UNIDAD DIDÁCTICA 4

DISTRIBUCIONES DE PROBABILIDAD

(2ª parte)

Contenidos

Distribuciones de probabilidad continuas

Función de densidad

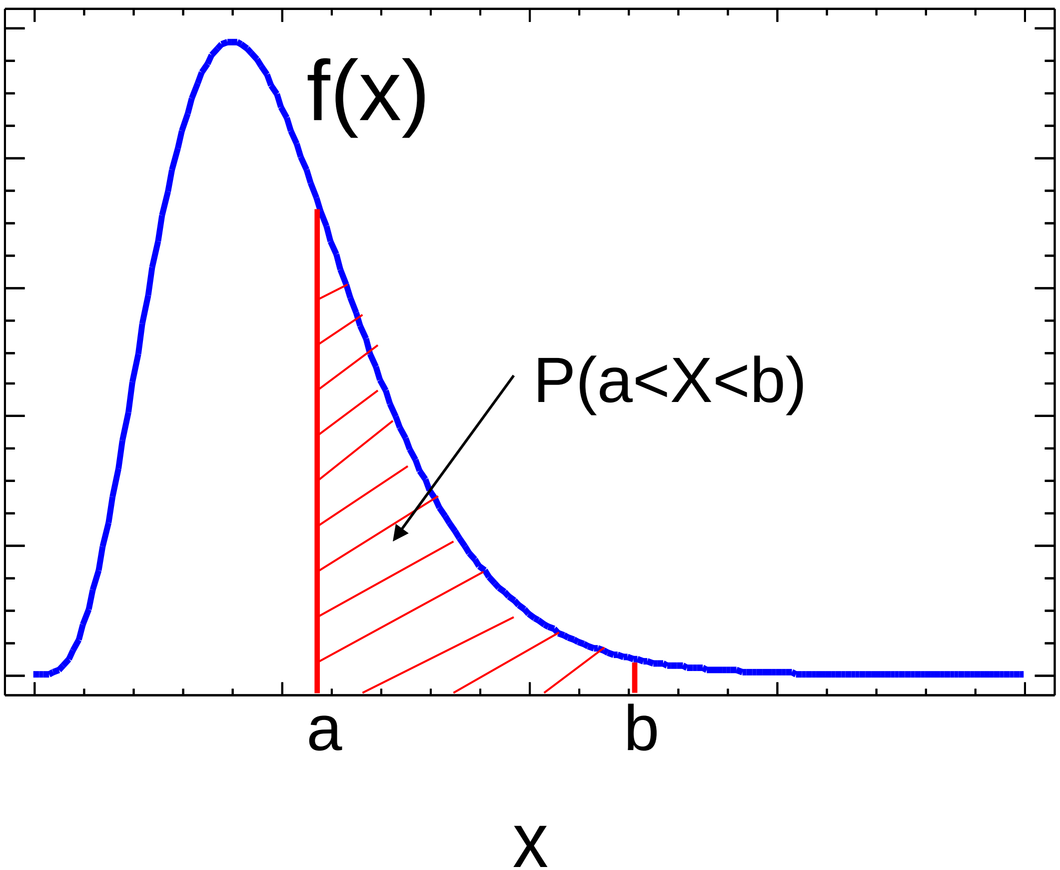
- La distribución de probabilidad de una variable continua **X,** se caracteriza con la función de densidad**:**

f(x)

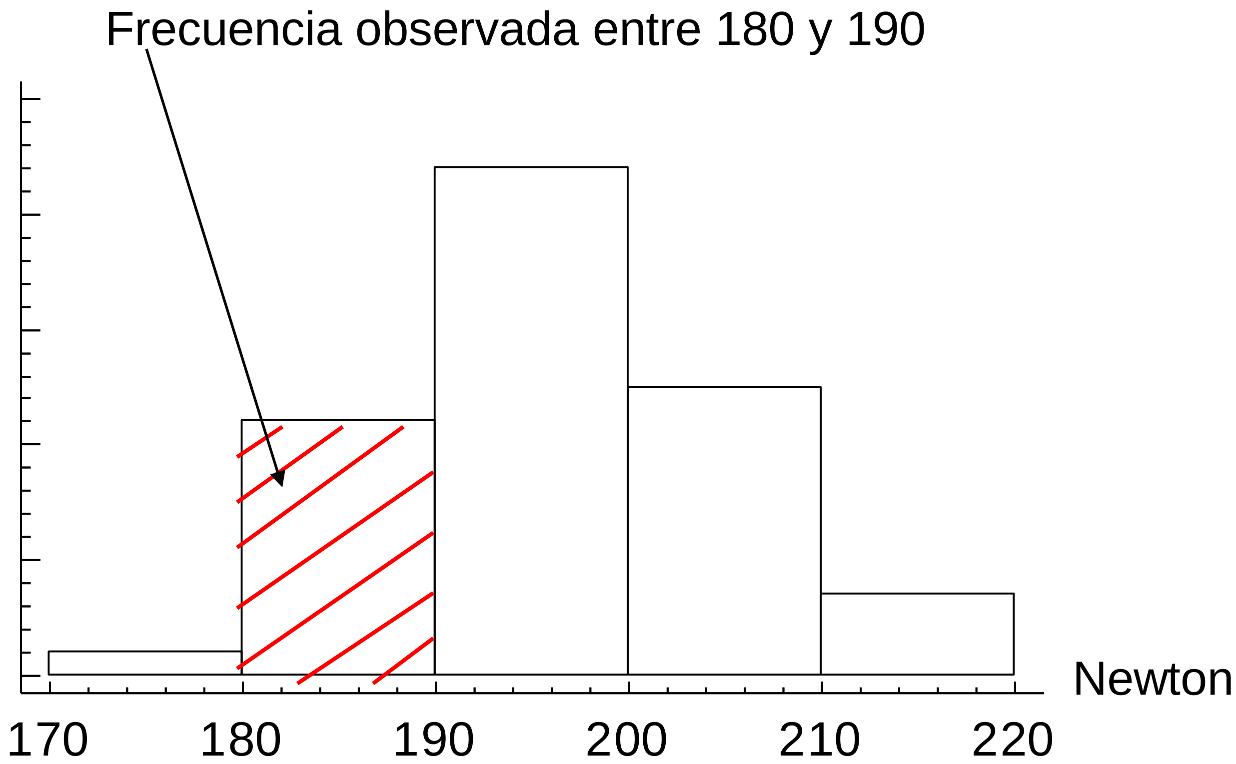
- El área comprendida bajo la función de densidad entre dos valores "a" y "b", da la probabilidad de que la variable tome valores en dicho intervalo :

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

Distribuciones de probabilidad continuas

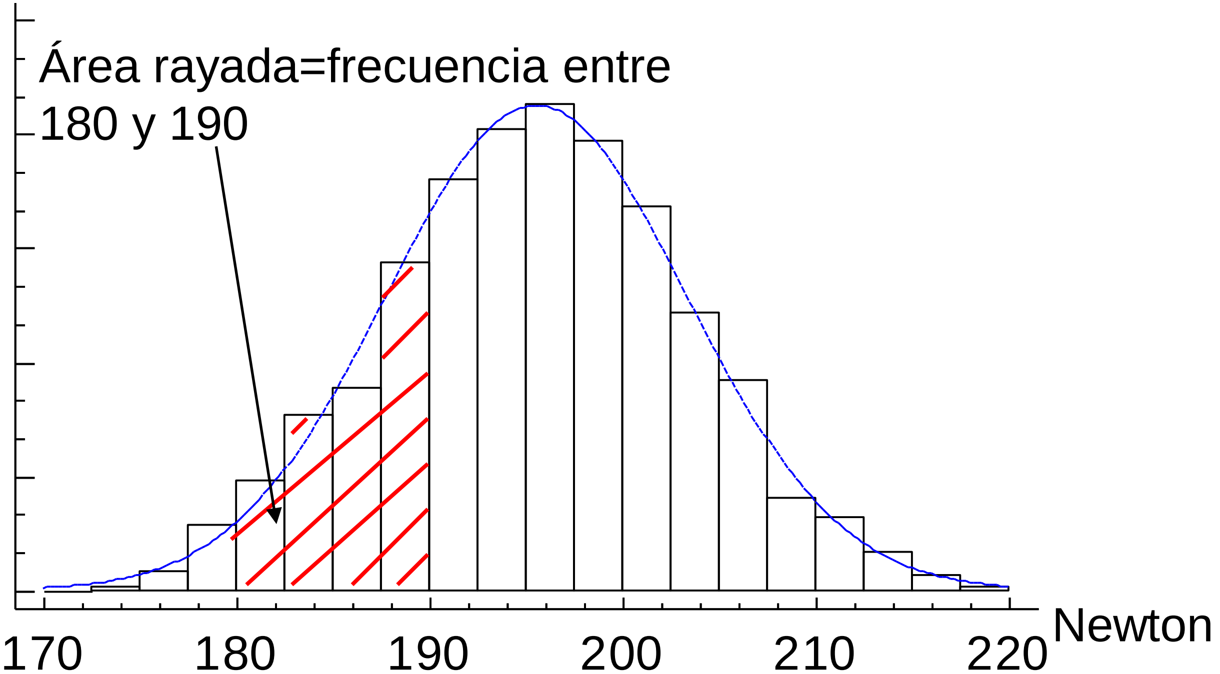


Distribuciones continuas. Función de densidad



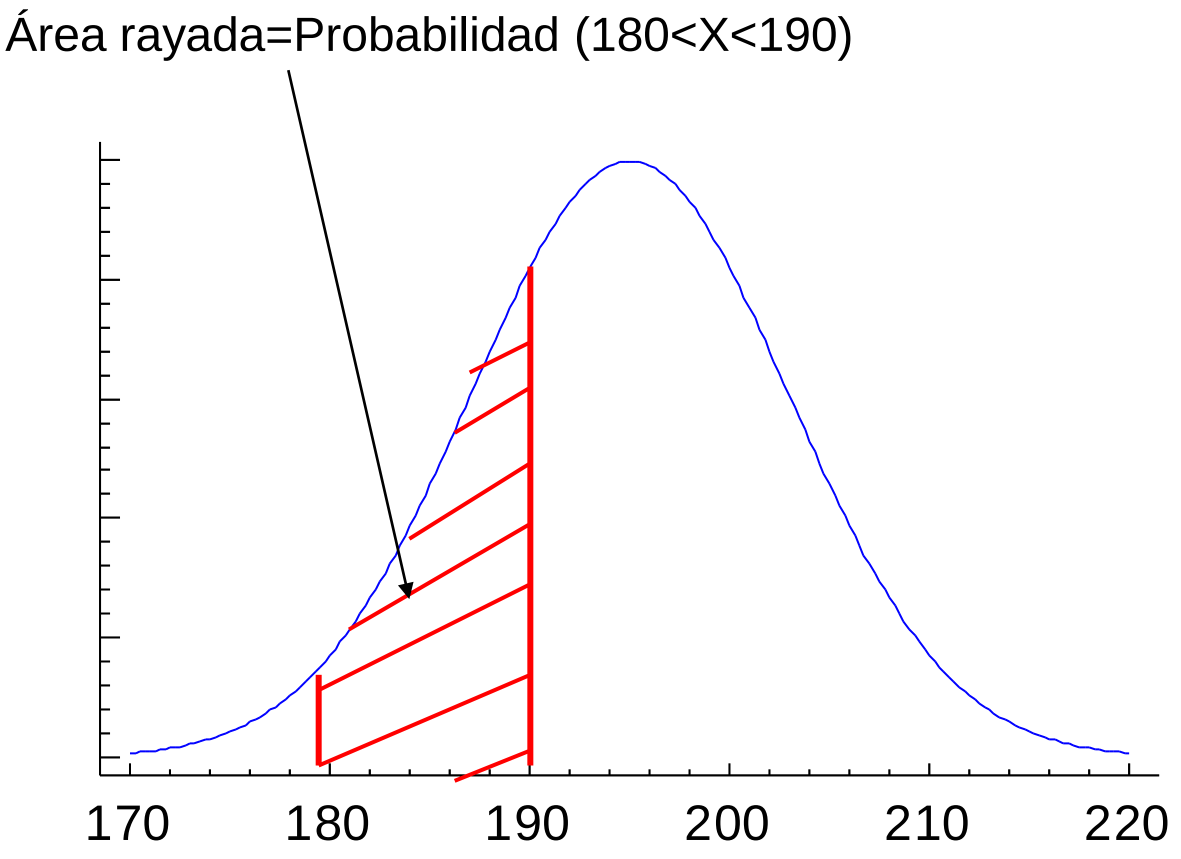
Histograma para N=100

Distribuciones continuas. Función de densidad



Histograma para N=1000

Distribuciones continuas. Función de densidad



"Histograma" para N=∞

Distribución Uniforme

- Variable aleatoria **X** continua con función de densidad constante en un intervalo (a,b):

Distribución Uniforme en (a,b)

$$\boxed{X \sim U(a,b)}$$

Distribución Uniforme

Ejemplos:

- Función que genera números aleatorios en el intervalo (0, 1) **Uniforme(0,1)**
- La distribución uniforme puede utilizarse también para modelizar la variabilidad del tiempo de acceso a datos en unidades de discos, tiempo que será mayor o menor según la posición de la cabeza de lectura respecto a la situación del dato buscado.
- También puede usarse para caracterizar el comportamiento del tamaño de un determinado tipo de fichero (entre 100 y 1000 Kb por ejemplo) generado por una aplicación.

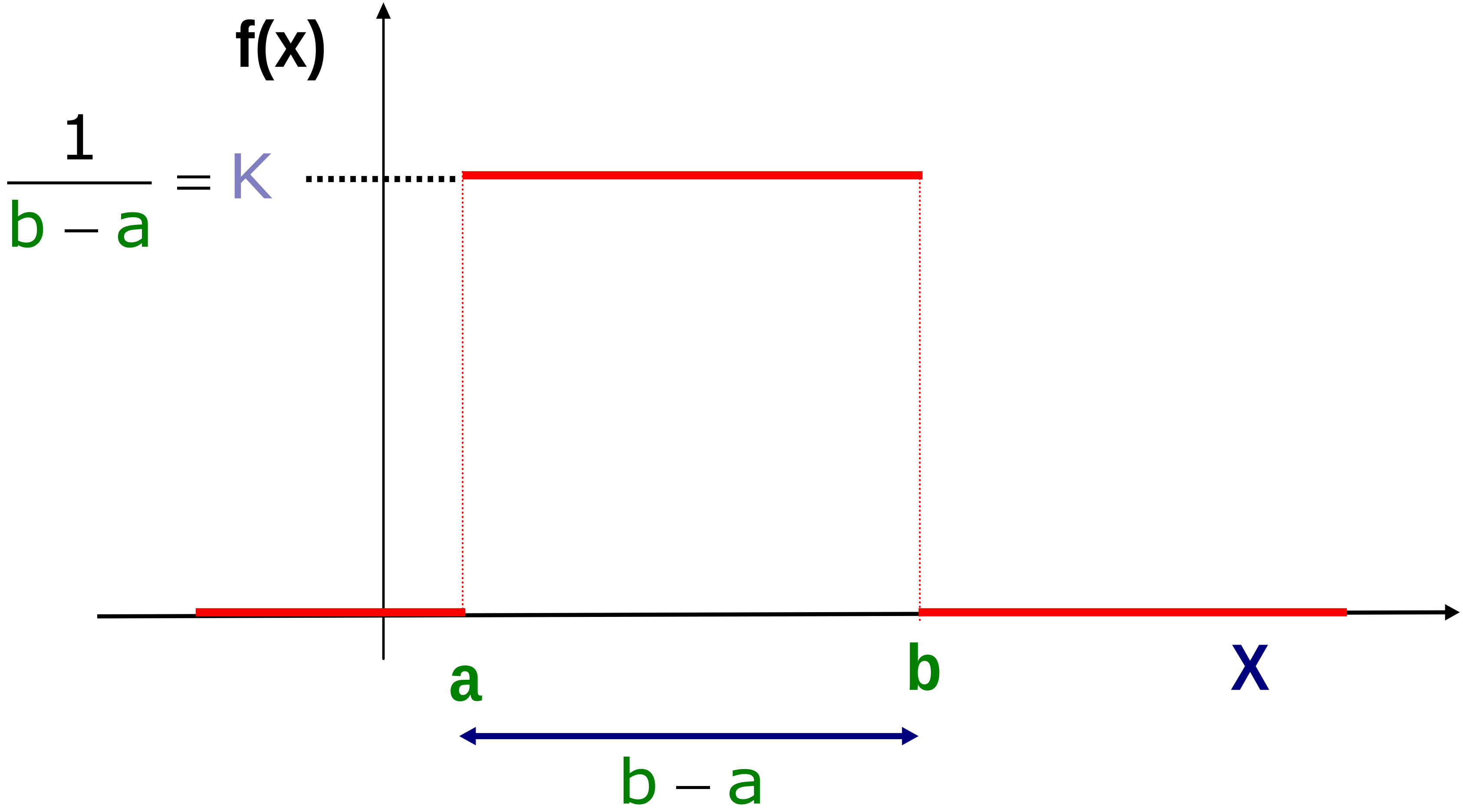
Distribución Uniforme

- La **función de densidad** de este tipo de variables es:

$$\mathbf{f(x) = K = \frac{1}{b - a} \quad x \in [a, b]}$$

$$\mathbf{f(x) = 0 \quad x \notin [a, b]}$$

Distribución Uniforme



Distribución Uniforme

- Por tanto, la **probabilidad P(X≤x)** puede calcularse, integrando f(x), lo que da la siguiente función:

$$P(X \leq x) = 0 \quad x < a$$

$$P(X \leq x) = \frac{x - a}{b - a} \quad a < x \leq b$$

$$P(X \leq x) = 1 \quad x > b$$

Distribución Uniforme

- **Media**

$$E(X) = \frac{a + b}{2}$$

- **Varianza**

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

Distribución Uniforme

Ejercicio 12: *el tiempo de acceso a un fichero fluctúa* ***uniformemente*** *entre 0,1 y 0,5 segundos.*

a) ¿Cuál es la probabilidad de que el tiempo de acceso sea inferior a 0,125 segundos?

b) ¿Cuál es la media del tiempo de acceso? ¿Y la desviación típica?

c) ¿Cuál es el valor del tiempo t que no es superado por el 50% de los accesos?

Distribución Uniforme

SOLUCIÓN:

Distribución $\boldsymbol{X} \sim U(0{,}1,\ 0{,}5)$.

a) $P(\boldsymbol{X}<0{,}125)=P(U(0{,}1,\ 0{,}5)<0{,}125)=$

$$=(0{,}125-0{,}1)/(0{,}5-0{,}1)=\mathbf{0{,}0625}$$

Distribución Uniforme

SOLUCIÓN:

b) Media $m=E(\mathbf{X})=(0,1+0,5)/2=\mathbf{0,3}$

Desviación típica$=\sigma=[(0,5-0,1)^2/12]^{½}=$

$=\mathbf{0,1154}$

c) $x?$ $P(\mathbf{X}<x)=0,5 \Rightarrow (x-0,1)/(0,5-0,1)=0,5 \Rightarrow$

$\Rightarrow x=0,5 \text{ x } 0,4+0,1=\mathbf{0,3}$

Distribución Exponencial

Variable aleatoria continua **X**

- La probabilidad de que la variable sea mayor que un valor de **X** cualquiera (**t**) disminuye exponencialmente conforme dicho valor **t** aumenta.
- Los valores que puede tomar son ≥ 0

$$X \sim \mathrm{Exp}(\alpha)$$

Distribución Exponencial

La distribución Exponencial se utiliza en:

- Vida o duraciones de equipos y/o componentes **(Fiabilidad)** como el tiempo hasta el fallo en equipos industriales o el tiempo hasta el fallo de componentes electrónicos, …
- Tiempo que se tarda en realizar un proceso **(Sistemas de Colas)** como tiempo que se tarda en atender una consulta en una BDD, tiempo de ejecución de una instrucción, tiempo que puede transcurrir en un servicio de urgencias para la llegada de un paciente, tiempo que transcurre entre la ocurrencia de dos sucesos consecutivos en un Proceso de Poisson…

Distribución Exponencial

- **Función de densidad:**

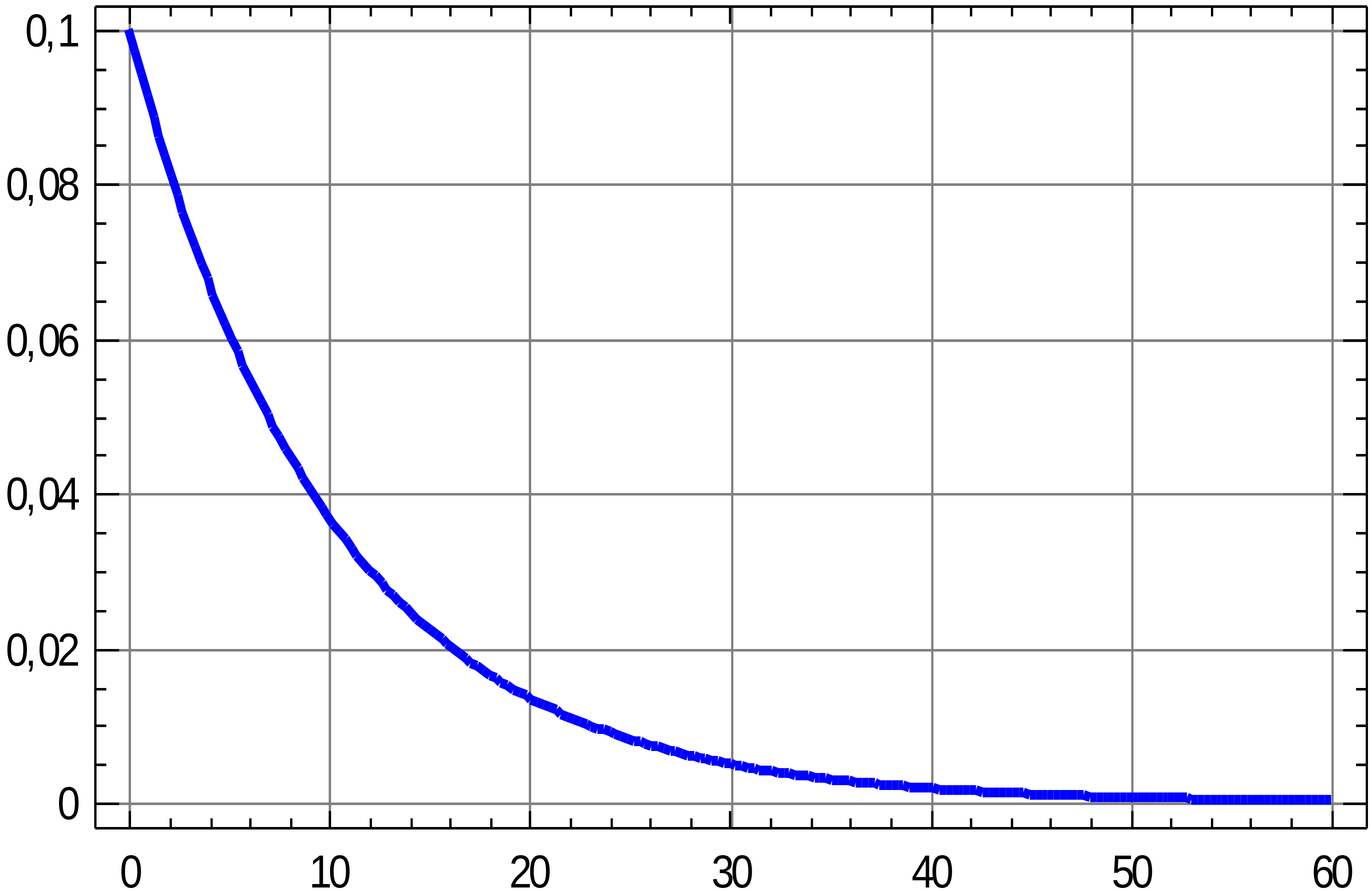
$$f(x) = \alpha e^{-\alpha x} \qquad x \geq 0$$

$$f(x) = 0 \qquad x < 0$$

Distribución Exponencial

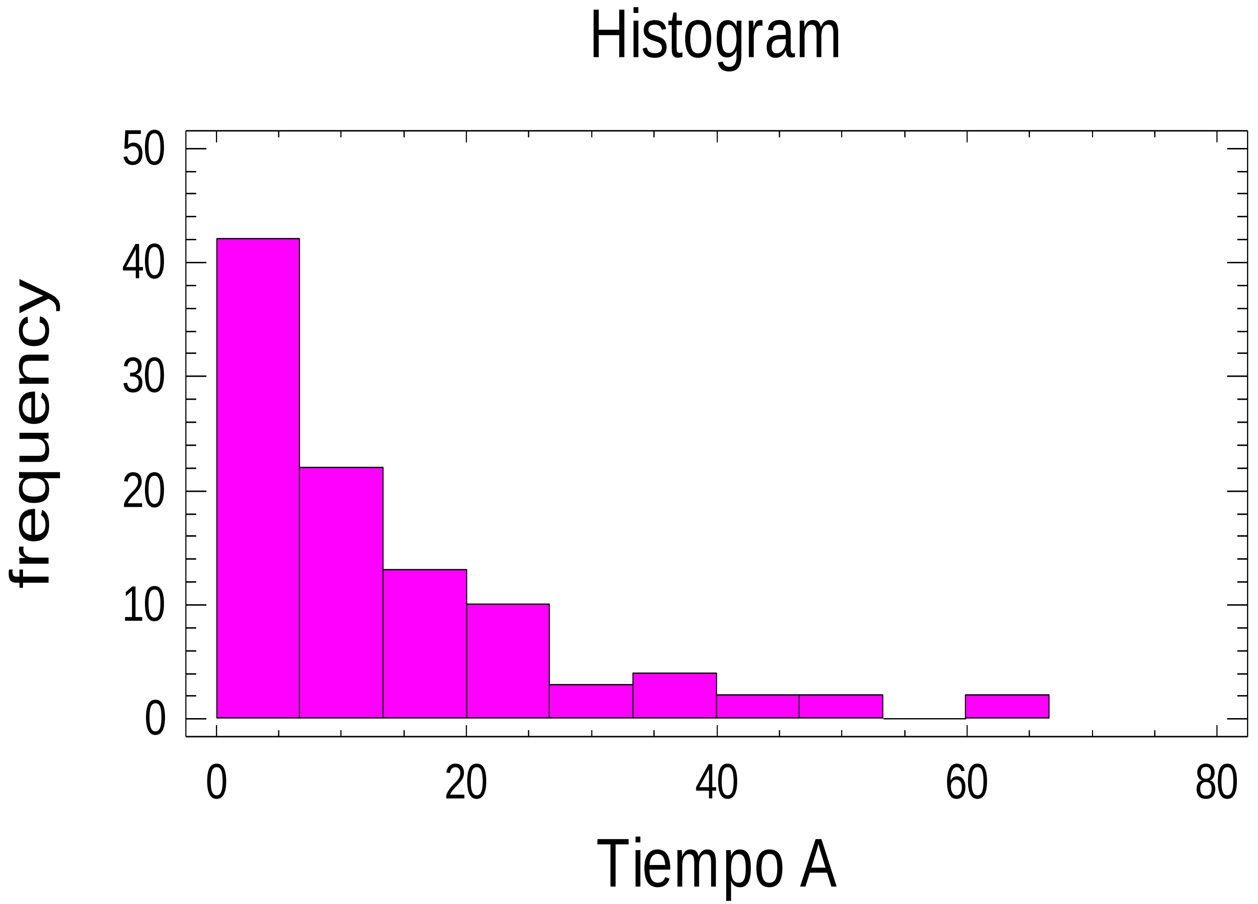
f(x) =

$= 0{,}1e^{-0{,}1x}$

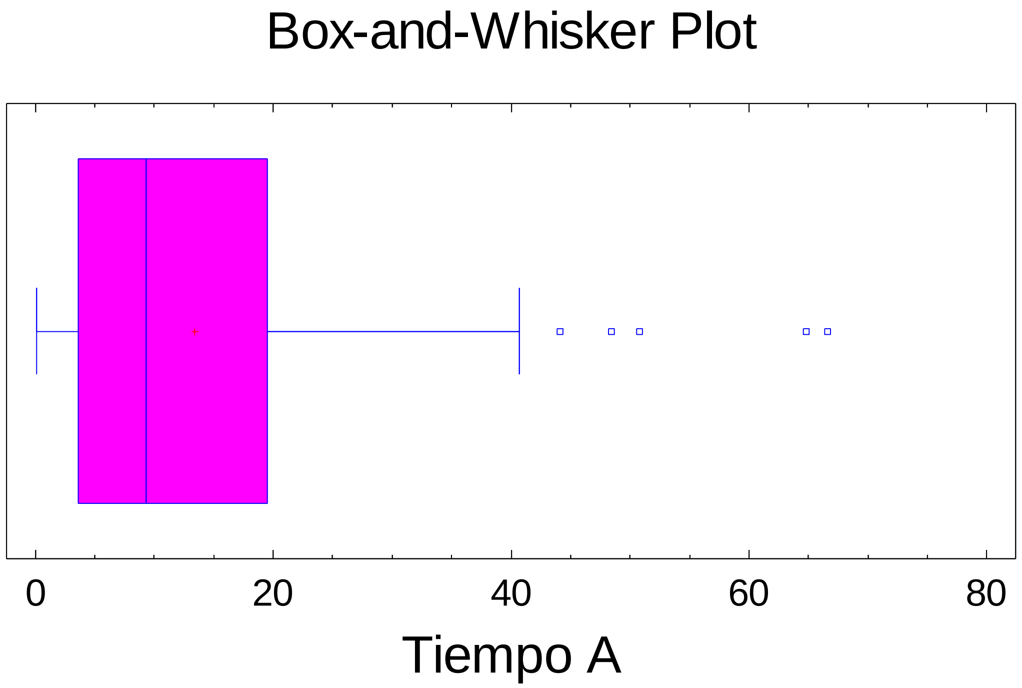


X ≥ 0

Ejemplo



Ejemplo



Summary Statistics for Tiempo A

Count = 100

Average = 13,3625

Median = 9,35744

Variance = 193,792

Standard deviation = 13,9209

Minimum = 0,128046

Maximum = 66,5925

Range = 66,4645

Lower quartile = 3,57336

Upper quartile = 19,5418

Interquartile range = 15,9684

Stnd. skewness = 7,28525

Stnd. kurtosis = 6,80917

Coeff. of variation = 104,179%

Distribución Exponencial

$$P(X \leq x) = 1 - e^{-\alpha x} \qquad x \geq 0$$

$$P(X \leq x) = 0 \qquad x < 0$$

Distribución Exponencial

- **Media** y **varianza:**

$$E(X) = \frac{1}{\alpha}$$

$$\sigma^2(X) = \frac{1}{\alpha^2}$$

Distribución Exponencial

Ejercicio 13: *La duración T (tiempo de funcionamiento hasta que fallan) de las pantallas fluctúa exponencialmente. Se sabe que la vida media de las pantallas es 40 años.*

a) *¿Cuánto vale la desviación típica?*

b) *¿Qué porcentaje dura más de 10 años?*

c) *Suponiendo que las pantallas tienen 1 año de garantía, ¿Qué porcentaje de las mismas tendrá que usar la garantía?*

d) *¿Cuál es la mediana de la duración?*

Distribución Exponencial

SOLUCIÓN:

T*=horas de funcionamiento hasta el fallo de las pantallas.*

Distribución ***T****~Exp(α)*

Media m = ***40 años*** *(α=1/40)*

a)Desviación típica σ = ***40 años***

b) $P(T>10)=e^{-10/40}=0{,}779 \Rightarrow$ ***77,9%***

Distribución Exponencial

SOLUCIÓN:

c) $P(\mathbf{T}<1)=1-e^{-1/40}=0{,}025 \Rightarrow$ **2,5%**

d) $P(T>mediana)=0{,}5 \Rightarrow e^{-mediana/40}=0{,}5 \Rightarrow$

$(-mediana/40)=\ln 0{,}5 \Rightarrow$

$mediana=-(\ln 0{,}5) \times 40=$ **27,73 años**

Distribución Exponencial

Ejercicio 14: *La vida **X** de las pilas eléctricas de una determinada marca se distribuye exponencialmente con media 50 horas:*

a) *Si una pila lleva ya funcionando sin fallo 50 horas, ¿cuál es la probabilidad de que falle a lo largo de las 10 horas siguientes?.*
b) *Y si una pila lleva ya funcionando sin fallo 100 horas, ¿cuál es la probabilidad de que falle a lo largo de las 10 horas siguientes?.*

Distribución Exponencial

SOLUCIÓN:

a) P(falle a lo largo de las 10 horas siguientes/lleva ya funcionando 50 horas)= =P(**X**<60/**X**>50)

$$P(X<60/X>50)=\frac{P(\{X<60\}\{X>50\})}{P(X>50)}=$$

$$=\frac{P(50<X<60)}{P(X>50)}=\frac{P(X<60)-P(\ X\ <50)}{1-P(\ X\ \leq 50)}=\frac{1-e^{-\frac{60}{50}}-1+e^{-\frac{50}{50}}}{1\text{-}1+e^{-\frac{50}{50}}}=-e^{-\frac{10}{50}}+1=0{,}18$$

Distribución exponencial

SOLUCIÓN:

b) P(falle a lo largo de las 10 horas siguientes/lleva ya funcionando 100 horas) =P(X<110/X>100)

$$P(X<110/X>100)=\frac{P(\{X<110\}\{X>100\})}{P(X>100)}=$$

$$=\frac{P(100<X<110)}{P(X>100)}=\frac{P(X<110)-P(X<100)}{1-P(X\leq 100)}=\frac{1-e^{-\frac{110}{50}}-1+e^{-\frac{100}{50}}}{1\text{-}1+e^{-\frac{100}{50}}}=-e^{-\frac{10}{50}}+1=0{,}18$$

Distribución Normal

Distribución de frecuencias

- Acumulación de valores en zona central.
- Decrecen de forma aproximadamente simétrica a medida que éstos se alejan del centro.

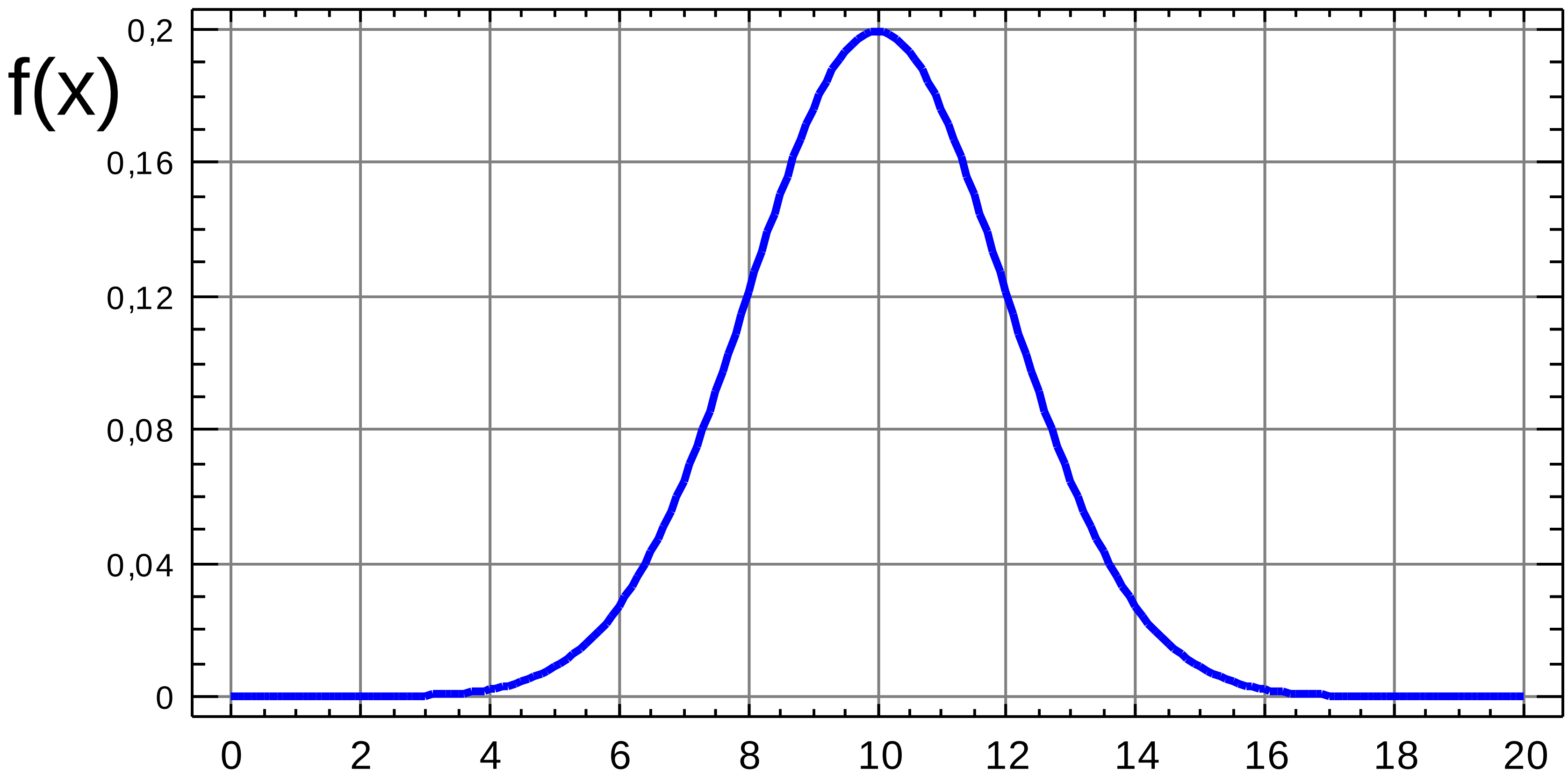
$$\boxed{\mathbf{X} \sim N(m,\sigma)}$$

Distribución Normal

- Función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$$

Distribución Normal



media = mediana = moda = 10

X~N(m=10, σ=2)

Distribución Normal. Propiedades

Propiedades de la distribución normal:

1. Coeficientes de asimetría y de curtosis nulos.

2. Si una variable aleatoria **X** se distribuye normalmente con media m y desviación típica σ:

$$P(m-\sigma < \mathbf{X} < m+\sigma)=0{,}683$$

$$P(m-2\sigma < \mathbf{X} < m+2\sigma)=0{,}954$$

$$P(m-3\sigma < \mathbf{X} < m+3\sigma)=0{,}997$$

Distribución Normal. Propiedades

3. Si $\mathbf{X \sim Normal(m_x, \sigma_x)}$

$Y = a + bX \Rightarrow Y \sim \text{Normal}(a + bm_x,\ b\sigma_x)$

4. Si $X_1 \sim N(m_{x1}, \sigma_{x1})$ $X_2 \sim N(m_{x2}, \sigma_{x2})$ **independientes**

$Y = X_1 + X_2 \Rightarrow Y \sim N(m_{x1} + m_{x2}, \sqrt{\sigma_{X1}^2 + \sigma_{X2}^2}\ )$

5. Si $X_i \sim N(m_{xi}, \sigma_{xi})$ $i = 1 \ldots N$ **independientes**

$Y = X_1 + \ldots + X_N \Rightarrow Y \sim N(m_{x1} + \ldots + m_{xN}, \sqrt{\sigma_{X1}^2 + \ldots + \sigma_{XN}^2})$

Tabla de la distribución Normal

- Una variable normal es **tipificada** si su media es cero y su desviación típica es uno.

$$Z \sim N[0,1]$$

- La probabilidad P(N(0,1)>z) (Tabla)

$$P(N(0,1) > z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Tabla de la distribución normal

Ejercicio 15: *utilizando la tabla de probabilidades de la N[0,1] calcular:*

a) La probabilidad de que una variable N[0,1] sea mayor que 2.

b) La probabilidad de que una variable N[0,1] sea mayor que -1.

c) La probabilidad de que una variable N[0,1] esté comprendida entre -1 y 2

Tabla de la distribución normal

SOLUCIÓN:

a) $P(N(0,1)>2) = \mathbf{0{,}0228}$

b) $P(N(0,1)>-1) = 1-P(N(0,1)<-1) = 1-0{,}1587 =$

$= \mathbf{0{,}8413}$

c) $P(-1 < N[0,1] < 2) =$

$= P(N[0,1] < 2) - P(N[0,1] < -1)=$

$=1 - P(N[0,1] > 2) - P(N[0,1] < -1)=$

$=1- 0{,}0228 – 0{,}1587 = \mathbf{0{,}8185}$

Tabla de la distribución normal

- Si **X** es N[m,σ]:

$$P(X > z) = P\left(\frac{x - m}{\sigma} > \frac{z - m}{\sigma}\right) = P\left(N(0,1) > \frac{z - m}{\sigma}\right)$$

Tabla de la distribución normal

Ejercicio 16: *La dureza de los asientos fabricados en una factoría fluctúa normalmente con media 185 newtons y desviación típica 12 newtons.*

¿Qué porcentaje de los asientos fabricados cumplirá las especificaciones establecidas que son de 180 ± 20 newtons?

Tabla de la distribución normal

SOLUCIÓN:

$$P(160 < dureza < 200) = P(160 < N[185,12] < 200) =$$

$$= P\left(\frac{160-185}{12} < N(0,1) < \frac{200-185}{12}\right) =$$

$$= P(-2{,}08 < N(0,1) < 1{,}25) =$$

$$= P(N(0,1) < 1{,}25) - P(N(0,1) < -2{,}08) =$$

$$= 1 - P(N(0,1) > 1{,}25) - P(N(0,1) < -2{,}08) =$$

$$= 1 - 0{,}1056 - 0{,}0188 = 0{,}8756$$

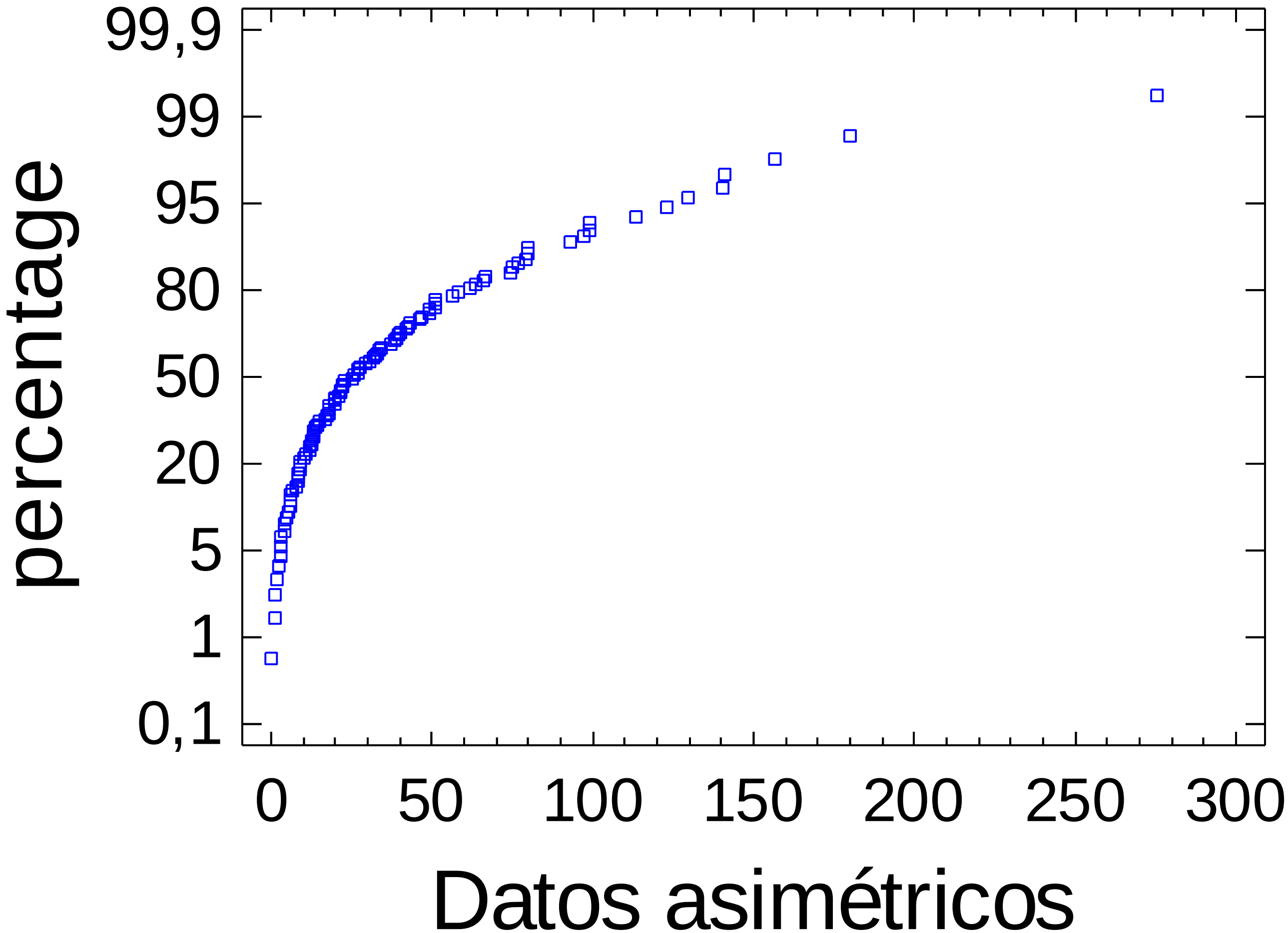
$\Rightarrow$ Un **87,56%** de los asientos cumple las especificaciones

Papel probabilístico normal

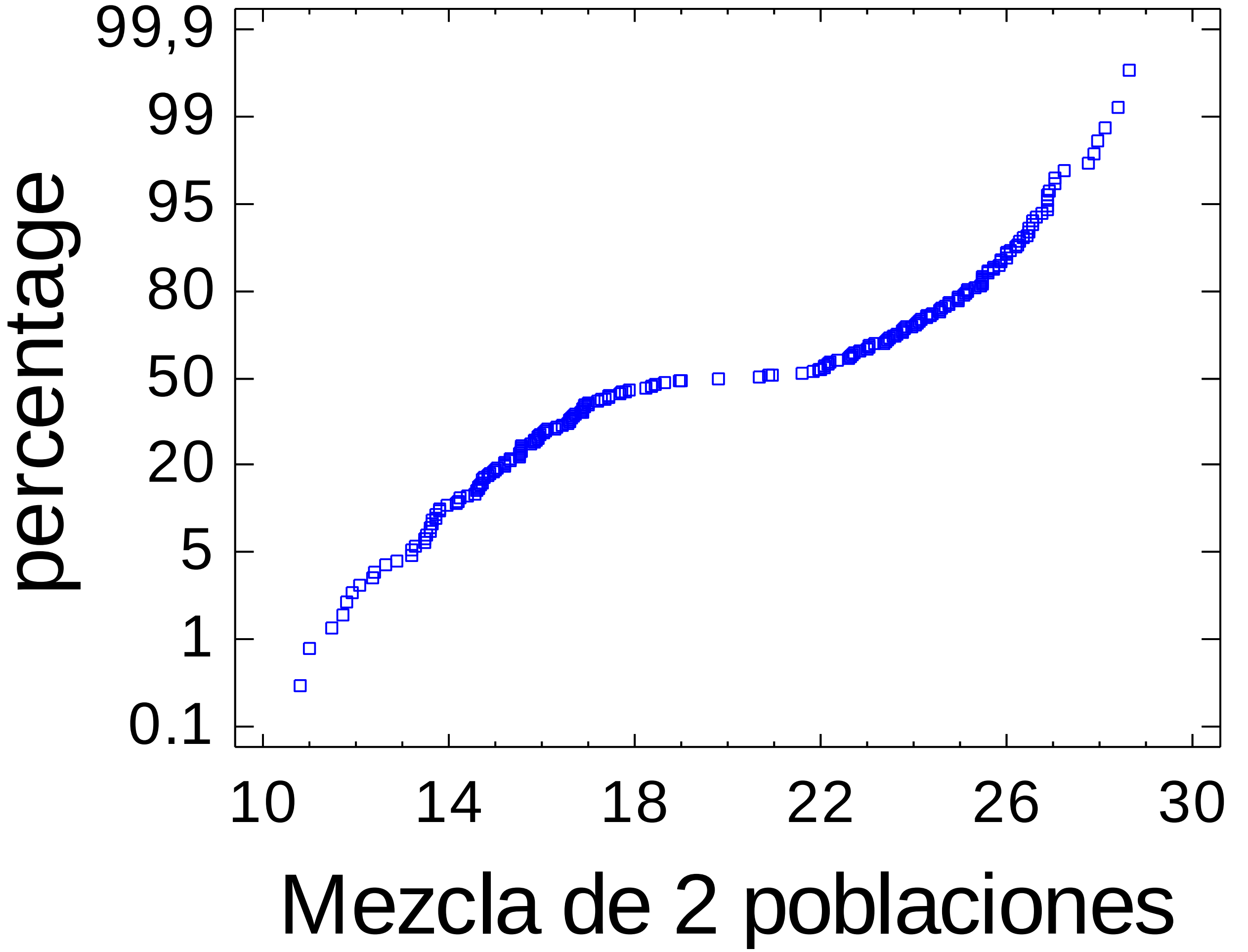
- Abcisa del punto: el valor ordenado

- Ordenada: porcentaje de valores en la muestra que son menores o iguales que el considerado. Escala modificada para la distribución normal.

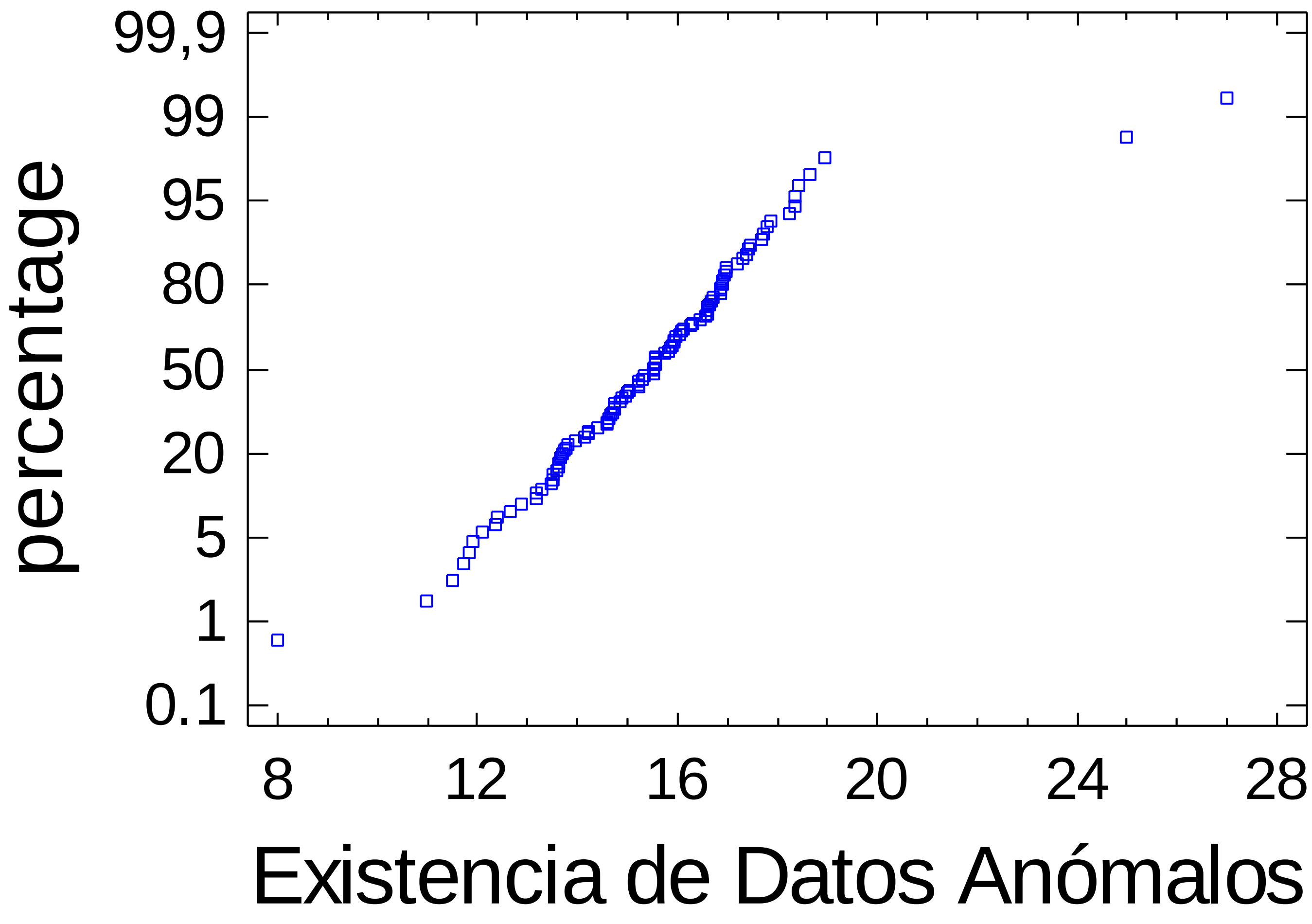
Datos con asimetría positiva:



Mezcla de poblaciones:



Los valores anormalmente altos o bajos



Papel probabilístico normal

Ejercicio 17:

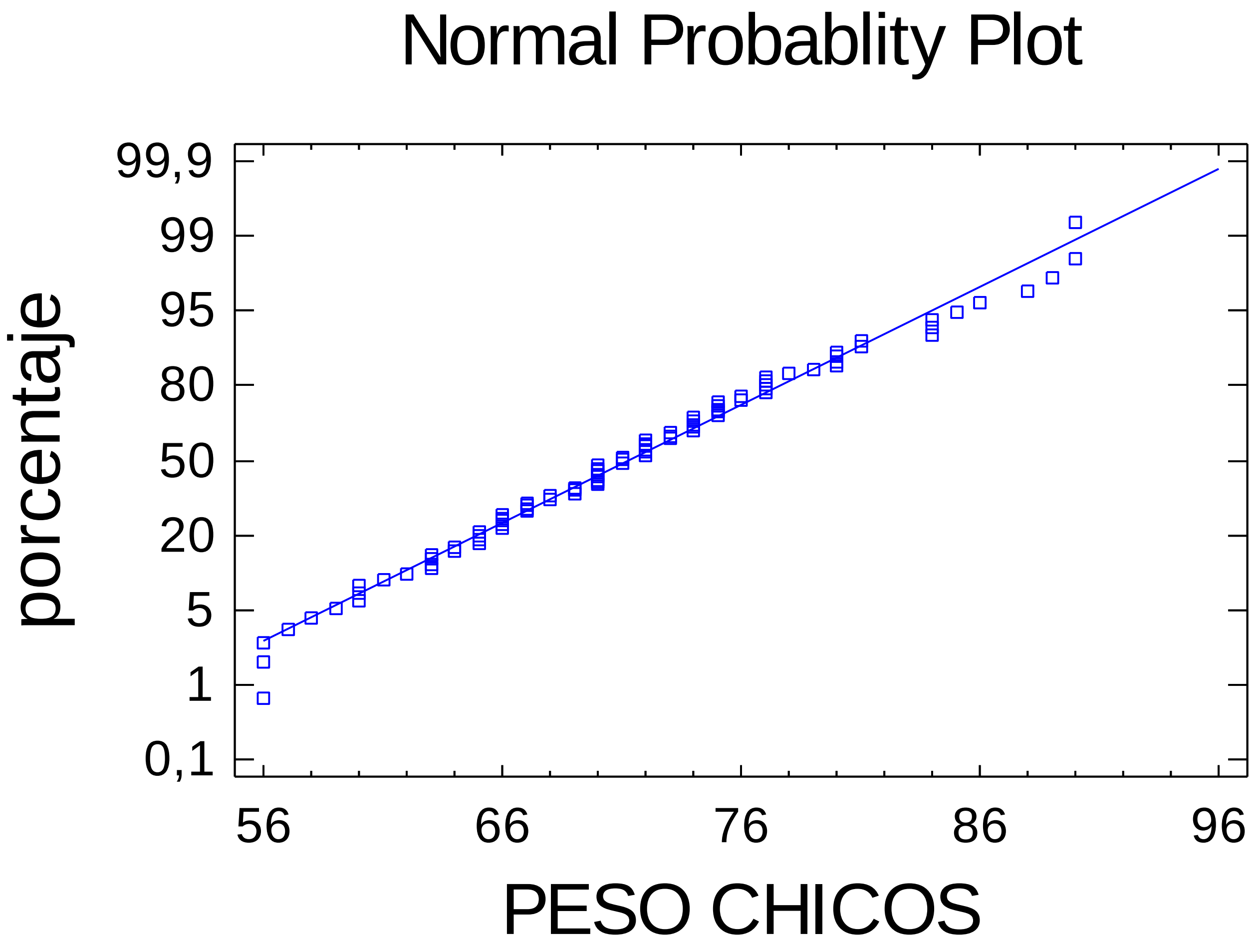
Obtener una representación en papel probabilístico normal de los datos de PESO en los chicos de la encuesta.

¿Se distribuye el peso de los chicos de forma aproximadamente normal?

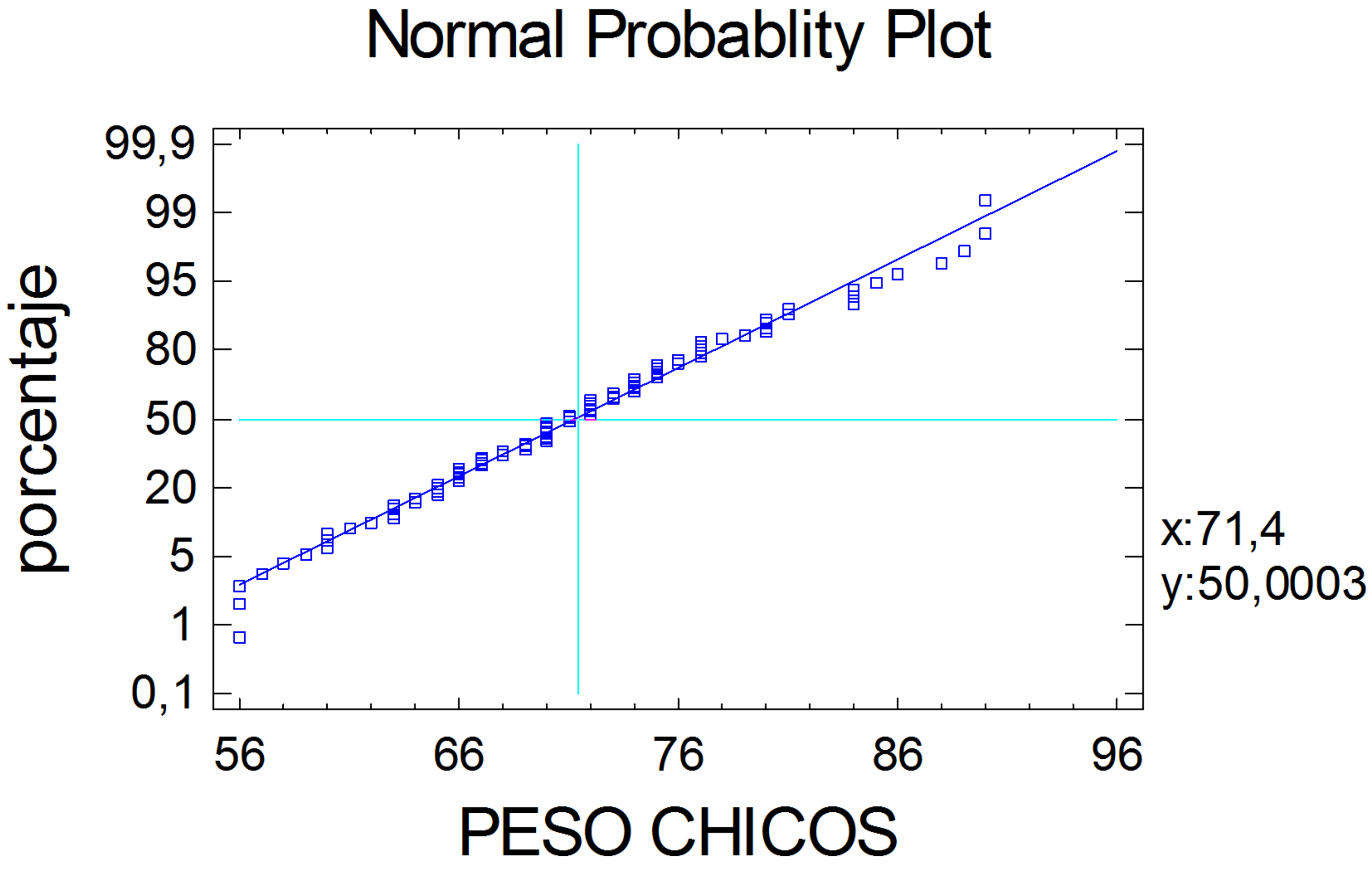
¿Cómo podría estimarse aproximadamente el peso medio de los chicos a partir de la representación anterior?

¿Cómo podría estimarse aproximadamente la desviación típica de la distribución a partir de la representación considerada?

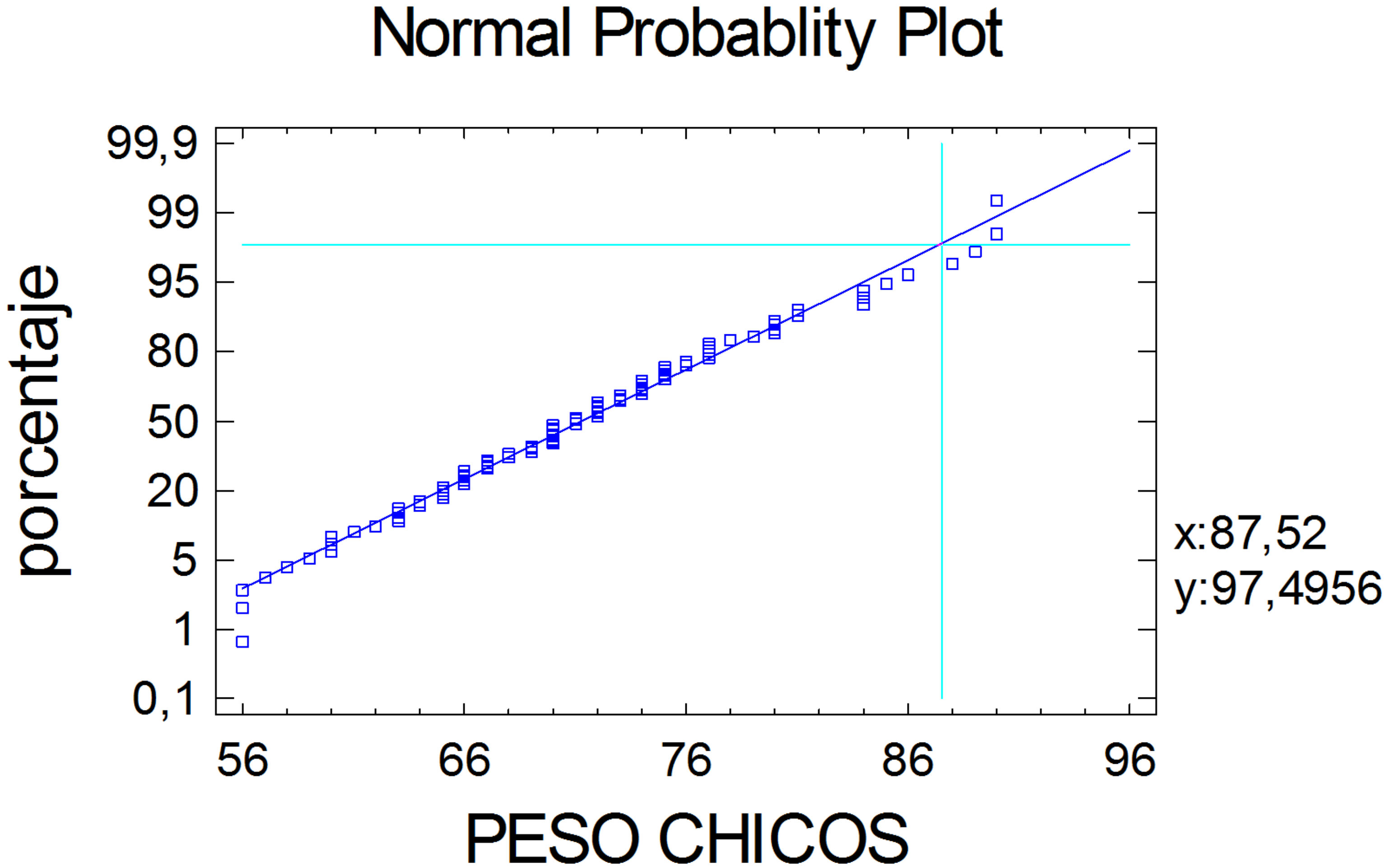
Papel probabilístico normal



Estimación aproximada de la media entrando con porcentaje 50% $\Rightarrow m \simeq$ **71,4**



Estimación aproximada de la desviación típica entrando con porcentaje 97,5%



$$\sigma \simeq (87{,}52-71{,}4)/2=\mathbf{8{,}06}$$

Aproximaciones normales

- **Teorema Central de Límite:**

$X_i \sim g_i(m_{xi}, \sigma_{xi})$ $i=1\ldots N$

independientes

Siendo g cualquier distribución (Binomial, Poisson, etc.). $N \to \infty$ (N muy grande)

$$Y=X_1+\ldots+X_N \Rightarrow Y \sim N(m_{x1}+\ldots+m_{xN}, \sigma^2_{X1}+\ldots+\sigma^2_{XN})$$

Aproximaciones normales

Ejercicio 18: *El tiempo de acceso o búsqueda de un fichero en una unidad de disco fluctúa* ***uniformemente*** *entre 0,1 y 0,5 segundos. Calcula la probabilidad de que el tiempo de acceso a 100 ficheros consecutivos supere los 31 segundos.*

SOLUCIÓN: t_i =tiempo de acceso a un fichero

$T_i \sim U(0{,}1,\ 0{,}5) \Rightarrow m_{ti}=0{,}3 \quad \sigma_{ti}^2=(0{,}5-0{,}1)^2/12=0{,}013$

Aproximaciones normales

SOLUCIÓN:

T=tiempo de acceso a 100 ficheros $\Rightarrow T=\sum_{i=1}^{100} t_i \Rightarrow$

T por el teorema central del límite se aproxima a una distribución normal con m=100x0,3=30 y σ^2*=100x0,013=1,3*

P(T>31)=P(N(30, 1,14)>31)=P(N(0,1)>1/1,14)=

=P(N(0,1)>0,88)=0,1894 $\Rightarrow$ *18,94%*

Aproximaciones normales

Distribución Binomial

- Si **X** es Binomial (n,p)

  y su varianza $np(1-p) \geq 9 \Rightarrow$

$$Y = \frac{X - np}{\sqrt{np(1-p)}} \approx N(0,1)$$

Aproximaciones normales

Ejercicio 19: *Cuando se prueban las tarjetas de circuito que se usan en la fabricación de reproductores de discos compactos, el porcentaje de defectos a largo plazo es 5%. Suponiendo que se recibe un lote de 250 tarjetas, ¿cuál es la probabilidad aproximada de que al menos el 10% de las tarjetas en el lote sean defectuosas?*

Aproximaciones normales

SOLUCIÓN:

X=nº tarjetas defectuosas en un lote de 250 $\Rightarrow$

X $\sim$ *B(n=250, p=0,05)* $\Rightarrow$ *np(1-p)=11,875>9* $\Rightarrow$

X *se puede aproximar con el modelo Normal de m=np=12,5 y* $\sigma = \sqrt{11{,}875} = 3{,}45$

El 10% de 250 tarjetas es 25.

*P(**X**≥25)≈P(N(12,5, 3,45) >24,5)=P(N(0,1)>*$\frac{24{,}5-12{,}5}{3{,}45}) =$

=P(N(0,1) > 3,48)=0,00025

Aproximaciones normales

Distribución de Poisson

- Si **X** es Poisson (λ) y su parámetro λ es ≥9 $\Rightarrow$

$$\mathrm{Y} = \frac{\mathrm{X} - \lambda}{\sqrt{\lambda}} \approx \mathrm{N}(0,1)$$

Aproximaciones normales

Ejercicio 20: *Los códigos CRC utilizados en el envío de paquetes a través de la red son capaces de corregir como máximo 1000 errores en cada bloque de 200 paquetes. Se sabe que el número medio de errores en cada bloque enviado es de 400. Si tomamos al azar un bloque de 200 paquetes, ¿cuál es la probabilidad de que el número total de errores sea inferior a 380?*

Aproximaciones normales

SOLUCIÓN:

X=nº errores en un bloque de 200 paquetes $\Rightarrow$

X~Poisson(λ=400) $\Rightarrow$ λ>9 $\Rightarrow$ ***X***~N(400, $\sqrt{400}$)

P(***X***<380)≈P(N(400, 20)<379,5)=

=P(N(0,1)< $\frac{379,5-400}{20}$) $=$ P(N(0,1)<-1,02)=0,1539

$\Rightarrow$ 15,39%