Teoría de la Probabilidad. Distribuciones de probabilidad.

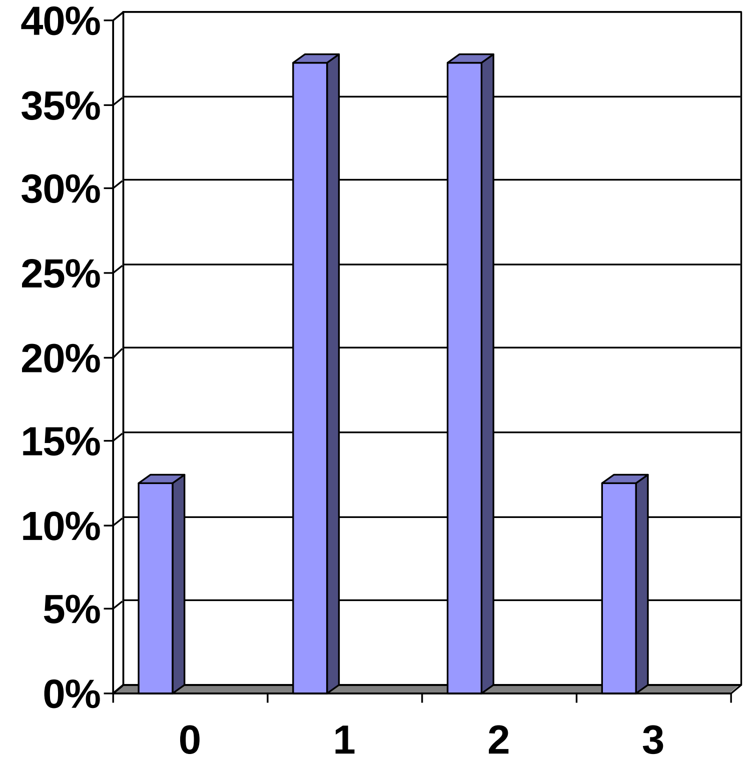
Variable aleatoria

- El resultado de un experimento aleatorio puede ser descrito en ocasiones como una cantidad numérica.
- En estos casos aparece la noción de variable aleatoria
  - Función que asigna a cada suceso un número.
- Las variables aleatorias pueden ser discretas o continuas.

Función de probabilidad (Variable Discreta)

- Asigna a cada posible valor de una variable discreta su probabilidad.

- Ejemplo
  - Número de caras al lanzar 3 monedas.



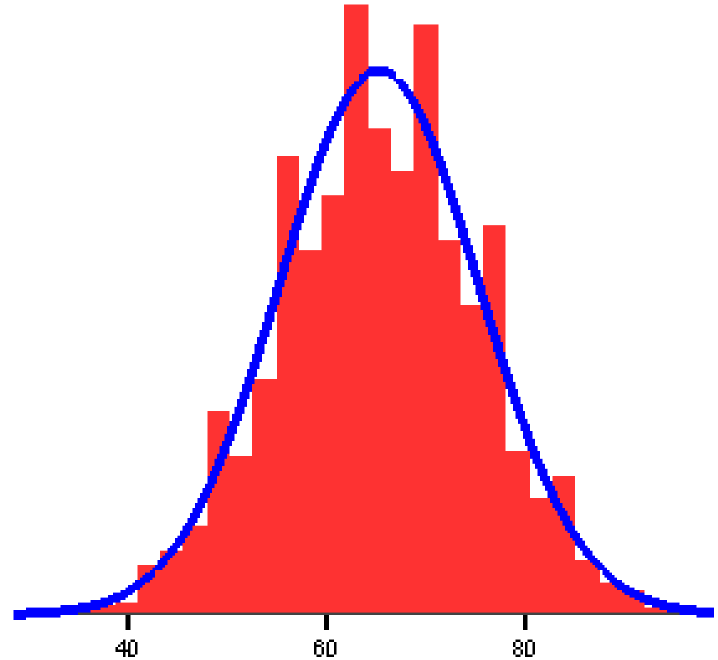
Función de densidad o de probabilidad de una **variable discreta X**: es la función que asigna a cada valor que puede tomar la variable, la probabilidad con la que eso sucede. Se puede expresar mediante una fórmula f(x), o mediante una tabla. La función de densidad cumple:

1) f(x)≥0 para todo valor que pueda tomar la variable.

2) $\sum_{x=-\infty}^{\infty} f(x) = 1$

Función de densidad (Variable Continua)

- Definición
  - Es una función no negativa de integral 1.
    - Es como la generalización de un histograma con frecuencias relativas para variables continuas.



- ¿Para qué se utiliza?
  - Nunca se utiliza directamente.
  - Sus valores no representan probabilidades.

Función de densidad: Dada una **variable aleatoria continua** X, decimos que f(x) es una función de densidad, si la probabilidad de que X tome valores en el intervalo (a,b) es igual al área encerrada por la gráfica de f(x), el eje x y las rectas x=a, x=b. Se cumplen que f(x)≥0 para todo valor de x

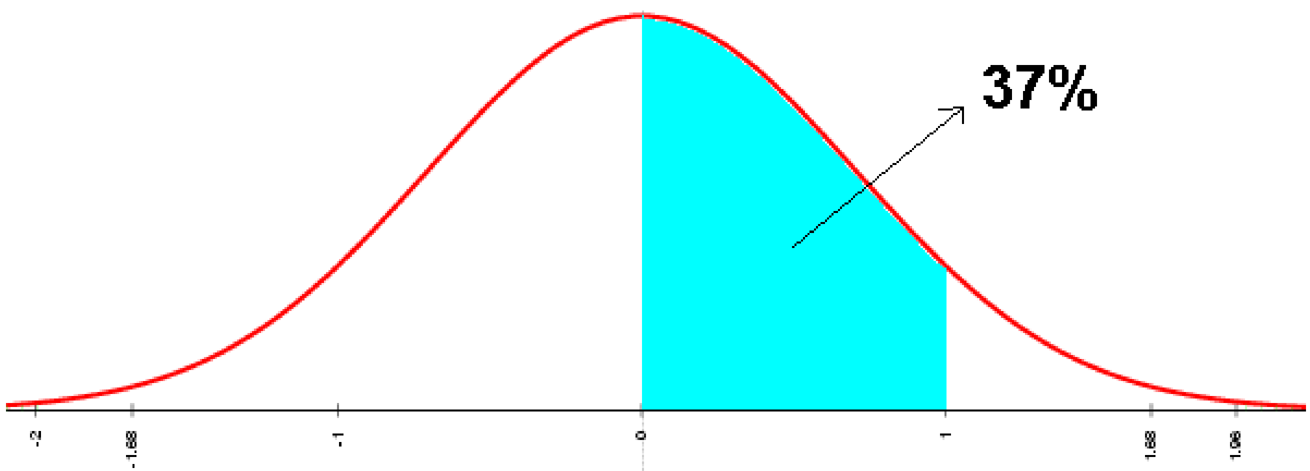
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

¿Para qué sirve la función de densidad?

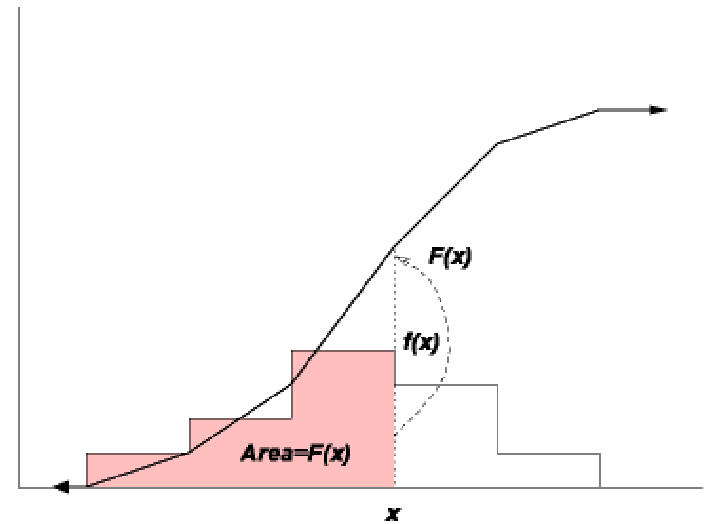
- Muchos procesos aleatorios vienen descritos por variables de forma que son conocidas las probabilidades en intervalos.

- La integral definida de la función de densidad en dichos intervalos coincide con la probabilidad de los mismos.

- Es decir, identificamos la probabilidad de un intervalo con el **área** bajo la función de densidad.



Función de distribución



- Es la función que asocia a cada valor de una variable, la probabilidad acumulada de los valores inferiores o iguales.
  - Es como la generalización de las frecuencias acumuladas.
    - A los valores extremadamente bajos les corresponden valores de la función de distribución cercanos a cero.
    - A los valores extremadamente altos les corresponden valores de la función de distribución cercanos a uno.

La **función de distribución** de una **variable discreta** X es la función que asigna a cada valor que puede tomar la variable, la probabilidad de que tome ese valor, o cualquier valor inferior.

$$F(x) = \sum_{k \le x} P(X = k)$$

¿Para qué sirve la función de distribución?

- Contrastar lo anómalo de una observación concreta.
    - Una persona de altura 210cm es “anómala” porque la función de distribución en 210 es muy alta.
    - Una persona adulta que mida menos de 140cm es “anómala” porque la función de distribución es muy baja para 140cm.
    - Una persona que mida 170cm no posee una altura nada extraña pues su función de distribución es aproximadamente 0,5.
- En un contexto de contrastes de hipótesis, se pueden observar unos resultados experimentales y contrastar lo “anómalos” que son en conjunto con respecto a una hipótesis de terminada.

Valor esperado y varianza de una variable aleatoria X

- Valor esperado
  - Se representa mediante E[X] ó μ
  - Es el equivalente a la media

- Varianza
  - Se representa mediante VAR[X] o σ^2
  - Es el equivalente a la varianza
  - Se llama desviación típica a σ

DISTRIBUCIÓN BINOMIAL

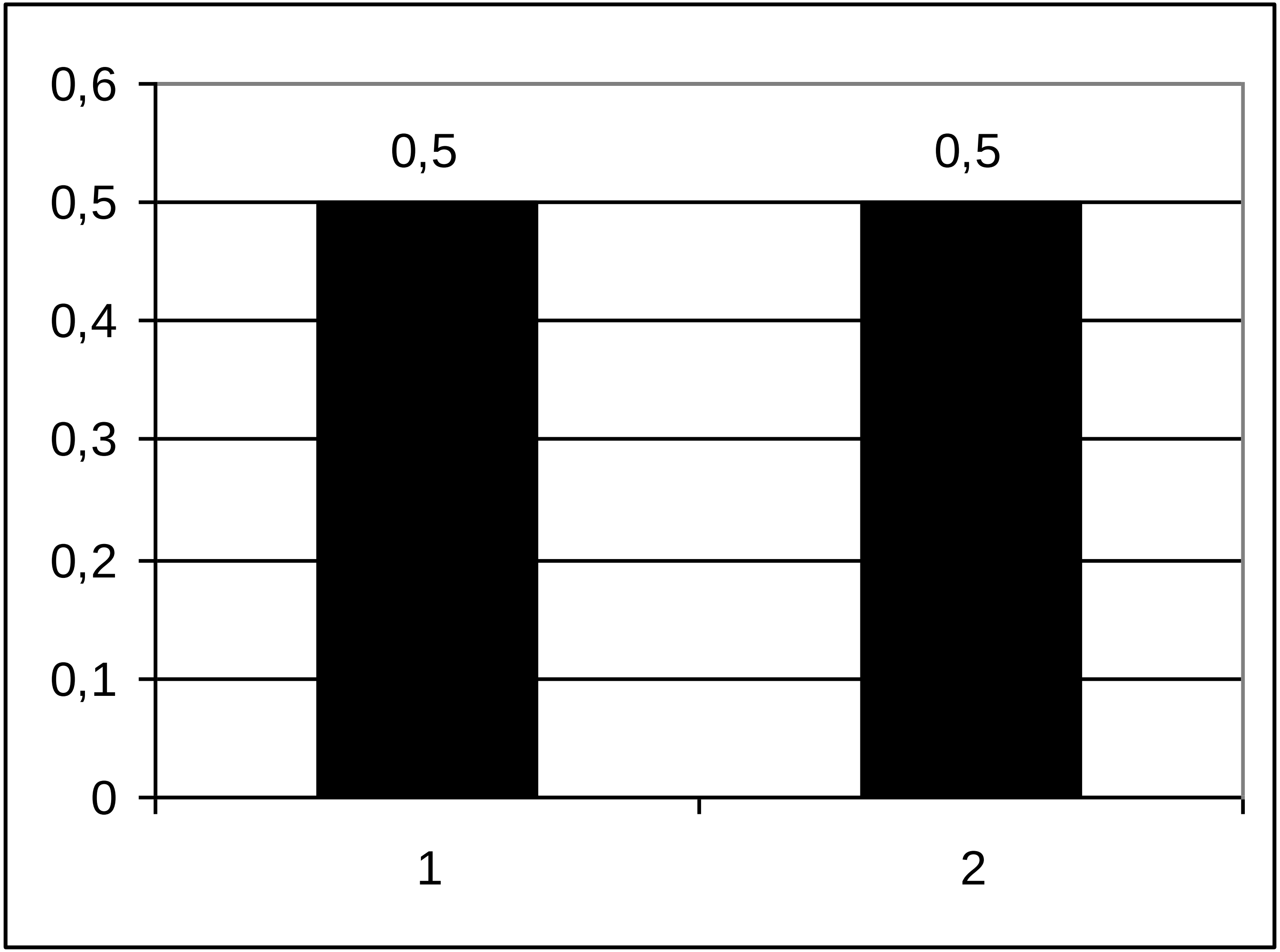
Se representa como B(n, p)

CINCO CONDICIONES (supuestos de Bernoulli):

1) Existe una serie de **n** experimentos.
2) En cada experimento hay sólo dos posibles resultados.
3) En cada experimento, los dos resultados posibles son mutuamente excluyentes (una probabilidad **p** y la otra **q=1-p**).
4) Los resultados de cada experimento son independientes entre sí.
5) La probabilidad de cada resultado posible en cualquier ensayo es la misma de un ensayo a otro (siempre **p** y **q=1-p**).

DISTRIBUCIÓN BINOMIAL

Gráficos [1 moneda, p=0,50] (simetría, normalidad)



DISTRIBUCIÓN BINOMIAL

Dos monedas:

¿Posibles resultados?

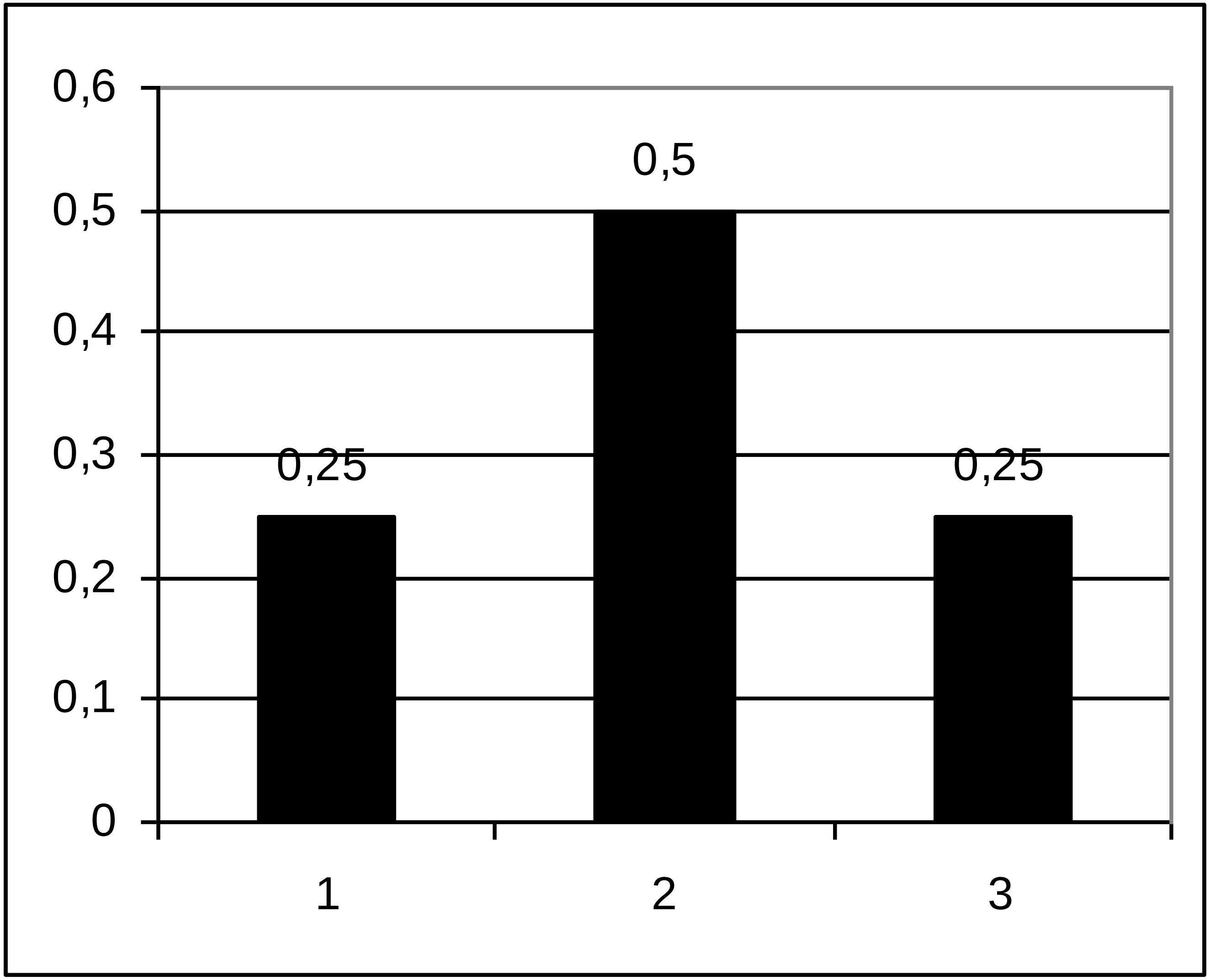
$$p(2\ caras) = p(cara\ cara) = \frac{Número\ de\ resultados\ clasificables\ como\ 2\ caras}{Cantidad\ total\ de\ resultados} = \frac{1}{4} = 0.25$$

$$p(1\ cara) = p(cara\ cruz\ o\ cruz\ cara) = \frac{Número\ de\ resultados\ clasificables\ como\ 1\ cara}{Cantidad\ total\ de\ resultados} = \frac{2}{4} = 0.5$$

$$p(0\ caras) = p(cruz\ cruz) = \frac{Número\ de\ resultados\ clasificables\ como\ 0\ caras}{Cantidad\ total\ de\ resultados} = \frac{1}{4} = 0.25$$

DISTRIBUCIÓN BINOMIAL

Gráficos [2 monedas, p=0,50] (simetría, normalidad)



DISTRIBUCIÓN BINOMIAL

Tres monedas:

$$p(3 \text{ caras}) = \frac{1}{8} = 0{,}1250$$

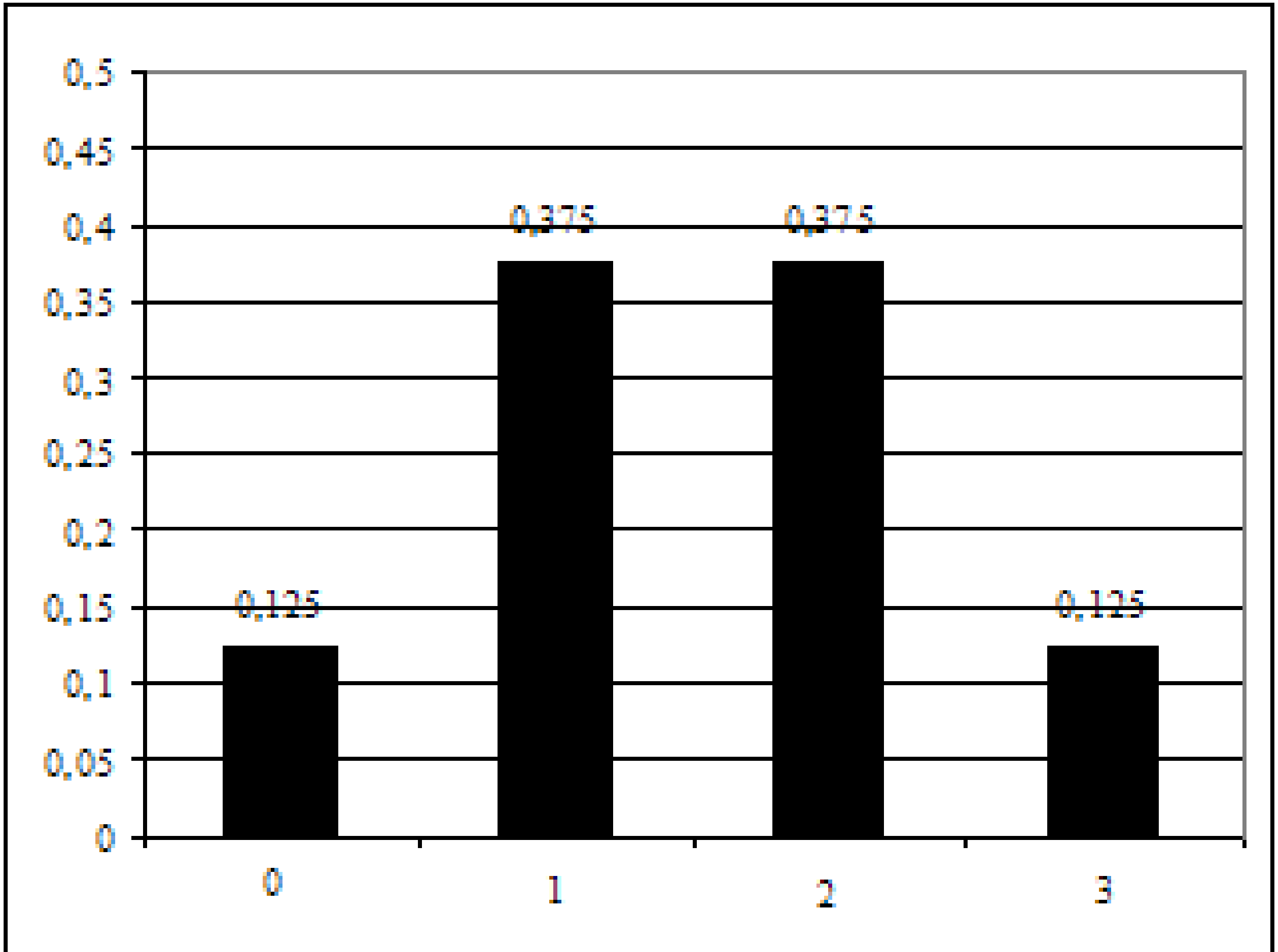
$$p(2 \text{ caras}) = \frac{3}{8} = 0{,}3750$$

$$p(1 \text{ cara}) = \frac{3}{8} = 0{,}3750$$

$$p(0 \text{ caras}) = \frac{1}{8} = 0{,}1250$$

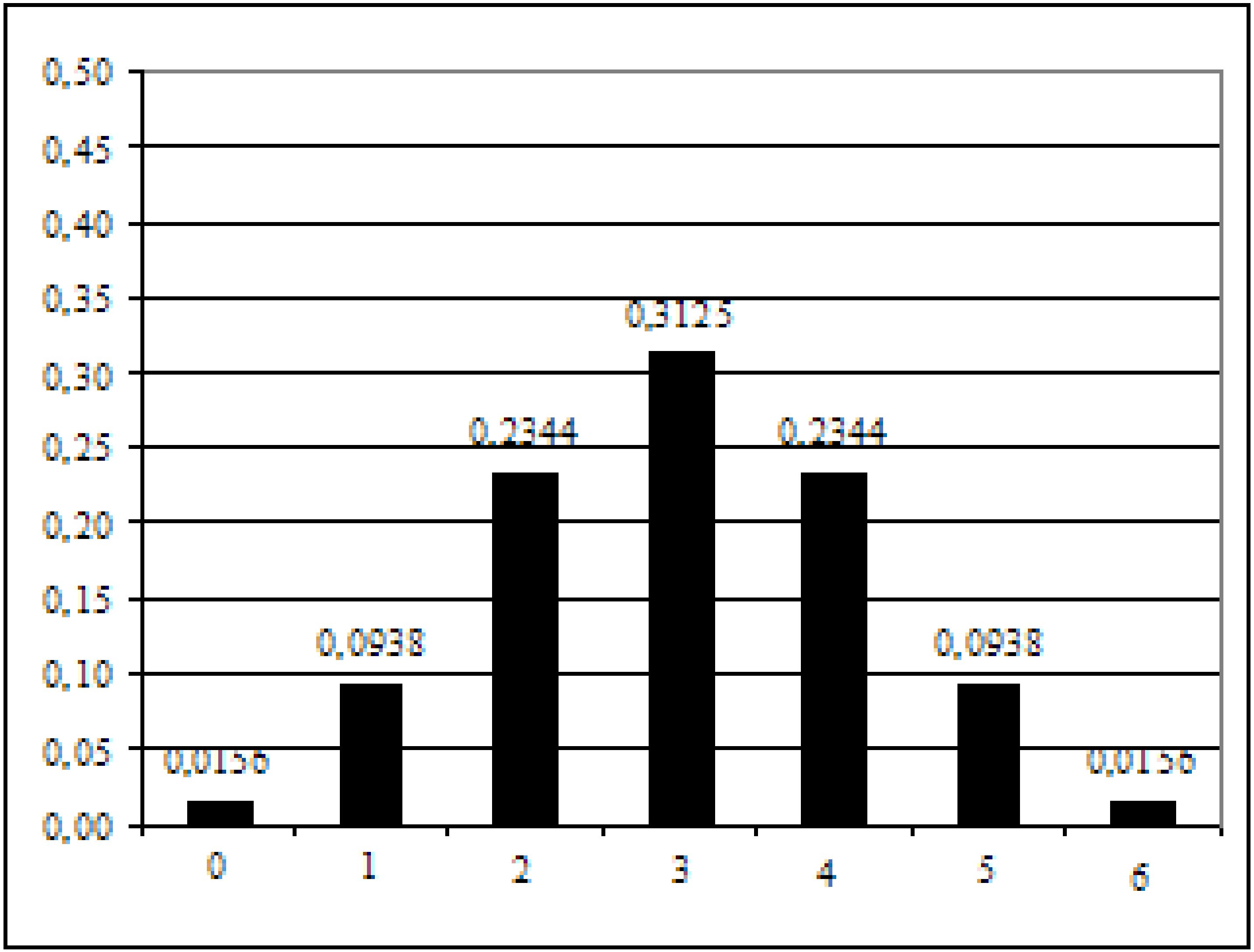
DISTRIBUCIÓN BINOMIAL

Gráficos [3 monedas, p=0,50] (simetría, normalidad)



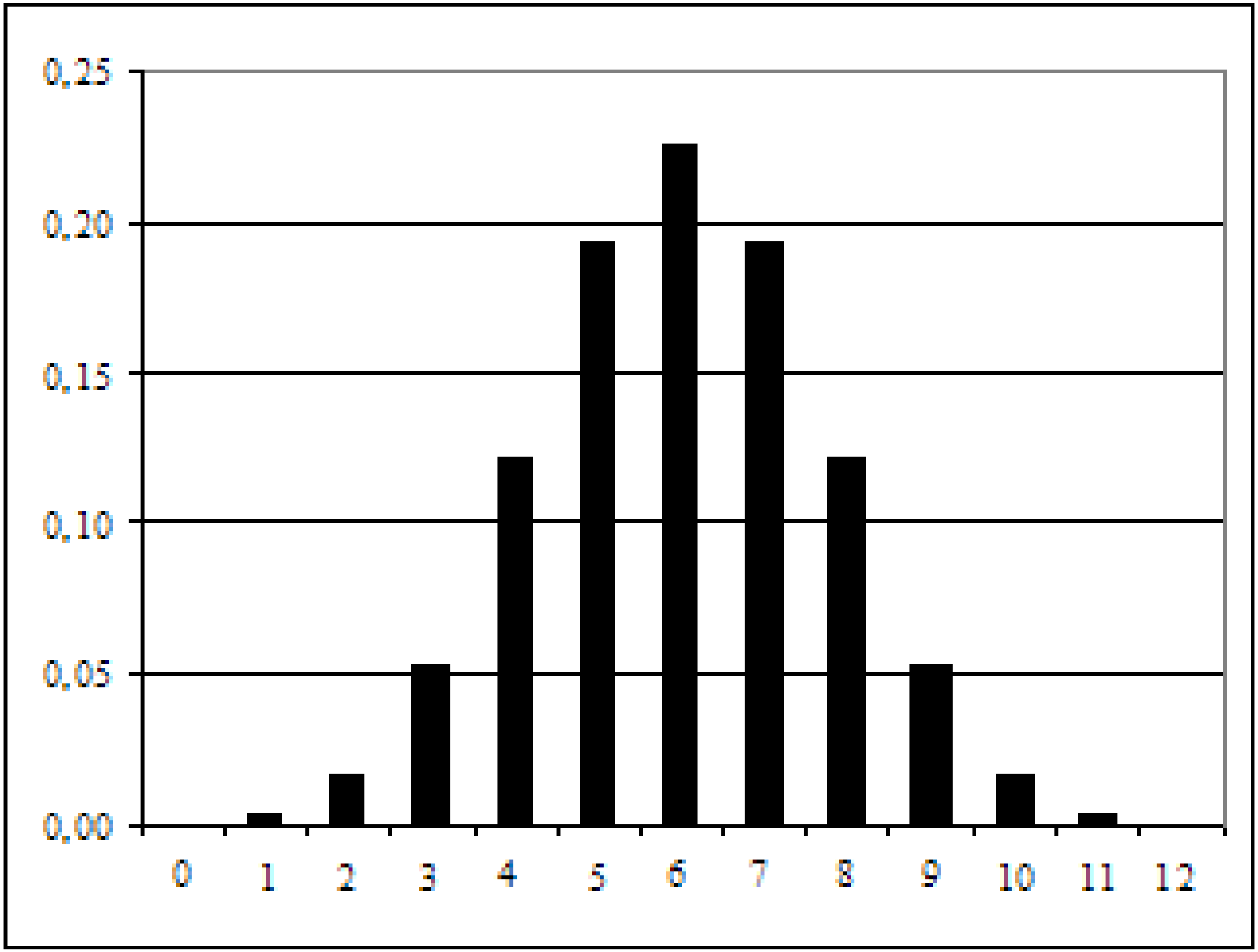
DISTRIBUCIÓN BINOMIAL

Gráficos [6 monedas, p=0,50] (simetría, normalidad)



DISTRIBUCIÓN BINOMIAL

Gráficos [12 monedas, p=0,50] (simetría, normalidad)



DISTRIBUCIÓN BINOMIAL

$$p(r,n) = \frac{n!}{(n-r)!\,r!} p^r q^{n-r}$$

n = número de intentos o experimentos
r = número de éxitos obtenidos
p = probabilidad a favor
q = probabilidad en contra (q = 1-p)

Media y Varianza de una distribución binomial B(n, p)

1) Media $\mu = n \cdot p$
2) Varianza $\sigma^2 = n \cdot p \cdot q$

DISTRIBUCIÓN BINOMIAL

$$p(r,n) = \frac{n!}{(n-r)!\,r!} p^r q^{n-r}$$

Media y Varianza de una distribución binomial B(n, p)

1) Media $\mu = n \cdot p$
2) Varianza $\sigma^2 = n \cdot p \cdot q$

La distribución binomial nos proporciona el número de éxitos, pero no el orden en el que suceden

DISTRIBUCIÓN BINOMIAL. EJEMPLOS

1) El 2% de los DVDs de una determinada marca son defectuosos. Si se venden lotes de 25 unidades, calcular la probabilidad de que haya como máximo dos defectuosos.

Es una distribución binomial con n = 25, p = 0,02, q = 1 - p = 0,98.

X es B(25 , 0,02), por lo que:

$$p(\mathrm{X} \le 2) = p(\mathrm{X}=0) + p(\mathrm{X}=1) + p(\mathrm{X}=2) =$$

$$= \binom{25}{0} 0,02^0 \cdot 0,98^{25} + \binom{25}{1} 0,02^1 \cdot 0,98^{24} + \binom{25}{2} 0,02^2 \cdot 0,98^{23} =$$

$$= 1 \cdot 1 \cdot 0,6035 + 25 \cdot 0,02 \cdot 0,6158 + \frac{25 \cdot 24}{2} 0,0004 \cdot 0,6283 =$$

$$= 0,6035 + 0,3079 + 0,0754 = 0,9868$$

DISTRIBUCIÓN BINOMIAL. EJEMPLOS

2) Se supone que la probabilidad de nacer niño es del 0,50. Calcula la probabilidad de que en una familia de seis hijos sean:

a) Todos varones.
b) Al menos, dos varones.
c) Tres varones.
d) Calcula la media y la desviación típica.

DISTRIBUCIÓN BINOMIAL. EJEMPLOS

3) Un tirador acierta en el 95% de las veces. Si realiza 7 lanzamientos, ¿cuál es la probabilidad de que falle alguno?

DISTRIBUCIÓN DE POISSON

- Es una distribución **discreta** que se obtiene como aproximación de una distribución binomial con la misma media, para 'n grande' ($n>30$) y 'p pequeño' ($p<0,1$).
- Describe la probabilidad de encontrar exactamente *k* eventos, en un lapso de tiempo, si los acontecimientos se producen de forma independiente a una velocidad constante . Es una de las distribuciones más utilizadas en estadística con varias aplicaciones como: describir el número de fallos en un lote de materiales o la cantidad de llegadas por hora a un centro de servicios.
- Queda caracterizada por un único parámetro μ (que es a su vez su media y varianza.)
- Función de probabilidad:

$$P[X = k] = e^{-\mu} \frac{\mu^k}{k!}, \ k = 0,1,2,\ldots$$

DISTRIBUCIÓN DE POISSON. EJEMPLOS

- El número de individuos que será atendido un día cualquiera en el servicio de urgencias de un hospital.
  - En Alicante hay 300.000 habitantes (n grande)
  - La probabilidad de que cualquier persona tenga un accidente es pequeña, pero no nula. Supongamos que es 1/10.000
  - **B**(n=300.000,p=1/10.000) **≈ Poisson**(μ=np=30)

- Sospechamos que diferentes hospitales pueden tener servicios de traumatología de diferente "calidad". Es difícil compararlos pues cada hospital atiende poblaciones de tamaños diferentes (ciudades, pueblos,…)
  - Tenemos en cada hospital n, nº de pacientes atendidos o nº individuos de la población que cubre el hospital.
  - Tenemos p pequeño calculado como frecuencia relativa de secuelas con respecto al total de pacientes que trata el hospital, o el tamaño de la población,…
  - Se puede modelar mediante Poisson(μ=np)

DISTRIBUCIÓN POISSON. EJEMPLOS

1) Una empresa electrónica observa que el número de componentes que fallan antes de cumplir 100 horas de funcionamiento es una variable aleatoria de Poisson. Si el número promedio de estos fallos es ocho,

a. ¿cuál es la probabilidad de que falle un componente en 25 horas?

b. ¿y de que fallen no más de dos componentes en 50 horas?

c. ¿cuál es la probabilidad de que fallen por lo menos diez en 125 horas?

a) Considerando que se cumplen ciertas condiciones de regularidad, podemos asumir que una variable η que mide el número de componentes que fallan antes de cumplir 25 horas de funcionamiento sigue una distribución de Poisson con parámetro μ=8/4=2

$$P = (\eta = 1) = \frac{2^1}{1!} * e^{-2} = 0{,}27067$$

b) Análogamente, definimos una variable aleatoria U con distribución de Poisson de parámetro μ=8/2=4, que mide el número de componentes que fallan antes de cumplir las 50 horas de funcionamiento. Se tiene entonces que:

$$P(U \leq 2) = \sum_{i=0}^{2} \frac{4^i}{i!} * e^{-4} = 0{,}2381$$

DISTRIBUCIÓN POISSON. EJEMPLOS

c) De la misma forma, definiendo una variable aleatoria V con distribución de Poisson de Parámetro $\mu=10$

$$P(V \geq 10) = 1 - P(V < 10) = 1 - \sum_{i=0}^{10} \frac{10^i}{i!} * e^{-10} = 0{,}41696$$

DISTRIBUCIÓN GEOMÉTRICA

- Es una distribución discreta que cumple con los supuestos de Bernoulli.
- Expresa la probabilidad de tener que esperar exactamente *k* pruebas hasta encontrar el primer éxito si la probabilidad de éxito en una sola prueba es *p*.
- Los dos parámetros que caracterizan la distribución son
  - $\mu = \frac{1}{p}$; $\sigma^2 = \frac{(1-p)}{p^2}$
- La función de probabilidad

$$P(X = k) = (1-p)^{k-1}p$$

$$P(X \leq k) = 1 - (1-p)^k$$

DISTRIBUCIÓN GEOMÉTRICA. EJEMPLOS

1) Un paciente sufre insuficiencia renal y necesita un trasplante de un donante adecuado. La probabilidad de que un donante aleatorio cumpla los requisitos de este paciente es 0,2.

a) Supongamos que ningún donante cumple los requisitos del paciente hasta que llega un quinto donante. ¿Cuál es la probabilidad en este caso?

b) Halla la probabilidad de que el paciente necesite 10 donantes o menos hasta que se encuentre un donante compatible.

c) ¿Cuál es el número esperado de donantes necesarios para conseguir un donante compatible?

d) Halla la varianza.

SOLUCIÓN:

a) $P(X=k)=(1-p)^{k-1}p=(1-0{,}2)^{k-1}\cdot 0{,}2;\ P(X=5)=(1-0{,}2)^{5-1}\cdot 0{,}2=0{,}08192$

b) $P(X\leq 10)=1-(1-0{,}2)^{10}=0{,}898625$

c) $\mu=\frac{1}{p}=\frac{1}{0{,}2}=5$

d) $\sigma^2=\frac{(1-p)}{p^2}=\frac{(1-0{,}2)}{0{,}2^2}=\frac{0{,}8}{0{,}04}=20$

DISTRIBUCIÓN GEOMÉTRICA. EJEMPLOS

La probabilidad de una alineación óptica con éxito en el montaje de un producto óptico, según especificaciones para su almacenamiento, es de 0.8. Supongamos que los intentos de montaje son independientes.

a. ¿Cuál es la probabilidad de que la alineación con éxito se produzca por primera vez en el quinto ensayo?

b. ¿Cuál es la probabilidad de que la primera alineación con éxito requiera como mínimo cuatro intentos?

c. ¿Cuál es la probabilidad de que la primera alineación con éxito requiera a lo menos cuatro y a lo más 6 intentos?

d. ¿Cuántos intentos se espera realizar para lograr el primer éxito?

e. Calcular la desviación estándar del número de intentos hasta lograr éxito en la alineación.

Esquema de solución:

Paso 1. Identificar el experimento:

Estudiar el número de intentos hasta lograr realizar con éxito la alineación de un producto óptico en su montaje.

Paso 2. Identificar el suceso E que se desea estudiar y que determina el éxito en la experiencia, y su contrario, E^C:

E : Se realiza con éxito la alineación de un producto óptico en el montaje.

E^C : No se realiza con éxito la alineación de un producto óptico en el montaje.

Paso 3. Identificar las probabilidades de E y E^C:

$$P(E) = p = 0.8 \qquad P(E^C) = p = 0.2\,.$$

Paso 4. Existe independencia entre los intentos de "alineación del producto óptico en el montaje" .

Paso 5. Identificación de la variable aleatoria:

X: Número de intentos hasta lograr realizar con éxito la alineación de un producto óptico en su montaje.

Paso 6. Deducción de la distribución: $X \sim Geométrica(0.8)$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. $P[X=5]=(0.2)^4(0.8)=0.0013$

Resp.: La probabilidad de que la alineación se produzca con éxito por primera vez en el quinto ensayo es de 0.0013 (1.3%).

b. $P[X\geq 4]=1-P[X\leq 3]=1-P[X=1]-P[X=2]-P[X=3]$

$$=1-(0.2)^0(0.8)-(0.2)^1(0.8)-(0.2)^2(0.8)=1-0.992=0.08$$

Resp.: La probabilidad de que la primera alineación con éxito requiera como mínimo cuatro intentos es de 0.08 (8.00%).

c. $P[4\leq X\leq 6]=P[X\leq 6]-P[X\leq 3]=F_X(6)-F_X(3)$

$$=0.999936-0.992=0.007936$$

Esta forma de solución necesita de la Función de distribución acumulativa. Otra forma más simple es:

$$P[4\leq X\leq 6]=P[X=4]+P[X=5]+P[X=6]=$$
$$=0.9984+0.99968+0.999936=0.007936$$

Resp.: La probabilidad de que la primera alineación con éxito requiera a lo menos cuatro y a lo más 6 intentos 0.0079 aprox. (7.9%, casi 8%).

d. $E(X)=\dfrac{1}{p}=\dfrac{1}{0.8}=1.25$ intentos.

Resp.: Se espera realizar entre 1 y 2 intentos para lograr realizar con éxito la alineación del producto óptico en el montaje. Esto explica las bajas probabilidades logradas en los cálculos de las preguntas anteriores.

e. $V(X)=\sigma_X^2=\dfrac{1-p}{p^2}=\dfrac{1-0.2}{(0.8)^2}=0.3125\Rightarrow \sigma_X=0.5590$ intentos.

Distribución normal o de Gauss

- Aparece de manera natural:
  - Errores de medida.
  - Distancia de frenado.
  - Altura, peso, propensión al crimen…
  - Distribuciones binomiales con n grande (n>30) y ‘p ni pequeño’ (np>5) ‘ni grande’ (nq>5).

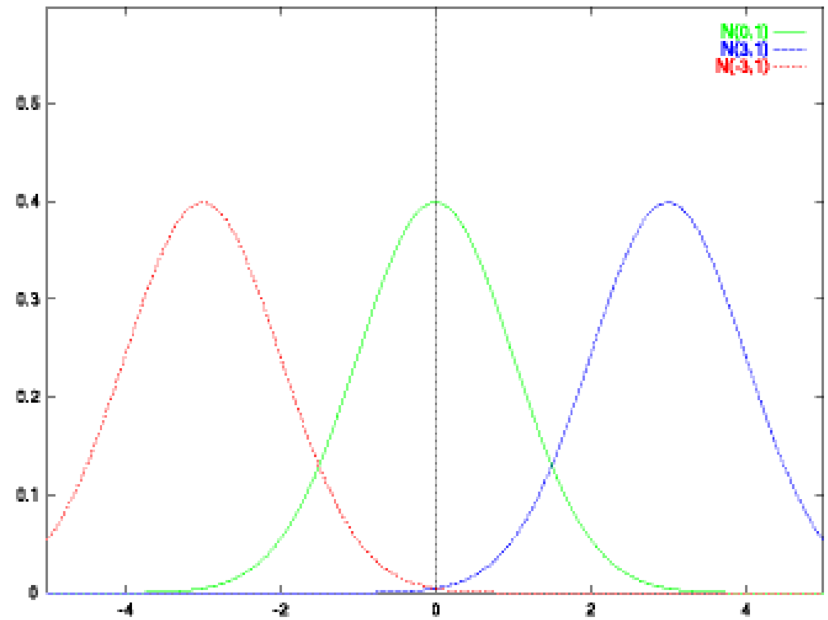
- Está caracterizada por dos parámetros: La media, μ, y la desviación típica, σ.

- Su función de densidad es:

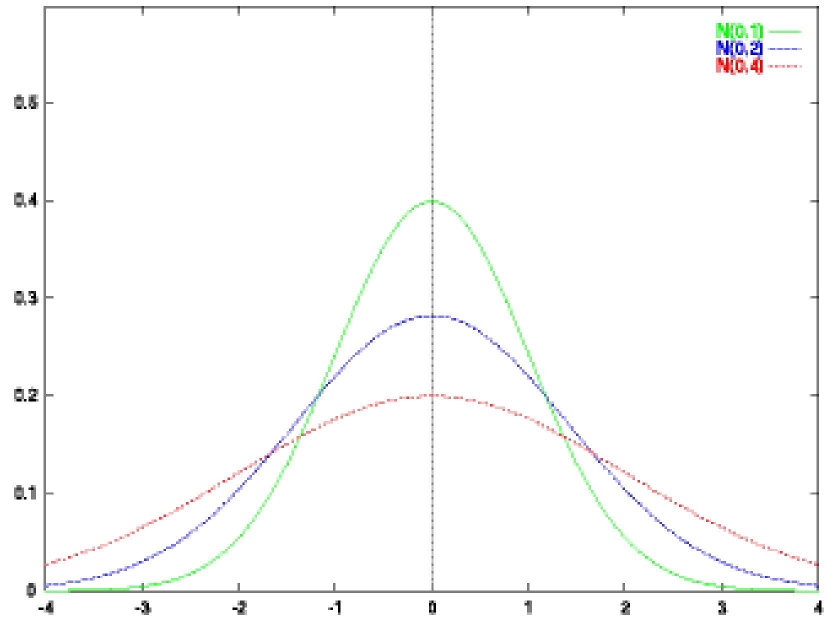
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

N(μ, σ): Interpretación geométrica

- Podéis interpretar la media como un factor de traslación.

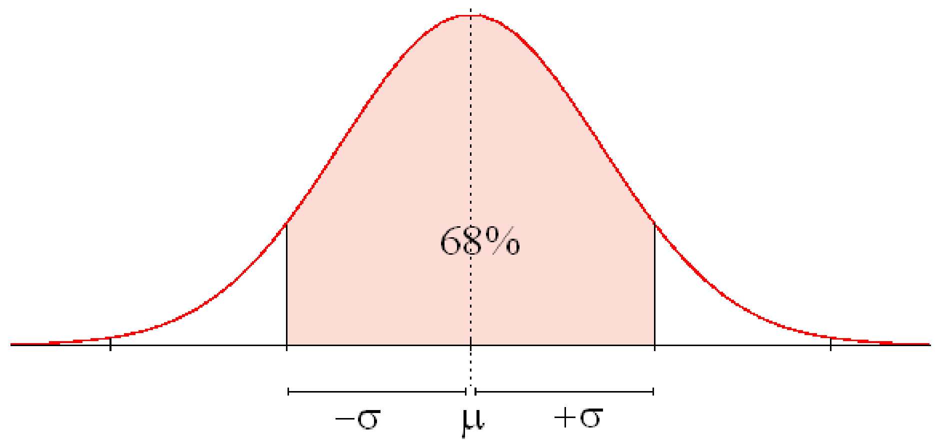


- Y la desviación típica como un factor de escala, grado de dispersión,…

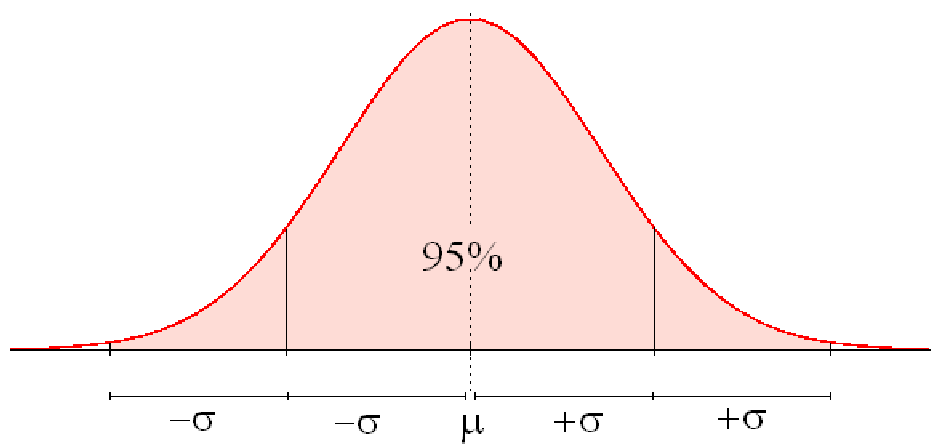


N(μ, σ): Interpretación probabilista

- Entre la media y una desviación típica tenemos siempre la misma probabilidad: aprox. 68%



- Entre la media y dos desviaciones típicas aprox. 95%



Algunas características

- La función de densidad es simétrica, los valores se concentran en torno a la media y tiene solo un valor que ocurre con mayor frecuencia.
  - Media, mediana y moda coinciden.

- Los puntos de inflexión de la función de densidad están a distancia **σ** de μ.

- Si tomamos intervalos centrados en μ, y cuyos extremos están…
  - a distancia **σ**, ➔ tenemos probabilidad **68%**
  - a distancia **2 σ**, ➔ tenemos probabilidad **95%**
  - a distancia **2'5 σ** ➔ tenemos probabilidad **99%**

- No es posible calcular la probabilidad de un intervalo simplemente usando la primitiva de la función de densidad, ya que no tiene primitiva expresable en términos de funciones 'comunes'.

- Todas las distribuciones normales N(μ, σ), pueden ponerse mediante una traslación μ, y un cambio de escala σ, como **N(0,1)**. Esta distribución especial se llama **normal tipificada**.

Tipificación

- Dada una variable de media μ y desviación típica σ, se denomina valor tipificado, z, de una observación x, a la distancia (con signo) con respecto a la media, medido en desviaciones típicas, es decir

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

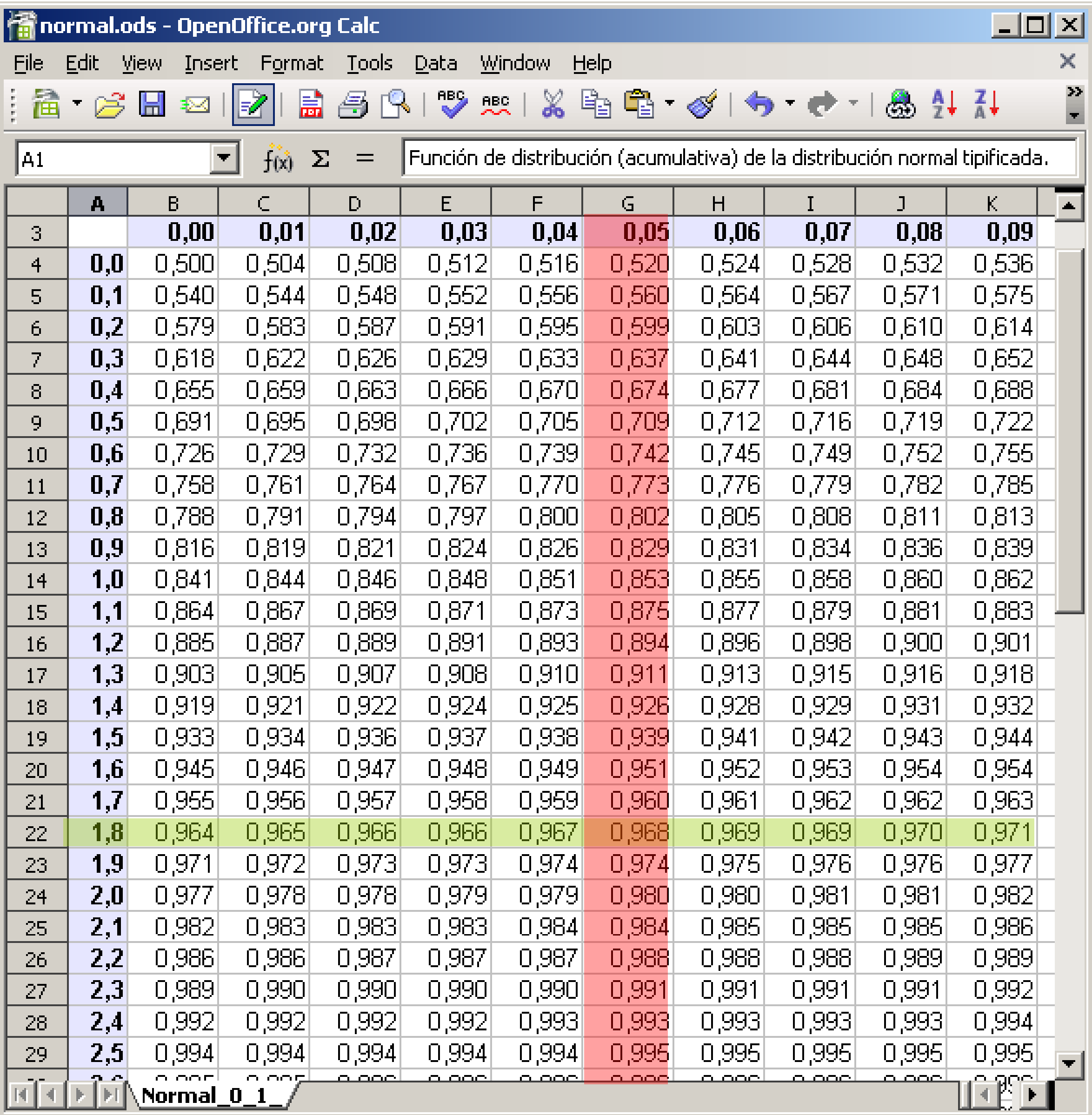
- En el caso de variable X normal, la interpretación es clara: Asigna a todo valor de $N(\mu, \sigma)$, un valor de $N(0,1)$ que deja **exactamente** la misma probabilidad por debajo.

- Permite comparar entre dos valores de dos distribuciones normales diferentes, para saber cuál de los dos es más extremo.

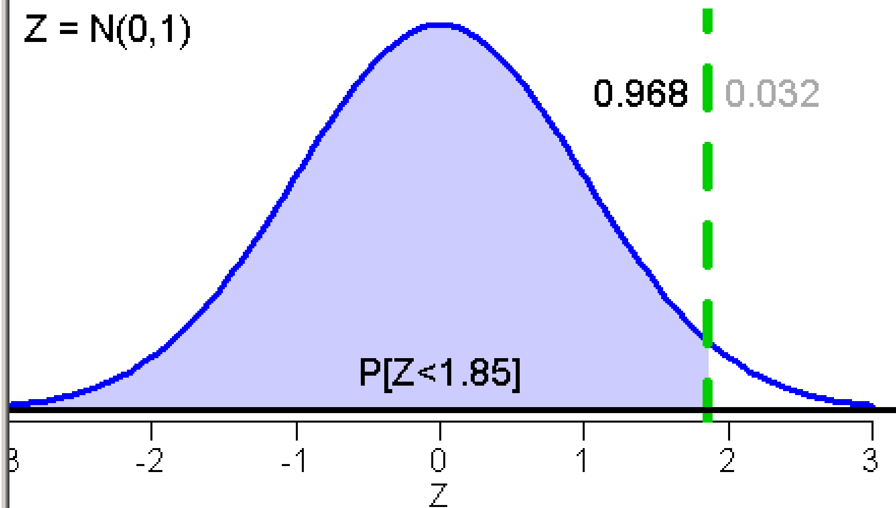
Tabla N(0,1)

Z es normal tipificada.

Calcular P[Z<1,85]



	B	C	D	E	F	G	H	I	J	K
	0,00	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995



Solución: 0,968 = 96,8%

Tabla N(0,1)

Z es normal tipificada.

Calcular P[Z<-0,54]

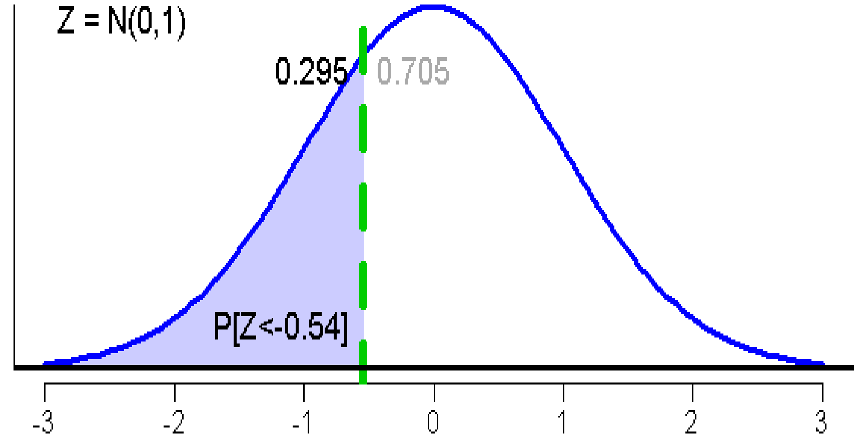
normal.ods - OpenOffice.org Calc

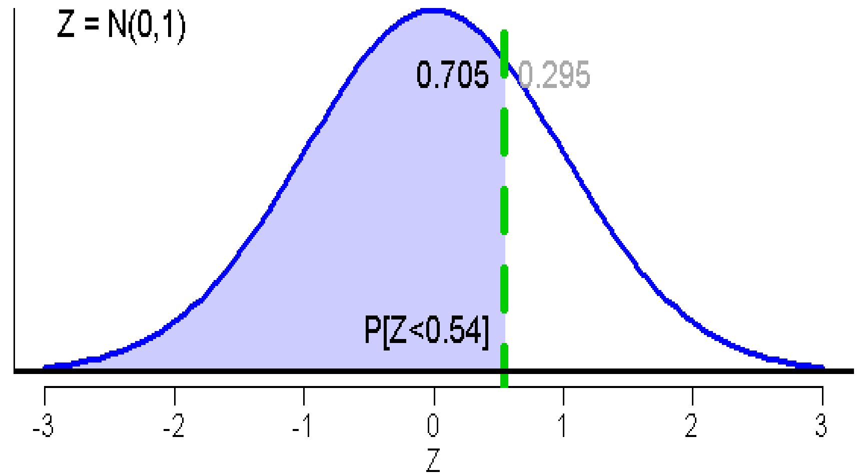
File Edit View Insert Format Tools Data Window Help

A1 | Función de distribución (acumulativa) de la distribución normal tipificada.

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995

Normal_0_1



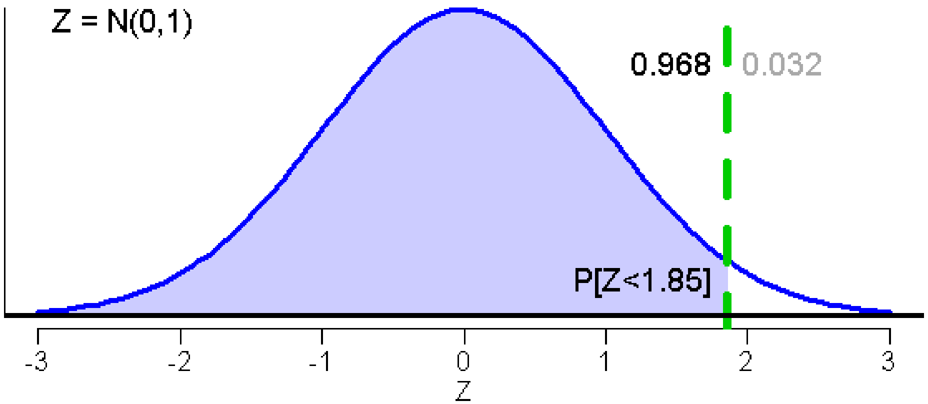


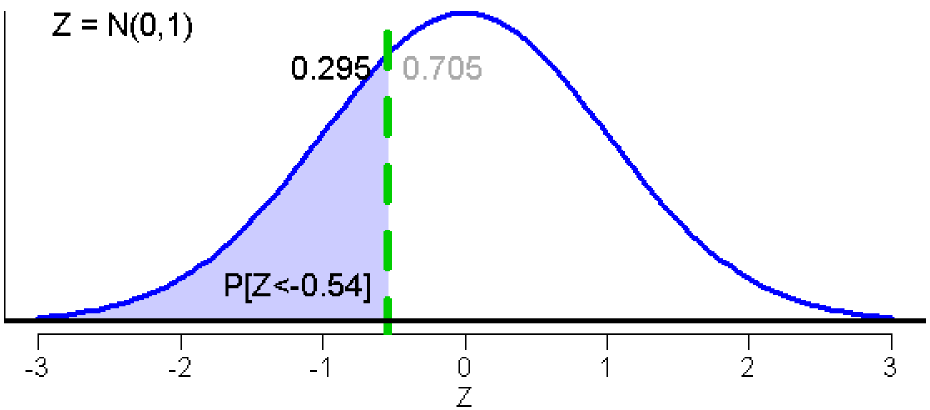
Solución: 1-0,705 = 0,295

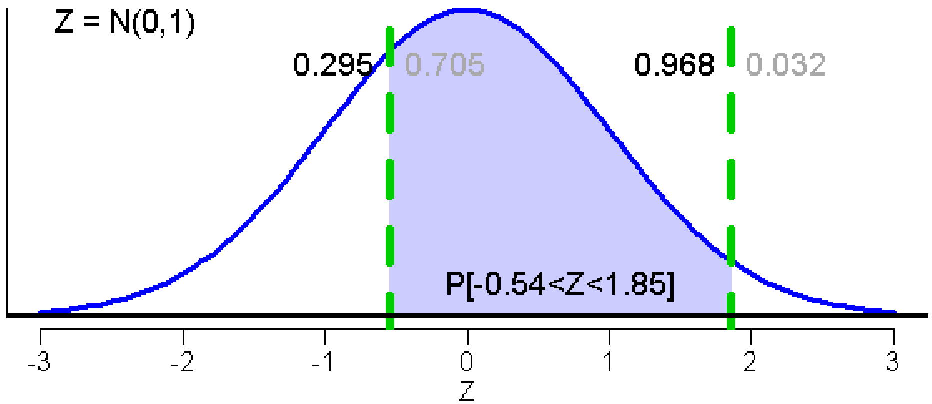
Tabla N(0,1)

Z es normal tipificada.

Calcular P[-0,54<Z<1,85]





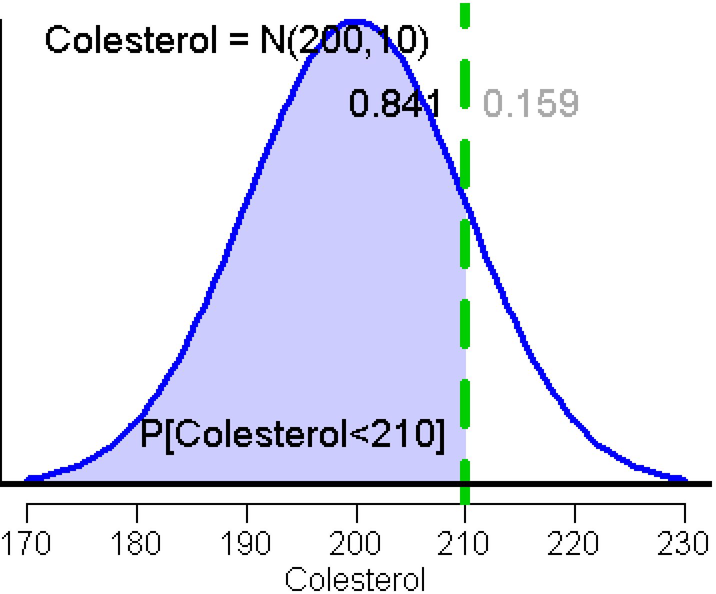


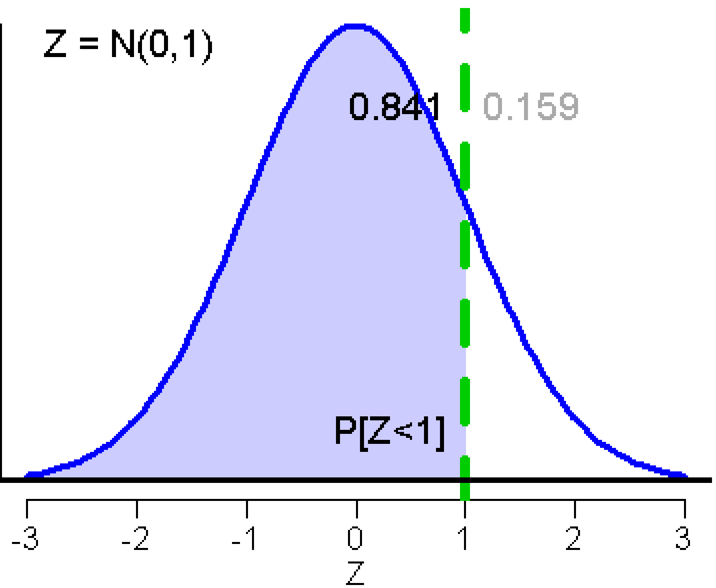
Solución: 0,968-0,295= 0,673

Ejemplo: Cálculo con probabilidades normales

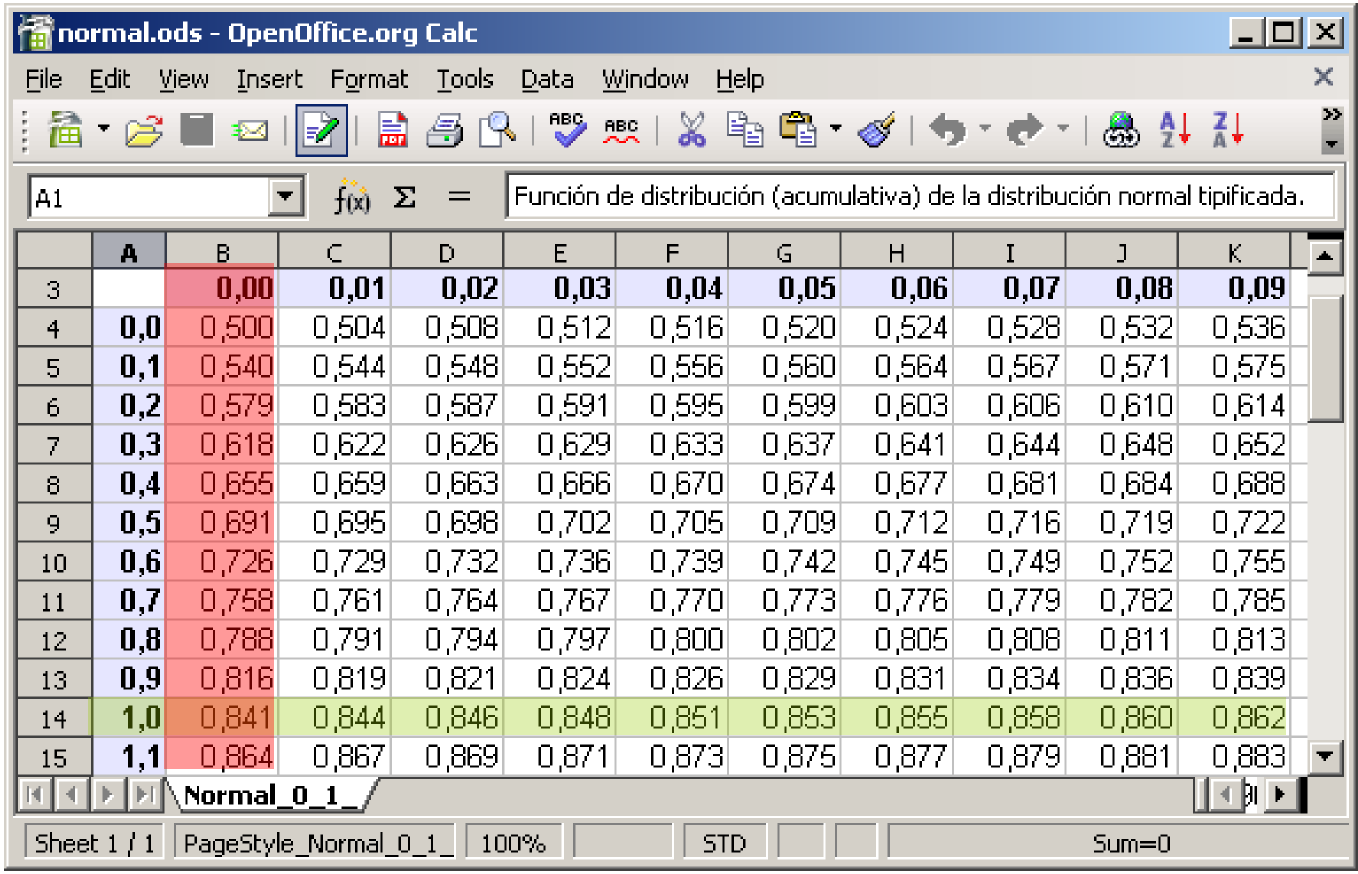
- El colesterol en la población tiene distribución normal, con media 200 y desviación 10.
- ¿Qué porcentaje de indivíduos tiene colesterol inferior a 210?
- Qué valor del colesterol sólo es superado por el 10% de los individuos.

- Si tipificamos.





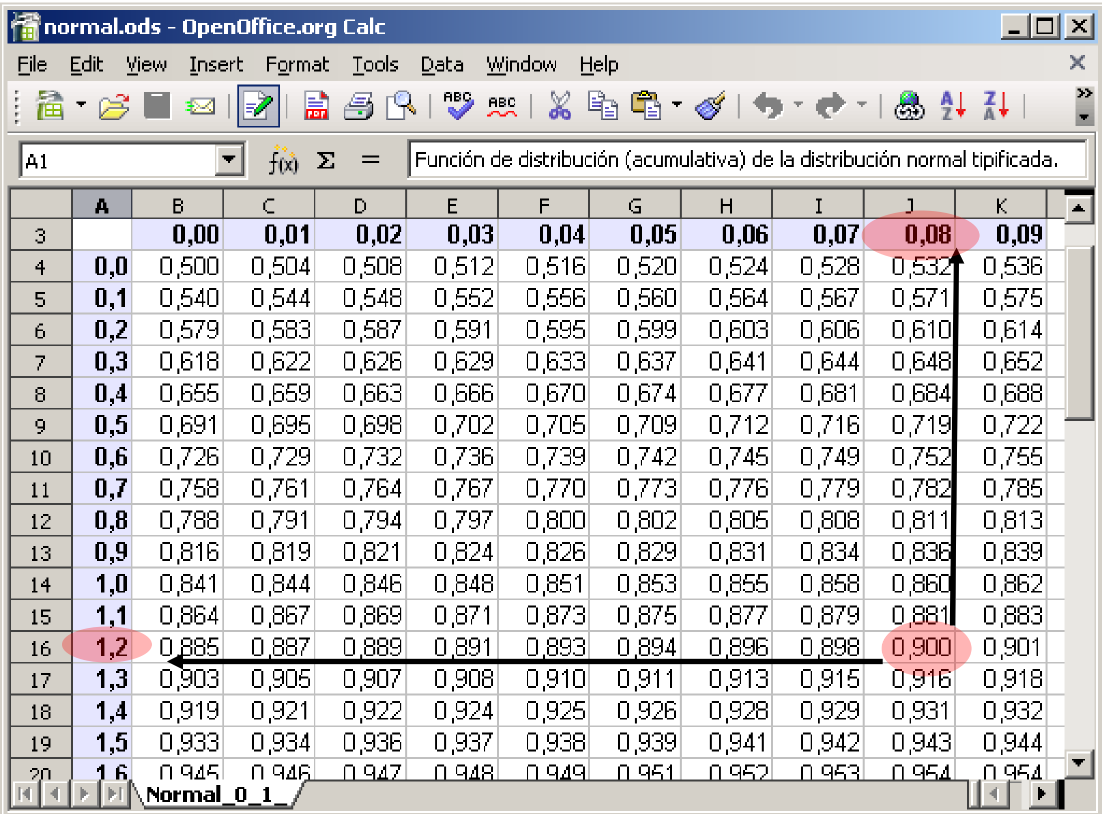
$$z = \frac{x-\mu}{\sigma} = \frac{210-200}{10} = 1$$



	A	B	C	D	E	F	G	H	I	J	K
3		0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
4	0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
5	0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
6	0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
7	0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
8	0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
9	0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
10	0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
11	0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
12	0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
13	0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
14	1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
15	1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883

$$P[Z < 1{,}00] = (\text{ver tabla}) = 0{,}841$$

- El valor del colesterol que sólo supera el 10% de los individuos es el percentil 90. Calculemos el percentil 90 de la N(0,1) y deshacemos la tipificación.

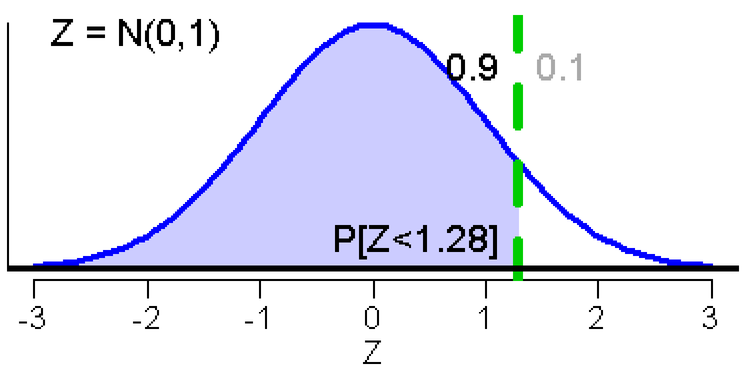


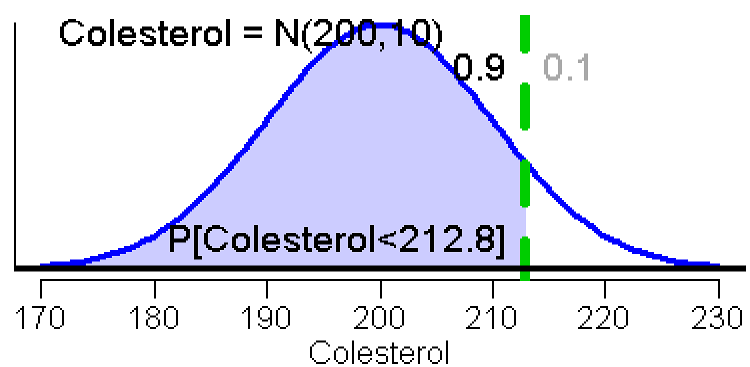
	A	B	C	D	E	F	G	H	I	J	K
3		0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
4	0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
5	0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
6	0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
7	0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
8	0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
9	0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
10	0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
11	0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
12	0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
13	0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
14	1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
15	1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
16	1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
17	1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
18	1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
19	1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
20	1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954

$$z = \frac{x-\mu}{\sigma}$$

$$1{,}28 = \frac{x-200}{10}$$

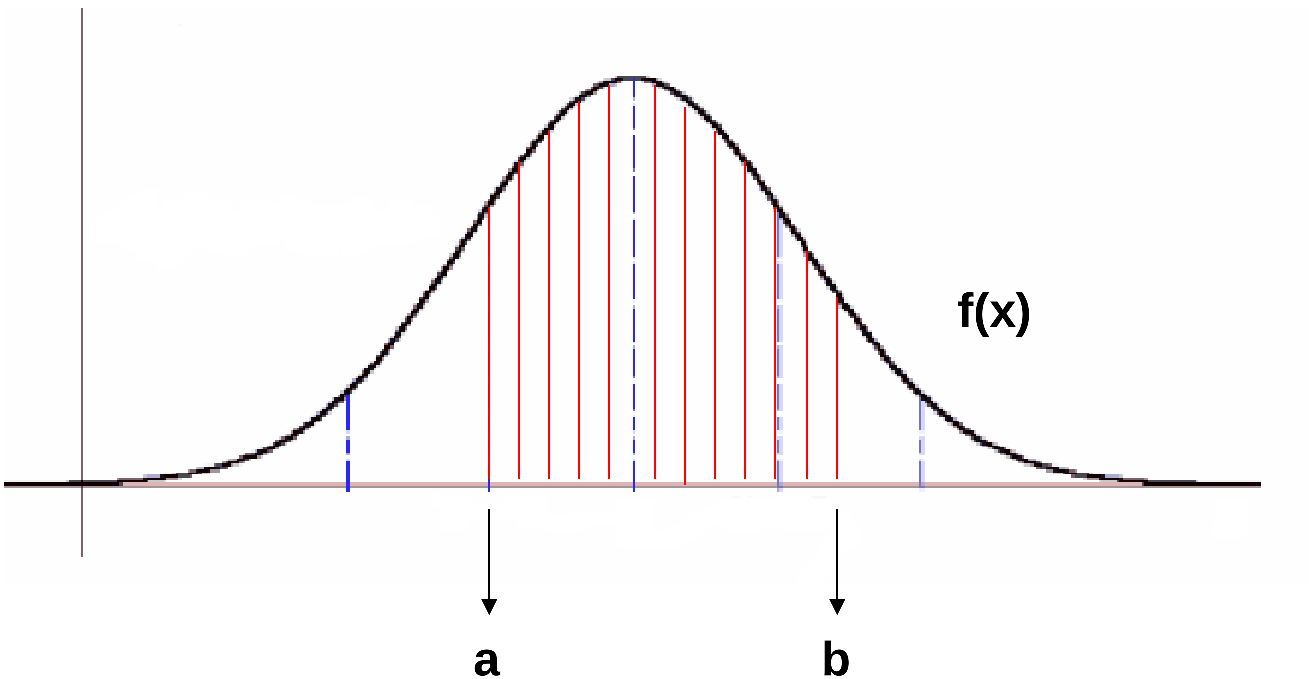
$$x = 200 + 10 \times 1{,}28 = 212{,}8$$





P(a ≤ X ≤ b), es decir, la probabilidad de que la variable X esté entre los valores a y b, se calcula como:

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



IMPORTANTE: En consecuencia, la probabilidad de que la variable X tome un determinado valor es CERO:

$$P(X = a) = \int_a^a f(x)dx = 0$$

Por lo tanto,

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

DEFINICION (Función de distribución): Dada una variable aleatoria continua X, con función de densidad f(x), la función de distribución F(x) es la función que para cada valor de la variable nos da la probabilidad de que X tome ese valor, o cualquier otro inferior.

$$F(x) = P(X \leq x) = \int_a^x f(t)dt$$

La función de distribución cumple:

1. La derivada de la función de distribución, es la función de densidad.

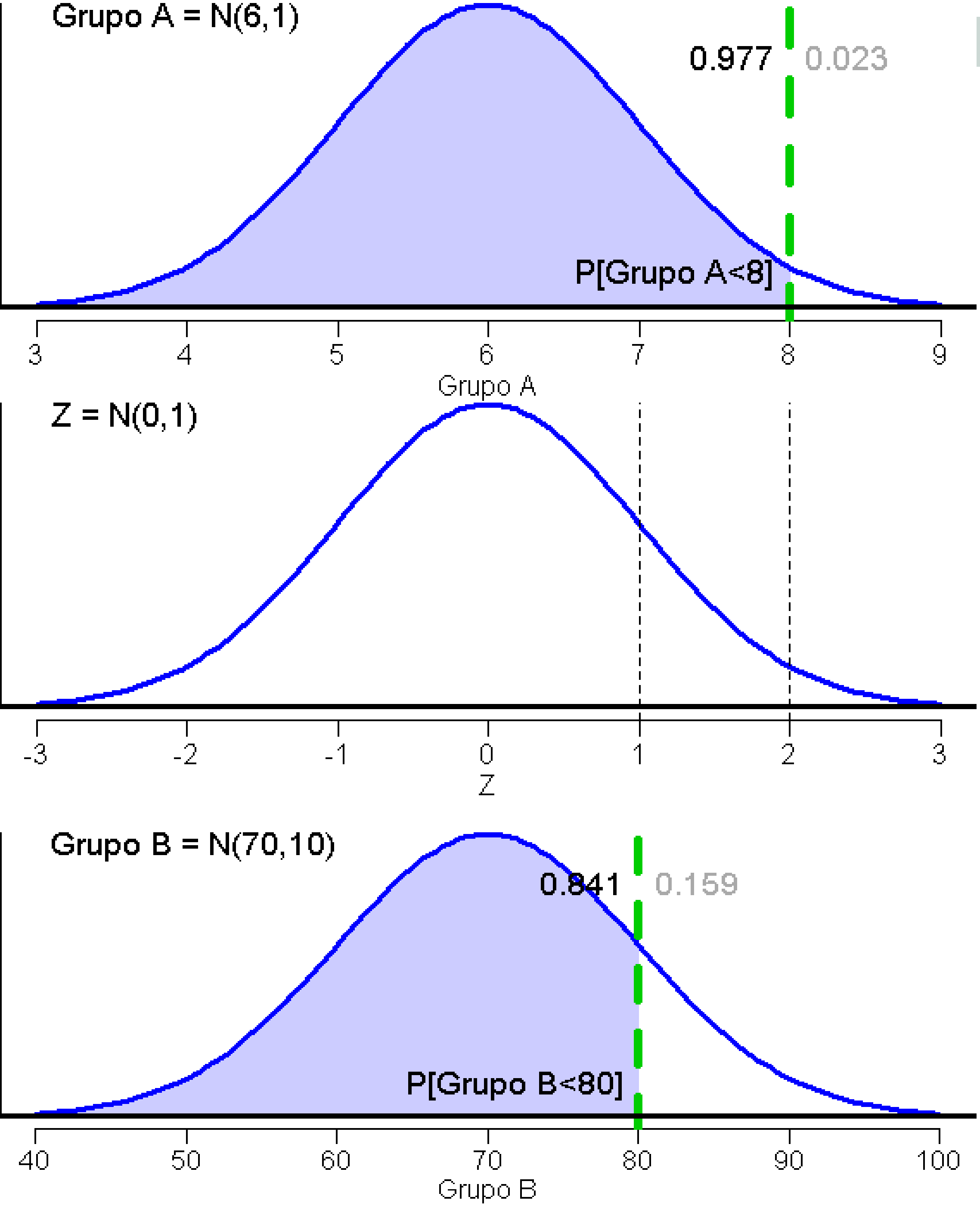
$$F'(x) = f(x)$$

2. Se verifica:

$$P(a \leq X \leq b) = F(b) - F(a)$$

Ejemplo: Tipificación

- Se quiere dar una beca a uno de dos estudiantes de sistemas educativos diferentes. Se asignará al que tenga mejor expediente académico.
  - El estudiante A tiene una calificación de 8 en un sistema donde la calificación de los alumnos se comporta como N(6,1).
  - El estudiante B tiene una calificación de 80 en un sistema donde la calificación de los alumnos se comporta como N(70,10).
- **Solución**
  - No se puede comparar directamente 8 puntos de A frente a los 80 de B. Como ambas poblaciones se comportan de modo normal, se puede tipificar y observar las puntuaciones sobre una distribución de referencia N(0,1)



$$z_A = \frac{x_A - \mu_A}{\sigma_A} = \frac{8-6}{1} = 2$$

$$z_B = \frac{x_B - \mu_B}{\sigma_B} = \frac{80-70}{10} = 1$$

Como $Z_A > Z_B$, se puede decir que el porcentaje de compañeros del mismo sistema de estudios que ha superado en calificación el estudiante A es mayor que el que ha superado B.

¿Se podría pensar que A es mejor candidato para la beca?

DISTRIBUCIÓN NORMAL. EJEMPLOS

1) Una empresa conservera enlata alimentos y el peso neto de las latas sigue la distribución normal N(400, 10), calcula:

a. La probabilidad de que una lata pese menos de 380 gramos.
b. La probabilidad de que una lata pese entre 385 y 410 gramos

Será necesario tipificar la variable para poder utilizar las tablas:

a)

$$p(X \leq 380) = p\left(\frac{X - 400}{10} \leq \frac{380-400}{10}\right) = p(Z \leq -2) =$$

$$= p(Z \geq 2) = 1 - p(Z \leq 2) = 1 - 0{,}9772 = 0{,}0228$$

b)

$$p(385 \leq X \leq 410) = p\left(\frac{385-400}{10} \leq \frac{X - 400}{10} \leq \frac{410-400}{10}\right) =$$

$$= p(-1{,}5 \leq Z \leq 1) = p(Z \leq 1) - p(Z \leq -1{,}5) =$$

$$= p(Z \leq 1) - p(Z \geq 1{,}5) =$$

$$= p(Z \leq 1) - [1 - p(Z \leq 1{,}5)] =$$

$$= 0{,}8413 - (1 - 0{,}9332) = 0{,}7745$$

¿Por qué es importante la distribución normal?

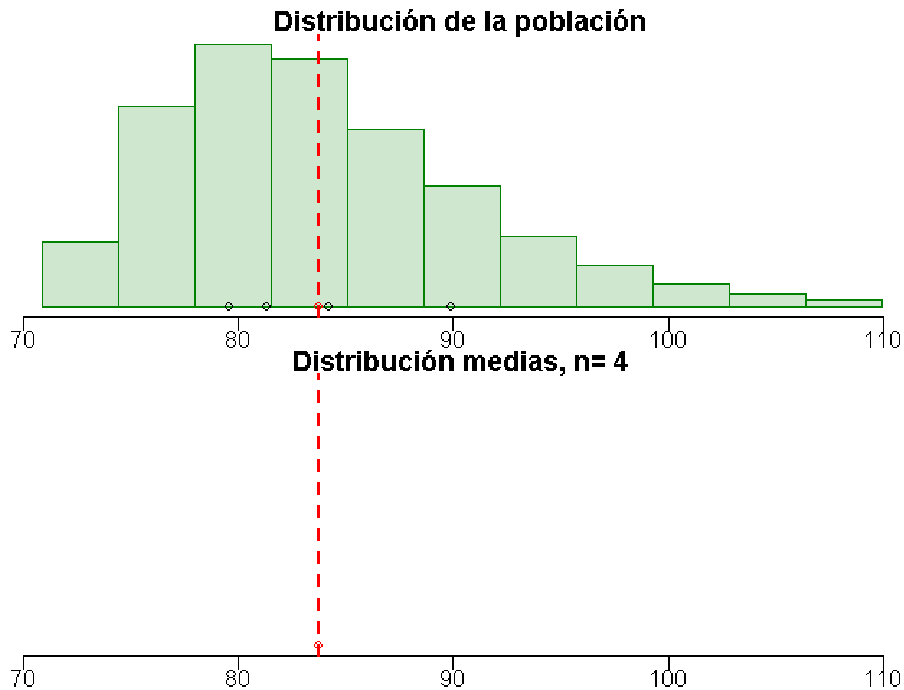
- Las propiedades que tiene la distribución normal son interesantes ya que, aunque una variable aleatoria no posea distribución normal, ciertos estadísticos/estimadores calculados sobre muestras elegidas al azar sí que poseen una distribución normal.

- Es decir, sea cual sea la distribución que tengan nuestros datos, los ‘objetos’ que resumen la información de una muestra, posiblemente tengan distribución normal (o asociada).

Aplic. de la normal: Estimación en muestras

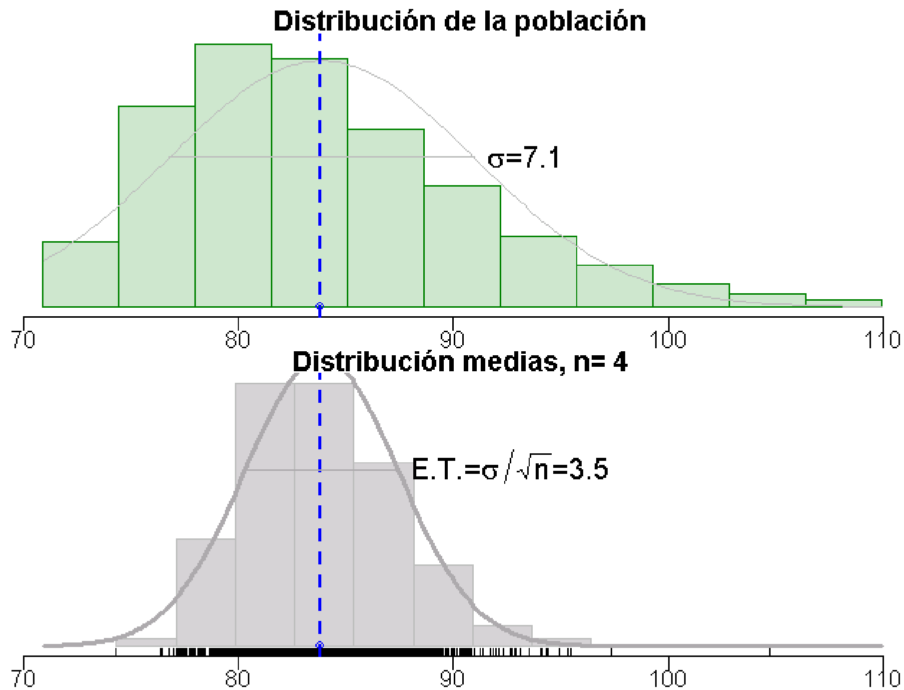
- Como ilustración se puede observar una variable que presenta valores distribuidos de forma muy asimétrica. Claramente no normal.

- Se va a seleccionar muestras de diferentes tamaños, y utilizar la media de cada muestra para estimar la media de la población.



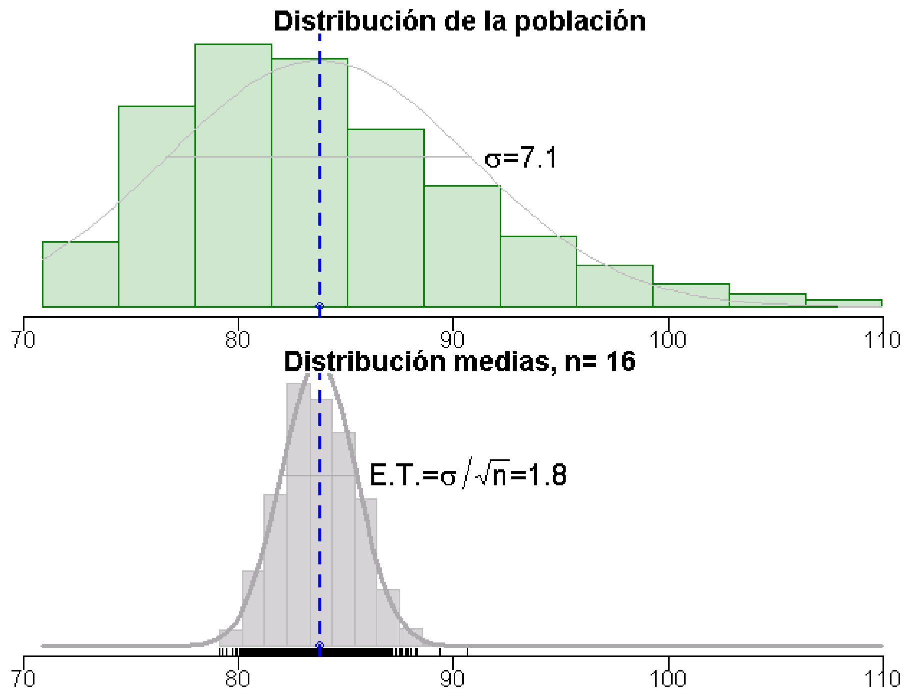
Aplic. de la normal: Estimación en muestras

- Cada muestra ofrece un resultado diferente: La media muestral es variable aleatoria.
- Su distribución es más parecida a la normal que la original.
- También está menos dispersa. A su dispersión ('desviación típica del estimador media muestral') se le suele denominar **error típico**.



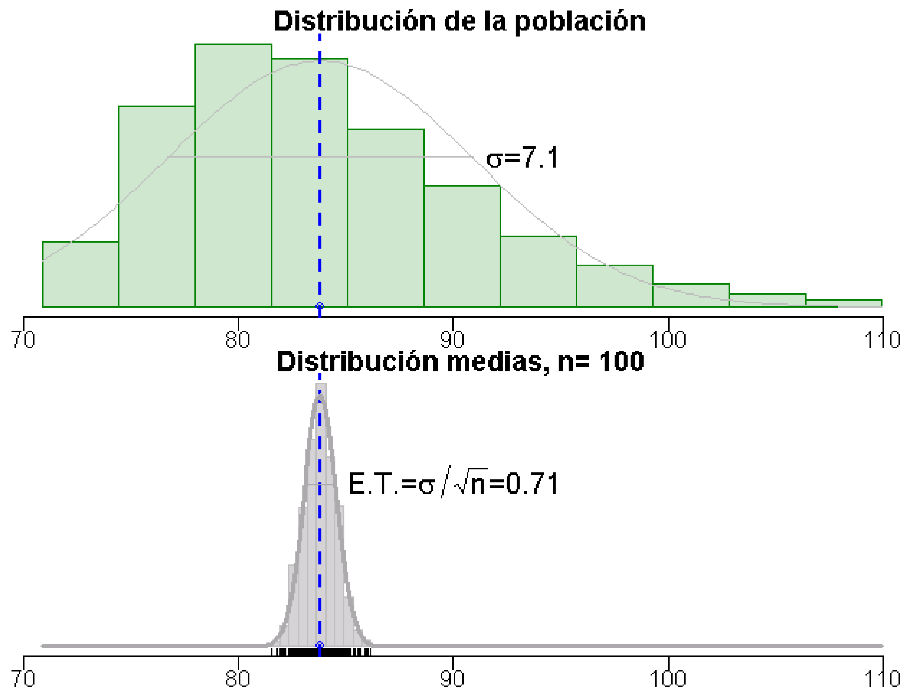
Aplic. de la normal: Estimación en muestras

- Al aumentar el tamaño, n, de la muestra:
  - La normalidad de las estimaciones mejora
  - El error típico disminuye.



Aplic. de la normal: Estimación en muestras

- Se puede 'garantizar' medias muestrales tan cercanas como quiera a la verdadera media, sin más que tomar 'n bastante grande'
- Por ejemplo, se utiliza esta propiedad para dimensionar el tamaño de una muestra antes de empezar una investigación.



Resumen: Teorema del límite central

- Dada una variable aleatoria **cualquiera**, si extraemos muestras de tamaño n, y calculamos los promedios muestrales, entonces:
  - dichos promedios tienen distribución aproximadamente normal;
  - La media de los promedios muestrales es la misma que la de la variable original.
  - La desviación típica de los promedios disminuye en un factor “raíz de n” (error estándar).
  - Las aproximaciones anteriores se hacen exactas cuando n tiende a infinito.
- Este teorema justifica la importancia de la distribución normal.
  - Sea lo que sea lo que midamos, cuando se promedie sobre una muestra grande (**n>30**) va a aparecer de manera natural la distribución normal.

Distribuciones asociadas a la normal

- Cuando se desea hacer inferencia estadística se ha visto que la distribución normal aparece de forma casi inevitable.
- Dependiendo del problema, podemos encontrar otras (asociadas):
  - X^2 (chi cuadrado)
  - t- student
  - F-Snedecor
- Estas distribuciones resultan directamente de operar con distribuciones normales. Típicamente aparecen como distribuciones de ciertos estadísticos.
- Sobre todo, interesa saber qué valores de dichas distribuciones son “atípicos”.