Capítulo 11

LA CURVA NORMAL

Sumario

Objetivos específicos

Al finalizar el estudio del capítulo, el estudiante será capaz de:

1. Distinguir entre variables aleatorias discretas y continuas.
2. Comprender y explicar el concepto de valores esperados.
3. Conocer y explicar las propiedades de la curva normal.
4. Calcular probabilidades usando la tabla de la normal estándar.
5. Aplicar la curva normal a problemas sencillos de naturaleza descriptiva.

Resumen

En este capítulo, se explica el concepto de variable aleatoria y su clasificación en discretas y continuas. Luego, se procede a introducir la distribución normal; presentar sus características y propiedades y a explicar los procedimientos para el cálculo de probabilidades usando dicha distribución. Finalmente, se ilustran aplicaciones de la curva normal a problemas de naturaleza descriptiva.

11.1. VARIABLE ALEATORIA

Cuando los valores que puede asumir una variable dependen del azar, es decir, no se logran predecir con exactitud, se dice que es una variable aleatoria o al azar; pueden ser discretas o continuas, como ejemplo se considera el caso discreto, más adelante se hará referencia a las continuas.

11.1.1. Variable aleatoria discreta

Se le da el nombre de variable aleatoria, al azar o estadística cuando existe una determinada probabilidad de tomar cada uno de los valores que puede asumir.

Los dos aspectos más importantes en la definición son:

a) Existe una probabilidad determinada de que la variable asuma cada uno de los valores posibles.

b) El que asuma un valor específico depende del azar.

Como ejemplos de variables aleatorias discretas, pueden citarse: a) el número de puntos que se obtiene al lanzar un dado, b) el de escudos obtenidos al lanzar cuatro monedas al aire, c) el de tornillos defectuosos en una muestra al azar de 10 tornillos, d) el de personas, en una muestra al azar, que compra por internet.

Si para una variable aleatoria x se conocen los valores que puede tomar y la probabilidad de asumir cada uno de ellos, se sabe la distribución de probabilidad de la variable x o simplemente la distribución de x.

La distribución de x es, por lo tanto, una tabla o una función que indica, para cada valor posible de x, la probabilidad de que lo asuma.

Como ejemplo, considere la experiencia, tantas veces citada, de la suma de puntos obtenidos al lanzar dos dados corrientes. Utilice la siguiente notación:

x: número de puntos obtenidos al lanzar dos dados.

$f(x)$: la distribución de probabilidad de x.

$f(t) = P(x = t)$: probabilidad de que x asuma el valor de t.

$F(t) = P(x \leq t)$: probabilidad de que x sea menor o igual a t.

La variable x puede asumir valores enteros entre 2 y 12, y la ocurrencia de uno cualquiera dependerá del azar. Para obtener la probabilidad de un cierto valor x, deben sumarse las de los diferentes puntos muestrales que producen esa operación; así, por ejemplo, $x = 4$, o sea, una suma de valor 4, la producen los puntos muestrales (1,3); (2,2) y (3,1); cada uno tiene una probabilidad de $\frac{1}{36}$; por consiguiente, $F(4) = P(x \leq 4) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$. De igual forma se pueden obtener las probabilidades de que ocurran los otros valores de x. En el cuadro y gráfico de la figura 11.1 se representa la distribución de probabilidades de x. En la primera columna aparecen los valores posibles de x y en la segunda, la probabilidad de que ocurran $f(x)$. También se incluye la acumulada "menos de", $F(x)$, la cual representa la posibilidad de que suceda ese valor o uno menor.

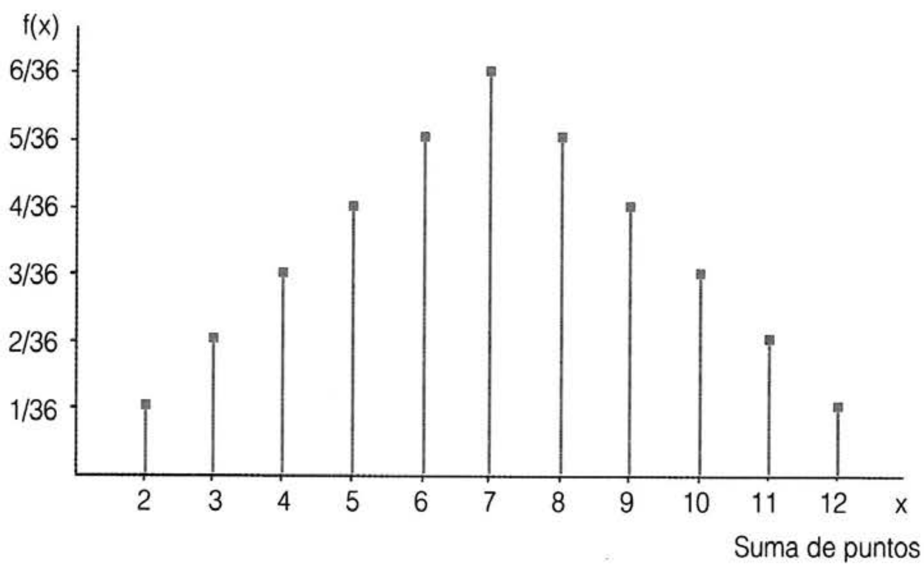


Figura 11.1. Gráfico que muestra la distribución de probabilidad de la variable suma de puntos al lanzar dos dados

Cuadro 11.1

DISTRIBUCIÓN DE PROBABILIDAD DE LA VARIABLE SUMA DE PUNTOS AL LANZAR DOS DADOS

x	$f(x)$	$F(x)$
2	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{2}{36}$	$\frac{3}{36}$
4	$\frac{3}{36}$	$\frac{6}{36}$
5	$\frac{4}{36}$	$\frac{10}{36}$
6	$\frac{5}{36}$	$\frac{15}{36}$
7	$\frac{6}{36}$	$\frac{21}{36}$
8	$\frac{5}{36}$	$\frac{26}{36}$
9	$\frac{4}{36}$	$\frac{30}{36}$
10	$\frac{3}{36}$	$\frac{33}{36}$
11	$\frac{2}{36}$	$\frac{35}{36}$
12	$\frac{1}{36}$	$\frac{36}{36}$
	$\frac{36}{36} = 1,00$	

A continuación, se usa el cuadro 11.1 para calcular:

a) La probabilidad de que la suma de puntos sea igual a 6.

$$P(x = 6) = f(6) = \frac{5}{36}.$$

b) La probabilidad de que la x sea menor o igual a 5.

$$P(x \leq 5) = F(5) = \frac{10}{36}.$$

Este resultado también puede obtenerse de la siguiente manera:

$$P(x \leq 5) = P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$$

$$= f(2) + f(3) + f(4) + f(5)$$

$$= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}.$$

Es evidente, sin embargo, que si se conoce $F(x)$ no hay motivo alguno para hacer los cálculos utilizando este último sistema.

c) ¿Cuál es la probabilidad de que x sea mayor o igual a 4, pero menor o igual a 10?

$$P(4 \leq x \leq 10) = P(x \leq 10) - P(x \leq 3) = F(10) - F(3) = \frac{33}{36} - \frac{3}{36} = \frac{30}{36}.$$

Como ejercicio, compruebe el valor obtenido calculando las probabilidades una a una.

d) ¿Cuál es la probabilidad de que x sea mayor o igual a 4, pero menor que 10?

$$P(4 \leq x < 10) = P(x < 10) - P(x \leq 3) = P(x \leq 9) - P(x \leq 3) = F(3)$$

$$= \frac{30}{36} - \frac{3}{36} = \frac{27}{36} = \frac{3}{4}.$$

Considere, como otro ejemplo, el juego denominado *más, menos y completo*; podría consistir en lanzar dos dados y el jugador apuesta bajo las siguientes condiciones, si apuesta a:

MÁS = más de 7 (gana si sale 8, 9, 10, 11 o 12 puntos).

COMPLETO = 7 exacto (gana si sale 7 puntos).

MENOS = menos de 7 (gana si sale 2, 3, 4, 5 o 6 puntos).

a) ¿Cuáles son las probabilidades correspondientes a cada una de las alternativas?

Probabilidad de que ocurra el evento MENOS

$$P(x < 7) = P(x \leq 6) = F(6) = \frac{15}{36}.$$

Probabilidad de que ocurra el evento COMPLETO

$$P(x = 7) = F(7) = \frac{6}{36}.$$

Probabilidad de que ocurra el evento MÁS

$$P(x > 7) = 1 - P(x \leq 7) = 1 - F(7) = 1 - \frac{21}{36} = \frac{15}{36}.$$

Es evidente que los eventos MÁS y MENOS tienen una probabilidad igual y mayor que la correspondiente al de COMPLETO. Por lo tanto, en idénticas condiciones, debe jugarse a MÁS o MENOS en vez de hacerlo a COMPLETO.

b) ¿Cómo debe interpretarse la probabilidad $\frac{15}{36}$ para el evento MÁS?

La probabilidad $\frac{15}{36}$ significa que si un jugador participa en un número grande de jugadas, debe esperarse que, en promedio, en cada 36 partidas gane 15 y pierda 21. En un número reducido, sin embargo, la proporción de veces ganadas puede ser diferente de $\frac{15}{36}$.

c) Si la apuesta es 100 colones y el premio 200 (le devuelven los 100 apostados y le dan 100 adicionales), ¿cuánto dinero habrá ganado o perdido un jugador que apueste a MÁS después de participar en 180 partidas?

Para participar en 180 jugadas, debe pagar 18 000 colones.[1]

Se espera que gane en $180\left(\frac{15}{36}\right) = 75$ partidas, o sea, recibirá en premios $75 \cdot 200 = 1500$ colones. Por lo tanto, después de 180 partidas se espera que haya perdido 15 000 colones.

11.1.2. Valores esperados

El concepto de *valor esperado o esperanza matemática* es muy importante en estadística; para introducirlo, se considerará el caso de valores esperados para variables discretas.

El valor esperado o esperanza matemática, $E(x)$, de una variable x, es simplemente la media aritmética de su distribución de probabilidad. Equivale a μ, símbolo que, como se indicó, representa la media de la población o media teórica.

$$\mu = E(x) = \sum xf(x)$$

en donde:

$x =$ variable aleatoria

$f(x) =$ probabilidad de que suceda x

$E(x) =$ valor esperado de x.

También, se puede calcular el valor esperado de $(x-\mu)^2$; el cual constituye la variancia teórica de x o de la población y se define de la siguiente forma:

$$\sigma^2 = E(x-\mu)^2 = \sum (x-\mu)^2 f(x).$$

1. Obviamente, esta es una pérdida en términos monetarios, y no contempla la satisfacción –en emociones, por ejemplo– que pueda haber recibido el jugador de su participación en el juego.

Ejemplo 1

Cálculo de la media y la variancia teórica, para x número de puntos, al lanzar un dado.

Cuadro 11.2

CÁLCULO DE LA MEDIA Y LA VARIANCIA TEÓRICA

x	$f(x)$	$xf(x)$	$x-\mu$	$(x-\mu)^2 f(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$	2,5	$(6{,}25)\ \frac{1}{6}$
2	$\frac{1}{6}$	$\frac{2}{6}$	–1,5	$(2{,}25)\ \frac{1}{6}$
3	$\frac{1}{6}$	$\frac{3}{6}$	–0,5	$(0{,}25)\ \frac{1}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$	0,5	$(0{,}25)\ \frac{1}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$	1.5	$(6{,}25)\ \frac{1}{6}$
6	$\frac{1}{6}$	$\frac{6}{6}$	2,5	$(6{,}25)\ \frac{1}{6}$
		$\frac{21}{6}$		$(17{,}50)\ \frac{1}{6}$

$$\mu = E(x) = \frac{21}{6} = 3{,}5 \quad \sigma^2 = \frac{17{,}50}{6} = 2{,}91.$$

La naturaleza teórica del valor esperado queda muy clara en este caso, si se tiene en cuenta que el valor 3,5 no puede darse en la realidad. Este indica, simplemente, si al lanzar el dado un número muy grande de veces, se anota el resultado obtenido y luego se calcula el promedio de esos valores, se obtendrá 3,5.

Ejemplo 2

En una urna hay 30 pequeños paquetes, todos iguales excepto en su contenido: 15 con una moneda de veinticinco colones, 7 con una de 50, 5 con una de 100 y 3 con una de quinientos.

a) Si un paquete es escogido al azar, ¿cuál es el valor esperado de la moneda contenida en él?

x	$f(x)$
25	$\frac{15}{30}$
50	$\frac{7}{30}$
100	$\frac{5}{30}$
500	$\frac{3}{30}$
	$\frac{30}{30} = 1$

$$\begin{aligned} E(x) &= 25\frac{15}{30} + 50\frac{7}{30} + 100\frac{5}{30} + 500\frac{3}{30} \\ &= \frac{375 + 350 + 500 + 1500}{30} \\ &= \frac{2725}{30} = 90,83\,. \end{aligned}$$

Por lo tanto, $E(x) = 90,83$ colones.

b) ¿Cómo se interpreta el valor obtenido en a)?

Se interpreta de la siguiente forma: si la experiencia de seleccionar un paquete se repite un número grande de veces, bajo las mismas condiciones, y luego se calcula el monto promedio por cada uno de los elegidos, este será de 91,8 colones, aproximadamente.[2]

◇◇◇

11.2. LA VARIABLE ALEATORIA CONTINUA

Una variable es continua en un intervalo si dados dos valores en él, tan cercanos como se quiera, siempre hay otro valor entre ellos que podrá ser tomado por la variable. También, es continua en un intervalo si puede adquirir cualquier valor del número infinito de este. El peso, la estatura, la velocidad, la duración de una conversación telefónica y la

2. El valor esperado un promedio, entonces asume valores que no son posibles en la realidad del ejemplo, como 90,83 colones.

edad son ejemplos de variables continuas. En el último caso, evidentemente, la persona que pasa de una edad a otra debe tener todas las intermedias.

En el caso de las variables discretas, es posible asociar una probabilidad de ocurrencia a cada uno de sus valores. Esto no puede hacerse, sin embargo, en las continuas; por más pequeño que sea un intervalo, contendrá un número infinito de valores y el cálculo de la probabilidad de uno específico daría cero. Debido a esto, se aborda la definición de la probabilidad en términos de un intervalo.

Se define, entonces, una variable aleatoria continua como aquella para la cual existe una determinada probabilidad de tomar valores en cada uno de los intervalos comprendidos dentro de su campo de variación. Estas probabilidades determinan la denominada función de densidad de la variable, la cual es similar a la de probabilidad, en el caso de las discretas, y se representan gráficamente por curvas continuas.

Si se tiene un cierto número de observaciones de una característica continua, lo apropiado es construir una distribución de frecuencias y un histograma con el fin de lograr una representación gráfica del comportamiento de la variable. Lógicamente, una mejor apreciación de su patrón se logra si el histograma se basa en las frecuencias relativas.

Ahora bien, usualmente el histograma se basa en una muestra de observaciones y, por lo tanto, evidencia una idea aproximada de la forma como se distribuye según la característica de la población. Para conocer la curva exacta, sería necesario tener las observaciones de toda la población de interés; por consiguiente, el histograma (construido con las frecuencias relativas) es una estimación de la curva continua y debe concebirse como el límite al que tiende al aumentarse el número de observaciones y reducirse la amplitud de las clases de la distribución.

En resumen, conforme se aumenta el número de observaciones y se reduce el intervalo de clases de la distribución de frecuencias, el histograma tiende a un límite que es una curva continua. Para ilustrar este punto, se presenta seguidamente, mediante una distribución de frecuencias, un histograma *(cuadro 11.2 y el gráfico de la figura 11.2)* de la talla, estatura en centímetros, de 3600 recién nacidos en hospitales de Costa Rica, en enero de 1971. Estos datos excluyen a niños enfermos o con defectos congénitos.

Puede notarse que la distribución es bastante simétrica, además existe una tendencia a la concentración de valores en la parte central, mientras los extremos no se presentan muy continuamente.

Esta distribución de frecuencias y su histograma pueden tomarse como una estimación de la curva continua que representa la verdadera distribución de las tallas de los recién nacidos de Costa Rica, la cual se desconoce.

Cuadro 11.3

TALLA EN CENTÍMETROS DE 3600 RECIÉN NACIDOS EN HOSPITALES
COSTA RICA ENERO DE 1971

Talla (cm)	Punto medio	Número de recién nacidos	Porcentajes
42,5 a menos de 43,5	43	20	0,55
43,5 a menos de 44,5	44	33	0,92
44,5 a menos de 45,5	45	75	2,08
45,5 a menos de 46,5	46	135	3,75
46,5 a menos de 47,5	47	250	6,94
47,5 a menos de 48,5	48	387	10,75
48,5 a menos de 49,5	49	537	14,92
49,5 a menos de 50,5	50	719	19,97
50,5 a menos de 51,5	51	532	14,78
51,5 a menos de 52,5	52	394	10,94
52,5 a menos de 53,5	53	245	6,81
53,5 a menos de 54,5	54	146	4,06
54,5 a menos de 55,5	55	76	2,11
55,5 a menos de 56,5	56	32	0,89
56,5 a menos de 57,5	57	19	0,25
		3600	100,00

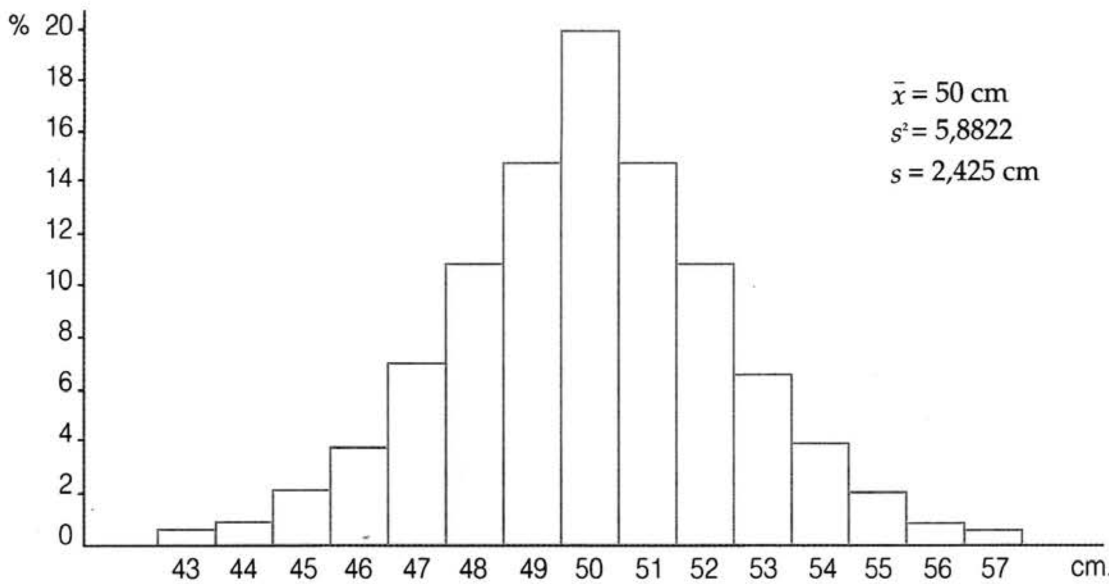


Figura 11.2. Gráfico que muestra la talla en centímetros de 3600 recién nacidos en hospitales de Costa Rica, enero de 1971

La distribución se basa en 3600 recién nacidos y utiliza un intervalo de clase de un centímetro. Obviamente, si se aumenta el número de observaciones al doble, por ejemplo, se pueden establecer más clases pero de medio centímetro de amplitud y si se ampliaran aún más, por ejemplo a un cuarto de centímetro. Si se continúa el proceso de acrecentar el número de observaciones, los intervalos de clase serían cada vez más estrechos y el histograma iría asemejándose cada vez más a una curva. En este ejemplo, que se refiere a la variable talla, el límite al cual tiende el histograma es, seguramente, la curva normal.

En el histograma, las áreas de las barras son proporcionales a las frecuencias relativas. En el presentado en el gráfico de la figura 11.2, por ejemplo, el área de la sexta barra representa la frecuencia relativa correspondiente a la clase 47,5 – 48,5. De acuerdo con la relación existente entre frecuencia relativa y probabilidad, 0,11 se interpreta como una estimación de la posibilidad de que la variable estatura esté entre 47,5 y 48,5 centímetros. El valor verdadero de esa probabilidad lo daría el área sombreada bajo la curva, tal y como se indica en el gráfico de la figura 11.3.

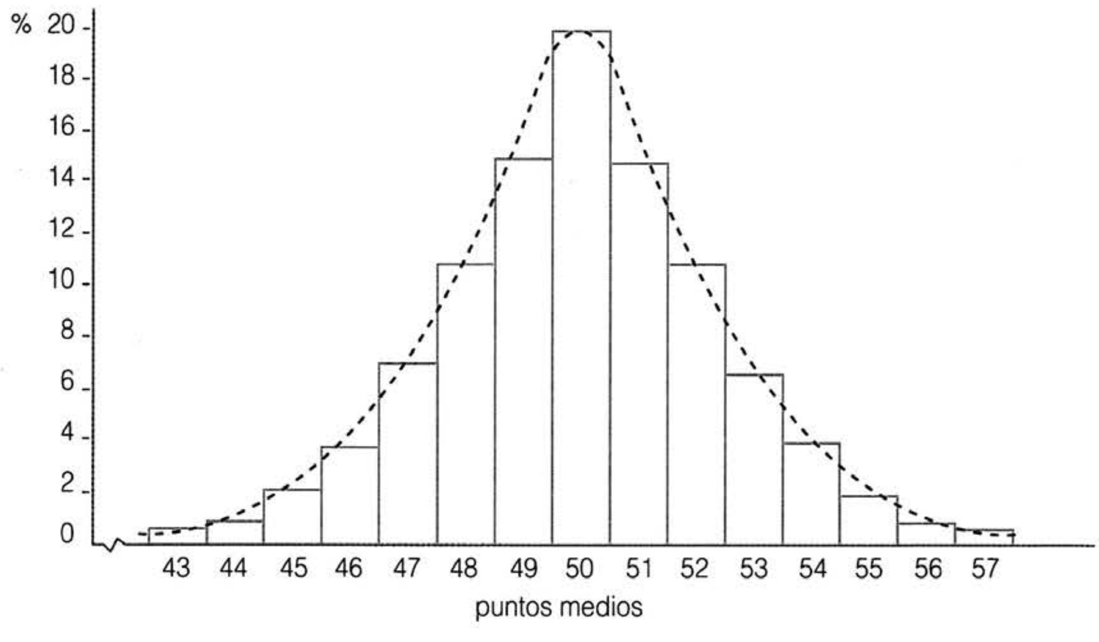


Figura 11.3. Gráfico de la curva continua como límite del histograma

Si se quiere estimar la probabilidad de que la talla de un recién nacido, escogido al azar, esté entre 49,5 y 53,5 cm, deben sumarse las frecuencias relativas correspondientes a las clases octava, novena, décima y undécima $(0,20 + 0,15 + 0,11 + 0,07 = 0,53)$, lo que gráficamente está representado por la suma de las áreas de las barras correspondientes del histograma. El valor 0,53 constituye, por lo tanto, una estimación del área bajo la curva continua, entre los puntos 49,5 y 53,5.

La suma de las frecuencias relativas de todas las clases de una distribución es igual a uno; también, el área bajo una curva, dentro de los límites de variación de la variable considerada, es igual a uno.

En un histograma, $f_{\frac{i}{n}}$ da el área correspondiente a la clase i y $F_{\frac{i}{n}}$ representa el área acumulada hasta el límite superior de esa clase. En una curva continua, $f(x)$ representa el valor de la ordenada y o, lo que es lo mismo, el de la función para el punto de abscisa x. Por otra parte, $F(x)$ representa el área acumulada hasta el punto x. Si se quiere conocer el área entre dos puntos a y b ($b > a$) se debe hacer:

$$F(b) - F(a).$$

Esta identidad entre áreas bajo la curva no debe olvidarse, pues facilita la comprensión, el razonamiento y la resolución de los problemas que se considerarán más adelante.

La función $f(x)$ se conoce con el nombre de *función de densidad de probabilidad de la variable* x y, dentro del intervalo total de variación de x, debe satisfacer dos condiciones fundamentales:

a) Debe ser positiva: $f(x) \geq 0$.

b) El área bajo la curva debe ser igual a uno.

En estadística se utiliza una gran cantidad de curvas continuas como modelos para representar diversas situaciones y fenómenos. El más empleado, por diversas razones analíticas y prácticas, es la curva normal también llamada *normal de error o campana de Gauss.*

11.3. CURVA NORMAL

Es uno de los primeros modelos utilizados, alrededor del cual se ha desarrollado casi toda la teoría y la práctica de la estadística. Recibe el nombre de "normal de error", proviene cuando se encuentra que los errores, en las mediciones y en las observaciones, se distribuyen siguiendo un patrón.

Expresión algebraica $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < \infty.$

x = variable de interés.

σ = desviación estándar de la variable x.

μ = media aritmética de la variable x.

e = constante matemática, base de los logaritmos naturales $e \approx 2,71828$.

Como puede apreciarse, conociendo los valores de μ y σ , la curva normal queda completamente definida. Por ejemplo, si la desviación estándar es 2 y la media 4, la expresión algebraica de esa curva normal sería:

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-4}{\sigma}\right)^2}.$$

11.3.1. Características

Es conveniente destacar algunas de las características más importantes de la curva normal:

a) El área bajo la curva es igual a uno. Se trata de una función de densidad.

b) Es simétrica con respecto a la media aritmética (por lo tanto, con respecto a la moda y la mediana).

c) Queda perfectamente determinada cuando se conoce μ y σ.

11.3.2. Importancia

Hay varias razones relevantes, las cuales han contribuido a que la normal sea tan vital en la teoría y la práctica de la estadística:

a) La curva normal y su integral reúnen muchas ventajas de tipo matemático y analítico.

b) Son bastantes los fenómenos que se comportan siguiendo la curva normal o aproximándose a ella suficientemente.

c) Las técnicas basadas en la curva normal se pueden aplicar a las curvas unimodales, aunque no sean normales, los resultados son bastante satisfactorios.

d) Muchas variables, que tienen distribuciones bastante asimétricas, pueden convertirse en normales o en curvas muy próximas a las normales utilizando una transformación adecuada.

e) La media muestral $\overline{X}$, que varía de una muestra a otra, se distribuye en forma normal o aproximadamente normal.

11.4. NORMAL ESTÁNDAR

Por procedimientos matemáticos bastante conocidos, se puede calcular el área bajo una curva entre dos valores de la variable; lo cual es cuestión de conocer la expresión algebraica de la función a la que corresponde.

El área bajo la curva normal es siempre igual a la unidad (se trata de una función de densidad); por lo tanto, si se conoce la expresión algebraica, puede determinarse en un intervalo y esta será igual a la probabilidad de que la variable asuma un valor dentro de él.

Una curva normal queda definida si se conocen su media aritmética y su desviación estándar; no obstante, como los valores posibles para μ y σ son prácticamente infinitos, es fácil concluir que el número de curvas normales viables también lo es. Debido a esto, el uso de esa distribución sería imposible en la práctica, pues habría que construir una tabla de probabilidades para cada una de las potenciales curvas normales, o sea, para cada pareja diferente μ y σ. Sin embargo, el problema se ha resuelto de una manera muy simple y satisfactoria: se ha calculado una tabla de probabilidades para una curva normal con media cero $(\mu = 0)$ y desviación estándar uno $(\sigma = 1)$; cualquier situación referente a una variable normal se traslada o refiere a la tabla mencionada, mediante simples operaciones algebraicas.

La normal, con media cero y desviación estándar uno, se conoce con el nombre de *normal estándar* y se representa por N(0,1); en cualquier libro de estadística se encuentran tablas de ordenadas y probabilidades para ella, como la que aparece en el cuadro 11.4.

Si se quiere calcular alguna probabilidad para una variable normal, que tenga el promedio diferente de cero, la desviación estándar diferente de uno o ambos parámetros distintos, se realiza una transformación algebraica sencilla denominada *estandarización*, mediante la cual el problema se traslada a la normal estándar y la probabilidad se obtiene de esa tabla única que es la de N(0,1).

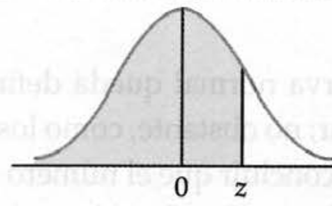
Primero, se abordará el uso de la tabla normal estándar para el cálculo de probabilidades; luego la resolución del problema cuando la curva es normal, pero no estándar.

Cuadro 11.4

TABLA DE DOS COLAS PARA ÁREAS DE LA CURVA NORMAL
TABLA DE ÁREAS BAJO LA CURVA NORMAL

Distribución normal estándar acumulada

Valores de $F(z)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6180	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7519
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8035	.8078	.8106	.8133
.9	.8159	.8186	.8512	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8636	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9174	.9184	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	9750	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9703	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9816	.9850	.9854	.9857
2.2	.9661	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9986	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9987	.9997	.9997	.9998

Fuente. Universidad de Costa Rica. Facultad de Ciencias Económicas. Escuela de Estadística. 1980

11.4.1. Uso de la tabla de la normal estándar

Para facilitar las indicaciones sobre el manejo de la tabla de probabilidades de la N(0,1), se utilizará la siguiente notación:

z = variable normal estándar $-\infty < z < \infty$.

$N(t)$ = área acumulada hasta t (t es un valor particular de z).

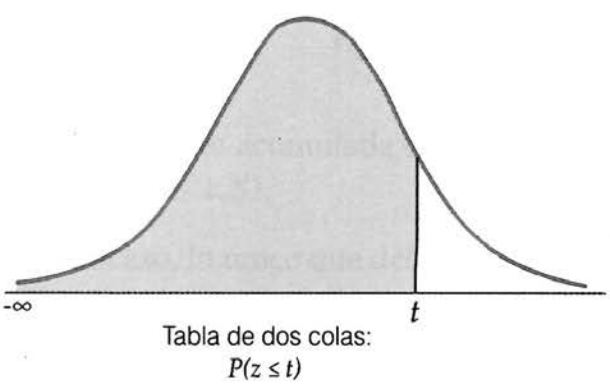
$N(t) = P(z \leq t)$ = área acumulada hasta t = probabilidad de que z sea menor o igual a t.

De acuerdo con la notación anterior, $P(a \leq z \leq b) = N(b) - N(a)$.

En algunas tablas se da el área acumulada desde $-\infty$ hasta cada uno de los valores de z; se conocen como *tablas de dos colas*, porque permiten conocer directamente las probabilidades acumuladas, tanto para valores positivos como para los negativos de la variable z. En otras, denominadas de *una cola*, se dan las probabilidades acumuladas solo para los valores de z ubicados en la parte derecha de la curva (positivos). En una variante, solo se proporciona el área bajo la curva entre $z = 0$ y $z = t$ y el analista debe agregar 0,50 para tener el área desde $-\infty$ a 0. En otra variante, la tabla ya incluye el ajuste de 0,50, haciéndola prácticamente equivalente a la de dos colas.

En el caso de las tablas de una cola, el cálculo de los valores de $N(z)$, para valores negativos de la variable z se realiza usando la expresión $N(-z) = 1 - N(z)$, la cual se deriva de la propiedad de simetría de la normal con respecto a $z = 0$ y en el entendido de que el área bajo la curva es igual a 1.

Se utilizará una tabla de una cola con el 0,50 ya incorporado, esta aparece en el cuadro 11.4.[3]



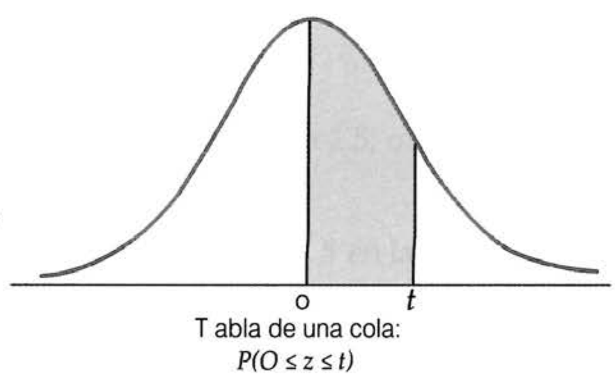


Figura 11.4. Gráfico que muestra una tabla de dos colas $P(z \leq t)$ y tabla de una cola $P(0 \leq z \leq t)$

3. Los cálculos pueden hacerse usando las funciones correspondientes, disponibles en las hojas de cálculo y en los programas de proceso estadístico de datos. Una posibilidad son las funciones DISTR NORM ESTAND y DISTR NORM ESTAND INV de Excel, estas dan iguales resultados que una tabla de dos colas.

En el empleo de la tabla se debe considerar lo siguiente:

a) En la primera columna aparecen los valores de z con un decimal.

b) En la primera fila aparece el segundo decimal para dar mayor precisión a la tabla.

c) En el cuerpo de la tabla se muestra el área acumulada desde $-\infty$ hasta un valor positivo dado de z.

Como se dijo, esta tabla da el área acumulada desde $-\infty$ hasta un valor de z; los valores de z alcanzan hasta 3,00. La situación se ilustra gráficamente para el caso de $z = 1,42$.

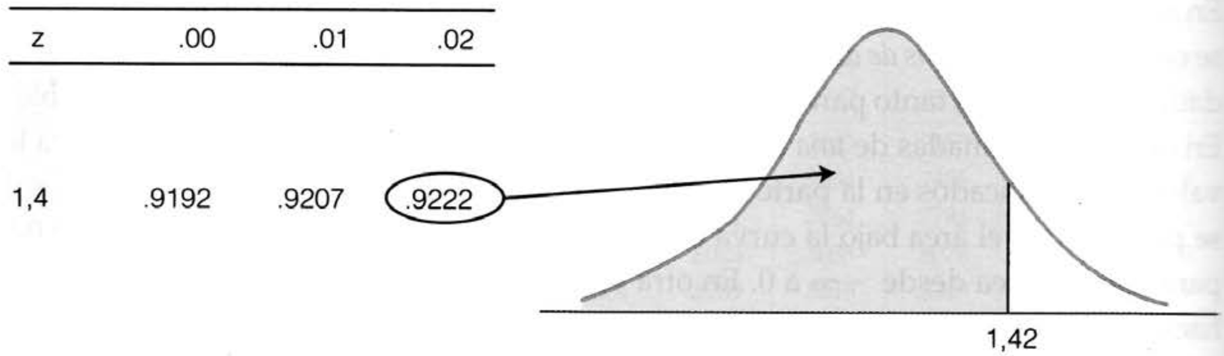


Figura 11.5. Área acumulada para $z = 142$

El hecho de que la tabla dé solamente áreas acumuladas, desde $-\infty$ hasta valores positivos de z, no constituye ningún problema especial, pues al ser simétrica la curva, las áreas acumuladas desde $-\infty$ hasta cierto z negativo son exactamente iguales a las que van desde $-\infty$ hasta el mismo valor z positivo. Esto puede apreciarse en los siguientes gráficos para el caso de $z = 1,5$.

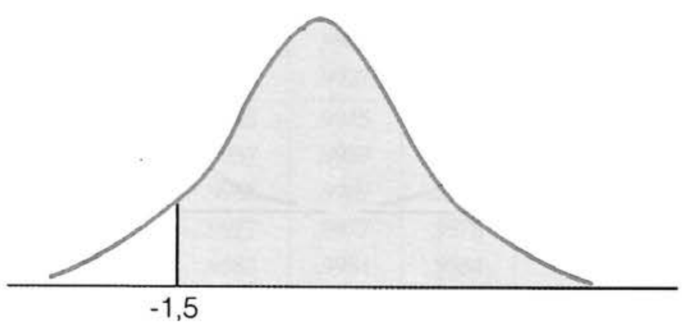


Figura 11.6. Gráfico de $P(z \geq -1,5) = 0,9332$

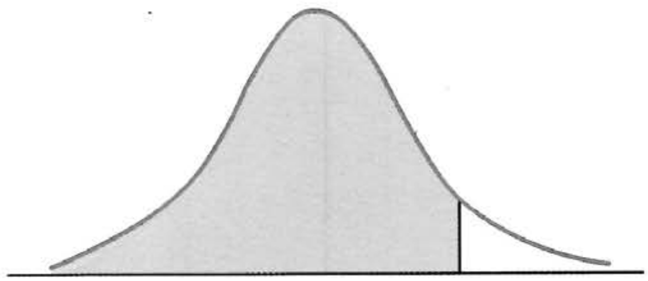


Figura 11.7. Gráfico de $P(z \leq 1,5) = 0,9332$

Es evidente que cualquier problema en el cual aparezcan cálculos de probabilidades o áreas, para valores negativos de z, pueden realizarse fácilmente con la tabla, recurriendo a la propiedad de simetría respecto a la media que tiene la curva normal. Aunque z puede asumir valores negativos y positivos $(-\infty < z < +\infty)$, las áreas o probabilidades siempre deben ser no–negativas y nunca superiores a uno.

Para ilustrar el uso de la tabla, a continuación se desarrollan ejemplos concretos.

1.

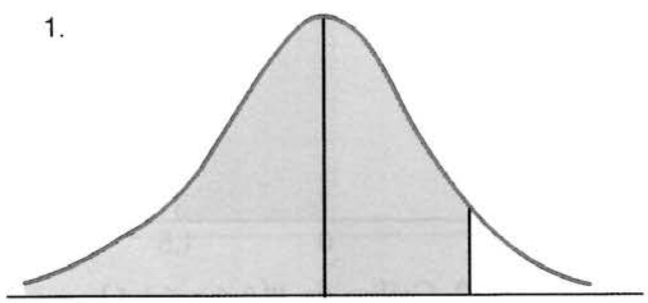


Figura 11.8. Gráfico de $P(z \leq 1,5)$

Se busca el área acumulada bajo la curva normal desde $-\infty$ hasta 1,5, o sea simbólicamente: $P(z \leq 1,5)$.

En este caso, lo único que debe hacerse es buscar el valor de $z = 1,5$ en la tabla y leer el valor del área acumulada correspondiente, que es igual a 0,9332.

2.

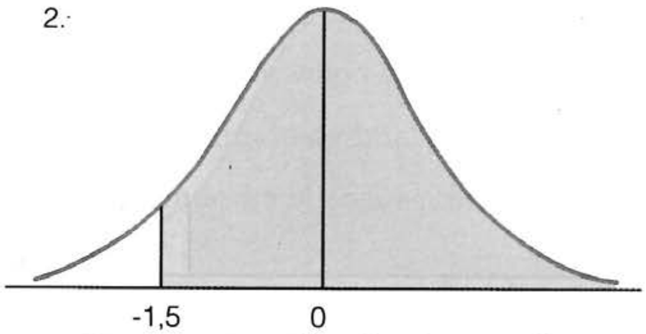


Figura 11.9. Gráfico de $P(z \geq -1,5)$

Se averigua el área bajo la curva entre –1,5 y $+\infty$, es decir, $P(z \geq -1,5)$.

Por simetría, el área bajo la curva entre –1,5 y $+\infty$ es igual al área bajo la curva entre $-\infty$ y 1,5. El problema es, evidentemente, similar al planteado en 1) y el área es igual a 0,9332. Por lo tanto, $P(z \geq -1,5) = 0,9332$.

3.

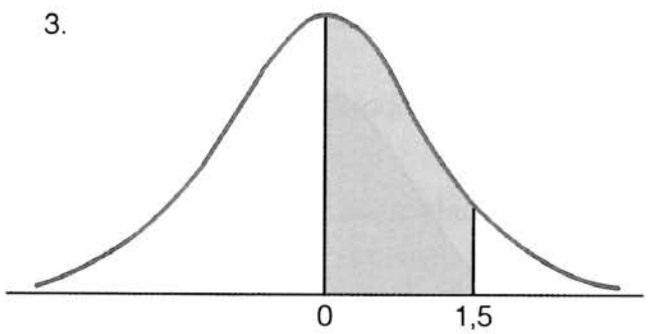


Figura 11.10. Gráfico de $P(0 \leq z \leq 1,5)$

Se busca el área bajo la curva entre 0 y 1,5, o sea, $P(0 \leq z \leq 1,5)$.

En este caso se sabe que el área acumulada bajo la curva, entre $-\infty$ y 1,5, es igual a 0,9332, a la cual se debe restar el área bajo la curva entre $-\infty$ y 0, que es igual a 0,5. Por lo tanto:

$$\begin{aligned} P(0 \leq z \leq 1,5) &= P(z \leq 1,5) - P(z \leq 0) \\ &= 0,9332 - 0,5 = 0,4342. \end{aligned}$$

4.

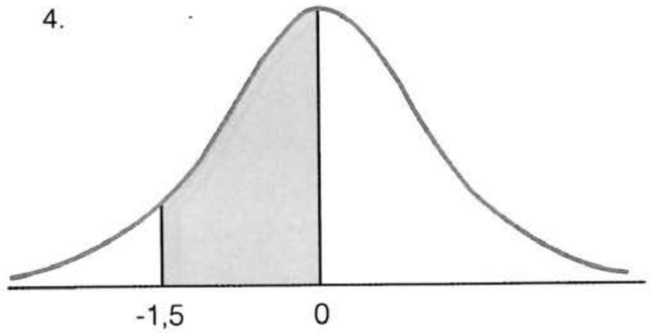


Figura 11.11. Gráfico de $P(-1,5 \leq z \leq 0)$

Se busca el área bajo la curva entre –1,5 y 0, simbólicamente, $P(-1,5 \leq z \leq 0)$.

Por simetría, el área bajo la curva entre –1,5 y $+\infty$ es igual al área de la curva entre $-\infty$ y 1,5, lo que corresponde al caso anterior y se produce, entonces, el mismo resultado de 0,4332. Por lo tanto, $P(-1,5 \leq z \leq 0) = 0,4332$.

5.

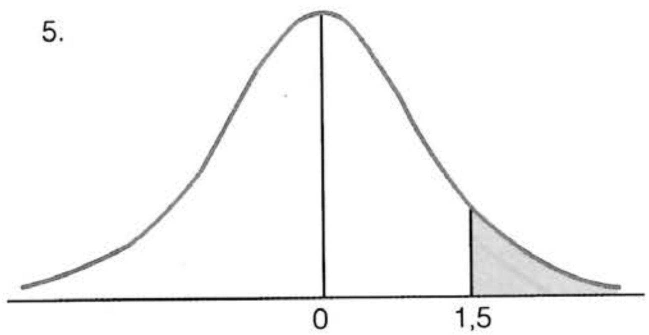


Figura 11.12. Gráfico de $P(z \geq 1,5)$

Se busca el área bajo la curva entre 1,5 y $+\infty$, esto es $P(z \geq 1,5)$.

Se sabe que el área bajo toda la curva es 1 y además que el área bajo la curva entre $-\infty$ y 1,5 es igual a 0,9332 (leyendo la tabla). Entonces, el área acumulada a la derecha de 1,5 se obtiene restando de 1 el área bajo la curva entre $-\infty$ y 1,5:

$$\begin{aligned} P(z \geq 1,5) &= 1 - P(z \leq 1,5) \\ &= 1 - 0,9332 = 0,0668. \end{aligned}$$

6.

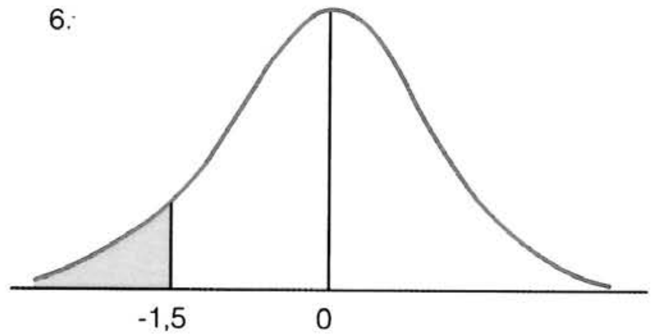


Figura 11.13. Gráfico de $P(z \leq -1,5)$

Se busca el área bajo la curva entre $-\infty$ y –1,5, es decir, el área da hasta –1,5. En símbolos, $P(z \leq -1,5)$.

Por simetría, se sabe que el área acumulada hasta –1,5 es igual al área entre 1,5 y $+\infty$.

$$P(z \leq -1,5) = P(z \geq 1,5) = 0,0668.$$

7.

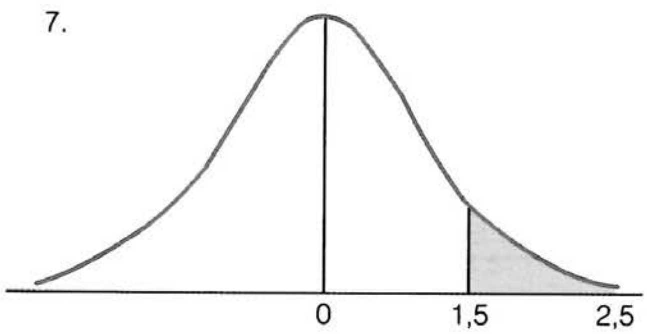


Figura 11.14. Gráfico de $P(1,5 \leq z \leq 2,5)$.

Se busca el área entre 1,5 y 2,5. En símbolos es $P(1,5 \leq z \leq 2,5)$.

En este caso, la tabla da las áreas acumuladas de $-\infty$ a 2,5 igual a 0,9938 y de $-\infty$ a 1,5 igual a 0,9332. El área deseada debe obtenerse por diferencia:

$$\begin{aligned} P(1,5 \leq z \leq 2,5) &= P(z \leq 2,5) - P(z \leq 1,5) \\ &= 0,9938 - 0,9332 = 0,0606. \end{aligned}$$

8.

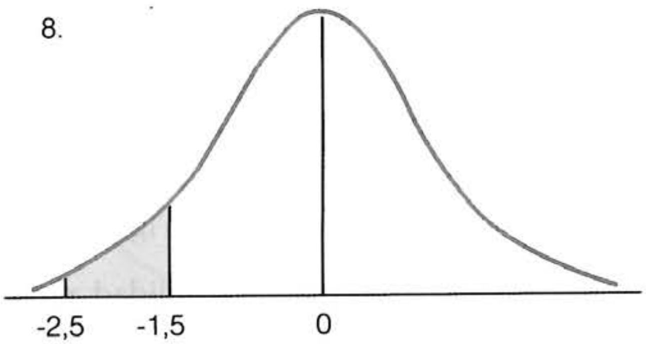


Figura 11.15. Gráfico de $P(-2,5 \leq z \leq -1,5)$

Se busca el área entre –2,5 y –1,5, o sea $P(-2,5 \leq z \leq -1,5)$.

En este caso, por simetría,

$P(-2,5 \leq z \leq -1,5) = P(1,5 \leq z \leq 2,5)$. Esto corresponde al anterior y produce el mismo resultado de 0,0606.

$$P(-2,5 \leq z \leq -1,5) = 0,0606.$$

9.

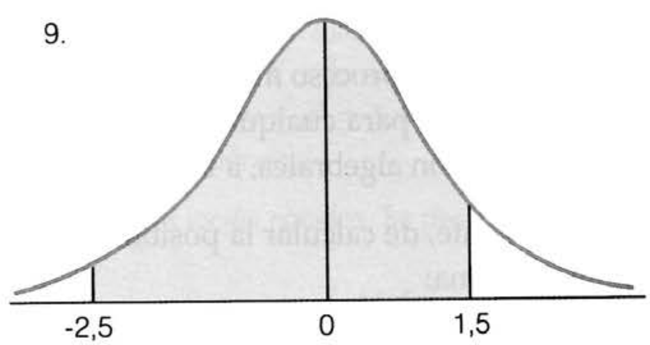


Figura 11.16. Gráfico de $P(-2,5 \leq z \leq 1,5)$

Se busca el área entre -2,5 y 1,5, o sea $P(-2,5 \leq z \leq 1,5)$.

En este caso, por simetría,

$P(z \geq -2,5) = P(z \leq 2,5) = 0,9938$ y que $P(z \leq -2,5) = P(z \geq 2,5) = 0,0062$.

Entonces:

$$\begin{aligned} &P(-2,5 \leq z \leq 1,5) = \\ &N(1,5) - N(-2,5) = \\ &0,9938 - 0,0062 = \\ &0,9270. \end{aligned}$$

11.4.2. Estandarización de una curva normal

En la práctica, casi la totalidad de las curvas normales tienen μ diferente de cero y σ diferente de uno, es decir, son normales pero no estándar. Un problema común es tener una variable x con distribución normal, media μ y desviación estándar σ, para la cual se quiere obtener la probabilidad de que x sea menor o igual que un cierto valor a.

En símbolos, $P(x \leq a) = ?$; para encontrar esta probabilidad, no se puede utilizar directamente la tabla por la sencilla razón de que la distribución de x no es $N(0,1)$.

¿Cómo debe resolverse el problema? Recurriendo al proceso denominado "estandarización", por medio del cual una variable cualquiera se transforma en una nueva, con media cero y desviación estándar uno.[4]

Estandarizar una curva normal es transformarla de manera que su media sea cero y su desviación estándar uno. Esto se logra restando a la variable x su promedio (μ) y dividiendo la diferencia por su desviación estándar (σ).

En general, si x tiene distribución normal con media μ y desviación estándar σ la variable $z = \frac{x-\mu}{\sigma}$ tiene distribución normal estándar. Es decir,

$$z = \frac{x-\mu}{\sigma} = N(0,1).$$

La estandarización es, por lo tanto, un proceso muy simple, que permite calcular probabilidades y resolver otros problemas para cualquier curva normal trasladándolos, por medio de una sencilla transformación algebraica, a la normal estándar.

El problema planteado inicialmente, de calcular la posibilidad de que x sea menor que a, se resolvería en la siguiente forma:

$$P(x \leq a) = P\left(\frac{x-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(z \leq \frac{a-\mu}{\sigma}\right).$$

La probabilidad de que z sea menor que $\frac{a-\mu}{\sigma}$ es calculable directamente en la tabla, porque z es normal estándar.

Como ejemplo, suponga que $\mu = 10$, $\sigma = 4$ y $a = 12$.

$$\begin{aligned} P(x \leq 12) &= P\left(\frac{x-10}{4} \leq \frac{12-10}{4}\right) = P(z \leq 0,5) \\ &= N(0,5) = 0,6915. \end{aligned}$$

El área a la izquierda de $x = 12$ es equivalente a la de la izquierda de $z = 0,5$ e igual a 0,6915.

4. El concepto de estandarización, aplicado al caso de notas o puntajes, se estudió en la sección 9.5.

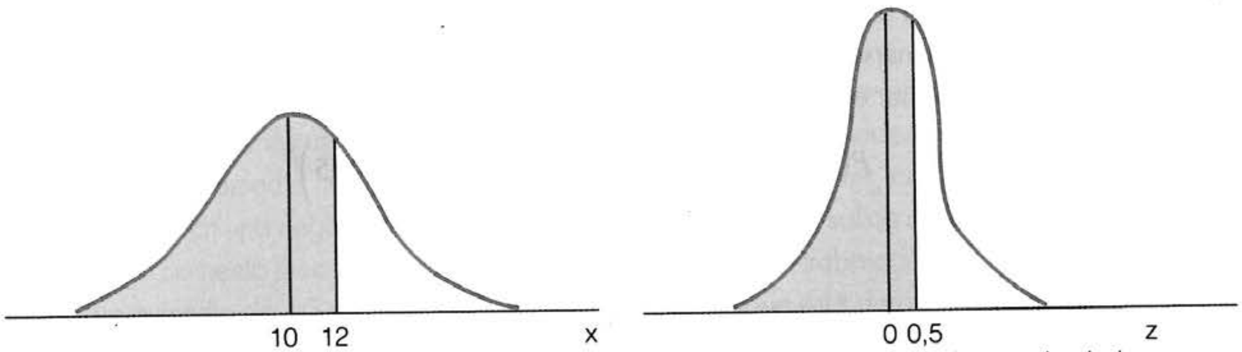


Figura 11.17. Gráfico que muestra el área a la izquierda de $x = 12$ equivalente a la de la izquierda de $z = 0,5$

11.5. APLICACIONES PRÁCTICAS DE LA CURVA NORMAL

Seguidamente, se harán algunas aplicaciones del uso de la curva normal a problemas concretos. Se trata de casos en los cuales se supone que los datos siguen una curva normal; luego, esa información se utilizará para derivar otra acerca de la distribución.

Ejemplo 3

Considere el caso de la distribución según talla de los 3600 recién nacidos, presentada en el cuadro 11.2. Se observa que un 2,08% mide entre 44,5 y 45,5 centímetros. Suponga ahora que la distribución es normal y $\mu = 50$ y $\sigma^2 = 5,88$, derive una estimación de ese porcentaje empleando la tabla de la normal estándar.

Sea x la variable estatura de los recién nacidos. Se desea calcular $P(44,5 \leq x \leq 45,5)$. Hay que estandarizar la normal:

$$\begin{aligned} P(44,5 \leq x \leq -1,86) &= P\left(\frac{44,5}{2,425} \leq \frac{x-50}{2,425} \leq \frac{45,5-50}{2,425}\right) \\ &= P(-2,27 \leq z \leq -1,86) \\ &= P(1,86 \leq z \leq 2,27) \\ &= P(z \leq 2,27) - P(z \leq 1,86) \\ &= N(2,27) - N(1,86) = 0,9884 - 0,9686 \\ &= 0,0198.^{5} \end{aligned}$$

Puede observarse que la proporción de recién nacidos, cuya medida está entre 44,5 y 45,5 cm, se estima usando la curva normal, en 1,98%, valor muy cercano al 2,08% que arroja la distribución de los 3600. -

5. El cálculo puede hacerse más directamente usando la función DIST.NORM.ESTAND del Excel. Se suministra al programa el valor $z = 1,86$ y se obtiene $N(-1,86) = 0,0314$, con el valor $N(-2,27) = 0,0116$. La diferencia $N(-1,86) - N(-2,27) = 0,0314 - 0,0116 = 0,0198$ coincide exactamente con el resultado al utilizar la tabla.

Suponga que interesa estimar la proporción de niños que al nacer miden menos de 45 cm. Este valor no puede obtenerse directamente de la distribución, porque 45 no es un límite real de clase. Se puede estimar con la normal estándar.

$$\begin{aligned} P(x \leq 45) &= P\left(\frac{x-50}{2,425} \leq \frac{45-502}{}2,425\right) \\ &= (z \leq -2,06) \\ &= 1 - P(z \leq 2,06) \\ &= 1 - N(2,06) \\ &= 1 - 0,9803 \\ &= 0,0197. \end{aligned}$$

Este resultado indica que, aproximadamente, un 2% de los niños miden, al nacer, menos de 45 cm.

◇◇

Ejemplo 4

Una universidad técnica realiza anualmente una prueba de admisión para decidir sobre el ingreso de los alumnos nuevos; se aplica a un número relativamente grande de estudiantes. El año anterior, según los datos de la oficina de admisiones, la nota promedio fue un 66 y la desviación estándar 16.

a) Suponga que este año la Comisión de Mejoramiento Académico propone recibir solamente a los estudiantes que obtengan una nota superior a 50 puntos. Si los resultados de la prueba son bastante estables a través del tiempo, ¿qué proporción de los aspirantes serían admitidos si se acepta la propuesta de la comisión?

En este caso, x es la nota en la prueba; $\mu = 66$ y $\sigma = 16$. Se quiere $P(x > 50)$ y se estandariza de la siguiente forma:

$$P(x > 50) = P\left(\frac{x-66}{16} > \frac{50-66}{16}\right) = P(z > -1) = 1 - P(z \leq -1) = N(1).$$

Al buscar en la tabla se encuentra que $N(1) = 0,8413$. Esto significa que si mantienen las mismas condiciones del año anterior, aproximadamente un 84% de los aspirantes será admitido si se aplica la propuesta de la comisión.

b) Si la universidad decide dar una beca especial de 50 000 colones mensuales a quienes obtengan un puntaje de 95 o mayor, ¿cuántas becas cabría esperar que se otorgaran? Suponga que el número de aspirantes es $N = 6000$.

$$\begin{aligned} P(x \geq 95) &= P\left(\frac{x-66}{16} \geq \frac{95-66}{16}\right) = P(z \geq 1,81) = 1 - (P(z \leq 1,81)) \\ &= 1 - N(1,81) = 1 - 0,9649 = 0,0351. \end{aligned}$$

Aproximadamente un 3,5% de los candidatos obtendría una nota de 95 o más. Si hay 6000, resultarían 5000 · 0,0351= 211 aspirantes con nota de 95 puntos o superior y, por lo tanto, la universidad debería otorgar 211 becas.

Ejemplo 5

En un sector de la industria de productos de cuero, los obreros ganan un sueldo semanal básico fijado por ley, más un incentivo por rendimiento (número de unidades producidas). Actualmente, el salario promedio mensual de los obreros es de 240 mil colones, con una desviación estándar de 7 mil. En una negociación salarial, el sindicato de obreros solicita a las empresas que concedan un aumento mensual, de 20 mil colones, al 15% de los trabajadores con salarios más bajos y un aumento de 10 mil al resto de los empleados, aunque está de acuerdo con que al 5%, con mayores salarios, se les dé solo 5 mil colones adicionales.

Para decidir sobre este planteamiento, un asesor de las empresas quiere determinar, aproximadamente, los montos salariales que marcan los límites de las categorías a los cuales se les aplicarán los aumentos solicitados. Como no tiene la información detallada de todas las empresas del sector industrial, decide realizar los cálculos suponiendo que la variable salario mensual sigue una distribución relativamente normal. Este supuesto lo considera legítimo porque las diferencias en los salarios obedecen a desigualdades en el rendimiento de los obreros, el cual, a su vez depende en gran medida de sus habilidades manuales.

En primer término, si $x =$ salario mensual, $\mu = 240\,000$, $\sigma = 7000$ y $P(x \leq x_0) = 0,15$ (donde x_0 representa el monto que delimita los salarios más bajos), entonces hay un valor z_0 tal que $P(z \leq z_0) = 0,15$.

Este z_0 no puede buscarse directamente en la tabla, porque solo da las áreas para valores positivos de z. Entonces, se busca el valor de z que corresponde al complemento $1 - 0,15 = 0,85$ Lo cual se hace por simple inspección, en la tabla normal estándar. Al revisarla, se encuentra lo siguiente:

Z	0,03	0,04	0,05
1,0	0,8485	0,8508	0,8531

Esta información señala que una buena aproximación al valor buscado de z es 1,04. Una más precisa, siempre usando la tabla, se logra interpolando linealmente el valor de z, con ayuda de la información para 1,03. Si se realiza este ejercicio, el valor de z resultante es 1,0365. Todo este trabajo puede evitarse si se usa la función DIST.NORM.ESTAND.INV del Excel, si se le suministra el valor 0,15 devuelve −1,0364 para z.

Conociendo que $P(z \leq -1,0364) = 0,15$, el paso siguiente es despejar el valor de x a partir de la fórmula básica de estandarización.

Si $z = \dfrac{x-\mu}{\sigma}$, puede despejarse x y calcular su valor mediante la siguiente expresión:

$$x = \mu + z\sigma.$$

Al aplicarla, con el valor de z conseguido y los otros datos conocidos se obtiene:

$$x = 240\,000 + (-1,0364)\cdot 7000 = 240\,000 - 7255 = 167\,745.$$

O sea, que 167 745 colones es el monto que delimita el 15% de obreros con los salarios más bajos y, por ello, todos los que ganen esa suma o menos recibirán 20 000 colones de aumento.

Para decidir sobre la segunda propuesta del sindicato, se debe encontrar el salario a partir del cual se ubica el 5% de empleados que más gana. Entonces, se busca en la tabla normal estándar el valor de z hasta el cual se acumula el 95% del área total. La inspección señala $z = 1{,}645$, pues $P(z \geq 1{,}645) = 0{,}05$. Aplicando el mismo procedimiento anterior, se tiene que:

$$x = \mu + z\sigma = 240\,000 + 1{,}645 \cdot 7000 = 240\,000 + 11\,515 = 251\,515.$$

Este es el salario a partir del cual el empleado recibirá solo 5000 colones de aumento.

Por simple diferencia de los dos niveles anteriores, se observa que los trabajadores, con salarios entre 167 745 y 251 515 colones, tendrán derecho a un aumento de 10 000 colones.

La principal condición para resolver este tipo de problemas es una adecuada selección del valor z, correspondiente al área fijada, lo cual se realiza mediante el análisis de que si se fijan topes en niveles inferiores de la distribución, z será negativo; si se hacen en niveles superiores del área, entonces z será positivo. Note que el uso de las funciones correspondientes del Excel simplifica y aligera el cálculo de probabilidades, dado un z o el de un z conociendo el área.

EJERCICIOS DE AUTOEVALUACIÓN

1. Indique si cada una de las variables aleatorias que se le presentan es discreta o continua. Para las discretas cite 3 posibles eventos muestrales que pueden ocurrir y para las continuas mencione un intervalo de posibles valores.
   a) Número de materias matriculadas por estudiantes universitarios en el primer semestre de este año.
   b) Duración de las llamadas telefónicas internacionales registradas por el ICE en el mes de febrero del 2009.
   c) Estatura de los jugadores que han participado por Costa Rica en la eliminatoria para la Copa Mundial de Futbol del 2010.
   d) Años requeridos, por los estudiantes de Administración Educativa, para concluir la carrera.
   e) Extensión de las fincas ganaderas ubicada en el cantón de Santa Cruz de Guanacaste.
2. Usando la tabla normal estándar determine las siguientes probabilidades:
   a) $P(z \geq 1,26)$
   b) $P(-1,43 \leq z \leq -0,48)$
   c) $P(1,09 \leq z \leq 2,13)$
3. Se tiene una variable x distribuida normalmente, con $\mu = 7$ y $\sigma = 3$. Calcule las siguientes probabilidades:
   a) $P(x \leq 10)$
   b) $P(x \geq 2,5)$
   c) $P(3 \leq x \leq 8)$
4. En un estudio se analizó la variable y, cociente de inteligencia (CI) de estudiantes de secundaria de un país, se usó la siguiente curva normal para representar su distribución.

$$f(y) = \frac{1}{12\sqrt{2\pi}} e^{-\left(\frac{y-90}{12}\right)^2}.$$

   a) ¿Cuál es el cociente de inteligencia promedio de los estudiantes de secundaria?
   b) ¿Cuál es la desviación estándar?

c) ¿Qué proporción de los estudiantes tiene un CI menor de 90 puntos?

d) ¿Qué proporción de los estudiantes tiene un CI mayor de 114 puntos?

5. En un colegio privado se utiliza una prueba de admisión para decidir el ingreso de los estudiantes y se sabe que el promedio es de 84 puntos, con una desviación estándar de 12 puntos. El puntaje máximo de la prueba son 120 puntos.

   a) Si este año se estableció que no podría ingresar el 20% de aspirantes con notas más bajas, ¿cuál fue la nota mínima de ingreso?

   b) Si se otorga beca al 10% de los mejores aspirantes, ¿a partir de qué nota se obtuvo ese derecho este año?

   c) Cuál fue el supuesto fundamental que hizo para realizar los cálculos pedidos en los puntos a) y b). Escríbalo.

6. Se aplica una escala de actitudes religiosas a mujeres residentes en áreas urbanas del país, la cual tiene un promedio de 120 puntos y una desviación estándar de 30, y sigue el modelo normal. En la escala, el mayor puntaje indica mayor religiosidad.

   a) ¿A partir de qué puntaje se localiza la cuarta parte de mujeres más religiosas?

   b) ¿Qué proporción de las mujeres presenta un índice de religiosidad de 200 puntos o mayor?

   c) ¿Qué puntaje marca el decil de las mujeres con menor religiosidad?

7. Un hombre viaja en su auto cada día al trabajo y encuentra que el tiempo que dura en el trayecto tiene $\mu = 35,5$ y $\sigma = 3,11$. Sale de su casa a las 7:20 a. m. y debe estar en la oficina a las 8:00 a. m.

   a) ¿Cuántas veces al año se espera que llegue tarde? Suponga que hace 240 viajes al año.

   b) Si quiere que su probabilidad de llegar tarde no sea mayor de 5%, ¿a qué hora debe salir?

RESPUESTA EJERCICIOS DE AUTOEVALUACIÓN

1. a) Variable discreta

   Evento muestral N. 1: estudiantes que se matricularon en **una** materia en el primer semestre de 1985.

   Evento muestral N. 2: estudiantes que se matricularon en menos de tres materias en el primer semestre de 1985.

   Evento muestral N. 3: estudiantes que se matricularon en más de cuatro materias en el primer semestre de 1985.

   b) Variable continua

   Intervalo: llamadas telefónicas internacionales registradas en el mes de febrero, con una duración de entre 5 y 10 minutos.

   c) Variable continua

   Intervalo: número de jugadores con estaturas entre 1,75 m y 1,90 m.

   d) Variable continua

   Intervalo: estudiantes que requirieron 6 y 8 años para concluir la carrera.

   e) Variable continua

   Intervalo: fincas ganaderas ubicadas en el cantón de Santa Cruz de Guanacaste, con una extensión de 100 a 499 hectáreas.

2. a) $$\begin{aligned} P(z \geq 1,26) &= 1 - P(z \leq 1,26) \\ &= 1 - 0,8962 \\ &= 0,1038. \end{aligned}$$

   b) $$\begin{aligned} P(z \geq -0,28) &= P(z \leq 0,28) \\ &= 0,6103. \end{aligned}$$

   c) $$\begin{aligned} P(-1,43 \leq z \leq -0,48) &= P(0,48 \leq z \leq 1,43) \\ &= P(z \leq 1,43) - P(z \leq 0,48) \\ &= 0,9236 - 0,6844 \\ &= 0,2392. \end{aligned}$$

   d) $P(z \leq 2,03) = 0,9788.$

e) $P(1,09 \leq z \leq 2;13) = P(z \leq 2,13) - P(z \leq 1,09)$
$= 0,9834 - 0,8621$
$= 0,1213.$

f) $P(z \geq -1,28) = P(z \leq 1,28)$
$= 08997.$

3. a) $P(x \leq 10) = P\left(\frac{x-\mu}{\sigma} \leq \frac{10-7}{3}\right)$
$= P(x \leq 1) = 0,8413.$

b) $P(x \geq 2,5) = P\left(\frac{x-\mu}{\sigma} \geq \frac{2,5-7}{3}\right) = P(z \geq -1,5)$
$= P(z \leq 1,5) = 0,9332.$

c) $P(3 \leq x \leq 8) = P\left(\frac{3-7}{3} \leq \frac{x-\mu}{\sigma} \leq \frac{8-7}{3}\right)$
$= P(-1,33 \leq z \leq 0,33)$
$= P(0 \leq z \leq 0,33) + P(0 \leq z \leq 1,33)$
$= 0,1293 + 0,4082$
$= 0,5375.$

4. a) El cociente de inteligencia (CI) promedio es igual a 90.

b) La desviación estándar es igual a 12.

c) $P(x < 90) = P\left(\frac{x-\mu}{\sigma} < \frac{90-90}{12}\right) = P(z < 0) = 0,5.$

El 50% del grupo de estudiantes de secundaria tiene un cociente de inteligencia menor a 90. La respuesta puede darse sin necesidad de hacer cálculos, si se nota que CI = 90 es igual μ, y que la curva normal es simétrica con respecto a la media.

d) $P(x > 114) = P\left(\frac{x-\mu}{\sigma} > \frac{114-90}{12}\right) = P(z > 2)$
$= 1 - P(z2)$
$= 1 - 0,9772 = 0,0228.$

El 2,3% del grupo de estudiantes de secundaria tiene un cociente de inteligencia mayor a 114.

5. $\mu = 84, \sigma = 12.$

   a) $p = 0,2$ (área a la derecha de la nota mínima de ingreso).

   Se debe encontrar el valor de z de la normal estándar hasta el cual se acumula esa área de 0,2. Se sabe, por la propiedad de simetría, que es igual, pero con signo negativo, al valor en el que se acumula el 80% del área bajo la curva. Por simple inspección, en la tabla normal estándar, se encuentra que al valor $z = 0,84$ le corresponde un área acumulada de 0,7995. Esto significa que, al valor $z = -0,84$, le corresponde el complemento 0,2005, cifra muy cercana al valor fijado de 20%. El paso siguiente es una simple operación algebraica a partir de la fórmula de estandarización.

   Se toma $z = \frac{x - \mu}{\sigma}$, se sustituyen valores conocidos y luego se despeja

$$x - 0,84 = \frac{x - 84}{12}$$
$$x = -0,84 \cdot 12 + 84$$
$$x = 73,92$$

   donde $73,92 \approx 74$ puntos es la nota mínima de ingreso.

   b) $p = 0,1$ (área a la derecha de la nota mínima para adquirir la beca). Para encontrar z, se procede de manera similar al caso anterior, encontrándose que z asume el valor de 1,28.

$$z = \frac{x - \mu}{\sigma}$$
$$1,28 = \frac{x - 84}{12}$$
$$x = 1,28 \cdot 12 + 84$$
$$x = 99,36$$

   donde 99,36 es la nota a partir de la cual se obtiene el derecho a la beca.

6. $\mu = 120, \ \sigma = 30.$

   a) $p = 0,25$ (área a la derecha del puntaje mínimo, donde se localiza la cuarta parte de las mujeres más religiosas).

   Para encontrar el valor de z, se procede de forma semejante al caso 6a), z toma el valor de 0,67. Luego se despeja x en la fórmula de estandarización:

$$0,67 = \frac{x - 120}{30}$$
$$x = 0,67 \cdot 30 + 120$$
$$x = 140,1$$

   donde $140,1 \approx 140$ es el puntaje a partir del cual se localiza la cuarta parte de mujeres más religiosas.

b) Para responder a esta pregunta, debe calcularse la probabilidad de que x sea mayor de 200. Si se realiza la estandarización, se encuentra

$$P(x200) = P(z2,67) = 1 - N(2,67) = 1 - 0,9962 = 0,0038.$$

Solo un 0,38% presenta un índice de religiosidad de 200 puntos o más. Esto significa que, si se seleccionan mujeres al azar, solo cerca de 4 de cada 1000 alcanzaría un nivel tan alto en la escala de religiosidad.

c) ¿Qué proporción de las mujeres presenta un índice de religiosidad de 200 puntos o mayor?

$p = 0,1$ (área a la izquierda del puntaje mínimo donde se localiza la décima parte de las mujeres menos religiosas).

Para hallar el valor asumido por z, se aplica un procedimiento similar al caso 5a); mediante el cual $x = -1,28$.

Luego, se procede a determinar el puntaje solicitado (x), empleando la fórmula de estandarización de la siguiente manera:

$$\begin{aligned}
z &= \frac{x-\mu}{\sigma} \\
-1,28 &= \frac{x-120}{30} \\
x &= -1,28 \cdot 30 + 120 \\
x &= 81,6
\end{aligned}$$

el valor de 81,6 marca el límite hasta el cual llega la décima parte de mujeres con menor religiosidad.

7. $\mu = 35,5,\ \sigma = 3,11$

$$\left.\begin{array}{l}\text{Hora de salida: 7:20} \\ \\ \text{Hora de llegada: 8:00}\end{array}\right\} \quad \text{Duración 40}$$

a) Probabilidad de que llegue tarde:

$$\begin{aligned}
P(x > 40) &= P\left(\frac{x-\mu}{\sigma} > \frac{40-35,5}{3,11}\right) = (z > 1,45) \\
&= P(z0) - P(0 \leq z \leq 1,45) \\
&= 0,5 - 0,4265 \\
&= 0,0735.
\end{aligned}$$

Número de llegadas tardías esperadas $= 240 \cdot 0,0735 = 17,64$.

Se espera que el hombre llegue 17,64 veces tarde al año.

b) Se quiere que:

$$P(x > x_0) = 0,5$$
$$P\left(\frac{x-\mu}{\sigma} > \frac{x_0 - 35,5}{3,11}\right) = 0,05$$
$$P(z > 1,64) = 0,05.$$

Igualando términos queda:

$$\frac{x_0 - 35,5}{3,11} = 1,64$$
$$x_0 = 1,64(3,11) + 35,5$$
$$x_0 = 40,6 \text{ minutos}$$

el hombre debe salir a las 7 horas con 19,4 minutos para que la probabilidad de llegar tarde no sea mayor de 5%.