embargo, eso implicaría que el primer día subiría ya de 1000 a 2000, el segundo a 4000 y el tercero a 8000, contra lo dicho.

Denotemos el crecimiento medio diario por r. Entonces

$$\text{Población de bacterias tras 1 día} = 1000 + 1000r = 1000(1 + r)$$
$$\text{Población de bacterias tras 2 días} = 1000(1 + r) + 1000(1 + r)r = 1000(1 + r)^2$$
$$\text{Población de bacterias tras 3 días} = 1000(1 + r)^2 + 1000(1 + r)^2 r = 1000(1 + r)^3$$

Esta última expresión debe dar 4000. Por tanto, $1000(1 + r)^3 = 4000$, $(1 + r)^3 = 4$, $1 + r = \sqrt[3]{4}$ y $r = \sqrt[3]{4} - 1 = 1.587 - 1 = 0.587$, así que $r = 58.7\%$.

En general, si arrancamos con una cantidad P y crece a razón constante r por unidad de tiempo, tendremos, tras n unidades de tiempo, la cantidad

$$A = P(1 + r)^n$$

Esta es la *fórmula del interés compuesto* (véanse Probs. 3.94 y 3.95).

LA MEDIA ARMONICA

3.39. Hallar la media armónica de los números 3, 5, 6, 6, 7, 10 y 12.

Solución

$$\frac{1}{H} = \frac{1}{N}\sum \frac{1}{X} = \frac{1}{7}\left(\frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6} + \frac{1}{7} + \frac{1}{10} + \frac{1}{12}\right) = \frac{1}{7}\left(\frac{140 + 84 + 70 + 70 + 60 + 42 + 35}{420}\right)$$

$$= \frac{501}{2940}$$

y $$H = \frac{2940}{501} = 5.87$$

A menudo conviene expresar antes las fracciones en forma decimal. Así

$$\frac{1}{H} = \tfrac{1}{7}(0.3333 + 0.2000 + 0.1667 + 0.1667 + 0.1429 + 0.1000 + 0.0833)$$

$$= \frac{1.1929}{7}$$

y $$H = \frac{7}{1.1929} = 5.87$$

La comparación con el Problema 3.35 ilustra el hecho de que la media es menor que la media geométrica, la cual a su vez es menor que la media aritmética.

3.40. Durante cuatro años sucesivos, una familia compró el fuel para su calefacción a \$0.80, \$0.90, \$1.05 y \$1.25 por galón (gal), respectivamente. Hallar el coste medio del fuel en ese período.

Solución

Caso I

Supongamos que consumieron todos los años la misma cantidad, digamos 1000 gal. Entonces

$$\text{Coste medio} = \frac{\text{coste total}}{\text{cantidad total adquirida}} = \frac{\$800 + \$900 + \$1050 + \$1250}{400 \text{ gal}} = \$1.00/\text{gal}$$

Eso es lo mismo que la media aritmética del coste por galón; es decir, $\tfrac{1}{4}(\$0.80 + \$0.90 + \$1.05 + \$1.25) = 1.00/\text{gal}$. El resultado sería el mismo si consumieran x galones al año.

Caso 2

Supongamos que la familia gasta cada año la misma cantidad de dinero en fuel, digamos \$1000. Entonces

$$\text{Coste medio} = \frac{\text{coste total}}{\text{cantidad total adquirida}} = \frac{\$4000}{(1250 + 1111 + 952 + 800)\ \text{gal}} = \$0.975/\text{gal}$$

Esto es lo mismo que la media armónica de los costes por galón:

$$\frac{4}{\dfrac{1}{0.80} + \dfrac{1}{0.90} + \dfrac{1}{1.05} + \dfrac{1}{1.25}} = 0.975$$

El resultado sería el mismo si gastasen y dólares al año.

Ambos procedimientos de promediar son correctos, cada uno en ciertas circunstancias.

Debe observarse que en caso de que el consumo en galones cambiase de año en año, la media aritmética del primer caso vendría sustituida por la media aritmética ponderada. Análogamente, ante un gasto variable en dólares de año en año, la media armónica del segundo caso sería reemplazada por una media armónica ponderada.

3.41. Una persona viaja de A a B con una velocidad media de 30 millas por hora(mi/h) y regresa de B a A a una velocidad media de 60 mi/h. Hallar su velocidad media en el viaje completo.

Solución

Supongamos que A y B distan 60 millas (aunque cualquier distancia valdría). Entonces

$$\text{Tiempo para ir de } A \text{ a } B = \frac{60\ \text{mi}}{30\ \text{mi/h}} = 2\ \text{h} \qquad \text{Tiempo para ir de } B \text{ a } A = \frac{60\ \text{mi}}{60\ \text{mi/h}} = 1\ \text{h}$$

y

$$\text{Velocidad media del viaje total} = \frac{\text{distancia total}}{\text{tiempo total}} = \frac{120\ \text{mi}}{3\ \text{h}} = 40\ \text{mi/h}$$

El promedio anterior es la media armónica de 30 y 60; esto es,

$$\frac{2}{1/30 + 1/60} = 40\ \text{mi/h}$$

Si las distancias recorridas no son iguales, se llega a una media armónica ponderada, donde los pesos son las distancias (véase Prob. 3.102).

Nótese que uno hubiera estado tentado de tomar la media aritmética de 30 y 60 mi/h obteniendo 45 mi/h, lo cual es incorrecto.

LA MEDIA CUADRATICA

3.42. Hallar la media cuadrática de los números 3, 5, 6, 6, 7, 10 y 12.

Solución

$$\text{Media cuadrática} = \text{MQ} = \sqrt{\frac{3^2 + 5^2 + 6^2 + 6^2 + 7^2 + 10^2 + 12^2}{7}} = \sqrt{57} = 7.55$$

3.43. Probar que la media cuadrática de dos números positivos distintos, a y b, es mayor que su media geométrica.

Solución

Tenemos que probar que $\sqrt{\frac{1}{2}(a^2 + b^2)} > \sqrt{ab}$. Si eso es verdad, entonces completando el cuadrado de ambos lados, $\frac{1}{2}(a^2 + b^2) > ab$, de manera que $a^2 + b^2 > 2ab$, $a^2 - 2ab + b^2 > 0$, o sea $(a - b)^2 > 0$. Pero esta última desigualdad es cierta, pues el cuadrado de todo número real no nulo es positivo.

La demostración consiste en volver hacia atrás esos pasos. Así, partiendo de $(a - b)^2 > 0$, que sabemos es cierta, podemos probar que $a^2 + b^2 > 2ab$, $\frac{1}{2}(a^2 + b^2) > ab$, y finalmente $\sqrt{\frac{1}{2}(a^2 + b^2)} > \sqrt{ab}$, como se quería.

Nótese que $\sqrt{\frac{1}{2}(a^2 + b^2)} = \sqrt{ab}$, si y sólo si, $a = b$.

CUARTILES, DECILES Y PERCENTILES

3.44. Hallar: (*a*) los cuartiles Q_1, Q_2 y Q_3, y (*b*) los deciles $D_1, D_2, \ldots, D_9$ para los salarios de los 65 empleados de la empresa P&R (véase Prob. 2.3).

Solución

(*a*) El primer cuartil Q_1 es el salario obtenido contando $N/4 = 65/4 = 16.25$ de los casos, comenzando con la primera clase (la más baja). Como la primera clase contiene 8 casos, debemos tomar 8.25 (16.25 − 8) de los 10 casos de la segunda clase. Por interpolación lineal se tiene

$$Q_1 = \$259.995 + \frac{8.25}{10}(\$10.00) = \$268.25$$

El segundo cuartil Q_2 se obtiene contando los primeros $2N/4 = N/2 = 65/2 = 32.5$ casos. Como las dos primeras clases contienen 18 casos, hay que tomar $32.5 - 18 = 14.5$ de los 16 casos de la tercera clase, es decir

$$Q_2 = \$269.995 + \frac{14.5}{16}(\$10.00) = \$279.06$$

Notemos que Q_2 es la mediana.

El tercer cuartil Q_3 se obtiene contando los primeros $3N/4 = \frac{3}{4}(65) = 48.75$ casos. Ya que las cuatro primeras clases contienen 48 casos, hemos de tomar $48.75 - 48 = 0.75$ de los 10 casos de la quinta; luego

$$Q_3 = \$289.995 + \frac{0.75}{10}(\$10.00) = \$290.75$$

Por tanto, el 25% de los empleados ganan \$268.25 o menos, el 50% \$279.06 o menos, y el 75% \$290.75 o menos.

(*b*) Los deciles primero, segundo y noveno se obtienen contando $N/10, 2N/10, \ldots, 9N/10$ casos a partir de la primera clase. Así pues,

$$D_1 = \$249.995 + \frac{6.5}{8}(\$10.00) = \$258.12 \qquad D_6 = \$279.995 + \frac{5}{14}(\$10.00) = \$283.57$$

$$D_2 = \$259.995 + \frac{5}{10}(\$10.00) = \$265.0 \qquad D_7 = \$279.995 + \frac{11.5}{14}(\$10.00) = \$288.21$$

$$D_3 = \$269.995 + \frac{1.5}{16}(\$10.00) = \$270.94 \qquad D_8 = \$289.995 + \frac{4}{10}(\$10.00) = \$294.00$$

$$D_4 = \$269.995 + \frac{8}{16}(\$10.00) = \$275.00 \qquad D_9 = \$299.995 + \frac{0.5}{5}(\$10.00) = \$301.00$$

$$D_5 = \$269.995 + \frac{14.5}{16}(\$10.00) = \$279.06$$

Por tanto, el 10% de los empleados ganan \$258.12 o menos, el 20% ganan \$265.00 o menos, ..., el 90% ganan \$301.00 o menos.

Nótese que el quinto decil es la mediana. El segundo, cuarto, sexto y octavo deciles, que dividen la distribución en cinco partes iguales, se llaman *quintiles* y a veces son utilizados en la práctica.

3.45. Determinar: (*a*) el 35.º percentil y (*b*) el 60.º percentil para la distribución del Problema 3.44.

Solución

(*a*) El 35.º percentil P_{35} se obtiene contando los primeros $35N/100 = 35(65)/100 = 22.75$ casos, comenzando por la primera clase (la más baja). Entonces, como en el Problema 3.44,

$$P_{35} = \$269.995 + \frac{4.75}{16}(\$10.00) = \$272.97$$

Eso significa que el 35% de los empleados cobran \$272.97 o menos.

(*b*) El 60.º percentil es $P_{60} = \$279.995 + \frac{5}{14}(\$10.00) = \$283.57$. Coincide con el 6.º decil y el tercer quintil.

3.46. Probar que los resultados de los Problemas 3.44 y 3.45 se pueden deducir de una ojiva de porcentajes.

Solución

La ojiva de porcentajes correspondiente a los datos de los Problemas 3.44 y 3.45 se muestra en la Figura 3.5.

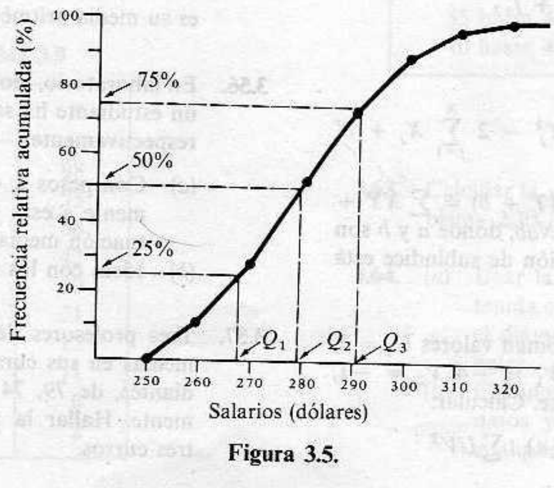


Figura 3.5.

El primer cuartil es la abscisa del punto de la ojiva cuya ordenada es 25%, y análogamente, los cuartiles segundo y tercero son las abscisas de aquellos puntos de la ojiva con ordenadas respectivas 50% y 75%.

De modo parecido se obtienen los deciles y percentiles. Por ejemplo, el 7.° decil y el 35.° percentil son las abscisas de aquellos puntos de la ojiva que tienen ordenadas respectivas 70% y 35%.

PROBLEMAS SUPLEMENTARIOS

NOTACION DE SUMA

3.47. Escribir los términos de cada suma indicada:

(*a*) $\sum_{j=1}^{4}(X_j+2)$ (*d*) $\sum_{k=1}^{N}(Y_k^2-4)$

(*b*) $\sum_{j=1}^{5} f_jX_j^2$ (*e*) $\sum_{j=1}^{4} 4X_jY_j$

(*c*) $\sum_{j=1}^{3} U_j(U_j+6)$

3.48. Expresar en notación abreviada de suma:

(*a*) $(X_1+3)^3+(X_2+3)^3+(X_3+3)^3$

(*b*) $f_1(Y_1-a)^2+f_2(Y_2-a)^2+\cdots+f_{15}(Y_{15}-a)^2$

(*c*) $(2X_1-3Y_1)+(2X_2-3Y_2)+\cdots+(2X_N-3Y_N)$

(*d*) $(X_1/Y_1-1)^2+(X_2/Y_2-1)^2+\cdots+(X_8/Y_8-1)^2$

(*e*) $\dfrac{f_1a_1^2+f_2a_2^2+\cdots+f_{12}a_{12}^2}{f_1+f_2+\cdots+f_{12}}$

3.49. Demostrar que

$$\sum_{j=1}^{N}(X_j-1)^2=\sum_{j=1}^{N}X_j^2-2\sum_{j=1}^{N}X_j+N$$

3.50. Probar que $\sum(X+a)(Y+b)=\sum XY+a\sum Y+b\sum X+Nab$, donde a y b son constantes. ¿Qué notación de subíndice está implícita?

3.51. Dos variables, U y V, toman valores $U_1=3$, $U_2=-2$, $U_3=5$ y $V_1=-4$, $V_2=-1$, $V_3=6$, respectivamente. Calcular:

(*a*) $\sum UV$ (*e*) $\sum UV^2$

(*b*) $\sum(U+3)(V-4)$ (*f*) $\sum(U^2-2V^2+2)$

(*c*) $\sum V^2$ (*g*) $\sum(U/V)$

(*d*) $(\sum U)(\sum V)^2$

3.52. Dado $\sum_{j=1}^4 X_j=7$, $\sum_{j=1}^4 Y_j=-3$ y $\sum_{j=1}^4 X_jY_j=5$, calcular: (*a*) $\sum_{j=1}^4(2X_j+5Y_j)$ y (*b*) $\sum_{j=1}^4(X_j-3)(2Y_j+1)$.

LA MEDIA ARITMETICA

3.53. Las notas de un estudiante han sido 85, 76, 93, 82 y 96. Hallar su media aritmética.

3.54. Los tipos de reacción de un individuo ante diversos estímulos, medidos por un psicólogo, fueron 0.53, 0.46, 0.50, 0.49, 0.52, 0.53, 0.44 y 0.55 segundos, respectivamente. Determinar su tiempo medio de reacción.

3.55. Un conjunto de números contiene 6 seises, 7 sietes, 8 ochos, 9 nueves y 10 dieces. ¿Cuál es su media aritmética?

3.56. En laboratorio, teoría y problemas de Física, un estudiante ha sacado 71, 78 y 89 puntos, respectivamente.

(*a*) Con pesos 2, 4, 5 asignados respectivamente a esas pruebas, ¿cuál es su puntuación media?

(*b*) Idem con los tres pesos iguales.

3.57. Tres profesores de Economía dieron notas medias en sus cursos, con 32, 25 y 17 estudiantes, de 79, 74 y 82 puntos, respectivamente. Hallar la puntuación media de los tres cursos.

3.58. El salario medio anual en una empresa es de \$15,000. Los de hombres y mujeres fueron, respectivamente, de \$15,600 y \$12,600 en media. Hallar el porcentaje de mujeres empleadas en esa empresa.

3.59. La Tabla 3.8 muestra la distribución de cargas máximas en toneladas cortas (1 tonelada corta = 2000 lb) que soportan los cables producidos en cierta fábrica. Determinar la carga máxima media, usando: (*a*) el «método largo» y (*b*) el método de compilación.

Tabla 3.8

Carga máxima (toneladas cortas)	Número de cables
9.3-9.7	2
9.8-10.2	5
10.3-10.7	12
10.8-11.2	17
11.3-11.7	14
11.8-12.2	6
12.3-12.7	3
12.8-13.2	1
	Total 60

X M
6 6
7
8
9
10

3.60. Hallar $\bar{X}$ para los datos de la Tabla 3.9, usando: (*a*) el «método largo» y (*b*) el método de compilación.

Tabla 3.9

X	f
462	98
480	75
498	56
516	42
534	30
552	21
570	15
588	11
606	6
624	2

3.61. La Tabla 3.10 muestra la distribución de los diámetros de los remaches salidos de una fábrica. Calcular el diámetro medio.

Tabla 3.10

Diámetro (cm)	Frecuencia
0.7247-0.7249	2
0.7250-0.7252	6
0.7253-0.7255	8
0.7256-0.7258	15
0.7259-0.7261	42
0.7262-0.7264	68
0.7265-0.7267	49
0.7268-0.7270	25
0.7271-0.7273	18
0.7274-0.7276	12
0.7277-0.7279	4
0.7280-0.7282	1
	Total 250

3.62. Calcular la media para los datos de la Tabla 3.11.

Tabla 3.11

Clase	Frecuencia
10 hasta 15	3
15 hasta 20	7
20 hasta 25	16
25 hasta 30	12
30 hasta 35	9
35 hasta 40	5
40 hasta 45	2
	Total 54

X_i
12,5
17,5
22,5
27,5
32,5
37,5
42,5

3.63. Calcular la vida media de los tubos del Problema 2.20.

3.64. (*a*) Usar la distribución de frecuencias obtenida en el Problema 2.27 para calcular el diámetro medio de las bolas de cojinetes.

(*b*) Calcular la media directamente de los datos y comparar con (*a*), explicando cualquier discrepancia.

LA MEDIANA

3.65. Hallar la media y la mediana de estos conjuntos de números: (*a*) 5, 4, 8, 3, 7, 2, 9 y (*b*) 18.3, 20.6, 19.3, 22.4, 20.2, 18.8, 19.7, 20.0.

3.66. Hallar la puntuación media del Problema 3.53.

3.67. Hallar el tiempo de reacción medio en el Problema 3.54.

3.68. Hallar la mediana del conjunto de números del Problema 3.55.

3.69. Hallar la mediana de las cargas máximas del Problema 3.59 (Tabla 3.8).

3.70. Hallar la mediana $\tilde{X}$ para la distribución del Problema 3.60 (Tabla 3.9).

3.71. Hallar el diámetro medio de los remaches de la Tabla 3.10, Problema 3.61.

3.72. Hallar la mediana de la distribución de la Tabla 3.11 del Problema 3.62.

3.73. La Tabla 3.12 muestra el número de bodas (incluidas posibles repeticiones) en EE.UU. para hombres y mujeres de distintos grupos de edad durante 1984.

(*a*) Hallar la mediana de edad de hombres y mujeres en esas bodas.

(*b*) ¿Por qué la mediana es una medida de tendencia central más adecuada que la media en este caso?

Tabla 3.12

Edad (años)	Varones (miles)	Hembras (miles)
18-19	121	481
20-24	2,441	4,184
25-29	5,930	6,952
30-34	6,587	7,193
35-44	11,788	11,893
45-54	9,049	9,022
55-64	8,749	8,171
65-74	5,786	4,654
75 y más	2,581	1,524

Fuente: U.S. Bureau of Census

3.74. Hallar la mediana de las ventas del Problema 2.31.

3.75. Hallar la mediana de las vidas medias de los tubos del Problema 2.20.

LA MODA

3.76. Hallar la media, la mediana y la moda de cada uno de estos conjuntos: (*a*) 7, 4, 10, 9, 15, 12, 7, 9, 7 y (*b*) 8, 11, 4, 3, 2, 5, 10, 6, 4, 1, 10, 8, 12, 6, 5, 7.

3.77. Hallar la puntuación modal del Problema 3.53.

3.78. Hallar el tiempo de reacción modal en el Problema 3.54.

3.79. Hallar la moda del conjunto de números del Problema 3.55.

3.80. Hallar la moda de las cargas máximas de los cables del Problema 3.59.

3.81. Hallar la moda $\hat{X}$ para la distribución de la Tabla 3.9 del Problema 3.60.

3.82. Hallar el diámetro modal de los remaches de la Tabla 3.10, del Problema 3.61.

3.83. Hallar la moda de la distribución del Problema 3.62.

3.84. Hallar la vida media modal de los tubos del Problema 2.20.

3.85. ¿Es posible determinar la moda para las distribuciones de los Problemas 3.73 y 2.31? Razonar la respuesta.

3.86. Usar la fórmula empírica media − moda = = 3(media − mediana) para calcular la moda de las distribuciones de los Problemas 3.59, 3.60, 3.61, 3.62 y 2.20. Comparar los resultados con los que da la fórmula (9) de este capítulo, explicando los acuerdos y las discrepancias.

3.87. Probar la afirmación del final del Problema 3.32.

LA MEDIA GEOMETRICA

3.88. Hallar la media geométrica de los números: (*a*) 4.2 y 16.8 y (*b*) 3.00 y 6.00.

3.89. Hallar (*a*) la media geométrica G y (*b*) la media aritmética $\bar{X}$ del conjunto 2, 4, 8, 16, 32.

3.90. Hallar la media geométrica de los conjuntos: (*a*) 3, 5, 8, 3, 7, 2 y (*b*) 28.5, 73.6, 47.2, 31.5, 64.8.

3.91. Hallar la media geométrica de las distribuciones en: (*a*) Problema 3.59 y (*b*) Problema 3.60. Verificar que la media geométrica es menor o igual que la media aritmética en estos casos.

3.92. Si el precio de un artículo se duplica en un período de 4 años, ¿cuál es el porcentaje medio de crecimiento anual?

3.93. En 1970 y 1980 la población de EE.UU. era de 203.3 y 226.5 millones, respectivamente.

(*a*) Hallar el porcentaje medio de crecimiento anual.
(*b*) Estimar la población en 1974.
(*c*) Si el porcentaje medio de crecimiento entre 1980 y 1990 es el de la parte (*a*), ¿cuál será la población en 1990?

3.94. ¿Qué capital final se tendrá al cabo de 6 años, si se invierten $1000 al 8% de interés anual?

3.95. Si en el problema anterior se compone el interés trimestralmente (o sea, el capital aumenta un 2% cada trimestre), ¿cuál sería el capital final?

3.96. Hallar dos números cuya media aritmética es 9.0 y cuya media geométrica es 7.2.

LA MEDIA ARMONICA

3.97. Hallar la media armónica de los números: (*a*) 2, 3 y 6 y (*b*) 3.2, 5.2, 4.8, 6.1 y 4.2.

3.98. Hallar (*a*) la media aritmética, (*b*) la media geométrica y (*c*) la media armónica de los números 0, 2, 4 y 6.

3.99. Si $X_1, X_2, X_3, \ldots$ representan las marcas de clase de una distribución de frecuencias con correspondientes frecuencias de clase $f_1, f_2, f_3, \ldots$, probar que la media armónica H de esa distribución viene dada por

$$\frac{1}{H} = \frac{1}{N}\left(\frac{f_1}{X_1} + \frac{f_2}{X_2} + \frac{f_3}{X_3} + \cdots\right) = \frac{1}{N}\sum \frac{f}{X}$$

donde $N = f_1 + f_2 + \cdots = \sum f$.

3.100. Usar el Problema 3.99 para hallar la media armónica de las distribuciones de: (*a*) Problema 3.59 y (*b*) Problema 3.60. Comparar con el Problema 3.91.

3.101. Las ciudades A, B y C están equidistantes entre sí. Un motorista viaja desde A hasta B a 30 mi/h, desde B hasta C a 40 mi/h, y desde C hasta A a 50 mi/h. Determinar su velocidad media en el viaje completo.

3.102. (*a*) Un avión vuela d_1, d_2 y d_3 millas a velocidades v_1, v_2 y v_3 mi/h, respectivamente. Probar que su velocidad media es V, dada por

$$\frac{d_1 + d_2 + d_3}{V} = \frac{d_1}{v_1} + \frac{d_2}{v_2} + \frac{d_3}{v_3}$$

Es una media armónica ponderada.
(*b*) Calcular V si $d_1 = 2500$, $d_2 = 1200$, $d_3 = 500$, $v_1 = 500$, $v_2 = 400$ y $v_3 = 250$.

3.103. Demostrar que la media geométrica de dos números positivos a y b es: (*a*) menor o igual que la media aritmética y (*b*) mayor o igual que la media armónica de esos números. ¿Puede extender la demostración a más de dos números?

LA MEDIA CUADRATICA

3.104. Hallar la media cuadrática de los números: (*a*) 11, 23 y 35 y (*b*) 2.7, 3.8, 3.2 y 4.3.

3.105. Probar que la media cuadrática de dos números positivos a y b es: (*a*) mayor o igual que la media aritmética y (*b*) mayor o igual que la media armónica. Extienda, si le es posible, la demostración a más de dos números.

3.106. Deducir una fórmula que sirva para hallar la media cuadrática de datos agrupados y aplíquese a alguna distribución de frecuencias ya considerada.

CUARTILES, DECILES Y PERCENTILES

3.107. La Tabla 3.13 muestra una distribución de frecuencias de puntuaciones de un examen final de álgebra. (*a*) Hallar los cuartiles de la distribución y (*b*) interpretar su significado.

Tabla 3.13

Grado	Número de estudiantes
90-100	9
80-89	32
70-79	43
60-69	21
50-59	11
40-49	3
30-39	1
	Total 120

3.108. Hallar los cuartiles Q_1, Q_2 y Q_3 para la distribución del: (*a*) Problema 3.59 y (*b*) Problema 3.60. Interpretar su significado.

3.109. Hallar: (*a*) el segundo decil, (*b*) el cuarto decil, (*c*) el 90.° percentil y (*d*) el 68.° percentil, para los datos del Problema 3.73, interpretando cada uno de ellos.

3.110. Hallar: (*a*) P_{10}, (*b*) P_{90}, (*c*) P_{25} y (*d*) P_{75} para los datos del Problema 3.59, interpretando cada uno de ellos.

3.111. (*a*) ¿Pueden todos los cuartiles ser expresados como percentiles? Explíquese.
(*b*) Idem con los quintiles.

3.112. Para los datos del Problema 3.107, determinar: (*a*) la puntuación más baja alcanzada por el 25% más alto del curso y (*b*) la más alta alcanzada por el 20% más bajo del curso. Interpretar la respuesta en términos de percentiles.

3.113. Interpretar los resultados del Problema 3.107 gráficamente usando: (*a*) un histograma de porcentajes, (*b*) un polígono de frecuencias en porcentajes y (*c*) una ojiva de porcentajes.

3.114. Resolver el Problema 3.113 con los datos del Problema 3.108.

3.115. (*a*) Desarrollar una fórmula, similar a la (8) de este capítulo, para calcular percentiles de una distribución de frecuencias.
(*b*) Ilustrar su uso obteniendo los resultados del Problema 3.110.

CAPITULO 4

La desviación típica y otras medidas de dispersión

DISPERSION O VARIACION

La *dispersión* o *variación* de los datos intenta dar una idea de cuán esparcidos se encuentran éstos. Hay varias medidas de tal dispersión, siendo las más comunes el rango, la desviación media, el rango semi-intercuartil, el rango percentil 10-90 y la desviación típica.

EL RANGO

El *rango* de un conjunto de números es la diferencia entre el mayor y el menor de todos ellos.

EJEMPLO 1. El rango del conjunto 2, 3, 3, 5, 5, 5, 8, 10, 12 es $12 - 2 = 10$. A veces el rango se indica dando el par de valores extremos; así, en este ejemplo, sería 2-12.

LA DESVIACION MEDIA

La *desviación media* o *desviación promedio*, de un conjunto de N números $X_1, X_2, \ldots, X_N$ es abreviada por MD y se define como

$$\text{Desviación media (MD)} = \frac{\sum_{j=1}^{N} |X_j - \bar{X}|}{N} = \frac{\sum |X - \bar{X}|}{N} = \overline{|X - \bar{X}|} \qquad (1)$$

donde $\bar{X}$ es la media aritmética de los números y $|X_j - \bar{X}|$ es el valor absoluto de la desviación de X_j respecto de $\bar{X}$. (El *valor absoluto* de un número es el número sin signo y se denota con dos barras verticales; así $|-4| = 4$, $|+3| = 3$, $|6| = 6$ y $|-0.84| = 0.84$.)

$\bar{X} = 6$

EJEMPLO 2. Hallar la desviación media del conjunto 2, 3, 6, 8, 11.

$$\mathrm{MD} = \frac{\sum_{j=1}^{N} |X - \bar{X}|}{N} = \overline{|X - \bar{X}|}$$

$$\text{Media aritmética } (\bar{X}) = \frac{2 + 3 + 6 + 8 + 11}{5} = 6$$

$$\mathrm{MD} = \frac{|2-6| + |3-6| + |6-6| + |8-6| + |11-6|}{5} = \frac{|-4| + |-3| + |0| + |2| + |5|}{5} = \frac{4+3+0+2+5}{5} = 2.8$$

Si $X_1, X_2, ..., X_K$ ocurren con frecuencias $f_1, f_2, ..., f_K$, respectivamente, la desviación media se puede escribir como

$$\text{MD} = \frac{\sum_{j=1}^{K} f_j|X_j - \bar{X}|}{N} = \frac{\sum f|X - \bar{X}|}{N} = \overline{|X - \bar{X}|} \tag{2}$$

donde $N = \sum_{j=1}^{K} f_j = \sum f$. Esta forma es útil para datos agrupados, donde los X_j representan las marcas de clase y los f_j son las correspondientes frecuencias de clase.

Ocasionalmente se define la desviación media en términos de desviaciones absolutas respecto de la mediana u otro promedio, en vez de la media. Una propiedad interesante de la suma $\sum_{j=1}^{N} |X_j - a|$ es que es mínima cuando a es la mediana (o sea, la desviación media respecto de la mediana es mínima).

Nótese que sería más apropiado usar la terminología *desviación media absoluta* que *desviación media*.

EL RANGO SEMI-INTERCUARTIL

El *rango semi-intercuartil*, o *desviación cuartil*, de un conjunto de datos se denota por Q y se define como

$$Q = \frac{Q_3 - Q_1}{2} \tag{3}$$

donde Q_1 y Q_3 son el primer y tercer cuartil de esos datos (véanse Probs. 4.6 y 4.7). El rango intercuartil $Q_3 - Q_1$ también se usa a veces, pero menos que el rango semi-intercuartil, como medida de dispersión.

EL RANGO PERCENTIL 10-90

El *rango percentil 10-90* de un conjunto de datos se define por

$$\text{rango percentil 10-90} = P_{90} - P_{10} \tag{4}$$

donde P_{10} y P_{90} son los décimo y nonagésimo percentiles de esos datos (véase Prob. 4.8). Puede usarse también el rango percentil semi 10-90 $\frac{1}{2}(P_{90} - P_{10})$, pero no es frecuente.

LA DESVIACION TIPICA O DESVIACION STANDAR.

La *desviación típica* de un conjunto de N números $X_1, X_2, ..., X_N$ se denota por s y se define como

$$s = \sqrt{\frac{\sum_{j=1}^{N} (X_j - \bar{X})^2}{N}} = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\overline{(X - \bar{X})^2}} \tag{5}$$

donde x representa las desviaciones de cada uno de los números X_j respecto de la media $\bar{X}$. Así que s es la raíz cuadrada de la media de las desviaciones cuadráticas, o como se le llama en ocasiones, la *desviación raíz-media-cuadrado.*

Si $X_1, X_2, \ldots, X_K$ ocurren con frecuencias $f_1, f_2, \ldots, f_K$, respectivamente, la desviación típica puede expresarse

$$s = \sqrt{\frac{\sum_{j=1}^{K} f_j(X_j - \bar{X})^2}{N}} = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{\sum fx^2}{N}} = \sqrt{\overline{(X - \bar{X})^2}} \tag{6}$$

donde $N = \sum_{j=1}^{K} f_j = \sum f$. En esta forma resulta útil para datos agrupados.

A veces se define la desviación típica de los datos de una muestra con $(N - 1)$ reemplazando a N en los denominadores de (5) y (6), porque el valor resultante da una mejor estimación de la desviación típica de la población total. Para grandes valores de N (ciertamente para $N > 30$), no hay prácticamente diferencia entre ambas definiciones. Además, cuando se necesita esa mejor estimación, siempre podemos obtenerla multiplicando la aquí definida por $\sqrt{N/(N-1)}$. Por tanto, nos quedaremos con la elección (5) y (6).

$\sqrt{(x-\bar{x})^2}$

$s^2 = (x-\bar{x})^2$

LA VARIANZA

La *varianza* de un conjunto de datos se define como el cuadrado de la desviación típica y viene dada en consecuencia por s^2 en las ecuaciones (5) y (6).

Cuando sea necesario distinguir la desviación típica de una población de la de una muestra de dicha población, usaremos el símbolo s para esta última y σ (*sigma* griega minúscula) para la primera. De modo que s^2 y σ^2 representarían la *varianza de la muestra* y la *varianza de la población*, respectivamente.

METODOS CORTOS PARA CALCULAR LA DESVIACION TIPICA

Las ecuaciones (5) y (6) se pueden escribir, respectivamente, en las formas equivalentes

$$s = \sqrt{\frac{\sum_{j=1}^{N} X_j^2}{N} - \left(\frac{\sum_{j=1}^{N} X_j}{N}\right)^2} = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\overline{X^2} - \bar{X}^2} \tag{7}$$

$$s = \sqrt{\frac{\sum_{j=1}^{K} f_j X_j^2}{N} - \left(\frac{\sum_{j=1}^{K} f_j X_j}{N}\right)^2} = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2} = \sqrt{\overline{X^2} - \bar{X}^2} \tag{8}$$

donde $\overline{X^2}$ denota la media de los cuadros de los diversos valores de X, mientras $\bar{X}^2$ denota el cuadrado de la media de los valores de X (véanse Probs. 4.12 a 4.14).

Si $d_j = X_j - A$ son las desviaciones de X_j respecto de alguna constante arbitraria A, los resultados (7) y (8) se convierten, respectivamente, en

$$s = \sqrt{\frac{\sum_{j=1}^{N} d_j^2}{N} - \left(\frac{\sum_{j=1}^{N} d_j}{N}\right)^2} = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2} = \sqrt{\overline{d^2} - \bar{d}^2} \tag{9}$$

$$s = \sqrt{\frac{\sum_{j=1}^{K} f_j d_j^2}{N} - \left(\frac{\sum_{j=1}^{K} f_j d_j}{N}\right)^2} = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\overline{d^2} - \bar{d}^2} \tag{10}$$

(Véanse Probs. 4.15 y 4.17.)

Cuando se tienen los datos agrupados en una distribución de frecuencias cuyos intervalos de clase tienen la misma anchura c, tenemos $d_j = cu_j$ o sea $X_j = A + cu_j$ y (10) pasa a ser

$$s = c\sqrt{\frac{\sum_{j=1}^{K} f_j u_j^2}{N} - \left(\frac{\sum_{j=1}^{K} f_j u_j}{N}\right)^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = c\sqrt{\overline{u^2} - \bar{u}^2} \tag{11}$$

Esta última fórmula proporciona un método muy breve para calcular la desviación típica y debe usarse para datos agrupados con igual anchura en sus intervalos de clase. Se llama *método de compilación* y es similar al utilizado en el Capítulo 3 para el cálculo de la media aritmética de datos agrupados. (Véanse Probs 4.16 a 4.19.)

PROPIEDADES DE LA DESVIACION TIPICA

1. La desviación típica puede definirse como

$$s = \sqrt{\frac{\sum_{j=1}^{N} (X_j - a)^2}{N}}$$

donde a es un promedio distinto de la media aritmética. De tales desviaciones típicas, la mínima es aquella para la cual $a = \bar{X}$, debido a la Propiedad 2 del Capítulo 3. Esta propiedad da una buena razón para adoptar la definición del comienzo. Su demostración se verá en el Problema 4.27.

2. Para distribuciones normales (véase Cap. 7), resulta (como sugiere la Fig. 4.1):

   (*a*) 68.27% de los casos están entre $\bar{X} - s$ y $\bar{X} + s$ (o sea, una desviación típica a cada lado de la media).

   (*b*) 95.45% de los casos están entre $\bar{X} - 2s$ y $\bar{X} + 2s$ (o sea, dos desviaciones típicas a cada lado de la media).

   (*d*) 99.73% de los casos entre $\bar{X} - 3s$ y $\bar{X} + 3s$ (o sea, tres desviaciones típicas a cada lado de la media).

   Para distribuciones poco asimétricas, los anteriores porcentajes son aproximadamente válidos (véase Prob. 4.24).

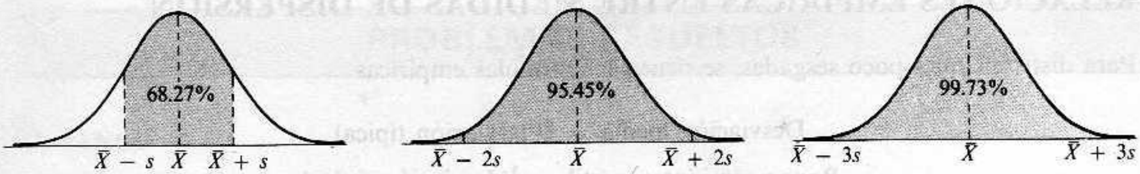


Figura 4.1.

3. Supongamos que dos conjuntos de N_1 y N_2 números (o dos distribuciones de frecuencias con frecuencias totales N_1 y N_2 tienen varianza dadas por s_1^2 y s_2^2, respectivamente, y tienen la *misma* media $\bar{X}$. Entonces la *varianza combinada* de ambos conjuntos (o de ambas distribuciones de frecuencias) vendrá dada por

$$s^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2} \tag{12}$$

Nótese que esto es una medida aritmética ponderada de las varianzas. El resultado admite generalización a más conjuntos.

COMPROBACION DE CHARLIER

La comprobación de Charlier en cálculos de la media y de la desviación típica por el método de compilación hace uso de las identidades

$$\sum f(u + 1) = \sum fu + \sum f = \Sigma\, fu + N$$

$$\sum f(u + 1)^2 = \sum f(u^2 + 2u + 1) = \sum fu^2 + 2 \sum fu + \sum f = \sum fu^2 + 2 \sum fu + N$$

(Véase Prob. 4.20.)

CORRECCION DE SHEPPARD PARA LA VARIANZA

El cálculo de la desviación típica es algo erróneo como resultado del agrupamiento de datos en clases (error de agrupamiento). Para corregirlo, se usa la fórmula

$$\text{Varianza corregida} = \text{varianza de los datos agrupados} - \frac{c^2}{12} \tag{13}$$

donde c es la anchura del intervalo de clase. La corrección $c^2/12$ (que se resta) se llama *corrección de Sheppard.* Se usa para distribuciones de variables continuas donde las «colas» van hacia cero en ambas direcciones.

Los estadísticos discrepan en cuanto a *si debe* aplicarse antes de examinar con corrección y *cuándo*. Ciertamente no debe aplicarse antes de examinar con cuidado la situación, pues a menudo tiende a *corregir en demasía*, con lo que sustituye un error por otro. En este libro, salvo indicación expresa, no la usaremos.

RELACIONES EMPIRICAS ENTRE MEDIDAS DE DISPERSION

Para distribuciones poco sesgadas, se tienen las fórmulas empíricas

$$\text{Desviación media} = \tfrac{4}{5}(\text{desviación típica})$$

$$\text{Rango semi-intercuartil} = \tfrac{2}{3}(\text{desviación típica})$$

Son consecuencia de que para la distribución normal la desviación media y el rango semi-intercuartil son iguales, respectivamente, a 0.7979 y 0.6745 veces la desviación típica.

DISPERSION ABSOLUTA Y RELATIVA: COEFICIENTE DE VARIACION

La variación o dispersión real, tal como se determina de la desviación típica u otra medida de dispersión, se llama la *dispersión absoluta*. Sin embargo, una dispersión (o variación) de 10 pulgadas (in) en la medida de 1000 pies es muy diferente de esa misma dispersión al medir una distancia de 20 pies. Una medida de este efecto la da la *dispersión relativa*, a saber

$$\text{Dispersión relativa} = \frac{\text{dispersión absoluta}}{\text{promedio}} \tag{14}$$

Si la dispersión absoluta es la desviación típica s y el promedio es la media $\bar{X}$, entonces la dispersión relativa se llama el *coeficiente de variación*, o *coeficiente de dispersión;* se denotará por V y se define como

$$\text{Coeficiente de variación } (V) = \frac{s}{\bar{X}} \tag{15}$$

y se expresa en general en forma de porcentaje. Hay otras posibilidades (véase Prob. 4.30).

Nótese que el coeficiente de variación es independiente de las unidades usadas. Por esa razón es útil al comparar distribuciones con unidades diferentes. Una desventaja del coeficiente de variación es que pierde su utilidad cuando $\bar{X}$ es próxima a cero.

VARIABLES TIPIFICADAS: UNIDADES ESTANDAR

La variable que mide la desviación de la medida en unidades de la desviación típica se llama una *variable tipificada*, es adimensional (independiente de las unidades usadas) y viene dada por

$$z = \frac{X - \bar{X}}{s} \tag{16}$$

Si las desviaciones de la media se dan en unidades de la desviación típica, se dicen expresadas en *unidades estándar*, o *recuentos estándar*. Son de gran valor al comparar distribuciones (véase Problema 4.31).

PROBLEMAS RESUELTOS

Rm -

EL RANGO

4.1. Hallar el rango de los conjuntos (*a*) 12, 6, 7, 3, 15, 10, 18, 5 y (*b*) 9, 3, 8, 8, 9, 8, 9, 18.

18 - 3

Solución

En ambos casos, rango = número mayor − número menor = 18 − 3 = 15. Sin embargo, como se ve de sus ordenaciones (*a*) y (*b*),

(*a*) 3, 5, 6, 7, 10, 12, 15, 18 (*b*) 3, 8, 8, 8, 9, 9, 9, 18

hay mucha más dispersión en (*a*) que en (*b*). De hecho, (*b*) consiste esencialmente de ochos y nueves.

Como el rango no indica diferencia entre esos conjuntos, no es buena medida de la dispersión en este caso. Cuando hay valores muy extremos, el rango es una pobre medida de la dispersión.

Se mejora eliminando los valores extremos, 3 y 18. Entonces para (*a*) el rango es (15 − 5) = 10, y para (*b*) es (9 − 8) = 1, que muestran claramente que el (*a*) tiene más dispersión que el (*b*). No obstante, no es así como se define el rango. El rango semi-intercuartil y el rango percentil 10-90 están pensados para mejorar el rango suprimiendo los valores extremos.

4.2. Hallar el rango de las alturas de los estudiantes de la Tabla 2.1.

Solución

Hay dos formas de definir el rango para datos agrupados.

Primer método

Rango = marca de clase de la clase más alta − marca clase más baja = 73 − 61 = 12 in

Segundo método

Rango = frontera superior de la clase más alta − frontera inferior de la clase más baja = = 74.5 − 59.5 = 15 in

El primer método tiende a eliminar los casos extremos en cierto grado.

LA DESVIACION MEDIA

4.3. Hallar la desviación media de los conjuntos de números del Problema 4.1.

Solución

(*a*) La media aritmética es

$$\bar{X} = \frac{12 + 6 + 7 + 3 + 15 + 10 + 18 + 5}{8} = \frac{76}{8} = 9.5$$

La desviación media es

$$MD = \frac{\sum |X-\bar{X}|}{N} =$$

$$= \frac{|12-9.5|+|6-9.5|+|7-9.5|+|3-9.5|+|15-9.5|+|10-9.5|+|18-9.5|+|5-9.5|}{8} =$$

$$= \frac{2.5+3.5+2.5+6.5+5.5+0.5+8.5+4.5}{8} = \frac{34}{8} = 4.25$$

(*b*) $$\bar{X} = \frac{9+3+8+8+9+8+9+18}{8} = \frac{72}{8} = 9$$

$$MD = \frac{\sum |X-\bar{X}|}{N} =$$

$$= \frac{|9-9|+|3-9|+|8-9|+|8-9|+|9-9|+|8-9|+|9-9|+|18-9|}{8} =$$

$$= \frac{0+6+1+1+0+1+0+9}{8} = 2.25$$

La desviación media indica que el conjunto (*b*) tiene menos dispersión que el (*a*), como debía ocurrir.

4.4. Hallar la desviación media de las alturas de los 100 estudiantes de la Universidad XYZ (Tabla 3.2 del Problema 3.20).

Solución

Del Problema 3.20, $\bar{X} = 67.45$ in. El trabajo se realiza como en la Tabla 4.1. Es posible diseñar un método de compilación para calcular la desviación media (véase Prob. 4.47).

Tabla 4.1

Altura (in)	Marca de clase (X)	$\|X-\bar{X}\| = \|X-67.45\|$	Frecuencia (f)	$f\|X-\bar{X}\|$
50-62	61	6.45	5	32.25
63-65	64	3.45	18	62.10
66-68	67	0.45	42	18.90
69-71	70	2.55	27	68.85
72-74	73	5.55	8	44.40
			$N = \sum f = 100$	$\sum f\|X-\bar{X}\| = 226.50$

$$MD = \frac{\sum f|X-\bar{X}|}{N} = \frac{226.50}{100} = 2.26 \text{ in}$$

4.5. Determinar el porcentaje de estudiantes del Problema 4.4 que miden entre (*a*) $\bar{X} \pm$ MD, (*b*) $\bar{X} \pm 2$ MD, (*c*) $\bar{X} \pm 3$ MD.

Solución

(*a*) El rango entre 65.19 y 69.71 in es $\bar{X} \pm \text{MD} = 67.45 \pm 2.26$. Este rango incluye a todos los individuos de la tercera clase; $+\frac{1}{3}(65.5 - 65.19)$, de los de la segunda; $+\frac{1}{3}(69.71 - 68.5)$, de los de la cuarta (como la anchura del intervalo de clase es 3 in, la frontera superior de la segunda clase es 65.5 in, y la inferior de la cuarta 68.5 in). El número de estudiantes en el rango $\bar{X} \pm 2$ MD es

$$42 + \frac{0.31}{3}(18) + \frac{1.21}{3}(27) = 42 + 1.86 + 10.89 = 54.75 \quad \text{o sea} \quad 55$$

que es el 55% del total.

(*b*) El rango desde 62.93 a 71.97 in es $\bar{X} \pm 2\text{ MD} = 67.45 \pm 2(2.26) = 67.45 \pm 4.52$. El número de estudiantes en el rango $\bar{X} \pm 2$ MD es

$$18 - \left(\frac{62.93 - 62.5}{3}\right)(18) + 42 + 27 + \left(\frac{71.97 - 71.5}{3}\right)(8) = 85.67 \quad \text{o sea} \quad 86$$

que es el 86% del total.

(*c*) El rango desde 60.67 a 74.23 in es $\bar{X} \pm 3\text{ MD} = 67.45 \pm 3(2.26) = 67.45 \pm 6.78$. El número de estudiantes en el rango $\bar{X} \pm 3$ MD es

$$5 - \left(\frac{60.67 - 59.5}{3}\right)(5) + 18 + 42 + 27 + \left(\frac{74.5 - 74.23}{3}\right)(8) = 97.33 \quad \text{o sea} \quad 97$$

que es el 97% del total.

EL RANGO SEMI-INTERCUARTIL

4.6. Hallar el rango semi-intercuartil para la distribución de alturas de la Universidad XYZ (Tabla 4.1 del Problema 4.4).

Solución

Los cuartiles inferior y superior son $Q_1 = 65.5 + \frac{2}{42}(3) = 65.64$ in y $Q_3 = 68.5 + \frac{10}{27}(3) = 69.61$ in, respectivamente, y el rango semi-intercuartil (o desviación cuartil) es $Q = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(69.61 - 65.64) = 1.98$ in. Nótese que el 50% de los casos cae entre Q_1 y Q_3 (o sea, 50 estudiantes miden entre 65.64 y 69.61 in).

Podemos considerar $\frac{1}{2}(Q_1 + Q_3) = 67.63$ in como una medida de tendencia central (o sea, un promedio de alturas). Se sigue que el 50% de las alturas caen en el rango 67.63 ± 1.98 in.

4.7. Hallar el rango semi-intercuartil para los salarios de los 65 empleados de la empresa P&R (Tabla 2.5 del Próblema 2.3).

Solución

Del Problema 3.44, $Q_1 = \$268.25$ y $Q_3 = \$290.75$. Así pues, el rango semi-intercuartil $Q = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(\$290.75 - \$268.25) = \$11.25$. Como $\frac{1}{2}(Q_1 + Q_3) = \$279.50$, podemos concluir que el 50% de los empleados cobra en el rango $\$279.50 \pm \$11.25$.

EL RANGO PERCENTIL 10-90

4.8. Hallar el rango percentil 10-90 de las alturas de la Tabla 2.1.

Solución

Aqui $P_{10} = 62.5 + \frac{5}{18}(3) = 63.33$ in, y $P_{90} = 68.5 + \frac{23}{27}(3) = 71.27$ in. Luego el rango percentil 10-90 es $P_{90} - P_{10} = 71.27 - 63.33 = 7.94$ in. Como $\frac{1}{2}(P_{10} + P_{90}) = 67.30$ in y $\frac{1}{2}(P_{90} - P_{10}) = 3.97$ in, podemos concluir que el 80% de los estudiantes tiene alturas en el rango 67.30 ± 3.97 in.

LA DESVIACION TIPICA

4.9. Hallar la desviación típica s de los conjuntos de números del Problema 4.1.

Solución

(*a*)
$$\bar{X} = \frac{\sum X}{N} = \frac{12 + 6 + 7 + 3 + 15 + 10 + 18 + 5}{8} = \frac{76}{8} = 9.5$$

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} =$$

$$= \sqrt{\frac{(12-9.5)^2 + (6-9.5)^2 + (7-9.5)^2 + (3-9.5)^2 + (15-9.5)^2 + (10-9.5)^2 + (18-9.5)^2 + (5-9.5)^2}{8}} =$$

$$= \sqrt{23.75} = 4.87$$

(*b*)
$$\bar{X} = \frac{9 + 3 + 8 + 8 + 9 + 8 + 9 + 18}{8} = \frac{72}{8} = 9$$

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} =$$

$$= \sqrt{\frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8}} =$$

$$= \sqrt{15} = 3.87$$

Los resultados anteriores deben compararse con los del Problema 4.3. Se apreciará que la desviación típica indica que (*b*) es menos disperso que (*a*). Sin embargo, el efecto está enmascarado por el hecho de que los valores extremos afectan a la desviación típica mucho más que a la desviación media. Era de esperar, desde luego, porque las desviaciones se elevan al cuadrado al calcular la desviación típica.

4.10. Hallar la varianza de los conjuntos de números del Problema 4.1.

Solución

Varianza $= s^2$. Luego del Problema 4.9 deducimos (*a*) $s^2 = 23.75$ y (*b*) $s^2 = 15$.

4.11. Hallar la desviación típica de las alturas de estudiantes de la Tabla 2.1.

Solución

De los Problemas 3.15, 3.20 ó 3.22, $\bar{X} = 67.45$ in. El método de trabajo se recoge en la Tabla 4.2.

Tabla 4.2

Altura (in)	Marca de clase (X)	$X - \bar{X} = X - 67.45$	$(X - \bar{X})^2$	Frecuencia (f)	$f(X - \bar{X})^2$
60-62	61	-6.45	41.6025	5	208.0125
63-65	64	-3.45	11.9025	18	214.2450
66-68	67	-0.45	0.2025	42	8.5050
69-71	70	2.55	6.5025	27	175.5675
72-74	73	5.55	30.8025	8	246.4200
				$N = \sum f = 100$	$\sum f(X - \bar{X})^2 = 852.7500$

$$s = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{852.7500}{100}} = \sqrt{8.5275} = 2.92 \text{ in}$$

CALCULO DE LA DESVIACION TIPICA PARA DATOS AGRUPADOS

4.12. (*a*) Probar que

$$s = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\overline{X^2} - \bar{X}^2}$$

(*b*) Usar la fórmula en (*a*) para hallar la desviación típica del conjunto de números 12, 6, 7, 3, 15, 10, 18, 5.

Solución

(*a*) Por definición:

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Entonces
$$s^2 = \frac{\sum (X - \bar{X})^2}{N} = \frac{\sum (X^2 - 2\bar{X}X + \bar{X}^2)}{N} = \frac{\sum X^2 - 2\bar{X}\sum X + N\bar{X}^2}{N} =$$

$$= \frac{\sum X^2}{N} - 2\bar{X}\frac{\sum X}{N} + \bar{X}^2 = \frac{\sum X^2}{N} - 2\bar{X}^2 + \bar{X}^2 = \frac{\sum X^2}{N} - \bar{X}^2 =$$

$$= \overline{X^2} = \bar{X}^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2$$

o sea
$$s = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\overline{X^2} - \bar{X}^2}$$

Obsérvese que en las sumas precedentes hemos usado la forma abreviada, con X sustituyendo a X_j y $\sum$ a $\sum_{j=1}^{N}$.

Otro método

$$s^2 = \overline{(X - \bar{X})^2} = \overline{X^2 - 2X\bar{X} + \bar{X}^2} = \overline{X^2} - \overline{2X\bar{X}} + \overline{\bar{X}^2} = \overline{X^2} - 2\bar{X}\bar{X} + \bar{X}^2 = \overline{X^2} - \bar{X}^2$$

(*b*)

$$\overline{X^2} = \frac{\sum X^2}{N} = \frac{(12)^2 + (6)^2 + (7)^2 + (3)^2 + (15)^2 + (10)^2 + (18)^2 + (5)^2}{8} = \frac{912}{8} = 114$$

$$\bar{X} = \frac{\sum X}{N} = \frac{12 + 6 + 7 + 3 + 15 + 10 + 18 + 5}{8} = \frac{76}{8} = 9.5$$

Así pues, $s = \sqrt{\overline{X^2} - \bar{X}^2} = \sqrt{114 - 90.25} = \sqrt{23.75} = 4.87$

Compárese este método con el del Problema 4.9(*a*).

4.13. Modificar la fórmula del Problema 4.12(*a*) para permitir frecuencias asignadas a los diferentes valores de X.

Solución

La modificación adecuada es

$$s = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2} = \sqrt{\overline{X^2} - \bar{X}^2}$$

Como en el Problema 4.12(*a*), ésta puede probarse partiendo de

$$s = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}}$$

Entonces $s^2 = \dfrac{\sum f(X - \bar{X})^2}{N} = \dfrac{\sum f(X^2 - 2\bar{X}X + \bar{X}^2)}{N} = \dfrac{\sum fX^2 - 2\bar{X}\sum fX + \bar{X}^2\sum f}{N} =$

$$= \frac{\sum fX^2}{N} - 2\bar{X}\frac{\sum fX}{N} + \bar{X}^2 = \frac{\sum fX^2}{N} - 2\bar{X}^2 + \bar{X}^2 = \frac{\sum fX^2}{N} - \bar{X}^2 =$$

$$= \frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2$$

es decir $s = \sqrt{\dfrac{\sum fX^2}{N} - \left(\dfrac{\sum fX}{N}\right)^2}$

Nótese que en las anteriores sumas se ha empleado la forma abreviada, con X y sustituyendo a X_j y f_j, $\sum$ a $\sum_{j=1}^{K}$ y $\sum_{j=1}^{K} f_j = N$.

4.14. Mediante la fórmula del Problema 4.13, hallar la desviación típica de los datos de la Tabla 4.2 del Problema 4.11.

Solución

Hágase como sugiere la Tabla 4.3, donde $\bar{X} = (\sum fX)/N = 67.45$ in, como se sigue del Problema 3.15. Nótese que este método, al igual que el del Problema 4.11, exige cálculos tediosos. El Problema 4.17 enseña que el método de compilación los simplifica en gran medida.

Tabla 4.3

Altura (in)	Marca de clase (X)	X^2	Frecuencia (f)	fX^2
60-62	61	3271	5	18,605
63-65	64	4096	18	73,728
66-68	67	4489	42	188,538
69-71	70	4900	27	132,300
72-74	73	5329	8	42,632
			$N = \sum f = 100$	$\sum fX^2 = 455{,}803$

$$s = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2} = \sqrt{\frac{455{,}803}{100} - (67.45)^2} = \sqrt{8.5275} = 2.92 \text{ in}$$

4.15. Si $d = X - A$ son las desviaciones de X respecto de una constante arbitraria A, probar que

$$s = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

Solución

Como $d = X - A$, $X = A + d$ y $\bar{X} = A + \bar{d}$ (véase Prob. 3.18), entonces

$$X - \bar{X} = (A + d) - (A + \bar{d}) = d - \bar{d}$$

así que
$$s = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{\sum f(d - \bar{d})^2}{N}} = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

usando el Problema 4.13 y sustituyendo X y $\bar{X}$ por d y $\bar{d}$, respectivamente.

Otro método

$$s^2 = \overline{(X - \bar{X})^2} = \overline{(d - \bar{d})^2} = \overline{d^2 - 2\bar{d}d + \bar{d}^2} =$$
$$= \overline{d^2} - 2\bar{d}^2 + \bar{d}^2 = \overline{d^2} - \bar{d}^2 = \frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2$$

y el resultado se sigue tomando la raíz cuadrada positiva.

4.16. Probar que si cada marca de clase X en una distribución de frecuencias con intervalos de clase de igual anchura c se compila en un valor asociado u según la relación $X = A + cu$, donde A es una marca de clase dada, entonces la desviación típica se escribe

$$s = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = c\sqrt{\overline{u^2} - \bar{u}^2}$$

Solución

Se deduce del Problema 4.15, ya que $d = X - A = cu$. Luego, al ser c constante,

$$s = \sqrt{\frac{\sum f(cu)^2}{N} - \left(\frac{\sum f(cu)}{N}\right)^2} = \sqrt{c^2\frac{\sum fu^2}{N} - c^2\left(\frac{\sum fu}{N}\right)^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2}$$

Otro método

También se puede demostrar directamente sin apelar al Problema 4.15. Como $X = A + cu$, $\bar{X} = A + c\bar{u}$, y $X - \bar{X} = c(u - \bar{u})$, entonces

$$s^2 = \overline{(X - \bar{X})^2} = \overline{c^2(u - \bar{u})^2} = c^2\overline{(u^2 - 2\bar{u}u + \bar{u}^2)} = c^2(\overline{u^2} - 2\bar{u}^2 + \bar{u}^2) = c^2(\overline{u^2} - \bar{u}^2)$$

y
$$s = c\sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2}$$

4.17. Hallar la desviación típica de las alturas de estudiantes de la Universidad XYZ (Tabla 2.1) mediante (*a*) la fórmula del Problema 4.15 y (*b*) el método del Problema 4.16.

Solución

En las Tablas 4.4 y 4.5, A se ha tomado arbitrariamente como la marca de clase 67. Nótese que en la Tabla 4.4 las desviaciones son todas múltiplos de la anchura del intervalo de clase $c = 3$. Ese factor se ha suprimido en la Tabla 4.5. En consecuencia, se simplifican muchos los cálculos de la Tabla 4.5 (a comparar con los de los Problemas 4.11 y 4.14). Por tal razón, el método de compilación es muy recomendable.

(*a*) Véase Tabla 4.4.

Tabla 4.4

Marca de clase (X)	$d = X - A$	Frecuencia (f)	fd	fd^2
61	−6	5	−30	180
64	−3	18	−54	162
$A \rightarrow 67$	0	42	0	0
70	3	27	81	243
73	6	8	48	288
		$N = \sum f = 100$	$\sum fd = 45$	$\sum fd^2 = 873$

$$s = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{873}{100} - \left(\frac{45}{100}\right)^2} = \sqrt{8.5275} = 2.92 \text{ in}$$

(*b*) Véase Tabla 4.5.

Tabla 4.5

Marca de clase (X)	$u = \dfrac{X - A}{c}$	Frecuencia (f)	fu	fu^2
61	−2	5	−10	20
64	−1	18	−18	18
$A \rightarrow 67$	0	42	0	0
70	1	27	27	27
73	2	8	16	32
		$N = \sum f = 100$	$\sum fu = 15$	$\sum fu^2 = 97$