

DISTRIBUCIÓN NORMAL

RESUMEN

Características generales

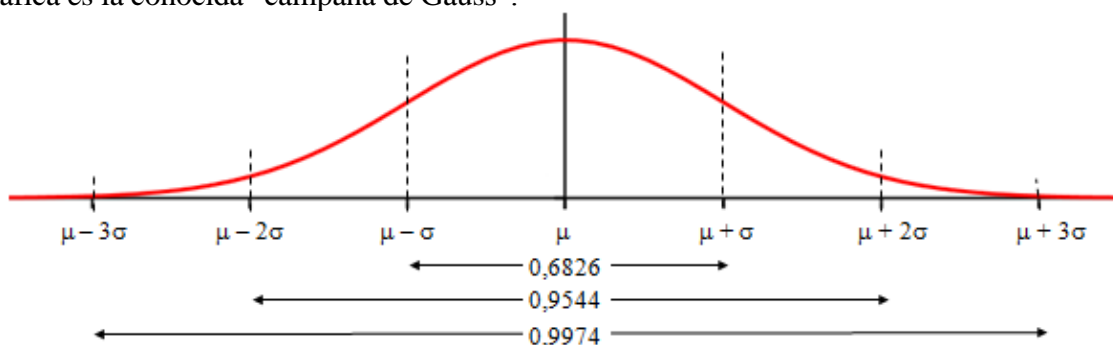
Es una distribución de probabilidad continua asociada (teóricamente) a multitud de fenómenos naturales y cotidianos (cociente intelectual, talla o peso de las personas; tamaño de los frutos de cualquier tipo de árbol...), que se caracterizan porque la mayoría de los resultados tienden a agruparse en 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 $(-\infty, +\infty)$: la variable puede tomar cualquier valor; siendo $f(x) > 0$ para todo x .
- El área del recinto plano limitado por la curva y el eje OX vale 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 $-\infty$ y $+\infty$, 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 $\mu - \sigma$ y $\mu + \sigma$ es 0,6826; entre $\mu \pm 2\sigma$ es de 0,9544; y entre $\mu \pm 3\sigma$ es de 0,9974. De hecho, a los valores que están a una distancia superior a 3,5 σ 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 $\mu = 175$ cm y desviación típica $\sigma = 9$ cm: $N(175, 9)$. Entonces, puede asegurarse, con las probabilidades que se indican, que:

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

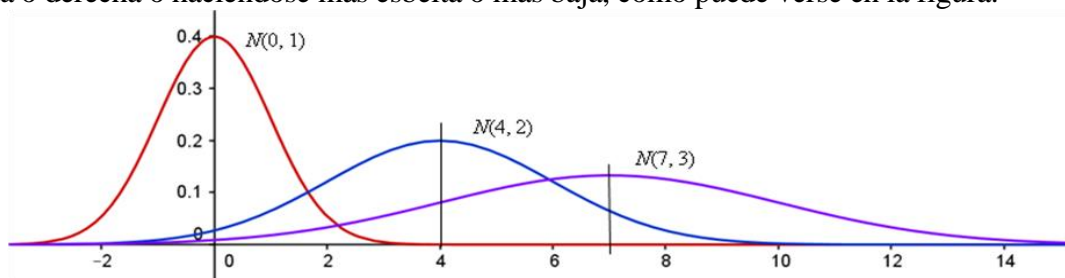
$$P(\mu - \sigma < X < \mu + \sigma) = P(175 - 9 < X < 175 + 9) = P(166 < X < 184) = 0,6826 \rightarrow$$

(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.

Distribución normal de media 0 y desviación típica 1: $N(0, 1)$

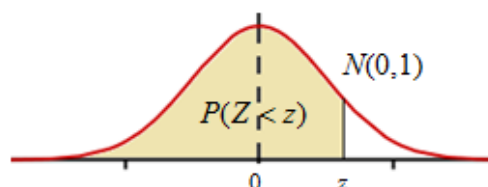
El comportamiento estadístico normal hace que puedan asignarse valores de probabilidad a cualquier suceso de la variable estudiada. Esto es, se puede saber (pues está tabulado) la probabilidad de que la variable tome valores comprendidos entre los extremos de un intervalo dado.

Lo que está tabulado es la función de distribución en el caso de la curva normal de media $\mu = 0$ y desviación típica $\sigma = 1$, la normal $N(0, 1)$.

La función de distribución, $F(x)$, da la superficie del recinto limitado por la función de densidad (la curva $y = f(x)$ definida más arriba) y el eje OX , desde $-\infty$

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).

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 $-\infty$ 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:

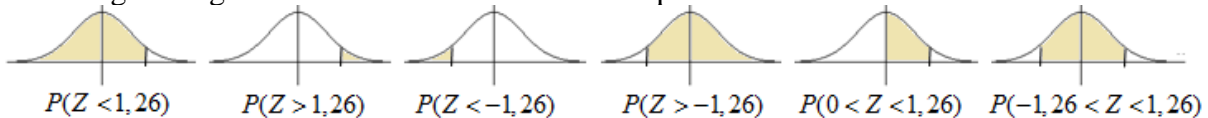
La probabilidad de que $Z < 1,26$, $P(Z < 1,26) = 0,8962$.

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.

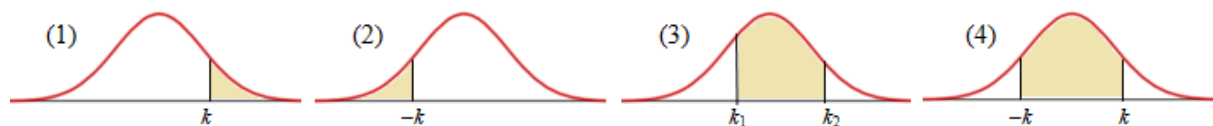


→ En general se cumple:

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



Ejemplos:

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

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

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.

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

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

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.

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

$$Z = \frac{X - \mu}{\sigma}, \text{ siendo } Z \text{ la } N(0, 1).$$

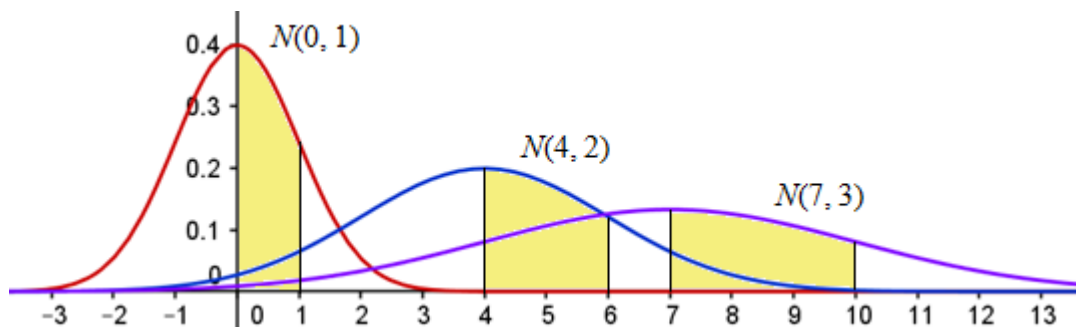
Con esto, las diferencias de X respecto de su media μ se “tipifican”: se calculan en desviaciones típicas, pues lo significativo no es el valor que tome X , sino cuántas desviaciones típicas es mayor o menor que su media. (En cada caso se utiliza una unidad relativa que es la desviación típica σ).

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 $Z = \frac{59 - 45}{7} = \frac{14}{7} = 2$.

- 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 $Z = \frac{X - 4}{2}$ 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 $Z = \frac{X - 7}{3}$.

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

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

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

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

Observación. [Ampliación de este resumen](#), pp. 297 y ss.