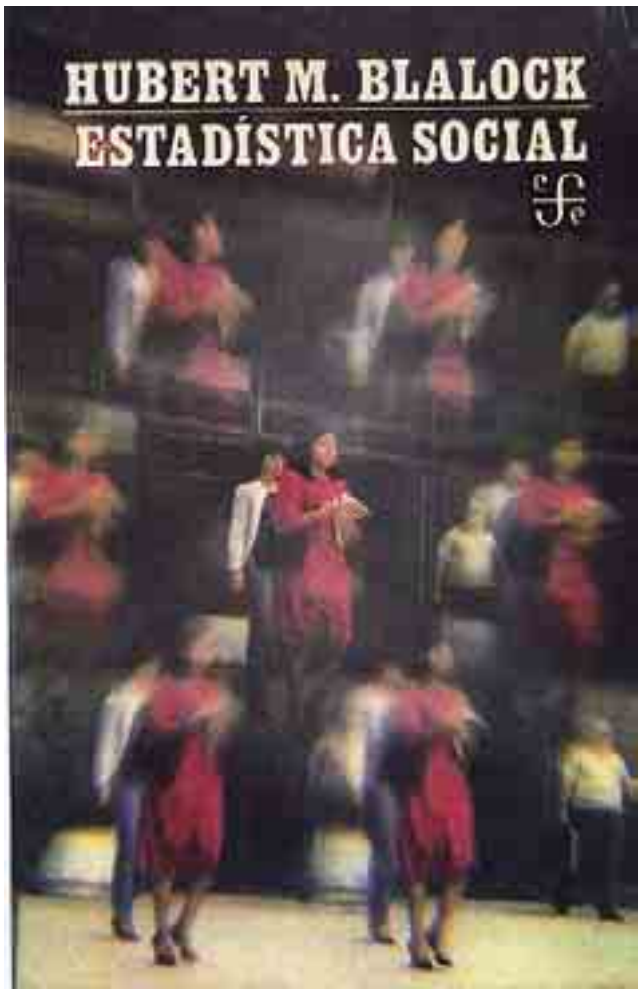
ESTADÍSTICA SOCIAL

HUBERT M. BLALOCK



FONDO DE CULTURA ECONÓMICA

México, 1986

Este material se utiliza con fines
exclusivamente didácticos

ÍNDICE

Prefacio 9

Primera Parte
INTRODUCCIÓN

1. Introducción: objetivos y límites de la estadística 15
1.1. Funciones de la estadística 16
1.2. El lugar de la estadística en el proceso de la investigación 19
1.3. Advertencia 20
Bibliografía 21

II. Teoría, medición y matemáticas 22
II. l. Teoría e hipótesis: definiciones operativas 22
II.2. El nivel de medición: escalas nominales, ordinales y de intervalo 26
II.3. Medición y estadística 32
II.4. Organización del libro 37
Bibliografía 40

Segunda Parte
ESTADÍSTICA DESCRIPTIVA UNIVARIADA

III. Escalas nominales: proporciones, porcentajes y razones 43
III.1. Proporciones 43
III.2. Porcentajes 45
111.3. Razones 49

Bibliografía 52

IV. Escalas de intervalo: distribuciones de frecuencia y representación gráfica 53
IV.1. Distribuciones de frecuencia: agrupamiento de los datos 53
IV.2. Distribuciones de frecuencia cumulativa 60
IV.3. Presentación gráfica: histogramas, polígonos de frecuencia y ojivas 61
Bibliografía 66

V. Escalas de intervalo: medidas de tendencia central 67
V.1. La media aritmética 67
V.2. La mediana 71
V.3. Cálculo de la media y la mediana de datos agrupados 73
V4. Comparación de la media y la mediana 81
V.5. Otras medidas de tendencia central 85
V.6. Deciles, cuartiles y percentiles 86
Bibliografía 88

VI. Escalas de intervalo: medidas de dispersión 90
VI. l. El recorrido 90
VI.2. La desviación cuartil 92
VI.3. La desviación media 92
VI.4. La desviación estándar 93
VI.5. El coeficiente de variabilidad 101
VI.6. Otras medidas resumidas 102
Bibliografía 103

VII. La distribución normal 104
VII. l. Distribuciones de frecuencias finitas versus infinitas 104
VII.2. Forma general de la curva normal 107
VII.3. Áreas bajo la curva normal 109

VII.4. Ilustraciones suplementarias del empleo de la tabla normal 113
Bibliografía 116

Tercera Parte
ESTADÍSTICA INDUCTIVA

VIII. Introducción a la estadística inductiva 119
VIII. l. Estadística y parámetros 119
VIII.2. Pasos en la verificación de una hipótesis 120
VIII.3. La falacia de afirmar el consecuente 123
VIII.4. La forma de las hipótesis estadísticas 124
Bibliografía 127

IX. Probabilidad 128
IX.1. Probabilidad a priori 129
IX.2. Propiedades matemáticas de las probabilidades 133
IX.3. Permutas 145
IX4. Valores esperados 151
IX.5. Independencia y muestreo aleatorio 153
Bibliografía 159

X. Pruebas de hipótesis: la distribución binomial 160
X1. La distribución de muestreo binomial 160
X.2. Pasos en las pruebas estadísticas 164
X.3. Aplicaciones de la binomial 177
X.4. Extensiones del binomio 181
X.5. Sumario 183
Bibliografía 186

XI. Pruebas de muestras simples que implican medias y proporciones 187
XI. l. Distribución en muestreo de las medias 187
XI.2. Prueba para la media de la población, conociendo. 194
XI.3. La distribución t de Student 199
XI.4. Pruebas que comportan proporciones 204
Bibliografía 210

XII. Estimación de punto e intervalo 211
XII. l. Estimación del punto 212
XII.2. Estimación del intervalo 215
XII.3. Intervalos de confianza para otros tipos de problemas 221
XII.4. Determinación del tamaño de la muestra 224
Bibliografía 227

Cuarta Parte
ESTADISTICAS BIVARIADAS Y MULTIVARIADAS

XIII. Pruebas de dos muestras: diferencia de las medias y las proporciones 231
XIII. l. Prueba de la diferencia de las medias 231
XIII.2. Diferencia de proporciones 240
XIII.3. Intervalos de confianza 245
XIII.4. Muestras dependientes: pares asociados 246
XIII.5. Comentarios a propósito de los esquemas experimentales y pruebas de significación 248
Bibliografía 255

XIV. Escalas ordinales: pruebas no paramétricas de dos muestras. 256
XIV1. Fuerza y eficiencia de la fuerza 257
XIV.2. La prueba de las secuencias (runs) de Wald-Wolfowitz 263

XIV.3. La prueba de Mann-Whitney o de Wilcoxon 269
XIVA. La prueba de Kolmogorov-Smirnov 277
XIV.5. La prueba de Wilcoxon de pares asociados y órdenes provistos de signo 280
XIV.6. Resumen 284
Bibliografía 288

XV. Escalas nominales: problemas de contingencia 289
XV.1. La prueba de la X-cuadrada 289
XV.2. La prueba exacta de Fisher 301
XV.3. Medidas de la fuerza de la relación 306
XV.4. Control de otras variables 319
Bibliografía 330

XVI. Análisis de la variancia 332
XVI. l. Análisis simple de la variancia 332
XVI.2. Comparación de medias específicas 343
XVI.3. Análisis bimodal de la variancia 349
XVIA. Alternativas no paramétricas del análisis de variancia 365
XVI.5. Medidas de asociación: correlación intraclase 370
Bibliografía 376

XVII. Correlación y regresión 377
XVII. l. Regresión lineal y mínimos cuadrados 377
XVII.2. Correlación 393
Bibliografía 413

XVIII. Correlación y regresión [conclusión] 414
XVIII .l. Prueba de significación e intervalos de confianza 414
XVIII.2. Correlación no lineal y regresión 425
XVIII.3. Efectos de los errores de medición 431
XVIII.4. Escalas ordinales: correlación de rangos 433
Bibliografía 445

XIX. Correlación múltiple y parcial 447
XIX.1. Regresión múltiple y mínimos cuadrados- 447
XIX.2. Correlación parcial 451
XIX.3. Correlación parcial e interpretaciones causales 461
XIX.4. Mínimos cuadrados múltiples y los coeficientes beta 469
XIX.5. ¡Correlación múltiple 473
XIX.6. Regresión múltiple y no linealidad 479
XIX.7. Pruebas de significación e intervalos de confianza 484
Bibliografía 489

XX. Análisis de covariancia y variables simuladas 491
XX.1. Relación de dos escalas de intervalo, control de la escala nominal 492
XX.2. Relación de una escala de intervalo y una escala nominal, control de la escala de intervalo ... 510
XX.3. Extensiones del análisis de covariancia 516
XX.4. Análisis de la variable simulada 517
XX.5. Observaciones finales 521
Bibliografía 526

Quinta Parte
MUESTREO

XXI. Muestreo 531
XXI.1. Muestreo aleatorio sencillo 532
XXI.2. Muestreo sistemático 537

XXI.3. Muestreo estratificado 539
XXIA. Muestreo por conglomerados 546
XXI.S. Muestreo sin probabilidad 552
XXI.6. Errores no de muestreo y tamaño de la muestra 553
Bibliografía 554

APÉNDICES

I. Resumen de operaciones algebraicas 559
Cuadros 565
Índice de figuras 599
Índice de cuadros 603

VII. LA DISTRIBUCIÓN NORMAL

La noción de la distribución de frecuencias es ya familiar. El presente capítulo se ocupa de un tipo muy importante de distribución de frecuencia: la curva normal. Esta distribución es muy útil, no sólo porque un gran número de distribuciones empíricas se encuentran ser aproximadamente normales, sino debido también a su significado teórico en la estadística inductiva. En este momento, el lector no debe preocuparse por las aplicaciones en las que se emplea la curva normal. En efecto, el objeto del presente capítulo está en indicar las propiedades de la curva en cuestión y en familiarizar al lector con el empleo de cuadros basados en la misma. Esta distribución se examina en la estadística descriptiva más que en la inductiva por dos razones principales. Primero, la curva normal puede emplearse para proporcionar una interpretación de la desviación estándar. Y en segundo lugar, serán útiles al lector para familiarizarse con la distribución normal algunos capítulos antes de exponerse a pruebas estadísticas que requieren facilidad en la manipulación de la misma. Por lo tanto, cuanto mejor se comprenda la materia expuesta en este capítulo, tanto menos dificultad se experimentará más adelante.

VII. 1. Distribuciones de frecuencia finitas versus infinitas

Las distribuciones de frecuencia hasta aquí examinadas comportaban un número finito de casos. De hecho, por supuesto, todas las distribuciones empíricas comportan necesariamente un número finito de casos, aunque tal vez muy grande. Sin embargo los matemáticos consideran ventajoso a menudo pensar en términos de distribuciones basadas en un número de casos infinitamente grande. Más bien que tratar con distribuciones empíricas de aspecto anguloso, como las que ejemplifican el histograma o el polígono de frecuencia, resulta posible concebir curvas lisas basadas en un número indefinidamente grande de casos y susceptibles de ser expresadas en términos de ecuaciones matemáticas relativamente sencillas. La distribución normal es una de tale curvas. Antes de examinar esta distribución específica, convendrá estudiar la naturaleza del proceso a través del cual se desarrolla una curva lisa semejante.

Empecemos con un histograma que comprende cinco intervalos (figura VII. 1a.). Con fines de simplicidad supondremos que la distribución de frecuencia es simétrica. Ya vimos que si el número de intervalos aumentaba sin cambiar N, la forma del histograma tiende a hacerse irregular. Supóngase, sin embargo que el número de casos se ha aumentado asimismo. En tal caso, como en la figura VII.1b, será posible servirse de un mayor número de intervalos más angostos, cada uno de los cuales tenga un número suficiente de casos para mantener la regularidad. Si el número de casos sigue aumentando, pueden emplearse todavía más rectángulos, conservando, con todo, el tipo regular (fig.VII.1c). Las curvas lisas se han trazado por los puntos medios del lado superior de cada rectángulo. Resulta claro que los rectángulos van formando aproximaciones cada vez mejores a la curva lisa a medida que el número de los mismos aumenta, esto es, a medida que disminuye el ancho de cada intervalo. Imaginemos ahora un número de casos en aumento incesante, con intervalos cada vez más angostos, hasta que los rectángulos se aproximen tan íntimamente a la curva lisa que ya no podamos apreciar diferencia alguna entre aquéllos y ésta. Designamos la curva lisa a la que se van acercando incesantemente los rectángulo cada vez más angosto como *límite* de la distribución de frecuencia.[1] Pese a que no podamos imaginarnos un número infinito de casos, podemos, sin embargo, concebir un número tan grande de ellos, que los rectángulos se acerquen a la curva lisa con el grado de exactitud deseado.

Se recordará que área de cada rectángulo puede utilizarse para representar la proporción de casos comprendidos en el intervalo. Como ya se indicó en el capítulo IV, el área total de todos los rectángulos se suele hacer igual a la unidad. Así, pues, si la proporción de los casos del primer intervalo es .10, entonces dicho mismo número representa el área real del primer rectángulo.

[1] La noción de límite se examina también en la sección IX. 1.

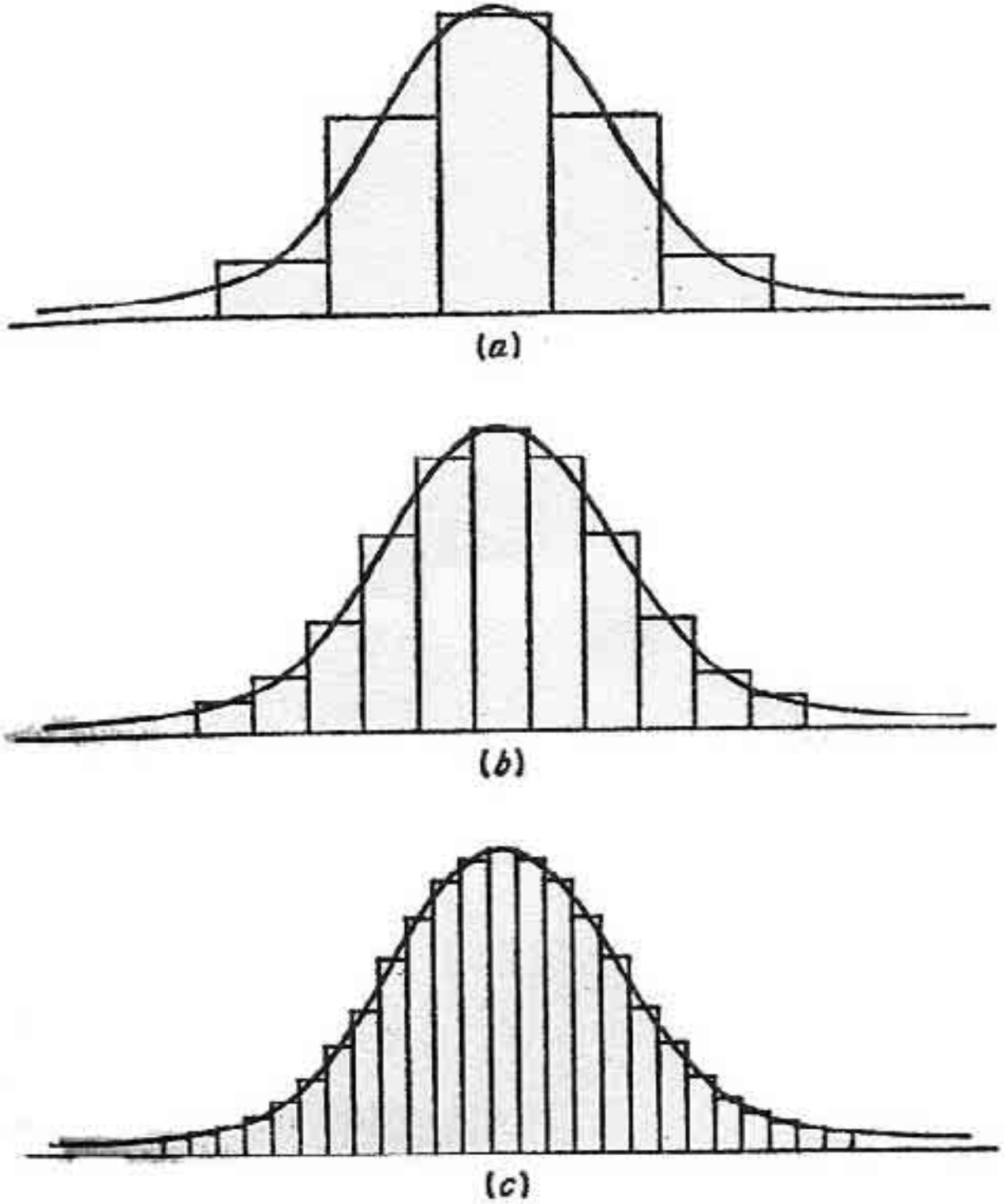


Fig. VII.1. ***Comparaciones de curvas lisas con histogramas de amplitudes diferentes de intervalo.***

Observamos ahora que el área del rectángulo correspondiente puede aproximarse al área que queda debajo de la curva lisa al interior de cualquier intervalo dado. Es lo que indica la figura VII. 2. A medida que el número de rectángulos aumenta, el área total de los rectángulos se convierte en una aproximación cada vez mejor al área que queda bajo la curva lisa. Esto puede verse observando que las áreas achuradas se van haciendo cada vez más pequeñas. En el límite, pues el área debajo de la curva lisa puede obtenerse sumando las áreas de un número indefinidamente grande de rectángulos. Y como quiera que el área debajo de los rectángulos es la unidad, el área debajo de la curva lisa será asimismo igual a la unidad. El proceso que acabamos de describir es exactamente la clase de proceso que se halla en la rama de las matemáticas designada como cálculo.

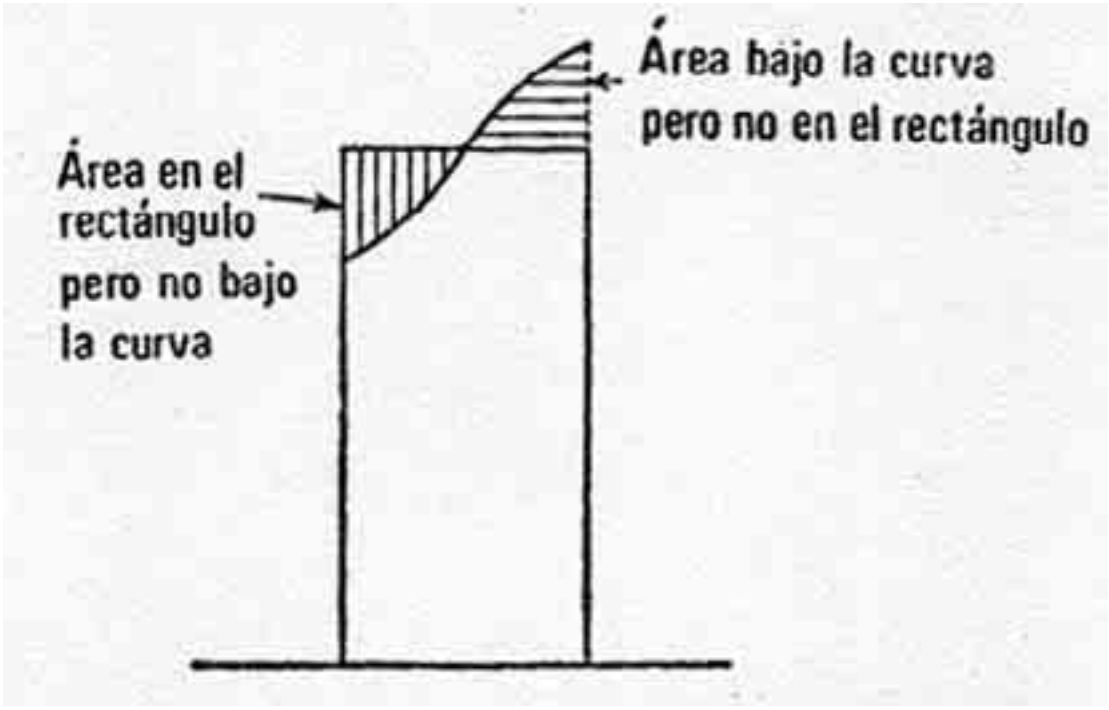


FIG. VII.2. ***Comparación de las áreas debajo de la curva y debajo del rectángulo***

VII.2. Forma general de la curva normal

La curva normal es un tipo especial de curva lisa simétrica. Como quiera que la curva normal es lisa, perfectamente simétrica y se basa en un número indefinidamente grande de casos, sólo es posible aproximarse a la misma mediante distribuciones de frecuencia que comportan datos efectivos. Tiene forma de campana y posee cierto número de propiedades matemáticas notables, algunas de las cuales se señalarán brevemente. Como quiera que es simétrica y unimodal, su media, mediana y modo coinciden. La forma general de la distribución normal se indica en la figura VII. 3.

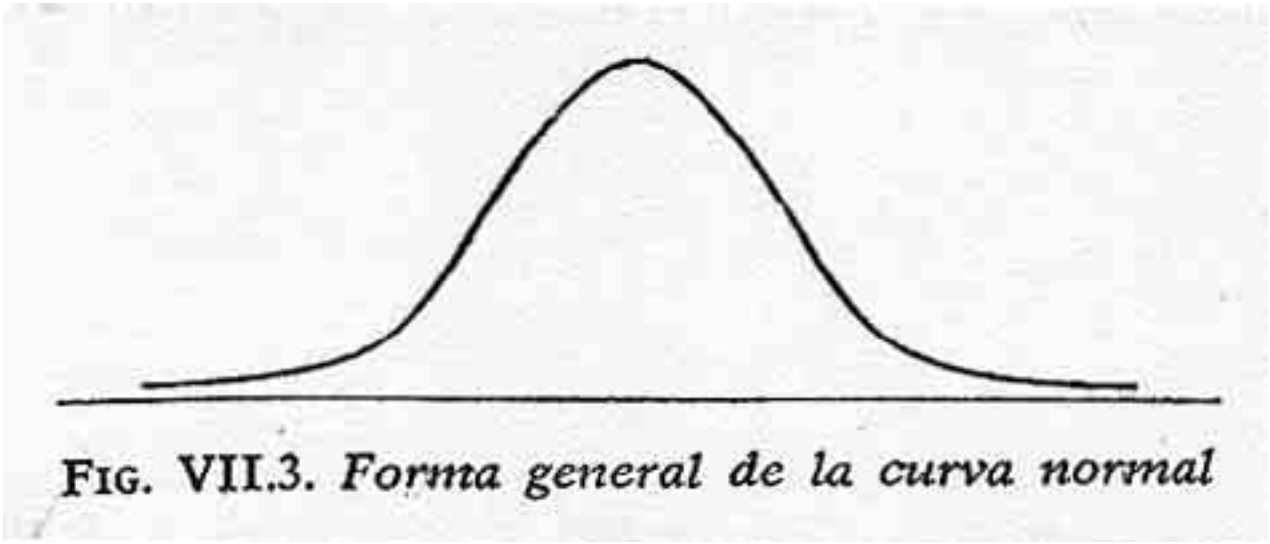


FIG. VII.3. ***Forma general de la curva normal***

*La ecuación matemática de la curva normal es relativamente sencilla en las normas de los matemáticos. Aunque el lector no habrá de emplear nunca dicha fórmula, ya que se han confeccionado cuadros con tal objeto, será útil, sin embargo, que la vea, para señalar y verificar algunas de las propiedades de esta distribución teórica. La fórmula es como sigue:

$$Y = \frac{1}{s\sqrt{2\pi}} e - \left(X - \overline{X}\right)^{2/2s^2}$$

en donde Y es la altura de la curva para un valor determinado de X. Como quiera que tanto π como e son constantes (iguales respectivamente a 3.14 y 2.72), la fórmula sólo comporta dos medidas de resumen, la media $\overline{X}$ y la desviación estándar s.[2] Por lo tanto, la forma exacta de la curva normal será conocida si se nos dan los valores de dichas medidas. En otros términos: hay muchas curvas normales, una para cada combinación de la media y de la desviación estándar.

*Recordando que una cantidad afectada de un exponente negativo puede escribirse como la recíproca de dicha cantidad elevada a la potencia positiva, podemos escribir la fórmula como sigue:

$$Y = \frac{1}{s\sqrt{2\pi}}\left(\frac{1}{2.72^{(\bar{x}-x)^2 2_8 2}}\right)$$

en la que e ha sido sustituida por su valor numérico. Supongamos que el valor de s es fijo, y busquemos el valor de X para el cual Y será un máximo. Es obvio que Y será máxima cuando el denominador incluido en los paréntesis sea mínimo. Pero dicho denominador consta de un número positivo mayor que la unidad elevado a una potencia que no puede ser negativa, ya que un número real cuadrado no puede ser nunca menor que cero. Por consiguiente, el denominador alcanzará su mínimo cuando el exponente sea cero. Y esto ocurrirá cuando X adopte el valor de $\overline{X}$, ya que tendremos X-$\overline{X}$=0. Esto muestra que el modo (y, por consiguiente, la media y la mediana) es realmente $\overline{X}$, hecho que ya se había señalado, pero sin demostrarlo. Podemos ver, asimismo, que la ecuación da una curva que es simétrica alrededor de X. Como quiera que la cantidad $X-\overline{X}$ está al cuadrado y no puede, por consiguiente, ser negativa, las desviaciones respecto de $\overline{X}$ en una u otra dirección producirán valores idénticos de Y.

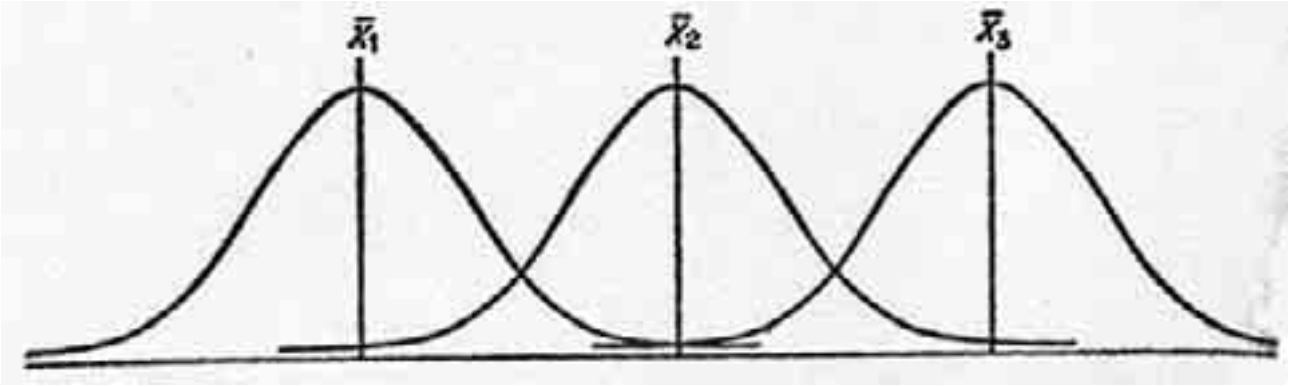


FIG. VII.4. ***Comparación de curvas normales de igual desviación estándar pero de medias diferentes***.

La ecuación específica para toda curva normal particular puede obtenerse empleando los valores propios de $\overline{X}$ y s. En la figura VII. 4 pueden verse curvas normales de la misma desviación estándar, pero de medias distintas. Por otra parte, las curvas de desviaciones estándar distintas variarán en la configuración de las cúspides, tal como se indica en la figura VII.5. Cuanto menor sea la desviación normal, tanto más puntiaguda resultará la curva. Habría que señalar que no todas las curvas simétricas en forma de campana son normales. Aunque las curvas de la figura VII. 5 difieran en cuanto a las cúspides, esto se debe únicamente a diferencias en sus desviaciones normales. Todas ellas son normales en cuanto a la forma. Por regla general, las curvas simétricas unimodales pueden ser más o menos puntiagudas o aplanadas que la curva normal, aun siendo sus desviaciones estándar las mismas. Algunas de estas curvas pueden verse en la figura XII.6. Las que son más puntiagudas que la normal se designan como *leptocúrticas* y las más planas que aquélla como *platicúrticas*. A diferencia de la curva normal, las ecuaciones de las curvas leptocúrticas y platicúrticas tienden a comportar medidas de resumen, además de la media y la desviación estándar.

[2] Cuando lleguemos a la estadística inductiva se introducirá otra notación para la media y la desviación estándar. La fórmula de la curva normal suele escribirse en términos de una media de μ y una desviación estándar de σ.

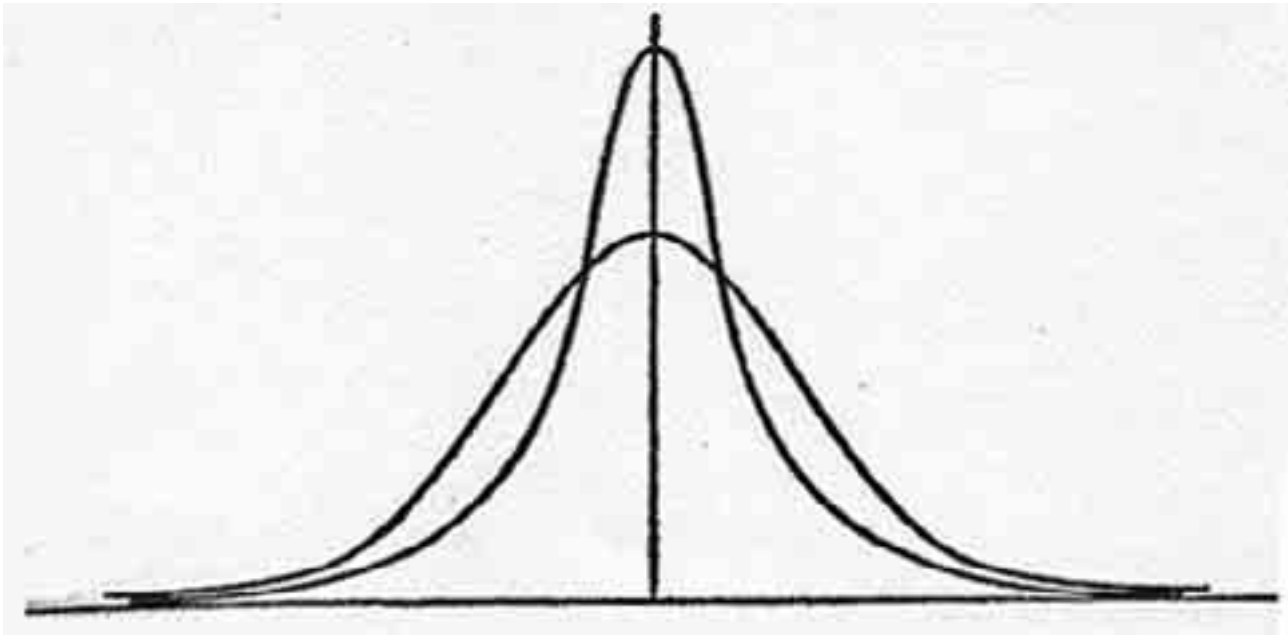


FIG. VII.5. ***Comparación de dos curvas normales de medias iguales pero con desviaciones estándar diferentes.***

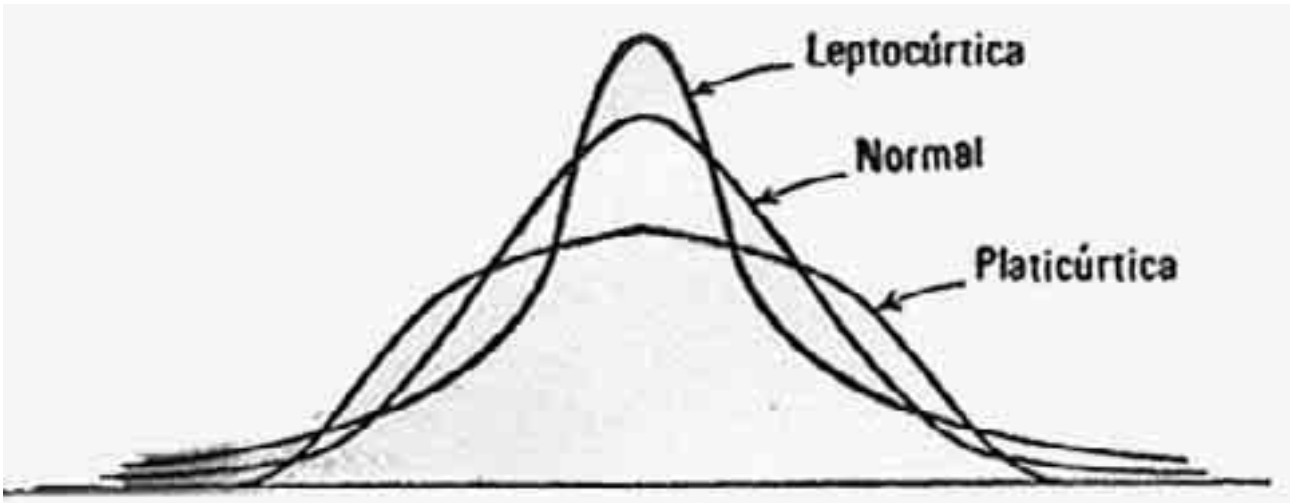


FIG. VII.6. ***Comparación de una curva normal con curvas de su misma desviación estándar pero distintas en cuanto a las cimas.***

VII.3. Áreas bajo la curva normal

Con frecuencia es necesario determinar la proporción de casos que quedan al interior de un intervalo dado. Afortunadamente, la curva normal posee una propiedad importante que hace que dicha tarea resulte relativamente sencilla. En efecto, resulta que independientemente de la media o de la desviación normal que

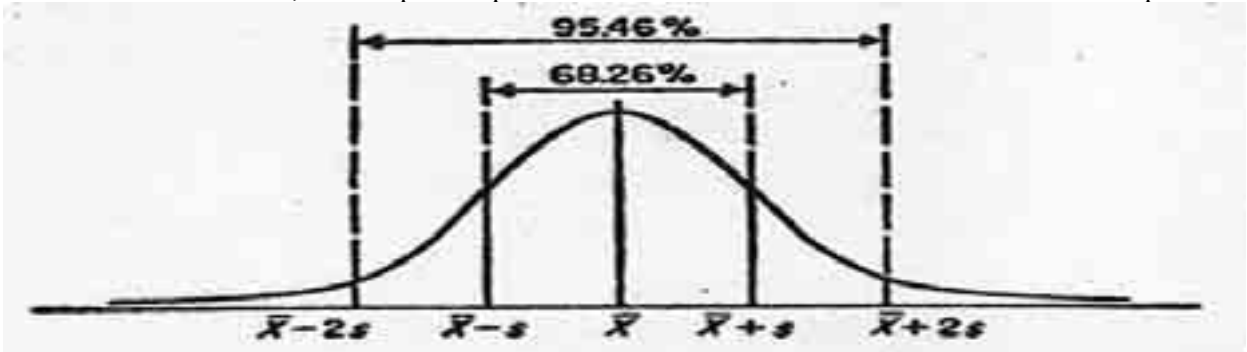


FIG. VII.7. *Areas debajo de la curva normal*

una curva ostente, habrá un *área constante* (o proporción de casos) *entre la media y una ordenada, que es una distancia determinada a partir de la media en términos de unidades de desviación estándar*. La figura VII.7 ayuda a ilustrar el sentido de esta afirmación.

Así, que, si vamos en una desviación estándar a la derecha de la media, encontraremos siempre .3413 del área incluida entre la media y la ordenada en dicho punto. Por consiguiente, dos veces dicha área, o .6826 estarán incluidas entre las dos ordenadas situadas a una desviación estándar a ambos lados de la media. En otros términos: un poco más de dos tercios de los casos se encontrarán siempre en el interior de una desviación estándar de la media. Y en forma análoga, el área comprendida entre la media y la ordenada a dos desviaciones estándar de aquélla será siempre .4773 y, por lo tanto, un poco más del 95 por ciento del área estará comprendido entre la pareja de ordenadas a dos desviaciones estándar a ambos lados de la media. Prácticamente todos los casos estarán comprendidos en el interior de tres desviaciones estándar de la media, aunque la curva normal se extienda teóricamente al infinito en ambas direcciones. Por supuesto, las distancias de la media no necesitan ser siempre múltiplos exactos de la desviación estándar. Mediante un procedimiento que vamos a describir en breve, es posible determinar las áreas entre dos ordenadas cualesquiera. Por ejemplo, si nos apartamos en 1.96 desviaciones estándar a ambos lados de la media, comprenderemos casi exactamente el 95 por ciento de área, en tanto que entre las ordenadas a 2.58 desviaciones normales de la media quedará incluido el 99 por ciento del área.

Esta propiedad de la curva normal brinda una interpretación de la desviación normal y un método para representar en forma visual el significado de esta medida de dispersión. Cierto número de distribuciones empíricas de frecuencia son lo bastante semejantes para que estas relaciones entre las áreas y la desviación normal se verifiquen razonablemente bien. Inclusive en el caso de distribuciones de ingresos, que propenden a distorsionarse en la dirección de los ingresos elevados, encontramos normalmente dos tercios de los casos en el interior de una desviación estándar de la media. Hay que tener presente, con todo, que aunque la curva normal proporciona una *interpretación* de la desviación estándar, esta propiedad no puede emplearse para *definir* lo que se entiende por desviación estándar. La definición se hace en términos de la fórmula. La propiedad en cuestión sólo se verifica en el caso de distribuciones normales o aproximadamente tales.

Resulta posible tomar cualquier curva normal y transformar sus valores numéricos de tal forma que pueda utilizarse un simple cuadro para evaluar la proporción de casos al interior de cualquier intervalo deseado. Vamos a ilustrar este proceso por medio de un ejemplo numérico. Supongamos que tenemos una curva normal con una media de 50 y una desviación estándar de 10. Busquemos la proporción de los casos en el intervalo de 50 a 65. Empezamos por determinar a cuántas desviaciones estándar se halla 65 de la media 50. Para ello tomamos la diferencia entre estos dos valores, esto es, 15, y dividimos entre la magnitud de la desviación estándar. En el presente caso el resultado es 1.5. De modo general podemos servirnos de la fórmula:

$$Z = \frac{X - \overline{X}}{s}$$

$$= \frac{65 - 50}{10} = 1.5$$

en donde X es el valor de la ordenada y Z representa la desviación con respecto a la media en unidades de desviación estándar.

*Antes de examinar cómo puede utilizarse el valor numérico de Z para determinar la proporción de los casos entre la media y la ordenada correspondiente a Z, permítasenos dar una interpretación alternativa de ésta. Podemos pensar en términos de una transformación efectiva de la variable X en la variable Z. En tanto que la distribución de la variable X es normal con una media de $\overline{X}$ y una desviación estándar de s, la nueva variable, en cambio, es normal con una media de cero y una desviación estándar de uno. [3] Esta desviación con una media cero y una desviación estándar de uno se designa como *forma estándar*, y la Z se designa a menudo como la *marca*. La transformación de variables se ilustra en la figura VII. 8. Sustraemos de cada X la constante $\overline{X}$. Al sustraer este valor constante (aquí 50) de cada X, hemos corrido cada marca original en

[3] La verificación de este hecho se deja como ejercicio (*véase* ejercicio 3).

50 unidades a la izquierda y, por lo tanto, hemos desplazado efectivamente la curva normal original a una posición directamente sobre el origen. Este tiene en cuenta el numerador en la expresión de Z. Dividimos

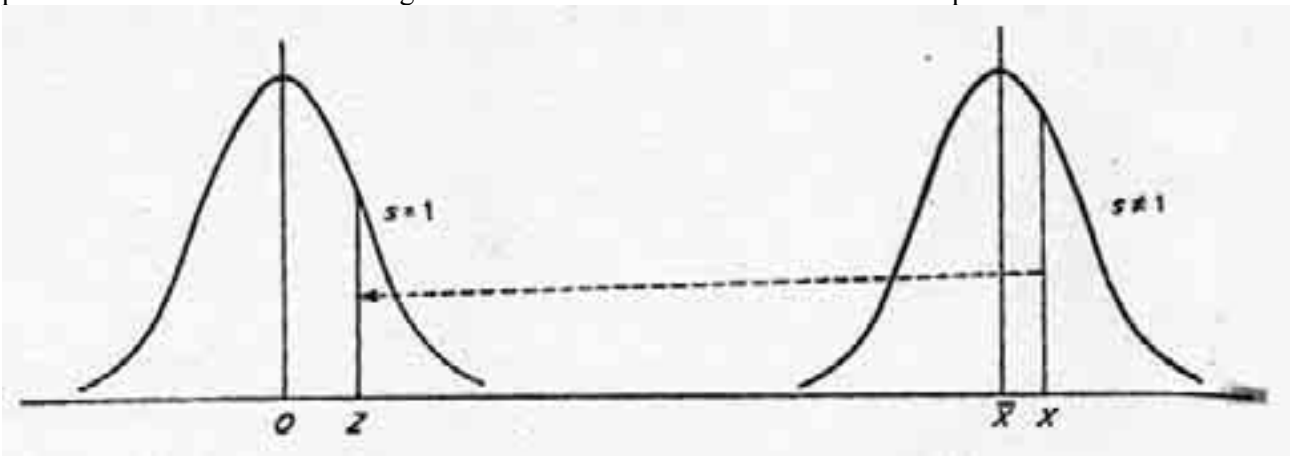


FIG. VII. 8. ***Comparación de las formas estándar y general de la curva normal***

ahora cada diferencia X- $\overline{X}$ entre la magnitud de la desviación estándar. Al hacerlo, o estrechamos la curva o la ensanchamos, según que su desviación estándar sea o no mayor que la unidad. Podemos, pues, pensar que hemos desplazado primero la posición de la curva normal original y que luego hemos cambiado la magnitud de la desviación estándar, de modo que quede sobre la forma estándar. Al dividir entre la desviación estándar de 10, hemos cambiado esencialmente las unidades a lo largo del eje horizontal, de modo que una distancia de 10 sobre el eje de X corresponde a la distancia de 1 sobre el eje de Z.

Independientemente de la interpretación que se dé, un valor de Z= 1.5 indica que la ordenada se encuentra a 1.5 desviaciones estándar de la media. En el caso de la forma estándar, esto significa, por supuesto, que la ordenada misma coincide con el valor 1.5 de la escala Z. Se han construido tablas que muestran áreas exactas para la forma estándar de la curva normal. El cuadro C del Apéndice 2 es una de ellas. Los valores de Z se dan de arriba abajo en el margen izquierdo, y horizontalmente arriba. Los dos dígitos de Z se obtienen leyendo de arriba abajo, y el tercero leyendo horizontalmente. Las cifras del cuerpo del cuadro indican la proporción del área entre la media (o sea cero) y la ordenada correspondiente a Z. En el ejemplo anterior, vemos que se hallan contenidas en dichos límites las .4332 del área. Si Z hubiera sido 1.52, el área correspondiente habría sido .4351.

VII.4. Ilustraciones suplementarias del empleo de la tabla normal

Supongamos que queremos hallar el área achurada de la curva normal indicada en la figura VII.9. En este caso el valor de Z es:

$$Z = \frac{143 - 168}{12} = \frac{-25}{12} = -2.08$$

El hecho de que Z sea negativa indica simplemente que el área achurada se sitúa a la izquierda de la media. Al utilizar la tabla normal, el signo de Z puede ignorarse, ya que la curva es perfectamente simétrica.

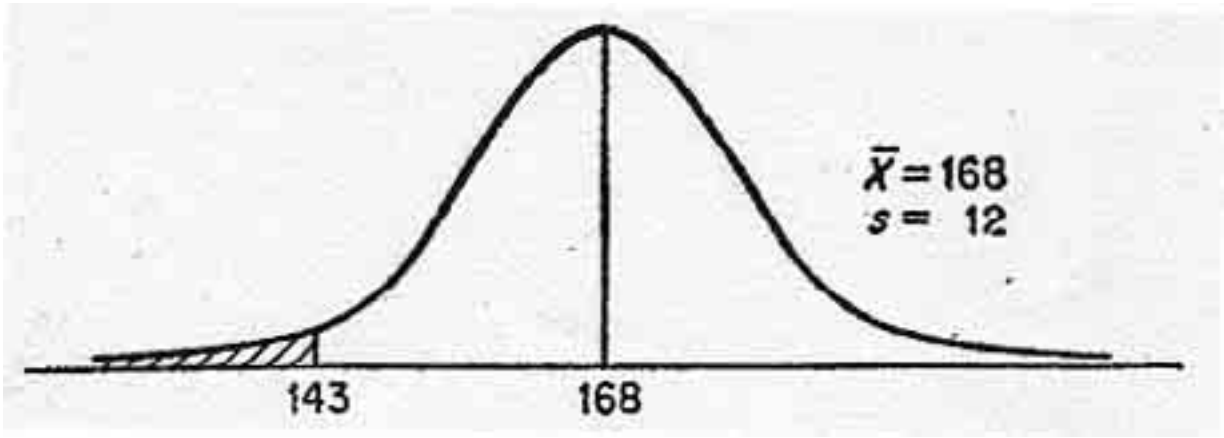


FIG. VII. 9. ***Curva normal, con porción achurada representando el área en una sola cola***

Del cuadro vemos que el área comprendida entre la media y una Z de 2.08 es .4812. Como quiera que el área total es la unidad, el área a la izquierda de la media ha de ser 5 (por simetría). Por consiguiente, el área achurada puede obtenerse restando el área comprendida entre la media y la ordenada del área total a la izquierda de la media. Así, pues:

(Proporción de casos ≤ 143) = .5000 – .4812= .0188

Por lo tanto, menos del 2 por ciento de los casos tienen marcas inferiores o iguales a 143.[4] El tipo de problema ilustrado en este ejemplo es muy corriente, debido al hecho de que las comprobaciones de hipótesis casi siempre comprenden las colas de una distribución de frecuencia. Si hubiéramos querido hallar el área total fuera de la región definida por 168 ± 25 (como la indican las áreas achuradas de la figura VII.10), habríamos doblado simplemente el resultado anteriormente obtenido, ya que las dos áreas achuradas son exactamente del mismo tamaño.

Para tomar otro ejemplo, supongamos que necesitamos, obtener el área achurada indicada en la figura VII. 11. Esta área se calcula hallando primero la proporción de casos entre la media y la ordenada B y

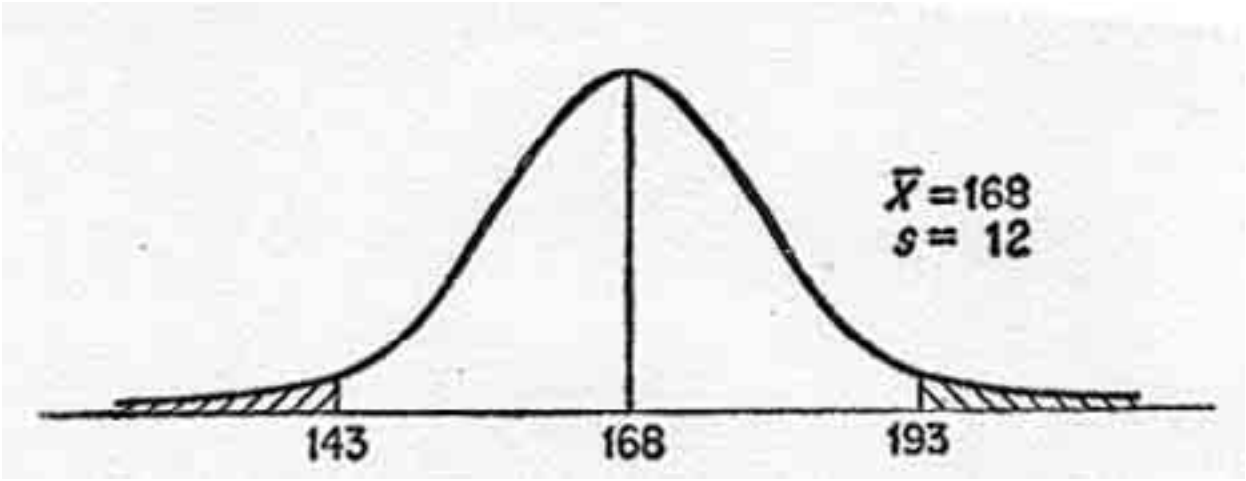


FIG. VII.10. ***Curva normal, con porciones achuradas presentando áreas en ambas colas***

sustrayendo luego la proporción de casos entre la media y la ordenada A. Las Z correspondientes a B y A son respectivamente 2.0 y 1.2. Tenemos, pues:

[4] En una distribución continua, la proporción de los casos que sean exactamente 143.0 será cero. Esto puede verse si imaginamos dos ordenadas extremadamente próximas una de otra. La proporción de casos entre estas dos ordenadas será también muy pequeña. Y si a continuación dejamos que las dos ordenadas se vayan aproximando indefinidamente, la proporción de los casos se hará infinitamente pequeña. Recuérdese que la línea matemática no tiene grueso. En la práctica podrá haber algunos casos con marcas de 143.0, debido a defectos de medición. Sin embargo, como quiera que estamos tratando de una distribución teórica, no importa que la ordenada ella misma se incluya o no en el intervalo. En adelante, nos referiremos simplemente al área entre dos ordenadas (pero sin comprender a éstas), o área inferior a un valor dado.

Proporción entre B y la media	.4773
Proporción entre A y la media	.3849
Proporción entre A y B	.0924

Por consiguiente, ligeramente más del 9 por ciento de los casos quedan entre .42 y .46 . Obsérvese que si se hubiera deseado obtener el área entre ordenadas a ambos lados de la media, el resultado se habría obtenido más fácilmente por adición que por sustracción.

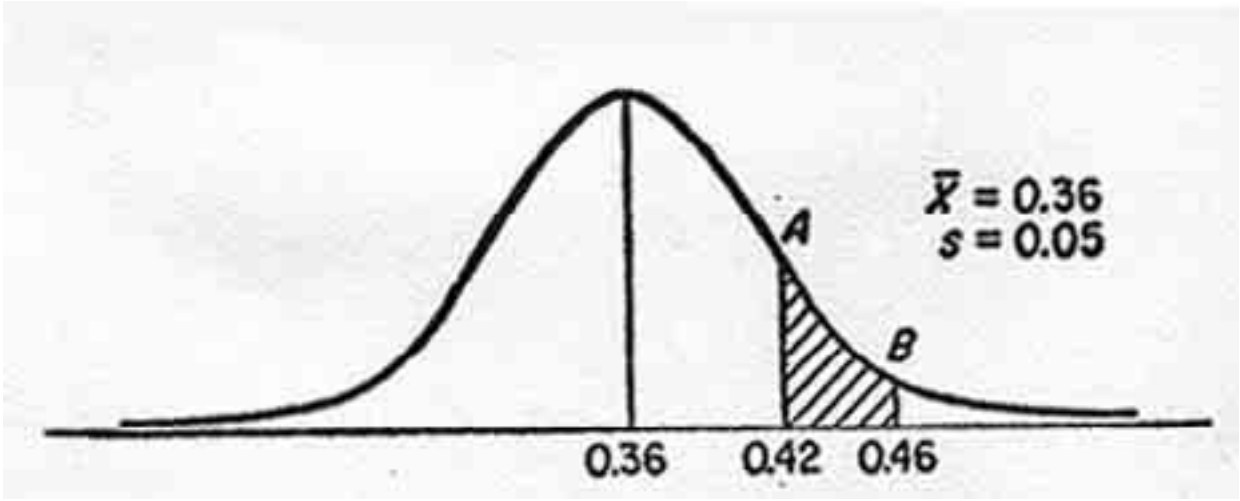


FIG. VII.11. *Curva normal con posición achurada, representando el área entre dos ordenadas.*

Glosario

Leptocúrtico
Distribución de frecuencia límite
Curva normal
Platicúrtico
Marca estándar

Ejercicios

1. Ya se calcularon la media y la desviación estándar de los datos del ejercicio 1, capítulo IV. ¿Cuál fue la proporción de los casos dentro de una desviación estándar de la media? ¿Al interior de dos desviaciones estándar? ¿De tres desviaciones estándar? ¿Con cuánta aproximación corresponden dichas cifras a las que encontraríamos si la distribución fuese exactamente normal? Contéstese a las mismas preguntas en relación con el ejercicio 2, capítulo IV. Compárense y explíquense diferencias entre los resultados de los dos grupos de datos.

2. Si la media de una distribución normal es de 80 y su desviación estándar de 12,

a) ¿Qué proporción de casos se halla entre 80 y 93? Respuesta, .3606.
b) ¿Qué proporción de casos se halla entre 90 y 105? ¿Entre 70 y 105? Respuesta, .1838.
c) ¿Qué proporción de casos es inferior a 68?
d) ¿Cuántas desviaciones estándar se necesitarían a ambos lados de media para obtener dos colas que comprendan cada una el 2 por ciento exactamente del área total? ¿El 10 por ciento del área total? Respuesta, 2.054.
e) ¿Cuál marca tiene por encima de ella el 4 por ciento de los casos? (en otros términos, sitúese la percentil 96).

3. Verifíquese que la forma estándar de la curva normal tiene una media de cero y una desviación estándar igual a la unidad. (Indicación: vuelva a escribirse la fórmula de la curva normal en términos de Z, aprovechando el hecho de que $Z=(X-\overline{X})/s$.)

4. Las calificaciones primarias de diversas pruebas de aptitud y actitud son tratadas a menudo por los psicólogos como escalas de intervalo. Dichas calificaciones suelen a menudo convertirse luego en calificaciones estándar con medias y desviaciones estándar convenientes. Supóngase que la calificación media primaria en un examen de admisión en la universidad es de 117 con una desviación estándar de 28.5. Supóngase que la calificación media primaria en un examen de admisión en la universidad es de 117 con una desviación estándar de 28.5. Supóngase, además, que esas calificaciones primarias están distribuidas normalmente.

a) ¿Cuál es la proporción de calificaciones por encima de 131? ¿Debajo de 79?
b) ¿Cuáles son las calificaciones primarias correspondientes a los cuartiles primero, segundo y tercero?
c) En los exámenes de la universidad, las calificaciones primarias se normalizan de modo que la media de la distribución normal sea exactamente de 500 y la desviación estándar de 100. Concretamente, ¿cómo se normalizarán los grupos de datos anteriores para obtener una media de 500 y una desviación estándar de 100? (Indicación: ¿cómo se normalizaría para obtener una media igual a cero y una desviación estándar igual a la unidad?)

BIBLIOGRAFÍA

Downie, N. M., y R. W. Heath: *Basic Statistical Methods*, 2ª ed., Harper and Row, Publishers, Incorporated, Nueva York, 1965, cap. 6.

Hagood, M. J., y D. 0. Price: *Statistics for Sociologists*, Henry Holt and Company, Inc., Nueva York, 1952, cap. 14.

Mueller, J. H., K. Schuessler y H. L. *Costner: Statistical Reasoning in Sociology*, 2ª ed., Houghton Mifflin Company, Boston, 1970, cap. 6.

Weinberg, G. H., y J. A. Schumaker: *Statistics: An Intuitive Approach*, Wadsworth Publishing Company, Inc., Belmont, Cal., 1962, cap. 8.

Weiss, R. S.: *Statistics in Social Research*, John Wiley & Sons, Inc., Nueva York, 1968, pp. 147-156.