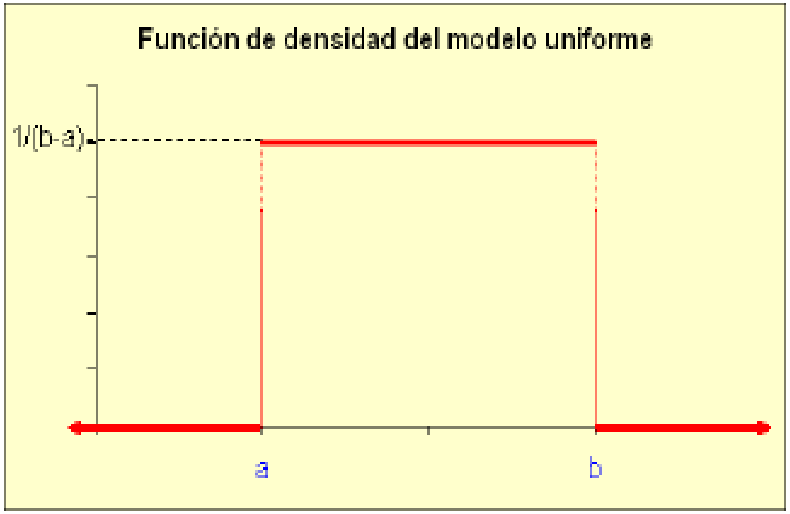
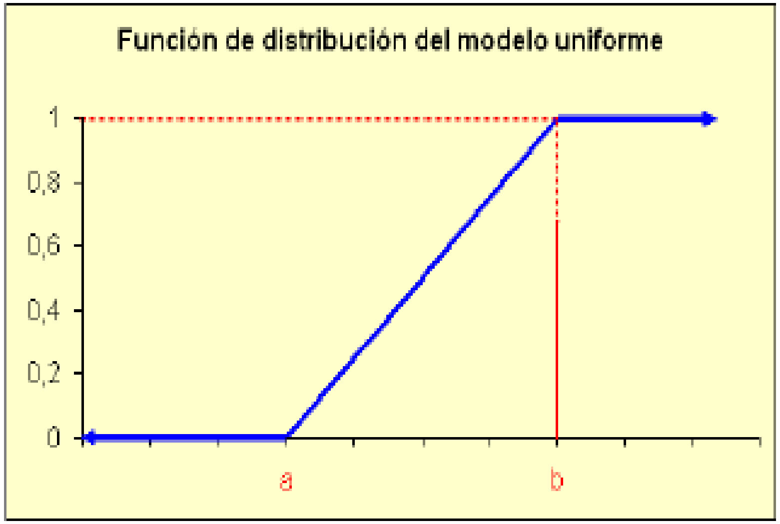
DISTRIBUCIÓN UNIFORME

Una variable aleatoria X tiene distribución Uniforme en el intervalo [a, b] si su función de densidad es:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & en\ otro\ caso \end{cases}$$

Si X tiene distribución Uniforme lo notaremos como $X \sim U(a;\ b)$





Si $X \sim U(a;\ b)$, ¿cuál es su función de distribución acumulada?

Si $a \leq x \leq b$ $\longrightarrow$ $F(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$

Por lo tanto queda que:

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Si $X \sim U(a;\ b)$, ¿cuál es su esperanza y su varianza?.

$$E(X) = \int_a^b x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{b-a}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

Distribución normal

Pierre Simon de Laplace
(1749-1827)

Karl F. Gauss
(1777-1855)

Sin duda la distribución continua de probabilidad más importante, por la frecuencia con que se encuentra y por sus aplicaciones teóricas, es la **distribución normal, gaussiana o de Laplace- Gauss.** Fue descubierta y publicada por primera vez en 1733 por De Moivre. A la misma llegaron, de forma independiente, Laplace (1812) y Gauss (1809), en relación con la teoría de los errores de observación astronómica y física .

Distribución normal o gaussiana

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

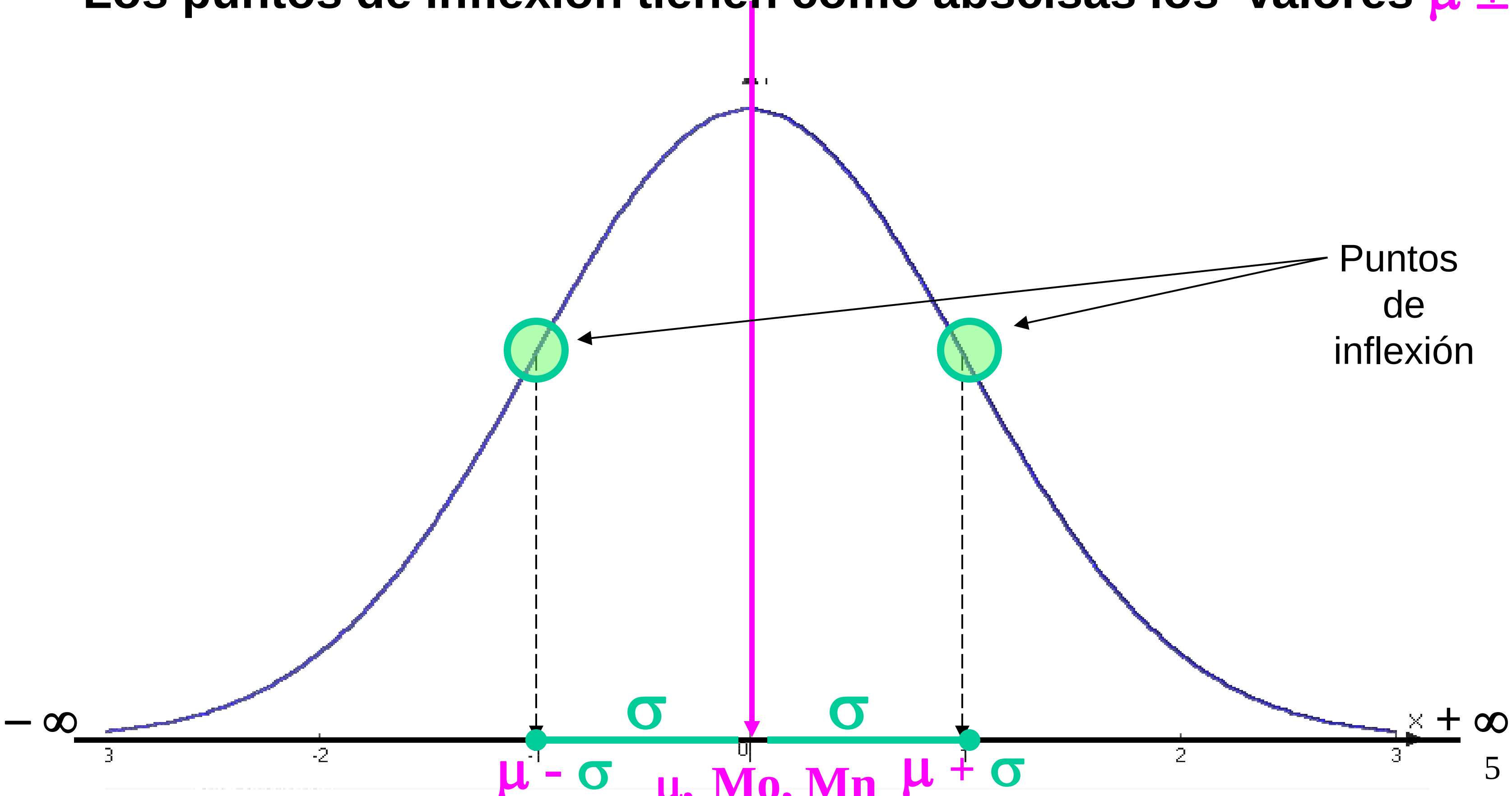
- Su función de densidad es:

$$\boxed{N(\mu,\sigma) = P(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (\sigma > 0)}$$

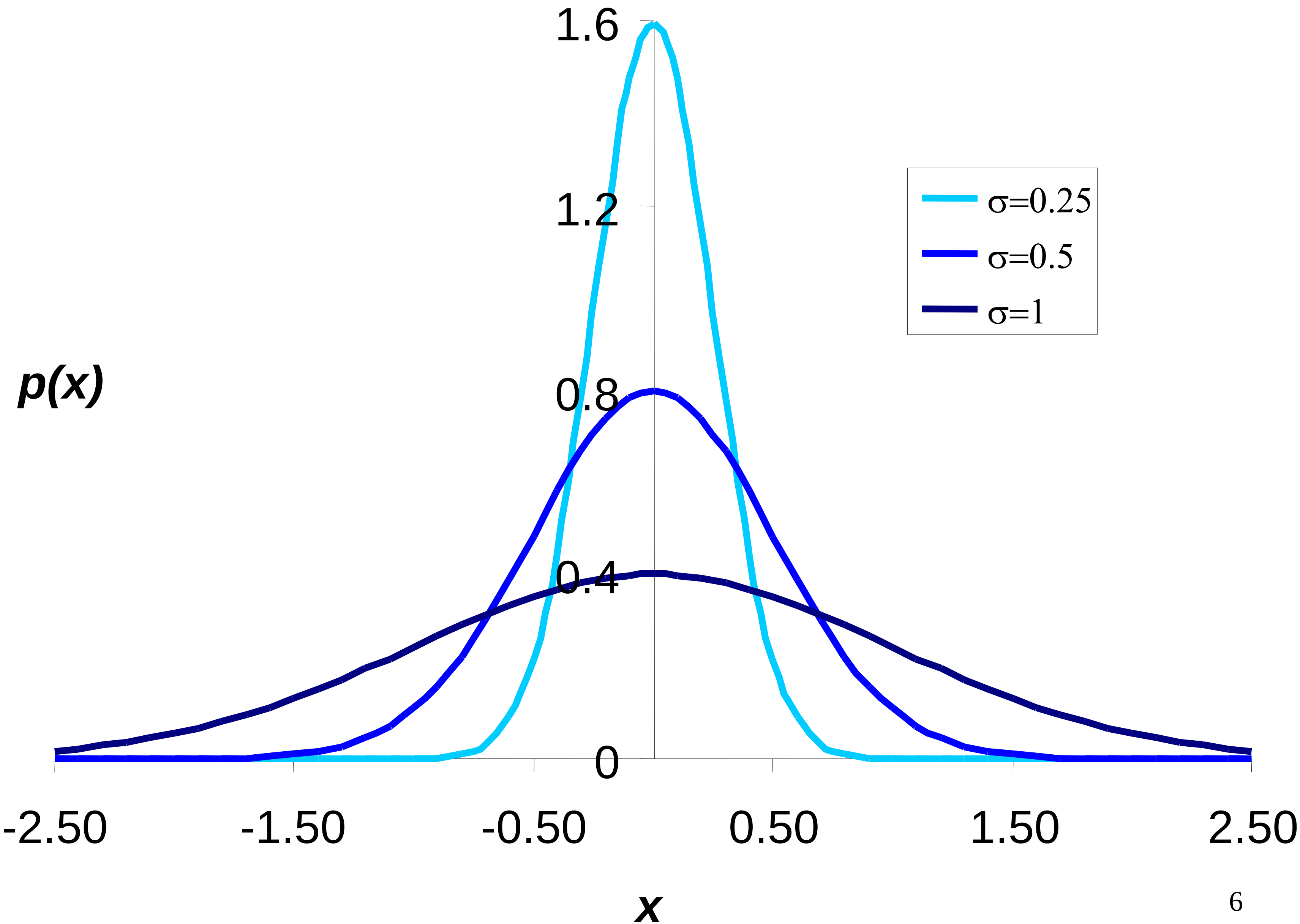
La curva normal adopta un número infinito de formas, determinadas por sus parámetros μ y σ.

Características de la distribución Normal

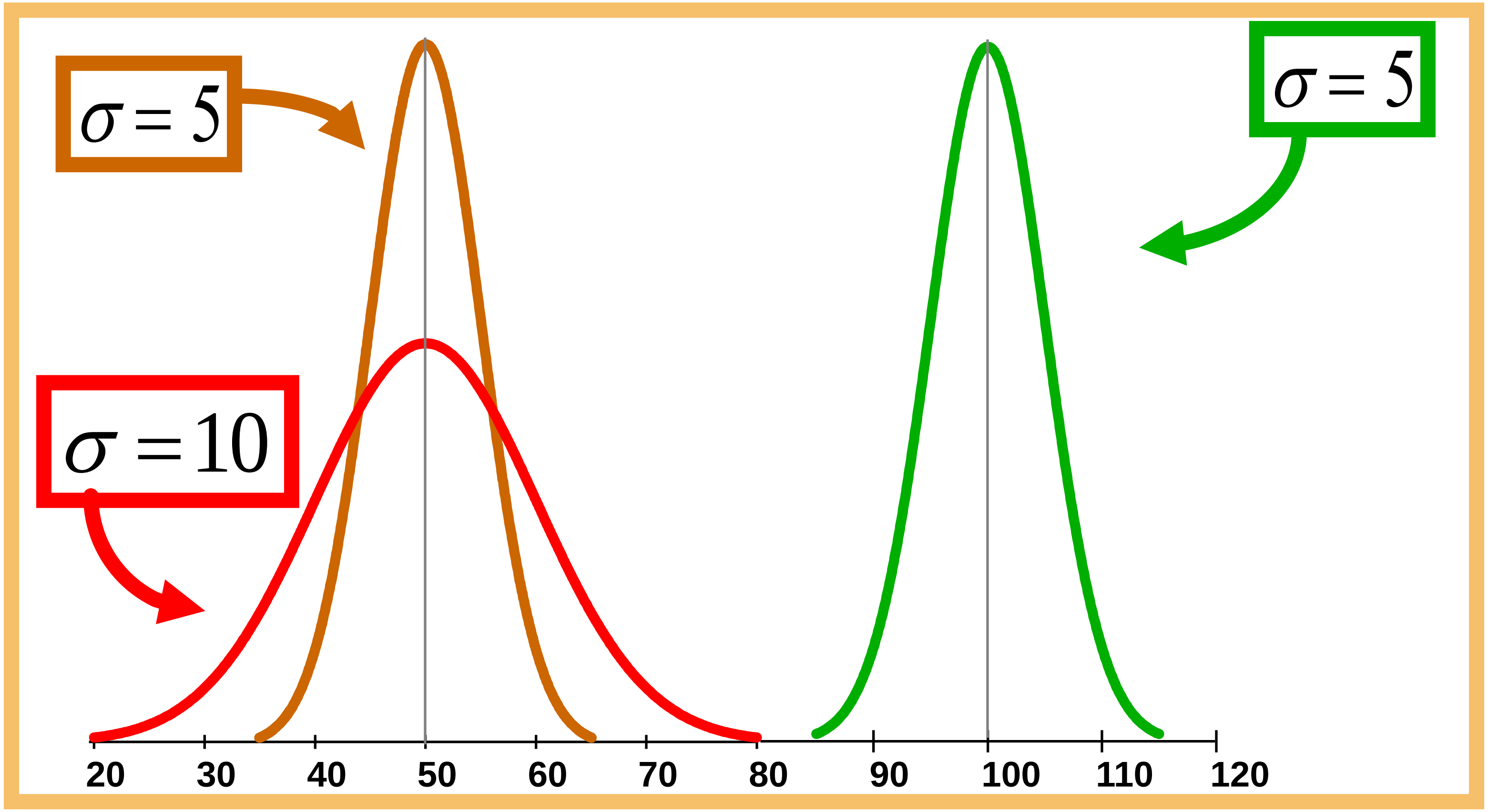
- **Tiene forma de campana, es asintótica al eje de las abscisas (para $x = \pm\infty$)**
- **Simétrica con respecto a la media (μ) donde coinciden la mediana (Mn) y la moda (Mo)**
- **Los puntos de inflexión tienen como abscisas los valores $\mu \pm \sigma$**



Distribución normal con μ=0 para varios valores σ

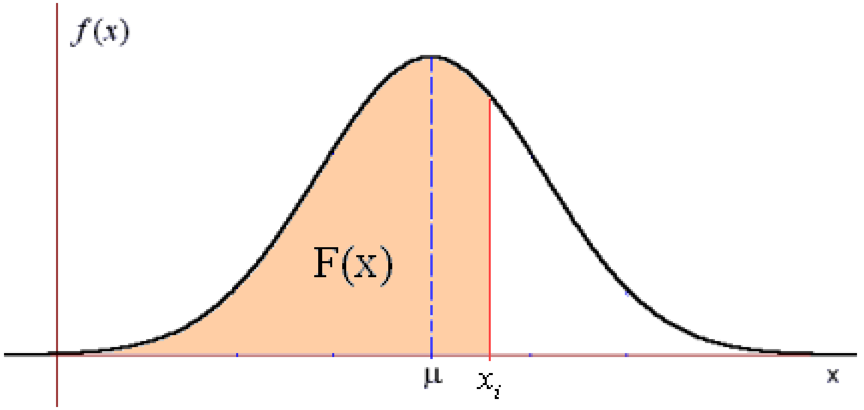


$$N(\mu,\sigma) = P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (\sigma > 0)$$



Curvas normales con distintas medias y desviaciones estándar.

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



Podemos obtener la función de distribución *F*(*x*) integrando la función de densidad de probabilidad:

$$\boxed{F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv}$$

De modo que la probabilidad de una variable aleatoria normal *X* en un intervalo $a \leq x \leq b$ es:

$$P(a \leq X \leq b) = F(b) - F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv$$

En particular: $$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv = 1$$

¡No podemos calcular analíticamente el valor de la integral! Tabularemos sus valores numéricos...

¿Cómo calcular probabilidades asociadas a una curva normal específica?

Dado que tanto μ como σ pueden asumir **infinitos valores** lo que hace impracticable tabular las probabilidades para todas las posibles distribuciones normales, se utiliza la **distribución normal reducida o tipificada o normal estándar.**

Se define una variable $Z = \boxed{\dfrac{x - \mu}{\sigma}}$

Es una traslación , y un cambio de escala de la variable original.

La nueva variable **z** se distribuye como una

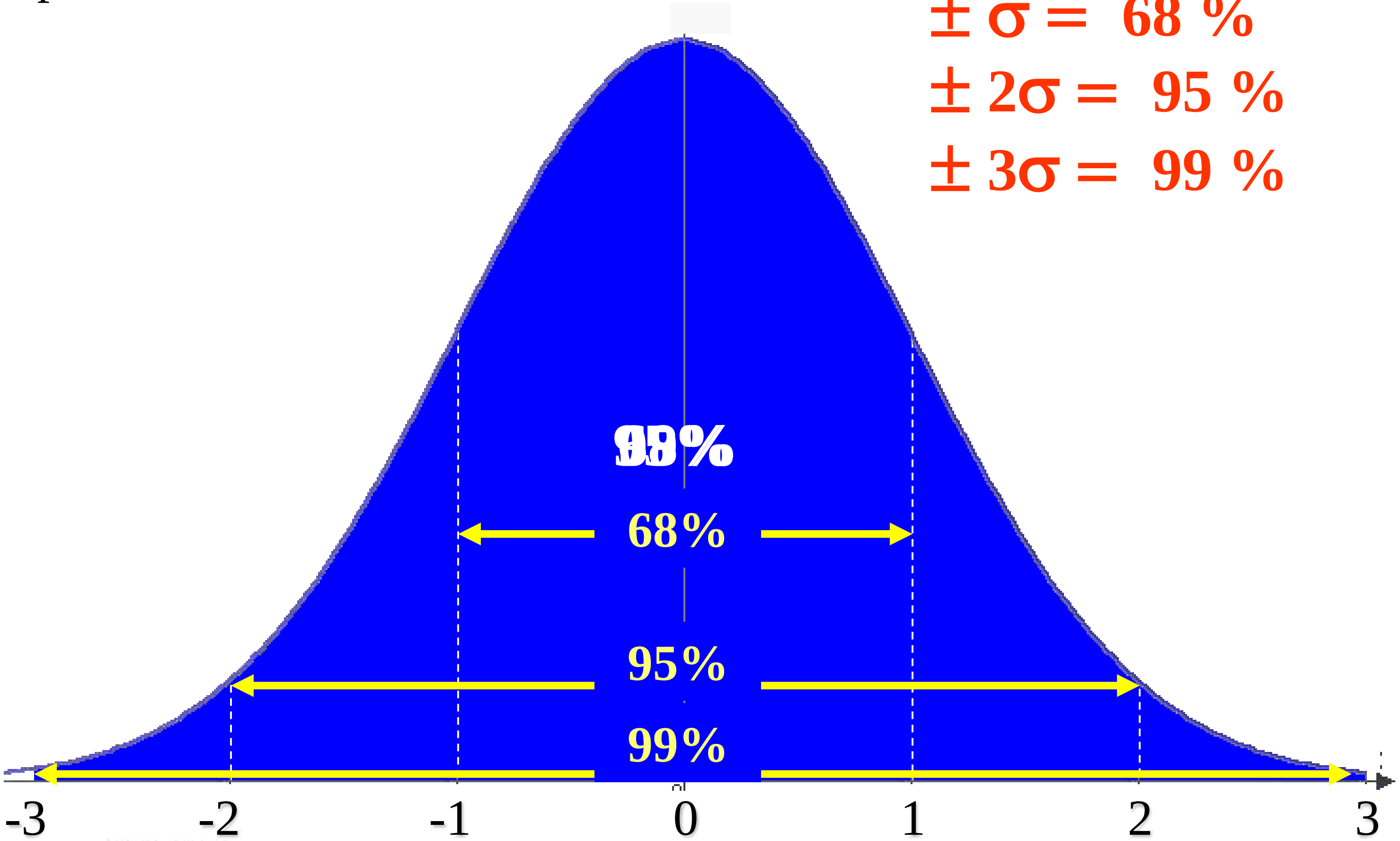
NORMAL con media μ **= 0** y desviación típica σ **= 1**

Recordemos de nuevo que en <u>cualquier</u> distribución normal las probabilidades delimitadas entre :

$\pm\,\sigma = 68\ \%$

$\pm\,2\sigma = 95\ \%$

$\pm\,3\sigma = 99\ \%$



Tipificación

- Dada una variable de media μ y desviación típica σ, se denomina valor tipificado z, de una observación x,

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

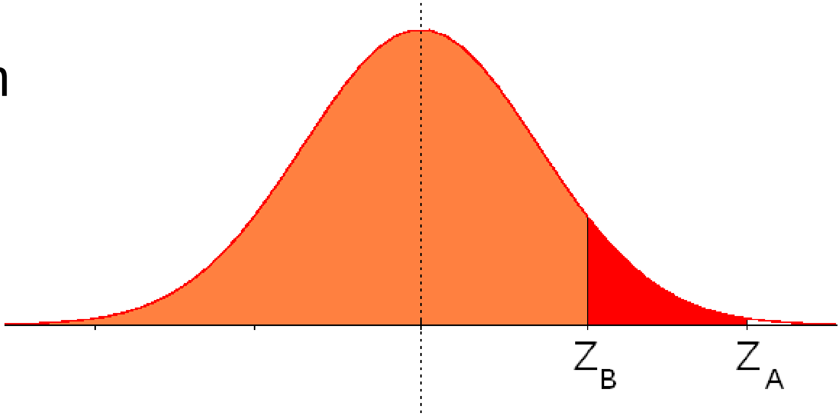
- En el caso de variable X normal, la interpretación es clara: asigna a todo valor de N(μ, σ), un valor de N(0,1) que deja **exáctamente** la misma probabilidad por debajo.

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

Se quiere dar una beca a uno de dos estudiantes de sistemas educativos diferentes y 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).

- No podemos comparar directamente 8 puntos de A frente a los 80 de B, pero como ambas poblaciones se comportan de modo normal, podemos tipificar y observar las puntuaciones sobre una distribución de referencia N(0,1).



- Como $z_A > z_B$, podemos 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. En principio A es mejor candidato para la beca.

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

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

Apliquemos el cambio de variable tipificada a la función de distribución $F(x)$:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv$$

$$\left. \begin{aligned} z &= \frac{v-\mu}{\sigma} \\ dv &= \sigma\, dz \end{aligned} \right\}$$

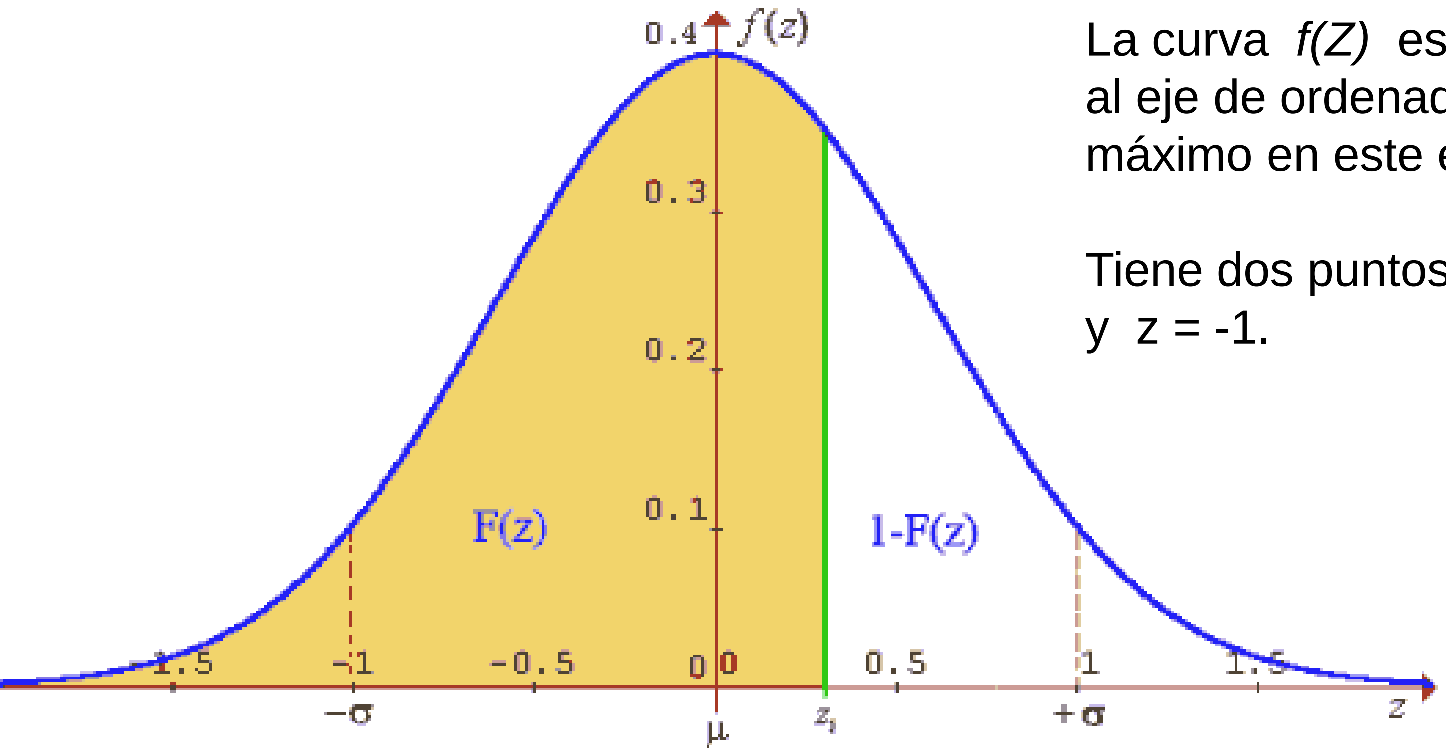
$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \; ; \; -\infty < z < \infty$$

$$\Phi(z) = p(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$$

Las probabilidades de la variable tipificada (z) están tabuladas para los diferentes valores de la variable. Para calcular probabilidades, una vez transformada, la variable a valores de z, se busca en una tabla el área correspondiente.

$$p(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\ ;\ -\infty<z<\infty$$

$$\Phi(z)=p(Z\le z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{u^2}{2}}du$$



Característica de la distribución normal estándar:

No depende de ningún parámetro.

Su media es 0, su varianza es 1 y su desviación típica es 1.

La curva *f(Z)* es simétrica respecto al eje de ordenadas y tiene un máximo en este eje.

Tiene dos puntos de inflexión en z =1 y z = -1.

La tabla consta de:

*Margen izquierdo : Los enteros de z y su primer decimal.

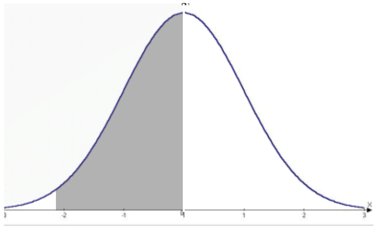
* Margen superior: segundo decimal

* Cuerpo de la tabla: áreas correspondientes, acumuladas, desde 0 hasta 3.99

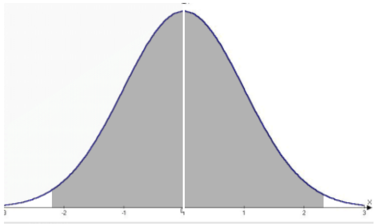
Z	0.00	0.01	0.02	0.03	0.04	Z
0.0	.50000	0.50399	0.50798	0,51197	0.41595	0.0
0.1	0.53983 .	0.54379	0.54776	0.55172	0.55567	0.1
0.2	. ...	......				0.2
0.3	....	......	......			0.3
0.4	....	....				0.4
0.5	....					0.5

EJEMPLOS:

1.-¿Cuál es la probabilidad de que un valor de z esté entre 0 y -2.03?

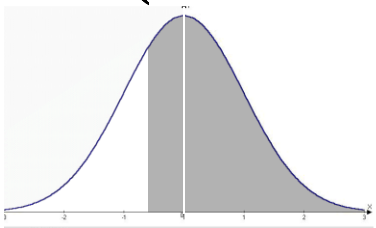


2.-¿Cuál es la probabilidad de que un valor de z esté entre -2.03 y +2.03?

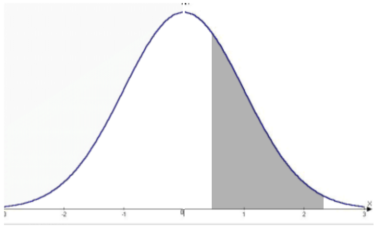


3. Hallar $P(z > 1.25)$

4. Hallar $P(-0.34 < z < \infty)$



5. Hallar $P(0.34 < z < 2.30)$



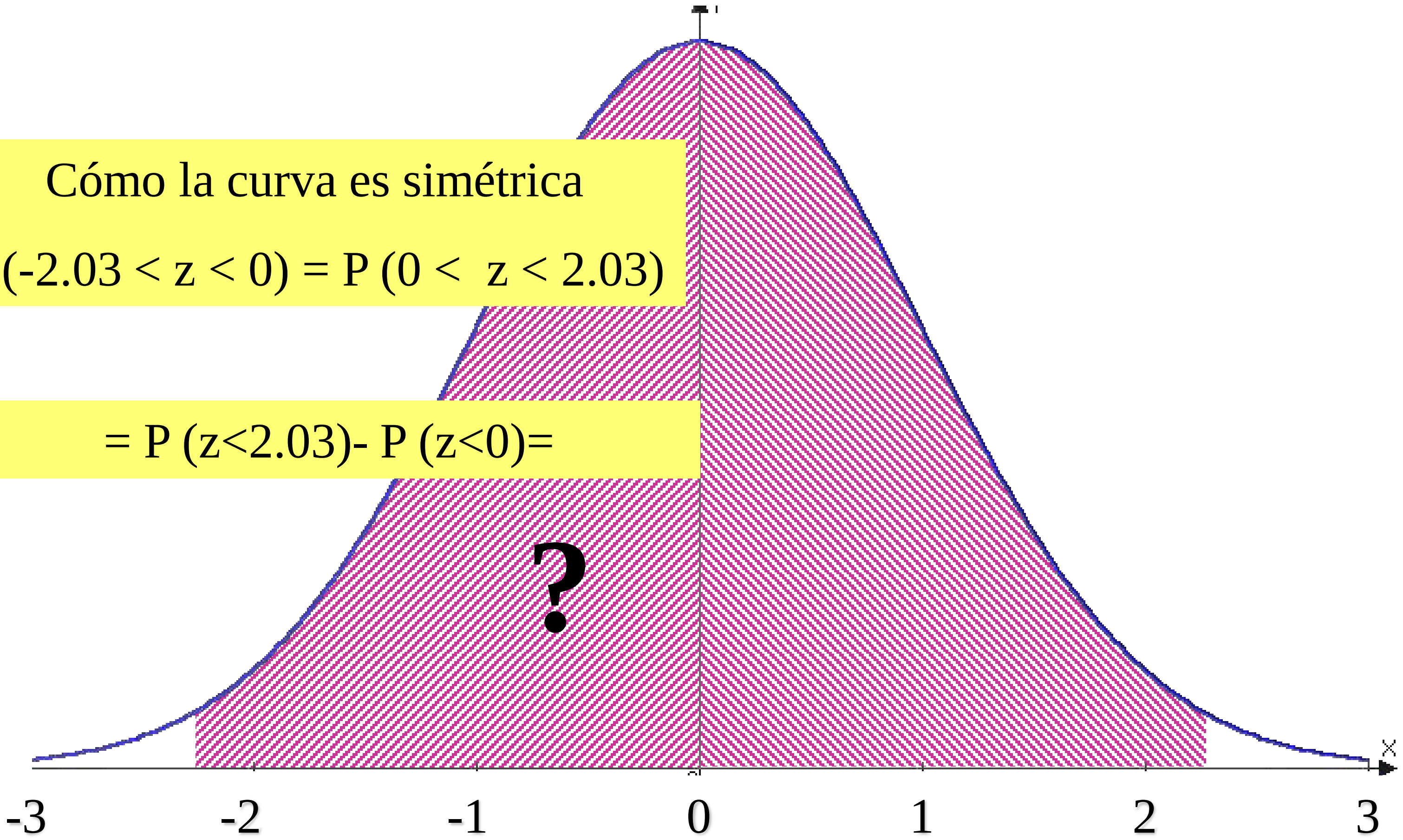
Ejemplo 1

¿Cuál es la probabilidad de que un valor de z esté entre 0 y -2.03?

Cómo la curva es simétrica

$P(-2.03 < z < 0) = P(0 < z < 2.03)$

$= P(z<2.03) - P(z<0) =$

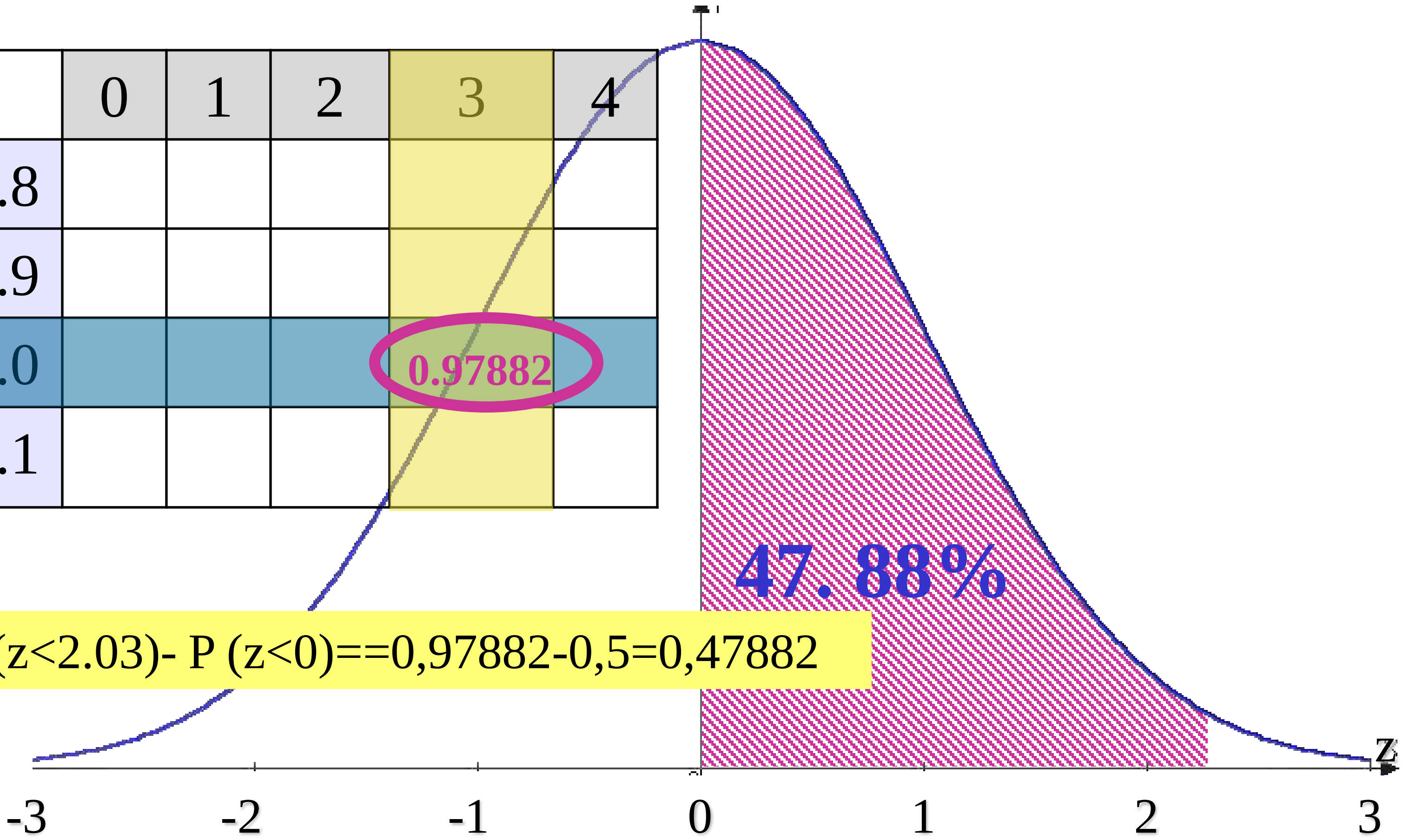


Ejemplo 1

¿Cuál es la probabilidad de que un valor de z esté entre 0 y -2.03?

Se busca en la tabla el área correspondiente a z = 2.03

	0	1	2	3	4
1.8					
1.9					
2.0				**0.97882**	
2.1					



$= P(z<2.03) - P(z<0) == 0{,}97882 - 0{,}5 = 0{,}47882$

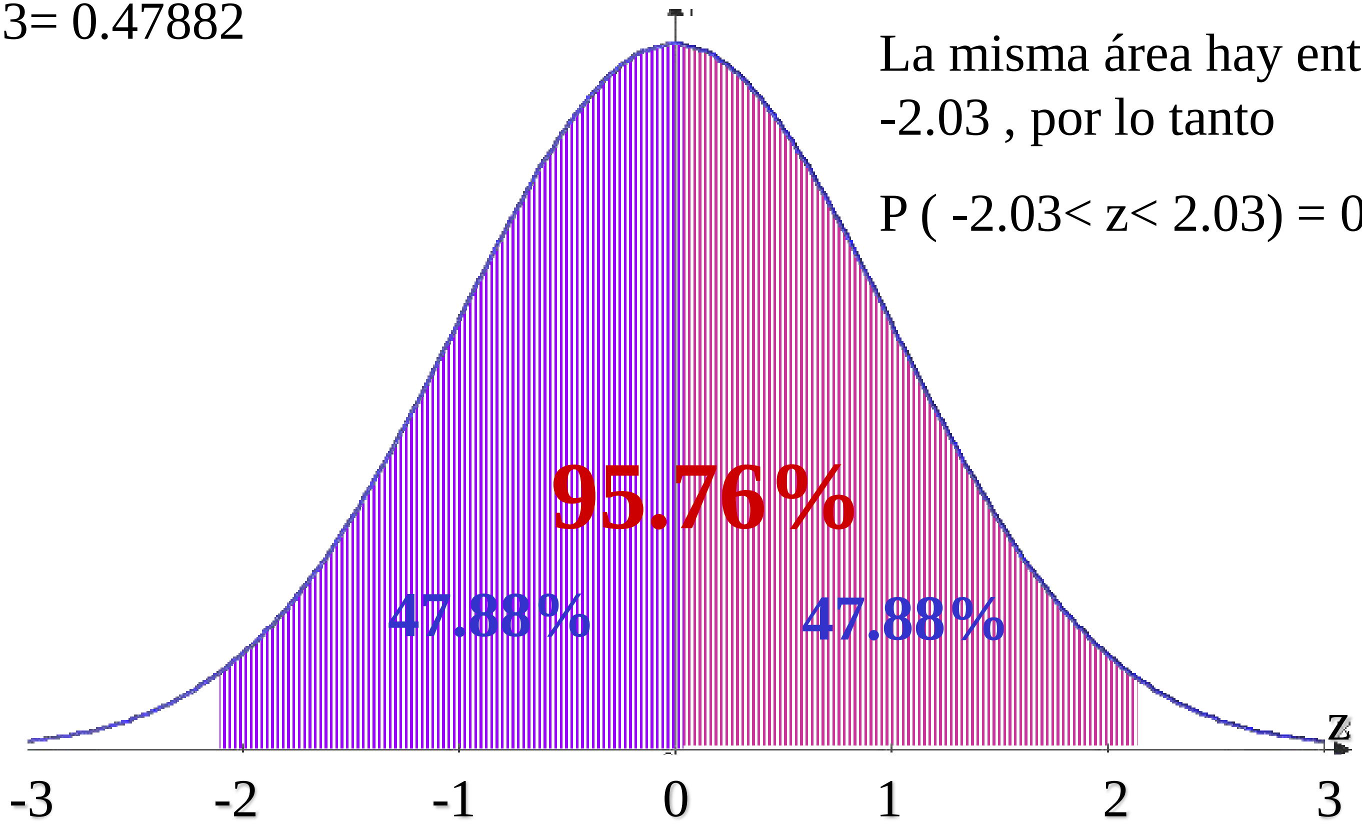
Ejemplo 2

¿Cuál es la probabilidad de que un valor de z esté entre -2.03 y 2.03 ?

En el ejemplo 1, vimos que la probabilidad de que z estuviera entre 0 y 2.03= 0.47882

La misma área hay entre 0 y -2.03 , por lo tanto

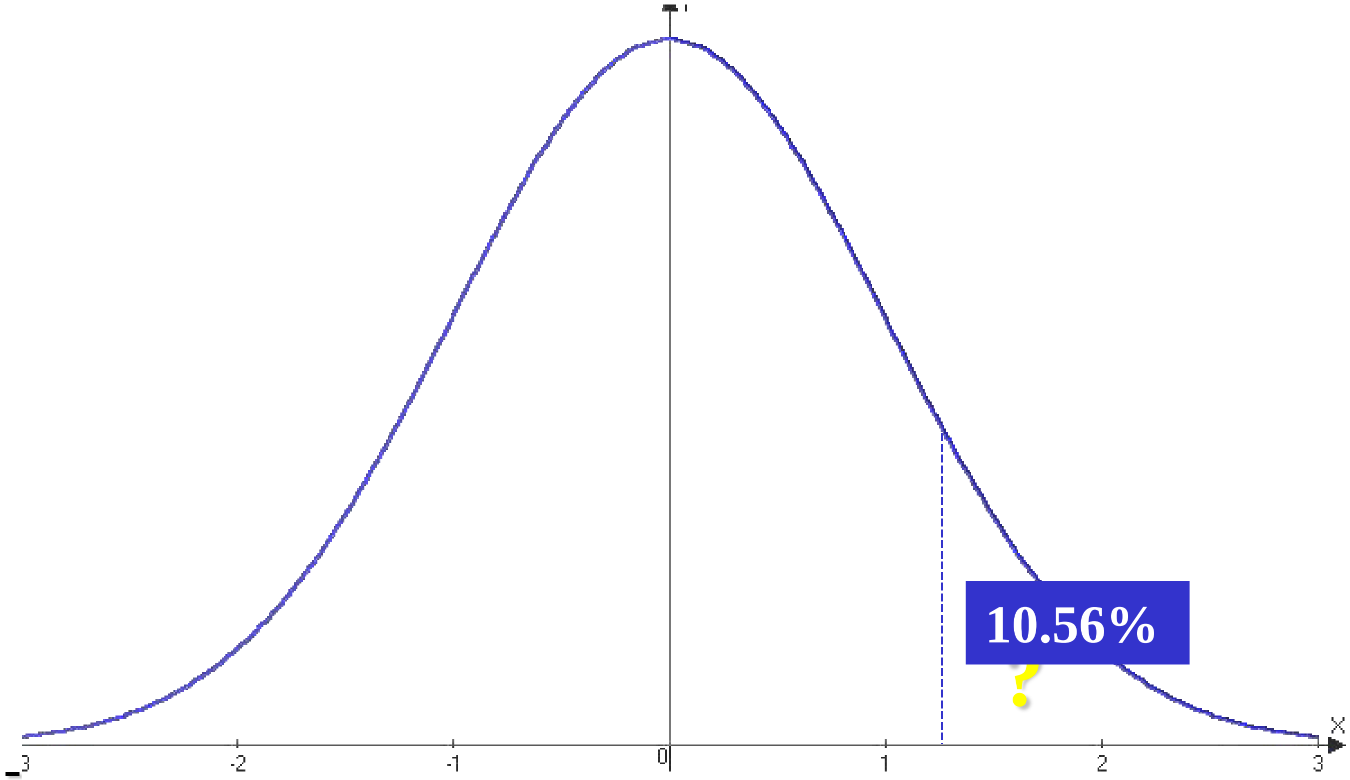
P (-2.03< z< 2.03) = 0.95764



Ejemplo 3

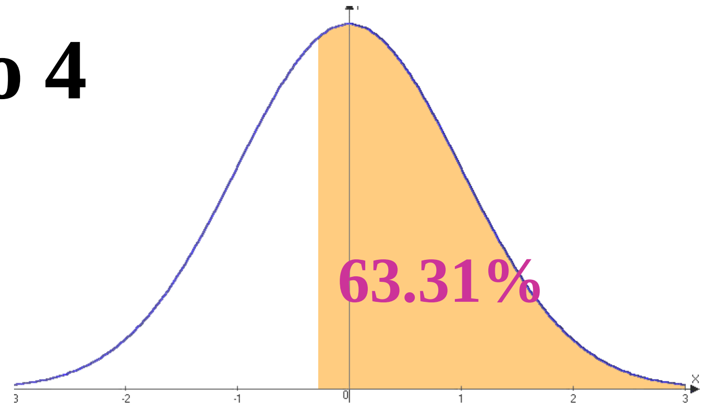
¿Cuál es la probabilidad de que un valor de z sea mayor a 1.25 ?

P(z>1.25) =1-P(z<1,25)=1- 0.8944= 0.10565



Ejemplo 4

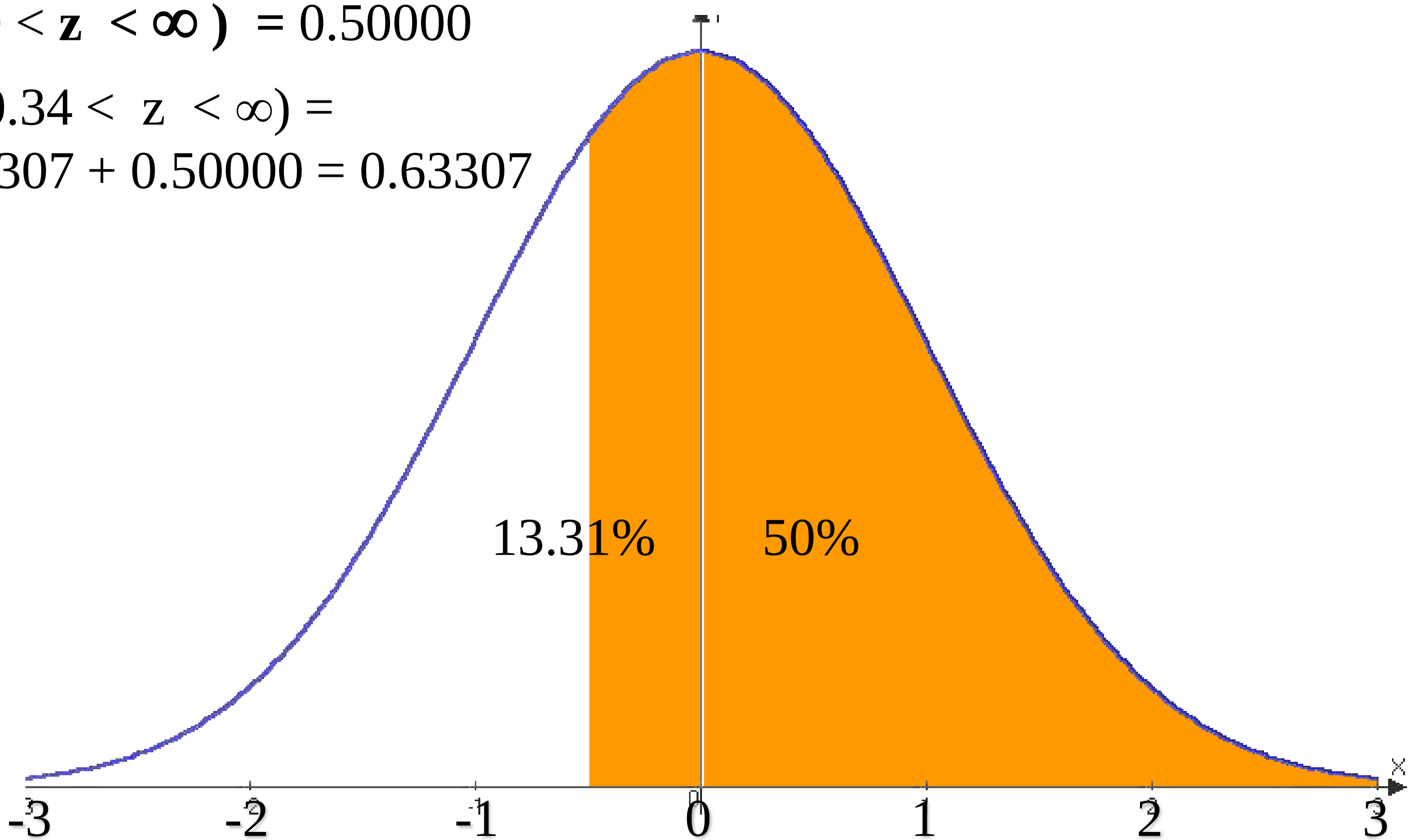
Hallar P(-0.34 < z)



$P(0 < z < 0.34) = 0.13307 = P(-0.34 < z < 0)$

$P(0 < \mathbf{z} < \infty) = 0.50000$

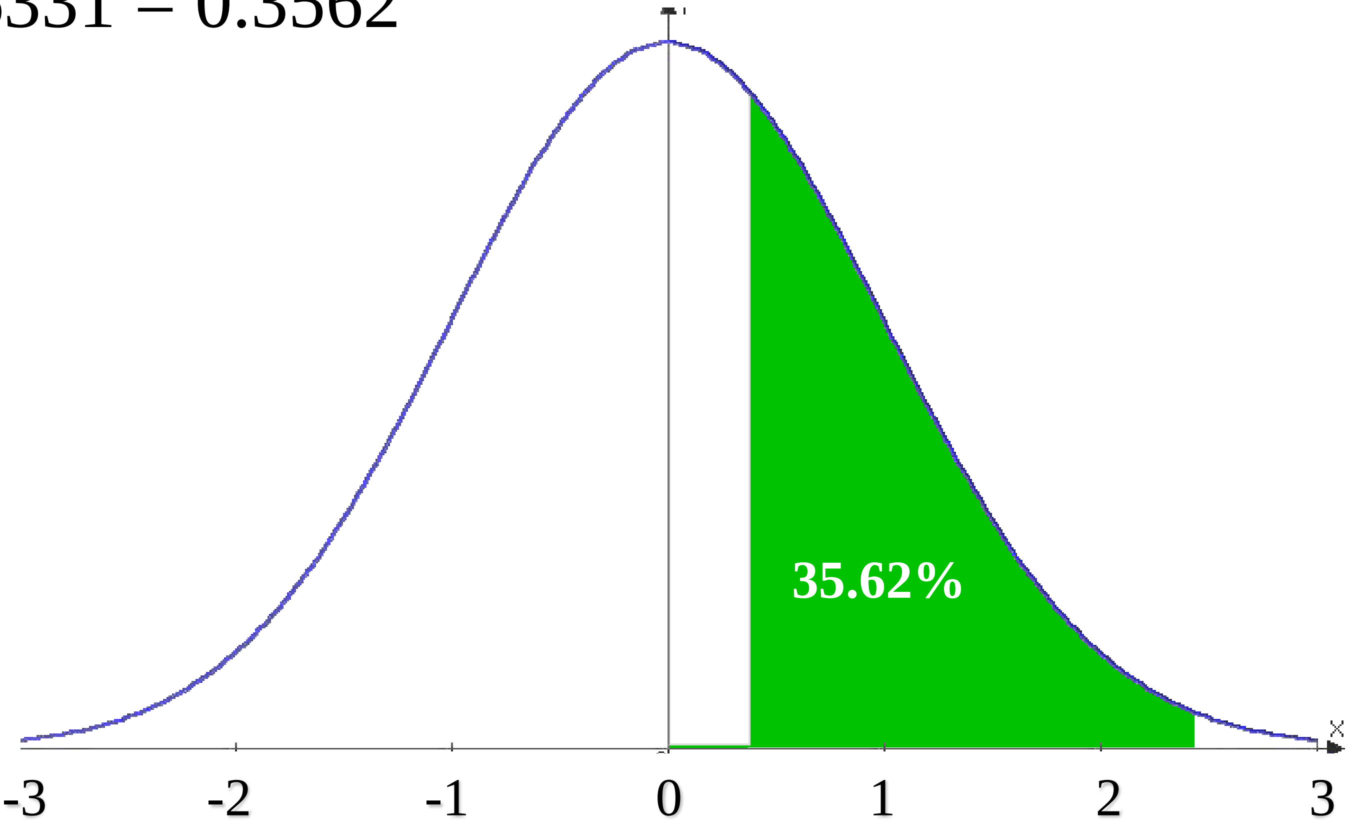
$P(-0.34 < z < \infty) =$
$0.13307 + 0.50000 = 0.63307$



Ejemplo 5

Hallar P(0.34 < z < 2.30)

P (0.34 < z < 2.30) =p(z<2,30)+P(z<0,34)= 0.9893 - 0.6331 = 0.3562



EJEMPLO

Sea una variable distribuida normalmente con media $\mu = 4$ y desviación típica $\sigma = 1.5$.

¿Cuál es la probabilidad de encontrar un valor $x \geq 6$ ($P(x \geq 6)$)?

$\mu = 4$ $\sigma = 1.5$ **Hallar P (x > 6)**

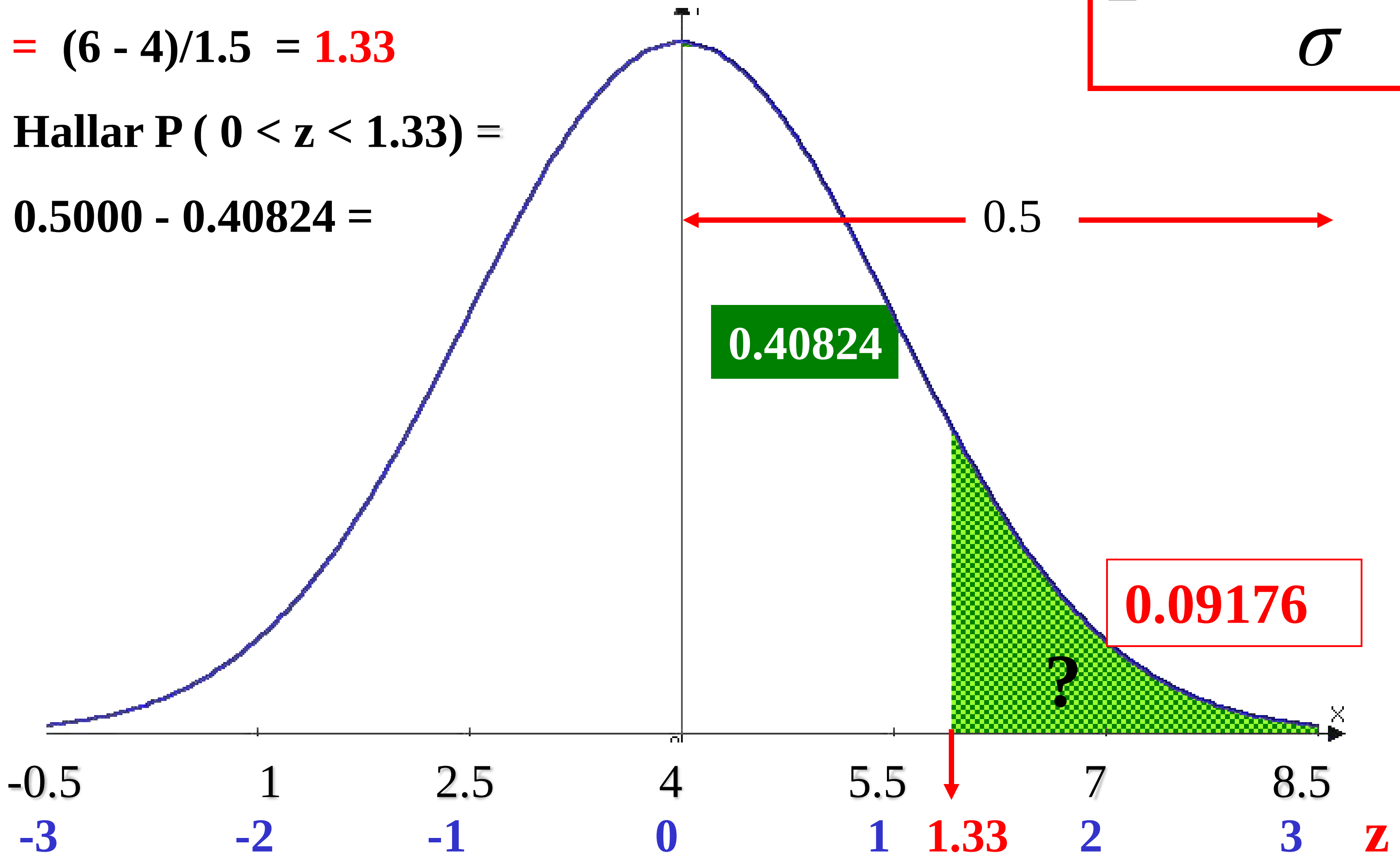
$$z = \frac{x - \mu}{\sigma}$$

1.- **transformar x en un valor de z**

z = (6 - 4)/1.5 = 1.33

2.- Hallar P (0 < z < 1.33) =

3.- 0.5000 - 0.40824 =



Hasta ahora vimos como dado un valor x de la variable, hallar probabilidades transformando (estandarización) la variable en valores de

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

¿Cómo hallar un valor de *x*, dada la probabilidad?

Ejemplo: Sea una variable distribuida normalmente con $\mu = 4$ y $\sigma = 2$. Hallar el valor de x que deja por encima de él un 38.20% (0.3820)

Se debe **desestandarizar** :

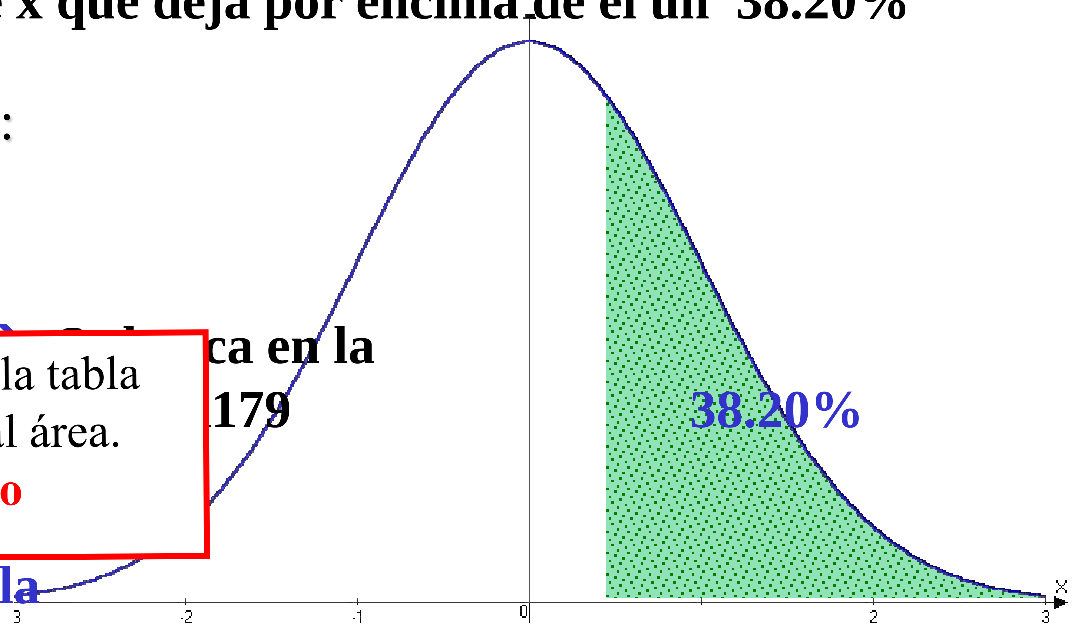
$$x = z\,\sigma + \mu$$

0.5000 - 0.38 … ca en la

tabla el valor … 179

corresponde a

Se busca en la tabla de acuerdo al área.
Con su signo



Sustituyendo en la fórmula

0.30x2+4 =4.60

4.60

Los diseñadores de un nuevo tipo de cabina de avión quieren colocar un interruptor de forma que la mayoría de los pilotos puedan alcanzarlo sin tener que cambiar de posición. Se sabe que la distribución de la distancia máxima (en cms) que pueden alcanzar los pilotos sin moverse (medida) desde el respaldo del asiento es $N(125, 10)$

a) *Si se pone el interruptor a 120cm del respaldo del asiento, ¿qué proporción de pilotos no podrá alcanzarlo sin moverse del asiento?*

La proporción viene dada por $\Pr(X < 120)$ donde $X \sim N(125, 10)$:

$$\begin{aligned} \Pr(X < 120) &= \Pr\left(Z < \frac{120 - 125}{10}\right) = \Pr(Z < -0{,}5) \\ &= 1 - \Pr(Z < 0{,}5) = 0{,}3085 \end{aligned}$$

b) *¿Cuál es la distancia máxima desde el respaldo a la que se podría poner el interruptor si queremos que el 95 % de los pilotos pueda alcanzarlo sin moverse?*

Se busca la distancia d que cumple con

$$0{,}95 = \Pr(X > d) = \Pr\left(Z > \frac{d - 125}{10}\right)$$

de donde sabemos que

$$\frac{d - 125}{10} = -1{,}96$$

y por tanto

$$d = 125 - 19{,}6 = 104{,}6.$$

DISTRIBUCIÓN NORMAL.
Aproximación de una Binomial a la Normal

- Se puede demostrar **(teorema central del límite)** que una v.a. discreta con distribución binomial, $X \rightsquigarrow \mathbf{B}\,(n, p)$ se puede aproximar mediante una distribución normal si *n* es suficientemente grande y *p* no está ni muy próximo a 0 ni a 1. Como el valor esperado y la varianza de *X* son respectivamente np y npq, la aproximación consiste en decir que

$$X \sim B(n,p) donde \begin{cases} n > 30 \\ np > 5 \\ nq > 5 \end{cases} \quad \rightarrow \quad X \sim N(np, npq$$

Se están aproximando probabilidades de una distribución discreta por probabilidades de una distribución continua, se debe aplicar un ***Factor de corrección por continuidad , antes de calcular las probabilidades.*** Cuando el tamaño de la muestra es muy grande entonces el efecto de considerar el factor de corrección por continuidad es insignificante.

Fórmulas de aproximación Normal a la Binomial.

Si X es una Binomial con parámetros n y p, entonces

i) $$P(X = k) \cong P(k - .5 < X < k + .5) = P\left(\frac{k - .5 - np}{\sqrt{npq}} < Z < \frac{k + .5 - np}{\sqrt{npq}}\right)$$

ii) $$P(a < X < b) = P(a + .5 < X < b - .5) = P\left(\frac{a + .5 - np}{\sqrt{npq}} < Z < \frac{b - .5 - np}{\sqrt{npq}}\right)$$

iii) $$P(a \le X \le b) = P(a - .5 < X < b + .5) = P\left(\frac{a - .5 - np}{\sqrt{npq}} < Z < \frac{b + .5 - np}{\sqrt{npq}}\right)$$

Ejemplo. Durante cierta epidemia de gripe, enferma el 30% de la población. En un aula con 200 estudiantes de Medicina, ¿cuál es la probabilidad de que al menos 40 padezcan la enfermedad? Calcular la probabilidad de que haya 60 estudiantes con gripe.

- La V.A. X: el número de alumnos que padece la gripe es

$$X \rightsquigarrow \mathbf{B}\,(n = 200, p = 0,3)$$

cuya media es μ = 200x0,3=60 y cuya σ =200x0,3x0,7=42

Realizar los cálculos con la ley binomial es muy engorroso, ya que intervienen números combinatorios de gran tamaño, y potencias muy elevadas. Por ello utilizamos la aproximación normal de *X*, teniendo en cuenta que se verifican las condiciones necesarias para que el error sea aceptable:

$$X \sim B(n,p) donde \begin{cases} n = 200 > 30 \\ np = 60 > 5 \\ nq = 140 > 5 \end{cases} \rightarrow \quad X \sim N(60, \sigma^2 = 42)$$

- Con una N(60; 6,48)

$$z = \frac{X - \mu}{\sigma} = \frac{40 - 60}{6{,}48} = -3{,}09$$

$$P(X \geq 40) = P(z \geq -3{,}09) = P(z < 3{,}09) = 0{,}999$$

Ejemplo: Se sabe que el 40 % de los habitantes de cierta ciudad consumen diariamente café.

a) Se encuesta a veinte personas sobre su consumo diario de café. Sea X el número de personas que consumen café cada día. Calcular la probabilidad de que X sea igual a 12. hallar la media y la desviación típica de x. Obtener la probabilidad de que ninguna tome café y de que al menos dos lo hagan.

b) Si se encuesta a 1000 personas ¿cuál es la probabilidad de que entre el 35,3% y el 44,7% consuman café?

c) Si se pregunta a 10.000 personas ¿la probabilidad anterior es mayor o menor que la anterior?

P{consumir café}=0.4

a) $X_i = \begin{cases} 1, & \text{toma café} \\ 0, & \text{no toma café} \end{cases}$ es una Bernoulli con probabilidad p=0.4.

Su media es $E[X_i] = p = 0.4$, y su varianza es $var(X_i) = pq = 0.24$, por lo que su desviación típica es $\sigma(X_i) = 0.4899$.

b) X es una binomial con $n = 20$, $p = 0.4$, con media $E[X] = np = 8$, $var(X) = npq = 4.8$; $\sigma(X) = 2.1909$.

La probabilidad de que X sea 12 es:

$P\{X = 12\} = \binom{20}{12} 0.4^{12} 0.6^{8} = 0.0355;$

La probabilidad de que ninguna tome café diariamente es:

$P\{X = 0\} = \binom{20}{0} 0.4^{0} 0.6^{20} = 3.65 * 10^{-5};$

y $P\{\text{al menos dos tomen café}\} = 1 - P\{X = 0\} - P\{X = 1\} =$

$= 1 - \binom{20}{0} 0.4^{0} 0.6^{20} - \binom{20}{1} 0.4^{1} 0.6^{19} = 0.995;$

c) Si n=1000, podemos aproximar la binomial B(1000,0.4) a una normal N(400,15.4919).

$$\begin{aligned} P\{353 &\leq X \leq 447\} \approx P\{\frac{353-400}{15.4919} \leq Z \leq \frac{447-400}{15.4919}\} = \\ &= P\{-3.0338 \leq Z \leq 3.0338\} = \\ &= 2P\{Z \leq 3.0338\} - 1 = \\ &= 2 \times 0.9988 - 1 = 0.9976 \end{aligned}$$

con $Z \equiv N(0,1)$.

d) Si n=10000, aproximamos la binomial B(10000,0.4) a una normal N(4000,48.9898) y tenemos.

$$\begin{aligned} P\{3530 &\leq X \leq 4470\} \approx P\{\frac{3530-4000}{48.9898} \leq Z \leq \frac{4470-4000}{48.9898}\} = \\ &= P\{-9.5938 \leq Z \leq 9.5938\} = \\ &= 2P\{Z \leq 9.5938\} - 1 = \\ &= 2 \times 1 - 1 = 1 \end{aligned}$$

con $Z \equiv N(0,1)$.

Al aumentar la muestra, la probabilidad aumenta.