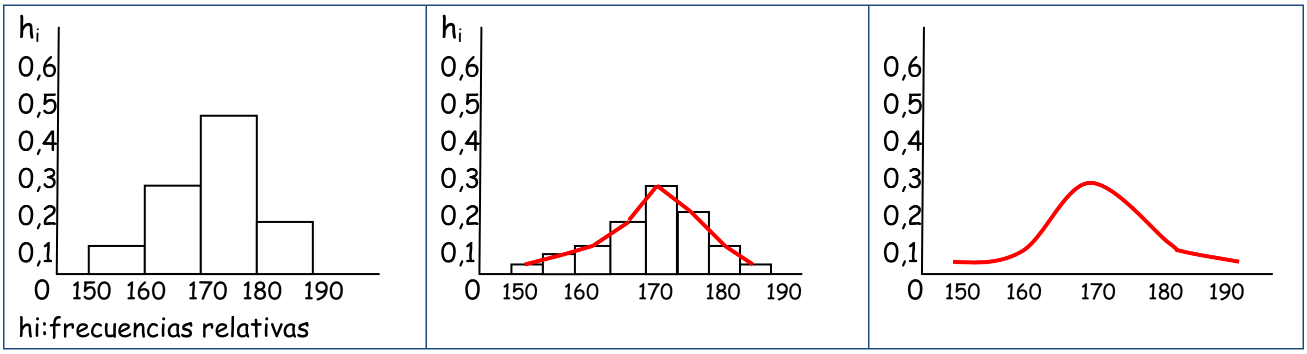
Distribuciones continuas. Distribución normal

Variable continua. Función de densidad

Ejemplo: Se ha medido la talla de 40 alumnos de bachillerato y se representan los datos en un histograma de frecuencias relativas con una amplitud de los intervalos de 10, correspondiente a la variable X = "altura en cm" (Fig. A).

Después se talla a 60 alumnos más, representando los datos de las 100 mediciones en un nuevo histograma de amplitud 5 (Fig. B)



Estas gráficas representan las distribuciones de observaciones procedentes de variables aleatorias continuas.

A medida que los intervalos de clase van siendo más pequeños y el tamaño de las muestras es mayor, el polígono de frecuencias se aproxima a una curva continua. La función cuya gráfica es esa curva límite se llama **FUNCIÓN DE DENSIDAD** de la variable X

Las alturas de los rectángulos en los histogramas son mayores o iguales que cero ($f(x) \geq 0$)

La suma de las áreas de los rectángulos es 1. Por tanto:

Una función f(x) es función de densidad de una variable aleatoria continua X, si cumple:

1.- $f(x) \geq 0$ en todo dominio

2.- El área limitada por la gráfica de f(x) y por el eje X es igual a 1

En el caso de variables aleatorias continuas no tiene sentido hablar de probabilidad de un punto, por ser siempre 0 ya que $\int_a^a f(x)dx = 0$; en cambio, tiene interés conocer la probabilidad correspondiente a un intervalo.

Función de distribución

DEFINICIÓN

Sea X una variable aleatoria continua con función de densidad f(x), se define la función de distribución, F(x), como:

Distribuciones continuas. Distribución normal

$$F(x) = p(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

PROPIEDADES

1. Ya que F(x) es el valor de una probabilidad, se verifica que $0 \leq F(x) \leq 1$

2. La función de distribución es nula para todo valor de x anterior al menor valor de la variable aleatoria.

3. La función de distribución F(x) es igual a la unidad para todo valor posterior al mayor de la variable aleatoria.

4. La función de distribución es creciente

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

6. f(x) = F´(x) → La función de densidad es la derivada de la función de distribución

*PARA CALCULAR LA PROBABILIDAD hacemos: $p(a \leq x \leq b) = \int_a^b f(x)dx = F(b) - F(a)$

Parámetros en distribuciones continuas

Media o esperanza matemática

Sea X una variable aleatoria continua con recorrido el intervalo [a, b] y sea f(x) su función de densidad, se llama media al valor de la siguiente integral:

$$\mu = \int_a^b x\, f(x)dx$$

Varianza

$$\sigma^2 = \left(\int_a^b x^2 f(x)dx\right) - \mu^2$$

Desviación típica: raíz cuadrada positiva de la varianza, se representa por "σ"

Distribución normal

Una variable aleatoria continua sigue una **distribución normal** de media μ y desviación típica σ y se designa por **N(μ, σ)** si su función de densidad es:

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

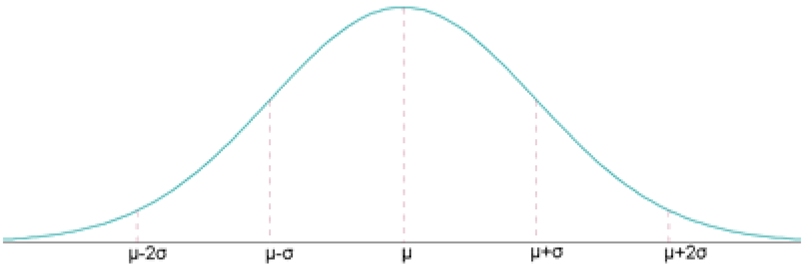
Distribuciones continuas. Distribución normal

Y la variable X puede tomar cualquier valor real.

Donde e = 2,7182….. ; π = 3,1415….. ; x = abscisa, valor cualquiera de la variable ;

μ = media ; σ = desviación típica ; f(x) = ordenada de la curva

CARACTERÍSTICAS DE LA FUNCIÓN DE DENSIDAD (campana de Gauss)



1.- Dominio: es cualquier valor real, es decir, (-∞, +∞).

2.- Simetría: Es simétrica respecto a la media; x = μ.

3.- Cortes con los ejes:

Con el eje X no hay puntos de corte

Con el eje Y: para x = 0 → $f(0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}}$

4.- Asíntotas:

$\lim_{x \to \pm\infty} f(x) = 0$; por tanto el eje X es una asíntota horizontal (y = 0)

5.- Monotonía: La función crece hasta x = μ y decrece desde x = μ

6.- Máximos y mínimos: Tiene un máximo en x = μ.

7.- Curvatura

Convexa e los intervalos $(-\infty, \mu - \sigma)\ y\ (\mu + \sigma, +\infty)$

Cóncava $(\mu - \sigma, \mu + \sigma\ )$

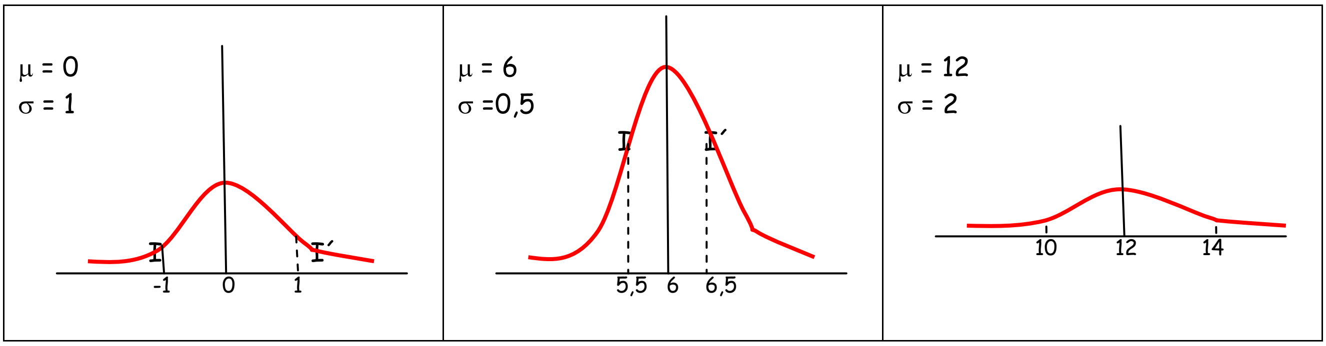
8.- Puntos de inflexión: En los puntos x = $\mu - \sigma$ y x = $\mu + \sigma$ presenta puntos de inflexión.

9.- Área encerrada bajo la curva y el eje X es igual a la unidad, ya que es una función de densidad.

Distribuciones continuas. Distribución normal

Familia de distribuciones normales

La forma de la campana de Gauss depende de los parámetros μ y σ. La media μ determina la situación del centro de la distribución y la desviación típica σ expresa la concentración de los datos de la distribución alrededor de la media. Es evidente que para cada valor de μ y de σ tendremos una función de densidad distinta, como se puede observar en estas gráficas:



Así pues, la expresión **N(μ, σ)** representa una familia de distribuciones normales.

Cuando la desviación típica es elevada, aumenta la dispersión y, en consecuencia, la gráfica es menos estilizada y más abierta.

Por el contrario, para valores de s muy pequeños, la dispersión disminuye y, en consecuencia, la gráfica de la función es mucho más estilizada y concentrada en torno a la media.

Ahora bien, el área encerrada bajo cualquiera de las curvas corresponde a la unidad.

Distribución normal estándar

La **distribución normal estándar, o tipificada o reducida**, es aquella que tiene por **media** el valor **cero, $\mu = 0$**, y por **desviación típica la unidad, $\sigma = 1$**.

Su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Tipificación de la variable

La distribución **N(0, 1)**, que se representa por Z, se encuentra tabulada, lo cual permite un cálculo rápido de las probabilidades asociadas a la misma.

Aunque hay muchos fenómenos que se comportan como una distribución normal, ninguno de ellos se comporta como una N(0, 1).

Distribuciones continuas. Distribución normal

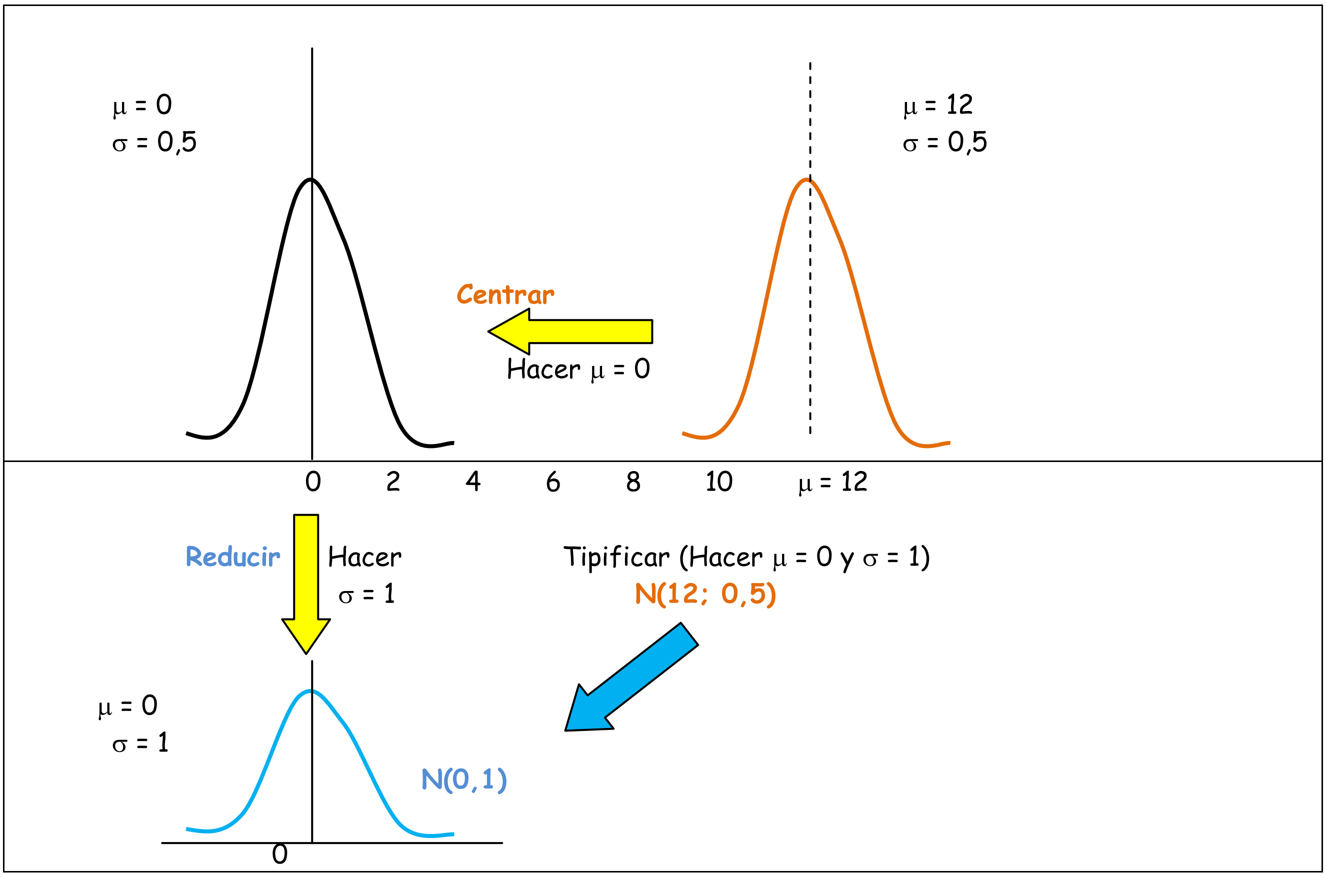
Para resolver el problema, se transforma la variable X que sigue una distribución normal N(μ,σ) en otra variable Z que siga una distribución normal N(0, 1). Esta transformación se conoce como **TIPIFICACIÓN DE LA VARIABLE**.

Para llevar a cabo esta transformación, hay que realizar dos pasos:

1º **CENTRAR**, es decir, trasladar la media de la distribución al origen de coordenadas. Esto equivale a hacer $\mu = 0$

2º **REDUCIR** la desviación típica a 1 ($\sigma = 1$). Esto equivale a **dilatar** o **contraer** la gráfica de la distribución para que coincida con la estándar.

Podemos observar estos dos procesos en el siguiente esquema:



Estos dos pasos se consiguen simultáneamente sólo con hacer el siguiente cambio de variable:

X: N(μ, σ) → Z: N(0, 1)

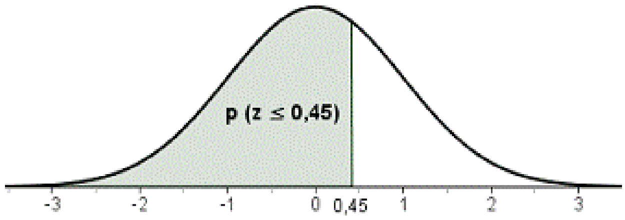
Cambio de variable $$Z = \frac{X - \mu}{\sigma}$$

Distribuciones continuas. Distribución normal

Uso de tablas

1 Cálculo de P (Z ≤ a)

Cuando la probabilidad pedida se encuentra directamente en las tablas



Ejemplo: Hallar la probabilidad **P (Z ≤ 0,45)**

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

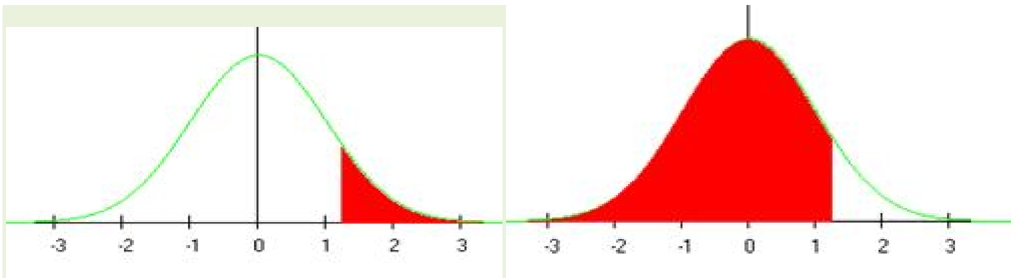
En la 1ª columna buscamos el valor de las unidades y las décimas.

En la 1ª fila el valor de las centésimas.

Basta buscar 0,4 en la columna y 0,05 en la fila. Su intersección nos da la probabilidad.

Leemos y nos da **0,6736. La probabilidad P (Z ≤ 0,45) = 0,6736**

2 Cálculo de P (Z > a)



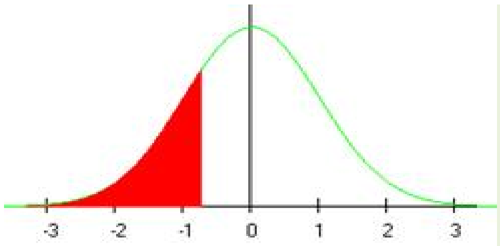
Ejemplo: **Probabilidad de un valor positivo P (Z > 1,24)**

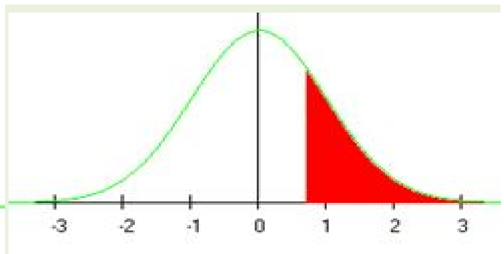
En este caso la probabilidad pedida no está en las tablas.

Sin embargo, si tenemos en cuenta que el área total bajo la gráfica ha de ser 1, deducimos de la figura que:

$p(z > 1{,}24) = 1 - p(z \leq 1{,}24) = 1 - 0{,}8925 = 0{,}1075$

3 Cálculo de P (Z $\leq$ -a)





Ejemplo: Probabilidad de un valor negativo p (Z $\leq$-0,72)

Como la gráfica es simétrica respecto al eje de ordenadas, $p(Z \leq -0{,}72) = p(Z \geq +0{,}72)$

Calculamos $p(Z \geq +0{,}72)$ igual que en el caso 2.

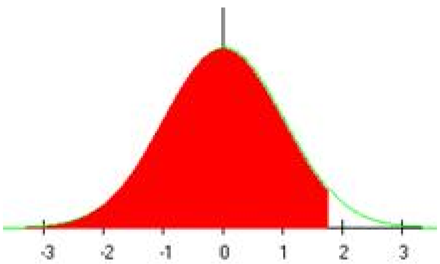
$p(Z \geq +0{,}72) = 1 - p(Z < +0{,}72) = 1 - 0{,}7642 = 0{,}2358$

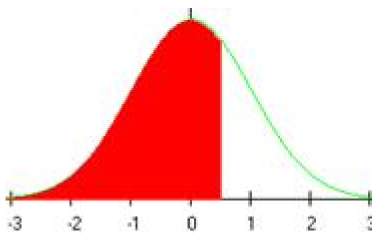
$p(Z \leq -0{,}72) = p(Z \geq +0{,}72) = 1 - p(Z < +0{,}72) = 1 - 0{,}7642 = 0{,}2358$

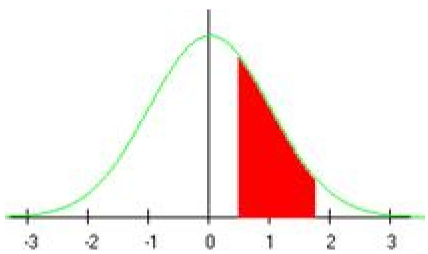
4 Probabilidad entre dos valores positivos p (0,5 $\leq$ z $\leq$ 1,76)

Leemos directamente en la tabla la $p(Z \leq 1{,}76)$ y la $p(Z \leq 0{,}5)$.

La diferencia entre ellas es la probabilidad que nos piden.



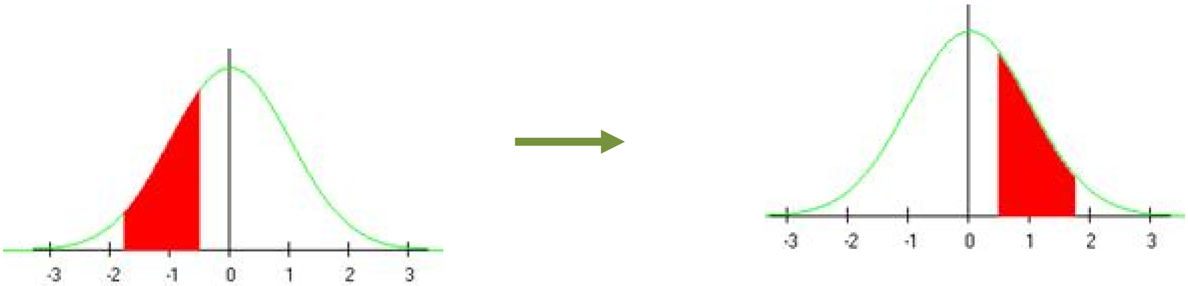




$p(0{,}5 \leq Z \leq 1{,}76) = p(Z \leq 1{,}76) - p(Z \leq 0{,}5) = 0{,}9608 - 0{,}6915 = 0{,}2693$

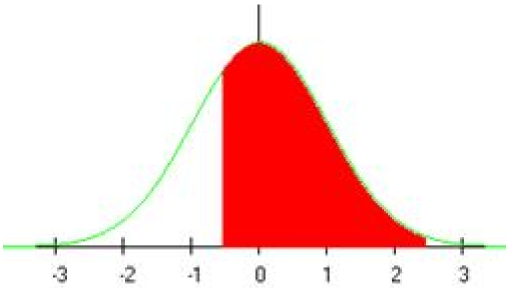
5 Probabilidad entre dos valores negativos p (- 1,76 $\leq$ Z $\leq$ - 0,5)

Por simetría cambiamos los dos valores negativos a positivos y calculamos sus probabilidades.



p (- 1,76 ≤ Z≤ - 0,5) = p (0,5 ≤ Z ≤ 1,76) = 0,9608 - 0,6915 = 0,2693

6 Probabilidad entre un valor positivo y uno negativo p(- 0,53 ≤ z ≤ 2,46)



p(- 0,53 ≤ Z≤ 2,46) = p (Z ≤ 2,46) - p (Z ≤ - 0,53)

p (Z ≤ - 0,53) = p (Z ≥ 0,53) = 1 - p (Z < 0, 53)= 1 - 0,7019 = 0,2981 (Caso **3**)

p(- 0,53 ≤ Z ≤ 2,46) = p (Z ≤ 2,46) - p (Z ≤ - 0,53) = 0,9931 - 0,2981 = 0,695

Aproximación de la Binomial por una Normal

TEOREMA DE "DE MOIVRE"

Una distribución binomial ***B(n, p)*** se puede aproximar por una distribución normal, siempre que *n* sea grande y p constante. La aproximación consiste en utilizar una distribución normal con la misma media y desviación típica que la distribución binomial.

En la práctica se considera la aproximación buena cuando: $\boldsymbol{n \cdot p \geq 5 \quad y \quad n \cdot q \geq 5}$

En cuyo caso: $\boldsymbol{X: B(n, p) \rightarrow X': N(\mu, \sigma)}$

$$siendo \ \mu = n \cdot p \ \ y \ \ \sigma = \sqrt{n \cdot p \cdot q}$$

Y tipificando se obtiene la normal estándar correspondiente:

$$\boldsymbol{X': N(\mu, \sigma) \rightarrow Z: N(0, 1)}$$
$$Z = \frac{X' - \mu}{\sigma}$$

Distribuciones continuas. Distribución normal

CORRECCIÓN POR CONTINUIDAD

La distribución binomial es de variable aleatoria discreta y, por tanto, tiene sentido calcular probabilidades puntuales. En cambio, la distribución normal es de variable aleatoria continua, y no tiene sentido calcular probabilidades puntuales, pues son todas nulas.

La aproximación de una variable discreta, X, por una continua, X´, genera un error que se corrige modificando el intervalo cuya probabilidad se quiere calcular:

Binomial = Normal	Ejemplos
$p\ (X = a) = p\ (a - 0{,}5 < X' \leq a + 0{,}5)$	$p\ (X = 13) = p\ (13 - 0{,}5 < X' \leq 13 + 0{,}5)$
$p\ (a \leq X \leq b) = p\ (a - 0{,}5 < X' \leq b + 0{,}5)$	$p\ (10 \leq X \leq 13) = p\ (10 - 0{,}5 < X' \leq 13 + 0{,}5)$
$p\ (X < a) = p\ (X' \leq a - 0{,}5)$	$p\ (X < 11) = p\ (X' \leq 11 - 0{,}5)$
$p\ (X \leq a) = p\ (X' \leq a + 0{,}5)$	$p\ (X \leq 11) = p\ (X' \leq 11 + 0{,}5)$
$p\ (X > a) = p\ (X' \leq a + 0{,}5)$	$p\ (X > 12) = p\ (X' \leq 12 + 0{,}5)$
$p\ (X \geq a) = p\ (X' \leq a - 0{,}5)$	$p\ (X \geq 12) = p\ (X' \leq 12 - 0{,}5)$

Problemas resueltos

VARIABLE CONTINUA

1.- En un taller de piezas de un motor, una vez terminada la pieza, se analiza la medida (altura) para mirar si hay algún error, siendo el error máximo permitido de 3 mm. La función de densidad es:

$$f(x) = \begin{cases} 0 & si \quad x < 0 \\ 1/3 & si \quad 0 \leq x \leq 3 \\ 0 & si \quad x > 3 \end{cases}$$

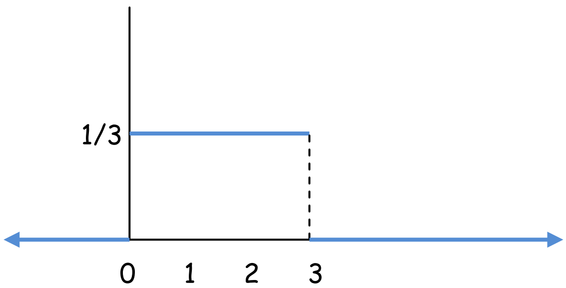
a) Comprueba que es una función de densidad y represéntala.

b) Calcula la función de distribución y represéntala.

c) Calcula la probabilidad de que el error de fabricación esté entre 1 y 2 mm.

Solución:

a)



Distribuciones continuas. Distribución normal

Para que sea función de densidad → $\int_{-\infty}^{+\infty} f(x) = 1$ (El área limitada por la gráfica de f(x) y por el eje X es igual a 1)

$$\int_{-\infty}^{0} 0\,dx + \int_{0}^{3} \frac{1}{3}dx\ + \int_{3}^{+\infty} 0\,dx = \left[\frac{x}{3}\right]_0^3 = \frac{3}{3} - \frac{0}{3} = 1$$

b) Función de distribución?

Calculo una primitiva

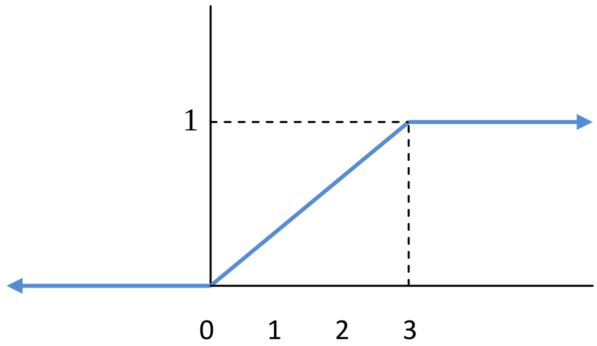
$$\int_0^3 \frac{1}{3}dx = \frac{x}{3} + C$$

La función de distribución será:

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ x/3 & si \quad 0 \le x \le 3 \\ 1 & si \quad x > 3 \end{cases}$$

(La derivada de esta función F(x) me da la función de densidad f(x))

Representación:



c) Para calcular la probabilidad se puede hacer de dos formas:

1.- A partir de la función de densidad

$$p(1 \le X \le 2) = \int_1^2 f(x)dx = \int_1^2 \frac{1}{3}dx = \left[\frac{x}{3}\right]_1^2 = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$

2.- A partir de la función de distribución (en la que ya está calculada la integral)

$$p(1 \le X \le 2) = p(X \le 2) - p(X \le 1) = F(2) - F(1) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$

DISTRIBUCIÓN NORMAL

2.- El peso de los bebés al nacer se distribuye normalmente con media 3,2 kg y desviación típica de 0,5 kg. Calcular las siguientes probabilidades:

a) Que el peso de un bebé esté entre 2,7 y 3,7 kg (ambos incluidos).

b) Que el peso de un bebé sea mayor de 4 kg.

c) Que el peso de un bebé sea menor de 3 kg.

Solución:

$$X: N(3{,}2\ ;0{,}5) \to Z: N(0,1)$$

$$Tipifico: Z = \frac{X-\mu}{\sigma}$$

a) $p(2{,}7 \le X \le 3{,}7) = p\left(\frac{2{,}7-3{,}2}{0{,}5} \le Z \le \frac{3{,}7-3{,}2}{0{,}5}\right) = p(-1 \le Z \le 1) = p(Z \le 1) - p(Z \le -1) =$

$$= 0{,}8413 - (1 - 0{,}8413) = 0{,}6826$$

b) $p(X > 4) = p\left(Z > \frac{4-3{,}2}{0{,}5}\right) = p(Z > 1{,}6) = 1 - p(Z \le 1{,}6) = 1 - 0{,}9452 = 0{,}0548$

c) $p(X < 3) = p\left(Z < \frac{3-3{,}2}{0{,}5}\right) = p(Z < -0{,}4) = p(Z > 0{,}4) = 1 - p(Z \le 0{,}4) = 1 - 0{,}6554 = 0{,}3446$

APROXIMACIÓN DE LA BINOMIAL POR LA NORMAL

3.- Después de realizar varios sondeos sobre cierta población, se ha conseguido averiguar que únicamente el 15% de la misma es favorable a los tratamientos de psicoterapia. Elegida al azar una muestra de 50 personas de dicha población, se desea saber:

a) La probabilidad de que haya más de 5 personas favorables a dichos tratamientos.

b) La probabilidad de que como mucho haya 6 personas favorables.

Solución:

Distribución binomial B (50, 0,15)
1.- A= "ser favorable" y $\bar{A}$= " no ser favorable"
2.- Que una persona sea favorable, es independiente de que otra lo sea
3.- p = 0,15 (constante)

Se cumple que n·p $\geq$ 5 y n·q $\geq$ 5

50 · 0,15 $\geq$ 5 y 50 · 0,85 $\geq$ 5 Aplicamos el ajuste de la variable y después se tipifica

$$X: B(50;\ 0{,}15) \to X'(7{,}5; 2{,}52) \to Z(0,1)$$

Ya que $\mu = n \cdot p = 50 \cdot 0{,}15 = 7{,}5$; $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{50 \cdot 0{,}15 \cdot 0{,}85} = 2{,}52$

a) $p(X > 5) = p(X' > 5 + 0{,}5) = p(Z > \frac{5{,}5-7{,}5}{2{,}52} = p(Z > -0{,}79) = p(Z \le 0{,}79) = 0{,}7852$

b) $p(X \le 6) = p(X' \le 6 + 0{,}5) = p(Z \le \frac{6-7{,}5}{2{,}52} = p(Z \le -0{,}4) = p(Z > 0{,}4) = 1 - p(Z \le 0{,}4) = 1 - 0{,}6554 = 0{,}3446$