Tema 13. Distribuciones de probabilidad

1. Distribución de probabilidad

Una distribución de probabilidad es un modelo teórico que trata de explicar el comportamiento de un fenómeno real. Actúa como una función que asigna a cada suceso, cuantificado mediante una variable aleatoria, la probabilidad correspondiente.

Una variable aleatoria asocia a cada suceso del espacio muestral un número real. Así, por ejemplo, en el experimento consistente en lanzar dos dados numerados del 1 al 6 y hallar su suma (*X*), la variable *X* puede tomar cualquier valor entero entre 2 y 12.
Las probabilidades de esos valores pueden calcularse mediante la regla de Laplace, teniendo en cuenta que los casos posibles son 36 y que los casos favorables se contabilizan en las diagonales de la tabla de sumas adjunta.

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Por ello, la distribución de probabilidad de *X* se resume en la siguiente tabla:

X (suma)	2	3	4	5	6	7	8	9	10	11	12
$P(X=i)$ $2 \le i \le 12$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

→ Una variable aleatoria puede ser discreta o continua.
Se llama discreta cuando solo puede tomar ciertos valores aislados (generalmente un número finito de valores).
Es continua cuando puede tomar todos los valores de un intervalo.

Ejemplos:
a) La suma de los resultados obtenidos al lanzar dos dados es una variable aleatoria discreta.

b) El número de veces que hay que lanzar uno de esos dados hasta que salga el número 6 es también discreta.

c) El tiempo que una persona tiene que esperar al autobús es una variable continua. Igualmente, la estatura de esa persona o su peso son variables continuas.

1.1. Función de probabilidad (variable discreta)

Es la que asigna a cada uno de los valores de la variable aleatoria discreta su probabilidad correspondiente.
Puede definirse como sigue:

$$f(x_i) = P\left(X = x_i\right) = p_i .$$

Al tratarse de de una función de probabilidad debe cumplir que:

$$0 \le f(x_i) \le 1 .$$

$$\sum_{i=1}^{n} p_i = 1$$ → la suma de las probabilidades de todos los sucesos es igual a 1.

Si *x* no es alguno de los valores de la variable aleatoria, $f(x) = 0$.

Ejemplo:
Para la suma de los resultados de dos dados:

$$f(2)=P\left(X=2\right)=\frac{1}{36};\ f(3)=P\left(X=3\right)=\frac{2}{36},\ldots,\ f(7)=P\left(X=7\right)=\frac{6}{36};$$

$$f(1)=P\left(X=1\right)=0;\ f(7,3)=0;\ f(15)=0.$$

También puede verse que

$$\sum_{i=1}^{n}P\left(X=i\right)=\frac{1}{36}+\frac{2}{36}+\frac{3}{36}+\ldots+\frac{6}{36}+\ldots+\frac{1}{36}=\frac{36}{36}=1.$$

1.2. Función de distribución

A partir de la distribución de probabilidad de la variable X se define la función de distribución, $F(x)$, de dicha variable como sigue:

$$F(x)=P\left(X\le x\right)$$

A cada valor x, $F(x)$ le asigna la probabilidad de que la variable aleatoria tome valores menores o iguales que x. La función $F(x)$ acumula probabilidades: $F(x)=\sum_{x_i\le x}f\left(x_i\right)$.

Por tanto, si se conoce la función de distribución de una variable aleatoria, es posible determinar la probabilidad de que tome uno de sus valores, pues:

$$P\left(X=x_i\right)=F(x_i)-F(x_{i-1}).$$

Igualmente: $P\left(X>x_i\right)=1-F(x_i)$.

Ejemplos:
a) Para la suma de los resultados de dos dados:

$$F(2)=P\left(X=2\right)=\frac{1}{36},\ F(3)=P\left(X=2\right)+P\left(X=3\right)=\frac{1}{36}+\frac{2}{36}=\frac{3}{36},\ldots$$

$$F(12)=P\left(X\le 12\right)=1.$$

b) Si se considera el experimento aleatorio consistente en lanzar 4 monedas y contar el número de caras que se obtienen, se tiene:
→ La variable aleatoria X que da el número de caras obtenido toma los valores: x_i = 0, 1, 2, 3, 4.
→ Las probabilidades de cada suceso (función de probabilidad) es:

$$f(0)=P\left(X=0\right)=\frac{1}{16};\ f(1)=P\left(X=1\right)=\frac{4}{16};\ f(2)=P\left(X=2\right)=\frac{6}{16};$$

$$f(3)=P\left(X=3\right)=\frac{4}{16};\ f(4)=P\left(X=4\right)=\frac{1}{16}.$$

1.3. Media y varianza de una distribución de probabilidad discreta

Los parámetros estadísticos más usuales se calculan como sigue.

→ La media de una distribución es un valor central que indica la cantidad que correspondería a cada suceso en una repartición igualitaria.
Si una variable aleatoria toma los valores x_1, x_2, …, x_n, con probabilidades p_1, p_2, …, p_n, la media, que suele denotarse por la letra griega μ (mu), se calcula mediante la expresión:

$$\mu=x_1\cdot p_1+x_2\cdot p_2+\ldots+x_n\cdot p_n=\sum_{i=1}^{n}x_i\cdot p_i\ .$$

Ejemplos:
a) La media de la variable aleatoria que mide la suma de las puntuaciones de dos dados es:

$$\mu = 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + \ldots + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} = \frac{252}{36} = 7\,.$$

b) La media de la distribución de probabilidad del número de caras que se obtienen al lanzar cuatro monedas es:

X = nº de caras	0	1	2	3	4
$P(X = x_i)$	1/16	4/16	6/16	4/16	1/16

$$\mu = 0\cdot\frac{1}{16} + 1\cdot\frac{4}{16} + 2\cdot\frac{6}{16} + 3\cdot\frac{4}{16} + 4\cdot\frac{1}{16} = \frac{32}{16} = 2\,.$$

→ La media μ de una variable aleatoria también se llama valor esperado o esperanza matemática de dicha variable y se suele escribir $\mu = E(X)$. El valor esperado de un juego es el baremo que se establece para determinar su equidad: si su esperanza matemática es cero, $E(X) = \mu = 0$, no existe ventaja para ninguno de los jugadores participantes.

Ejemplo:
En los exámenes de tipo test (con respuesta múltiple) suele penalizarse el error, pues un examinando puede tener un número importante de aciertos por el mero hecho de contestar al azar. Así, en el caso de preguntas con tres respuestas posibles de las que solo una es correcta, cada acierto debe ser contrarrestado por 2 errores. En efecto:
→ Si un examinando no sabe nada, su puntuación justa (su calificación esperada) debe ser 0. Como la probabilidad de acierto es 1/3 y la de fallo 2/3, si por cada acierto suma 1 punto, por cada error debe restar x, ¿pero cuanto debe ser x?
La esperanza matemática es:

$$E(X) = 1\cdot\frac{1}{3} + x\cdot\frac{2}{3} = 0 \Rightarrow 1 + 2x = 0 \Rightarrow x = -\frac{1}{2}$$ → Por cada error hay que restar 0,5 puntos

→ La varianza es una medida de la dispersión (de la desigualdad) de los valores de la variable aleatoria respecto a la media. Se suele designar por σ^2.
→ La desviación típica, σ, es la raíz cuadrada positiva de la varianza.
Sus valores vienen dados por las expresiones:

$$\sigma^2 = (x_1-\mu)^2\cdot p_1 + (x_2-\mu)^2\cdot p_2 + \ldots + (x_n-\mu)^2\cdot p_n = \sum_{i=1}^{n}(x_i-\mu)^2\cdot p_i = \sum_{i=1}^{n} x_i^2\cdot p_i - \mu^2\,.$$

$$\sigma = \sqrt{\sum_{i=1}^{n}(x_i-\mu)^2\cdot p_i} = \sqrt{\sum_{i=1}^{n} x_i^2\cdot p_i - \mu^2}\,.$$

La desviación típica se utiliza con mayor frecuencia pues, al estar expresada en las mismas unidades que la variable X, permite establecer más claramente las comparaciones.

Ejemplos:
a) La varianza de la distribución de suma de las puntuaciones de dos dados es:

$$\sigma^2 = 2^2\cdot\frac{1}{36} + 3^2\cdot\frac{2}{36} + 4^2\cdot\frac{3}{36} + 5^2\cdot\frac{4}{36} + \ldots + 12^2\cdot\frac{1}{36} - 7^2 = 5{,}833 \Rightarrow \sigma = \sqrt{5{,}833} = 2{,}415\,.$$

b) La varianza de la distribución de suma del número de caras obtenidas al lanzar 4 monedas es:

$$\sigma^2 = 0^2\cdot\frac{1}{16} + 1^2\cdot\frac{4}{16} + 2^2\cdot\frac{6}{16} + 3^2\cdot\frac{4}{16} + 4^2\cdot\frac{1}{16} - 2^2 = 1$$; su desviación típica es: $\sigma = \sqrt{1} = 1$.

2. Distribución binomial

Es una de las distribuciones de probabilidad más utilizadas en la práctica estadística. Se emplea cuando el fenómeno de estudio queda determinado por dos sucesos complementarios: si/no; hombre/mujer; nacional/extranjero; trabajador en activo/parado; … En general, esas dos situaciones pueden considerarse resultados de un experimento aleatorio y a los sucesos contrarios, sin que indique valoración alguna, suelen llamárseles éxito y fracaso.

Las características básicas de una distribución binomial son:

- Cada prueba del experimento aleatorio presenta dos únicas opciones, que puede designarse como éxito (*E*) y fracaso (*F*).
- Se realizan *n* ensayos del experimento, independientes unos de otros e idénticos.
- La probabilidad de éxito es constante a lo largo de las *n* pruebas: $P(E)=p$.
- La probabilidad de fracaso también es constante: $P(F)=q=1-p$.

→ Una distribución de estas características también recibe el nombre de pruebas de Bernoulli. (Jacob Bernoulli: suizo, 1654–1705. Es interesante "entrar" en Familia Bernoulli).

La variable aleatoria *X*, cuenta el número *r* de éxitos en las *n* pruebas: *r* = 0, 1, …, *n*. Por tanto, los valores que puede tomar *X* son: 0, 1, 2, ..., *n*.
La distribución binomial queda determinada por los parámetros *n* y *p* (número de veces que se realiza el experimento y probabilidad de éxito en cada prueba).
Se indica simbólicamente por *B*(*n*, *p*).

Ejemplos:
a) La variable que cuenta el número de caras obtenidas al lanzar 8 monedas es una binomial de parámetros *n* = 8 y *p* = (probabilidad de cara) = 1/2 = 0,5. Se denota por *B*(8, 0,5).
b) Si en una determinada región, la tasa de paro entre su población activa es del 12%, si se pregunta a 10 personas de esa población, elegidos al azar, por su situación laboral, el número de parados viene descrito por la binomial de parámetros *n* = 10 y *p* = 0,12: *B*(10, 0,12).

2.1. Probabilidad de *r* éxitos

La función de probabilidad que mide el número *r* de éxitos cuando una prueba de carácter binomial se realiza *n* veces, *B*(*n*, *p*), viene dada por:

$$P(X=r)=\binom{n}{r}p^r\cdot q^{n-r}, \quad r=0, 1, 2,..., n.$$

Ejemplos:
a) Para la *B*(8, 0,5), que cuenta el número de caras obtenidas al lanzar 8 monedas, se tiene:

$$P(\text{obtener 3 caras}) = P(X=3)=\binom{8}{3}\cdot 0,5^3\cdot 0,5^5=\frac{8\cdot 7\cdot 6}{3\cdot 2}\cdot 0,5^8=56\cdot\frac{1}{2^8}=\frac{56}{256}.$$

Análogamente:

$$P(X=6)=\binom{8}{6}\cdot 0,5^6\cdot 0,5^2=\frac{8\cdot 7}{2}\cdot\frac{1}{2^8}=\frac{28}{256};\ P(X=8)=\binom{8}{8}\cdot 0,5^0\cdot 0,5^8=1\cdot 0,5^8=\frac{1}{256}.$$

Recuerda que $\binom{n}{r}=\frac{n!}{r!(n-r)!}\to\binom{8}{3}=\frac{8!}{3!(8-3)!}=\frac{8\cdot 7\cdot 6\cdot 5!}{3!\cdot 5!}=\frac{8\cdot 7\cdot 6}{3\cdot 2\cdot 1}=56$.

- Para este mismo ejemplo, si se plantea la probabilidad de que al menos salgan dos caras, habrá que calcular: $P(X=2)+P(X=3)+...+P(X=8)=1-P(X=0)-P(X=1)$.

b) Para la binomial B(10, 0,12) que estudia el número de parados entre 10 personas elegidas al azar cuando la tasa de paro es del 12%, se tendrá:

P(2 parados) = $P(X=2)=\binom{10}{2}\cdot 0,12^2\cdot 0,88^8=\frac{10\cdot 9}{2}\cdot 0,12^2\cdot 0,88^8=45\cdot 0,0051787...=0,233$.

- Para este mismo ejemplo, la probabilidad de que haya menos de 3 parados es:

$$P(X<3)=P(X=0)+P(X=1)+P(X=2)=$$
$$=\binom{10}{0}\cdot 0,12^0\cdot 0,88^{10}+\binom{10}{1}\cdot 0,12^1\cdot 0,88^9+\binom{10}{2}\cdot 0,12^2\cdot 0,88^8=$$
$$=1\cdot 0,88^{10}+10\cdot 0,12\cdot 0,88^9+45\cdot 0,12^2\cdot 0,88^8=$$
$$= 0,278500976 + 0,379774058 + 0,23043172 = 0,891318206.$$

Con esto, la probabilidad de que haya 3 o más parados, $P(X\geq 3)$, será:

$$P(X\geq 3)=1-P(X<3)=1-0,891318206=0,108681794.$$

(En todos los casos es suficiente el redondeo a diezmilésimas: $P(X\geq 3)=0,1087$).

Observación sobre la tabla binomial. Hasta hace poco tiempo era "obligado" incluir en los libros de texto una tabla binomial que facilitaba el cálculo de esas probabilidades. En dicha tabla se dan los resultados para algunos valores de n y p. Pienso que la generalización de calculadoras y ordenadores hace innecesario el empleo de la tabla binomial. Por ejemplo, tecleando en Excel: =DISTR.BINOM(2;10;0,12;0) se obtiene 0,23304317, que es el valor correspondiente a los 2 parados del ejemplo visto anteriormente. Si se teclea =DISTR.BINOM(2;10;0,12;2) da la probabilidad acumulada correspondiente a 0, 1 o 2 parados, que es 0,891318206.

2.2. Media y varianza de la binomial *B*(*n*, *p*)

La media y varianza de la distribución $B(n, p)$ se obtiene a partir de sus parámetros, siendo:

→ Media: $\mu=n\cdot p$.

→ Varianza: $\sigma^2=n\cdot p\cdot q$.

En consecuencia, la desviación típica vale $\sigma=\sqrt{npq}$.

Ejemplos:

a) La media y desviación típica de la binomial B(8, 0,5) valen:

μ = 8 · 0,5 = 4; $\sigma=\sqrt{8\cdot 0,5\cdot 0,5}=\sqrt{2}$.

Por tanto, cuando se tiran 8 monedas cabe esperar 4 caras.

b) Si se considera la binomial B(50, 0,12), que puede servir para determinar el número de parados en muestras de tamaño n = 50, se tiene:

Media: μ = 50 · 0,12 = 6 parados; desviación típica: $\sigma=\sqrt{50\cdot 0,12\cdot 0,88}=\sqrt{5,28}\approx 2,3$.

→ Para valores grandes de n, la probabilidad de cada uno de los posibles sucesos (de un número r de éxitos) es muy pequeña, sobre todo para valores de X alejados de la media. Así, por ejemplo, para la binomial B(50, 0,12), pueden darse las siguientes probabilidades:

$P(X=2)$ = 0,03816514; $P(X=8)=0,10754701$; $P(X=12)$ = 0,0084088;

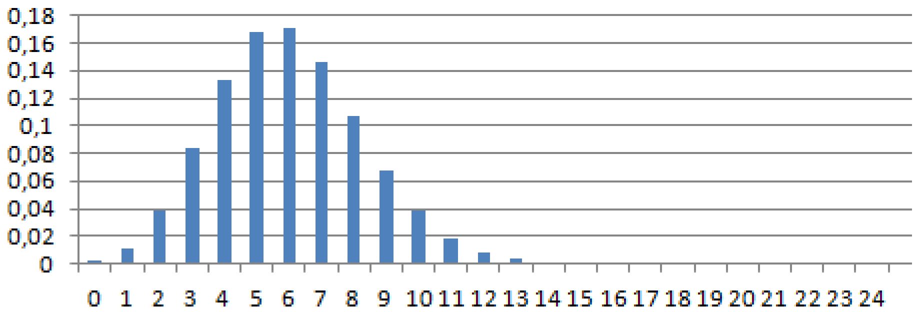
$P(X=15)$ = 0,00039533.

Estos resultados se han calculado con ayuda del ordenador (Excel).

Con detalle, la probabilidad de que haya 6 parados (que sería el número esperado, la media) es:

$$P(X=6)=\binom{50}{6}\cdot 0,12^{6}\cdot 0,88^{44}=15890700\cdot\left(2,985984\cdot 10^{-6}\right)\cdot\left(3,607759\cdot 10^{-3}\right)=0,171185935\,.$$

En el gráfico de barras que sigue se dan las probabilidades de cada uno de los sucesos asociados a la *B*(50, 0,12). En el eje *OX* se indican los valores de *X*; en *OY* sus probabilidades.



Observación: Al final de este tema se abordará un método más rápido para obtener un valor aproximado de estas probabilidades.

3. Algunos ejercicios de tipo binomial

Ejercicio 1. Si se tira 12 veces una moneda, ¿cuál es la probabilidad de obtener exactamente 8 caras?
Solución:
El experimento es de tipo binomial, con $n = 12$ y $p = 0,5$: *B*(12, 0,5).

$$P(X=8)=\binom{12}{8}\cdot 0,5^{8}\cdot 0,5^{4}=\frac{12\cdot 11\cdot 10\cdot 9}{4\cdot 3\cdot 2\cdot 1}0,5^{12}=495\cdot 0,5^{2}=0,12085\,.$$

Ejercicio 2. Un examen de tipo test consta de 6 preguntas con 4 posibles respuestas cada una, de las que solo una de ellas es correcta. Si se responde al azar, ¿cuál es probabilidad de que acertar 4 o más preguntas?
Solución:
Es un experimento binomial *B*(6, 0,25). Por tanto:

$$P(X\geq 4)=P(X=4)+P(X=5)+P(X=6)=\binom{6}{4}\cdot 0,25^{4}\cdot 0,75^{2}+\binom{6}{5}\cdot 0,25^{5}\cdot 0,75+\binom{6}{6}\cdot 0,25^{6}$$

$$=\frac{6\cdot 5}{2}0,25^{4}30,75^{2}+6\cdot 0,25^{5}\cdot 0,75+0,25^{6}=0,03296+0,00439+0,00024=0,03759\,.$$

Ejercicio 3. En una ciudad, el 15% de sus ciudadanos tocan algún instrumento musical. Si se eligen 8 personas al azar, ¿cuál es la probabilidad de que al menos 2 de ellas toquen algún instrumento?
Solución:
Se trata de una binomial *B*(8, 0,15). Por tanto:

$$P(X\geq 2)=1-P(X=0)-P(X=1)=$$

$$=1-\binom{8}{0}\cdot 0,15^{0}\cdot 0,85^{8}-\binom{8}{1}\cdot 0,15^{1}\cdot 0,85^{7}=1-0,27249-0,38469=0,34282\,.$$

4. Distribuciones de probabilidad continuas

Una variable estadística se llama continua cuando puede tomar todos los valores de un intervalo. Así, por ejemplo, son variables estadísticas continuas, las estaturas y pesos de los individuos, los tiempos de espera de un autobús, el tamaño de una determinada variedad de manzanas, etc.

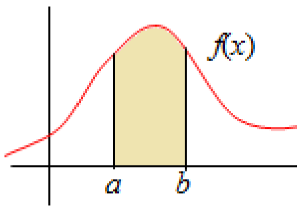
Para estas distribuciones, la probabilidad de un valor concreto es 0, pues el número de casos posibles es infinito. Por ejemplo, la probabilidad de que una persona mida exactamente 172,12345678910… cm es 0; altura tan improbable como que mida exactamente 172,000…cm. En cambio, la probabilidad de que una persona mida entre 171,5 cm y 172,5 cm sí podrá calcularse.
Esto es, si *X* es la variable que mide la estatura de una persona, se tendrá:

$$P(X = 172{,}12345...) = 0;\ \ P(X = 172{,}000...) = 0.$$

En cambio, $P(171{,}5 < X < 172{,}5) = ?$, valor que dependerá de la población de estudio.

4.1. Función de probabilidad

Esta función, que también se llama función de densidad, es la permite el cálculo de probabilidad para distribuciones continuas. La probabilidad de que la variable tome valores en un intervalo, $P(a < X < b)$, será el área del recinto plano limitado por la función de densidad y el eje *OX* cuando $X \in (a,\ b)$.



Esto es, si la función de densidad es $f(x)$,

$$P(a < X < b) = \int_a^b f(x)dx.$$

En general, para que $f(x)$ sea una función de densidad debe cumplir:

1) $f(x) \geq 0$ para todo valor *x* de su dominio: para todos los valores que pueda tomar la variable aleatoria.

2) El área limitada por la curva de $f(x)$ y el eje de abscisas, vale 1: $\int_{-\infty}^{+\infty} f(x)dx = 1$.

3) La probabilidad de que la variable tome valores en un intervalo [*a*, *b*], es:

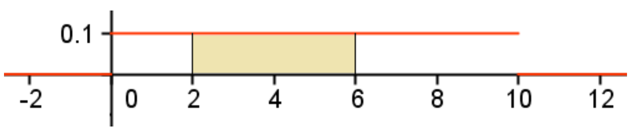
$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

Observación: La probabilidad de que *X* tome el valor *a*, $P(X = a) = \int_a^a f(x)dx = 0$, que se corresponde con el área de un rectángulo de base 0. Por tanto, son iguales las probabilidades:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Ejemplo:

La función $f(x) = \begin{cases} 0{,}1, & \text{si } 0 \leq x \leq 10 \\ 0, & \text{en otro caso} \end{cases}$, que puede describir el tiempo de espera, en minutos, hasta que llega un tren de cercanías, es una función de densidad.



La probabilidad de que haya que esperar entre 2 y 6 minutos es el área del rectángulo sombreado en la figura adjunta, que vale 4 · 0,1 = 0,4.

Naturalmente coincide con el valor $\int_2^6 0{,}1dx = 0{,}1x\big|_2^6 = 0{,}1 \cdot 6 - 0{,}1 \cdot 2 = 0{,}4$.

4.2. Función de distribución

La función de distribución, $F(x)$, asigna a cada valor *x* la probabilidad de que la variable *X* tome valores menores o iguales que *x*. Se define como sigue:

$$F(x)=P\left(X\le x\right)=\int_{-\infty}^{x} f(t)dt\,.$$

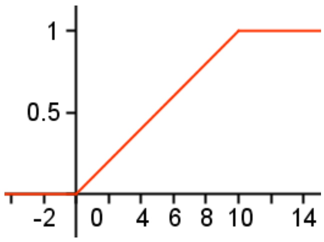
$\to$ $F(x)$ es una función acumulativa y creciente; toma valores comprendidos entre 0 y 1.

Ejemplo:

La función de distribución asociada a $f(x)=\begin{cases}0{,}1, & \text{si } 0\le x\le 10\\ 0, & \text{en otro caso}\end{cases}$ es

$$F\left(x\right)=\int_{0}^{x} 0{,}1dt=0{,}1t\Big|_{0}^{x}=0{,}1x\,,\ x\in[0,\,10].$$

O de manera completa: $F(x)=\begin{cases}0, & \text{si } x<0\\ 0{,}1x, & \text{si } 0\le x\le 10\\ 1, & \text{si } x>10\end{cases}$.



Su gráfica es la adjunta.
Con esto, por ejemplo:

$$F(6)=P\left(X\le 6\right)=0{,}6\,;\ F(4)=P\left(X\le 4\right)=0{,}4\Rightarrow$$
$$\Rightarrow P\left(4\le X\le 6\right)=F(6)-F(4)=0{,}6-0{,}4=0{,}2\,.$$

4.3. Media y varianza

Si una distribución de variable continua *X* tiene función de densidad $f(x)$, su media y varianza se determinan como sigue:
Media:

$$\mu=\int_{-\infty}^{+\infty} x\cdot f(x)dx\ \to \text{Si el dominio de } f(x) \text{ es } [a,\,b],\ \mu=\int_{a}^{b} x\cdot f(x)dx\,.$$

Varianza:

$$\sigma^2=\int_{-\infty}^{+\infty}\left(x-\mu\right)^2\cdot f(x)dx=\int_{-\infty}^{+\infty} x^2\cdot f(x)dx-\mu^2\,.$$

La desviación típica es σ, la raíz cuadrada de la varianza.

$\to$ Si el dominio de $f(x)$ es $[a, b]$: $\mu=\int_{a}^{b} x\cdot f(x)dx\,;\ \sigma^2=\int_{a}^{b} x^2\cdot f(x)dx-\mu^2\,.$

Ejemplo:

Para la distribución continua definida por $f(x)=\begin{cases}0{,}1, & \text{si } 0\le x\le 10\\ 0, & \text{en otro caso}\end{cases}$ se tiene:

$$\mu=\int_{0}^{10} 0{,}1xdx=0{,}1\cdot\frac{x^2}{2}\Bigg|_{0}^{10}=5\,;$$

$$\sigma^2=\int_{0}^{10} x^2\cdot 0{,}1dx-5^2=0{,}1\frac{x^3}{3}\Bigg|_{0}^{10}-25=\frac{25}{3}\Rightarrow\sigma=\sqrt{\frac{25}{3}}=\frac{5}{\sqrt{3}}\approx 2{,}89\,.$$

5. Distribución de probabilidad normal

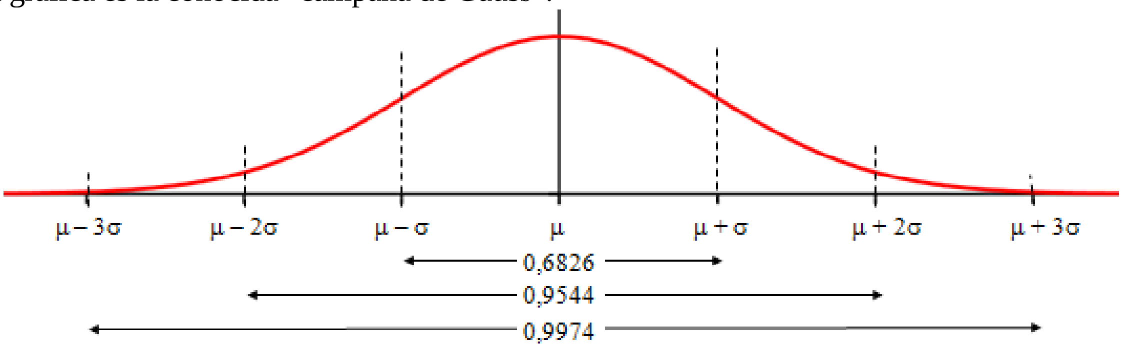
Es una distribución de probabilidad continua, asociada (teóricamente) a multitud de fenómenos naturales y cotidianos (cociente intelectual, talla o peso de las personas; tamaño de los frutos de cualquier tipo de árbol…), que se caracterizan porque la mayoría de los resultados tienden a agruparse en torno a su media.

Una variable con distribución normal queda totalmente definida por su media μ y por su desviación típica σ. Se denota como $N(\mu, \sigma)$.

→ La expresión analítica de la función de densidad de la distribución normal es

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

Su gráfica es la conocida “campana de Gauss”.



Esta función cumple las siguientes propiedades:

- Está definida para todo número real, es decir, en el intervalo (–∞, +∞): la variable puede tomar cualquier valor; siendo $f(x)>0$ para todo x.
- El área por debajo de la curva vale 1. Esto es: $\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=1$.
- Es simétrica respecto a su media μ.
- El eje de abscisas es una asíntota de la curva.
- Aunque la variable puede tomar cualquier valor entre –∞ y +∞, la probabilidad de que tome valores alejados de la media es prácticamente nula, pues se cumple que: el área delimitada por la curva y el eje OX entre μ – σ y μ + σ es 0,6826; entre μ ± 2σ es de 0,9544 y entre μ ± 3σ es de 0,9974. De hecho, a los valores que están a una distancia superior a 3,5σ de la media se les asigna una probabilidad 0.

Ejemplo:

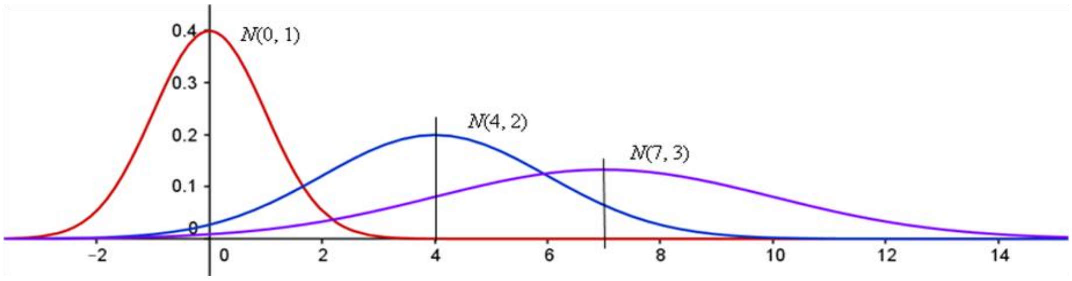
Supongamos que la estatura de los jóvenes de 20 años de una determinada región es una variable estadística X, que se distribuye de acuerdo con la normal de media μ = 175 cm y deviación típica σ = 9 cm: $N(175, 9)$. Entonces, puede asegurase, con las probabilidades que se indican, que:

P(de que un joven mida menos de 175 cm) = $P(X < 175) = 0{,}5$ → la mitad de los jóvenes tiene una estatura por debajo de la media; la otra mitad medirá más de 175 cm.

$P(\mu-\sigma<X<\mu+\sigma) = P(175-9<X<175+9) = P(166<X<184) = 0{,}6826$ →
(El 68,26% de los jóvenes de esa región tiene una estatura comprendida entre 166 y 184 cm).

$P(\mu-2\sigma<X<\mu+2\sigma) = P(175-18<X<175+18) = P(157<X<193) = 0{,}9544$.
(El 95,44% de los jóvenes de esa región tiene una estatura comprendida entre 157 y 193 cm).

→ La variación de la media y de la desviación típica originan cambios en la curva, desplazándose a izquierda o derecha o haciéndose más esbelta o más baja, como puede verse en la figura.



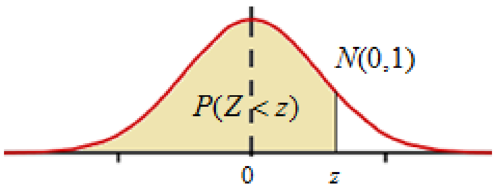
Recuérdese que la desviación típica es una medida de la dispersión de los elementos de una población. Una desviación típica más grande significa que los datos son más heterogéneos; por eso las curvas normales con mayor desviación típica son más planas, que es un indicador de que los datos pueden estar más alejados de la media. (Cuando la igualdad entre los datos es grande, la desviación típica es pequeña; y al revés).

Ejemplo:
En una clase de 2º de bachillerato la variable edad tiene una desviación típica muy pequeña. Puede admitirse que se distribuya como una normal *N*(17, 0,5).
En esa misma clase, la variable estatura puede distribuirse con media 171 cm y desviación típica 11 cm: *N*(171, 11). Los valores de estatura son más heterogéneos que los de edad.

5.1. Distribución normal de media 0 y desviación típica 1: *N*(0, 1)

El comportamiento estadístico normal hace que puedan asignarse valores de probabilidad a cualquier suceso de la variable estudiada. Esto es, se puede saber (pues está tabulado) la probabilidad de que la variable tome valores comprendidos entre los extremos de un intervalo dado.
Lo que está tabulado es la función de distribución en el caso de la curva normal de media μ = 0 y desviación típica σ = 1, la normal *N*(0, 1).
La función de distribución, $F(x)$, da la superficie del recinto limitado por la función de densidad, curva $y = f(x)$ definida más arriba, y el eje *OX*, desde –∞ hasta un valor determinado *z*, esto es

$$P(Z < z) = F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx .$$



Esta área da la probabilidad de que la variable *Z*, tome valores menores que *z*.
(Cuando se trata de la *N*(0, 1), la variable aleatoria suele designarse por la letra *Z*).

5.2. Tabla normal estándar: *N*(0, 1)

La tabla normal *N*(0, 1) puede encontrarse fácilmente en internet. Basta con teclear "tabla normal estándar". Aquí se muestra una parte de ella. Habitualmente los valores de la tabla indican la probabilidad de que la variable *Z*, *N*(0, 1), tome valores entre –∞ y +*z*, $P(Z < z)$; los demás valores se obtienen teniendo en cuenta la simetría de la curva y que el área por debajo de la curva vale 1. Así, a partir del valor $P(Z < z)$, pueden obtenerse los valores de $P(Z > z)$, $P(Z < -z)$, $P(Z > -z)$, $P(0 < Z < z)$ y $P(-Z < z < Z)$.
En esta tabla, la cifra de las unidades y de las décimas se muestran en la columna de la izquierda, la de las centésimas en la fila superior.

Los valores de Z están dados en desviaciones típicas: $Z = 1$, indica un valor superior en una desviación típica a la media; $Z = 1,26$ indica 1,26 desviaciones típicas más que la media. $Z = 0$ indica que la variable se sitúa justamente en la media.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319

Ejemplos:
La probabilidad de que $Z < 1,26$, $P(Z<1,26)=0,8962$.
A partir de ese valor se deducen:

$P(Z>1,26)=1-0,8962=0,1038$; $P(Z<-1,26)=0,1038$; $P(Z>-1,26)=0,8962$;

$P(0<Z<1,26)=0,8962-0,5=0,3962$; $P(-1,26<Z<1,26)=2\cdot 0,3962=0,7924$.

En las figuras siguientes se indican las áreas correspondientes a cada valor.

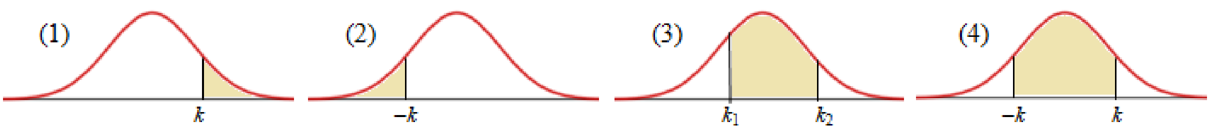
$P(Z<1,26)$ $\quad P(Z>1,26)$ $\quad P(Z<-1,26)$ $\quad P(Z>-1,26)$ $\quad P(0<Z<1,26)$ $\quad P(-1,26<Z<1,26)$

→ En general se cumple:

(1) $P(Z>k)=1-P(Z<k)$; (2) $P(Z<-k)=P(Z>k)=1-P(Z<k)$, válido para todo k

(3) $P(k_1<Z<k_2)=P(Z<k_2)-P(Z<k_1)$;

(4) $P(-k<Z<k)=P(Z<k)-P(Z<-k)=P(Z<k)-\left[1-P(Z<k)\right]=2\cdot P(Z<k)-1$.



Ejemplos:

a) $P(Z>0,8)=1-P(Z<0,8)=1-0,7881=0,2119$.

b) $P(Z<-075)=P(Z>0,75)=1-P(Z<0,75)=1-0,7734=0,2266$.

c) $P(-0,9<Z<1,25)=P(Z<1,25)-P(Z<-0,9)=0,8944-(1-P(Z<0,9))$ =

= 0,8944 – (1 – 0,8159) = 0,7103.

d) Como $P(Z<1)=0,8413$, puede deducirse que

$P(-1<Z<1)=P(Z<1)-P(Z<-1)=0,8413-(1-0,8413)=0,6826$,

que es el valor de probabilidad que se apuntó anteriormente.

5.3. Cálculo del valor de *Z* a partir de su probabilidad asociada

La tabla normal se emplea también en sentido contrario, para hallar la abscisa (el valor de Z) correspondiente a una probabilidad determinada.
Esto es, igual que se sabe que $P(Z < 1) = 0{,}8413$, en sentido contrario la pregunta sería: ¿cuánto debe valer z para que $P(Z < z) = 0{,}8413$? La respuesta es evidente: el valor de z debe ser 1.

Ejemplos:

a) El valor de z_1 tal que $P(Z < z_1) = 0{,}9207$ es $z_1 = 1{,}41$. Para determinarlo basta con buscar en la tabla normal el valor de Z correspondiente a una probabilidad de 0,9207.

b) Si la pregunta es: ¿cuánto debe valer Z para $P(-z_2 < Z < z_2) = 0{,}80$? Se procede así:
Como

$$P(-z_2 < Z < z_2) = P(Z < z_2) - P(Z < -z_2) = P(Z < z_2) - \left[1 - P(Z < z_2)\right] \Rightarrow$$

$$\Rightarrow P(-z_2 < Z < z_2) = 2 \cdot P(Z < z_2) - 1.$$

Así: $P(-z_2 < Z < z_2) = 0{,}80 \Rightarrow 2 \cdot P(Z < z_2) - 1 = 0{,}80 \Rightarrow P(Z < z_2) = 0{,}90 \Rightarrow z_2 = 1{,}28$.

c) Un caso que se presenta con frecuencia es encontrar el intervalo $(-z, z)$ que contiene el 95% de los datos de la variable estadística. Esto es, hallar el valor z tal que $P(-z < Z < z) = 0{,}95$.
Por el ejemplo anterior:

$$P(-z < Z < z) = 0{,}95 \Rightarrow 2 \cdot P(Z < z) - 1 = 0{,}95 \Rightarrow P(Z < z) = 0{,}9750 \Rightarrow z = 1{,}96.$$

→ Si el valor de probabilidad no figura en la tabla se tomará el más cercano. Así se acaba de hacer en el ejemplo b): en la tabla el valor más cercano a 0,9000 es 0,8997. (Suponemos que una diferencia de 3 diezmilésimas no será problemática).
También puede optarse por la interpolación. Así, para $P(Z < z) = 0{,}9950$ se toma $z = 2{,}575$, intermedio entre 2,57 y 2,58, cuyos valores de probabilidad respectivos son 0,9949 y 0,9951.

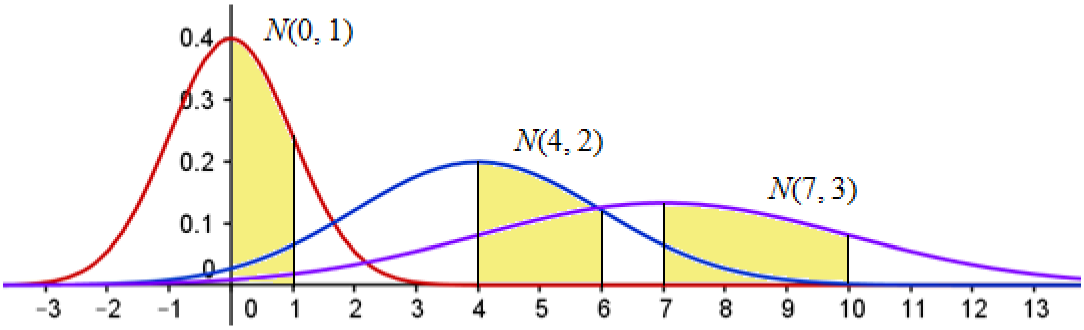
5.4. Tipificación

Las distribuciones normales con las que se trabaja en la práctica no son la estándar: la $N(0, 1)$. Son distribuciones con media μ (la que sea) y desviación típica σ: $N(\mu, \sigma)$.
Para calcular valores de probabilidad de una variable X, normal $N(\mu, \sigma)$, se hace el cambio de variable $Z = \dfrac{X - \mu}{\sigma}$, siendo Z la $N(0, 1)$.
Con esto, las diferencias de X respecto de su media μ se "tipifican": se calculan en desviaciones típicas, pues lo significativo no es el valor que tome X, sino cuántas desviaciones típicas es mayor o menor que su media. (En cada caso se utiliza una unidad relativa que es la desviación típica σ).

Ejemplo:

En la normal $N(0, 1)$, el valor de $Z = 2$ (dos unidades típicas a la derecha de la media $\mu = 0$) está, relativamente, lo mismo de alejado de la media 0, que en la normal $N(45, 7)$ el valor 59, que también es 2 desviaciones típicas mayor que $\mu = 45$, pues $59 = 45 + 2 \cdot 7$. Al tipificar, el valor $X = 59$ se transforma en $Z = 2$.
En efecto, para una $N(45, 7)$, si $X = 59$, el valor de Z tipificado es $Z = \dfrac{59 - 45}{7} = \dfrac{14}{7} = 2$.

- El proceso de tipificación se puede explicar gráficamente con ayuda de la siguiente figura. En ella, los recintos coloreados tienen la misma área.



→ En el caso de la *N*(0, 1), el área coloreada mide la probabilidad de que la variable *Z* tome valores entre 0 y 1: $P(0<Z<1)$.

Esta probabilidad vale:

$$P(0<Z<1)=P(Z<1)-P(Z<0)=0,8413-0,5000=0,3413.$$

→ Para la variable *X*, *N*(4, 2), el área coloreada entre la curva y los valores *X* = 4 y *X* = 6, mide la probabilidad $P(4<X<6)$. Haciendo el cambio de variable $Z=\dfrac{X-4}{2}$ se tiene:

$$P(4<X<6)=P\left(\frac{4-4}{2}<Z<\frac{6-4}{2}\right)=P(0<Z<1)=P(Z<1)-P(Z-0)=0,3413.$$

→Y lo mismo para la variable *X*, *N*(7, 3): el área coloreada entre la curva y los valores *X* = 7 y *X* = 10, mide la probabilidad $P(7<X<10)$, que también vale 0,3413. El cambio es $Z=\dfrac{X-7}{3}$.

- En general, para la variable $X \approx N(\mu, \sigma)$, la probabilidad de que *X* < *k*, se calcula así:

$$P(X<k)=P\left(\frac{X-\mu}{\sigma}<\frac{k-\mu}{\sigma}\right)=P\left(Z<\frac{k-\mu}{\sigma}\right)$$, donde *Z* es *N*(0, 1).

Ejemplos:

a) Si *X* es una variable normal *N*(45, 7), la probabilidad de que *X* tome valores menores de 52, mayores de 52; o entre 52 y 59 es:

$$P(X<52)=P\left(Z<\frac{52-45}{7}\right)=P(Z<1)=0,8413;$$

$$P(X>52)=1-P(Z<52)=1-P\left(Z<\frac{52-45}{7}\right)=1-0,8413=0,1587;$$

$$P(52<X<59)=P\left(\frac{52-45}{7}<Z<\frac{59-45}{7}\right)=P(1<Z<2)=P(Z<2)-P(Z-1)=$$

= 0,9772 – 0,8413 = 0,1359.

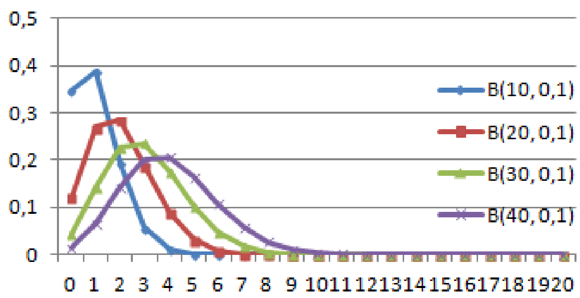
b) En sentido inverso, también se puede encontrar el valor de *X* a partir de la probabilidad asociada. Esto es, para la misma *N*(45, 7), si se desea conocer el valor de *k* tal que, por ejemplo, $P(X<k)=0,9772$, se resuelve

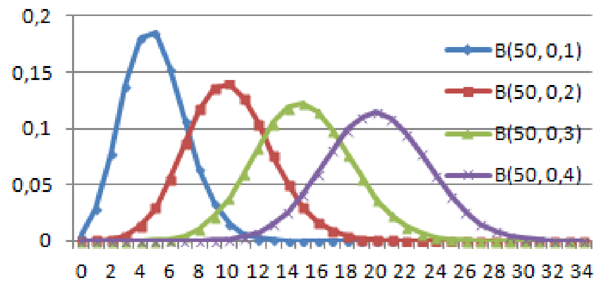
$$P(X<k)=P\left(Z<\frac{k-45}{7}\right)=0,9772 \Rightarrow \frac{k-45}{7}=2 \Rightarrow k=45+2\cdot 7=59.$$

6. Aproximación de la distribución binomial mediante una normal

Al estudiar la distribución binomial, $B(n, p)$, se ha indicado que para valores grandes de n el cálculo de probabilidades se hace engorroso; además, los valores de probabilidad pueden resultar muy pequeños. Así, por ejemplo, para la binomial $B(50, 0{,}12)$, las probabilidades de r = 6 y de r = 12 son: $P(X=6)=0{,}171185935$ y $P(X=12)=0{,}0084088$. Más complicada resulta, por ejemplo, calcular la probabilidad de que $X > 10$ o de que $4 < X < 12$.
Tales inconvenientes pueden obviarse, pues, para valores grandes de n, la distribución binomial se aproxima bastante bien mediante una distribución normal.

En las siguientes figuras puede observarse la evolución de las poligonales de frecuencias de la distribución binomial $B(n, 0{,}1)$ cuando n va creciendo, y de la distribución binomial $B(50, p)$ cuando p se acerca a 0,5.





(En estas figuras, que se han dibujadas con Excell, en el eje horizontal se indica el número r de éxitos para cada binomial; en el eje vertical se da su probabilidad: $P(X=r)$).

Como ves, cuando n aumenta la poligonal se parece más a una campana de Gauss; y es mucho más evidente cuando n es grande y p se acerca más a 0,5. En general, se admite que la aproximación es buena cuando $n \geq 25$ y el producto $np \geq 5$.
Así, para las poligonales de la figura de la izquierda la única que puede admitirse como aceptable es la $B(40, 0{,}1)$, aunque $np = 4$. En cambio, las binomiales representadas en la figura de la derecha se aproximan bastante bien a campanas de Gauss.

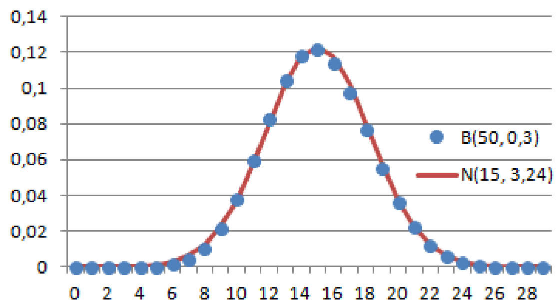
→ <u>Cuando el ajuste sea posible</u>, la distribución normal que mejor se aproxima a la $B(n, p)$ es la que tiene por media y desviación típica la de la distribución binomial. Como la media y desviación típica de la variable $X \approx B(n,p)$ son $\mu = np$ y $\sigma = \sqrt{npq}$, el ajuste se hace por la variable $X' \approx N\left(np,\ \sqrt{npq}\right)$, se tipifica haciendo $Z = \dfrac{X'-\mu}{\sigma} = \dfrac{X'-np}{\sqrt{npq}}$.

Con esto, cuando sea preciso calcular probabilidades de una variable X binomial $B(n, p)$ puede hacerse recurriendo a una variable normal X' asociada. Por tanto:

$$P(X<k) = P(X'<k) = P\left(Z < \frac{k-np}{\sqrt{npq}}\right).$$

En la figura adjunta se han dibujado la $B(50, 0{,}3)$ y la $N\left(50\cdot 0{,}3,\ \sqrt{50\cdot 0{,}3\cdot 0{,}7}\right) = N(15,\ 3{,}24)$.

Las probabilidades binomiales se representan por puntos; la normal es la curva continua. Es evidente la gran coincidencia.



Ejemplo:
La variable binomial $X = B(50,\ 0{,}3)$ se estudia mediante la variable normal $X' = N(15,\ \ 3{,}24)$, entonces:

$$P(X<18) = P(X'<18) = P\left(Z < \frac{18-15}{3,24}\right) = P(Z<0,93) = 0,8238.$$

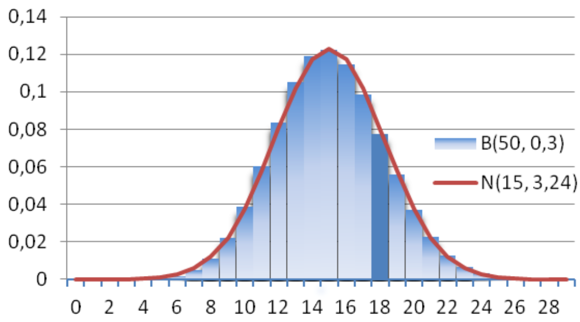
$$P(15<X<18) = P(X'<18) - P(X'<15) = P(Z<0,93) - P(Z<0) = 0,8238 - 0,5 = 0,3238$$

6.1. Corrección de continuidad

En la distribución normal la probabilidad de que la variable tome un valor concreto es 0, pero en una binomial no es así. Por seguir con el ejemplo anterior:

- Para la normal $X' = N(15,\ \ 3{,}24)$, $P(X'<18) = P(X'\le 18)$, pues $P(X'=18)=0$.
- Para la binomial $X = B(50,\ 0{,}3)$, $P(X<18) < P(X\le 18)$, pues $P(X=18) = 0,077247062$,

que es el resultado de $P(X=18) = \binom{50}{18} \cdot 0,3^{18} \cdot 0,7^{22}$. Este valor coincide con el área de la barra sobre $X = 18$ (sombreada con mayor intensidad), que se muestra en el gráfico adjunto.



Esta anomalía puede paliarse mediante la llamada corrección de continuidad, por la que probabilidades puntuales (de valor 0 en las distribuciones continuas) son sustituidas por probabilidades de intervalo de la forma siguiente:

$\to P(X=k) = P(k-0,5 < X' < k+0,5)$

$\to P(X\le k) = P(X' < k+0,5)$ $\qquad \to P(X<k) = P(X'<k-0,5)$

$\to P(X\ge k) = P(X' > k-0,5)$ $\qquad \to P(X>k) = P(X'>k+0,5)$

Ejemplos:
Para la binomial $X = B(50,\ 0{,}3) \approx X' = N(15,\ \ 3{,}24)$, se tiene:

a) $P(X=15) = P(14,5 < X' < 15,5) = P\left(\frac{14,5-15}{3,24} < Z < \frac{15,5-15}{3,24}\right) = P(-015 < Z < 0,15) =$

$= P(Z<0,15) - P(Z<-0,15) = 0,5596 - (1-0,5596) = 0,1192.$

b) $P(X=18) = P(17,5 < X' < 18,5) = P\left(\frac{17,5-15}{3,24} < Z < \frac{18,5-15}{3,24}\right) =$

$= P(Z<1,08) - P(Z<0,77) = 0,8599 - 0,7794 = 0,0805.$

c) $P(15<X<18) = P(X'<17,5) - P(X'<15,5) = P(Z<1,08) - P(Z<0,15) =$

$= 0,8599 – 0,5596 = 0,3003.$

Observación: En todos los casos, los valores de probabilidad obtenidos mediante la normal son aproximados. El resultado binomial exacto, obtenidos con Excel, es:

$P(X=15) = 0,122346862$; $P(X=18) = 0,077247062$. (La diferencia es de milésimas).

7. Ejercicios finales

Ejercicio 1. Utilizando la tabla normal *N*(0, 1) calcula el valor de *k* tal que:

a) $P(Z<1{,}5)=k$ b) $P(Z<-0{,}5)=k$ c) $P(Z<k)=0{,}67$

Solución:

a) $P(Z<1{,}5)=0{,}9332$ ⇒ k = 0,9332.

b) $P(Z<-0{,}5)=1-P(Z<0{,}5)=1-0{,}6915=0{,}3085$ ⇒ k = 0,3085.

c) $P(Z<k)=0{,}67$ ⇒ k = 0,44.

Ejercicio 2. Admitamos que el peso, en kg, de los habitantes adultos de una gran ciudad sigue una distribución normal de media 60 kg y desviación típica 5 kg. Si se elige una de las personas al azar, ¿qué probabilidad hay de que pese?:

a) Menos de 50 kg. b) Entre 52 y 65 kg.

c) Elegidas 100 personas, ¿cuántas cabe esperar que pesarán más de 65 kg?

Solución:

La variable *X*, que mide el peso de esas personas, se distribuye según la *N*(60, 5). Se tipifica haciendo el cambio $Z=\frac{X-60}{5}$. Con esto:

a) $P(X<50)=P\left(Z<\frac{50-60}{5}\right)=P(Z<-2)=1-P(Z<2)=1-0{,}9772=0{,}0228$.

b) $P(52<X<65)=P\left(\frac{52-60}{5}<Z<\frac{65-60}{5}\right)=P(-1{,}6<Z<1)=P(Z<1)-P(Z<-1{,}6)=$

= 0,8413 – (1 – 0,9452) = 0,7865.

c) La probabilidad de que una persona pese más de 65 kg es:

$P(X>65)=P\left(Z>\frac{65-60}{5}\right)=P(Z>1)=1-P(Z<1)=1-0{,}8413=0{,}1587$.

Entonces, para 100 personas: 100 · 0,1587 = 15,87 ≈ 16 pesarán más de 65 kilos.

Ejercicio 3. En un determinado país, se ha estimado en un 22% el porcentaje de hombres con problemas de alopecia (calvicie). Si se toma una muestra aleatoria de 40 hombres:

a) ¿Cuál es la probabilidad de que exactamente 10 de ellos tengan problemas de alopecia?

b) ¿Y de que haya más de 10?

Solución:

La variable *X* que computa el número de personas con alopecia es una variable *B*(40, 0,22) que puede aproximarse por la normal *X*': *N*(40 · 0,22, $\sqrt{40\cdot 0{,}22\cdot 0{,}78}$) = *N*(8,8, 2,62). Con esto, haciendo la corrección de continuidad y tipificando:

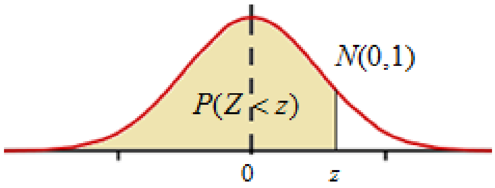
a) $P(X=10)=P(9{,}5<X'<10{,}5)=P\left(\frac{9{,}5-8{,}8}{2{,}62}<Z<\frac{10{,}5-8{,}8}{2{,}62}\right)=P(0{,}27<Z<0{,}65)=$

= 0,7422 – 0,6064 = 0,1358.

b) $P(X>10)=P(X'>10{,}5)=P(Z>0{,}65)=1-P(Z<0{,}65)=1-0{,}7422=0{,}2578$.

ÁREAS BAJO LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR

Los valores en la tabla representan el área bajo la curva normal hasta un valor positivo z.



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Problemas Propuestos

Distribución de Probabilidad

1. Una variable aleatoria discreta, *X*, se distribuye como se indica en la siguiente tabla:

X	1	2	3	4	5
$P(X = x_i) = p_i$	0,20	0,30	0,25	m	0,15

a) Halla el valor de *m*.
b) Calcula la media y la desviación típica de la variable.

2. La función de probabilidad de una variable aleatoria discreta *X* es:

X	0	1	2	3	4	5
$P(X = x_i)$	0,01	0,20	m	n	0,10	0,09

Calcula *m* y *n* si la media de *X* es 2,45.

3. a) En una lotería se pueden ganar 10000 € con probabilidad de 0,001; en los demás casos se pierde lo jugado. Si cada apuesta cuesta 12 €, ¿es un juego equitativo?
b) En otra lotería se pueden ganar 5000 € con probabilidad 0,002 o 15000 € con probabilidad 0,0001; en los demás casos se pierde lo jugado. Si cada apuesta cuesta 12 €, ¿es un juego equitativo?

4. Para cada una de las loterías anteriores, ¿cuánto debe valer cada apuesta si se quiere que el juego sea equitativo?

5. Dada $f(x) = \begin{cases} x^2 + kx, & \text{si } 0 \le x \le 2 \\ 0, & \text{en otro caso} \end{cases}$, ¿puede ser $f(x)$ función de densidad de una variable aleatoria continua para algún valor de *k*?

6. Dada $f(x) = \begin{cases} kx^2 + x, & \text{si } 0 \le x \le 2 \\ 0, & \text{en otro caso} \end{cases}$, ¿puede ser $f(x)$ función de densidad de una variable aleatoria continua para algún valor de *k*?
Si fuese una función de densidad calcula $P(0 < X < 1)$.

7. La variable aleatoria que mide el tiempo de espera, en minutos, para ser atendido en una empresa de telefonía móvil tiene función de densidad $f(x) = \frac{3}{512}x^2$, con $0 \le x \le 8$. (Si $x > 8$ la llamada se corta automáticamente). Halla:
a) La probabilidad de ser atendido en menos de 2 minutos.
b) La probabilidad de ser atendido después de 5 minutos de espera.

Distribución Binomial

8. Un dado, cuyas caras están numeradas del 1 al 6, se lanza cinco veces. Halla la probabilidad de que el número 3 salga:

a) Exactamente dos veces. b) Una vez a lo sumo. c) Más de una vez.

9. En un Centro Escolar el 25% de los alumnos son de origen extranjero. Si se eligen 6 estudiantes al azar, ¿cuál es la probabilidad de 4 o más sean de origen extranjero?

10. Se lanza una moneda correcta 10 veces y se mide el número de caras y cruces obtenidas.
a) ¿Cuántos resultados forman el espacio muestral? ¿Cuál es la probabilidad de cada uno de los resultados posibles?
b) ¿Cuál es la probabilidad de que salgan 4 caras?

11. En una moneda trucada la probabilidad de obtener cara es 0,4. Si se lanza 5 veces, calcula la probabilidad de obtener al menos 3 caras.

12. Un examen consta de 8 preguntas con 3 posibles respuestas cada una, de las que sólo una de ellas es correcta. Si un estudiante responde al azar marcando las respuestas aleatoriamente, calcula la probabilidad de que:

a) No acierte ninguna respuesta correcta. b) Acierte 6 o más preguntas.

13. Una compañía de seguros estima que la probabilidad de que un asegurado de motocicleta tenga algún tipo de accidente es 0,15. De 10 asegurados, ¿cuál es la probabilidad de que haya al menos 2 accidentados?

14. En un Centro Comercial el 35 % de los consumidores utiliza el coche para hacer la compra. Si se eligen al azar 7 consumidores que hayan realizado la compra en dicho Centro Comercial:
a) ¿Cuál es la probabilidad de 3 de ellos hayan ido en coche a comprar?
b) ¿Cuál es la probabilidad de todos hayan ido en coche?

Distribución Normal

15. Utilizando la tabla normal *N*(0, 1) calcula:

a) $P(Z<1,2)$ b) $P(Z<1,27)$

c) $P(Z<-1,2)$ d) $P(Z<-1,27)$

16. Utilizando la tabla normal *N*(0, 1) calcula interpolando:

a) $P(Z<1,325)$ b) $P(Z<1,645)$ c) $P(Z<0,666)$

d) $P(Z<1,863)$ e) $P(Z>1,45)$ f) $P(Z>-1,42)$

17. Utilizando la tabla normal *N*(0, 1), determina el valor de *k* que cumple:

a) $P(Z<k)=0,9115$ b) $P(Z<k)=0,9452$

c) $P(Z<k)=0,1587$ d) $P(Z<k)=0,95$

18. Para una distribución normal *N*(50, 5), halla:

a) $P(X<56)$ b) $P(X>58)$ c) $P(X<48)$ d) $P(48<X<56)$

19. Supongamos que la estatura media de las alumnas de bachillerato se distribuye normalmente con media μ = 166 cm y desviación típica 9 cm. Si se elige una alumna al azar halla la probabilidad de que su estatura sea:
a) Superior a 175 cm. b) Inferior a 155 cm. c) Esté entre 155 cm y 175 cm.

20. Para una distribución normal *N*(60, 5), determina el valor de *k* que cumple:

a) $P(X<k)=0{,}90$ b) $P(X>k)=0{,}95$ c) $P(60-k<X<60+k)=0{,}9544$

21. Supongamos que los chicos de 15 años de un determinado país tienen una estatura que se distribuye según una normal de media 168 cm y desviación típica 12 cm. Si se quieren seleccionar al 5% de los chicos más altos, ¿a partir de qué altura debe hacerse?

22. El diámetro de las ciruelas de una determina variedad se distribuye normalmente con media 4,5 cm y desviación típica 0,3 cm. Si se desea seleccionar, para su exportación, el 10% de las más grandes, ¿a partir de qué tamaño hay que cogerlas?

23. La edad de los habitantes de cierta ciudad se distribuye normalmente, con una media de 40 años. Se sabe además que el 2,28 % de los habitantes tiene más de 60 años.
a) ¿Cuál es la desviación típica?
b) ¿Cuál es el porcentaje de habitantes con menos de 35 años?

24. La duración de una determinada marca de lavadoras se ajusta a una normal de media 8,4 años y desviación típica 6 meses. El fabricante asegura que sus lavadoras duran más de 7 años, comprometiéndose a: "si una lavadora se estropea antes de 7 años le damos otra nueva". ¿Cuántas lavadoras nuevas tendrá que reponer por cada 10000 vendidas?

25. Los envases de cartón de una determinada marca de leche contienen 1 litro de media, siendo la desviación típica de 5 ml.
a) ¿Qué porcentaje de envases sobrepasan los 1005 ml.
b) Si el control de calidad rechaza los envases que contengan menos de 990 ml y más de 1010 ml, ¿qué porcentaje de envases habrá que rechazar?

Aproximación de la binomial mediante una normal

26. Mediante la aproximación normal de la binomial *B*(50, 0,12) calcula:

a) $P(X=6)$ b) $P(X=12)$ c) $P(6<X\le 12)$

27. El 42 % de los habitantes de un pueblo pasa cada día por la calle mayor. Elegidos 60 habitantes al azar, ¿qué probabilidad hay de que más de 30 de ellos pasen ese día por la calle mayor?

28. Un examen de respuesta múltiple consta de 80 preguntas, cada una con 4 opciones, una de ellas correcta y erróneas las otras tres. Si un estudiante contesta al azar, ¿cuál es la probabilidad de que acierte 25 o más preguntas? ¿Y menos de 10?

29. (Propuesto en ABAU 2018, Galicia)
En un bombo tenemos 10 bolas idénticas numeradas del 0 al 9 y cada vez que hacemos una extracción devolvemos la bola al bombo.
a) Si hacemos 5 extracciones, calcula la probabilidad de que el 7 salga menos de dos veces.
b) Si hacemos 100 extracciones, calcula la probabilidad de que el 7 salga menos de 9 veces.

30. (Propuesto en EvAU 2019, Madrid)
La probabilidad de que un pez de una determinada especie sobreviva más de 5 años es del 10%. Se pide:
a) Si en un acuario tenemos 10 peces de esta especie nacidos este año, hallar la probabilidad de que al menos dos de ellos sigan vivos dentro de 5 años.
b) Si en un tanque de una piscifactoría hay 200 peces de esta especie nacidos este mismo año, usando una aproximación mediante la distribución normal correspondiente, hallar la probabilidad de que al cabo de 5 años hayan sobrevivido al menos 10 de ellos.

Otros problemas

31. El peso de los estudiantes varones de una universidad se distribuye normalmente con media 68,5 kilos y desviación típica, 10 kilos. Halla:
a) El porcentaje de estudiantes que pesan entre 48 y 71 kilos.
b) El porcentaje de estudiantes que pesan más de 91 kilos
c) Si se eligen 5 alumnos al azar, ¿cuál es la probabilidad de que exactamente 2 de ellos pesen más de 75 kilos?

32. Supongamos que la estatura de los jóvenes de 20 años de una determinada región sigue una distribución normal de media 175 cm. Si se sabe además que los jóvenes que miden más de 190 cm representan el 6,68 % del total, calcula:
a) La desviación típica de la población considerada.
b) El porcentaje de jóvenes con estatura superior a 165 cm

33. El 1 % de los individuos de una población supera los 185 cm de estatura, mientras que el 3 % no llega a los 160 cm. Si se supone que la estatura sigue una distribución normal, calcula la media y la desviación típica de esa distribución.

34. En un test de inteligencia, las puntuaciones se distribuyen normalmente, con media 100 y desviación típica 25. Si el 10 % de las puntuaciones más altas corresponde al grupo de los superdotados, ¿qué puntuación mínima hay que alcanzar para entrar en el grupo de los superdotados?

35. Determina el valor de *k* para que la función $f(x)=\begin{cases} k\sin x, & \text{si } x\in[0,\ \pi] \\ 0, & \text{en otro caso} \end{cases}$ sea una función de densidad. Para ese valor de *k*:
a) Halla la expresión de la función de distribución y calcula la probabilidad de de *X* tome valores menores que $\pi/3$.
b) Calcula la media de la variable aleatoria que tiene por función de densidad a $f(x)$.

Soluciones

1. a) 0,10. b) 2,7; 1,31.
2. 0,40; 0,20.
3. a) $\mu = -1,988$; no. b) $\mu = -0,4748$; no.
4. a) 10,01. b) 11,52.
5. No.
6. –3/8; 5/16.
7. a) 8/512. b) 387/512.
8. a) 0,16075. b) 0,80375. c) 0,19625.
9. 0,03759.
10. a) 1024; 1/1024. b) 105/5122.
11. 0,31744.
12. a) 256/6561. b) 129/6561.
13. 0,4557.
14. a) 0,2679. b) 0,00064.
15. a) 0,8849. b) 0,8980. c) 0,1151. d) 0,1020.
16. a) 0,9074. b) 0,95. c) 0,74732. d) 0,96881. e) 0,0735. f) 0,9222.
17. a) 1,35. b) 1,6. c) –1. d) 1,645.
18. a) 0,8849. b) 0,0548. c) 0,3446. d) 0,5403.
19. a) 0,1587. b) 0,1112. c) 0,7301.
20. a) 66,4. b) 51,775. c) 10.
21. 187,74 cm.
22. 4,884 cm.
23. a) 10. b) 30,85 %.
24. 26.
25. a) 15,87 % b) 4,56 %
26. a) 0,1742. b) 0,0061. c) 0,5787.
27. 0,0823.
28. 0,1230; 0,0034.
29. a) 0,08146. b) 0,3085.
30. a) 0,264. b) 0,9934.
31. a) 57,85 % b) 1,22 % c) 0,2717.
32. a) 10. b) 84,13 %
33. 169,37; 4,99.
34. 132.
35. a) $F(x) = -\frac{1}{2}\cos x + \frac{1}{2}$; 1/4. b) $\pi/2$.