

Distribuciones continuas. Distribución normal.

Función de densidad.

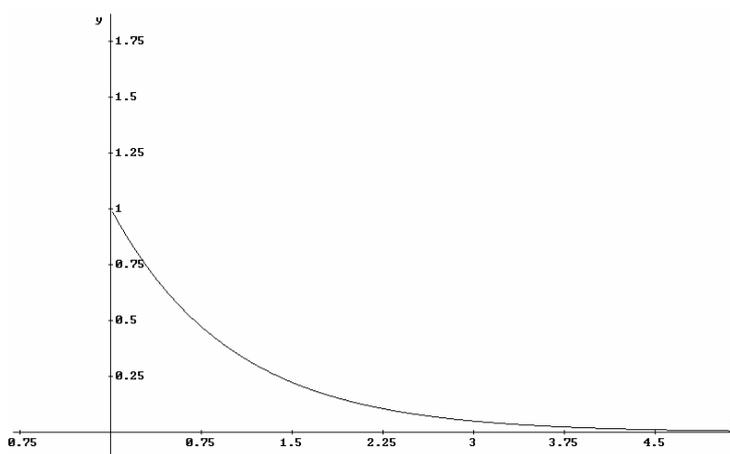
Cuando X es una variable aleatoria continua, existe una función f_x , denominada función de densidad asociada a la variable aleatoria X , que cumple

- a) $f_x(x) \geq 0 \quad \forall x \in \mathbb{R}$
- b) f_x admite un número finito de discontinuidades en cada intervalo
- c) Si P_x es la función de probabilidad de la v. a. X , se cumplirá

$$P_x(X < x) = \int_{-\infty}^x f_x(t) \cdot dt$$

Ejemplo.- Diremos que X es una distribución exponencial de parámetro $\theta=1$, y lo representamos por $\exp(1)$, si su función de densidad es de la forma

$$f_x(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-x} & \text{si } x > 0 \end{cases}$$



Que claramente se comprueba que verifica la Positividad, por ser f una función positiva o nula. Además, también se cumple la normalización, ya que

$$\int_{-\infty}^{+\infty} f_x(x) \cdot dx = \int_0^{+\infty} e^{-x} \cdot dx = 1$$

Así por ejemplo

$$P_x(1 \leq X \leq 2) = \int_1^2 e^{-x} \cdot dx = e^{-1} - e^{-2}$$

Media y varianza de una variable aleatoria discreta

Si X es una variable aleatoria continua de función de densidad f_X , denominamos

Media o esperanza matemática: $\mu = \int_{-\infty}^{+\infty} x \cdot f_X(x) \cdot dx$

Varianza: $\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f_X(x) \cdot dx$

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

Ejemplo.- Hallar la media, varianza y desviación típica de una variable aleatoria continua que tiene la siguiente función de densidad

$$f_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{3} & \text{si } 0 \leq x \leq 3 \\ 0 & \text{si } x > 3 \end{cases}$$

a) Media: $\mu = \int_0^3 x \cdot \frac{1}{3} = \frac{9}{6} = 1,5$

b) Varianza: $\sigma^2 = \int_0^3 (x - 1,5)^2 \cdot \frac{1}{3} = 0,75$

c) Desviación típica: $\sigma = \sqrt{0,75} = 0,866$

Variable aleatoria de distribución normal

La v. a. continua X sigue la distribución Normal de parámetros μ y σ , y se escribe de forma abreviada $X \equiv N(\mu, \sigma)$ si para cada $x \in \mathbb{R}$, la probabilidad P cumple

$$P(X \leq x) = \int_{-\infty}^x f(t) \cdot dt = \int_{-\infty}^x \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \cdot dt$$

Siendo μ y σ números reales, tal que $\sigma > 0$. Y a la función $f(x)$ se le denomina función de densidad.

La función de densidad $f(x)$ cumple

1.- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

2.- $\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$

Media y varianza de $N(\mu, \sigma)$

Por la propia definición de la distribución, se cumple

Media = μ ; Varianza = σ^2 ; Desviación típica = σ

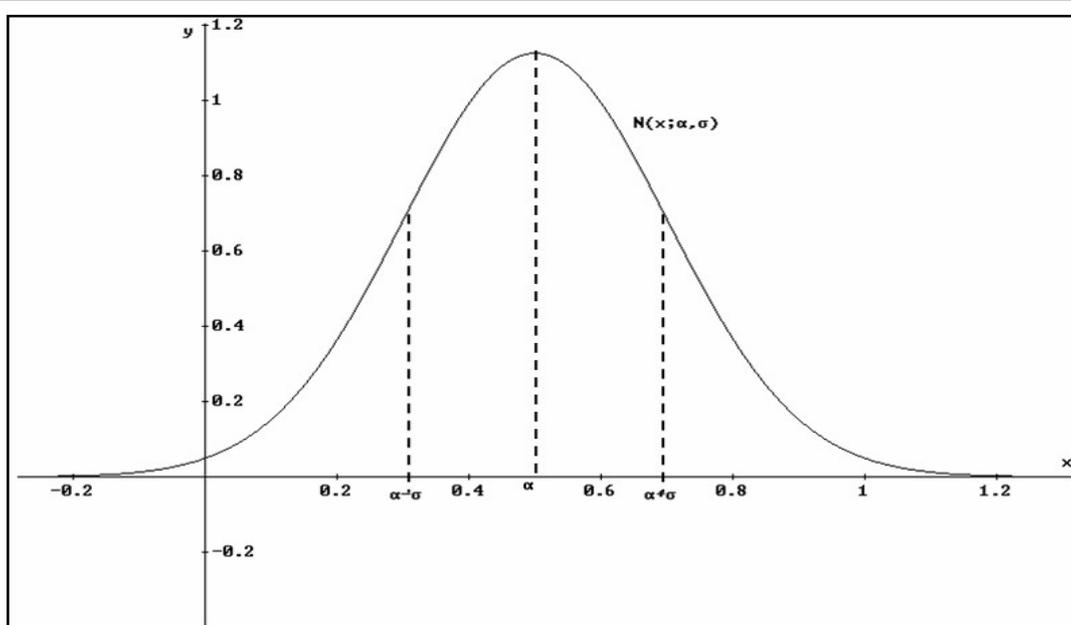
Propiedades de la función de densidad de una distribución $N(\mu, \sigma)$:

- La función $f(x)$ su valor máximo en $x=\mu$, es decir f alcanza el máximo en el punto

$$(\mu, f(\mu)) = \left(\mu, \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \right)$$

- La función $f(x)$ es simétrica respecto del punto $x=\mu$, es decir $\forall x \in \mathbb{R}$, se cumple $f(\mu-x) = f(\mu+x)$
- Cuando x tiende a $-\infty$ o $+\infty$, f(x) tiende a cero
- La función es convexa en $(-\infty, \mu-\sigma) \cup (\mu+\sigma, +\infty)$ y cóncava en $(\mu-\sigma, \mu+\sigma)$

Es decir, la función es de la forma



Distribución normal estándar

Teniendo en cuenta que dependiendo de los valores de μ y de σ , la gráfica de la función estará más desplazada hacia un lado y más o menos comprimida, y por tanto más o menos alta, siempre podemos utilizar la variable normal tipificada $N(0,1)$, tipificando la variable $X \equiv N(\mu, \sigma)$ a la variable $Y \equiv N(0,1)$ mediante el cambio o tipificación de la variable

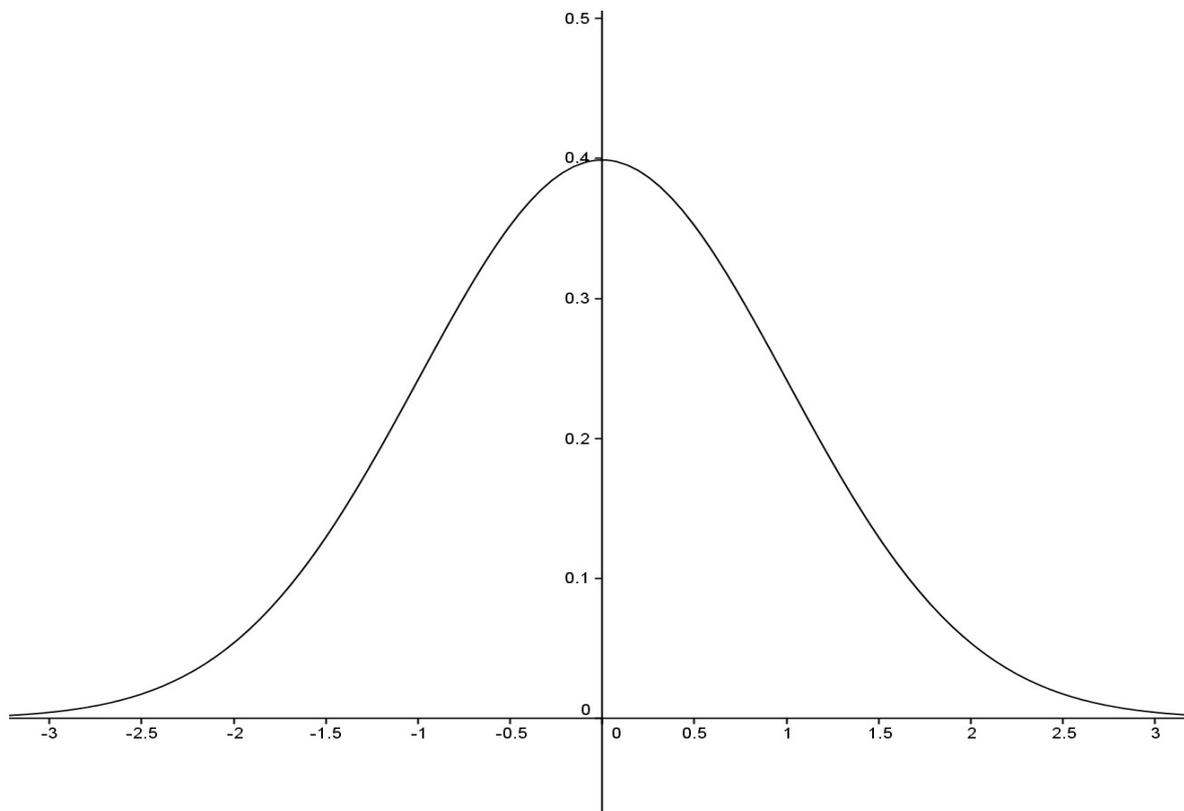
$$Y = \frac{X - \mu}{\sigma}$$

y para cada $y \in \mathbb{R}$, la probabilidad P cumple

$$P(Y \leq y) = \int_{-\infty}^y f(t) \cdot dt = \int_{-\infty}^y \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{t^2}{2}} \cdot dt$$

y cuya función de densidad $f(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}$ la podemos representar gráficamente

como



Habitualmente calculamos esta probabilidad de forma aproximada mediante la tabla tabulada de la distribución $N(0,1)$.

Ejemplo.- Si las calificaciones de los 500 alumnos de magisterio de Castilla la Mancha presentados a un examen, se distribuye normalmente con media 6,5 y varianza 4.

Para calcular la probabilidad de que un aspirante obtenga más de 8 puntos. Como las calificaciones X de los alumnos siguen una distribución normal $N(6,5, \sqrt{4}) = N(6,5, 2)$.

Y la variable $Z = \frac{X - \bar{x}}{\sigma} = \frac{X - 6,5}{2}$ sigue una distribución $N(0,1)$.

Tenemos $X > 8 \Leftrightarrow Z > \frac{8 - 6,5}{2}$, luego

$$P(X > 8) = P\left(\frac{8 - 6,5}{2}\right) = P(Z > 0,75) = 1 - P(Z \leq 0,75) = 1 - 0,7734 = 0,2266$$

Donde el valor $P(Z \leq 0,75)$ los hallamos con una tabla de la distribución normal $N(1,0)$

Ejemplos de Manejo de tablas de distribución $N(0,1)$

Ejemplos

- $P(Z \leq 0,45) = 0,6736$
- $P(Z > 1,24) = 1 - P(Z \leq 1,24) = 1 - 0,8925 = 0,1075$
- $P(Z \leq -0,72) = P(Z \geq 0,72) = 1 - P(Z < 0,72) = 1 - P(Z \leq 0,72) = 1 - 0,7642 = 0,2358$
- $P(0,5 \leq Z \leq 1,76) = P(Z \leq 1,76) - P(Z \leq 0,5) = 0,9608 - 0,6915 = 0,2693$
- $P(-1,76 \leq Z \leq -0,5) = P(Z \leq -0,5) - P(Z \leq -1,76) =$
 $= P(Z > 0,5) - P(Z > 1,76) = (1 - P(Z \leq 0,5)) - (1 - P(Z \leq 1,76)) =$
 $= P(Z \leq 1,76) - P(Z \leq 0,5) = 0,9608 - 0,6915 = 0,2693$
- $P(-0,53 \leq Z \leq 2,46) = P(Z \leq 2,46) - P(Z \leq -0,53) =$
 $= P(Z \leq 2,46) - P(Z > 0,53) = P(Z \leq 2,46) - (1 - P(Z \leq 0,53)) =$
 $= 0,9931 - (1 - 0,7019) = 0,9931 - 0,2981 = 0,695$

Intervalo de confianza para la media

Si el parámetro μ de una distribución normal $N(\mu, \sigma)$ es desconocido, y para estimarlo tomamos el estimador \bar{x} , al tomar una muestra aleatoria de tamaño n .

Como la variable aleatoria

$$\text{var } X \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Tipificando, resulta que

$$Z = \frac{X - \mu}{\sqrt{\frac{\sigma}{n}}} \equiv N(0,1)$$

y se cumplirá que para la probabilidad $1 - \alpha$ (*coeficiente de confianza*), existirá un valor

crítico $z_{\frac{\alpha}{2}}$, tal que

$$\begin{aligned} P\left[-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right] &= P\left[-z_{\frac{\alpha}{2}} < \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right] = \\ &= P\left[-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = \\ &= P\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \end{aligned}$$

Luego, el intervalo de confianza para μ , al nivel de confianza $1 - \alpha$ es

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Si σ es desconocido y $n \geq 30$, podemos tomar la desviación típica muestral

$$\hat{s} = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

en vez de la desviación típica σ y obtendremos el intervalo de

confianza

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}; \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} \right)$$

Ejemplo.- Se sabe que la desviación típica de las tallas de los alumnos de una Universidad es igual 5 cm. Se desea estimarla talla media de los alumnos, para lo cual se escoge una muestra de 100 estudiantes y se obtiene la media bar $x = 172$ cm. Para hallar el intervalo de confianza al nivel 0,90, 0,95 y 0,99.

Como

$$\bar{x} \equiv N\left(\mu, \frac{\sigma}{n}\right) = N\left(\mu, \frac{5}{\sqrt{100}}\right) = N(\mu; 0,5)$$

Luego, los intervalos de confianza para la talla media de los alumnos al nivel de confianza

$1 - \alpha$ será

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{5}{\sqrt{100}}; \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{5}{\sqrt{100}} \right) = N(\mu; 0,5)$$

que para 90%, 95% y 99%, serán

$$\alpha = 0,05 \quad \Rightarrow \quad (172 - 1,64 \cdot 0,5; 172 + 1,64 \cdot 0,5) = (171,18; 172,82)$$

$$\alpha = 0,025 \quad \Rightarrow \quad (172 - 1,96 \cdot 0,5; 172 + 1,96 \cdot 0,5) = (171,02; 172,98)$$

$$\alpha = 0,005 \quad \Rightarrow \quad (172 - 2,58 \cdot 0,5; 172 + 2,58 \cdot 0,5) = (170,71; 173,29)$$

Para hallar los intervalos de confianza, tanto para la proporción como para la media, mirando el la tabla $N(0,1)$, para cada valor $1 - \alpha$, obtenemos los valores más utilizados

$1 - \alpha$	α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0,8	0,2	0,1	1,28
0,9	0,1	0,05	1,64
0,95	0,05	0,025	1,96
0,99	0,01	0,005	2,58

Relación entre distribución Binomial $B(n, p)$ y distribución $N(\mu, \sigma)$

Cuando n es grande en la distribución discreta $B(n, p)$ podemos utilizar de forma aproximada la distribución $N(n.p, \sqrt{n.p.q})$, particularmente es fiable si $n.p \geq 0,5$ y $n.q \geq 0,5$

Ejemplo.- Sea un experimento de Bernoulli (solo dos resultados posibles) con resultados que denotaremos $\{a, b\}$, y probabilidades $p, q=1-p$. Dado un número entero $n > 1$, consideremos el experimento consistente en la repetición de n experimentos de Bernoulli. Si definimos la variable:

$x_i = 1$, cuando ocurre el suceso a en el experimento i -ésimo.

$x_i = 0$, cuando ocurre el suceso b en el experimento i -ésimo, $i = 1, 2, \dots, n$.

Teniendo en cuenta que:

$$\mu = p.1 + q.0 = p.$$

$$\sigma^2 = p.(1-p)^2 + q.(0-q)^2 = p.q(q+p) = p.q.$$

Y denominando:

$$s_n = (x_1 + x_2 + \dots + x_n)$$

Se cumplirá:

$$\lim_{n \rightarrow +\infty} B(n, p) \rightarrow N(p, \sqrt{n.p.q}).$$