ESTADÍSTICA

www.etsii.upm.es/ingor/estadistica/

Descriptiva

Probabilidad

Inferencia

Escuela Técnica Superior de Ingenieros Industriales
Universidad Politécnica de Madrid

Edición Curso 22/23

Estadística

Parte I. PEC1

1. Estadística Descriptiva
2. Fundamentos de Probabilidad
3. Variable Aleatoria

Parte II. PEC2

4. Modelos de Probabilidad
5. Estimación
6. Contraste de Hipótesis

Laboratorio de Estadística
Escuela Técnica Superior de Ingenieros Industriales
Universidad Politécnica de Madrid
http://www.etsii.upm.es/ingor/estadistica/

Profesores

Jesús Juan Ruiz.
María Jesús Sánchez Naranjo.
José Manuel Mira McWilliams.
Camino González Fernández.
Carolina García Martos.
Francisco Javier Cara Cañas.
Eduardo Caro Huertas.

Estadística

Parte I.

Tema 1. Estadística Descriptiva

Datos

Número	Consumo	Cilindrada	Potencia	Peso	Aceleración	Año	País	Nº Cilindros
	l/100Km	*cc*	*CV*	*kg*	*segundos*			
1	15	4982	150	1144	12	70	EEUU	8
2	16	6391	190	1283	9	70	EEUU	8
3	24	5031	200	1458	15	70	EEUU	8
4	9	1491	70	651	21	71	EEUU	4
5	11	2294	72	802	19	71	EEUU	4
6	17	5752	153	1384	14	71	EEUU	8
7	12	2294	90	802	20	72	EEUU	4
8	17	6555	175	1461	12	72	EEUU	8
9	18	6555	190	1474	13	72	EEUU	8
10	12	1147	97	776	14	72	Japón	3
11	16	5735	145	1360	13	73	EEUU	8
12	12	1868	91	860	14	73	Europa	4
13	9	2294	75	847	17	74	EEUU	4
14	8	1295	67	666	16	74	Europa	4
15	7	1163	65	612	21	74	Japón	4
16	7	1360	61	667	19	74	Japón	4
17	12	3802	90	1070	17	75	EEUU	6
18	13	3687	95	1261	19	75	EEUU	6
19	9	1475	71	741	17	75	Europa	4
20	9	1983	115	890	14	75	Europa	4
...	...	...	...	...	...	...	...	...
391	7	1753	75	735	15	82	Japón	4

Tipos de datos

- Cuantitativos
  - Continuos: *consumo, potencia,aceleración, peso*
  - Discretos: *nº de cilindros*
- Cualitativos
  - Ordinales: categoría
  - No ordinales: país, gasolina/gasoil

Lectura de datos

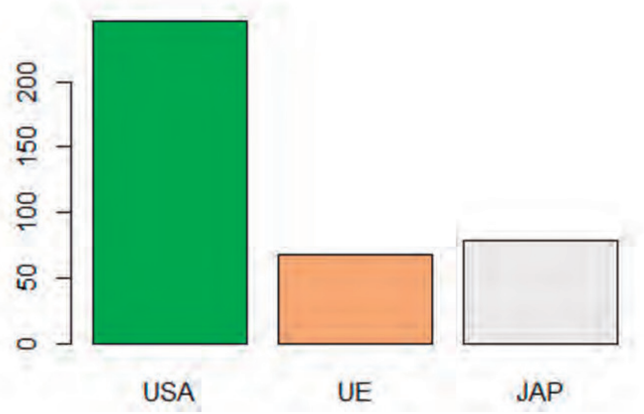
Para leer datos de un achivo de texto organizado en columnas se utiliza la instrucción `read.table()`.

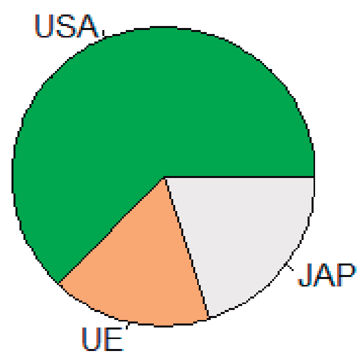
```
coches = read.table('coches2.txt',header=TRUE)
head(coches)
```

```
##   Consumo   CC  CV Peso Acel Ano Origen Cilindros
## 1      11 3277  85  862   16  70      1         6
## 2      11 3261  90  882   15  70      1         6
## 3      13 3261  97  924   16  70      1         6
## 4      11 3245  95  944   16  70      1         6
## 5      17 7456 225 1028   10  70      1         8
## 6      15 4982 150 1144   12  70      1         8
```

Descriptiva de variables cualitativas

```
coches$origen2 = factor(coches$Origen,labels=c('USA','UE','JAP'))
table(coches$origen2)
##
## USA  UE JAP
## 246  68  79
barplot(table(coches$origen2),col= terrain.colors(3))
pie(table(coches$origen2),col=terrain.colors(3))
```

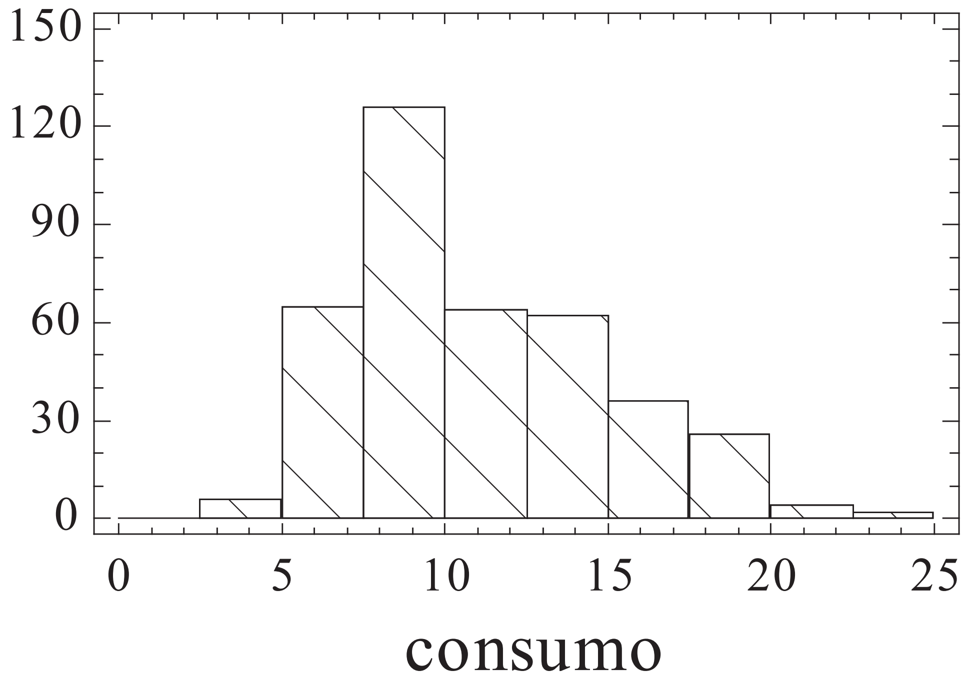




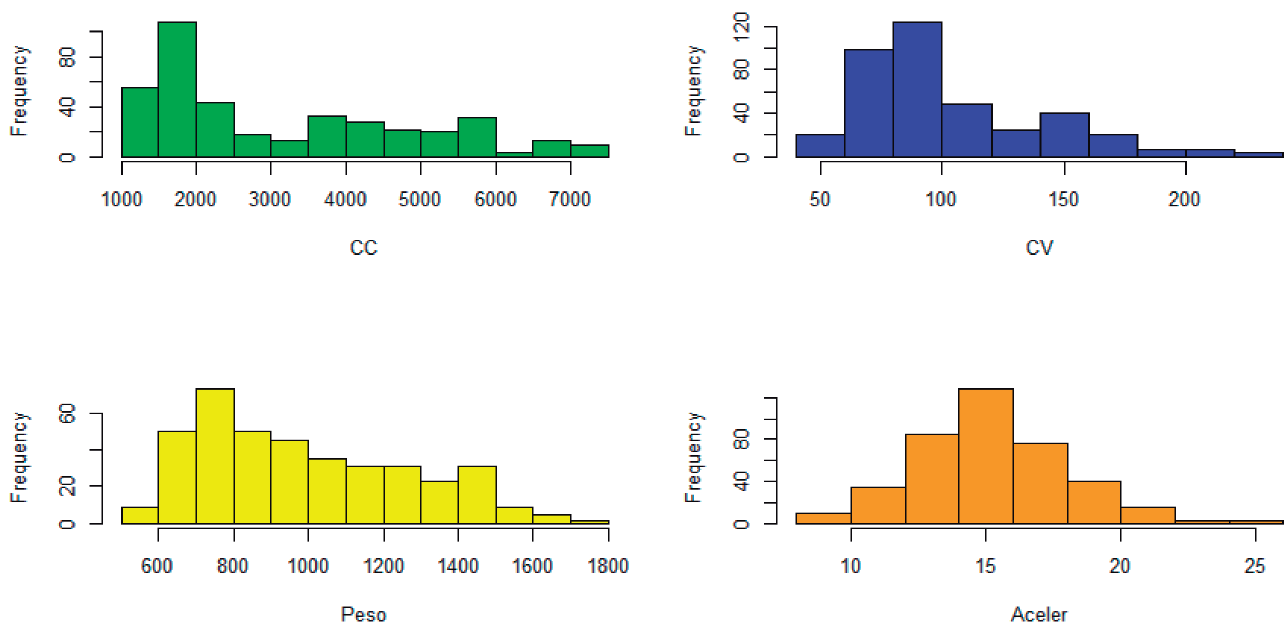
Distribución de frecuencias: consumo *l/100 km*

Clase	Limite Inferior	Limite Superior	Punto Medio	Frecuencia Absoluta
1	0,0	2,5	1,25	0
2	2,5	5,0	3,75	6
3	5,0	7,5	6,25	65
4	7,5	10,0	8,75	126
5	10,0	12,5	11,25	64
6	12,5	15,0	13,75	62
7	15,0	17,5	16,25	36
8	17,5	20,0	18,75	26
9	20,0	22,5	21,25	4
10	22,5	25,0	23,75	2
Total				391

Histograma

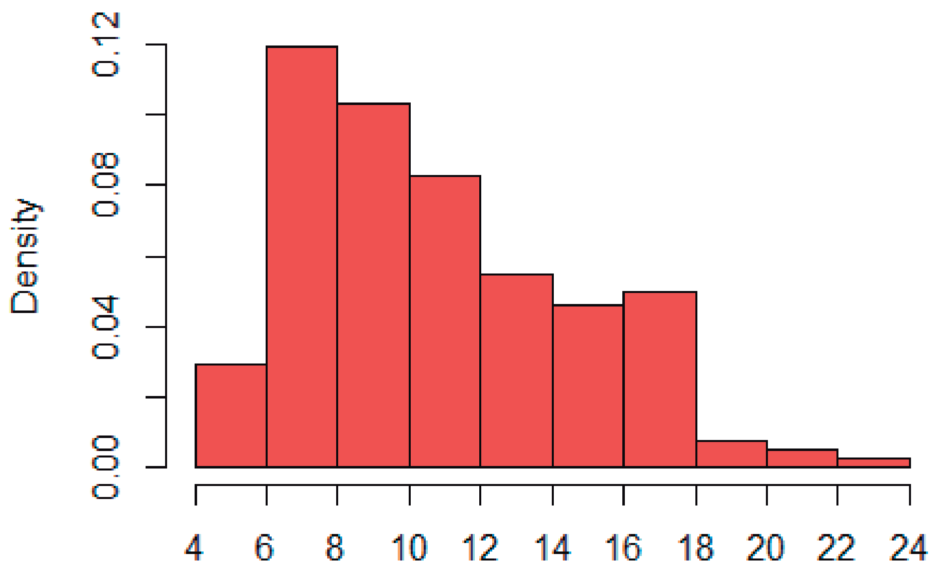


Histogramas para coches



Descriptiva de variables continuas

```
hist( coches$Consumo, color = 'tomato', probability =TRUE)
```



Medidas de centro

$$x_1, x_2, \ldots, x_n$$

Media aritmética

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Media geométrica

(si $x_i > 0$ para todo i)

$$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n}$$

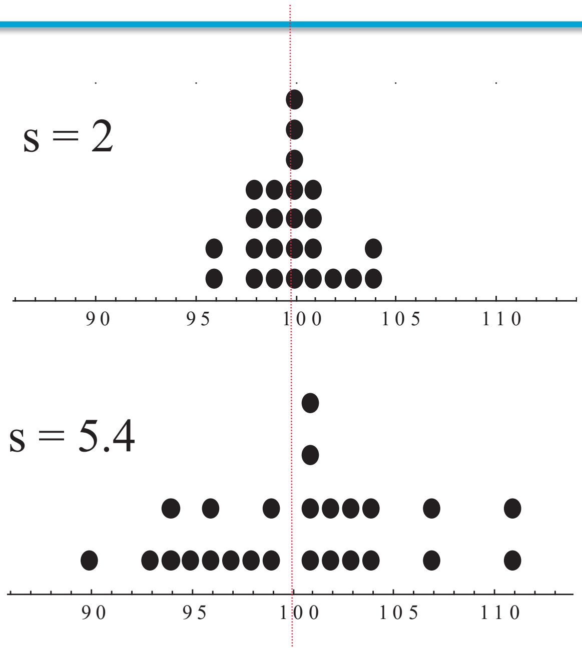
Media armónica

(si $x_i > 0$ para todo i)

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

$$\bar{x}_H \le \bar{x}_G \le \bar{x}$$

Medidas de dispersión



Media 100

$$x_1, x_2, \ldots, x_n$$

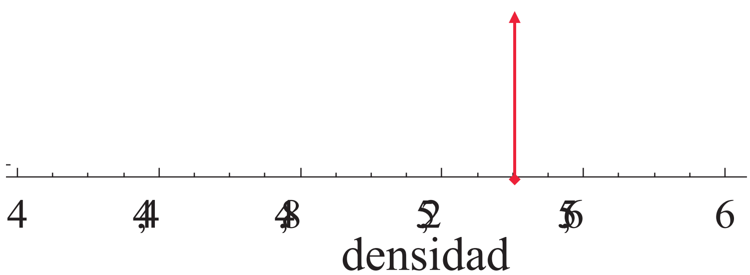
Desviación Típica

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

Varianza : s^2

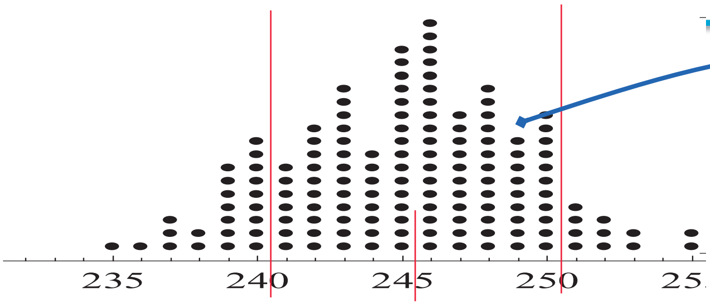
Densidad de la tierra (Cavendish, 1798)

5,5	5,47	5,55	5,75	5,29	5,27
5,57	4,88	5,34	5,29	5,34	5,85
5,42	5,62	5,3	5,1	5,26	5,65
5,61	5,63	5,36	5,86	5,44	5,39
5,53	4,07	5,79	5,58	5,46	



Media = 5.42 *Desv. Típ. = 0.338*

Desigualdad de Chebychev



$$\bar{x} - k \times s \qquad \bar{x} \qquad \bar{x} + k \times s$$

$$fr(|\, x_i - \bar{x} \,| \le ks) > 1 - \frac{1}{k^2}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{\sum_{|x_i - \bar{x}| \le ks} (x_i - \bar{x})^2}{n} + \frac{\sum_{|x_i - \bar{x}| > ks} (x_i - \bar{x})^2}{n}$$

$$s^2 \ge \frac{\sum_{|x_i - \bar{x}| > ks} (x_i - \bar{x})^2}{n} > \frac{\sum_{|x_i - \bar{x}| > ks} k^2 s^2}{n} = fr(|\, x_i - \bar{x} \,| > ks) k^2 s^2$$

$$fr(|\, x_i - \bar{x} \,| > ks) \le \frac{1}{k^2} \qquad \Leftrightarrow \qquad fr(|\, x_i - \bar{x} \,| \le ks) > 1 - \frac{1}{k^2}$$

Mediana y Cuartiles

$x_1, x_2, \ldots, x_n$

Datos ordenados

$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$

$$\text{Mediana} \begin{cases} x_{(p)} & p = \dfrac{n+1}{2} : n \quad \text{impar} \\[2ex] \dfrac{x_{(p)} + x_{(p+1)}}{2} & p = \dfrac{n}{2} : \quad n \text{ par} \end{cases}$$

Cuartiles

$$Q_1 = x_{(r)} \qquad Q_3 = x_{(s)}$$

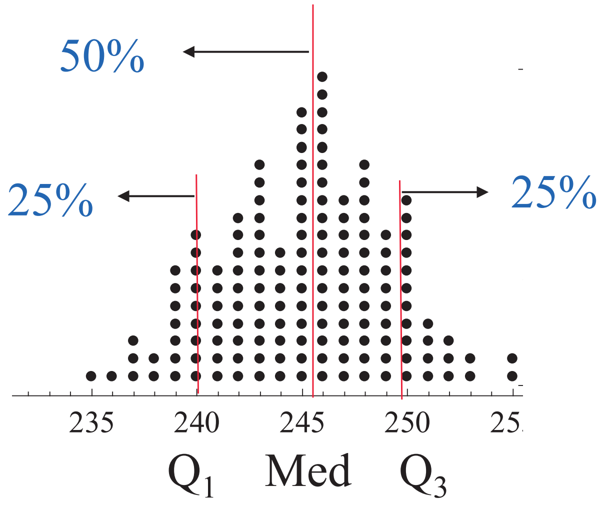
$$r = \left\lfloor \frac{p+1}{2} \right\rfloor \qquad s = n - r + 1$$

Mediana y Cuartiles

$x_1, x_2, \ldots, x_n$

Mediana : (Med)

$$fr(x_i \leq Med) = 0.50$$



Cuartiles

Q_1

$$fr(x_i \leq Q_1) = 0.25$$

Q_3

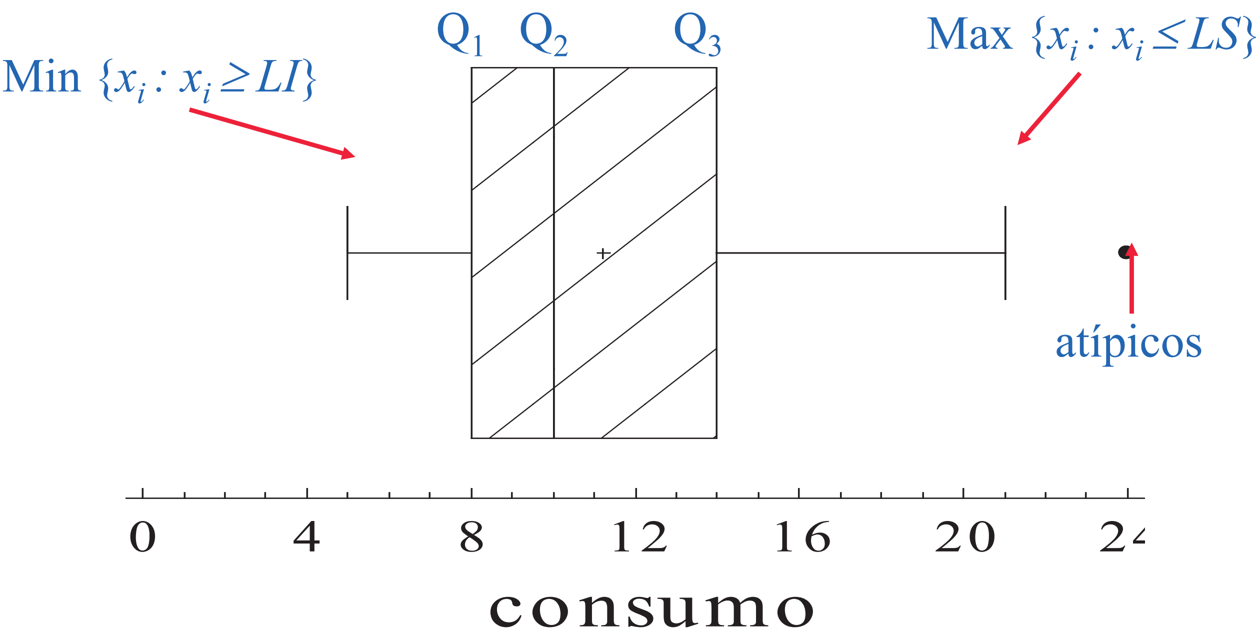
$$fr(x_i \leq Q_3) = 0.75$$

Medidas características

	Consumo	Cilindrada	Potencia	Peso	Aceleración
Media	11.2	3181.2	104.2	990.7	15.7
Desv. Típica	3.9	1714.6	38.3	281.9	2.8
Primer Cuartil	8	1721	75	741.5	14
Mediana	10	2474	93	933	16
Tercer Cuartil	13.5	4334	125	1203.5	17
Rango Intercuartílico	5.5	2613	50	462	3

```
summary(coches)

##     Consumo           CC              CV              Peso              Acel
##  Min.   : 5.00   Min.   :1114   Min.   : 46.0   Min.   : 537.0   Min.   : 8.00
##  1st Qu.: 8.00   1st Qu.:1721   1st Qu.: 75.0   1st Qu.: 742.0   1st Qu.:14.00
##  Median :10.00   Median :2474   Median : 93.0   Median : 935.0   Median :16.00
##  Mean   :11.22   Mean   :3179   Mean   :104.2   Mean   : 990.6   Mean   :15.67
##  3rd Qu.:13.00   3rd Qu.:4293   3rd Qu.:125.0   3rd Qu.:1203.0   3rd Qu.:17.00
##  Max.   :24.00   Max.   :7456   Max.   :230.0   Max.   :1713.0   Max.   :25.00
##                                 NA's   :2
```

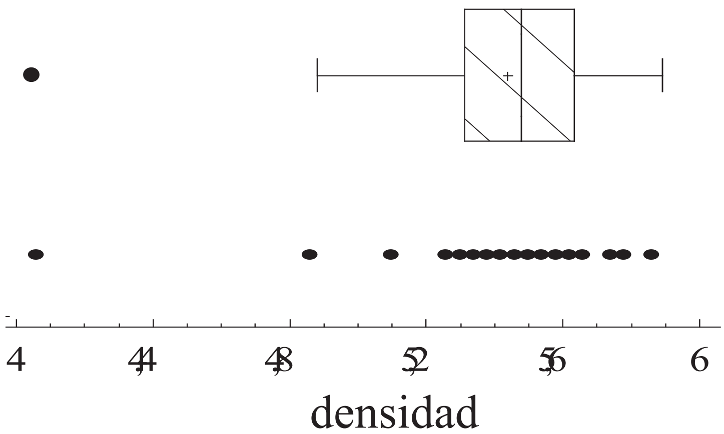
Diagrama de caja



$LI = Q_1 - 1.5\ RI$ $\quad RI = Q_3 - Q_1$ $\quad LS = Q_3 + 1.5\ RI$

Densidad de la tierra (Cavendish, 1798)

5,5	5,47	5,55	5,75	5,29	5,27
5,57	4,88	5,34	5,29	5,34	5,85
5,42	5,62	5,3	5,1	5,26	5,65
5,61	5,63	5,36	5,86	5,44	5,39
5,53	4,07	5,79	5,58	5,46	



Media = 5.42 *Desv. Típ. = 0.338*

Diagrama de caja múltiple

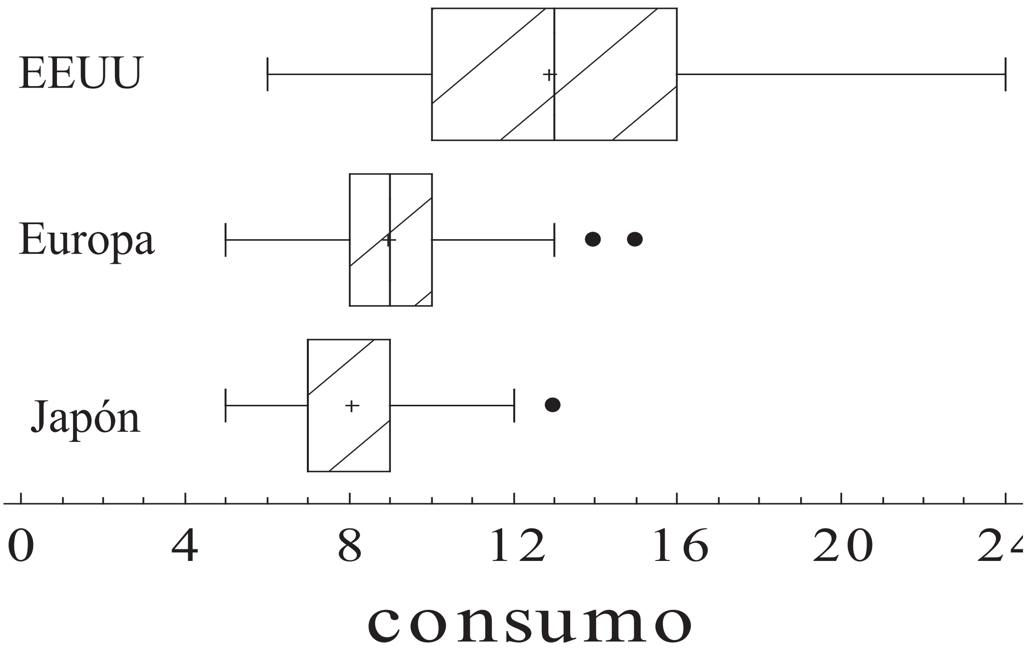
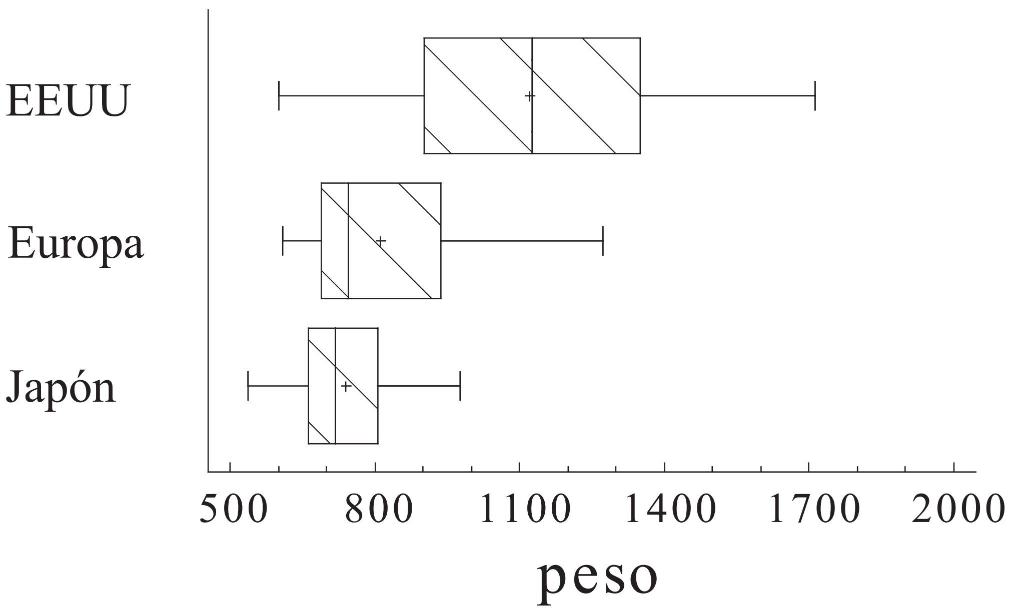
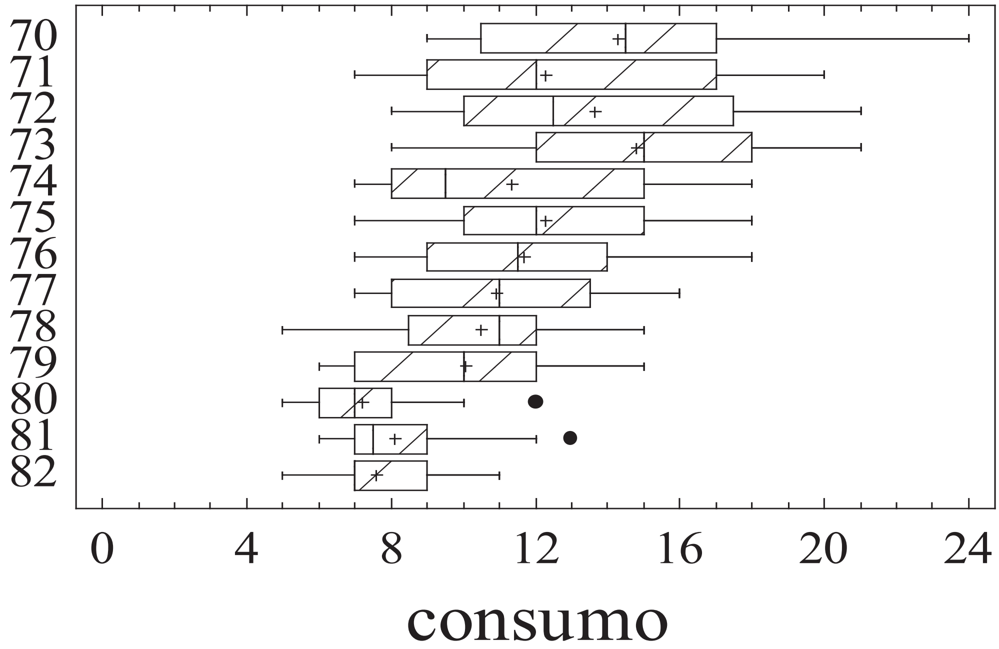


Diagrama de caja múltiple



Consumo según año de fabricación



Box plot en R

```
coches$origen2 = factor(coches$Origen,labels=c('USA','UE','JAP'))
boxplot(coches$Consumo~coches$origen2,horizontal=TRUE, col=terrain.colors(3))
```

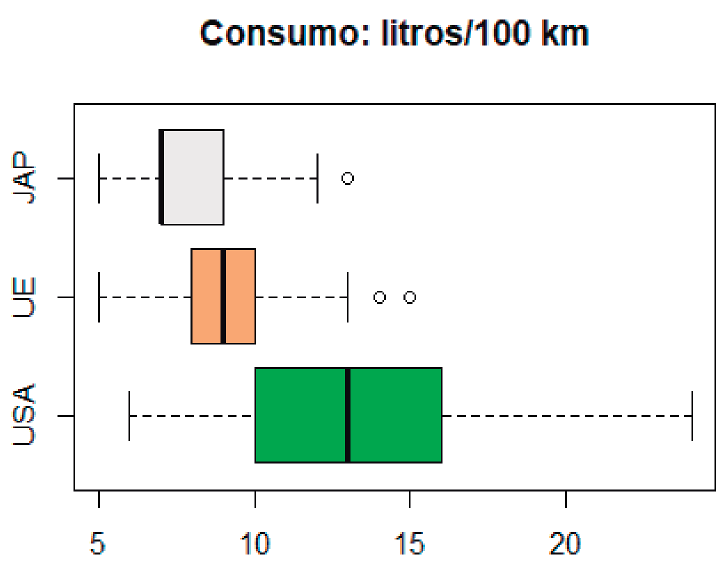


Diagrama de Caja Múltiple

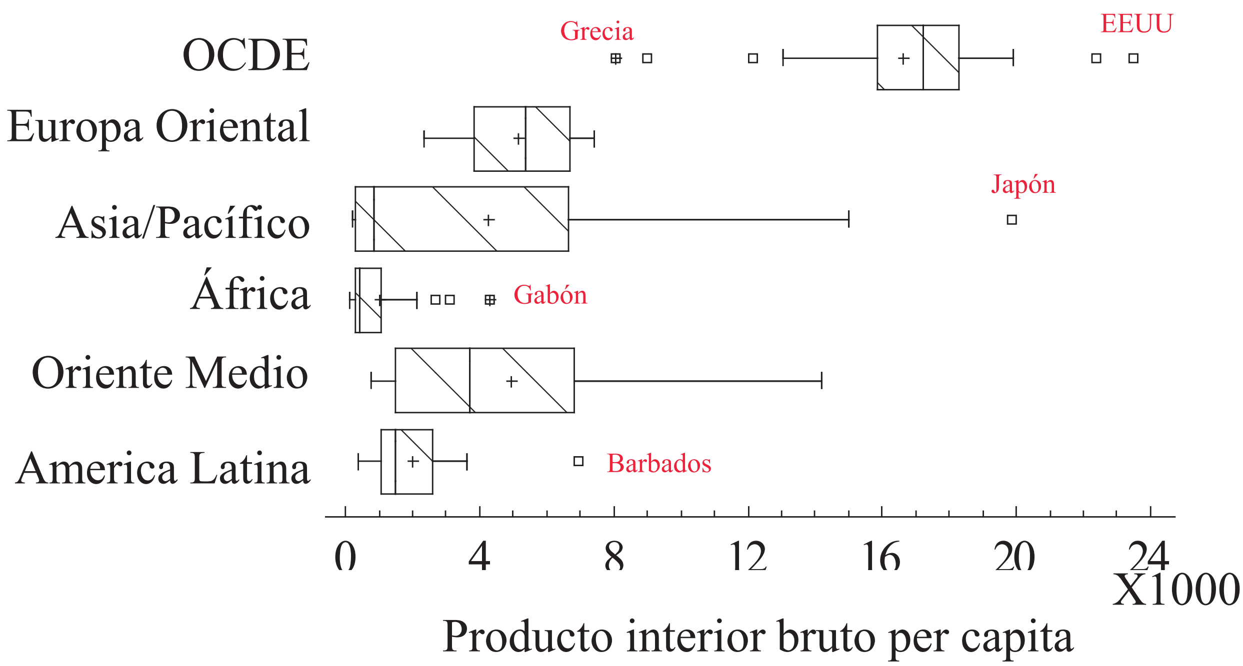


Diagrama de tallos y hojas

```
        LO|4,07

 1        4|
 1        4|
 1        4|
 1        4|
 2        4|8
 3        5|1
12        5|222233333
(9)       5|444455555
 8        5|666677
 2        5|88
```

- Media 5,419
- Des. Típica 0,339
- Mínimo 4,07
- Máximo 5,86
- Cuartil 1 5.3
- Mediana 5.46
- Cuartil 3 5.61

Diagrama de tallos y hojas con R

```
> stem(coches$Peso)

  The decimal point is 2 digit(s) to the right of the |

   5 | 45999
   6 | 0000111112223445555555556666666666666677777788888899999
   7 | 000001111111111111222222222223333334444444444455555666666677777888999
   8 | 00000000111222333444455555666666667777788888889999999
   9 | 0000111111222333344444556666667777777888888888899999
  10 | 000111223333455555566677777888899999
  11 | 00112222333444444444455555677888999
  12 | 0001112222223444555666677888899
  13 | 000011223555566677788889
  14 | 00111122333445555566667889999
  15 | 000445556788
  16 | 45557
  17 | 1
```


Medidas características de forma: asimetría y curtosis

Coeficiente de asimetría

$$C_{AS} = \frac{m_3}{s^3} = \frac{\sum (x_i - \bar{x})^3}{ns^3}$$

Coeficiente de curtosis o apuntamiento

$$C_{AP} = \frac{m_4}{s^4} = \frac{\sum (x_i - \bar{x})^4}{ns^4}$$

Momento respecto al origen

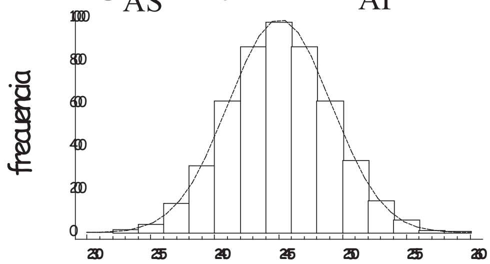
$$a_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$

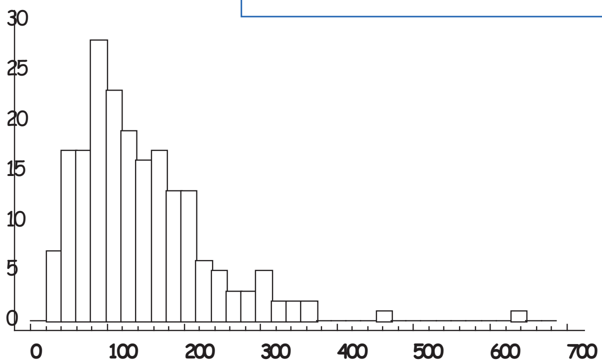
Momentos respecto a la media

$$m_k = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^k}{n}$$

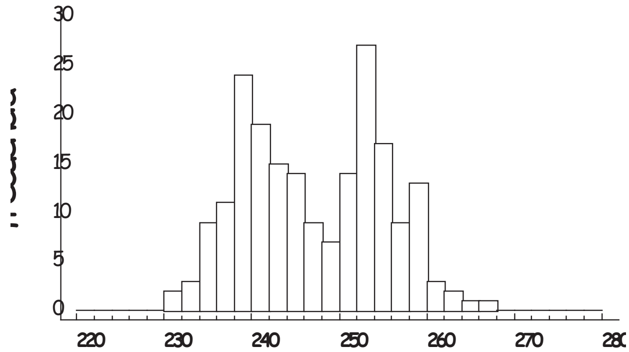
Modelo ideal

$C_{AS} = 0 \qquad C_{AP} = 3$





$C_{AS} > 0$



$C_{AP} < 3$

Transformaciones de datos

- Transformaciones Lineales

$$y_i = a + bx_i$$
$$\bar{y} = a + b\bar{x}$$
$$s_y = |b| s_x$$

La "*forma*" de la distribución no cambia

(Asimetría y curtosis no cambia)

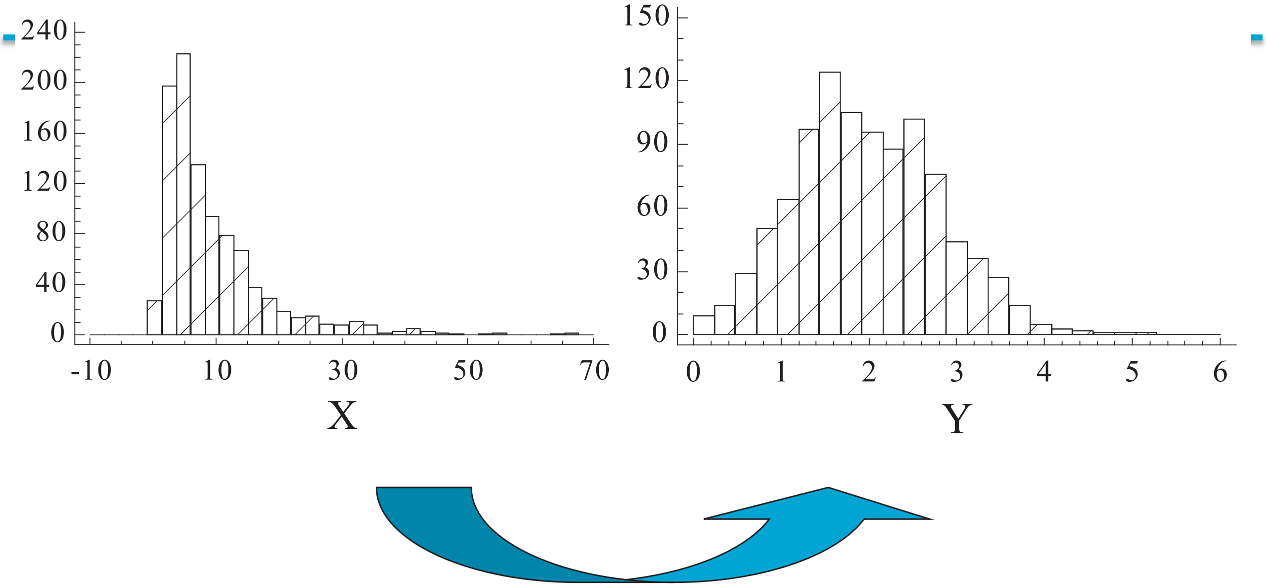
- Transformaciones no-lineales

$$y_i = h(x_i)$$
$$\bar{y} \neq h(\bar{x})$$

Cambia la "*forma*" de la distribución

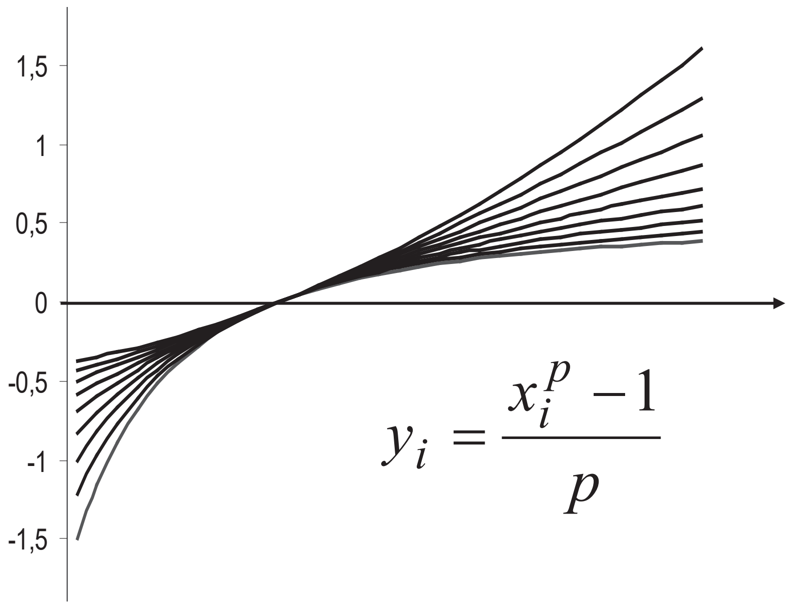
(coeficientes de asimetría y curtosis cambian)

Efecto de la transformación de datos



$$y_i = \log x_i$$

Transformaciones Box-Cox



$$y_i = \frac{x_i^p - 1}{p}$$

$$p = 0 \Rightarrow y_i = \log x_i$$

Datos Multivariantes

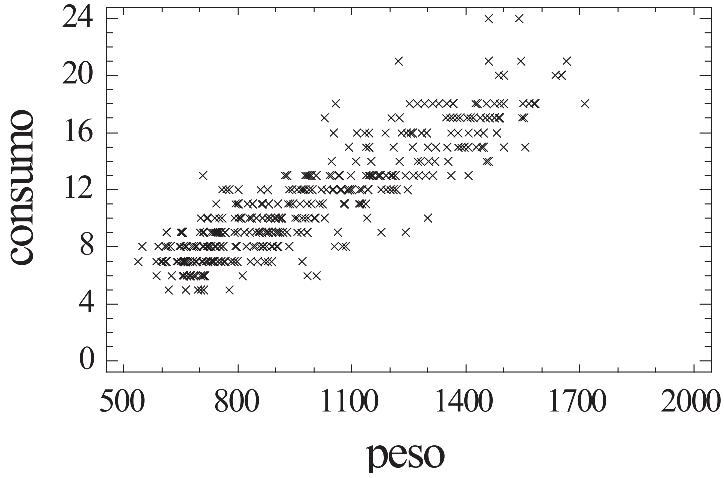
Variables

$$
\textbf{Observaciones}\quad
\begin{array}{c|cccccc}
 & Y_1 & Y_2 & \cdots & Y_k & & \\
\hline
1 & x_{11} & x_{21} & \cdots & x_{k1} & \longrightarrow & \mathbf{x_1} \\
2 & x_{12} & x_{22} & \cdots & x_{k2} & \longrightarrow & \mathbf{x_2} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \longrightarrow & \vdots \\
n & x_{1n} & x_{2k} & \cdots & x_{kn} & \longrightarrow & \mathbf{x}_{\mathrm{n}}
\end{array}
$$

Vector de Medias

$$
\mathbf{x_i} = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ki} \end{pmatrix}; \qquad \bar{\mathbf{x}} = \frac{\sum_{i=1}^{n} \mathbf{x_i}}{n} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{pmatrix}
$$

Covarianza



Coche	*Peso*	*Consumo*
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Matriz de Varianzas

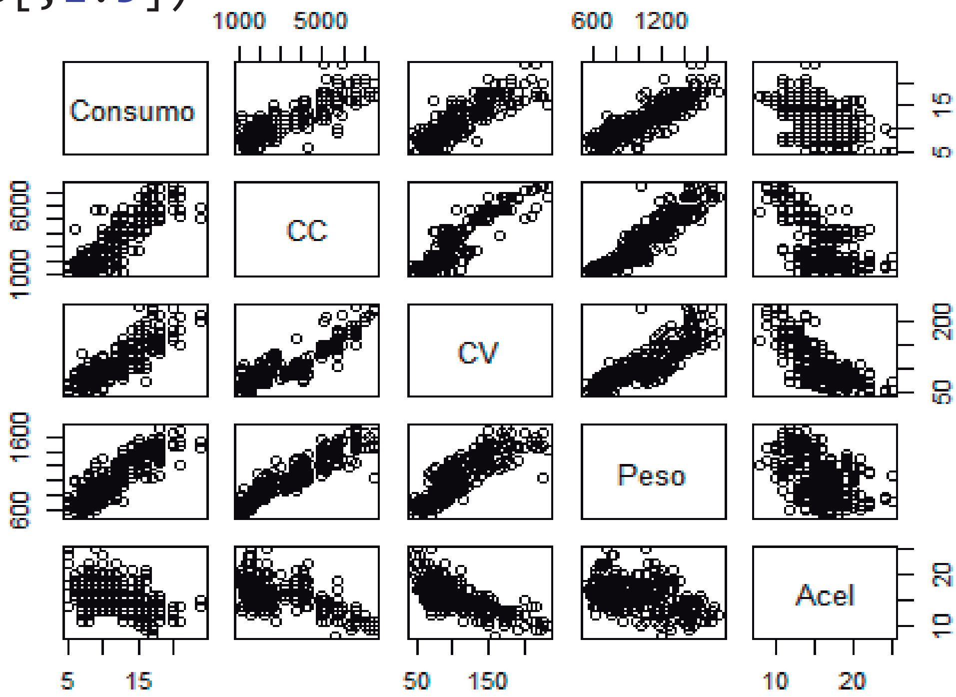
$$\mathbf{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_{1i}-\bar{x}_1 \\ x_{2i}-\bar{x}_2 \\ \vdots \\ x_{ki}-\bar{x}_k \end{pmatrix}\begin{pmatrix} x_{1i}-\bar{x}_1 & x_{2i}-\bar{x}_2 & \cdots & x_{ki}-\bar{x}_k \end{pmatrix}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} (x_{1i}-\bar{x}_1)^2 & (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2) & \cdots & (x_{1i}-\bar{x}_1)(x_{ki}-\bar{x}_k) \\ (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2) & (x_{2i}-\bar{x}_2)^2 & \cdots & (x_{2i}-\bar{x}_2)(x_{ki}-\bar{x}_k) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{1i}-\bar{x}_1)(x_{ki}-\bar{x}_k) & (x_{2i}-\bar{x}_2)(x_{ki}-\bar{x}_k) & \cdots & (x_{ki}-\bar{x}_k)^2 \end{pmatrix}$$

$$= \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1k} \\ s_{12} & s_2^2 & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1k} & s_{2k} & \cdots & s_k^2 \end{pmatrix}$$

Gráficos de dispersión: ejemplo coches

```
pairs(coches[,1:5])
```



Matriz de Varianzas

```
var(coches[,1:5],use="complete.obs")
```

##	Consumo	CC	CV	Peso	Acel
## Consumo	15.1557	5824.4	127.304	971.53	-5.0073
## CC	5824.4404	2939742.5	58965.409	451460.55	-2597.4073
## CV	127.3035	58965.4	1465.220	9312.79	-73.5510
## Peso	971.5264	451460.5	9312.785	79495.52	-328.0469
## Acel	-5.0073	-2597.4	-73.551	-328.05	7.6156

Propiedades de S^2

$$\tilde{\mathbf{X}} = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_2 & \cdots & x_{k1}-\bar{x}_k \\ x_{12}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{k2}-\bar{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}-\bar{x}_1 & x_{2n}-\bar{x}_2 & \cdots & x_{kn}-\bar{x}_k \end{pmatrix}$$

$$\mathbf{S}^2 = \frac{1}{n}\tilde{\mathbf{X}}^{\mathbf{T}}\tilde{\mathbf{X}} \longrightarrow \left\{ \begin{array}{l} \text{Cuadrada } k \text{ x } k \\ \text{Simétrica} \\ \text{Semidef. positiva} \end{array} \right.$$

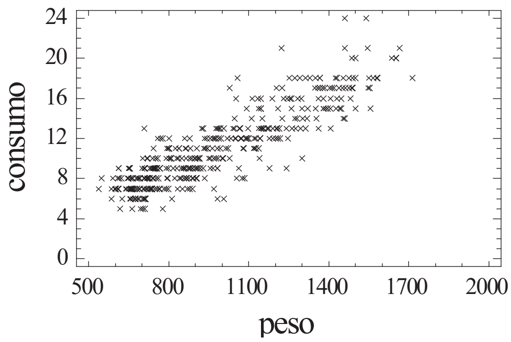
$\mathbf{S}^2$ es *semidefinida positiva* :

$$\forall \mathbf{w} \in \Re^k, \quad \mathbf{w}^{\mathbf{T}}\mathbf{S}^2\mathbf{w} \geq 0$$

$$\mathbf{w}^{\mathbf{T}}\mathbf{S}^2\mathbf{w} = \mathbf{w}^{\mathbf{T}}(\frac{1}{n}\tilde{\mathbf{X}}^{\mathbf{T}}\tilde{\mathbf{X}})\mathbf{w} = \frac{1}{n}(\tilde{\mathbf{X}}\mathbf{w})^{\mathbf{T}}(\tilde{\mathbf{X}}\mathbf{w})$$

$$\mathbf{v} = \tilde{\mathbf{X}}\mathbf{w}, \quad \Rightarrow \mathbf{w}^{\mathbf{T}}\mathbf{S}^2\mathbf{w} = \frac{1}{n}\mathbf{v}^{\mathbf{T}}\mathbf{v} = \frac{\sum_{i=1}^n v_i^2}{n} \geq 0$$

Correlación

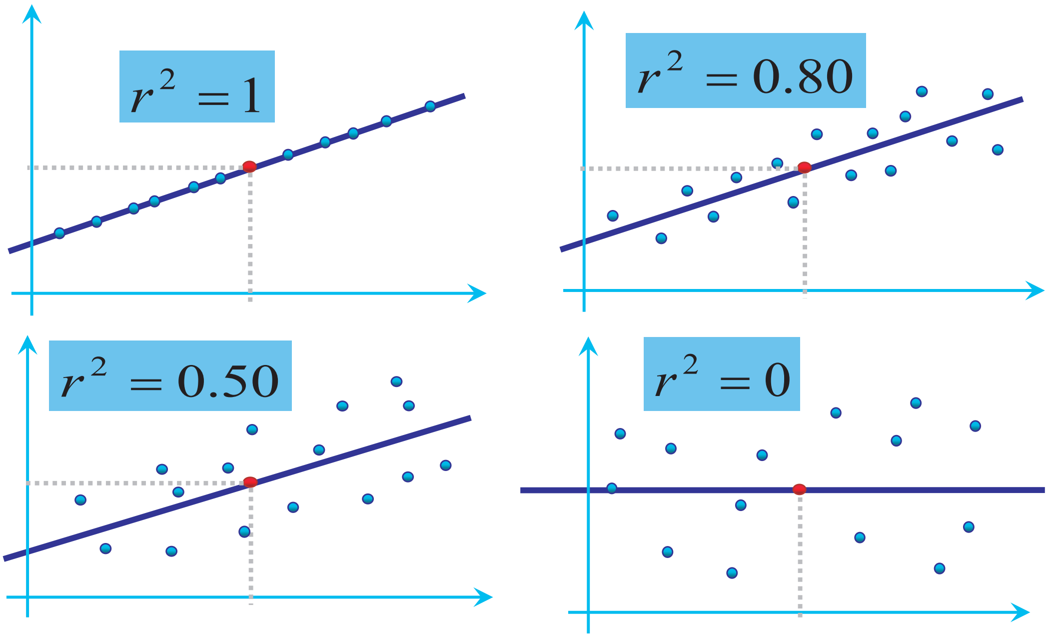


Obs.	*Var* -1	*Var* -2
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^n (x_i-\bar{x})^2}\sqrt{\sum_{i=1}^n (y_i-\bar{y})^2}}$$

- Adimensional
- $-1 \leq r_{xy} \leq +1$
- $|r_{xy}| = 1 \Leftrightarrow y_i = a + b\,x_i$

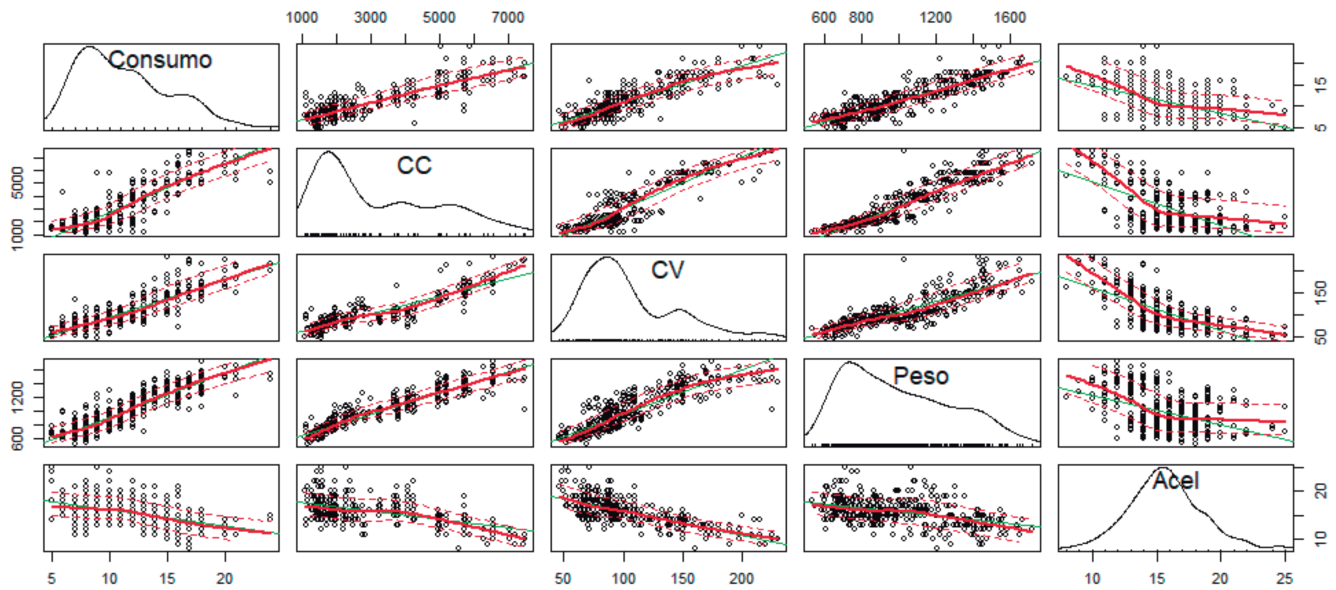
Coef. correlación



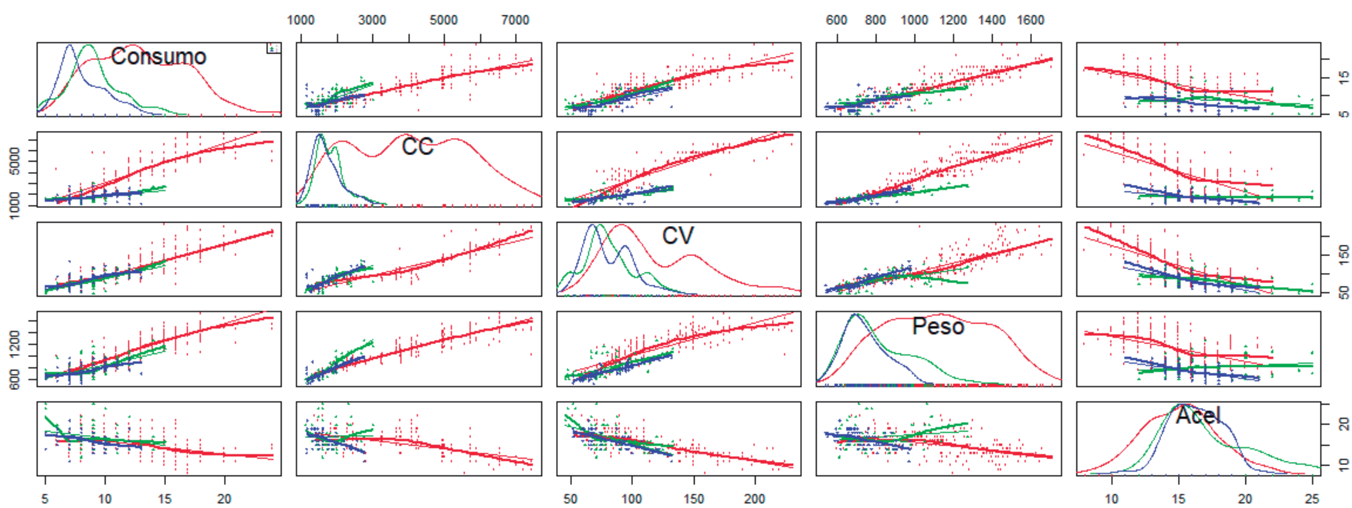
Matriz de Correlación

```
cor(coches[,1:5],use="complete.obs")
##          Consumo       CC       CV     Peso     Acel
## Consumo  1.00000  0.87259  0.85428  0.88510 -0.46609
## CC       0.87259  1.00000  0.89844  0.93389 -0.54895
## CV       0.85428  0.89844  1.00000  0.86289 -0.69628
## Peso     0.88510  0.93389  0.86289  1.00000 -0.42161
## Acel    -0.46609 -0.54895 -0.69628 -0.42161  1.00000
```

```r
library('car')
scatterplotMatrix(coches[1:5])
```




```r
scatterplotMatrix(coches[1:5],groups=coches$Origen,
cex=.2,by.groups = TRUE,col= rainbow(3),
main='USA=Rojo, UE=Verde, JAP= Azul')
```



Transformaciones Lineales

$$y_i = a_1 x_{1i} + a_2 x_{2i} + \cdots + a_k x_{ki} = \begin{pmatrix} a_1 & a_2 & \cdots & a_k \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ki} \end{pmatrix} = \mathbf{a}^{\mathrm{T}} \mathbf{x_i}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} \mathbf{a}^{\mathrm{T}} \mathbf{x_i}}{n} = \frac{\mathbf{a}^{\mathrm{T}} (\sum_{i=1}^{n} \mathbf{x_i})}{n} = \mathbf{a}^{\mathrm{T}} \overline{\mathbf{x}}$$

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})}{n} = \frac{\sum_{i=1}^{n} (\mathbf{a}^{\mathrm{T}} \mathbf{x_i} - \mathbf{a}^{\mathrm{T}} \overline{\mathbf{x}})(\mathbf{x_i^{\mathrm{T}}} \mathbf{a} - \overline{\mathbf{x}}^{\mathrm{T}} \mathbf{a})}{n}$$

$$= \mathbf{a}^{\mathrm{T}} \left(\frac{\sum_{i=1}^{n} (\mathbf{x_i} - \overline{\mathbf{x}})(\mathbf{x_i} - \overline{\mathbf{x}})^{\mathrm{T}}}{n} \right) \mathbf{a} = \mathbf{a}^{\mathrm{T}} \mathbf{S}^2 \mathbf{a}$$

$$\mathbf{S}^2 = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1k} \\ s_{12} & s_2^2 & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1k} & s_{2k} & \cdots & s_k^2 \end{pmatrix}$$

$$Var(a^T x) = (a_1, a_2, \cdots, a_k) \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1k} \\ s_{12} & s_2^2 & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1k} & s_{2k} & \cdots & s_k^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$$

$$= a_1^2 s_1^2 + a_2^2 s_2^2 + \cdots + a_k^2 s_k^2 + 2a_1 a_2 s_{12} + 2a_1 a_3 s_{13} + \cdots + 2a_{k-1} a_k s_{k-1,k}$$

Casos particulares

1. $y_i = x_{1i} + x_{2i}$

   $Var(y_i) = Var(x_{1i}) + Var(x_{2i}) + 2Cov(x_{1i}, x_{2i})$

2. $y_i = x_{1i} - x_{2i}$

   $Var(y_i) = Var(x_{1i}) + Var(x_{2i}) - 2Cov(x_{1i}, x_{2i})$

3. $y_i = a_1 x_{1i} + a_2 x_{2i}$

   $Var(y_i) = a_1^2 Var(x_{1i}) + a_2^2 Var(x_{2i}) + 2a_1 a_2 Cov(x_{1i}, x_{2i})$

Transformaciones lineales II

$$\left.\begin{array}{l} y_{1i} = a_{11}x_{1i} + a_{12}x_{2i} + \cdots + a_{1k}x_{ki} \\ y_{2i} = a_{21}x_{1i} + a_{22}x_{2i} + \cdots + a_{mk}x_{ki} \\ \quad\vdots \\ y_{mi} = a_{m1}x_{1i} + a_{m2}x_{2i} + \cdots + a_{mk}x_{ki} \end{array}\right\} \begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{mi} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ki} \end{pmatrix}$$

$$\mathbf{y_i} = \mathbf{A x_i}$$

$$\overline{\mathbf{y}} = \frac{\sum_{i=1}^{n} \mathbf{y_i}}{n} = \frac{\sum_{i=1}^{n} \mathbf{A x_i}}{n} = \frac{\mathbf{A}(\sum_{i=1}^{n} \mathbf{x_i})}{n} = \mathbf{A}\overline{\mathbf{x}}$$

Transformaciones lineales III

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{mi} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ki} \end{pmatrix} \longrightarrow \boldsymbol{y_i = Ax_i}$$

$$\mathbf{S}_\mathbf{Y}^2 = \frac{\sum_{i=1}^{n} (\mathbf{y}_\mathrm{i} - \bar{\mathbf{y}})(\mathbf{y}_\mathrm{i} - \bar{\mathbf{y}})^T}{n}$$

$$= \frac{\sum_{i=1}^{n} (\mathbf{Ax}_\mathrm{i} - \mathbf{A}\bar{\mathbf{x}})(\mathbf{x}_\mathrm{i}^\mathrm{T}\mathbf{A}^\mathrm{T} - \bar{\mathbf{x}}^\mathrm{T}\mathbf{A}^\mathrm{T})}{n} = \mathbf{A}\underbrace{\left(\frac{\sum_{i=1}^{n} (\mathbf{x}_\mathrm{i} - \bar{\mathbf{x}})(\mathbf{x}_\mathrm{i} - \bar{\mathbf{x}})^T}{n} \right)}_{\mathbf{S_X}^2}\mathbf{A}^\mathrm{T}$$

$$= \mathbf{AS}_\mathbf{X}^2\mathbf{A}^\mathrm{T}$$

Ejercicio:

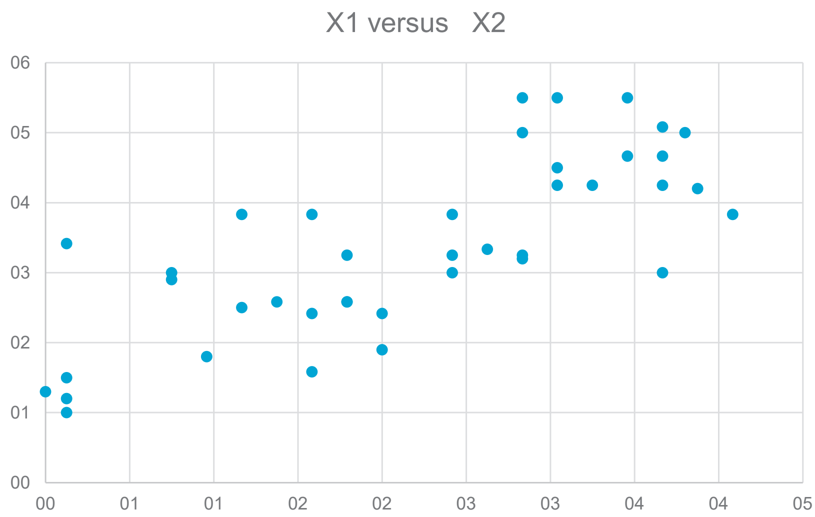
Notas de Diseño (X1) y Regresión (X2) y Nota Final (Y=X1+X2)

En la tabla siguiente se proporciona las notas de las dos pruebas de evaluación continua de la asignatura Diseño de Experimentos (X1) y Modelos de Regresión (X2) y la suma de ambas (Y=X1+X2). Además se proporcionan la media y la desviación típica de cada una de las tres variables.

1. Calcula la covarianza entre X1 y X2
2. Calcula la correlación entre X1 y X2
3. Proporciona la matriz de varianzas de X1 y X2 y la matriz de correlaciones de X1 y X2
4. Como las notas de X1 son en general inferiores a X2, se decide utilizar como nota final Z = 0.8 X1 + 1.2 X2. Calcula la media y la varianza de Z.
5. Calcula la matriz de correlación de las variables (Y y Z).

Notas de Diseño (X1) y Regresión (X2) y Nota Final (Y=X1+X2)

matrícula	X1	X2	Y	matrícula	X1	X2	Y
76020	3.3	4.3	**7.5**	76523	3.5	4.7	**8.1**
76032	0.1	3.4	**3.5**	76564	0.8	2.9	**3.7**
76056	3.7	5.1	**8.8**	76575	3.0	4.3	**7.3**
76077	3.0	4.5	**7.5**	76581	1.4	2.6	**4.0**
76079	3.7	4.3	**7.9**	76639	0.1	1.0	**1.1**
76084	1.0	1.8	**2.8**	76668	0.1	1.2	**1.3**
76133	3.0	5.5	**8.5**	76740	2.8	3.3	**6.1**
76151	1.8	3.3	**5.0**	76742	3.7	4.7	**8.3**
76218	2.8	5.0	**7.8**	76745	1.8	2.6	**4.4**
76235	2.4	3.8	**6.3**	76765	1.6	1.6	**3.2**
76241	1.8	2.6	**4.4**	76770	2.8	5.5	**8.3**
76251	2.4	3.3	**5.7**	76792	4.1	3.8	**7.9**
76268	3.7	3.0	**6.7**	76796	3.5	5.5	**9.0**
76272	2.0	2.4	**4.4**	76816	2.6	3.3	**6.0**
76303	0.1	1.5	**1.6**	76834	3.9	4.2	**8.1**
76354	1.2	3.8	**5.0**	76853	2.8	3.2	**6.0**
76377	0.8	3.0	**3.8**	76885	2.4	3.0	**5.4**
76433	0.0	1.3	**1.3**	76889	1.2	2.5	**3.7**
76435	1.6	3.8	**5.4**	76990	2.0	1.9	**3.9**
76517	1.6	2.4	**4.0**	76991	3.8	5.0	**8.8**
				Media	**2.19**	**3.37**	**5.56**
				DesvTipic	**1.21**	**1.24**	**2.29**



Solución:

Notas de Diseño (X1) y Regresión (X2) y Nota Final (Y=X1+X2)

$$1.\ Var(Y) = Var(X1) + Var(X2) + 2cov(X1, X2)$$

$$cov(X1, X2) = \frac{Var(X1) + Var(X2) - Var(Y)}{2} = 1.138$$

$$2.\ Cor(X1, X2) = \frac{Cov(X1, X2)}{\sqrt{Var(X1)}\sqrt{Var(X2)}} = 0.76$$

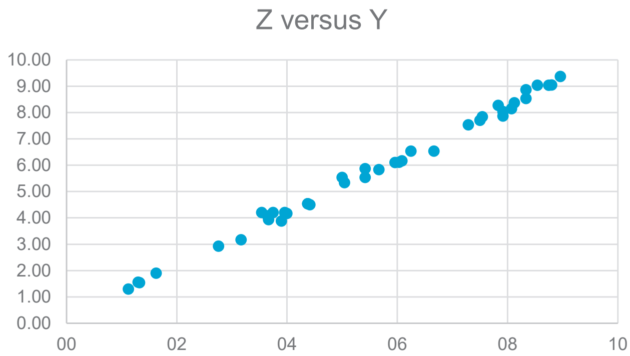
$$3.\ Var\begin{pmatrix} X1 \\ X2 \end{pmatrix} = \begin{pmatrix} 1.45 & 1.138 \\ 1.138 & 1.54 \end{pmatrix} \quad Cor\begin{pmatrix} X1 \\ X2 \end{pmatrix} = \begin{pmatrix} 1. & 0.76 \\ 0.76 & 1. \end{pmatrix}$$

4. $\bar{Z} = 0.8 \times 2.19 + 1.2 \times 3.37 = 5.79$

$$Var(Z) = 0.8^2 Var(X1) + 1.2^2 Var(X2) + 2 \times 0.8 \times 1.2 \times Cov(X1, X2) = 5.33$$

5. $Var\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ .8 & 1.2 \end{bmatrix} \begin{bmatrix} 1.45 & 1.138 \\ 1.138 & 1.54 \end{bmatrix} \begin{bmatrix} 1 & .8 \\ 1 & 1.2 \end{bmatrix} = \begin{bmatrix} 5.266 & 5.284 \\ 5.284 & 5.331 \end{bmatrix}$

$$Cor\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{bmatrix} 1 & .997 \\ .997 & 1 \end{bmatrix}$$



Capítulo 1: Estadstica descriptiva

1. Para los datos siguientes: 28, 22, 35, 42, 44, 53, 58, 41, 40, 32, 31, 38, 37, 61, 25, 35.
   (a) Calcula la mediana y los cuartiles.
   (b) Dibuja el box-plot
   (c) Construye un diagrama de tallo y hojas
   (d) Calcula la media y la desviación típica
   (e) Llamando x_i a los datos anteriores, calcula la media y la desviación típica para los datos y_i obtenidos como $y_i = 5 + 10x_i$.

2. A continuación se proporciona el consumo en *millas por galón* de 68 coches fabricados en Europa.

29.5	26.0	21.5	25.0	17.0	26.0	24.0	28.0	29.0	20.0
31.9	30.5	29.0	33.0	24.0	26.0	18.0	30.0	22.0	29.0
41.5	37.3	31.5	23.0	35.0	36.0	27.0	29.8	28.0	26.0
26.0	24.0	19.0	24.0	20.0	21.6	16.2	34.3	20.3	31.0
25.0	29.0	16.5	25.0	23.0	25.0	25.0	26.0	22.0	36.0
27.0	30.0	21.0	29.0	30.7	36.4	25.4	28.1	26.0	26.0
43.1	44.3	30.0	19.0	23.0	43.4	44.0	27.2		

   (a) Construye el histograma con 6 clases
   (b) Realiza el diagrama de tallos y hojas
   (c) Obtén la mediana y los cuartiles
   (d) Dibuja el box-plot. Indica las observaciones atípicas.
   (e) Transforma los datos con la función logarítmo: calcula la mediana y los cuartiles y dibuja el *box-plot* de los datos transformados.

3. En un departamento cuatro profesores imparten clases en grupos con 10, 18, 22 y 150 alumnos respectivamente. Si se pregunta a los profesores por el tamaño de su clase ¿cuál sería el valor medio y la desviación típica obtenida? ¿Y si se pregunta a todos los alumnos del departamento?

4. En la figura se presenta el diagrama de tallos y hojas de los residuos obtenidos de un diseño factorial. Representa el diagrama de caja (box plot) de los datos. (Nota.- La rama -6|91 representa los valores -0.69 y -0.61).

2	-6	|	91
2	-5	|	
4	-4	|	00
10	-3	|	766320
18	-2	|	98754310
29	-1	|	98654321100
(16)	-0	|	9977666554433211
36	0	|	015566677
27	1	|	2333478
20	2	|	134789
14	3	|	23455699
6	4	|	011355

5. Para n piezas se han medido dos dimensiones x_i e y_i. Es posible que la varianza de la variable x_i sea 4, la de y_i sea 9 y la de $z_i = x_i + y_i$ sea igual a 2? Justificar la respuesta.

6. Se toman 2 medidas x_i e y_i de n individuos. Dada dos constantes a y b positivas, demostrar que al multiplicar x_i por a e y_i por b, el coeficiente de correlación entre ambas no varía.

7. Demostrar que si entre dos variables x_i e y_i existe una relación lineal exacta $y_i = a + bx_i$, con $b > 0$, el coeficiente de correlación es uno.

8. Demostrar que el coeficiente de correlación es siempre en valor absoluto menor que uno.

9. En un proceso de fabricación se han medido tres variables y calculado la matriz de varianzas con el resultado siguiente:

$$\begin{pmatrix} 2 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

   (a) ¿Podemos afirmar que hay un error en los cálculos?

   (b) ¿Por qué?

10. A la variable x de media $\overline{x} = 100$ se le ha aplicado una transformación con el logaritmo decimal obteniéndose la nueva variable $y = \log_{10}(x)$. La media de la nueva variable es $\overline{y} = 2.5$. ¿Es posible este resultado?

Estadística

Parte I.

Tema 2. Fundamentos de Probabilidad

Experimento Aleatorio

El término “**experimento aleatorio**” se utiliza en la teoría de la probabilidad para referirse a un proceso cuyo resultado no es conocido de antemano con certeza.

“Suma de valores en el lanzamiento de 2 dados.”

Ejemplos

- Número de piezas defectuosas en una muestra de 100 piezas.
- Número de llamadas a una centralita telefónica en un día.
- Energía eléctrica consumida en Madrid durante un periodo de tiempo.

Espacio Muestral

Conjunto formado por todos los posibles resultados de un experimento aleatorio.

- DISCRETOS:
  - Lanzamiento de un DADO: S = {1,2,3,4,5,6}
  - Piezas defectuosas en una muestra de 100

    S = {0,1,2,...,100}
  - Llamadas a una centralita durante un día

    S = {0,1,2,3,...,∞}
- CONTINUOS:
  - Energía consumida en Madrid: S={[0, ∞)}

Suceso

Cualquier subconjunto del espacio muestral.

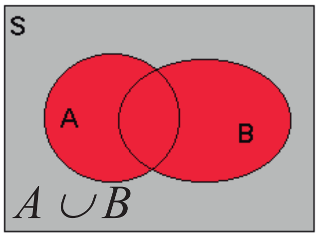
- "Obtener un número par al lanzar un dado":
  A = {2,4,6}
- "Observar menos de 5 piezas defectuosas en una muestra de 100": B = {0,1,2,3,4}
- "Tener más de 50 llamadas de teléfono en una hora": C = {51,52,...,∞}
- "Tener una demanda de energía eléctrica entre 300 Mwh y 400 Mwh" : D =(300,400)

Operaciones

Sean *A* y *B* dos subconjuntos de *S*

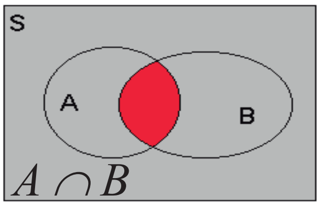
- Unión

$$A \cup B = \{x : (x \in A) \textbf{ o } (x \in B)\}$$



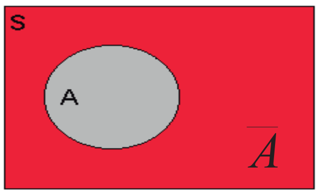
- Intersección

$$A \cap B = \{x : (x \in A) \textbf{ y } (x \in B)\}$$



- Complementario

$$\overline{A} = \{x : x \notin A\}$$



Axiomas de Probabilidad

Dado un espacio muestral S, una función de probabilidad asigna valores P(A) a cada suceso $A \subset S$ y satisface :

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. Para una secuencia de sucesos $A_1, A_2, \ldots, A_n$

que cumplen $A_i \cap A_j = \varnothing$ cuando $i \neq j$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Problema fundamental

Dado un espacio muestral discreto con resultados A_1, A_2, ..., A_n disjuntos, el experimento aleatorio queda caracterizado si asignamos un valor $P(A_i)$ no negativo a cada resultado A_i de forma que

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1.$$

- **Ejemplo**. Se lanza dos veces una moneda.

{XX,XC,CX,CC}

Se asigna probabilidad 1/4 a cada uno de los cuatro resultados anteriores.

¿ Es una asignación correcta?

Propiedades elementales

1. $P(\emptyset)=0.$
2. $P(\overline{A})=1-P(A).$
3. Si $A\subset B$ entonces $P(A)\leq P(B).$
4. Para dos sucesos cualesquiera $A,B\subset S$,

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

5. Para n sucesos $A_1,A_2,\ldots,A_n\subset S$,

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{i=1}^{n}\sum_{j>i}^{n}P(A_i\cap A_j)+$$

$$\sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{k>j}^{n}P(A_i\cap A_j\cap A_k)+\cdots+(-1)^{n+1}P(A_1\cap A_2\cap\cdots\cap A_n)$$

Asignación de probabilidades

1. Clásica (Laplace): Equiprobabilidad

2. Frecuencialista (von Mises, 1931)

3. Subjetiva

Clásica: sucesos equiprobables

Sea un experimento con un número finito *N* de resultados excluyentes y equiprobables, la probabilidad del suceso A es

$$P(A) = \frac{N(A)}{N},$$

donde N es el número de resultados posibles del experimento y N(A) el número de resultados favorables al suceso A.

Ejemplos (equiprobabilidad)

- Lanzamiento de una moneda. S={C,X}

$$P(C) = \frac{1}{2}$$

- Lanzamiento de un dado. S={1,2,3,4,5,6}

$$P(\text{"Número par"}) = \frac{3}{6} = \frac{1}{2}.$$

- Extracción de una de las 40 cartas de la baraja, S={1 Oros,2 Oros,...., Rey Bastos}

$$P(Bastos) = \frac{10}{40} = \frac{1}{4}.$$

Lanzamiento de dos dados

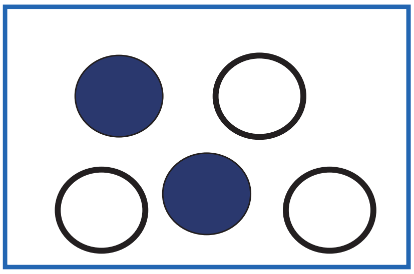
1er Dado

2º Dado

	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

P("suma 7") = 6/36 = 1/6

Urna: 2 Negras y 3 Blancas



1ª Bola

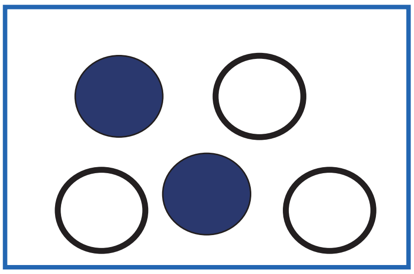
2ª Bola

	B1	B2	B3	N1	N2
B1		B2,B1	B3,B1	N1,B1	N2,B1
B2	B1,B2		B3,B2	N1,B2	N2,B2
B3	B1,B3	B2,B3		N1,B3	N2,B3
N1	B1,N1	B2,N1	B3,N1		N2,N1
N2	B1,N2	B2,N2	B3,N2	N1,N2	

Se extraen dos bolas al azar, una detrás de otra, sin reposición.

P("1ª Blanca y 2ª Negra") = 6/20 = 3/10

Urna: 2 Negras y 3 Blancas



1ª Bola (columnas) / 2ª Bola (filas)

	B1	B2	B3	N1	N2
B1	B1,B1	B2,B1	B3,B1	N1,B1	N2,B1
B2	B1,B2	B2,B2	B3,B2	N1,B2	N2,B2
B3	B1,B3	B2,B3	B3,B3	N1,B3	N2,B3
N1	B1,N1	B2,N1	B3,N1	N1,N1	N2,N1
N2	B1,N2	B2,N2	B3,N2	N1,N2	N2,N2

Se extraen dos bolas al azar, una detrás de otra, **con reposición.**

P("1ª Blanca y 2ª Negra") = 6/25

Combinatoria: 5 objetos tomados de dos en dos

SIN REEMPLAZAMIENTO — **CON REEMPLAZAMIENTO**

IMPORTA EL ORDEN

SIN REEMPLAZAMIENTO — Primera extracción

	1	2	3	4	5
1		(2,1)	(3,1)	(4,1)	(5,1)
2	(1,2)		(3,2)	(4,2)	(5,2)
3	(1,3)	(2,3)		(4,3)	(5,3)
4	(1,4)	(2,4)	(3,4)		(5,4)
5	(1,5)	(2,5)	(3,5)	(4,5)	

Número = 20

CON REEMPLAZAMIENTO — Primera Extracción

	1	2	3	4	5
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)
2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)
3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)
4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)

Número = 25

NO IMPORTA EL ORDEN

SIN REEMPLAZAMIENTO — Primera extracción

	1	2	3	4	5
1					
2	(1,2)				
3	(1,3)	(2,3)			
4	(1,4)	(2,4)	(3,4)		
5	(1,5)	(2,5)	(3,5)	(4,5)	

Número = 10

CON REEMPLAZAMIENTO — Primera extracción

	1	2	3	4	5
1	(1,1)				
2	(1,2)	(2,2)			
3	(1,3)	(2,3)	(3,3)		
4	(1,4)	(2,4)	(3,4)	(4,4)	
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)

Número = 15

Combinatoria: Número posible de reordenaciones de n objetos tomados de r en r

	SIN REEMPLAZAMIENTO	CON REEMPLAZAMIENTO
IMPORTA EL ORDEN	$\frac{n!}{(n-r)!}$	n^r
NO IMPORTA EL ORDEN	$\binom{n}{r}$	$\binom{n+r-1}{r}$


```
S=('A','B','C','D') elección de tres elementos
```

Sin Reemplazamiento — **Con Reemplazamiento**

Orden Importa

Sin Reemplazamiento:

```
##      X1 X2 X3    ##      X1 X2 X3
## 1    A  B  C     ## 13   A  C  D
## 2    A  C  B     ## 14   A  D  C
## 3    C  A  B     ## 15   D  A  C
## 4    C  B  A     ## 16   D  C  A
## 5    B  C  A     ## 17   C  D  A
## 6    B  A  C     ## 18   C  A  D
## 7    A  B  D     ## 19   B  C  D
## 8    A  D  B     ## 20   B  D  C
## 9    D  A  B     ## 21   D  B  C
## 10   D  B  A     ## 22   D  C  B
## 11   B  D  A     ## 23   C  D  B
## 12   B  A  D     ## 24   C  B  D
```

Con Reemplazamiento:

```
##      X1 X2 X3    ##      X1 X2 X3    ##      X1 X2 X3    ##      X1 X2 X3
## 1    A  A  A     ## 17   A  A  B     ## 33   A  A  C     ## 49   A  A  D
## 2    B  A  A     ## 18   B  A  B     ## 34   B  A  C     ## 50   B  A  D
## 3    C  A  A     ## 19   C  A  B     ## 35   C  A  C     ## 51   C  A  D
## 4    D  A  A     ## 20   D  A  B     ## 36   D  A  C     ## 52   D  A  D
## 5    A  B  A     ## 21   A  B  B     ## 37   A  B  C     ## 53   A  B  D
## 6    B  B  A     ## 22   B  B  B     ## 38   B  B  C     ## 54   B  B  D
## 7    C  B  A     ## 23   C  B  B     ## 39   C  B  C     ## 55   C  B  D
## 8    D  B  A     ## 24   D  B  B     ## 40   D  B  C     ## 56   D  B  D
## 9    A  C  A     ## 25   A  C  B     ## 41   A  C  C     ## 57   A  C  D
## 10   B  C  A     ## 26   B  C  B     ## 42   B  C  C     ## 58   B  C  D
## 11   C  C  A     ## 27   C  C  B     ## 43   C  C  C     ## 59   C  C  D
## 12   D  C  A     ## 28   D  C  B     ## 44   D  C  C     ## 60   D  C  D
## 13   A  D  A     ## 29   A  D  B     ## 45   A  D  C     ## 61   A  D  D
## 14   B  D  A     ## 30   B  D  B     ## 46   B  D  C     ## 62   B  D  D
## 15   C  D  A     ## 31   C  D  B     ## 47   C  D  C     ## 63   C  D  D
## 16   D  D  A     ## 32   D  D  B     ## 48   D  D  C     ## 64   D  D  D
```

Orden No Importa

Sin Reemplazamiento:

```
##      X1 X2 X3
## 1    A  B  C
## 2    A  B  D
## 3    A  C  D
## 4    B  C  D
```

Con Reemplazamiento:

```
##      X1 X2 X3    ##      X1 X2 X3    ##      X1 X2 X3    ##      X1 X2 X3
## 1    A  A  A     ## 6    A  B  C     ## 11   B  B  B     ## 16   B  D  D
## 2    A  A  B     ## 7    A  B  D     ## 12   B  B  C     ## 17   C  C  C
## 3    A  A  C     ## 8    A  C  C     ## 13   B  B  D     ## 18   C  C  D
## 4    A  A  D     ## 9    A  C  D     ## 14   B  C  C     ## 19   C  D  D
## 5    A  B  B     ## 10   A  D  D     ## 15   B  C  D     ## 20   D  D  D
```

Lotería Primitiva

1 - 6 - 21 - 29 - 33 -43

1. La primitiva. Se eligen 6 números distintos del 1 al 49, ambos inclusive.

- Probabilidad de acertar los 6.
- Probabilidad de acertar 5.
- Probabilidad de acertar 4.
- Probabilidad de no acertar ninguno.
- Probabilidad de que salga un número concreto, por ejemplo el número 1.

Ejemplo 1: Lotería Primitiva

$$P(\text{Acertar 6}) = \frac{1}{\binom{49}{6}} = \frac{1}{13.983.816} = 0{,}000000072$$

$$P(\text{Acertar 5}) = \frac{\binom{6}{5}\times\binom{43}{1}}{\binom{49}{6}} = \frac{258}{13.983.816} = 0{,}000018$$

$$P(\text{Acertar 4}) = \frac{\binom{6}{4}\times\binom{43}{2}}{\binom{49}{6}} = \frac{13.545}{13.983.816} = 0{,}00097$$

$$P(\text{Ninguno}) = \frac{\binom{43}{6}}{\binom{49}{6}} = \frac{6.096.454}{13.983.816} = 0{,}44$$

$$P(\text{Salga el 1}) = \frac{\binom{48}{5}}{\binom{49}{6}} = \frac{6}{49} = 0{,}1224$$

Ejemplo 2: Metro

En una estación de metro hay 5 pasajeros esperando a un tren con 10 vagones, si cada pasajero elige un vagón al azar, ¿cuál es la probabilidad de que todos elijan un vagón diferente?

$$P(A) = \frac{N(A)}{N} = \frac{10\times 9\times 8\times 7\times 6}{10^5} = 0.3024$$

Ejemplo 3: Fabricación por lotes

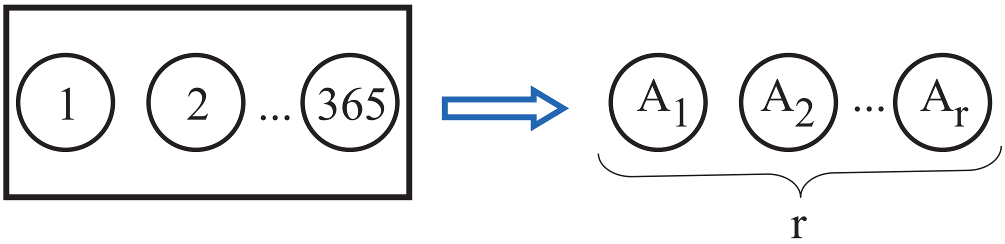
De un lote con 100 piezas se toman al azar 10, si todas las piezas elegidas son buenas se acepta el lote y se rechaza en caso contrario. ¿Cuál es la probabilidad de aceptar un lote con 10 piezas defectuosas?

$$N = \binom{100}{10} = \frac{100\,!}{10\,!\,90\,!}; \quad N(A) = \binom{90}{10} = \frac{90\,!}{80\,!\,10\,!}$$

$$P(A) = \frac{N(A)}{N} = \frac{90\,!\;90\,!}{80\,!\;100\,!} = \frac{90\times 89\times\cdots\times 81}{100\times 99\times\cdots\times 91} = 0.330$$

Ejemplo 4: Cumpleaños

Probabilidad de que en un grupo de *r = 25* personas haya al menos dos con el mismo cumpleaños.



$$A = \text{" No haya ninguna coincidencia"}$$

$$P(A) = \frac{365 \times (365-1) \times \cdots \times (365-r+1)}{365^r}$$

$$P(\overline{A}) = 1 - P(A), \qquad r = 25 \rightarrow P(\overline{A}) = 0.578$$

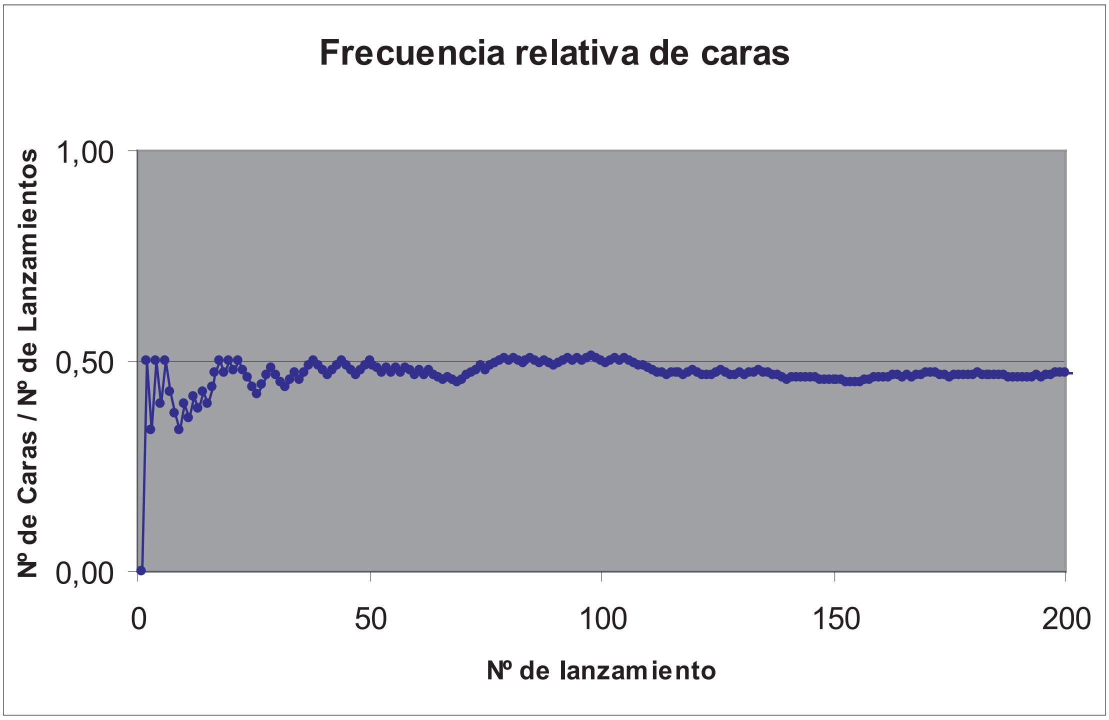
Probabilidad y Frecuencia Relativa

La probabilidad *P(A)* de un suceso *A* es el límite

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

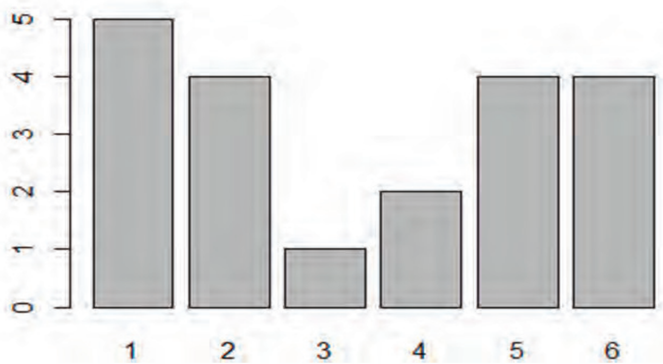
dónde n_A es el número de veces que ha ocurrido *A* al repetir ir el experimento *n* veces en *idénticas* condiciones.

Proporción de caras observadas al lanzar una moneda 200 veces



Simulación lanzamiento de un dado

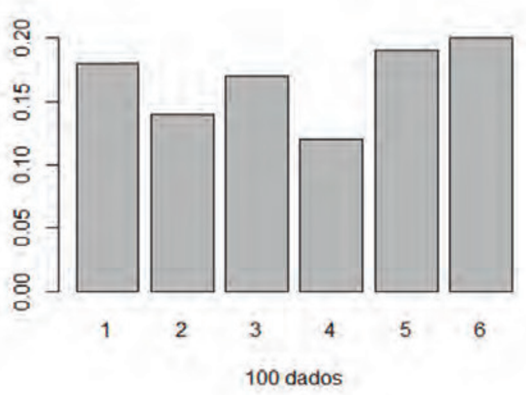
```
u = sample(1:6,size=20,replace=TRUE)
u
##  [1] 4 1 3 1 2 1 6 4 2 5 6 2 6 5 6 1 5 5 1 2
table(u)
## u
## 1 2 3 4 5 6
## 5 4 1 2 4 4
barplot(table(u))
```

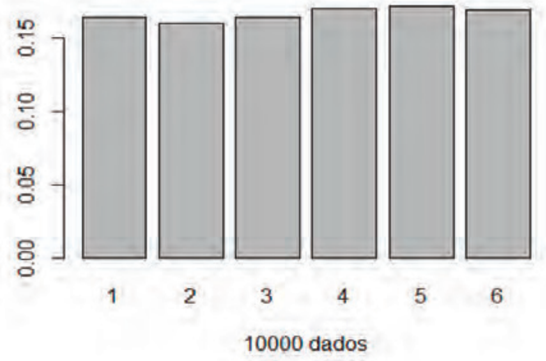


Simulación lanzamiento de un dado

```
u1 = sample(1:6, size=100, replace=TRUE)
table(u1)/100
## u1
##    1    2    3    4    5    6
## 0.18 0.14 0.17 0.12 0.19 0.20
barplot(table(u3)/100,xlab='100 dados’)
```

```
u2 = sample(1:6, size=10000, replace=TRUE)
table(u2)/10000
## u2
##      1      2      3      4      5      6
## 0.1645 0.1599 0.1644 0.1704 0.1715 0.1693
barplot(table(u2)/10000,xlab='10000 dados')
```





Resultados de 3036 sorteos de la Primitiva

Núm	Veces	%
1	376	0.12385
2	361	0.11891
3	400	0.13175
4	360	0.11858
5	381	0.12549
6	385	0.12681
7	367	0.12088
8	345	0.11364
9	366	0.12055
10	374	0.12319
11	382	0.12582
12	369	0.12154
13	357	0.11759
14	378	0.12451
15	376	0.12385
16	363	0.11957
17	365	0.12022

Núm	Veces	%
18	351	0.11561
19	354	0.11660
20	345	0.11364
21	369	0.12154
22	388	0.12780
23	386	0.12714
24	348	0.11462
25	356	0.11726
26	364	0.11989
27	361	0.11891
28	364	0.11989
29	380	0.12516
30	398	0.13109
31	363	0.11957
32	363	0.11957
33	358	0.11792
34	373	0.12286

Núm	Veces	%
35	374	0.12319
36	379	0.12484
37	378	0.12451
38	405	0.13340
39	408	0.13439
40	375	0.12352
41	386	0.12714
42	363	0.11957
43	361	0.11891
44	363	0.11957
45	393	0.12945
46	372	0.12253
47	403	0.13274
48	379	0.12484
49	351	0.11561
TOTAL	**18216**	**0.12245**

La frecuencia con la que ha aparecido cada número es muy próxima a 0.12245=6/49, que es la probabilidad de que aparezca un número concreto en cada sorteo

Simulación con R: muestras

```
sample( 1:49, size = 6, replace = FALSE) # Extrae 6 números al azar
        entre 1 y 49 sin reemplazamiento

## [1]  8 35  3 40 32 23
```

Calculo por simulación de la probabilidad de aparición del número 27

```
u = rep(0,10000)          # creo un vector con 10.000 ceros
for (k in 1:10000)        # Bucle k toma valores de 1 a 10.000
{                         # comienzo bucle
  x = sample(1:49, size = 6, replace = FALSE) # sorteo k
  v = x == 27 # por ejem. v es [FALSE FALSE FALSE FALSE TRUE FALSE],
                indicaría que el quinto número es el 27
  u[k] = sum(v) # sum(v) es cero o uno , según esté o no el 27
}                         # fin del bucle
sum(u)/10000 # Probabilidad de obtener 27

## [1] 0.1255
```


Simulación con R: cumpleaños

Simular con R una muestra al azar de 25 fechas de cumpleaños y comprobar si hay repeticiones.

```
x = sample(1:365,25,replace=TRUE) # extrae con reemplazamiento 25 num.
u = sort(x) # los ordena
print(u)

##  [1]   3   4   6  13  32  41  41  68  74  88  89 121 204 207 212 214 221
## [18] 232 272 299 312 318 330 330 353

duplicated(u) # indica si hay valores repetidos (FALSE no y TRUE sí)

##  [1] FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE TRUE FALSE

sum(duplicated(u)) # número de 'TRUES'

## [1] 2
```

Simulación con R: ejemplos

1. **Lanzamiento de 10 monedas**

```
x = sample(c('C','X'), size = 10, replace = TRUE)
print (x)
##  [1] "C" "X" "X" "C" "X" "X" "X" "X" "X" "C“
```

2. **Lanzamiento de 3 dados**

```
y = sample(1:6, size = 3, replace = TRUE)
print(y)
## [1] 2 4 4
```

3. **Quiniela al azar**

```
z = sample(c('1','X','2'),size = 14, replace = TRUE)
print (z)
##  [1] "X" "X" "2" "2" "X" "1" "X" "1" "X" "2" "X" "2" "2" "2“
```

4. **Quiniela al azar no equiprobable**

```
u = sample(c('1','X','2'),size = 14, replace = TRUE, prob = c(.7,.2,.1))
print (u)
##  [1] "2" "1" "X" "1" "2" "1" "X" "1" "1" "X" "X" "1" "1" "1"
```


Probabilidad Condicionada

Definición. Sea *B* un suceso con probabilidad distinta de cero, se define probabilidad del suceso *A* dado *B* a:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Probabilidad Condicionada

	Mujeres (M)	Hombres (H)	**TOTAL**
Fumadores (F)	0,12	0,18	**0,30**
No Fumadores (N)	0,39	0,31	**0,70**
TOTAL	**0,51**	**0,49**	**1,00**

$$P(F) = 0{,}30 \quad \begin{cases} P(F \mid H) = \dfrac{0{,}18}{0{,}49} = 0{,}37 \\ \\ P(F \mid M) = \dfrac{0{,}12}{0{,}51} = 0{,}24 \end{cases}$$

Utilidad

- Actualizar probabilidad del suceso *A* en función de la información disponible *I*

$$P(A|I) = P(A \cap I)/P(I)$$

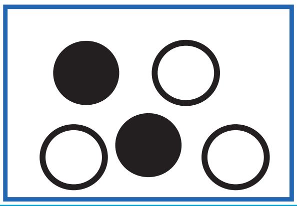
- Cálculo de la intersección de sucesos

$$P(A \cap B) = P(A|B)P(B)$$

- Cálculo de probabilidad de un suceso

$$P(A) = P((A \cap B) \cup (A \cap \overline{B})) = P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})$$

Ejemplo Urna



Probabilidad de "1ª Blanca y 2ª Negra"

- **Sin reemplazamiento**:

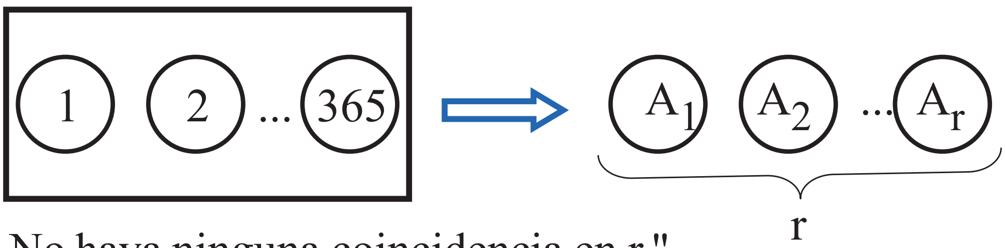
$$P(B1 \cap N2) = P(B1)\,P(N2|\,B1)$$
$$= (3/5)(2/4) = 3/10$$

- **Con reemplazamiento**:

$$P(B1 \cap N2) = P(B1)\,P(N2|\,B1)$$
$$= (3/5)(2/5) = 6/25$$

Cumpleaños

Probabilidad de que en un grupo de $r = 25$ personas haya al menos dos con el mismo cumpleaños.



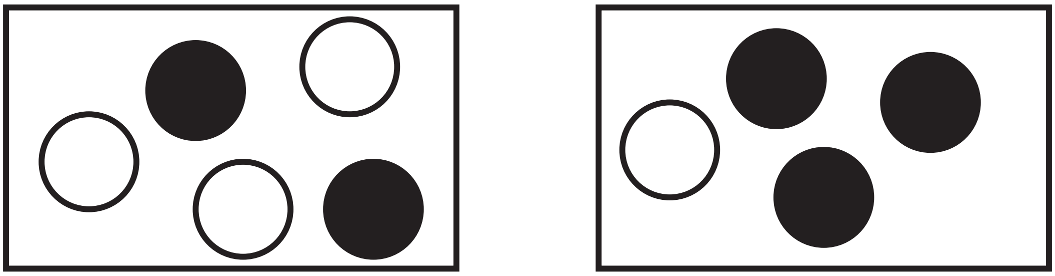
$$B_r = \text{" No haya ninguna coincidencia en r "}$$

$$P(B_r) = P(B_1)P(B_2 \mid A_1)P(B_3 \mid A_1 \neq A_2)\cdots P(B_r \mid A_1 \neq A_2 \neq \cdots \neq A_{r-1})$$

$$= 1 \times \frac{365-1}{365} \times \frac{365-2}{365} \times \cdots \times \frac{365-r+1}{365}$$

$$P(\overline{B_r}) = 1 - P(B_r), \qquad r = 25 \rightarrow P(\overline{B_r}) = 0.578$$

Ejemplo



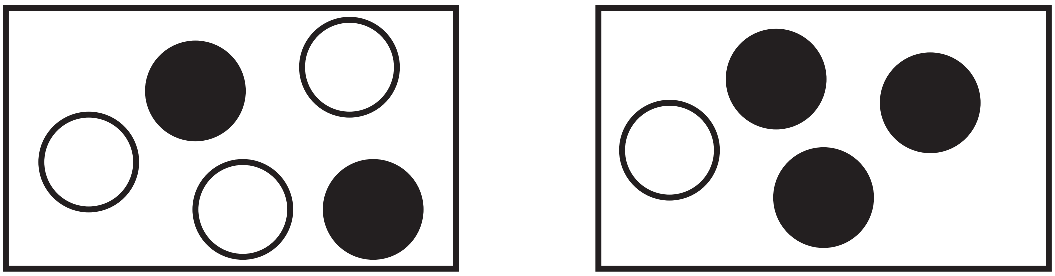
Urna **U1** Urna **U2**

Se elige una urna al azar y se extrae una bola: ¿ $\underline{P(\text{Blanca})}$?

$$P(B) = P(B \mid U1)P(U1) + P(B \mid U2)P(U2)$$

$$= \frac{3}{5} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{17}{40} = 0.425$$

Ejemplo (cont.)



Urna **U1** Urna **U2**

Se toma al azar una bola de U1 y se mete en U2. Se extrae una bola de U2: ¿ $\underline{P(\text{Blanca})}$?

$$P(B) = P(B \mid B1)P(B1) + P(B \mid N1)P(N1)$$

$$= \frac{2}{5} \times \frac{3}{5} + \frac{1}{5} \times \frac{2}{5} = \frac{8}{25} = 0.32$$

Cuentas de colores

Un coleccionista tiene 73 cuentas de 6 colores en un vaso: 8 rojas, 9 rosas, 11 verdes, 12 azules, 15 amarillas y 18 moradas. Saca dos cuentas al azar sin reposición:

1. ¿Cuál es la probabilidad de que sean las dos rojas?
2. ¿Cuál es la probabilidad de que una sea roja y otra verde?
3. ¿Cuál es la probabilidad de que ninguna sea roja?
4. ¿Cuál es la probabilidad de que sean del mismo color?
5. ¿Cuál es la probabilidad de que sean diferentes?

Rojas	8
Rosas	9
Verdes	11
Azules	12
Amarillas	15
Moradas	18

Cuentas de colores: Solución

1. $P(R_1 \cap R_2) = P(R_1)P(R_2|R_1) = \frac{8}{73} \times \frac{7}{72}$

2. $P(R_1 \cap V_2) + P(V_1 \cap R_2) = \frac{8}{73} \times \frac{11}{72} + \frac{11}{73} \times \frac{8}{72}$

3. $\mathrm{P}(\bar{R}_1 \cap \bar{R}_2) = \frac{65}{73} \times \frac{64}{72}$

4. $\frac{8 \times 7 + 9 \times 8 + 11 \times 10 + 12 \times 11 + 15 \times 14 + 18 \times 17}{73 \times 72}$

5. $1 - \frac{8 \times 7 + 9 \times 8 + 11 \times 10 + 12 \times 11 + 15 \times 14 + 18 \times 17}{73 \times 72}$

Independencia

Si el conocimiento de la ocurrencia de un suceso *B* cambia la probabilidad de que ocurra otro *A*, se dice que *A* y *B* son **dependientes**, en ese caso $P(A|B) \neq P(A)$.

Cuando el suceso *A* es independiente de *B*, la ocurrencia de *B* no cambia la probabilidad de *A*, es decir $P(A|B) = P(A)$.

Como $P(A|B) = P(A \cap B)/P(B)$,

$$\boxed{A \text{ y } B \textbf{ son independientes} \Leftrightarrow P(A \cap B) = P(A)\, P(B)}$$

Lanzamiento de dos monedas

$S = \{CC, CX, XC, XX\}$

Hipótesis:

- Monedas equilibradas: $P(C) = P(X)$
- Independientes

$$P(CC) = P(C)\,P(C) = (1/2)\times(1/2) = 1/4$$

$$P(CX) = P(C)\,P(X) = (1/2)\times(1/2) = 1/4$$

$$P(XC) = P(X)\,P(C) = (1/2)\times(1/2) = 1/4$$

$$P(XX) = P(X)\,P(X) = (1/2)\times(1/2) = 1/4$$

Independencia (3 o más sucesos)

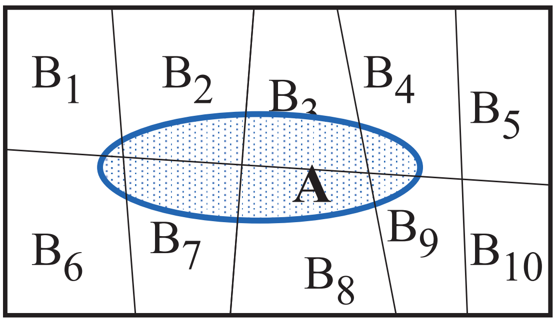
- Tres sucesos *A, B* y *C* son independientes si

$$\begin{cases} \bullet P(A \cap B \cap C) = P(A)\, P(B)\, P(C) \\ \bullet P(A \cap B) = P(A)\, P(B) \\ \bullet P(A \cap C) = P(A)\, P(C) \\ \bullet P(B \cap C) = P(B)\, P(C) \end{cases}$$

- Los sucesos $A_1, A_2, \ldots, A_n$ son independientes si cualquier subconjunto $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$, cumple

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k})$$

Probabilidad Total



Partición.

$$B_1, B_2, \ldots, B_n : \quad B_j \subset S$$

$$B_i \cap B_j = \varnothing, \quad \forall i \neq j$$

$$B_1 \cup B_2 \cup \cdots \cup B_n = S$$

$$\begin{aligned} P(A) &= P(A \cap S) \\ &= P[A \cap (B_1 \cup B_2 \cup \cdots \cup B_n)] \\ &= P[(A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)] \\ &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n) \end{aligned}$$

$$\boxed{P(A) = P(A \mid B_1) P(B_1) + P(A \mid B_2) P(B_2) + \cdots P(A \mid B_n) P(B_n)}$$

Teorema de Bayes

Sea $B_1, B_2, \ldots, B_n$ una partición del espacio S tal que $P(B_j) > 0$, para $j = 1,2,\ldots,n$ y sea A cualquier suceso con $P(A) > 0$, entonces para cualquier B_i :

$$\boxed{P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.}$$

Ejemplo Bayes 1: Almacén

M-1	M-2	M-3
5 % D	20 % D	10 % D
200 p/h	100 p/h	100 p/h

ALMACÉN

El porcentaje de piezas defectuosas fabricadas por tres máquinas es 5%, 20% y 10%. La primera fabrica 200 piezas por hora y las otras dos 100 piezas por hora. Todas las piezas fabricadas se llevan a un almacén. Al final del día se toma una pieza del almacén y es defectuosa, ¿cuál es la probabilidad de que proceda de M-1, M-2 o M-3?

$$P(M_1 \mid D) = \frac{P(D \mid M_1)P(M_1)}{P(D \mid M_1)P(M_1) + P(D \mid M_2)P(M_2) + P(D \mid M_3)P(M_3)}$$
$$= \frac{0.05 \times 0.5}{0.05 \times 0.5 + 0.20 \times 0.25 + 0.10 \times 0.25} = 0.25$$

$$P(M_2 \mid D) = \frac{P(D \mid M_2)P(M_2)}{P(D \mid M_1)P(M_1) + P(D \mid M_2)P(M_2) + P(D \mid M_3)P(M_3)}$$
$$= \frac{0.20 \times 0.25}{0.05 \times 0.5 + 0.20 \times 0.25 + 0.10 \times 0.25} = 0.50$$

$$P(M_3 \mid D) = \frac{P(D \mid M_3)P(M_3)}{P(D \mid M_1)P(M_1) + P(D \mid M_2)P(M_2) + P(D \mid M_3)P(M_3)}$$
$$= \frac{0.10 \times 0.25}{0.05 \times 0.5 + 0.20 \times 0.25 + 0.10 \times 0.25} = 0.25$$

$$P(M_1 \mid D) + P(M_2 \mid D) + P(M_3 \mid D) = 1$$

Ejemplo Bayes 2: Virus

Si una persona es portadora del virus A, un análisis de sangre lo detecta el 99% de las veces. Sin embargo, el test también proporciona "*falsos positivos*", indicando la presencia del virus en el 3% de personas sanas. Si sólo 5 de cada 1000 personas tienen el virus, ¿cuál es la probabilidad de que una persona tenga el virus realmente si el análisis ha dado positivo?

$$V = \text{"Tener el Virus"} \qquad S = \text{"El análisis es positivo"}$$

$$P(V \mid S) = \frac{P(V \cap S)}{P(S)} = \frac{P(S \mid V)P(V)}{P(S \mid V)P(V) + P(S \mid \overline{V})P(\overline{V})}$$
$$= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.03 \times 0.995} = 0.142$$

Ejemplo Virus
(Aplicado a 1.000.000 personas)

	SANOS	ENFERMOS	Total
NEGATIVO	965.150	50	965.200
POSITIVO	29.850	4.950	34.800
Total	995.000	5.000	1.000.000

Entre los **34.800** que han dado positivo, sólo **4.950** tienen el virus

$$P(V|S) = 4.950/34.800 = 0.142$$

Capítulo 2: Probabilidad Elemental

1. Sean A y B dos sucesos no disjuntos cualesquiera de un experimento aleatorio. Indicar cual de las siguientes probabilidades es mayor $P_1 = P(A \cap B \mid A)$ ó $P_2 = P(A \cap B \mid A \cup B)$. Justificar la respuesta.

2. Dado dos sucesos A y B, tal que $P(A) > 0$ y $P(B) > 0$ indicar, justificando la respuesta, si son ciertas o no las afirmaciones siguientes:

   a) Si A y B son sucesos mutuamente excluyentes entonces no pueden ser independientes.

   b) Si A y B son independientes entonces no pueden ser mutuamente excluyentes.

3. Demostrar que si dos sucesos A y B son independientes, también los son A y $\overline{B}$, y $\overline{A}$ y $\overline{B}$, donde $\overline{A}$ y $\overline{B}$ representan los sucesos complementarios de A y B respectivamente.

4. Sean dos sucesos equiprobables e independientes tal que la probabilidad de su unión es 0.75. Obtener la probabilidad de cada uno de ellos.

5. Se ha realizado una encuesta entre los estudiantes de grado del MIT (Massachusetts Institute of Technology) para conocer sus preferencias tecnológicas. El 35 % de los entrevistados tienen un iPhone y un IPad, el 80 % tienen al menos uno de estos dispositivos y el 60 % no tiene iPad. Se elige un estudiante al azar, calcula la probabilidad de que:

   a) Disponga de iPhone y no de iPad.

   b) Tenga un iPad pero no un Iphone.

   c) Tenga únicamente uno de los dos dispositivos.

   d) No Disponga de ninguno de los dos dispositivos.

6. Una encuesta realizada entre inmigrantes proporciona la siguiente información: el 80 % de los jóvenes entre 18 y 25 años no tiene trabajo, el 75 % de los jóvenes en esa edad no están matriculados en ningún centro educativo (no estudian). Si se toma un joven al azar de la muestra,¿cuál es la probabilidad de que ni estudie ni trabaje? (Otra forma de hacer la misma pregunta, ¿cuál es la proporción de jóvenes de la encuesta que ni estudian ni trabajan?

7. Demostrar que para cualquier par de sucesos, A_1, A_2 se cumple que

$$P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1,$$

   demostrar que la extensión a n sucesos es $P(A_1 \cap A_2 \cap \cdots \cap A_n) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$.

8. La probabilidad de que un componente de una máquina se averíe antes de 100 horas es 0.01. La máquina tiene 50 componentes; calcular la probabilidad de avería de la máquina antes de 100 horas en los casos siguientes:

   a) La máquina se avería cuando lo hace uno o más componentes.

   b) La máquina se avería cuando fallan dos o más componentes.

   c) La máquina sólo se avería cuando lo hacen todos los componentes.

9. Tres fichas están marcadas con las letras A, B, C, respectivamente, y una cuarta con las tres ABC. Se meten en una urna y se toma una de ellas al azar. Se pregunta si los tres sucesos consistentes en la presencia de la letra A, la letra B o la C sobre la ficha son o no independientes.

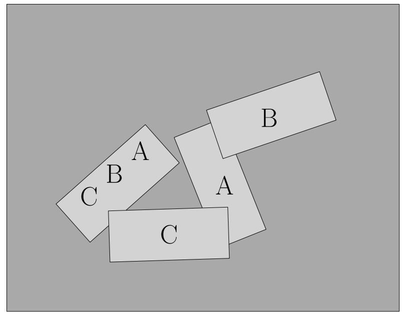


Figura 1: Cuatro fichas con letras A, B, C y ABC

10. En un campeonato de tenis usted tiene la opción de escoger la secuencia de partidos $A-B-A$ o la $B-A-B$, donde A y B indican sus oponentes. Para clasificarse debe usted ganar dos partidos consecutivos. El jugador A es mejor que el B. ¿ Qué secuencia será preferida?

11. Un jurado de tres miembros que decide por mayoría tiene dos miembros que deciden independientemente el veredicto correcto con probabilidad p y el tercero lanza una moneda. Si un juez tiene probabilidad p de acertar, ¿cuál de los dos sistemas tiene mayor probabilidad de acertar?

12. Una caja contiene r bolas rojas y b bolas blancas. Se elige una bola al azar de la caja (es decir todas las bolas tienen las misma probabilidad de ser elegidas), y después se extrae otra al azar entre las restantes de la caja. Calcular la probabilidad de

    a) Las dos bolas sean rojas
    b) Las dos bolas sean blancas
    c) La primera bola sea roja y la segunda blanca
    d) Las dos bolas sean de distinto color

13. Una caja contiene r bolas rojas y b bolas blancas. Se elige una bola al azar de la caja (es decir todas las bolas tienen las misma probabilidad de ser elegidas), después se devuelve a la caja, se mezcla y se extrae otra al azar. Calcular la probabilidad de

    a) Las dos bolas sean rojas
    b) Las dos bolas sean blancas
    c) La primera bola sea roja y la segunda blanca
    d) Las dos bolas sean de distinto color

14. Una urna contiene 3 bolas azules, 2 blancas y 5 rojas. Si se extraen 6 bolas con reposición, calcular la probabilidad de obtener 2 bolas de cada color.

15. El sorteo de la lotería primitiva consiste en acertar los seis números extraídos al azar y sin reposicin de una bombo que contiene 49 bolas numeradas del 1 al 49. Un participante ha realizado una apuesta múltiple y ha tachado 8 números. Calcular la probabilidad de acertar 6, 5, 4, 3, 2, 1 o ninguno de los números premiados. Si juega todos los días del año (365 sorteos) la misma apuesta ¿cuál es la probabilidad de que no acierte más de cuatro en ningún sorteo del año?

16. Si la probabilidad de que un interruptor cualquiera de la figura 2 se encuentre cerrado (es decir, pase corriente) es p, calcular la probabilidad de que pase corriente de A a B para las tres configuraciones de la figura (2).

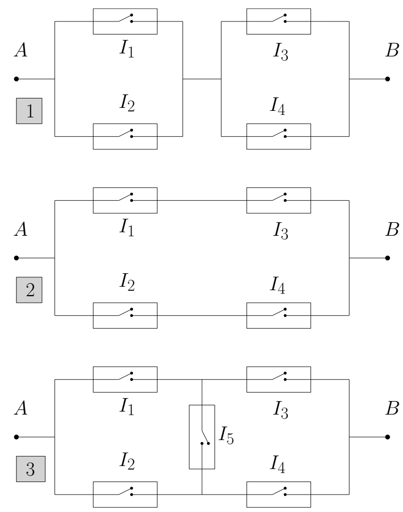
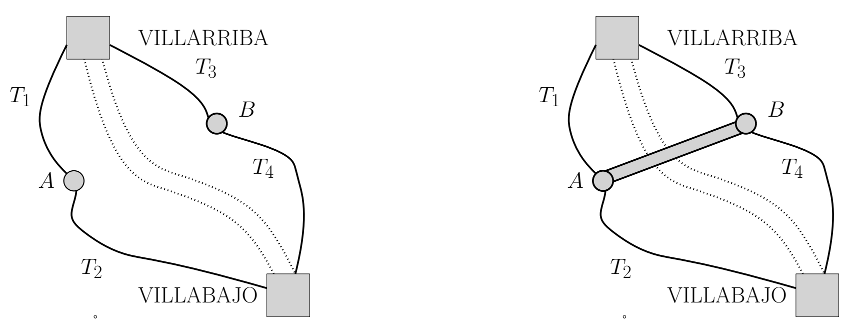


Figura 2: Tres circuitos

17. Un matrimonio tiene dos hijos y se sabe que uno de ellos es varón. Aceptando que la probabilidad de varón y hembra es la misma. ¿ Cuál es la probabilidad de que los dos sean varones?

18. Dos pueblos de alta montaña están comunicados por cuatro tramos de carretera según se muestra en la Figura 3a. Los días de invierno debido a la nieve es frecuente que las carreteras estén cortadas. Cuando nieva, la probabilidad de que cualquier tramo sea transitable es 0.8 con independencia de lo que ocurra con el resto. ¿Cuál es la probabilidad de que un vehículo pueda viajar de Villarriba a Villabajo un día con nieve? Para mejorar las comunicaciones se piensa construir un nuevo tramo de carretera entre los puntos A y B. Cuando nieva la probabilidad de accesibilidad es también 0.8 con independencia del resto. ¿Cuál es la probabilidad de que una persona pueda hacer el viaje anterior con esta nueva configuración 3b?



(a) Sin puente (b) Con puente

Figura 3: Red de carreteras entre Villarriba y Villabajo

19. Dos estudiantes A y B están matriculados en un curso. A asiste a las clases el 80 % de los días y B el 60 %, siendo las asistencias de ambos independientes. Si exactamente uno de los dos está en clase un día concreto, ¿cuál es la probabilidad de que sea A?

20. Cuatro estudiantes comparten piso, después de una comida deciden el siguiente procedimiento para elegir quién lava los platos: en un sombrero meten cuatro papeles doblados e indistinguibles, en

uno de ellos pone 'Sí' y en los otros tres 'No'. El primero coge un papel y lo abre, si el resultado es 'Sí' le toca fregar, si el resultado es que 'No' se libra. No se devuelve el papel al sombrero. A continuacin el segundo coge el papel, y se aplica el mismo criterio, y as sucesivamente hasta que uno de ellos obtiene el papel con el 'Sí'. ¿Quién tiene más probabilidad de lavar los platos, el primero o el último?

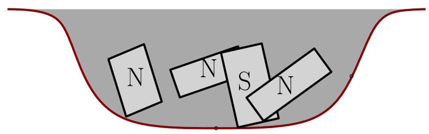


Figura 4: Sombrero con cuatro papeletas

21. Una comunidad de vecinos ha contratado un sistema de alarma para evitar robos en sus hogares. En caso de robo la alarma se pone en funcionamiento con seguridad. La probabilidad de que se produzca un robo en el vecindario es 0.001. Existen además otras causas (viento, golpes bruscos, ...) que ponen en funcionamiento el sistema de alarma, la probabilidad de que en una noche se produzca una falsa alarma es 0.01. Si se declara una señal de alarma, ¿cuál es la probabilidad de que sea falsa?

22. Una pareja que espera un hijo está preocupada porque un test practicado al feto ha dado positivo a una enfermedad muy rara que solo la padecen uno de cada 10000 individuos. Sin embargo, el test es bastante seguro: de acuerdo con el laboratorio acierta el 99 por ciento de los casos, tanto para los bebés con la enfermedad como para los sanos. ¿Cuál es la probabilidad de que el feto tenga la enfermedad teniendo en cuenta que el resultado del test ha sido positivo?

23. La probabilidad de que un componente se averíe en un período de tiempo dado es 0.01. Su estado (averiado o funcionando) se comprueba con un ensayo que cumple que cuando el componente funciona la probabilidad de que el ensayo diga lo contrario es 0.05, pero si el componente está averiado el ensayo no se equivoca. Si el ensayo indica que el componente está averiado, ¿cuál es la probabilidad de que realmente lo esté?

24. Un concursante debe elegir entre tres puertas, detrás de una de las cuales se encuentra un magnífico regalo. Hecha la elección, el presentador que sabe donde se encuentra el premio le abre una de las dos puertas no escogidas donde (lógicamente) no está el premio y a continuación le da al concursante la posibilidad de cambiar la puerta elegida. ¿Qué debe hacer el concursante?

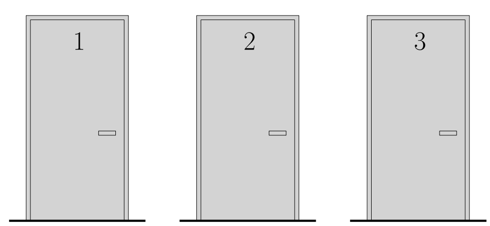
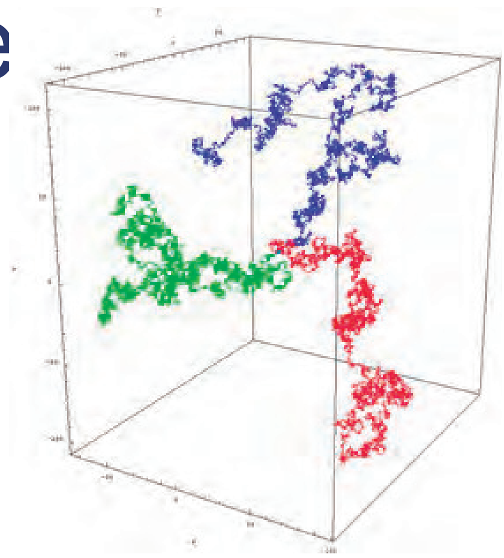


Figura 5: El concursante debe elegir una de las tres puertas

Estadística

Parte I.

Tema 3. Variable Aleatoria



Variable Aleatoria

Una **variable aleatoria** es una función que asigna un número real a cada uno de los resultados de un experimento aleatorio.

Lanzamiento de 2 monedas

$\boxed{X(s) \equiv \text{Número de CARAS}}$

s	$X(s)$
CC $\rightarrow$	2
CX $\rightarrow$	1
XC $\rightarrow$	1
XX $\rightarrow$	0

Variable Aleatoria Discreta

Cuando los valores que toma una variable aleatoria son *finitos o infinitos numerables* se dice que es **discreta**.

- Resultado obtenido al lanzar un dado

  $\{1,2,3,4,5,6\}$

- Número de veces que hay que lanzar una moneda hasta obtener una CARA

  $\{1,2,3,4, \ldots\}$

Distribución de probabilidad

Sea $\{ x_1 , x_2 , \ldots, x_n \}$ los valores que puede tomar la variable aleatoria X. Se denomina **distribución de probabilidad** de la variable aleatoria a $P(X=x_i)$ que cumple:

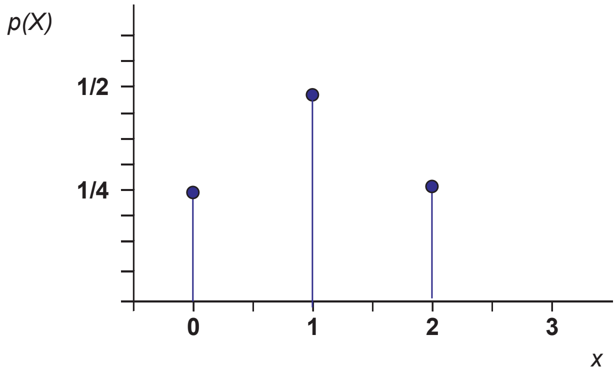
- $P (X = x_i) \geq 0$
- $\Sigma_{i=1} P (X = x_i) = 1.$

Nº de Caras al lanzar 2 monedas

x		$P(X=x)$
0	$\rightarrow$	1/4
1	$\rightarrow$	1/2
2	$\rightarrow$	1/4

Distribución de probabilidad

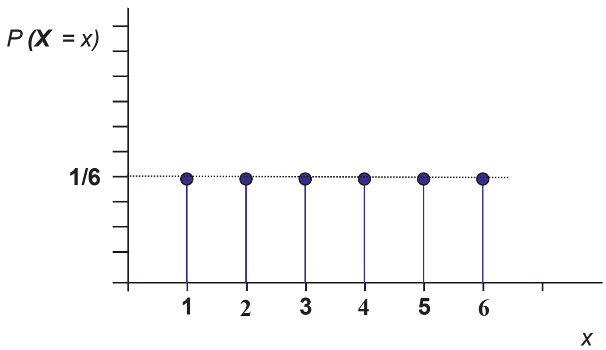
$$P(X=x)=\begin{cases}1/4 & x=0\\ 1/2 & x=1\\ 1/4 & x=2\end{cases}$$



Nº de Caras al lanzar 2 monedas

Lanzamiento de un dado

x	$P(X=x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6



$$P(X=x)=\frac{1}{6}, \qquad x=1,2,3,4,5,6$$

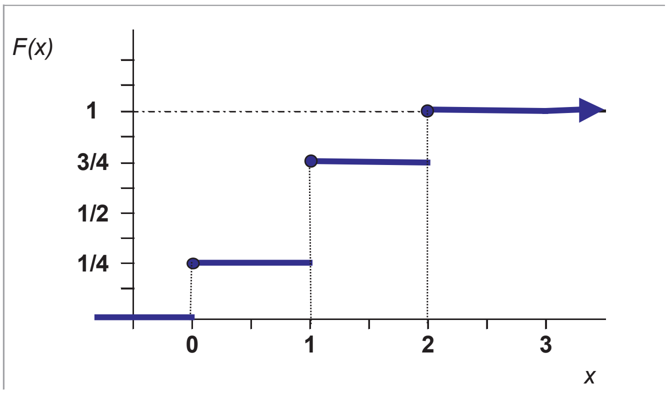
Función de distribución

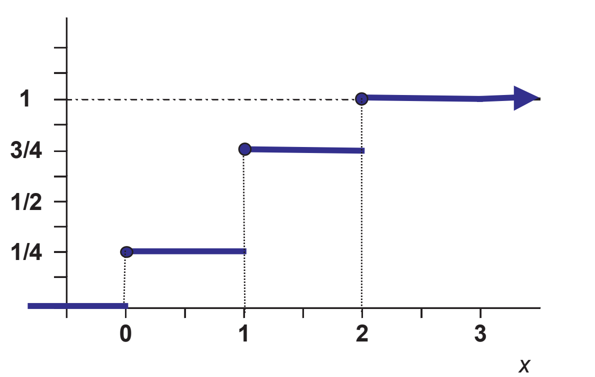
La función de distribución $F_X(x)$ de una variable aleatoria X se define para todo número real x como:

$$F_X(x) = P_X(X \leq x).$$

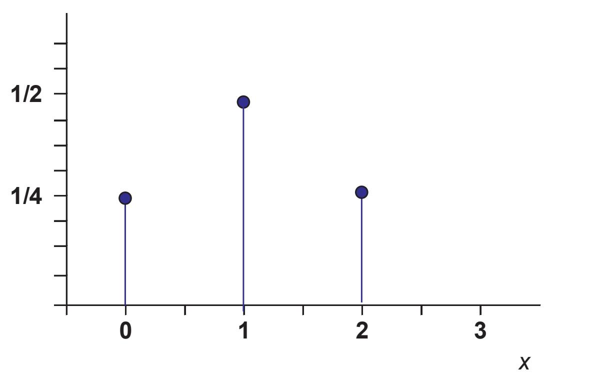
Ejemplo. X = Número de caras al lanzar 2 monedas

x	$F_X(x)$
$(-\infty,0)$	0
$[0,1)$	1/4
$[1,2)$	3/4
$[2,\infty)$	1



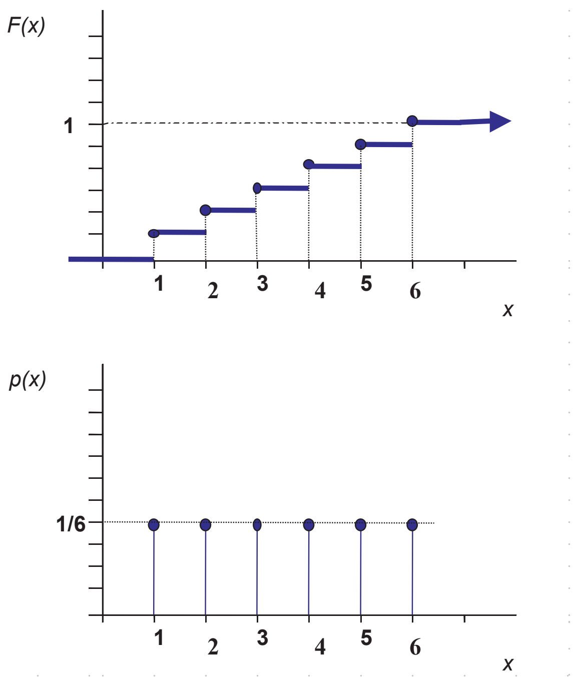


Función de Distribución



Distribución puntual de probabilidad

Lanzamiento de un dado



x	$P(X = x)$
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$

Función de Distribución *F(x)*

Una función F(x) es una **función de distribución** si y sólo si cumple las siguientes condiciones:

a. $\lim_{x \to -\infty} F(x) = 0$ y $\lim_{x \to +\infty} F(x) = 1.$

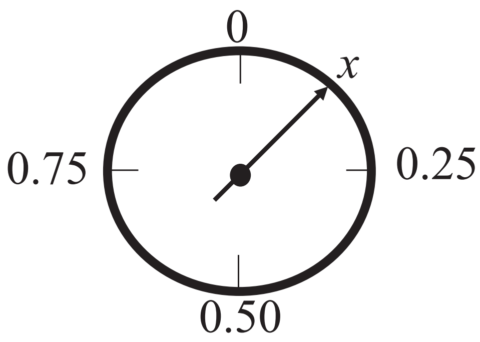
b. $F(x)$ es una función no decreciente.

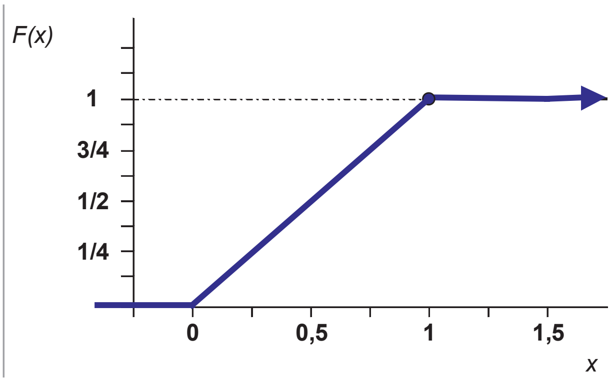
c. $F(x)$ es continua por la derecha:

$$\forall h > 0, \quad \lim_{h \to 0} F(x+h) = F(x).$$

Variable aleatoria continua

Una variable aleatoria X es continua si su función de distribución $F_X(x)$ es continua.





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

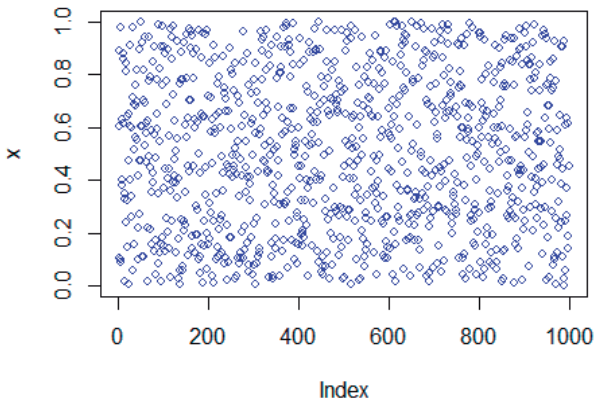
Generación números aleatorios

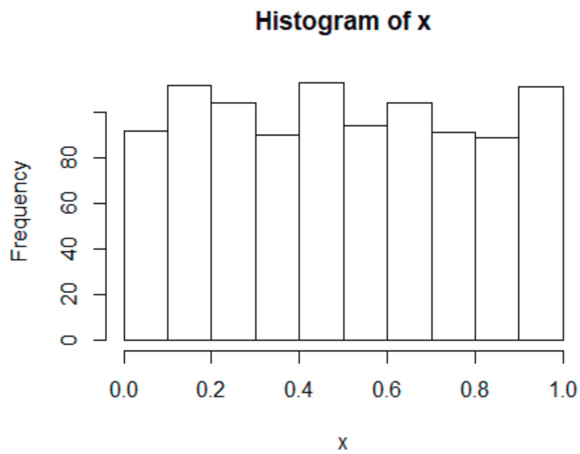
Generar valores al azar de la distribución uniforme

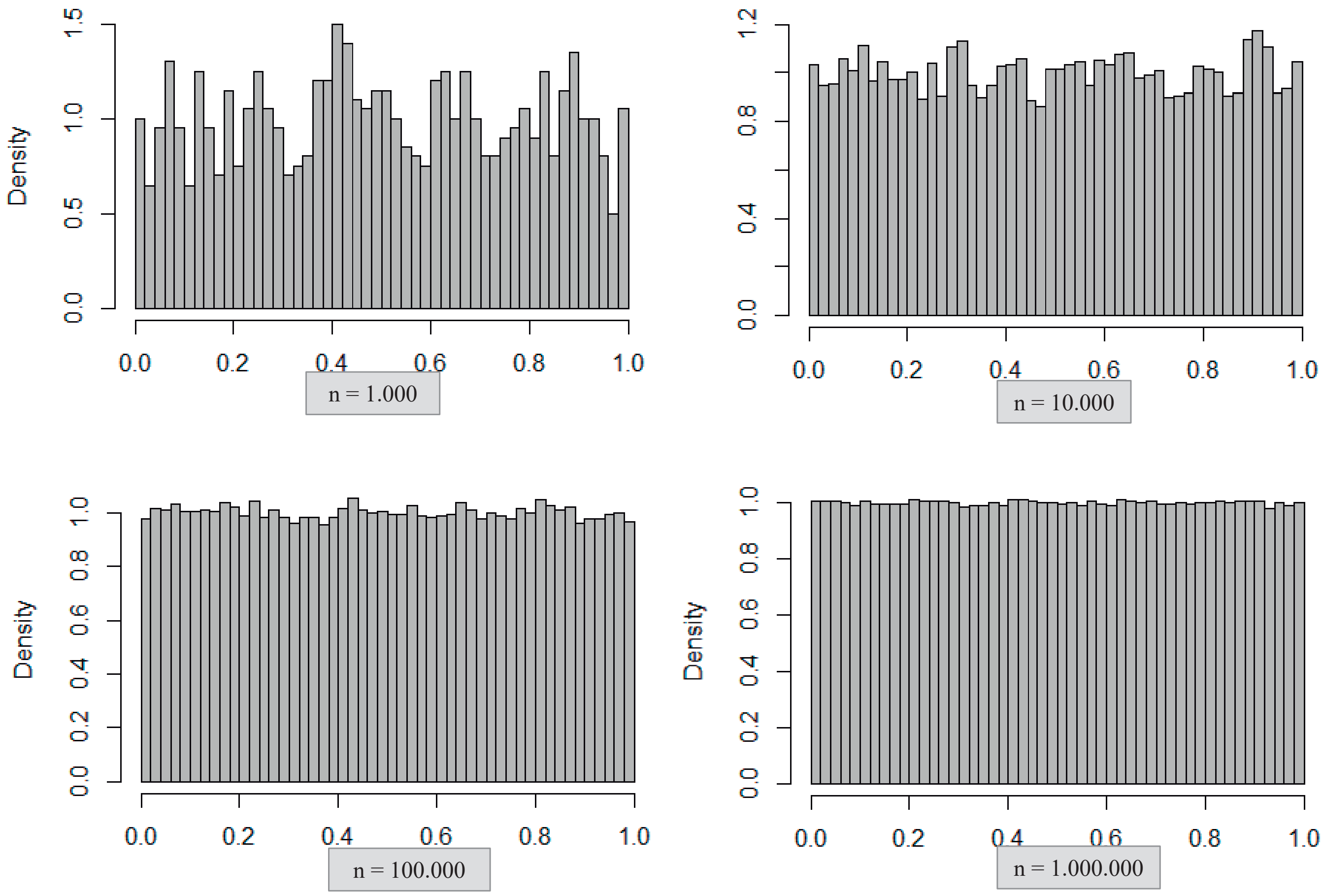
```
runif(1) #genera un valor al azar en el intervalo [0,1] con dist. uniforme

## [1] 0.5498192

runif(25) # genera 25 valores independientes con dist. uniforme en [0,1]

##  [1] 0.13841597 0.24418695 0.71302890 0.83867345 0.55214026 0.50127997
##  [7] 0.13076084 0.12667461 0.97046547 0.98281857 0.84036605 0.60393346
## [13] 0.90280394 0.30585911 0.89637537 0.33943981 0.62345635 0.87002649
## [19] 0.52071798 0.09812817 0.42877694 0.59504927 0.52124912 0.71727385
## [25] 0.34428384
```

Instrucciones en R

```
x = runif(1000) # crea un vector con 1000 valores con dist. uniforme en [0,1]
plot(x,cex=.8,col='blue') #dibuja en orden los 1000 valores de x en color azul (cex controla el tamaño de cada punto)
hist(x) # histograma de los 1000 valores
```








```
hist(x,probability = TRUE, nclass=50,col='grey')
```

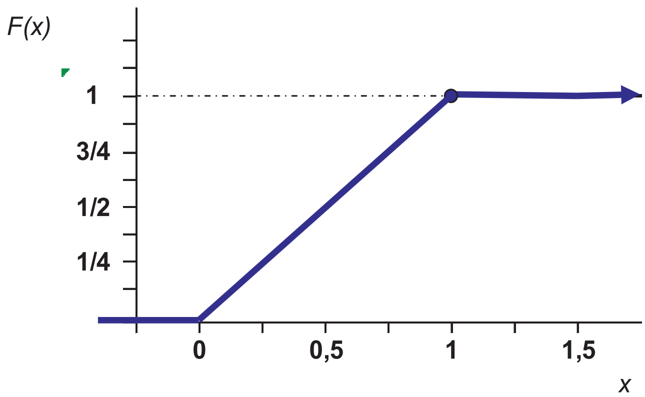
Función de densidad

La **función de densidad** de probabilidad $f_X(x)$ de una variable aleatoria continua X es la función que verifica

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \quad \forall\, x.$$

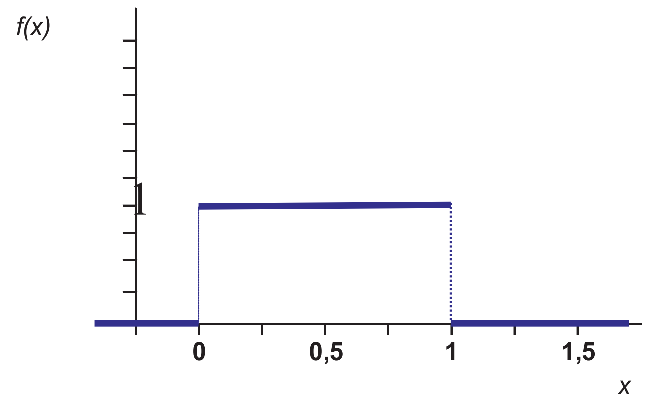
Si $F_X(x)$ es derivable, además

$$\frac{d}{dx} F_X(x) = f_X(x).$$



Función de distribución

$F_X(x) = x, \ \ x \in [0,1]$



Función de densidad

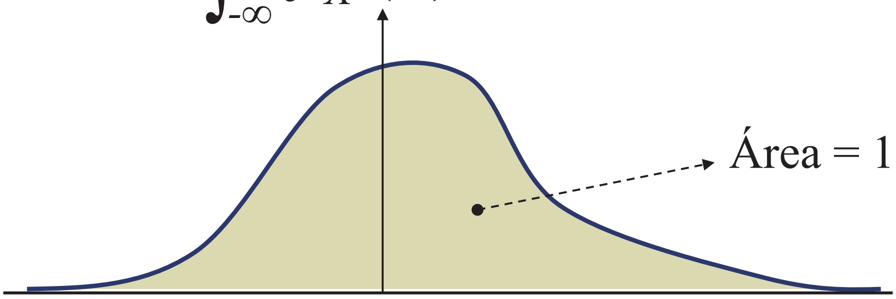
$f_X(x) = 1, \ \ x \in [0,1]$

Propiedades de la función de densidad

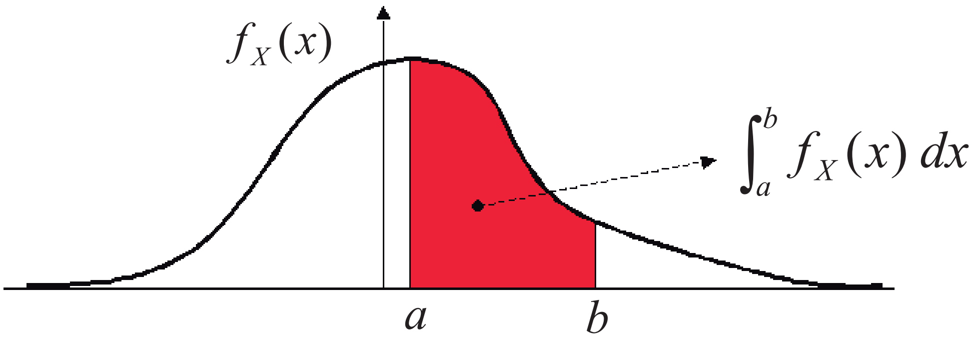
Una función $f_X(x)$ es una **función de densidad** de probabilidad de una variable aleatoria X si y sólo si cumple:

a. $f_X(x) \geq 0$ para todo x.

b. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1.$



Cálculo de probabilidades



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

$$= \int_a^b f_X(x)\, dx$$

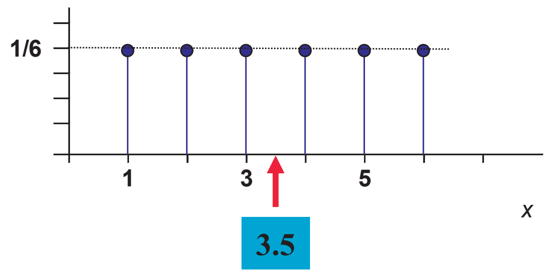
Esperanza de una variable aleatoria

Se define **esperanza** o **media** de una variable aleatoria discreta X y se representa por $E[X]$ al valor

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i).$$

Ejemplo: Lanzamiento de un dado

$$E[X] = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = 3.5$$



Centro de la distribución de probabilidad

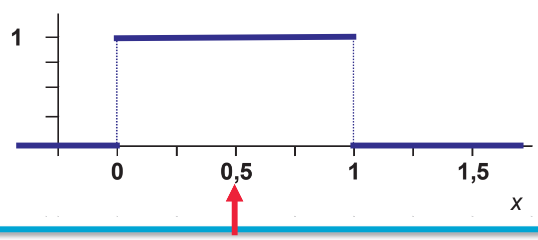
Esperanza de una variable aleatoria

Se define **esperanza** o **media** de una variable aleatoria continua X con función de densidad $f_X(x)$ y se representa por $E[X]$ al valor

$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

Ejemplo: Distribución uniforme $f_X(x) = 1,\ 0 \le x \le 1$

$$E[X] = \int_0^1 x \times 1\, dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}.$$



Centro de la distribución de probabilidad

Propiedades de E[X]

Transformaciones lineales Y = a X+b (a y b constantes)

$$E[aX+b]=a\,E[\,X\,]+b$$

Demostración:

$$E[Y]=\sum_{i=1}^{n}(ax_i+b)P(X=x_i)=a\underbrace{\sum_{i=1}^{n}x_iP(X=x_i)}_{\mathrm{E[X]}}+b\underbrace{\sum_{i=1}^{n}P(X=x_i)}_{1}$$

Varianza

Sea *X* una variable aleatoria con media μ, se denomina **varianza** a

Var (*X*) = E[(*X* - μ)^2].

- Variable aleatoria discreta

$$Var[X]=\sum_{x=-\infty}^{\infty}(x-\mu)^2P(X=x).$$

- Variable aleatoria continua

$$Var[X]=\int_{-\infty}^{\infty}(x-\mu)^2 f_X(x)\,dx$$

Propiedades de la varianza

1. $Var(X) = E[(X-\mu)^2] = E[X^2] - \mu^2.$

2. $Var(aX+b) = a^2 Var(X)$

Ejemplos

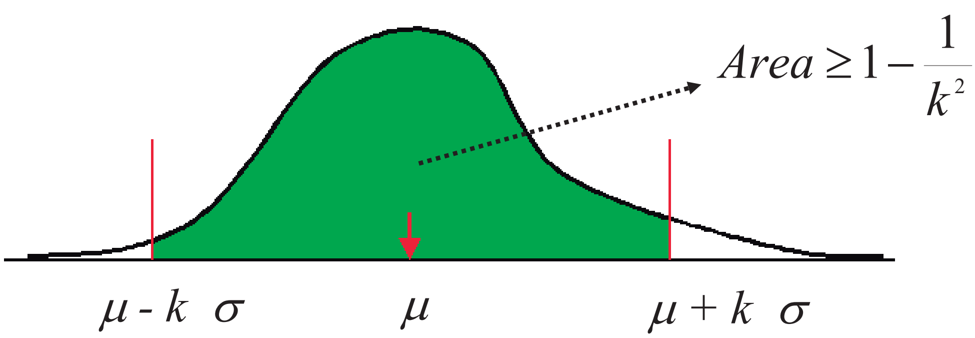
Lanzamiento de un dado

$$Var[X] = (1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6}) - (3.5)^2 = \frac{35}{12}.$$

Distribución uniforme

$$\mathrm{Var[X]} = \int_0^1 x^2 \times 1\,dx - (1/2)^2 = \left.\frac{x^3}{3}\right|_0^1 - \frac{1}{4} = \frac{1}{12}.$$

Desigualdad de Tchebychev



Para cualquier variable aleatoria

$$\mu = E[X] \qquad \sigma^2 = Var[X]$$

$$P(|X-\mu| \leq k\sigma) > 1 - \frac{1}{k^2}.$$

Momentos de una Variable Aleatoria

Momentos respecto al Origen

$$\mu_1 = E[X] = \mu$$

$$\mu_2 = E[X^2]$$

$$\mu_p = E[X^p]$$

Momentos respecto a la media

$$\alpha_1 = E[(X-\mu)] = 0$$

$$\alpha_2 = E[(X-\mu)^2] = \sigma^2$$

$$\alpha_p = E[(X-\mu)^p]$$

Transformaciones

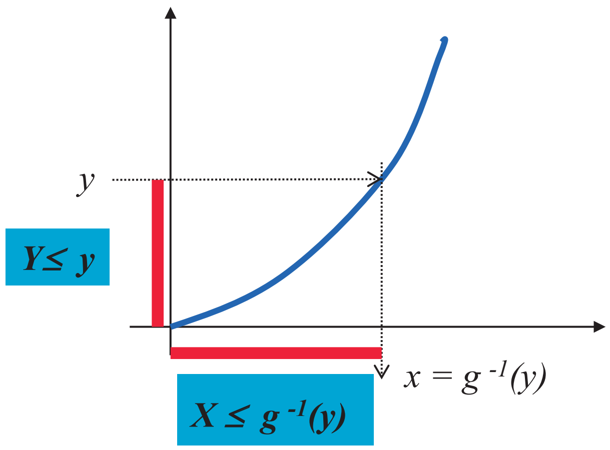
Dada una variable aleatoria X con función de densidad $f_X(x)$ vamos a ver como se obtiene la función de densidad $f_Y(y)$ de la variable aleatoria Y, definida como

$$Y = g(X)$$

Casos a considerar:

* La función g es monótona creciente
* La función g es monótona decreciente
* La función g no es monótona

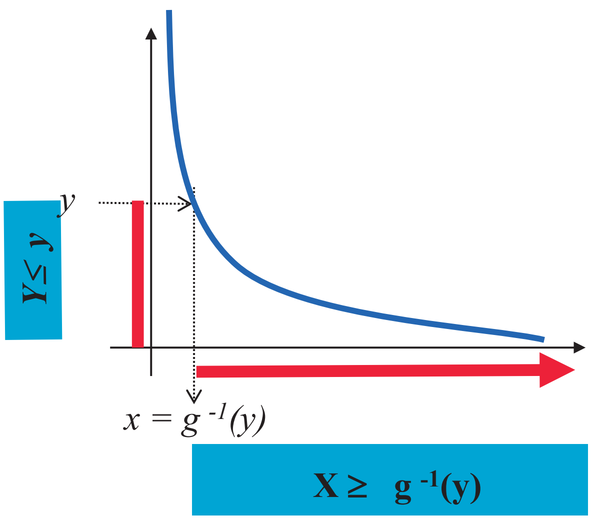
Función *g* creciente



$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(g(X) \le y) \\
&= P(\, X \le g^{-1}(y)\,) \\
&= F_X(g^{-1}(y))
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} \\
&= \frac{d\, g^{-1}(y)}{dy} \times f_X(g^{-1}(y))
\end{aligned}
}
$$

Función *g* decreciente



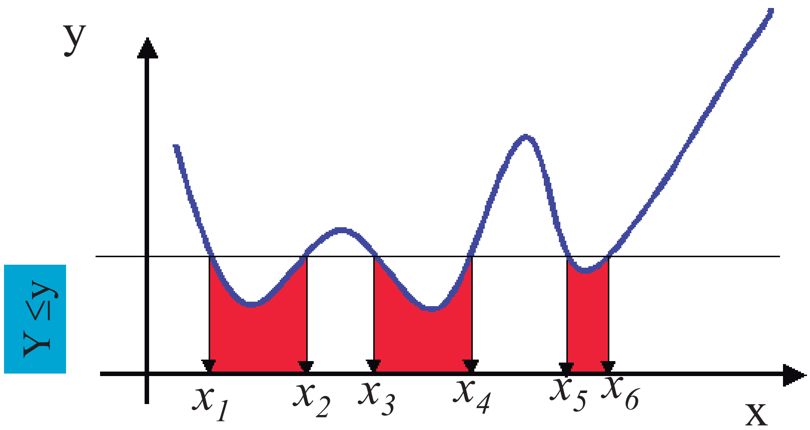
$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(g(X) \le y) \\
&= P(\, X \ge g^{-1}(y)\,) \\
&= 1 - F_X(g^{-1}(y))
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} \\
&= -\frac{d\, g^{-1}(y)}{dy} \times f_X(g^{-1}(y))
\end{aligned}
}
$$

Función monótona

$$f_Y(y) = \left| \frac{d\, g^{-1}(y)}{dy} \right| \times f_X(g^{-1}(y))$$

Transformación no monótona



$$F_Y(y) = P(Y \le y)$$
$$= P(x_1 \le X \le x_2) + P(x_3 \le X \le x_4) + P(x_5 \le X \le x_6)$$

Ejemplo de transformación

El radio de una esfera es una variable aleatoria cuya función de densidad es

$$f_X(x) = 3x^2, \quad 0 \le x \le 1.$$

¿Cuál es la función de densidad del volumen?

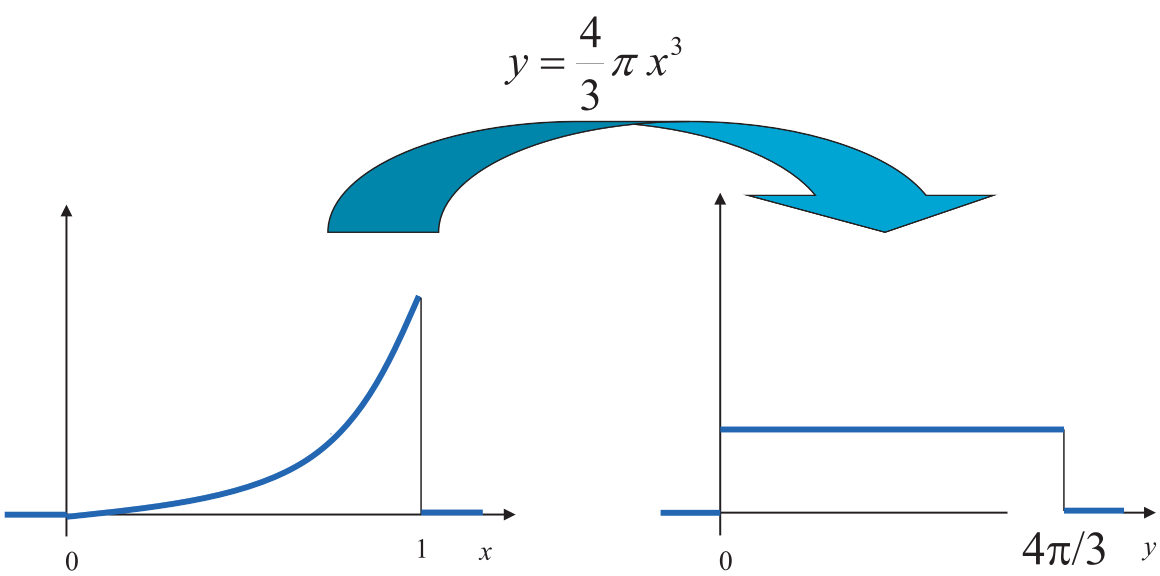
$Y =$ Volumen de la esfera

$$Y = \frac{4}{3}\pi X^3 \Rightarrow X = \left(\frac{3Y}{4\pi}\right)^{\frac{1}{3}}$$

$$g(x) = \frac{4}{3}\pi x^3; \Rightarrow g^{-1}(y) = \left(\frac{3y}{4\pi}\right)^{\frac{1}{3}};$$

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| \times f_X(g^{-1}(y))$$

$$= \frac{3}{4\pi}, \quad 0 \le y \le \frac{4\pi}{3}.$$



$$f_X(x) = 3x^2, \quad 0 \le x \le 1.$$

$$f_Y(y) = \frac{3}{4\pi}, \quad 0 \le y \le \frac{4\pi}{3}.$$

Un caso muy relevante del problema inverso

Dada una variable aleatoria *X* con distribución uniforme *(0,1)*, encontrar la transformación *Y = g(X)* que hace que Y tenga la función de densidad $f_Y(y)$

$$\frac{d\,g^{-1}(y)}{dy} = f_Y(y)$$

$$g(x) = F_y^{-1}(x)$$

Representación gráfica con ejemplo

Generar valores al azar con función de densidad

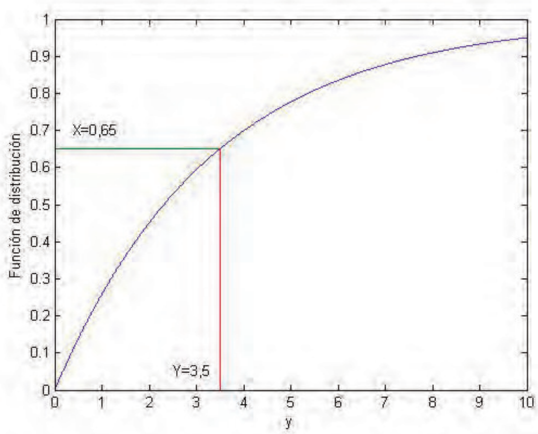
$$f_Y(y) = 0.3\,e^{-0.3\,y}, \quad y \geq 0$$

$$F_Y(y) = 1 - e^{-0.3\,y}$$

$$g(x) = F_Y^{-1}(x) = -\frac{1}{0.3}\log(1-x)$$

Generamos valores U[0,1]

$$x = 0.65 \rightarrow y = g(0.65) = 3.5$$



Distribución conjunta de variables aleatorias

Definiciones (v. a. discretas)

- Distribución de probabilidad conjunta de dos variables aleatorias X, Y

$$P(X = x, Y = y) \quad \begin{cases} P(X = x, Y = y) \geq 0, \quad \forall x, y \\ \sum_{x=-\infty}^{x=\infty} \sum_{y=-\infty}^{y=\infty} P(X = x, Y = y) = 1. \end{cases}$$

- Función de distribución conjunta:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

Lanzamiento de dos dados

Dado ROJO

Dado AZUL

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

X = " Resultado de dado ROJO"

Y = " Resultado de dado AZUL"

Distribución conjunta de probabilidad

P (X=i , Y=j) =1/36, (i,j de 1 a 6)

Ejemplo

Ejemplo: Se lanzan dos dados y se definen las variables aleatorias suma (S) y valor absoluto de la diferencia (D) de los resultados.

Distribución Conjunta P(S= x, D= y)

		S : SUMA DE DOS DADOS										
		2	3	4	5	6	7	8	9	10	11	12
	0	1/36		1/36		1/36		1/36		1/36		1/36
	1		1/18		1/18		1/18		1/18		1/18	
D : DIFERENCIA	2			1/18		1/18		1/18		1/18		
DE DOS DADOS	3				1/18		1/18		1/18			
	4					1/18		1/18				
	5						1/18					

Distribuciones Marginales

						S: SUMA DE DOS DADOS						
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	0	1/36		1/36		1/36		1/36		1/36		1/36	**6/36**
	1		1/18		1/18		1/18		1/18		1/18		**10/36**
***D*: DIFERENCIA**	**2**			1/18		1/18		1/18		1/18			**8/36**
DE DOS DADOS	**3**				1/18		1/18		1/18				**6/36**
	4					1/18		1/18					**4/36**
	5						1/18						**2/36**
		1/36	**2/36**	**3/36**	**4/36**	**5/36**	**6/36**	**5/36**	**4/36**	**3/36**	**2/36**	**1/36**	

Marginal de *S*

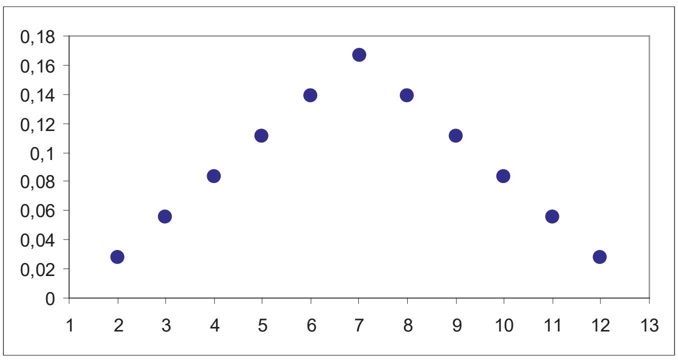
Marginal de *D*

Distribuciones Marginales

$$P(X=x)=\sum_{y=-\infty}^{y=\infty} P(X=x,Y=y)$$

$$P(Y=y)=\sum_{x=-\infty}^{x=\infty} P(X=x,Y=y)$$

Suma de dos dados



Distribuciones condicionadas

					***S*: SUMA DE DOS DADOS**							
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	0	1/36		1/36		1/36		**1/36**		1/36		1/36	6/36
	1		1/18		1/18		1/18		1/18		1/18		10/36
***D*: DIFERENCIA**	**2**			1/18		1/18		**1/18**		1/18			8/36
DE DOS DADOS	**3**				1/18		1/18		1/18				6/36
	4					1/18		**1/18**					4/36
	5						1/18						2/36
		1/36	2/36	3/36	4/36	5/36	6/36	**5/36**	4/36	3/36	2/36	1/36	

Distribución de la diferencia entre los dados condicionada a que la suma es 8.

P(D = y | S=8) = P(D=y, S=8) / P(S=8)

***D* | *S* = 8**	
0	**1/5**
1	**0**
2	**2/5**
3	**0**
4	**2/5**
5	**0**
	1

Independencia

Las variables aleatorias X, Y son independientes si y sólo si

$$P(\boldsymbol{X} = x, \boldsymbol{Y} = y) = P(\boldsymbol{X} = x) \times P(\boldsymbol{Y} = y)$$

Dado ROJO

Dado AZUL

	1	**2**	**3**	**4**	**5**	**6**
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

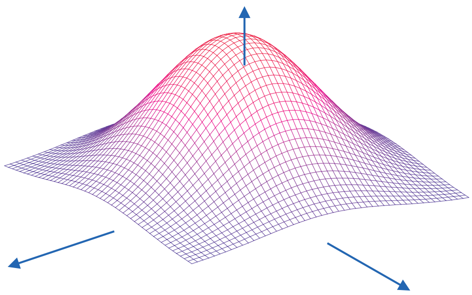
Variables aleatorias continuas

$$P(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f_{XY}(x,y)\,dx\,dy$$

Siendo $f_{XY}(x,y)$ la función de densidad conjunta, que cumple:

1) $f_{XY}(x,y) \ge 0, \quad \forall x,y$

2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$



Variables aleatorias continuas

- Función de distribución

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv$$

- Funciones de densidad marginales

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy$$

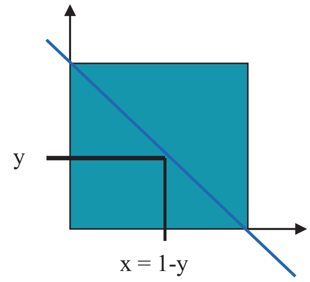
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx$$

Las variables aleatorias X, Y tienen como función de densidad conjunta

$$f_{XY}(x,y) = 6xy^2, \quad 0 < x < 1, 0 < y < 1$$

1. $P(X \le x, Y \le y) = \int_0^x \int_0^y 6uv^2 du dv = x^2 y^3$
2. $P(X + Y \le 1) = \iint_{x+y\le 1} 6xy^2 dxdy$

$$= \int_0^1 \int_0^{1-y} 6xy^2 dxdy = \frac{1}{5}$$



3. Marginales

$$f_X(x) = \int_0^1 6xy^2 dy = 2x, \quad 0 < x < 1$$

$$f_Y(y) = \int_0^1 6xy^2 dx = 3y^2, \quad 0 < y < 1$$

Independencia

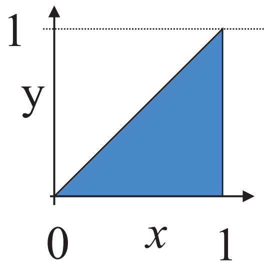
Las variables aleatorias X, Y son independientes si y sólo si

$$f_{XY}(x,y) = f_X(x) f_Y(y).$$

1. $f_{XY}(x,y) = 6xy^2, \quad 0 < x < 1, 0 < y < 1 \Rightarrow \begin{cases} f_X(x) = 2x, 0 < x < 1 \\ f_Y(y) = 3y^2, 0 < y < 1 \end{cases}$

Independientes

2. $f_{XY}(x,y) = \frac{1}{x}, \quad 0 \le y \le x \le 1 \quad \Rightarrow \begin{cases} f_X(x) = \int_0^x \frac{1}{x} dy = 1, \quad 0 < x < 1 \\ f_Y(y) = \int_y^1 \frac{1}{x} dx = -\log(y), \quad 0 < y < 1 \end{cases}$



No independientes

Funciones de densidad condicionadas

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad cuando\ f_Y(y) > 0$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad cuando\ f_X(x) > 0$$

Independencia -II

1. $f_{XY}(x,y) = 6xy^2, \quad 0<x<1, 0<y<1 \Rightarrow \begin{cases} f_X(x) = 2x, 0<x<1 \\ f_Y(y) = 3y^2, 0<y<1 \end{cases}$ Independientes

$$f_{X|Y}(x\,|\,y) = \frac{6xy^2}{3x^2} = 2x, \quad 0<x<1$$

2. $f_{XY}(x,y) = \frac{1}{x}, \quad 0 \le y \le x \le 1 \quad \Rightarrow \begin{cases} f_X(x) = \int_0^x \frac{1}{x} dy = 1, \quad 0<x<1 \\ f_Y(y) = \int_y^1 \frac{1}{x} dx = -\log(y), \quad 0<y<1 \end{cases}$

$$f_{X|Y}(x\,|\,y) = -\frac{1}{x\log(y)}, \quad y \le x \le 1$$

$$f_{Y|X}(y\,|\,x) = \frac{1}{x}, \quad 0 \le y \le x$$

No independientes

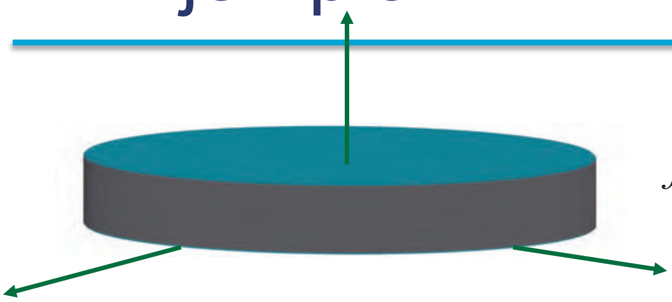
Independencia -III

Las variables aleatorias $X,\ Y$ son independientes si y sólo si

$$f_{X|Y}(x \mid y) = f_X(x)$$

$$f_{Y|X}(y \mid x) = f_Y(y)$$

Ejemplo



$$f(x,y) = \begin{cases} c, & x^2 + y^2 \le r^2 \\ 0, & x^2 + y^2 > r^2 \end{cases}$$

1. $c = \dfrac{1}{\pi r^2}.$

2. $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

$$= \frac{1}{\pi r^2} \int_{-\sqrt{r^2-x^2}}^{+\sqrt{r^2-x^2}} dy$$

$$= \frac{2}{\pi r^2}\sqrt{r^2 - x^2}, \quad -r \le x \le r$$

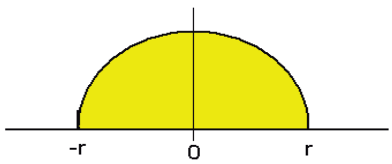
Ejemplo (cont.)

$$3. f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

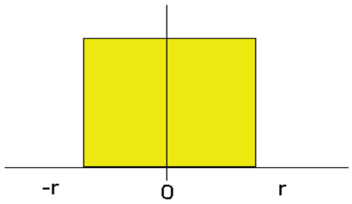
$$= \frac{1}{\pi r^2}\int_{-\sqrt{r^2-y^2}}^{+\sqrt{r^2-y^2}} dx$$

$$= \frac{2}{\pi r^2}\sqrt{r^2-y^2}, \quad -r \le y \le r$$

$$4. f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{r^2-y^2}}, \quad -\sqrt{r^2-y^2} \le x \le \sqrt{r^2-y^2}$$



$f_Y(y)$



$f_{X|Y}(x|y)$

Independencia

Las variables aleatorias X, Y son independientes si y sólo si

$$f_{XY}(x,y) = f_X(x)f_Y(y).$$

$$1. \quad f_{XY}(x,y) = 6xy^2, \quad 0<x<1, 0<y<1 \Rightarrow \begin{cases} f_X(x) = 2x, 0<x<1 \\ f_Y(y) = 3y^2, 0<y<1 \end{cases}$$

Independientes

$$2. \quad f(x,y) = \begin{cases} \frac{1}{\pi r^2}, & x^2+y^2 \le r^2 \\ 0, & x^2+y^2 > r^2 \end{cases} \Rightarrow \begin{cases} f_X(x) = \frac{2}{\pi r^2}\sqrt{r^2-x^2} & ,-r \le x \le r \\ f_Y(y) = \frac{2}{\pi r^2}\sqrt{r^2-y^2} & ,-r \le y \le r \end{cases}$$

No Independientes

Esperanza de g(X,Y)

(a) Si X, Y son variables aleatorias discretas, se define esperanza de la función $g(X,Y)$ como:

$$E[g(X,Y)] = \sum_{x=-\infty}^{x=\infty} \sum_{y=-\infty}^{y=\infty} g(x,y)P(X=x, Y=y)$$

(b) Si X, Y son variables aleatorias continuas, se define esperanza de la función $g(X,Y)$ como:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\,dxdy$$

Propiedades de E[g(X,Y)]

Si $g(X,Y) = g_1(X) + g_2(Y)$ se cumple

$$\begin{aligned}
E[g_1(X)+g_2(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(g_1(x)+g_2(y)\right) f_{XY}(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x) f_{XY}(x,y)\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(y)\, f_{XY}(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} g_1(x)\left(\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\right)dx + \int_{-\infty}^{\infty} g_2(y)\left(\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\right)dy \\
&= \int_{-\infty}^{\infty} g_1(x) f_X(x)\,dx + \int_{-\infty}^{\infty} g_2(y) f_Y(y)\,dy \\
&= E[g_1(X)] + E[g_2(Y)]
\end{aligned}$$

Ejemplo

$$E[X+Y] = E[X] + E[Y]$$

Covarianza

La covarianza de dos variables aleatorias X, Y, se denota por $Cov(X,Y)$ y se define como:

$$
\begin{aligned}
Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy
\end{aligned}
$$

donde $\mu_X = E[X]$ y $\mu_Y = E[Y]$.

Si las v.a's son discretas:

$$
\begin{aligned}
Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= \sum_i\sum_j(x_i-\mu_X)(y_j-\mu_Y)P(X=x_i, Y=y_j)
\end{aligned}
$$

Propiedades de la covarianza

La covarianza es una medida de la dependencia lineal entre las dos variables. Si las variables son **independientes, Cov(X,Y)=0**

$$
\begin{aligned}
Cov(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty}(x-\mu_X)f_X(x)dx\int_{-\infty}^{\infty}(y-\mu_Y)f_Y(y)dy = 0
\end{aligned}
$$

Propiedades

- $Cov(X,Y)=E[XY] - E[X]E[Y]$
- $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

Medias y Matriz de Varianzas

X, Y dos variables aleatorias con función de densidad conjunta $f(x,y)$

$$X : f_X(x) \quad \to \mu_X = E[X], \quad \sigma_X^2 = Var[X]$$

$$Y : f_Y(y) \quad \to \mu_Y = E[Y], \quad \sigma_Y^2 = Var[Y]$$

$$\sigma_{XY} = Cov(X,Y)$$

Vector aleatorio $\mathbf{U} = \begin{pmatrix} X \\ Y \end{pmatrix}$;

Vector de medias

$$E[U] = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$$

Matriz de varianzas

$$Var[U] = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$$

Correlación

Se define coeficiente de correlación ρ entre dos variables aleatorias X, Y como

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

Propiedades

- $-1 \leq \rho(X,Y) \leq +1$
- Si X e Y son independientes, entonces $\rho(X,Y) = 0$.
- $Y = a + b\,X \Leftrightarrow \rho(X,Y) = 1\ (b>0)$ o $\rho(X,Y) = -1\ (b<0)$

n variables aleatorias

Para calcular probabilidades de un suceso en el que intervienen variables aleatorias X_1, X_2, ..., X_n es preciso conocer la distribución de probabilidad conjunta de todas ellas.

Si las variables son continuas se emplea la función de densidad conjunta

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad f_X(x_1, x_2, \ldots, x_n)$$

Vector de variables aleatorias

$\mathbf{X} = (X_1, X_2, \ldots, X_n) \to$ Conjunto (vector) de n v.a. continuas

Su distribución de probabilidad viene determinada por

su función de densidad conjunta : $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$

o por la función de distribución conjunta $F_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$.

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \ldots, t_n) dt_1 dt_2 \cdots dt_n$$

Distribuciones marginales

$\mathbf{X} = (X_1, X_2, ..., X_n) \rightarrow$ Conjunto (vector) de n v.a. continuas

La función de densidad marginal de X_i, $f_i(x_i)$,

$$f_i(x_i) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n}_{\text{Todas menos } x_i}$$

La función de densidad marginal de (X_i, X_j), $f_{ij}(x_i, x_j)$,

$$f_{ij}(x_i, x_j) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n}_{\text{Todas menos } x_i \text{ y } x_j}$$

Esperanza

Esperanza de g($\mathbf{X}$) $= g(X_1, X_2, ..., X_n)$

$$E[g(\mathbf{X})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, ..., x_n) f_{\mathbf{X}}(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n$$

Es fácil demostrar:

$$E[X_i] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i f_{\mathbf{X}}(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n$$
$$= \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i$$

$$E[g(X_1, X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2) f_{\mathbf{X}}(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2) f_{12}(x_1, x_2) dx_1 dx_2$$

Vector de Medias y Matriz de Varianzas

$$\mathbf{X} = (X_1, X_2, ..., X_n)^T, \rightarrow \text{Vector de } n \text{ variables aleatorias}$$

$$E[\mathbf{X}] = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \rightarrow \begin{cases} \mu_1 = E[X_1] \\ \mu_2 = E[X_2] \\ \vdots \\ \mu_n = E[X_n] \end{cases}$$

$$Var[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix} \rightarrow \begin{cases} Var(X_i) = \sigma_i^2 \\ Cov(X_i, X_j) = \sigma_{ij} \quad i \neq j \end{cases}$$

Independencia

Las variables aleatorias $X_1, X_2, ..., X_n$ son independientes si y sólo si

$$f(x_1, x_2, ..., x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n).$$

Transformaciones Lineales

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T, \rightarrow \text{Vector de } n \text{ variables aleatorias}$$

$$\mathbf{a} = (a_1, a_2, \ldots, a_n)^T, \rightarrow \text{Vector de } n \text{ constantes}$$

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n = \mathbf{a}^{\mathrm{T}} \mathbf{X}$$

$$E[Y] = \mathbf{a}^{\mathrm{T}} E[\mathbf{X}] = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n$$

$$Var[Y] = \mathbf{a}^{\mathrm{T}}\, Var[\mathbf{X}]\mathbf{a} = \begin{pmatrix} a_1 & a_2 & \cdots a_n \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

Transformaciones Lineales Caso General

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T, \rightarrow \text{Vector de } n \text{ variables aleatorias}$$

$$\mathbf{a} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \rightarrow \text{Matriz de } m \times n \text{ constantes}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \mathbf{AX} \rightarrow E[Y] = \mathbf{A}E[\mathbf{X}]$$

$$Var[Y] = \mathbf{A}\, Var[\mathbf{X}]\mathbf{A}^T$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

Transformaciones Lineales (Independencia)

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T, \rightarrow \text{Vector de } n \text{ variables aleatorias independientes}$$

$$\mathbf{a} = (a_1, a_2, \ldots, a_n)^T, \rightarrow \text{Vector de } n \text{ constantes}$$

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

$$Var[Y] = \begin{pmatrix} a_1 & a_2 & \cdots a_n \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2$$

Ejemplo:

Calcular la media y la varianza de la suma de 12 variables aleatorias independientes con distribución uniforme en (0,1)

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_{12} \end{pmatrix}, \qquad U_i \rightarrow Uniforme(0,1) \quad \begin{cases} E[U_i] = 1/2 \\ Var[U_i] = 1/12 \\ Cov(U_i, U_j) = 0 \end{cases}$$

$$Y = U_1 + U_2 + \cdots + U_{12}$$

$$E[Y] = \sum_{i=1}^{12} E[U_i] = 6$$

$$Var[Y] = \sum_{i=1}^{12} Var[U_i] = 1$$

Ejemplo

Se dispone de *n* sobres con sus correspondientes cartas. Se extraen las cartas de los sobres, se sortean y se vuelven a introducir de forma aleatoria cada una en un sobre. ¿Cuál es el número esperado de cartas que coinciden con su sobre inicial?

$$X \equiv \text{Número de coincidencias} \qquad X = X_1 + X_2 + \cdots + X_n$$

$$X_i = \begin{cases} 0, & \text{Si la carta } i \textbf{ no} \text{ coincide con su sobre inicial} \\ 1, & \text{Si la carta } i \textbf{ sí} \text{ coincide con su sobre inicial} \end{cases}$$

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n]$$
$$= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = 1.$$

Media de *n* variables aleatorias independientes

$\mathbf{X} = (X_1, X_2, ..., X_n)^T, \rightarrow$ Vector de n variables aleatorias independientes

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \begin{cases} E[\overline{X}] = \dfrac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n} \\ Var[\overline{X}] = \dfrac{Var[X_1] + Var[X_2] + \cdots + Var[X_n]}{n^2} \end{cases}$$

Si las variables tienen la misma media y varianza

$$\mu = E[X_i] \quad \forall i \qquad\qquad \sigma^2 = Var[X_i] \quad \forall i$$

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \Rightarrow \begin{cases} E[\overline{X}] = \mu \\ Var[\overline{X}] = \dfrac{\sigma^2}{n} \end{cases}$$

Capítulo 3: Variable Aleatoria

1. Dada la variable aleatoria discreta X, cuya función de probabilidad viene definida por

$$P(X = x) = kx, \quad x = 1, 2, ..., 5$$

   a) Calcular el valor de la constante k
   b) Calcular $P(X > 2)$
   c) Calcular $E[X]$ y $Var[X]$.
   d) Calcular $E[Y]$ si $Y = 2X + 5$

2. Dada la variable aleatoria X, cuya función de densidad es

$$f(x) = \begin{cases} k(1 - x^2), & \text{si } 0 < x < 1 \\ 0, & \text{en el resto} \end{cases}$$

   a) Obtener k.
   b) Dibujar la función de densidad.
   c) Calcular la probabilidad $P(X < 0.3)$.
   d) Obtener la media y la varianza de X.
   e) Obtener la media y la varianza de la variable $Y = 3X - 1$.
   f) Obtener la media y la varianza de la variable $Z = 3X^2$.

3. Un gran almacén guarda cajas que contienen piezas de distinto tipo. La proporción X de piezas de tipo A en una caja se puede considerar una variable aleatoria con función de densidad:

$$f(x) = kx(1 - x) \text{ con } 0 \leq x \leq 1$$

   a) Calcular el valor de k, la media y la varianza de la variable aleatoria X.
   b) Si se toman 10 cajas al azar.¿Cuál es la probabilidad de que ninguna de ellas contenga una proporción de piezas de tipo A igual o superior al 75 % ?

4. Un modelo que habitualmente se utiliza en balística para comprobar la correcta calibración de las armas es

$$f(x) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right], \qquad x \geq 0, \sigma \geq 0,$$

   donde la variable aleatoria X es la distancia del punto de impacto del proyectil al centro del blanco al que iba dirigido y σ es el parámetro que mide la precisión. Si para una distancia determinada de disparo la precisión del arma es $\sigma = 10$ cm,

   a) ¿Cuál es la probabilidad de que un impacto esté a una distancia menor o igual de 5 cm del centro?
   b) Calcular la función de distribución.
   c) ¿Cuál es la probabilidad de que al lanzar 10 proyectiles, ninguno haya impactado a una distacia menor de 5 cm del centro del blanco?

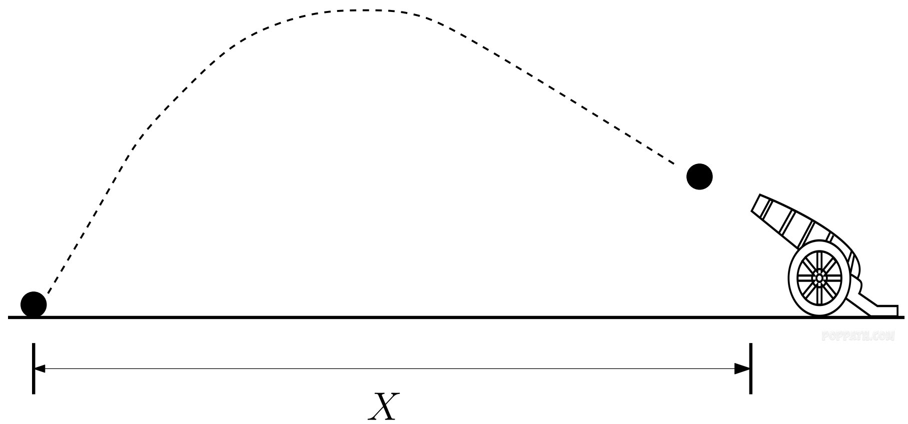


Figura 1: Disparo de un cañón

d) ¿Cuál es la probabilidad de que al lanzar 10 proyectiles, todos hayan impactado a una distacia menor de 5 cm del centro del blanco?

5. Una variable aleatoria tiene distribución uniforme en el intervalo $[-a, a]$. Calcular la media, la varianza, el coeficiente de asimetría y el coeficiente de apuntamiento (o curtosis).

6. Si X es una variable aleatoria con media μ. Demostrar que cuando $m = \mu$, $E[(X - m)^2]$ es mínima.

7. Demostrar la desigualdad de Markov

$$P(X \geq a) \leq \frac{1}{a} E\left[X\right]$$

donde X es una variable aleatoria positiva ($P(X > 0) = 1$).

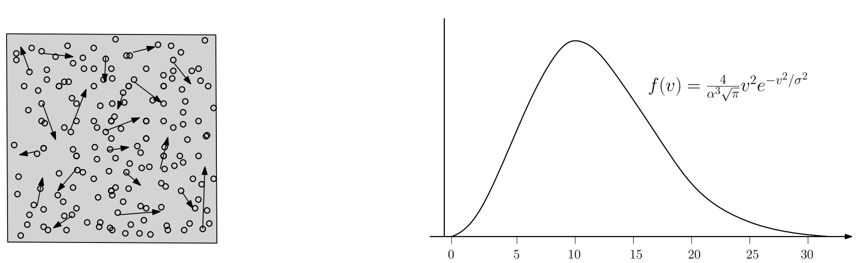


Figura 2: Distribución de Maxwell

8. De acuerdo con la teoría cinética de los gases, la velocidad V de una molécula de masa m de un gas a la temperatura (absoluta) T es una variable aleatoria con la siguiente función de densidad (distribución de Maxwell):

$$f(v) = \frac{4}{\alpha^3 \sqrt{\pi}} v^2 e^{-v^2/\alpha^2}, \quad v \geq 0$$

donde $\alpha = \sqrt{2kT/m}$, siendo k la constante de Boltzmann. Además,

$$E(V) = 2\alpha/\sqrt{\pi}, \quad \mathrm{Var}(V) = (3/2 - 4/\pi)\alpha^2$$

a) Calcular el valor medio de la energía cinética, $mV^2/2$, de una molécula. ¿ A una misma temperatura T, qué gas tiene mayor valor medio de energía cinética, uno ligero u otro más pesado?

b) Obtener la función de densidad de la energía cinética de una molécula. Indicar si depende de la masa molecular.

9. Sea X una variable aleatoria con distribución uniforme en $(0, 1)$. Calcular la probabilidad de que $Y > 0.8$ si $Y = e^{-X^2}$.

10. La función de distribución de la variable aleatoria X es $F_X(x)$. Obtener la función de densidad de la variable aleatoria $Y = F_X(X)$.

11. La variable aleatoria Z tiene como función de densidad

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad -\infty < z < \infty, \quad \sigma > 0$$

Obtener la función de densidad de $Y = |Z|$ y su media.

12. Si X es una variable aleatoria con distribución uniforme entre 0 y θ, obtener la función de densidad de la variable aleatoria $Y = +\sqrt{X}$.

13. Se elige un punto al azar interior a la circunferencia de ecuación $x^2 + y^2 = r^2$. Llamando Z a la variable aleatoria definida por la distancia entre el punto elegido y el centro de la circunferencia, calcular las funciones de densidad y distribución de Z.

14. Supóngase una diana circular con centro en el origen de coordenadas de radio r y X, Y las coordenadas de un punto elegido al azar (por ejemplo, el lanzamiento de un dardo). Supóngase que cualquier otro punto de la diana tiene la misma probabilidad de ser elegido. Calcule $f_{XY}(x, y)$, las funciones de densidad marginales y condicionadas.

15. Obtén la distribución de probabilidad del máximo, del mínimo y de la suma de los resultados obtenidos al lanzar dos dados equilibrados. Se acepta que los resultados de los dados son variables aleatorias independientes.

16. La función de densidad de una variable aleatoria bidimensional viene dada por la expresión:

$$f_{XY}(x, y) = \begin{cases} xy + ce^x, & \text{cuando } 0 < x < 1 \text{ y } 0 < y < 1 \\ 0, & \text{en el resto} \end{cases}$$

Indicar si son independientes las variables aleatorias X e Y.

17. La cantidad en miligramos de dos componentes contenidos en un producto es una variable aleatoria bidemensional, cuya función de densidad viene dada por la expresión

$$f_{XY}(x, y) = \begin{cases} 4xy, & \text{cuando } 0 \le x \le 1 \text{ y } 0 \le y \le 1 \\ 0, & \text{en el resto} \end{cases}$$

a) Obtener las funciones de densidad marginales

b) Calcular la covarianza

c) Calcular $P(X + Y \le 0.5)$

d) Calcular la probabilidad de que la cantidad del primer componentes sea menor que 0.3 miligramos cuando la del segundo es 0.8 miligramos.

18. La función de densidad de la variable aleatoria bidimensional (X, Y), bien dada por la expresión:

$$f_{XY}(x,y) = \begin{cases} kxy, & \text{cuando } 0 < x < y < 1 \\ 0, & \text{en el resto} \end{cases}$$

a) Calcular el valor de k y las distribuciones marginales.

b) Calcular $P(X < 0.5|Y = 0.5)$.

c) ¿Son independientes las variables aleatorias X e Y?

19. X e Y son variables aleatorias con coeficiente de correlación lineal $\rho = -1$. Si las varianzas son iguales, calcular la varianza de $Z = X + Y - 1$.

20. La distribución de probabilidad conjunta de las variables aleatorias Y_1 e Y_2 es la siguiente:

		Y_1		
		-1	0	1
	-1	1/16	3/16	1/16
Y_2	0	3/16	0	3/16
	1	1/16	3/16	1/16

Calcular su coeficiente de correlación e indicar si son independientes.

21. La función de densidad conjunta de X e Y viene dada por

$$f(x,y) = xy, \qquad 0 < x < 1, \;\; 0 < y < 2$$

a) Obtener las funciones de densidad marginales y decir si X e Y son independientes.

b) Calcular $\mathrm{P}(X + Y < 1)$.

c) Calcular $\mathrm{P}(X + Y < 2)$.

22. Un ordenador tarda un total de Y segundos en procesar un mensaje de correo electrónico, esta cantidad incluye el tiempo X durante el cual el mensaje está en la cola esperando a ser procesado $(Y \geq X)$. La función de densidad conjunta de las variables aleatorias X, Y es

$$f_{X,Y}(x,y) = e^{-y}, \quad 0 \leq x \leq y < \infty$$

Calcular la probabilidad de que un mensaje haya estado menos de un segundo en la cola si el tiempo total que ha durado su procesamiento ha sido mayor que dos segundos.

23. Sea X un valor elegido al azar de la distribución uniforme en el intervalo [0,1]. Se observa $X = x$ y a continuación se toma al azar otro valor Y de la distribución uniforme en el intervalo $[x, 1]$. Calcular la función de densidad marginal de Y.

24. Sean X, Y, U y V variables aleatorias, demostrar que si $Y = U + V$, entonces

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,U) + \mathrm{Cov}(X,V).$$

25. Sean dos variables aleatorias independientes X_1 y X_2 con la misma función de densidad $f(x) = e^{-x}, x \geq 0$

   a) Obtener la función de densidad conjunta.

   b) Obtener la función de densidad de $Y = X_1 + X_2$.

   c) Añadimos otra variable X_3 independiente e idénticamente distribuida que las anteriores, obtener la función de densidad de $Y_3 = X_1 + X_2 + X_3$.

   d) Dada n variables aleatorias independientes con la distribución anterior, se define $Y_n = X_1 + X_2 + \cdots + X_n$. Aplicando el procedimiento anterior de manera recurrente se obtiene que la función de densidad de Y_n es

$$f_{Y_n}(y) = \frac{y^{n-1}}{k(n)} e^{-y}, y \geq 0.$$

   Obtener $E[Y_n]$ y deducir la expresión de la función $k(n)$.

Estadística

Parte II.

Tema 4. Modelos de Probabilidad

Proceso de Bernoulli

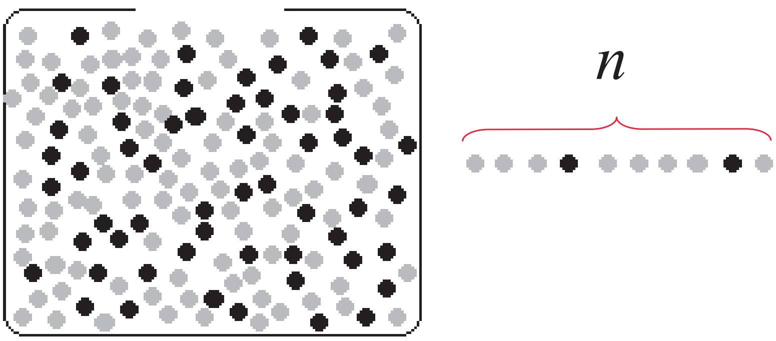
El resultado de un experimento admite dos categorías: "*Aceptable*" y "*Defectuoso*".

- Se repite el experimento *n* veces.
- La probabilidad de "*defectuoso*" es la misma *p* en todos los experimentos.
- Los experimentos son independientes.

Ejemplos de procesos de Bernoulli

- Lanzamiento de *n* monedas. Resultado: cara o cruz.
- Se extraen piezas al azar de un sistema continuo de fabricación. Se clasifican las piezas en aceptables o no.
- Lanzamiento de un dado *n* veces. En cada lanzamiento se clasifica como 6 o distinto de 6.

Distribución Binomial (*n*,*p*)



Proporción defectuosas $= p$

$X =$ " Nº de defectuosas al extraer n piezas"

Distribución binomial (n=10,p)

□□□□□□□□□□ $\longrightarrow$ $p^4(1-p)^6$

□□□□□□□□□□
□□□□□□□□□□
↓
□□□□□□□□□□

$\longrightarrow \binom{10}{4}$

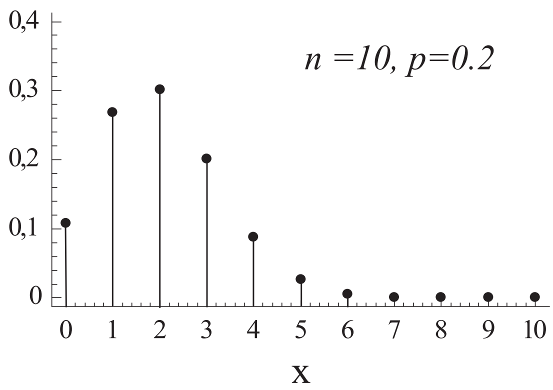
$$P(X=4)=\binom{10}{4}p^4(1-p)^6$$

Distribución binomial (n,p)

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,2,\dots,n$$

(1) $P(X=k)\geq 0,\ \forall\ k.$

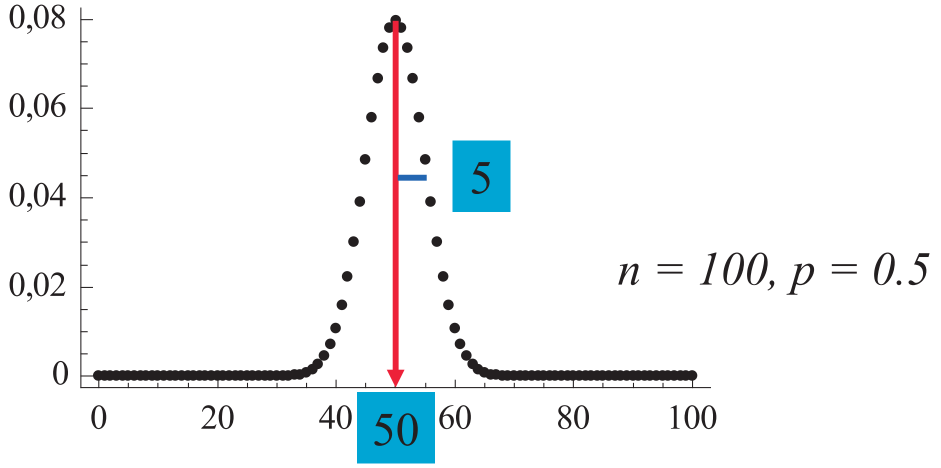
(2) $\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=1.$



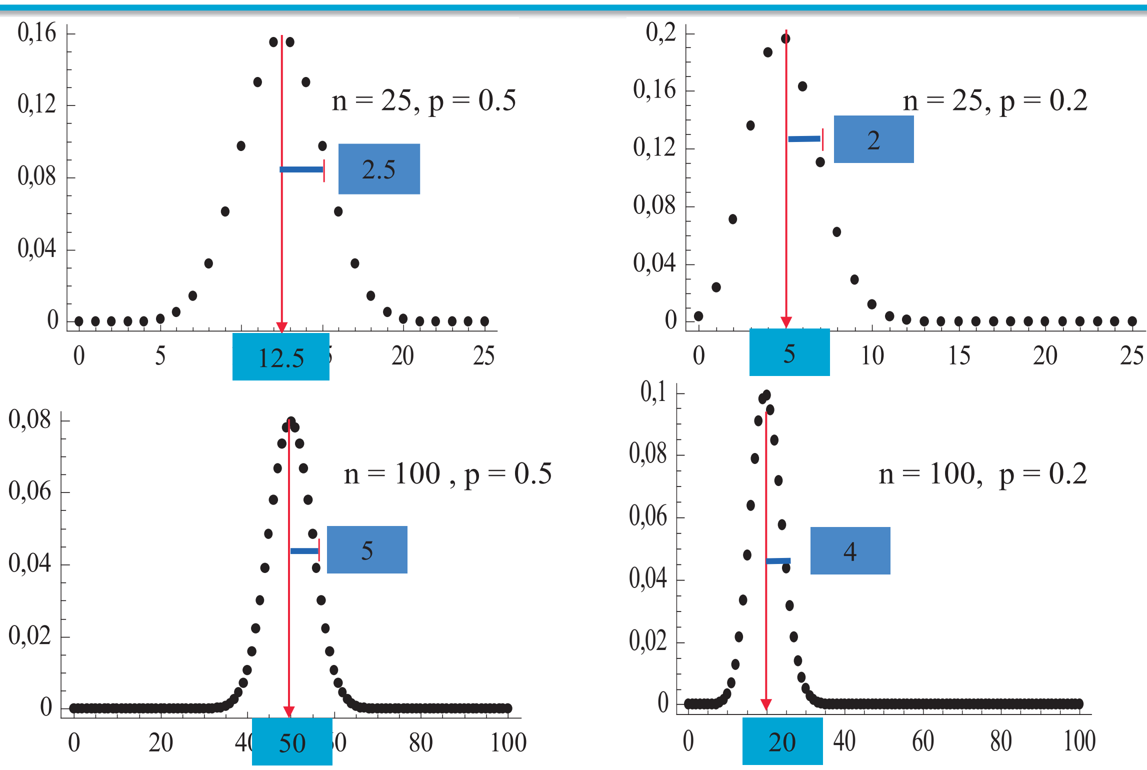
Propiedades de la dist. binomial

$$E[X] = \sum_{k=0}^{n} k \times P(X = k) = np.$$

$$Var[X] = E[X^2] - E[X]^2 = np(1-p)$$



Distribuciones binomiales



Ejemplo

Un contrato estipula la compra de componentes en lotes grandes que deben contener un máximo de 10% de piezas con algún defecto. Para comprobar la calidad se toman 11 unidades y se acepta el lote si hay como máximo 2 piezas defectuosas. ¿Es un buen procedimiento de control?

Sea p la proporción de piezas en un lote,

$$\text{X} \equiv \text{Número de defectuosas en la muestra}$$

$$\text{P(Aceptar)} = \text{P}(\text{X} \leq 2)$$

$$= \binom{11}{0} p^0 (1-p)^{11} + \binom{11}{1} p^1 (1-p)^{10} + \binom{11}{2} p^2 (1-p)^9$$

p	5%	10%	15%	20%	25%
$P(Aceptar)$	0.985	0.910	0.779	0.617	0.45

Distribución binomial

```
P1 = dbinom(27,size=50,prob=0.5) # o P1=dbinom(10,50,.5)
P2 = pbinom(27,size=50,prob=0.5)
Q1 = qbinom(.95,size=50,prob=0.5)
x  = rbinom(5,size=50,prob=0.5)
```

Si X es una variable con distribución binomial de parámetros $n = 50$ y $p = 0.5$,

$$P1 = P(X = 27),$$

$$P2 = P(X \leq 27),$$

y Q1 es el valor tal que

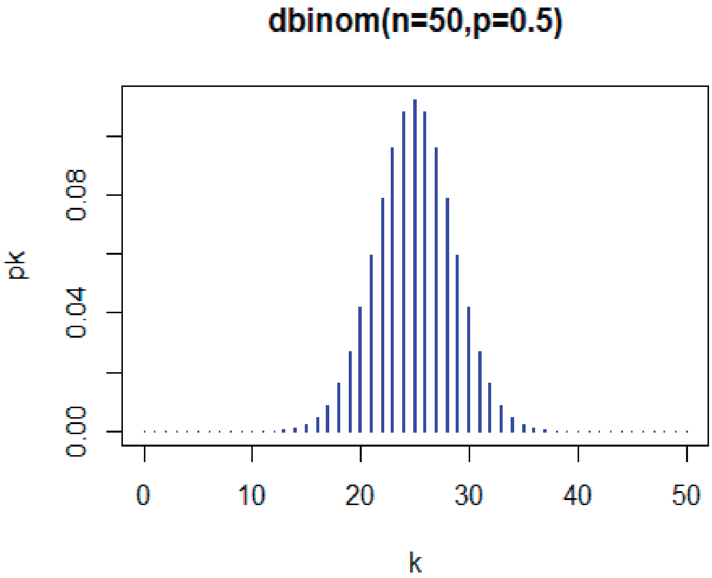
$$P(X \leq Q1) = 0.95$$

x es un vector con 5 valores generados al azar de X.

```
## [1] 25 26 26 22 32
```

Distribución binomial

```
k=0:50
pk = dbinom(k,size=50,prob=0.5)
plot(k,pk,col='blue',type='h',lwd=2,main='dbinom(n=50,p=0.5)')
```



Ejemplo: control de procesos

Las piezas de un proceso de fabricación pueden ser aceptables o defectuosas. Cuando el proceso está bajo control, el porcentaje de piezas defectuosas fabricadas es 3%. De la producción de cada hora se toma una muestra de 200 piezas al azar.

1. El ingeniero de calidad decide que si el número de defectuosas en la muestra es 8 o más, se detenga el proceso y se analice si está bajo control. ¿Cuál es la probabilidad de parar el proceso de manera injustificada?

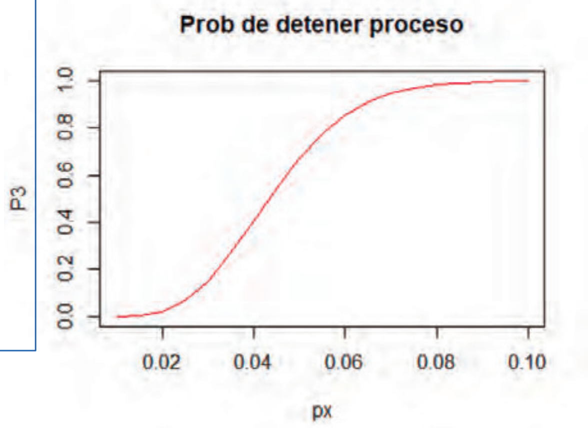
2. Calcular la probabilidad de parar el proceso, cuando está fabricando con un porcentaje de defectuosas del 5%.

3. Calcular la probabilidad de no detener el proceso si está fabricando un 6% de piezas defectuosas.

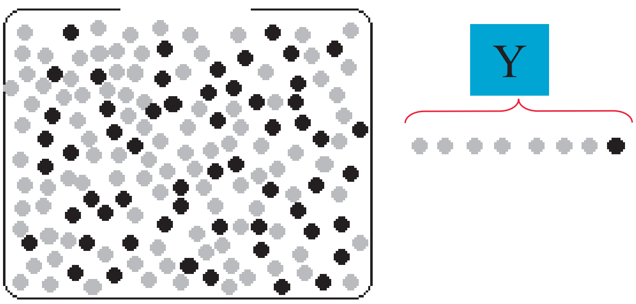
4. Repetir el cálculo del apartado 3 para p (porcentaje de defectuosas) entre 1% y 10%. Dibujar las probabilidades en función de p.

Solución

1. `P1 = pbinom(8,200,0.03)` igual a 0.8504038
2. `P2 = 1- pbinom(8,200,0.05)` igual a 0.6729755
3. `P3 = pbinom(8,200,0.06)` igual a 0.1469911
4.
```
px = seq(0.01,0.10,0.005)
P3 = 1 - pbinom(8,200,px)
plot(px,P3,type='l',col='red',
main='Prob de detener proceso')
```



Distribución Geométrica (*p*)



Proporción defectuosas = p

Y = "Piezas extraídas hasta que aparezca una defectuosa"

Distribución geométrica (p)

□ 1 $\longrightarrow$ p

□□ 2 $\longrightarrow$ $(1-p)p$

□□□ 3 $\longrightarrow$ $(1-p)^2p$

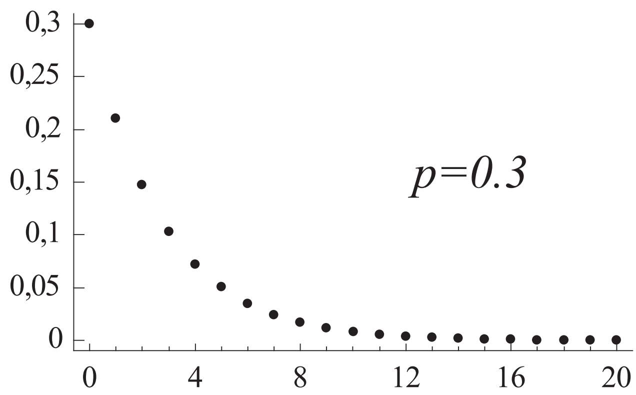
$\downarrow$ $\downarrow$

□□□...□ k $\longrightarrow$ $(1-p)^{k-1}p$

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1,2,3,\ldots$$

Propiedades de la v.a. geométrica

$$E[Y] = \frac{1}{p}, \qquad Var[Y] = \frac{1-p}{p^2}$$



Distribución Geométrica

```
p1 = dgeom(6, prob=0.10)
p2 = pgeom(6, prob=0.10)
q1 = qgeom(.90, prob=0.10)
 x = rgeom(5, prob=0.10)
```

Si X es una variable con distribución geométrica de parámetro $p = 0.10$,

$$p1 = P(X = 6) = 0.0531441,$$

$$\mathrm{p2} = P(X \leq 6) = 0.5217031,$$

y q1 es el valor tal que

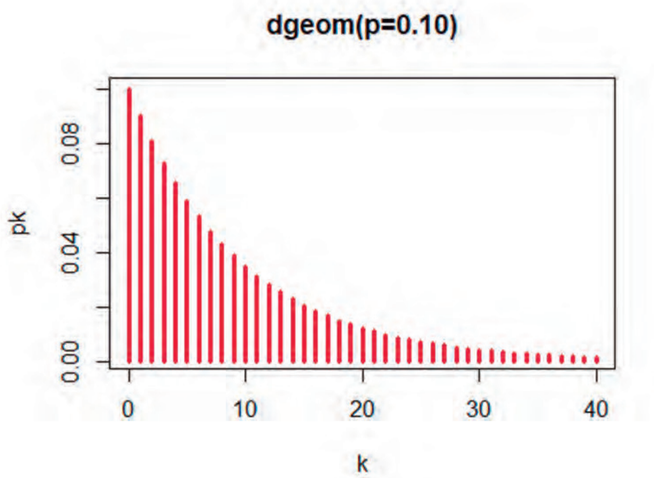
$$P(X \leq q1) = 0.90 \rightarrow q1 = 21$$

x es un vector con 5 valores generados al azar de X.

```
## [1] 11 21 2 3 64
```


Distribución Geométrica

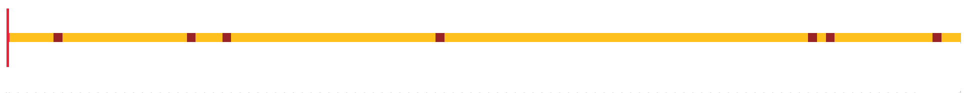
```
k=0:40
pk=dgeom(k,prob=0.10)
plot(k,pk,type='h',col='red',lwd=3,main='d
geom(p=0.10)')
```



Distribución de Poisson

- Número de defectos aparecidos en tramos de longitud fija de hilos de cobre.
- Número de partículas por centímetro cúbico en líquidos con sustancias en suspensión.
- Emisiones radiactivas: número de partículas emitidas en intervalos de tiempo fijo.
- Número de llamadas a una centralita de teléfonos en un día

Distribución de Poisson

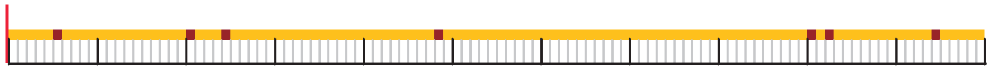


Ejemplo: Fabricación continua de conductor de cobre.

$\lambda \equiv$ Número medio de defectos cada 100 m

$X \equiv$ Número de defectos en un tramo de 100 m

Límite de la dist. binomial



$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad p = \frac{\lambda}{n}$$

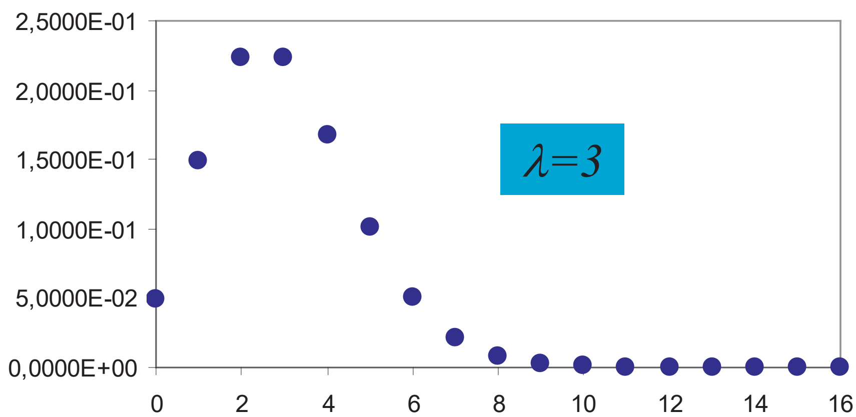
$$P_X(x) = \lim_{n\to\infty} \frac{n!}{(n-x)!x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\to 1}$$

$$= \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0,1,2,\ldots$$

Distribución de Poisson

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0,1,2,..$$



Media y Varianza

$$P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}, \quad x=0,1,2,\ldots$$

$$E[X]=\sum_{x=0}^{\infty} x\times\frac{\lambda^x}{x!}e^{-\lambda}=\sum_{x=1}^{\infty} x\times\frac{\lambda^x}{x!}e^{-\lambda}$$

$$=\lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}=\lambda \qquad \longrightarrow e^{\lambda}$$

$$\boxed{E[X]=\lambda \qquad\qquad Var[X]=\lambda}$$

Ejemplo

Una fuente radiactiva emite partículas según un proceso de Poisson de media 10 partículas por minuto. Se desea calcular:

- Probabilidad de 5 partículas en un minuto
- Probabilidad de 0 partículas en un minuto
- Probabilidad de más de 5 partículas en un minuto.
- Probabilidad de 30 o menos partículas en 5 minutos.

Ejemplo Poisson

1. $P(X=5)=e^{-10}\dfrac{10^5}{5!}=0.0378$
2. $P(X=0)=e^{-10}=4.54\times 10^{-5}.$
3. $P(X>5)=1-P(X\leq 5)$

$$=1-e^{-10}\sum_{x=0}^{5}\frac{10^x}{x!}=0.933.$$

4. $Y\equiv$ Nº de partículas en 5 minutos

$\lambda' = 5\times 10=50$

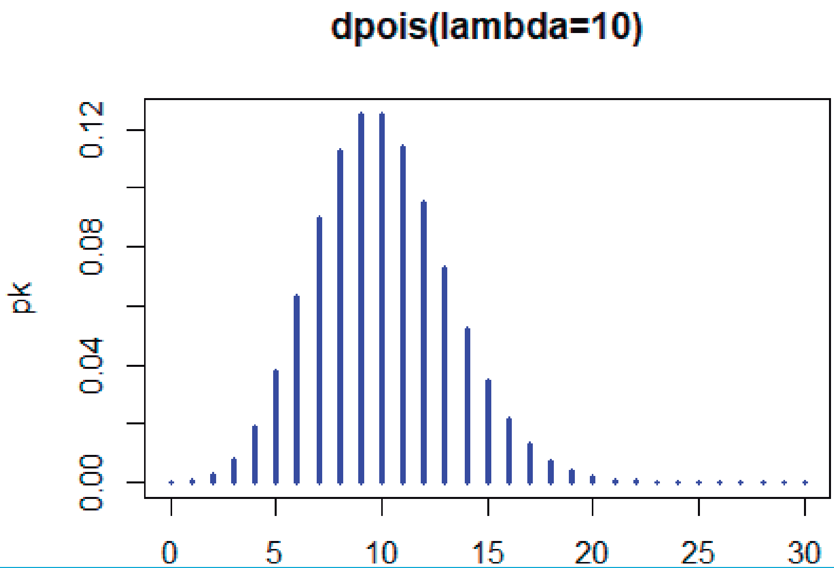
$$P(Y\leq 30)=e^{-50}\sum_{x=0}^{30}\frac{50^x}{x!}=0.0016$$

Ejemplo Poisson

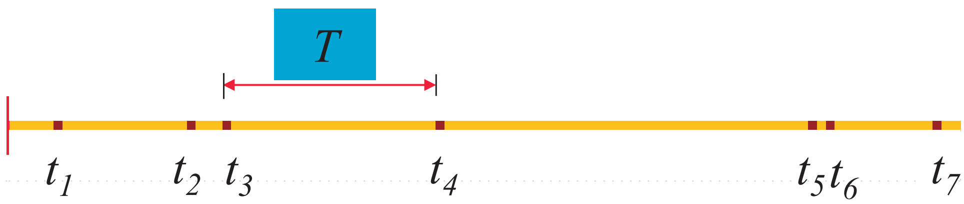
```
1. p1 = dpois(5, lambda = 10)     = 3.783327e-02

2. p2 = dpois(0, lambda = 10)     = 4.539993e-05

3. p3 = 1 - ppois (5, 10)          = 0.932914

4. p4 = ppois(30,lambda=50)       = 1.594027e-03
```

Poisson de media 10

```
k=0:30   pk=dpois(k,lambda=10)
plot(k,pk,type='h',col='blue',lwd=3,main='dpois(lambda=10)')
```



Distribución Exponencial

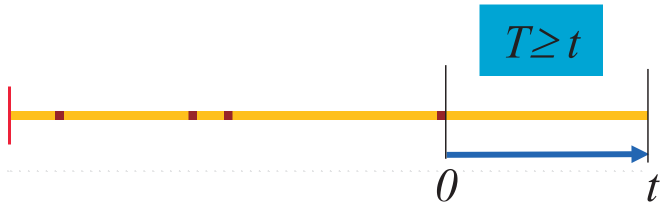


Ejemplo: Fabricación continua de conductor de cobre.

$\lambda \equiv$ Número medio de defectos cada 100 m

$T \equiv$ "Distancia entre dos defectos consecutivos"

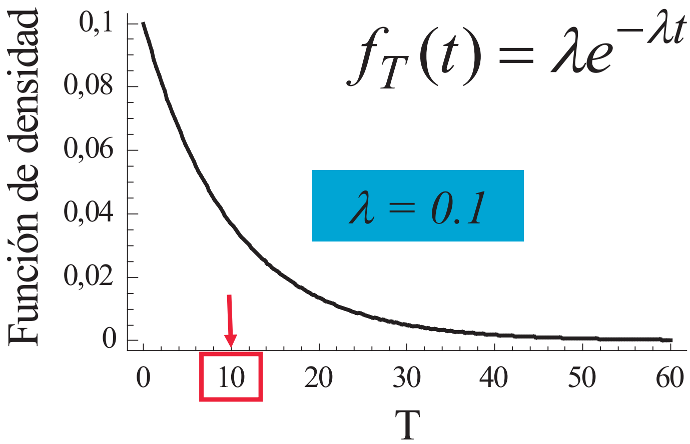
Distribución Exponencial



$$P(T \geq t) = P\{0 \text{ defectos en el intervalo } [0, t)\}$$
$$= e^{-\lambda t}, \quad t \geq 0$$
$$F_T(t) = P(T \leq t)$$
$$= 1 - e^{-\lambda t}, \quad t \geq 0$$

$$f_T(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Propiedades (Exponencial)



$$f_T(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

$$E[T] = \int_{-\infty}^{\infty} t f_T(t) dt = \int_0^{\infty} \lambda t e^{-\lambda t} dt = \frac{1}{\lambda}$$

$$Var[T] = E[T^2] - E[T]^2 = \int_0^{\infty} \lambda t^2 e^{-\lambda t} dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Ejemplo distribución exponencial

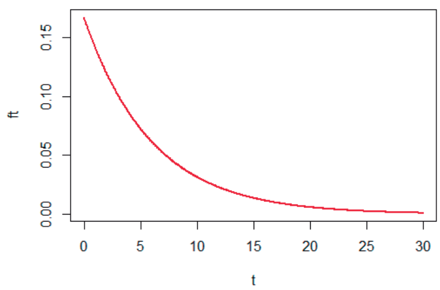
Una fuente radiactiva emite partículas según un proceso de Poisson de media 10 partículas por minuto. Se desea calcular:

- Tiempo medio entre partículas
- Calcula la función de densidad del tiempo en segundos entre dos partículas y dibújala
- Probabilidad de que aparezca la primera partícula antes de 4 segundos
- Probabilidad de un minuto sin partículas.

Ejemplo distribución exponencial (sol.)

1. $\lambda = 10\,\frac{part}{min}, \rightarrow \mu = \frac{1}{\lambda} = 0.1\,\frac{min}{part} = 6\,\frac{s}{part}$

2. $f_T(t) = \frac{1}{6}\exp\left(-\frac{t}{6}\right), t > 0$



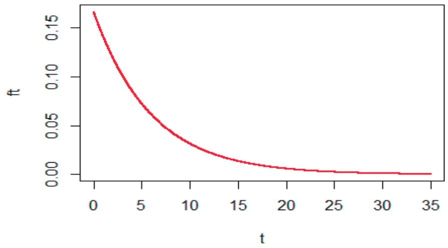
3. $P(T \leq 4) = \int_0^4 \frac{1}{6}\exp\left(-\frac{t}{6}\right) dt = 1 - e^{-\frac{4}{6}} = 0.486582$

4. $P(T > 60) = 1 - \int_0^{60} \frac{1}{6}\exp\left(-\frac{t}{6}\right) dt = e^{-10} = 4.54 \times 10^{-5}$

Ejemplo distribución exponencial

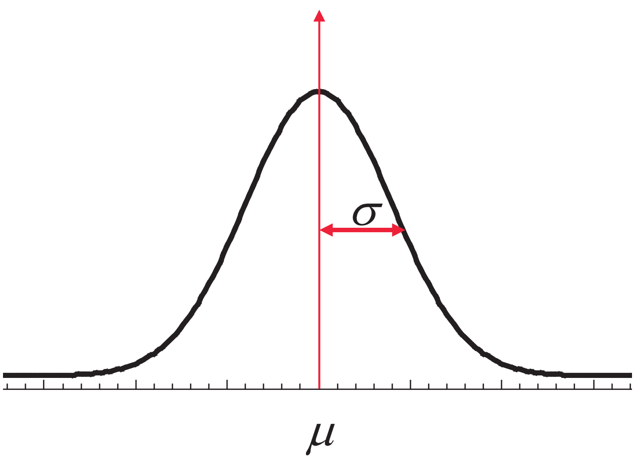
1. Rate = 10 part/min = 10 part/60 s= 1/6 part/s

```
2. t=seq(0,35,.01)
ft = dexp(t,rate = 1/6)
plot(t,ft,col='red',lwd='2',type='l')
```



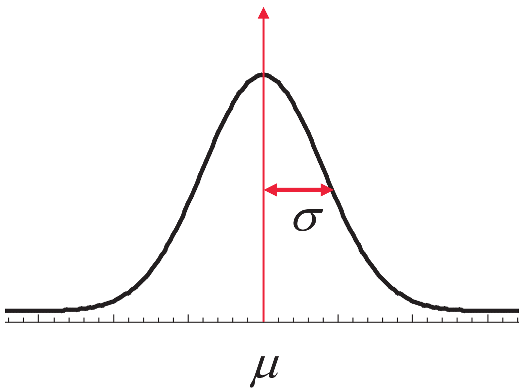

```
3. p3 = pexp(4,1/6) =  0.4865829
4. p4 = 1- pexp(60,1/6) = 4.539993e-05
```


Distribución Normal: Campana de Gauss



$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \Re$$

Medidas Características



$$X \rightarrow N(\mu, \sigma)$$

$E[X] = \mu$

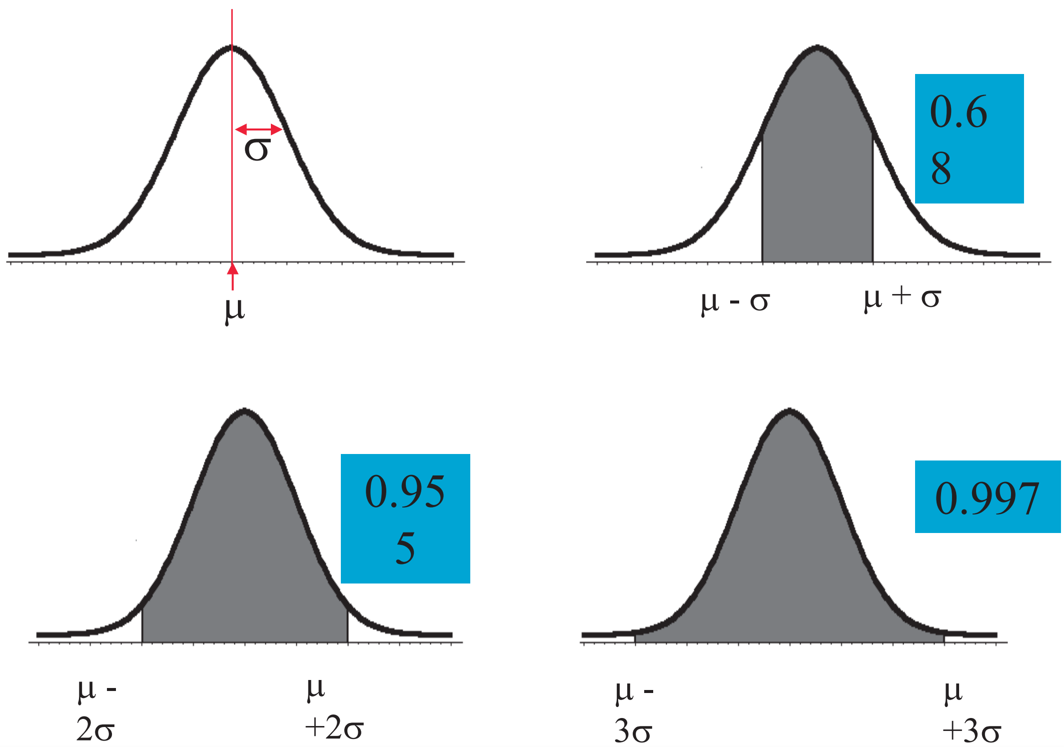
$Var[X] = \sigma^2$

$E[(X-\mu)^3] = 0$

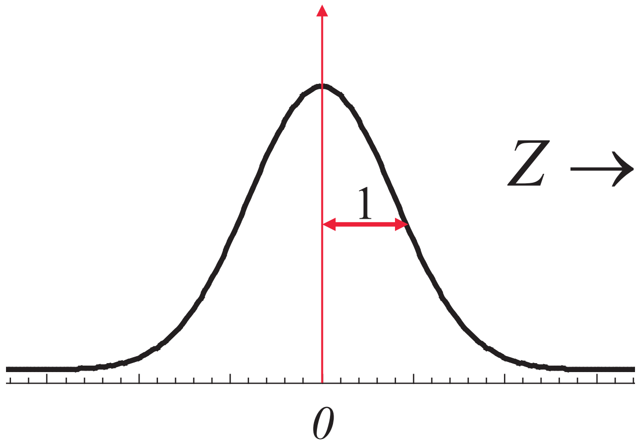
$E[(X-\mu)^4] = 3\sigma^4$

$$CA = Asimetría = \frac{E[(X-\mu)^3]}{\sigma^3} = 0$$

$$CAp = Curtosis = \frac{E[(X-\mu)^4]}{\sigma^4} = 3$$



Normal Estándar



$$Z \to N(0,1)$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \Re$$

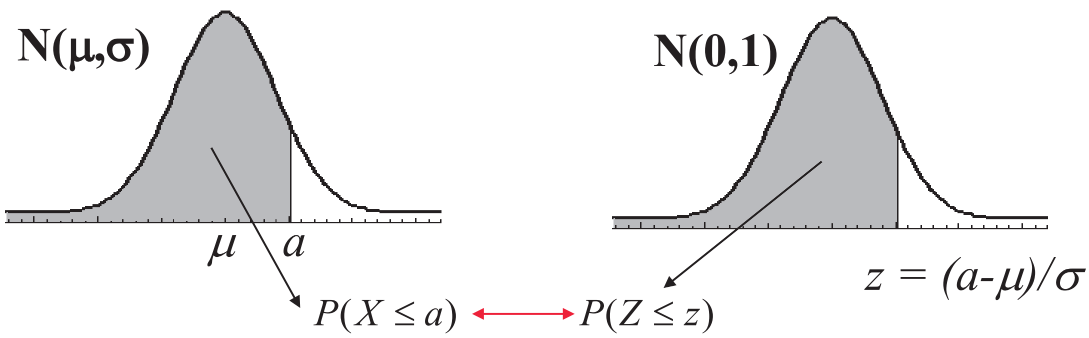
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

TABLAS

Estandarización

$$X \to N(\mu,\sigma) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \to N(0,1)$$

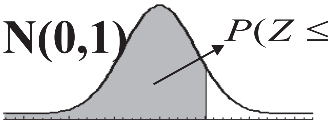
$$P(X \le a) = P(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}) = P(Z \le \frac{a-\mu}{\sigma}) = \Phi(\frac{a-\mu}{\sigma}).$$



TABLA

Normal Estándar

N(0,1) $P(Z \le z)$



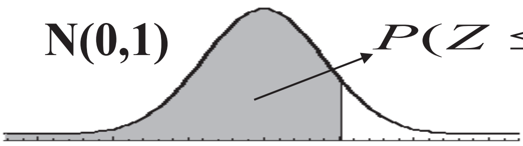
z

Ejemplo.

$P(Z \le 1.96) = 0.9750$

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0,1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0,2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0,3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0,4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0,5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0,6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0,7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0,8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0,9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1,0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1,1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1,2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1,3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1,4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1,5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1,6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1,7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1,8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1,9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2,0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2,1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2,2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2,3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2,4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2,5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2,6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2,7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2,8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2,9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3,0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

N(0,1) $P(Z \le z)$



z

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,1	.9990323	.9990645	.9990957	.9991259	.9991552	.9991836	.9992111	.9992377	.9992636	.9992886
3,2	.9993128	.9993363	.9993590	.9993810	.9994023	.9994229	.9994429	.9994622	.9994809	.9994990
3,3	.9995165	.9995335	.9995499	.9995657	.9995811	.9995959	.9996102	.9996241	.9996375	.9996505
3,4	.9996630	.9996751	.9996868	.9996982	.9997091	.9997197	.9997299	.9997397	.9997492	.9997584
3,5	.9997673	.9997759	.9997842	.9997922	.9997999	.9998073	.9998145	.9998215	.9998282	.9998346
3,6	.9998409	.9998469	.9998527	.9998583	.9998636	.9998688	.9998739	.9998787	.9998834	.9998878
3,7	.9998922	.9998963	.9999004	.9999042	.9999080	.9999116	.9999150	.9999184	.9999216	.9999247
3,8	.9999276	.9999305	.9999333	.9999359	.9999385	.9999409	.9999433	.9999456	.9999478	.9999499
3,9	.9999519	.9999538	.9999557	.9999575	.9999592	.9999609	.9999625	.9999640	.9999655	.9999669
4,0	.9999683	.9999696	.9999709	.9999721	.9999733	.9999744	.9999755	.9999765	.9999775	.9999784

Ejemplo (Normal)

La longitud X de ciertos tornillos es una variable aleatoria con distribución normal de media 30 mm y desviación típica 0.2 mm. Se aceptan como válidos aquellos que cumplen

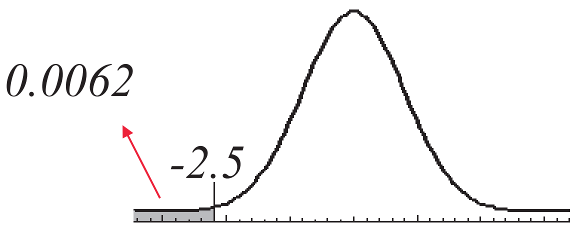
$$29.5 \leq X \leq 30.4.$$

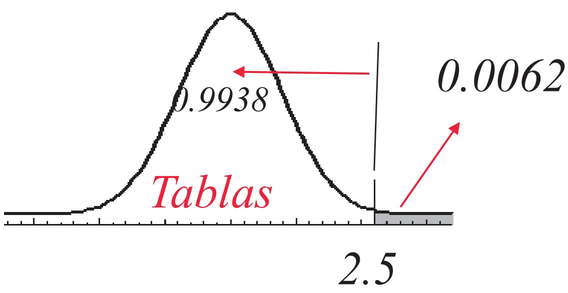
1. Proporción de tornillos no aceptables por cortos.
2. Proporción de tornillos no aceptables por largos.
3. Proporción de tornillos válidos.

Ejemplo (Solución2)

$$X \to N(30, 0.2) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \to N(0,1)$$

1. $$P(X \leq 29.5) = P\left(\frac{X-30}{0.2} \leq \frac{29.5-30}{0.2}\right)$$
$$= P(Z \leq -2.5) = \Phi(-2.5) = 0.0062$$





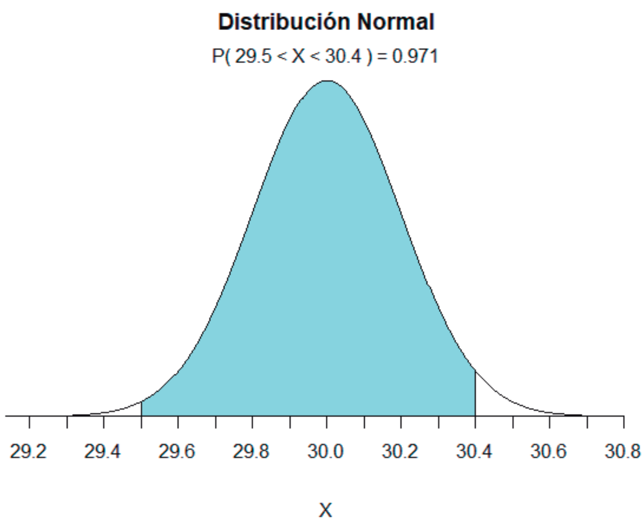
Ejemplo (Solución)

$$2. \quad P(X \leq 30.4) = P(\frac{X-30}{0.2} \leq \frac{30.4-30}{0.2})$$
$$= P(Z \leq 2) = \Phi(2) = 0.9772$$
$$P(X > 30.4) = 1 - 0.9772 = 0.0228$$

$$3. \quad P(29.5 \leq X \leq 30.4) = P(\frac{29.5-30}{0.2} \leq \frac{X-30}{0.2} \leq \frac{30.4-30}{0.2})$$
$$= P(-2.2 \leq Z \leq 2)$$
$$= \Phi(2) - \Phi(-2.2)$$
$$= 0.9772 - 0.0062 = 0.9710$$

Ejemplo (Solución R)

```
1. p1 = pnorm(29.5, mean = 30, sd = 0.2) = 0.006209665
2. p2 = 1- pnorm(30.4, mean = 30, sd = 0.2) = 0.022750132
3. p3 = pnorm(30.4, mean = 30, sd = 0.2)- pnorm(29.5, mean = 30, sd = 0.2) = 0.97104
```



Teorema Central del Límite

Sean $X_1, X_2, \dots, X_n$ variables aleatorias independientes con media μ y desviación típica σ, entonces cuando *n* tiende a infinito

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \to N(\mu, \frac{\sigma^2}{n})$$

Válido en muchos casos prácticos para n pequeños

Teorema Central del Límite

Sea $X_1, X_2, \dots, X_n$ una secuencia de variables aleatorias independientes con la misma distribución de probabilidad de media μ y varianza $\sigma^2 < \infty$. Entonces

$$\lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le t \right) = \Phi(t), \quad -\infty < t < \infty$$

donde $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$ y

$$\Phi(\mathrm{t}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$$

es la función de distribución de la normal estándar.

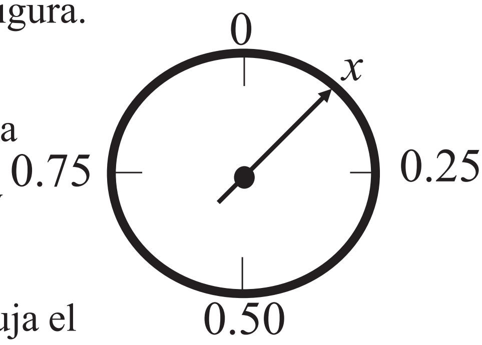
Aplicación: Media de 12 "Uniformes"

Se lanza 12 veces la ruleta circular que aparece en la figura. Se calcula la media de los doce valores.

(a) Genera por simulación 12 valores y obtén la media

(b) Repite el experimento anterior cuatro veces más y proporciona los valores medios obtenidos

(c) Repite el experimento anterior 10000 veces y dibuja el histograma

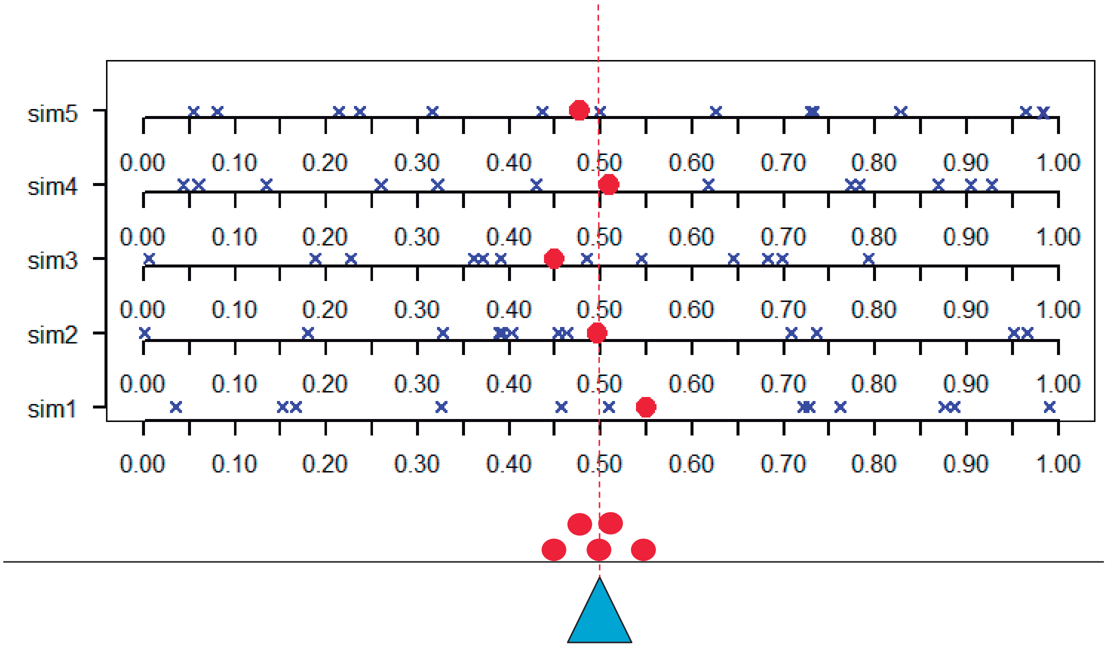


(d) Analiza teóricamente el experimento y determina la distribución de probabilidad de la media de 12 variables aleatorias independientes con distribución uniforme en [0,1]

Simulaciones de 12 uniformes

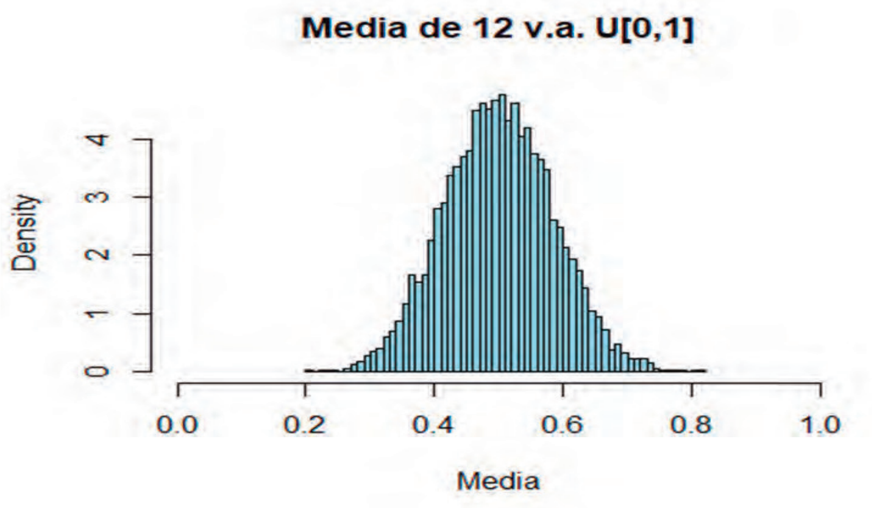
		Medias
`s1=runif(12)`	`## [1] 0.720904 0.875773 0.760982 0.886125 0.456481 0.166372 0.325095` `## [8] 0.509224 0.727705 0.989737 0.034535 0.152373`	0.55044
`s2=runif(12)`	`## [1] 0.735685 0.001136 0.391203 0.462494 0.388144 0.402485 0.1789636` `## [8] 0.951658 0.453728 0.326752 0.965415 0.707481`	0.49709
`s3=runif(12)`	`## [1] 0.644542 0.389828 0.698543 0.544057 0.226467 0.484557 0.793007` `## [8] 0.005987 0.187712 0.681833 0.370104 0.361625`	0.44902
`s4=runif(12)`	`## [1] 0.868795 0.904155 0.617425 0.134032 0.782193 0.429199 0.927274` `## [8] 0.773243 0.259681 0.321225 0.060195 0.043456`	0.51007
`s5=runif(12)`	`## [1] 0.055053 0.625542 0.964470 0.827302 0.315028 0.213025 0.732496` `## [8] 0.499241 0.729772 0.080336 0.435530 0.236580`	0.47619

Simulaciones y distribución de las medias



Simulación de 10000 medias e histograma

```
x = matrix(runif(120000),10000,12)     # crea una matriz de dimensión 10.000 x 12
con número aleatorios con dist. unif. [0,1]
m= rowMeans(x)                         # calcula las medias de las 10.000 filas
hist(rowMeans(x),nclass=50,col='light blue',xlab='Media',xlim=c(0,1),prob=T,main
='Media de 12 v.a. U[0,1]')
```



Aplicación Teorema Central del Límite
(media de 12 V.A. con dist. Uniforme [0,1])

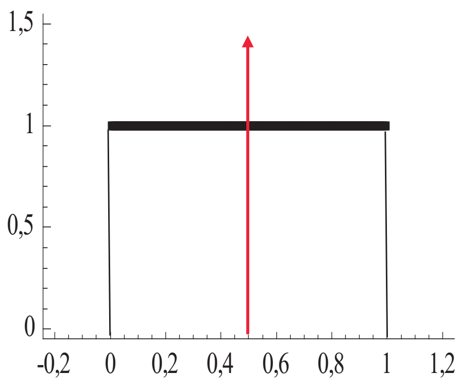
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{12}}{12} \qquad E[X_i] = \frac{1}{2} \quad y \qquad Var[X_i] = \frac{1}{12}$$

$$E[\bar{X}] = E\left[\frac{X_1 + X_2 + \cdots + X_{12}}{12}\right] = \frac{E[X_1] + E[X_2] + \cdots + E[X_{12}]}{12} = \frac{{}^1/_2 + {}^1/_2 + \cdots + {}^1/_2}{12} = \frac{1}{2}$$

$$Var[\bar{X}] = Var\left[\frac{X_1 + X_2 + \cdots + X_{12}}{12}\right] = \frac{Var[X_1] + Var[X_2] + \cdots + Var[X_{12}]}{12^2} = \frac{{}^1/_{12} + {}^1/_{12} + \cdots + {}^1/_{12}}{12^2} = \frac{1}{12^2}$$

$$\boxed{\bar{X} \sim N\left(0.5, \frac{1}{12}\right)}$$

Media y Varianza de Uniforme [0,1]



$$\boxed{E[X] = \int_o^1 x dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}}$$

$$Var[X] = E[X^2] - E[X]^2$$

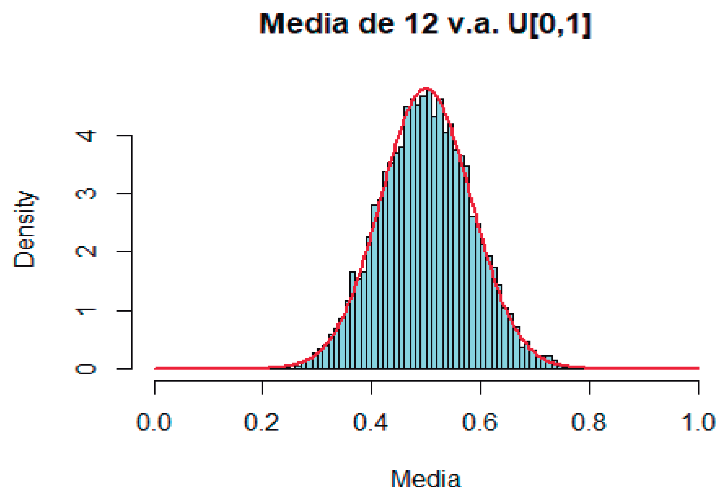
$$E[X^2] = \int_o^1 x^2 dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}$$

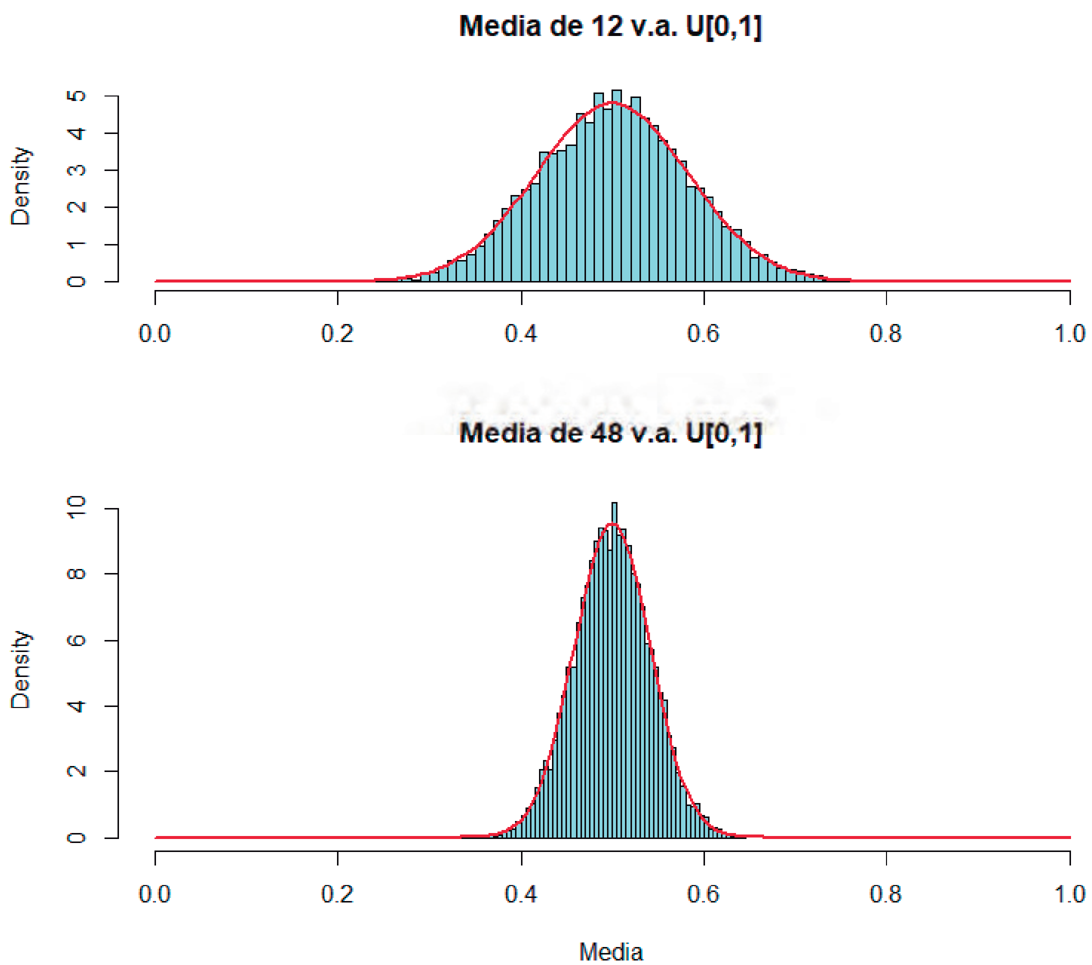
$$Var(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

Aplicación Teorema Central del Límite
(media de 12 V.A. con dist. Uniforme [0,1])

```
x = matrix(runif(120000),10000,12)     # crea una matriz de dimensión 10.000 x 12
con número aleatorios con dist. unif. [0,1]
m= rowMeans(x)                         # calcula las medias de las 10.000 filas
hist(rowMeans(x),nclass=50,col='light blue',xlab='Media',xlim=c(0,1),prob=T,main
='Media de 12 v.a. U[0,1]')
```

```
x1 = seq(0,1,.005)
lines(x1,dnorm(x1,mean=0.5,sd=1/12),col='red',lwd=2)
```

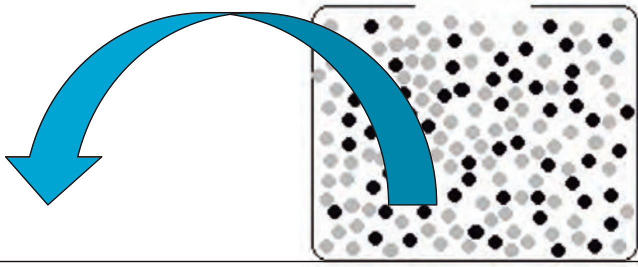




2. Aplicación Teorema Central del Límite

Binomial como suma de variables aleatorias

Proporción de defectuosas ***p***



Extraigo una pieza con reposición

$$X = \begin{cases} 0, Aceptable \\ 1, Defectuosa \end{cases} \qquad P(X=x) = \begin{cases} 1-p, x=0 \\ p, \qquad x=1 \end{cases}$$

Repito el experimento ***n*** veces, llamo $X_1, X_2, \dots, X_n$

El número de piezas defectuosas en la muestra de ***n*** es igual a $Y = X_1 + X_2 + \cdots + X_n$ **Y ->Binomial**

2. Aplicación Teorema Central del Límite

Binomial como suma de variables aleatorias

El número de piezas defectuosas en la muestra de ***n*** es igual a $Y = X_1 + X_2 + \cdots + X_n$

$$E[X_i] = 0 \times (1-p) + 1 \times p = p$$

$$Var[X_i] = E[X_i{}^2] - E[X_i]^2 = (0^2 \times (1-p) + 1^2 \times p) - p^2 = p(1-p)$$

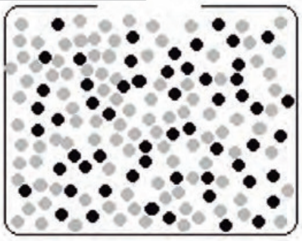
E[Y]= E$[X_1 + X_2 + \cdots + X_n] = np$

Var[Y]=Var$[X_1 + X_2 + \cdots + X_n] = np(1-p)$

2. Aplicación Teorema Central del Límite

Binomial como suma de variables aleatorias

Proporción de defectuosas ***p***



Extraigo $\boldsymbol{n}$ piezas con reposición

$$X_i = \begin{cases} 0, & P(X_i = 0) = 1 - p \\ 1, & P(X_i = 1) = p \end{cases}$$

El número de piezas defectuosas en la muestra de $\boldsymbol{n}$ es igual a $Y = X_1 + X_2 + \cdots + X_n$

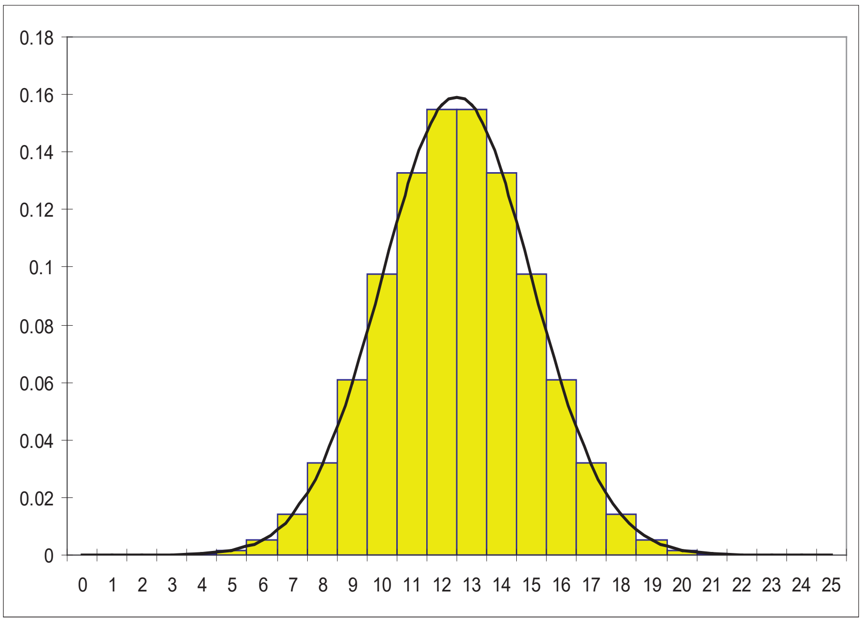
Y tiene distribución Binomial $(\boldsymbol{n}, \boldsymbol{p})$

Y tiende a distribución normal

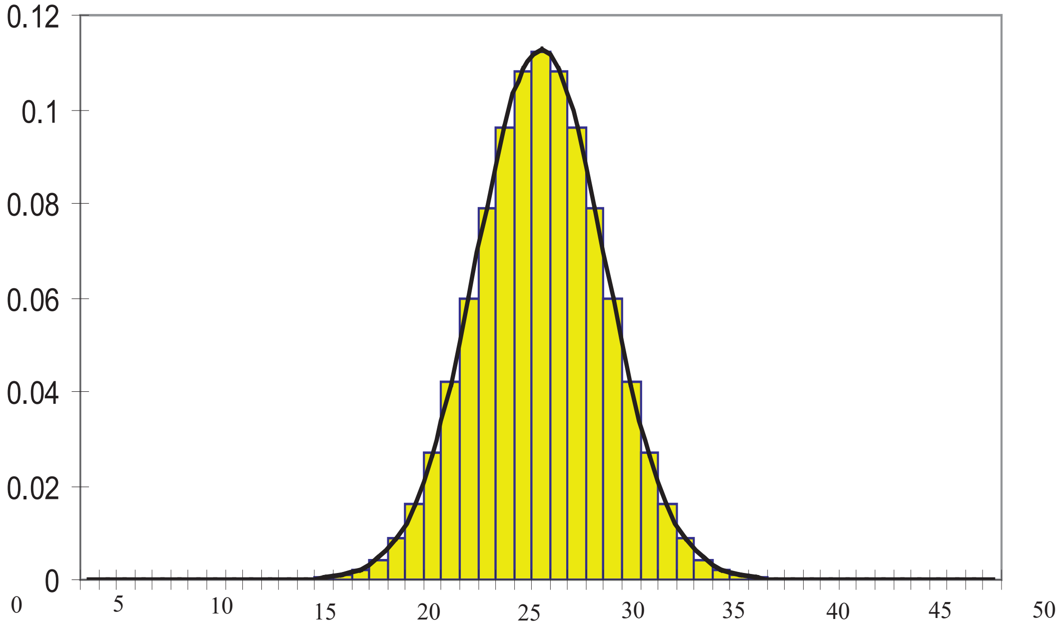
$$\boxed{Y \sim N\left(np, \sqrt{np(1-p)}\right)}$$

$$\boldsymbol{B(n,p) \sim N\left(np, \sqrt{np(1-p)}\right)}$$

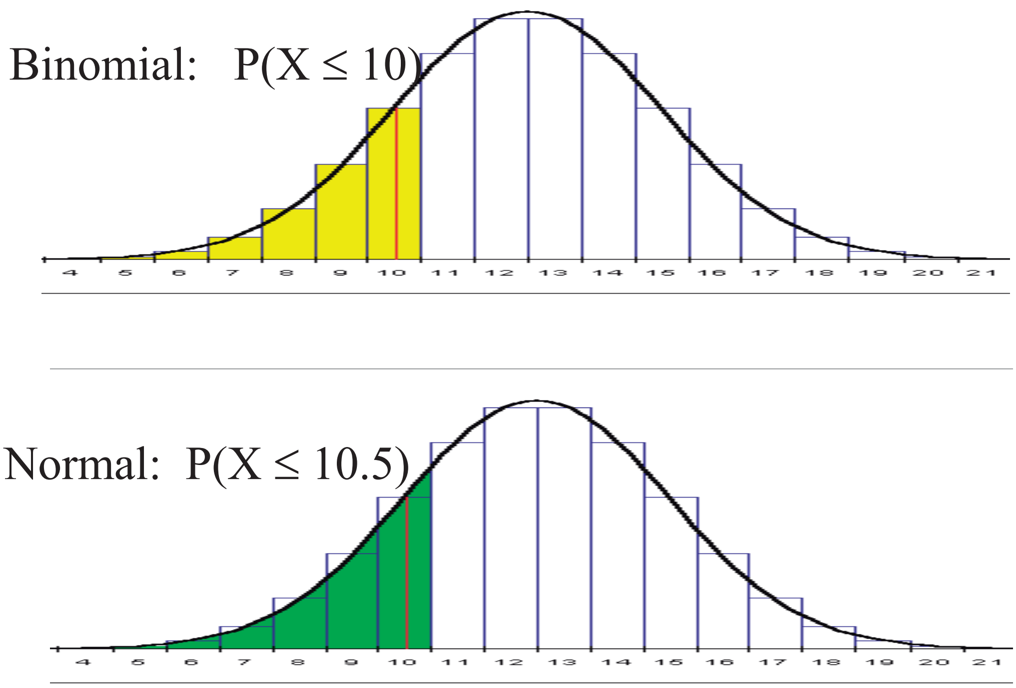
Aproximación Binomial-Normal n=25, p=1/2



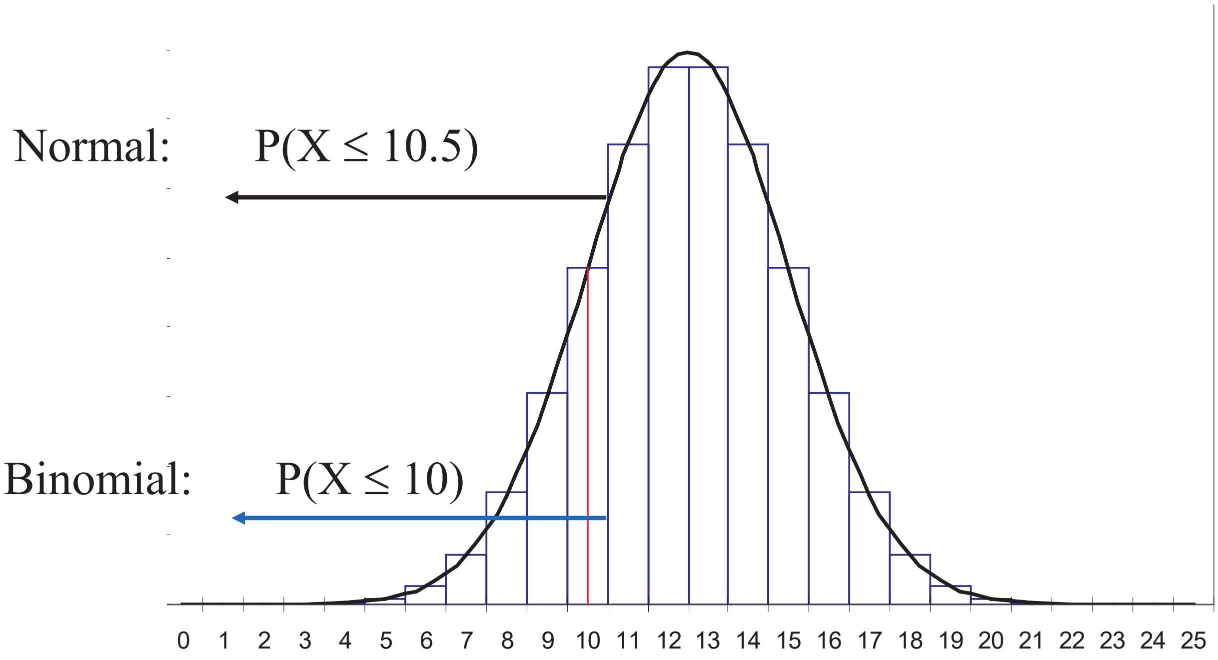
Aproximación Binomial-Normal n = 50, p=0.5



Corrección por continuidad



Corrección por continuidad



Se ha tomado una muestra de 45 piezas de un proceso que fabrica un promedio de 25% de piezas fuera de especificación.

1.¿Cuál es la probabilidad de que en la muestra haya exactamente 13 elementos defectuosos?

Cálculo exacto: $X \rightarrow$ Binomial$(n = 45,\ p = 0.25)$

$$P(X = 13) = \binom{45}{13} 0.25^{13} 0.75^{32} = 0.110$$

Aproximación Normal: $Y \rightarrow N(11.25, 2.9)$

$$P(X = 13) = P(12.5 \leq Y \leq 13.5) = 0.1142$$

2. ¿Cuál es la probabilidad de que la muestra contenga 13 o más piezas defectuosas?

$$E : P(X \geq 13) = P(X = 13) + P(X = 14) + \cdots + P(X = 45) = 0.325$$

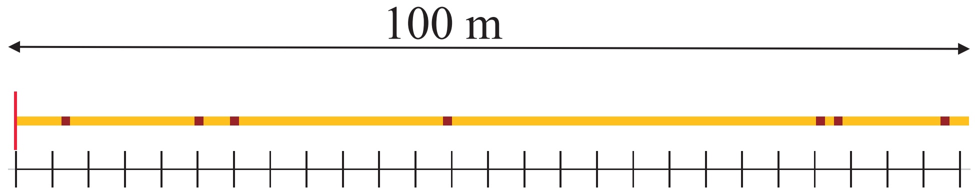
$$A : P(Y \geq 12.5) = 1 - \Phi\left(\frac{12.5 - 11.25}{2.9}\right) = 1 - \Phi(0.4310) = 0.333$$

3. Aplicación teorema central del límite

Aprox. de la distribución de Poisson

Ejemplo: Fabricación continua de conductor de cobre.

$\lambda \equiv$ Número medio de defectos cada 100 m



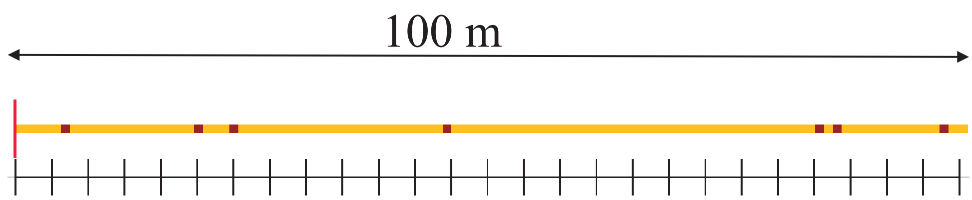
Dividimos los 100 metros en 100 tramos de 1 metro, sea X_1 el número de defectos en el primer metro, X_2 el número de defectos en el segundo metro, y así sucesivamente.

$X \equiv$ Número de defectos en un tramo de 100 m

$$\boldsymbol{X = X_1 + X_2 + \cdots + X_{100} \sim Poisson(\lambda)}$$

3. Aplicación teorema central del límite

Aprox. de la distribución de Poisson



$X \equiv$ Número de defectos en un tramo de 100 m

$$X = X_1 + X_2 + \cdots + X_{100}$$

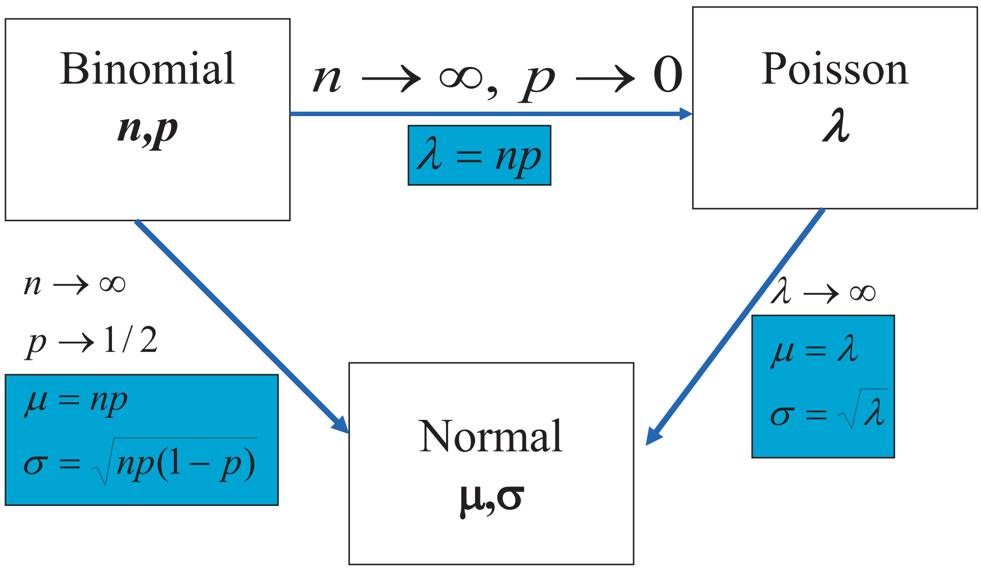
$$\mathrm{E}[X_i] = {}^{\lambda}/_{100}\,, Var[X_i] = {}^{\lambda}/_{100}$$

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_{100}] = \lambda$$

$$\mathrm{Var}[X] = Var[X_1] + Var[X_2] + \cdots + Var[X_{100}] = \lambda$$

$$\boxed{X \sim N(\lambda, \sqrt{\lambda})}$$

Binomial-Poisson-Normal



Aplicación a Control de Recepción

Plan de muestreo simple por atributos

Una compañía recibe lotes con un gran número de piezas. Según el contrato cada lote debe tener como máximo una proporción de piezas defectuosas igual p_A *(AQL).*

Un plan de muestreo simple por atributos consiste en determinar

n: número de piezas muestreadas

c: número máximo de piezas defectuosas en la muestra

De forma que si X es el número de piezas defectuosas en la muestra se aplica la siguiente regla:

$x \leq c$ se acepta el lote

$x > c$ se rechaza el lote

Riesgos del vendedor y comprador

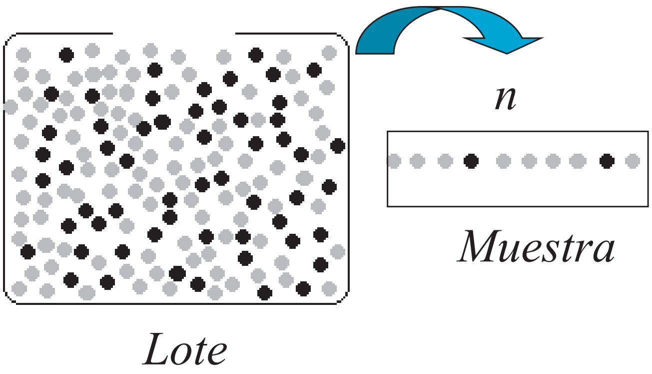
- **Riesgo del vendedor**: Probabilidad de rechazar un lote bueno (con porcentaje de defectuosas igual al p_A *(AQL))*

$$\alpha = P(X > c \mid p = p_A).$$

- **Riesgo del comprador**: Probabilidad de aceptar un lote malo (con un porcentaje de defectuosas $p_R >> p_A$)

$$\beta = P(X \leq c \mid p = p_R).$$

Planteamiento del problema



DATOS:

$p_A = AQL$ α: *Riesgo vendedor*

$p_R = RQL$ β: *Riesgo comprador*

OBJETIVO:

n: tamaño muestral

c: máximo defectuosas

Ecuación del vendedor

$p \equiv$ Proporción de piezas defectuosas en el lote

$X \equiv$ Número de piezas defectuosas en una muestra de n

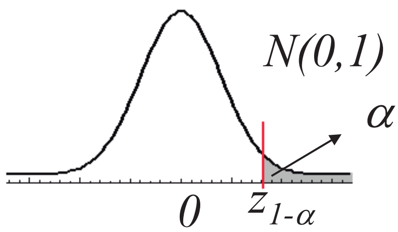
$X \to$ Binomial $(n, p) \approx N(np, \sqrt{np(1-p)})$

Si $p = p_A$

$$\alpha = P(X > c | p = p_A) = P\left(\frac{X - np_A}{\sqrt{np_A(1-p_A)}} > \frac{c - np_A}{\sqrt{np_A(1-p_A)}}\right)$$

$$= P(Z \geq z_{1-\alpha})$$

Conocido $\alpha \Rightarrow \boxed{z_{1-\alpha} = \frac{c - np_A}{\sqrt{np_A(1-p_A)}}}$



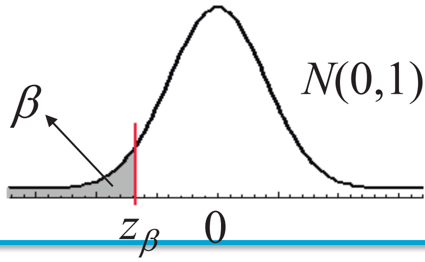
Ecuación del comprador

Si $p = p_R$:

$$\beta = P(X \le c | p = p_R) = P\left(\frac{X - np_R}{\sqrt{np_R(1-p_R)}} \le \frac{c - np_R}{\sqrt{np_R(1-p_R)}}\right)$$

$$= P(Z \le z_\beta)$$

Conocido $\beta \Rightarrow z_\beta = \dfrac{c - np_R}{\sqrt{np_R(1-p_R)}}$

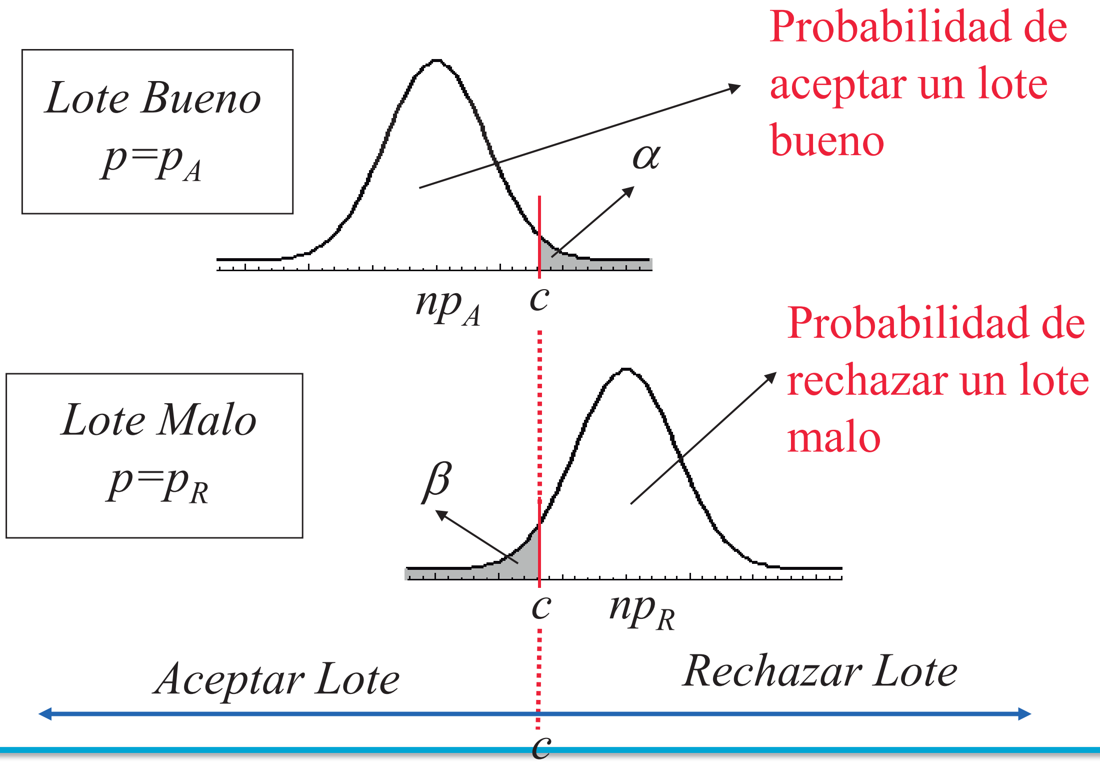


Valores de *n* y *c*

$$z_{1-\alpha} = \frac{c - np_A}{\sqrt{np_A(1-p_A)}} \qquad z_\beta = \frac{c - np_R}{\sqrt{np_R(1-p_R)}}$$

$$n = \left(\frac{z_{1-\alpha}\sqrt{p_A(1-p_A)} - z_\beta\sqrt{p_R(1-p_R)}}{p_R - p_A}\right)^2$$

$$c = np_A + z_{1-\alpha}\sqrt{np_A(1-p_A)}$$



Ejemplo: plan de muestreo

Diseñar un plan de muestreo para lotes de 10000 unidades con un AQL igual al 0.02, RQL igual a 0.08, riesgo de comprador de 0.10 y del vendedor igual al 0.05.

$$\text{Para } \alpha = 0.05 \Rightarrow z_{1\text{-}\alpha} = 1.64 \text{ y } \beta = 0.10 \Rightarrow z_{\beta} = -1.28$$

$$n = \left(\frac{1.65\sqrt{0.02 \times 0.98} + 1.28\sqrt{0.08 \times 0.92}}{0.08 - 0.02} \right)^2 \approx 93$$

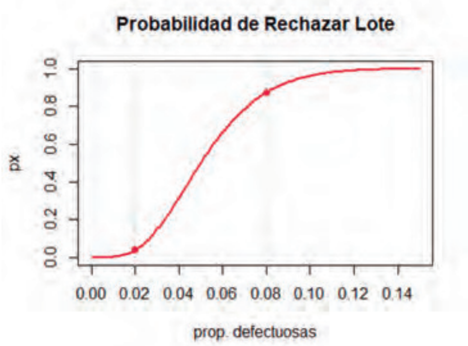
$$c = 93 \times 0.02 + 1.65\sqrt{93 \times 0.02 \times 0.98} \approx 4$$

Plan de muestreo simple por atributos

```
planSimple = function(pa, pr, alfa = 0.05, beta = 0.05)
  #
  # Plan de muestreo simple por atributos
  #
  # pa : AQL # pr : RQL   # alfa : riesgo del vendedor # beta : riesgo del comprador
  #
  # RESULTADOS: n = tamaño de muestra
  #             c = máximo número de piezas defectuosas para aceptar un lote
{         zalfa = qnorm (1-alfa ,0 ,1)
          zbeta = qnorm (beta, 0, 1)
          n = ( (zalfa*sqrt(pa*(1-pa))-zbeta*sqrt(pr*(1-pr)) )/( pr-pa ) )^2
          c = n*pa + zalfa * sqrt( n*pa*(1-pa) )
          n=round(n)
          c=round(c)
          names(n) = 'n'
          names(c) = 'c'
          sol = c(n,c)
          sol
}
```


Plan de muestreo simple por atributos en R

```
source('planSimple.R')
planSimple(pa = .02, pr=0.08, alfa = 0.05, beta = 0.10)

##  n  c
## 93  4

p = seq(0,.15,0.001)
px = 1-pbinom(4,size=93,prob=p)
plot(p,px,type='l',lwd=2,col='red',
xlab='prop. defectuosas',
main='Probabilidad de Rechazar Lote')
```



Distribución normal multivariante

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}; \quad \mu = E[X] = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}; \quad \mathbf{M} = Var[X] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix}$$

$$f(x_1, x_2, ..., x_n) = \frac{1}{(2\pi)^{n/2} |M|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \mathbf{M}^{-1} (\mathbf{x} - \mu) \right\}$$

$$\mathbf{x} = (x_1, x_2, ..., x_n)^T \in \Re^n$$

$$\mu = (\mu_1, \mu_2, ..., \mu_n)^T \in \Re^n$$

$\mathbf{M} \in$ Matriz $n \times n$ semidefinida positiva

Distribución normal bivariante

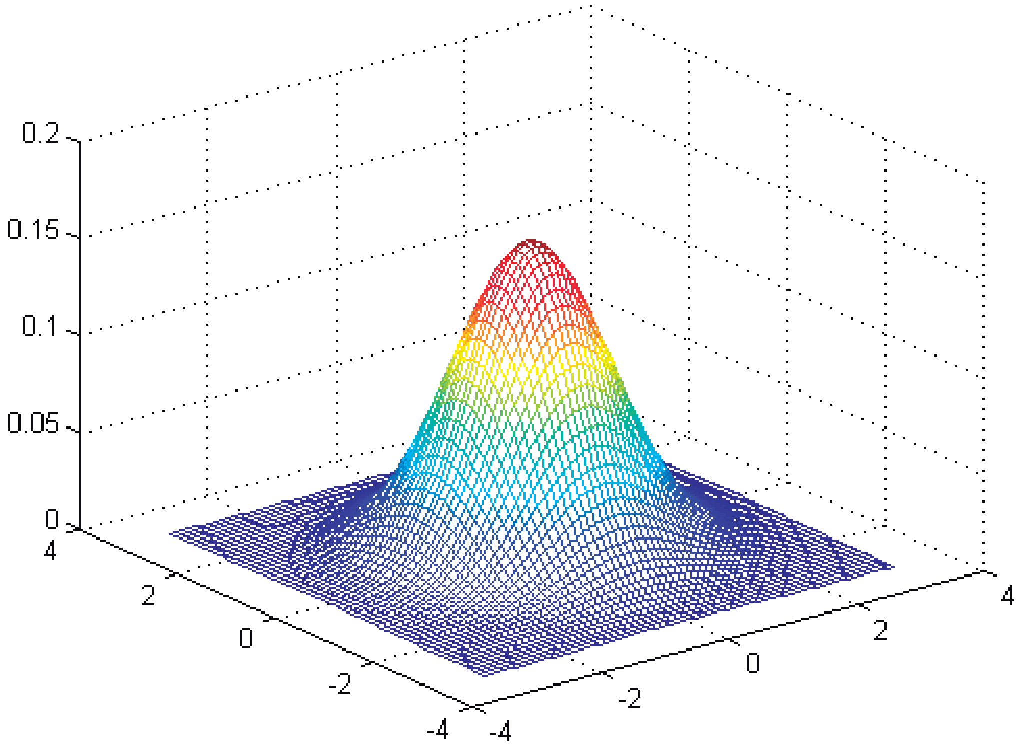
$$f(x_1, x_2) = \frac{1}{2\pi |M|^{1/2}} \exp\left\{ -\frac{1}{2} (x_1 - \mu_1 \; x_2 - \mu_2) \mathbf{M}^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right\}$$

$$\mathbf{x} = (x_1, x_2)^T \in \Re^2 \qquad \mathbf{M} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix};$$

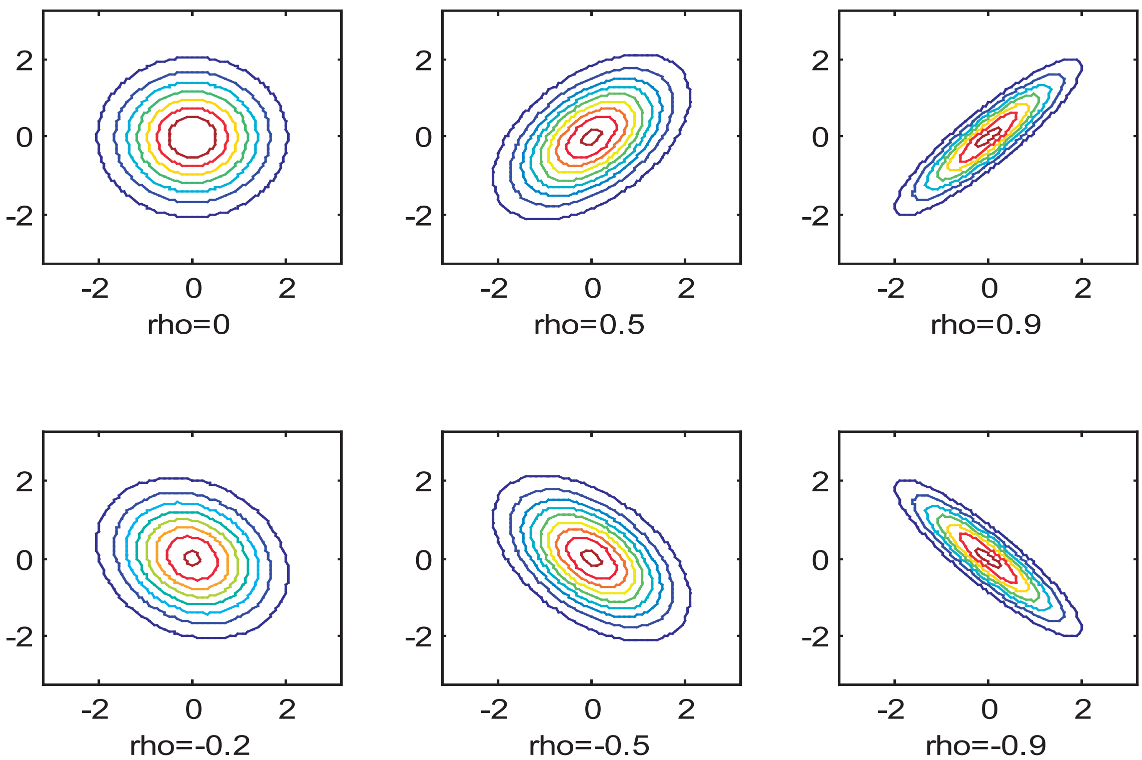
$$\mu = (\mu_1, \mu_2)^T \in \Re^2$$

$$|\mathbf{M}| = \sigma_1^2 \sigma_2^2 (1 - \rho^2), \quad \mathbf{M}^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1 \sigma_2} \\ \frac{-\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

$$\boxed{f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right] \right\}}$$



0.2
0.15
0.1
0.05
0
4
2
0
-2
-4
-4
-2
0
2
4



rho=0
rho=0.5
rho=0.9
rho=-0.2
rho=-0.5
rho=-0.9

Propiedades

- Las dist. marginales son normales $N(\mu_i, \sigma_i)$.
- Las dist. condicionadas son normales.
- $\rho=0 \Leftrightarrow$ Las variables son independientes
- Transformaciones lineales:

$$\boldsymbol{Y} = \boldsymbol{AX}$$

$$\boldsymbol{X} \text{ es } N(\mu, \boldsymbol{M}) \Rightarrow \boldsymbol{Y} \text{ es } \boldsymbol{N(A\ \mu, AMA^T)}$$

Ejemplo

Sea la (X_1, X_2, X_3) normal tridimensional de media $(10, 20, 30)$ y matriz de varianzas

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

¿$P(X_1 + X_2 \geq X_3 + 1)$?

$$Y = X_1 + X_2 - X_3 \rightarrow Normal \begin{cases} E[Y] = 10 + 20 - 30 = 0 \\ Var[Y] = Var(X_1) + Var(X_2) + Var(X_3) = 6 \end{cases}$$

$$\begin{aligned} P(X_1 + X_2 \geq X_3 + 1) &= P(X_1 + X_2 - X_3 \geq 1) \\ &= P(Y \geq 1) \\ &= P\left(\frac{Y-0}{\sqrt{6}} \geq 1/\sqrt{6}\right) \\ &= 1 - \Phi\left(1/\sqrt{6}\right) \end{aligned}$$

Distribuciones de Probabilidad en R

Cada distribución de probabilidad tiene cuatro funciones. Cada una tiene un nombre que la identifica, por ejemplo en el caso de la distribución normal, el nombre es *norm*. Al nombre se le añade un prefijo de una letra:

p para probabilidad *p=pnorm(x)*, $p = P(X \leq x)$
q para la inversa, *q = qnorm(p)* si $p = P(X \leq q)$
d para la función de densidad *fx = dnorm(x)*
r para generar valores aleatorios *x=rnorm(n)* genera *n* valores aleatorios

pnorm(x) se utiliza para la normal estándar, para otra normal hay que indicarles los parámetros, por ejemplo, *p=pnorm(x, mean = 100, sd= 5).* Lo mismo para las otras tres funciones.

Para las distribuciones dicretas, por ejemplo la binomial,
p=pbinom(x,size=100,prob=.5) proporciona $P(X = x)$.
Y *k=qbinom(q,size=100,prob=.5)*es mínimo valor k tal que $P(X \leq k)] \geq q$.

Principales Funciones de Distribución

Beta	`pbeta qbeta dbeta rbeta`
Binomial	`pbinom qbinom dbinom rbinom`
Cauchy	`pcauchy qcauchy dcauchy rcauchy`
Chi-Square	`pchisq qchisq dchisq rchisq`
Exponential	`pexp qexp dexp rexp`
F	`pf qf df rf`
Gamma	`pgamma qgamma dgamma rgamma`
Geometric	`pgeom qgeom dgeom rgeom`
Hypergeometric	`phyper qhyper dhyper rhyper`
Logistic	`plogis qlogis dlogis rlogis`
Log Normal	`plnorm qlnorm dlnorm rlnorm`
Negative Binomial	`pnbinom qnbinom dnbinom rnbinom`
Normal	`pnorm qnorm dnorm rnorm`
Poisson	`ppois qpois dpois rpois`
Student t	`pt qt dt rt`
Uniform	`punif qunif dunif runif`
Weibull	`pweibull qweibull dweibull rweibull`

Capítulo 4: Modelos de Probabilidad

1. Calcular la probabilidad de obtener al menos un seis al lanzar cuatro veces un dado.
2. Se lanza una pareja de dados 24 veces, calcular la probabilidad de por lo menos obtener una pareja de seises. (Este problema tiene interés histórico, fue resuelto por Pascal en el siglo XVIII para ayudar a un jugador llamado De Meré).
3. Una urna contiene 6 bolas, cuatro blancas y dos negras. Se extraen seis bolas con reposición, ¿Cuál es la probabilidad de que se obtengan 4 bolas blancas y dos negras? ¿Cuál es la probabilidad de que hayan extraído más bolas negras que blancas? ¿Cuál es la probabilidad de que hayan extraído más bolas blancas?
4. De un lote que contiene 200.000 piezas se extraen al azar 250. Si el número de piezas defectuosas en la muestra es mayor que c se rechaza el lote y se devuelve al proveedor. Calcular c si se desea que la probabilidad de rechazar un lote con 3 % de piezas defectuosas sea 0.05. (Nota: suponer que se toma la muestra con remplazamiento para hacer los cálculos).
5. Si las llamadas telefónicas a una centralita siguen una distribución de Poisson de parámetro $\lambda = 3$ llamadas/cinco minutos, calcular la probabilidad de:

   a) Seis llamadas en cinco minutos.

   b) Tres llamadas en diez minutos.

   c) Más de 15 en un cuarto de hora.

   d) Dos en un minuto.

6. La variable aleatoria X tiene distribución exponencial con media 1. Obtener la función de distribución y la función de densidad de

$$W = aX^{1/b}, \quad a > 0, b > 0$$

7. El número de averías diarias de una máquina sigue una distribución de Poisson de media 0.4 averías. Calcular la probabilidad de que haya tres días sucesivos sin averías.
8. A un puesto de servicio llegan de manera independiente, por término medio, 10 clientes/hora. Calcular la probabilidad de que lleguen 8 clientes en la próxima media hora sabiendo que en la última hora llegaron 14 clientes, y que la variable aleatoria *número de clientes que llegan en un hora* siguen una distribución de Poisson.
9. Sean X e Y dos variables independientes con distribución geométrica de parámetros 0.4 y 0.6 respectivamente. Calcular $P(X + Y = 3)$.
10. En una planta industrial (ver **??**) dos bombas B_1 y B_2 en paralelo conducen agua desde un pozo a una depuradora D, y posteriormente otras dos bombas B_3 y B_4, también en paralelo, la trasladan a un depósito como indica la figura.

    Los tiempos de vida de la depuradora y de las bombas son variables aleatorias independientes con distribución exponencial, siendo 20 mil horas la vida media de la depuradora y 30 mil horas la de cada bomba.

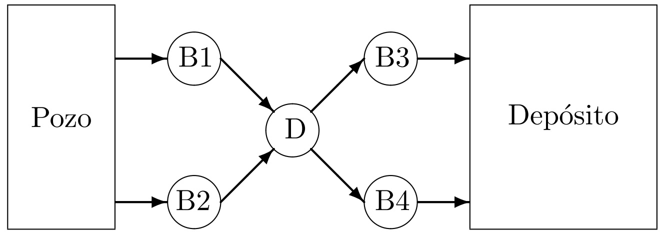


Figura 1: Planta industrial de depuración

a) Calcular la probabilidad de que llegue agua al depósito después de 20 mil horas de funcionamiento.

b) Calcular la probabilidad de que una depuradora que ha trabajado T horas falle antes de las mil horas siguientes. ¿Es razonable que para evitar fallos de la depuradora se renueve ésta cada 20 mil horas? ¿Por qué?

11. Un laboratorio de análisis realiza pruebas de sangre para detectar la presencia de un tipo de virus. Se sabe que una de cada 100 personas es portadora del virus. Se va a realizar un estudio en un colegio, para abaratar las pruebas se realiza un análisis combinado que consiste en: En lugar de analizar la sangre de cada individuo, se toman las muestras de 50 y se analiza la mezcla. Si el resultado del análisis es negativo, se concluye que los 50 individuos están sanos. Si el análisis es positivo, se repite a cada persona de manera individual. El análisis es infalible.

    a) Determinar el número esperado de pruebas (análisis) que se tendrá que realizar si se sigue este tipo de estrategia.

    b) ¿Cuál es la probabilidad de que un individuo determinado sea portador del virus, si el resultado del análisis realizado a su grupo de 50 ha resultado positivo?

12. De un lote con una proporción de piezas defectuosas p, se extraen piezas con reposición hasta que se observa la $k-$ésima defectuosa. Obtener la distribución de probabilidad de la variable aleatoria X *número total de piezas observadas.*

13. La función de densidad de una variable aleatoria X viene dada por la expresión

    $$f(x) = \begin{cases} x/8, & \text{si } 0 \leq x \leq 4 \\ 0, & \text{en el resto} \end{cases}$$

    Se generan secuencialmente valores de esta variable. ¿Cuántos valores de X habrá que generar por término medio hasta obtener un valor mayor que 3?

14. Una pareja decide tener hijos hasta el nacimiento de la primera niña. Calcular la probabilidad de que tengan más de 4 hijos. (Supóngase $P(niño) = P(niña) = 0.5$)

15. La distancia D entre dos vehículos consecutivos en una autopista sigue una distribución exponencial con media 200 metros. ¿Cuál es la probabilidad de que en un tramo de 1 km haya exactamente 5 vehículos?

16. Ricardo es un pescador experto que ha comprobado, después de una larga experiencia practicando su deporte favorito, que el número de peces capturados por la mañana puede ser representado por una variable aleatoria de Poisson de media 3 peces a la hora. Quiere ir a pescar el sábado próximo, si empieza a las 7 de la mañana, ¿cuál es la probabilidad de que capture el primer pez antes de las 7 h. 15 min.? ¿Cuál es la probabilidad de que capture 5 peces durante dos horas de pesca?

17. La variable aleatoria T representa la duración de vida de un componente electrónico. En teoría de la fiabilidad la probabilidad de que un componente falle en el instante t sabiendo que ha durado hasta t se denomina *tasa de fallo* y se representa por $\lambda(t)$, siendo su valor en función de t

$$\lambda(t) = \frac{f(t)}{1 - F(t)},$$

donde f y F son, respectivamente, las funciones de densidad y de distribución de la variable aleatoria T. Obtener la tasa de fallo en caso que T sea una variable aleatoria exponencial de media 1000 horas e interpretar el resultado.

18. Un examen consiste en 25 cuestiones. En cada cuestión, el alumno debe elegir entre 5 soluciones propuestas, de las que una (y sólo una) es cierta. El número mínimo de respuestas correctas que debe tener un alumno para aprobar es a. El profesor decide fijar a con el siguiente criterio: que la probabilidad de aprobar para un alumno que conteste todas las cuestiones al azar sea menor de 0.05. Obtener a. (Una cuestión es respondida al azar si cada uno de los cinco resultados propuestos tiene la misma probabilidad de ser escogido).

19. Obtener la función de densidad de una variable aleatoria χ^2 con un grado de libertad. (Si $X \rightsquigarrow N(0,1), Y = X^2$ es una χ^2_1.)

20. Dada una variable aleatoria X, cuya distribución es $N(0, \sigma^2)$, calcular la mediana de la variable $Y = |X|$.

21. La longitud L en milímetros de las piezas fabricadas en un proceso es una variable aleatoria que se distribuye según una $N(32, 0.3)$, considerándose aceptables aquellas cuya medida se encuentra dentro del intervalo (31.1, 32.6).

    a) Calcular la probabilidad de que una pieza elegida al azar sea aceptable.
    b) Si se toma al azar una muestra de tres piezas, ¿cuál es la probabilidad de que la primera y la tercera sean aceptables y la segunda no lo sea?
    c) ¿Cuál es la probabilidad de que en una muestra de tamaño 3 al menos una sea aceptable?
    d) Las piezas se embalan en lotes de 500. Calcular la probabilidad de que un lote tenga más de 15 defectuosas.

22. Un concesionario de automóviles recibe pedidos de un modelo según un proceso de Poisson de media 2 vehículos por semana. Los pedidos al fabricante se deben realizar con una antelación mínima de un mes, de forma que el concesionario pide en cada mes los vehículos que necesita para el mes siguiente. ¿Cuántos automóviles disponibles ha de tener a principios de un mes para satisfacer con probabilidad igual o mayor que 0.95 la demanda mensual? (Se considera que el mes tiene cuatro semanas).

23. Si la probabilidad de que un disparo impacte una diana es 0.0001, ¿cuál es la probabilidad de impactar en la diana 4 o más veces en 50000 disparos? Da un resultado numérico empleando la aproximación que consideres más adecuada. Se supone independencia.

24. En una urna hay 20 bolas, 5 son negras y 15 blancas. Se extraen 4 sin reposición y se define la variable aleatoria Y como el número de bolas negras.

a) Demostrar que la distribución de probabilidad de Y es

$$P(Y=k) = \frac{\binom{5}{k}\binom{15}{4-k}}{\binom{20}{4}}, \quad k = 0,1,2,3,4$$

b) Generalizar el resultado anterior si en la urna hay M bolas blancas, N bolas negras, se extraen K bolas e interesa conocer la distribución de probabilidad de Y número de bolas negras. La extracción es sin reposición. (Esta distribución se denomina *Hipergeométrica* y se utiliza en los problemas de muestreo sin reposición)

c) Plantear el problema del número de aciertos en la lotería primitiva como un caso particular de la distribución anterior. Aplicarlo para el caso en el que el participante tacha (elige) 6 números y en el caso en el que tacha 8 números.

25. Para controlar la calidad de un proceso textil se cuenta el número de defectos que aparecen en la tela fabricada. Según el fabricante, cuando el proceso funciona correctamente el número de defectos en una bobina de 100 metros cuadrados es una variable aleatoria de Poisson con media 4. Se ha instalado un equipo de visión artificial para realizar el recuento que permite inspeccionar 900 m^2 de tela cada hora. ¿Cuál es la probabilidad de que aparezcan más de 50 defectos en una hora si el proceso funciona bien?

26. Un compañía compra *chips* para montar en placas de ordenadores clónicos. Una empresa de reciclado le ofrece lotes de 10.000 chips a precios muy ventajosos pero con un porcentaje de defectuosos alto, alrededor del 10 %. Para realizar el control de calidad de los lotes recibidos está considerando dos alternativas: (a) Tomar 100 unidades al azar y rechazar el lote si existen más de 15 defectuosas. (b) Tomar 100 unidades al azar, dividirlas en 10 grupos y si en algún grupo hay más de una pieza defectuosa rechazar el lote.¿Cuál es la probabilidad de rechazar un lote con el 10 % de chips defectuosos para cada uno de los métodos? ¿Cuál es la probabilidad de aceptar un lote con el 20 % de chips defectuoso para cada método?

27. Una compañía para comprobar la calidad de ciertos lotes de 30000 piezas realiza el siguiente control: toma una muestra al azar de 300 piezas y si tiene 15 o más piezas defectuosas rechaza el lote, aceptándolo en caso contrario. La compañía cada mes aplica este control a 200 lotes, ¿cuál es el número esperado de lotes rechazados si todos los lotes de un mes tienen exáctamente un 4 % de piezas defectuosas?

28. Un servicio telefónico de urgencias recibe por término medio 10 llamadas cada minuto, ¿cuál es la probabilidad de recibir más de 550 llamadas en una hora? Se ha diseñado un "*call center*" con capacidad de respuesta de 650 llamadas a la hora, ¿cuál es el número esperado de horas al año con número de llamadas superior a su capacidad? (Se supone que las llamadas son independientes y todas las horas son similares)

29. A un congreso de medicina acuden 500 personas. Un laboratorio farmaceútico va a regalar corbatas a los hombres y pañuelos a las mujeres. Desgraciadamente no conocen el número exacto de cada sexo, aunque saben de otros congresos que la proporción es similar. Calcula el número mínimo de corbatas y de pañuelos que deben tener disponibles los organizadores para que todos los asistentes tengan el regalo que les corresponde con probabilidad de 0.99 (es decir ninguna mujer se quede sin pañuelo y ningún hombre sin corbata). Se supone que la probabilidad de hombre o mujer es igual a 0.5 y que la probabilidad de que un asistente sea de un determinado sexo es independiente del sexo de los restantes.

30. Federer y Nadal se encuentran empatados, 40-40 en un juego en el que está sacando Nadal. Según las estadísticas la probabilidad de que Nadal gane un punto determinado cuando tiene el saque es 0.6. ¿Cuál es la probabilidad de que el juego lo termine ganando Nadal? (Nota. Piensa en el desempate de la siguiente forma: se juegan dos puntos: si los gana un jugador ese jugador ha ganado el juego, si cada jugador gana un punto se juegan otros dos puntos y se vuelve a aplicar la misma regla).

31. *"Cibeles in Concert"* es una empresa que organiza viajes en autobus para asistir a actuaciones musicales. Para un concierto de *Bruce Springsteen* en Paris ha ofertado 300 plazas que salen de Madrid y Sevilla. Las reservas se hacen por Internet, el precio del viaje para los de Sevilla es de 60 euros y para los de Madrid 50 euros . Las 300 plazas se cubren con seguridad. Si la probabilidad de que un asistente salga de Sevilla es 1/3 y de Madrid 2/3, y se acepta independencia entre las 300 reservas, calcula los ingresos esperados por la compañía y la varianza de estos ingresos.

32. Una empresa de celulosa tiene dos líneas para fabricar pasta de papel en planchas de $1\text{m} \times 1\text{m}$. Una medida de su calidad es la limpieza, que se mide en número de impurezas (partículas) por m^2. La línea I, fabrica con una tasa media de 5 impurezas por m^2 y la línea II con 3 impurezas por m^2. El número de impurezas por plancha es una variable aleatoria con distribución de Poisson. Las planchas de pasta se empaquetan en balas de 2000 unidades y se almacenan clasificadas según la línea de procedencia. Cuando por algún motivo se encuentra en el almacén una bala sin clasificar se adopta el siguiente criterio: tomar una muestra aleatoria de 10 planchas y determinar el número medio de impurezas. Asignarla al grupo de la línea I si el número medio de impurezas es mayor que 4 y a la línea II en caso contrario. (Se supone que la probabilidad inicial de pertenecer a una u otra linea es la misma).

    a) Calcular la probabilidad de clasificar erróneamente una bala.

    b) En un caso concreto, el número de impurezas en cada una de las diez planchas han sido $1, 2, 3, 3, 4, 4, 5, 5, 6, 8$. Calcular la probabilidad de que pertenezca a cada una de las líneas.

33. Los billetes de banco son fabricados en pliegos. La impresión se realiza por dos máquinas iguales, una de ellas imprime el anverso y la otra el reverso. Sea X e Y, respectivamente, el número de defectos de impresión en el anverso y reverso de un pliego. Ambas variables son independientes con distribución de Poisson de parámetros λ_1 y λ_2.

    a) Demostrar que el número total de defectos en un pliego $Z = X + Y$ tiene distribución de Poisson. (Nota.- Utilizar que

    $$\mathrm{P}\{Z = n\} = \sum_{k=0}^{n} \mathrm{P}\{X = k\}\mathrm{P}\{Y = n - k\}$$

    y el desarrollo del binomio de Newton para $(\lambda_1 + \lambda_2)^n$.)

    b) Si el número total de defectos en un pliego es $Z = n$, ¿ cuál es la probabilidad de que haya exáctamente $X = k$ defectos en el anverso? (Obtener la expresión en función de λ_1, λ_2, n y k). ¿ De qué distribución de probabilidad se trata?

34. La llegada de los clientes a un banco se considera un proceso Poisson con parámetro λ. Sabiendo que en la última hora han llegado 2 clientes, ¿cuál es la probabilidad de que los dos entraran en los primeros 15 minutos?

35. Un equipo de radio tiene dos partes, el receptor y el amplificador. La duración del receptor es una variable aleatoria exponencial de media 500 horas y la duración del amplificador una variable exponencial de media 1000 horas. ¿Cuál es la probabilidad de que el fallo del equipo (cuando se produzca) sea debido a un fallo del receptor? (Se supone que las variables son independientes)

36. Sea X_1 una variable aleatoria N(10,1), X_2 una variable aleatoria N(20,1), y X_3 una variable aleatoria N(30,4). Se define

$$Z_1 = X_1 + X_2 - X_3$$
$$Z_2 = X_1 + X_2 + X_3$$
$$Z_3 = X_1 - X_2 - X_3$$

Si X_1, X_2, X_3 son independientes, calcular la matriz de varianzas de (Z_1, Z_2, Z_3).

37. Una oficina de correos tiene dos ventanillas de atención al público. Tres personas A,B y C llegan en el mismo instante a la oficina de correos y encuentran las dos ventanillas desocupadas. Los tiempos de servicio requeridos por las tres personas son variables aleatorias independientes con distribución exponencial de parámetro λ. Los tiempos de servicio de A y B comienzan de inmediato, mientras que C debe esperar a que termine el primero de los dos. ¿Cuál es la probabilidad de que C no sea el último en salir de la oficina de correos?

38. En cierta fabricación mecánica el 96 % de las piezas resultan con longitudes admisibles (dentro de tolerancias), un 3 % son piezas defectuosas cortas y un 1 % son defectuosas largas. Calcular la probabilidad de:

   a) En un lote de 250 piezas sean admisibles 242 o más.
   b) En un lote de 500 sean cortas 10 o menos.
   c) En 1000 piezas haya entre 6 y 12 largas. Todas las aproximaciones se calculan la distribución normal.

39. La estatura de los ciudadanos varones de un país sigue una distribución normal:

$$N(\mu = 175cm, \sigma = 5cm);$$

si se seleccionan al azar 100 ciudadanos, ¿cuál es la probabilidad de que al menos 30 superen los $180cm$?

40. En un proceso de fabricación de película fotográfica aparecen por término medio 1 defecto de cada 20 metros de película. Si la distribución de defectos es Poisson, calcular la probabilidad de 6 defectos en un rollo de 200 metros de pelicula (a) directamente; (b) utilizando la aproximación normal.

41. De un lote que contiene 200.000 piezas se extraen al azar 250. Si el número de piezas defectuosas en la muestra es mayor que c se rechaza el lote y se devuelve al proveedor. Calcular c si se desea que la probabilidad de rechazar un lote con 3 % de piezas defectuosas sea 0.05. (Nota: suponer que se toma la muestra con remplazamiento para hacer los cálculos y utilizar la aproximación normal para obtener c).

42. Para controlar la recepción de lotes de 10000 unidades de discos compactos se toma una muestra al azar de $n = 200$ discos clasificándolos como aceptables y defectuosos. Si el número de discos defectuosos es igual o inferior a $c = 15$ se acepta el lote, en caso contrario se rechaza. ¿Cuál es la probabilidad de aceptar un lote con el 12 % de discos defectuosos? ¿Cuál es la probabilidad de rechazar un lote con el 5 % de discos defectuosos? ¿Que valores de n y c deben utilizarse si se desea que las probabilidades anteriores sean iguales a 0.05?

43. La función de densidad conjunta de las variables aleatorias X, Y es la siguiente normal bidimensional:

$$f(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

Si $\rho = 0.3$, calcular $\Pr(Y \leq X + 1)$.

44. Un inversor tiene su dinero repartido en acciones de dos compañías. La rentabilidad (% de beneficio) de las compañías pueden ser consideradas como dos variables aleatorias, con media anual igual para las dos del 10 %, aunque el riesgo es muy diferente. Una tiene una desviación típica de 2.5 % y la otra del 1 %. Además se sabe que la correlación entre ellas es -0.50. ¿Qué proporción debe invertir en cada una para que el riesgo sea mínimo? (Nota: Mínimo riesgo es lo mismo que mínima varianza. $X \to N(10, 2.5)$ y $Y \to N(10, 1)$, $corr(X, Y) = -0.5$).

45. Cierto compuesto metálico se ha sometido durante una hora a una atmósfera de oxígeno a 200 grados centígrados. Una medida de su corrosión es la ganancia de peso durante este tiempo. Se ha comprobado que para un determinado compuesto esta ganancia X_1, en una hora, se distribuye segúun una normal $N(100, 5)$.

    a) Si se realizan los ensayos de manera secuencial, ¿cuántos se tendrán que hacer por término medio para encontrar una probeta con una ganancia mayor que 105?

    b) Se comprueba que sometiendo la probeta al ensayo durante dos horas, la ganancia en la segunda hora X_2 tiene una distribución normal $N(60, 5)$, siendo el coeficiente de correlación entre las variables X_1 y X_2, $\rho = -0.28$. Calcular la probabilidad de que una probeta tenga mayor ganancia de peso en la segunda hora que en la primera.

    c) Se forman los índices de oxidación $Z_1 = X_1 + X_2$ y $Z_1 = X_1 - X_2$. Calcular la función de densidad conjunta de estas dos nuevas variables. ¿Son independientes?

    d) Si la ganancia durante las dos horas ha sido 170, ¿cuál es el valor medio de la ganancia en la primera hora?

46. El abastecimiento de energía eléctrica de una comarca depende de tres centrales: una nuclear, una térmica de carbón y una hidráulica con potencias instaladas de 500 MW, 300 MW y 200 MW, respectivamente. Desde el punto de vista de fiabilidad, cada central sólo puede estar en uno de estos dos estados: disponible (con toda su potencia) o averiada (con potencia cero). En un día, la probabilidad de avería de la central nuclear es 0.10, de la térmica es 0.12 y de la hidráulica 0.05. Las averías son independientes y supondremos como hipótesis simplificadora que a lo largo de un día la central no cambia de estado.

    a) Para un día, calcular la función de distribución de la variable aleatoria X = “Potencia disponible en la comarca”.

b) Un pais tiene 30 centrales de cada uno de los tipos del apartado 1. Calcular la probabilidad de que en un día la potencia disponible en el pais sea menor que 24000 MW. (Utilizar la aproximación normal).

c) La potencia máxima diaria demandada en el pais es una variable aleatoria con distribución normal de media 23000 MW y desviación típica 1000 MW. ¿Cuál es la probabilidad de que en un día la demanda sea superior a la potencia disponible?

47. Una compañía desea aplicar un plan de muestreo para controlar la compra de lotes de 10000 unidades. La capacidad de inspección máxima que tienen es de 200 piezas. Determinar c, "*el número máximo de piezas defectuosas en la muestra que debe tener un lote aceptado*"si se desea que la probabilidad de rechazar un lote con el 10 % de piezas defectuosas sea igual 0.02. Un proveedor estima que sus lotes tienen un 5 % de piezas defectuosas, ¿qué probabilidad tiene de que un lote suyo sea aceptado?

48. Se tienen dos variables aleatorias independientes U y V, ambas con distribución uniforme en [0,1]. A partir de ellas se definen las variables aleatorias Z y W:

$$\begin{cases} Z = \sqrt{-2\log U} \\ W = 2\pi V \end{cases}$$

a) Calcular las probabilidades $P(Z < z_0)$ y $P(W < w_0)$. Utilizar estos resultados para deducir la expresión de la función de densidad de probabilidad conjunta $f_{Z,W}(z,w)$ para las variables aleatorias Z y W, así como el domino donde están definidas.

b) Una vez caracterizadas las variables aleatorias Z y W por su función de densidad de probabilidad conjunta $f_{Z,W}(z,w)$ se vuelve a definir 2 nuevas variables aleatorias X e Y mediante la siguiente transformación: $\begin{cases} X = Z \cdot \cos W \\ Y = Z \cdot senW \end{cases}$

Para estas nuevas variables se comprueba que la función de densidad de probabilidad conjunta es:

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(x^2+y^2\right)\right), \text{ con } (x,y) \in \mathbb{R}^2.$$

Demostrar que las variables aleatorias X e Y son independientes. Calcular la media y la varianza de las variables y sus funciones de densidad marginal identificando el tipo de variables de las que se trata.

c) Con la información del apartado (a) y (b) obtener el valor numérico de las siguientes probabilidades:

- $P(-1 < X < 1, -1 < Y < 1)$
- $P(X^2 + Y^2 < 1)$
- $P(Y < X)$

d) Finalmente, calcular los coeficientes de la siguiente tansformación lineal

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c & 0 \\ d & e \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

con $c, e > 0$, para que las nuevas variables aleatorias $(X', Y')^T$ tengan una distribución de probabilidad con vector de medias y matriz de varianzas dadas por:

$$\mu = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \qquad S = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}.$$

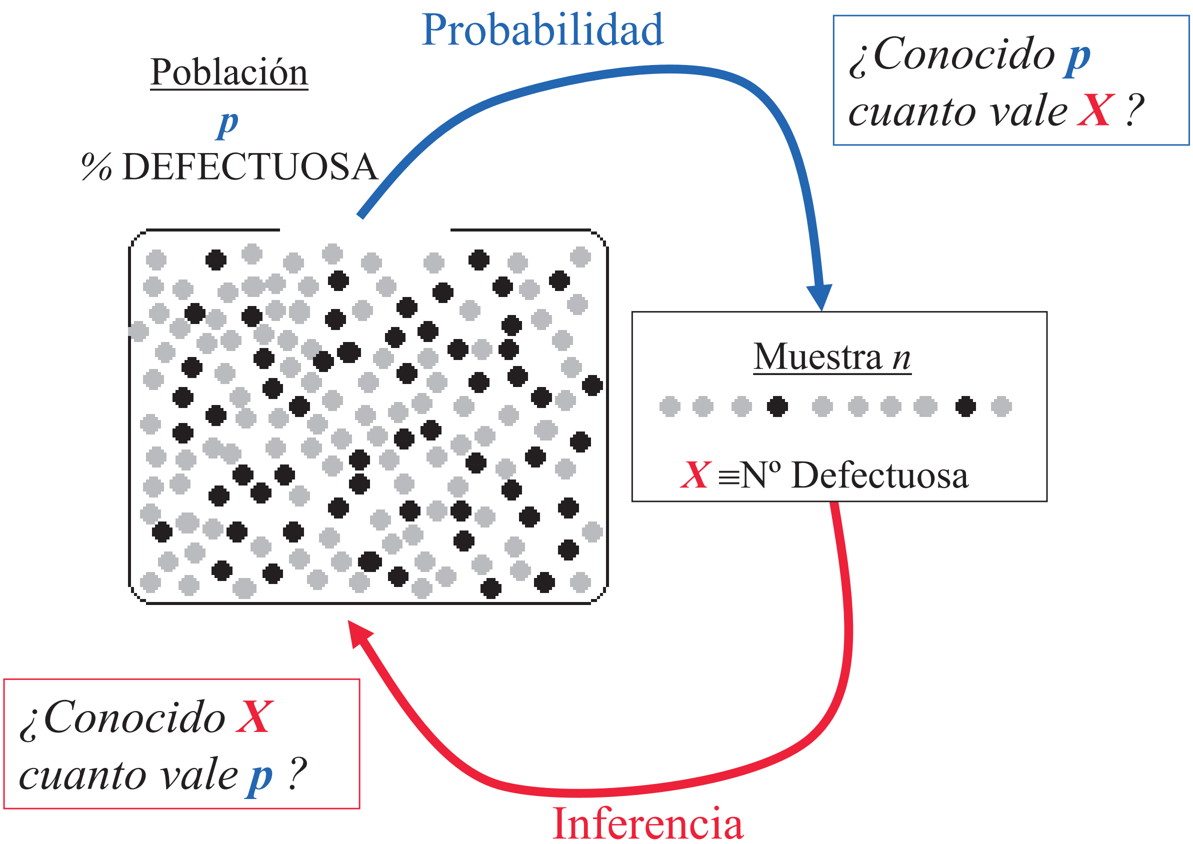
49. Hay un juego muy popular en EEUU que se juega en carnavales en los que un jugador paga 1 dólar, elige un número del 1 al 6 y lanza tres dados. La banca le paga tantos dólares como veces aparece el número elegido. En un día de carnaval participan en el juego 500 personas. Obtén la distribución de probabilidad de las ganancias de la banca para ese día, indicando su media y desviación típica.

50. Se lanzan 100 dados. Obtén la distribución de la suma de los números pares. Calcula la media y la varianza.

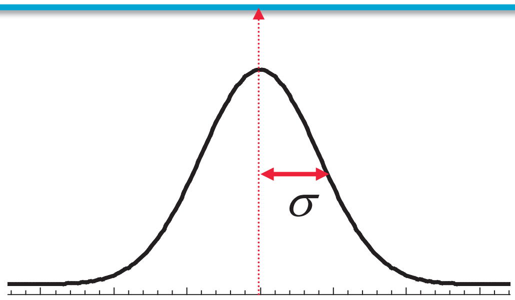
Estadística

Parte II.

Tema 5. Estimación



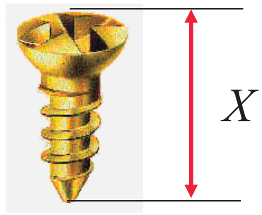
Inferencia



μ

$X \rightarrow N(\mu,\sigma)$

$\boxed{\begin{array}{c}\textit{Parámetros} \\ ¿\mu,\sigma?\end{array}}$



MUESTRA n

$X_1, X_2, \ldots, X_n$

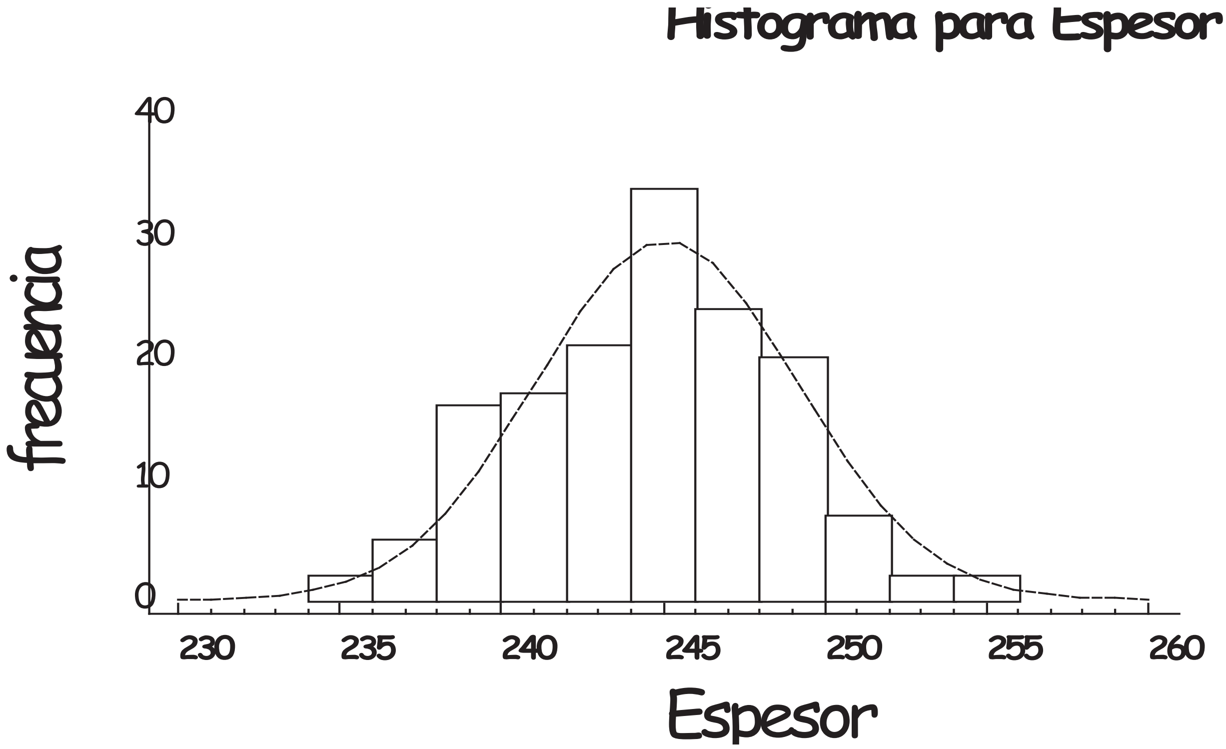
Datos Conocidos

?

Espesores de 150 obleas de Silicio (micras)

240	235	240	240	247	237	243	242	236	239
243	237	243	242	245	239	245	245	239	240
250	246	244	246	255	242	248	248	241	242
253	249	249	249	250	247	251	251	246	243
248	246	246	248	249	245	250	249	242	244
238	240	245	240	237	242	244	242	243	239
242	241	250	243	239	244	246	245	246	240
245	246	250	246	243	246	246	248	247	250
251	247	247	250	247	251	250	243	252	252
247	249	248	248	246	248	246	246	247	250
239	240	238	241	242	243	241	241	241	241
242	243	240	245	244	245	239	244	243	243
246	244	245	243	245	247	244	245	245	249
250	248	248	247	248	252	250	249	248	255
248	245	246	245	245	249	246	247	246	253

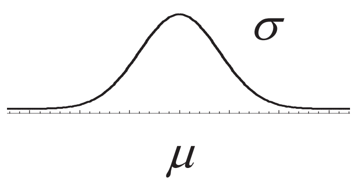
Histograma y función de densidad Normal



Distintos problemas de inferencia

Dado un modelo *para los datos:*

$$x_1, x_2, \ldots, x_n$$

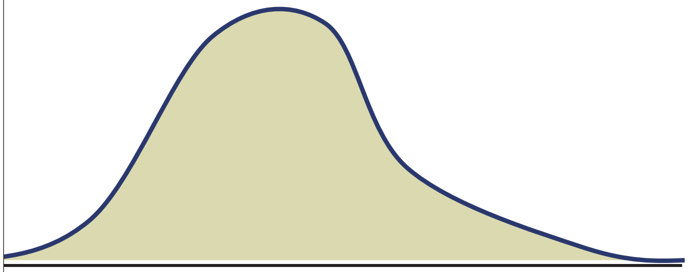


- *Estimar* μ y σ
- Dar un *intervalo de confianza* para μ y σ
- Elegir entre (*contraste de hipótesis*):

$$\mu \geq 250 \quad \text{o} \quad \mu < 250$$

- Comprobar la validez del modelo (*contraste de bondad de ajuste*).

Métodos de Estimación



$$X \rightarrow f_X(x, \theta_1, \theta_2, \ldots, \theta_r)$$

f_X conocida

Parámetros desconocidos

$$\theta_1, \theta_2, \ldots, \theta_r$$

Dada una muestra aleatoria simple de X

$$x_1, x_2, \ldots, x_n \quad \rightarrow \quad ¿\theta_1, \theta_2, \ldots, \theta_r?$$

1. Método de los momentos
2. Método de máxima verosimilitud

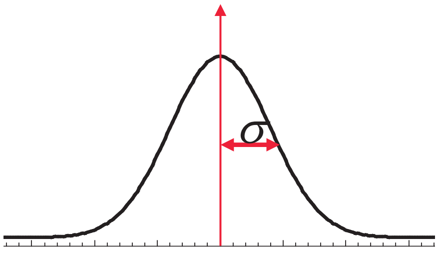
Métodos de los momentos

DATOS	VAR. ALEATORIA
$x_1, x_2, \ldots, x_n$	$f_X(x, \theta_1, \theta_2, \ldots, \theta_r)$
$a_1 = \dfrac{\sum_{i=1}^n x_i}{n}$	$\alpha_1 = E[X] = g_1(\theta_1, \theta_2, \ldots, \theta_r)$
$a_2 = \dfrac{\sum_{i=1}^n x_i^2}{n}$	$\alpha_2 = E[X^2] = g_2(\theta_1, \theta_2, \ldots, \theta_r)$
$\vdots$	$\vdots$
$a_r = \dfrac{\sum_{i=1}^n x_i^r}{n}$	$\alpha_r = E[X^r] = g_r(\theta_1, \theta_2, \ldots, \theta_r)$

Estimadores $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r$

$$\rightarrow \begin{cases} g_1(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r) = a_1 \\ g_2(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r) = a_2 \\ \vdots \\ g_r(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r) = a_r \end{cases}$$

Método de los momentos: Distribución normal



$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}, \quad x \in \Re$$

Parámetros : ¿ μ, σ ?

$x_1, x_2, \ldots, x_n$	$f_X(x, \mu, \sigma)$
$a_1 = \frac{\sum_{i=1}^n x_i}{n}$	$\alpha_1 = E[X] = \mu$
$a_2 = \frac{\sum_{i=1}^n x_i^2}{n}$	$\alpha_2 = E[X^2] = \sigma^2 + \mu^2$

$$\left.\begin{array}{l} a_1 = \alpha_1 \\ a_2 = \alpha_2 \end{array}\right\} \quad \boxed{\begin{array}{c} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2 = s^2 \end{array}}$$

Método de máxima verosimilitud

Una fuente radiactiva emite partículas según un proceso de Poisson con media λ desconocida. Durante 10 minutos se han contado el número de partículas emitidas:

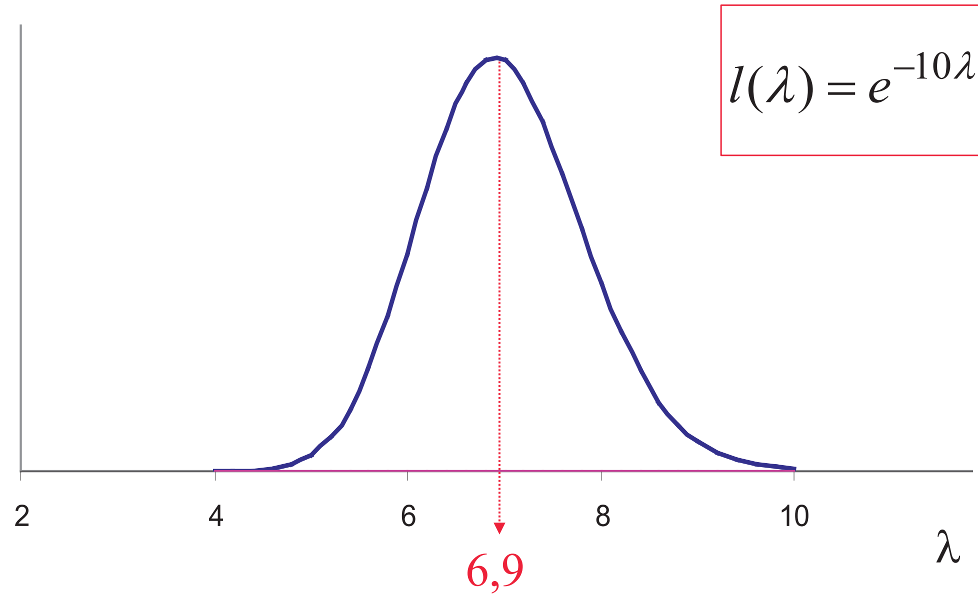
12, 6, 11, 3, 8, 5, 3, 9, 7, 5

$$P(X_1 = 12, X_2 = 6, \ldots, X_{10} = 5) = e^{-\lambda}\frac{\lambda^{12}}{12!} \times e^{-\lambda}\frac{\lambda^{6}}{6!} \times \cdots \times e^{-\lambda}\frac{\lambda^{5}}{5!}$$

$$= e^{-10\lambda}\frac{\lambda^{12+6+\cdots+5}}{12! \times 6! \times \cdots \times 5!} = e^{-10\lambda}\frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}$$

$$\boxed{l(\lambda) = e^{-10\lambda}\frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}}$$

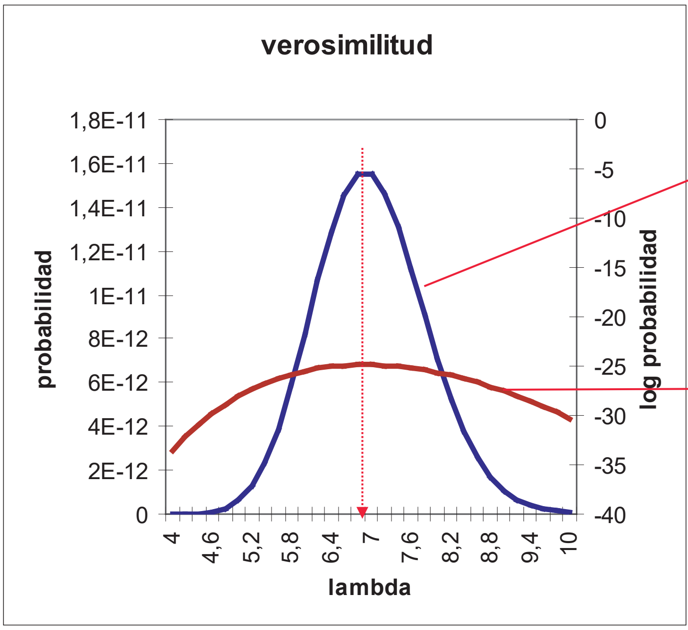
Función de verosimilitud



$$l(\lambda) = e^{-10\lambda} \frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}$$

Estimador máximo-verosímil: $\hat{\lambda} = 6{,}9$

Función de verosimilitud



$$l(\lambda) = e^{-10\lambda} \frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}$$

$$L(\lambda) = \log l(\lambda)$$

Estimador máximo-verosímil: $\hat{\lambda} = 6{,}9$

Estimación por máxima verosimilitud

$$X \to f_X(x,\theta_1,\theta_2,\dots,\theta_r)$$

$$f_X \quad \text{conocida}$$

Parámetros desconocidos: $\theta_1,\theta_2,\dots,\theta_r$

Muestra aleatoria simple: $x_1,x_2,\dots,x_n$

Distribución conjunta:

$$f(x_1,x_2,\dots,x_n;\theta_1,\dots,\theta_r) =$$

$$= f_X(x_1,\theta_1,\theta_2,\dots,\theta_r) f_X(x_2,\theta_1,\theta_2,\dots,\theta_r)\cdots f_X(x_n,\theta_1,\theta_2,\dots,\theta_r)$$

$$\log f(x_1,x_2,\dots,x_n) = \sum_{i=1}^{n} \log f_X(x_i,\theta_1,\theta_2,\dots,\theta_r) = L(\theta_1,\theta_2,\dots,\theta_r)$$

$$\hat\theta_1,\hat\theta_2,\dots,\hat\theta_r \quad \Rightarrow \quad L(\hat\theta_1,\hat\theta_2,\dots,\hat\theta_r) = \max L(\theta_1,\theta_2,\dots,\theta_r)$$

Máx. verosimilitud: Distribución normal

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x\in\Re \qquad \text{Parámetros: ¿}\mu,\sigma\text{?}$$

$x_1,x_2,\dots,x_n$: muestra aleatoria simple

$$f(x_1,x_2,\dots,x_n) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2}\cdots\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2} = \frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2}$$

$$L(\mu,\sigma^2) = \log f(x_1,x_2,\dots,x_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2$$

$$\max L: \left\{\begin{array}{l} \dfrac{\partial L}{\partial\mu} = \dfrac{1}{\sigma^2}\sum_{i=1}^n(x_i-\hat\mu) = 0 \\[2ex] \dfrac{\partial L}{\partial\sigma^2} = -\dfrac{n}{2}\dfrac{1}{\hat\sigma^2} + \dfrac{1}{2\hat\sigma^4}\sum_{i=1}^n(x_i-\hat\mu)^2 = 0 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \hat\mu = \bar x \\ \hat\sigma^2 = s^2 \end{array}\right.$$

Máx. Verosimilitud: Poisson

$$P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}, x=0,1,2,\ldots \rightarrow \quad \text{Parámetro } \lambda$$

Muestra aleatoria simple: $x_1, x_2, \ldots, x_n$

$$p(x_1,x_2,\ldots,x_n)=P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)$$

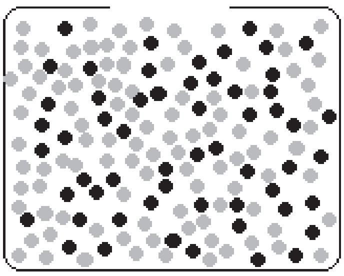
$$=e^{-\lambda}\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}\cdots e^{-\lambda}\frac{\lambda^{x_n}}{x_n!}=e^{-\lambda n}\frac{\lambda^{\Sigma x_i}}{x_1!x_2!\cdots x_n!}$$

$$L(\lambda)=-\lambda n+(\Sigma x_i)\log\lambda-\sum\log x_i!$$

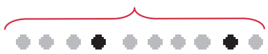
$$\max L(\lambda): \frac{dL(\lambda)}{d\lambda}=-n+\frac{\Sigma x_i}{\hat{\lambda}}=0 \Rightarrow \hat{\lambda}=\bar{x}$$

Máx. verosimilitud: Binomial(*n*,*p*)

Proporción defectuosas = ¿ p ?



n



Muestra (Bernoulli): $x_1, x_2, \ldots, x_n$

$$x_i=\begin{cases}0 & \text{si es } \textbf{Aceptable}\\ 1 & \text{si es } \textbf{Defectuosa}\end{cases}$$

$$p(x_1,x_2,\ldots,x_n)=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)$$

$$=P(X_1=x_1)P(X_2=x_2)\cdots P(X_n=x_n)$$

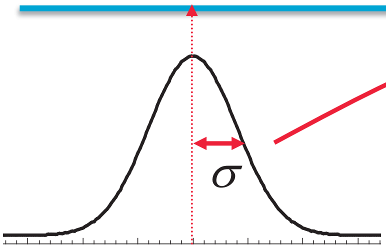
$$=p^{\sum x_i}(1-p)^{n-\sum x_i}$$

$$=p^r(1-p)^{n-r} \quad \text{donde } r=\sum x_i \text{ es el nº de defectuosas}$$

$$L(p)=r\log p+(n-r)\log(1-p)$$

$$\frac{dL(p)}{dp}=\frac{r}{\hat{p}}-\frac{n-r}{1-\hat{p}}=0 \quad \rightarrow \quad \hat{p}=\frac{r}{n}$$

Distribución de media (normal)



$$X \to N(\mu, \sigma)$$

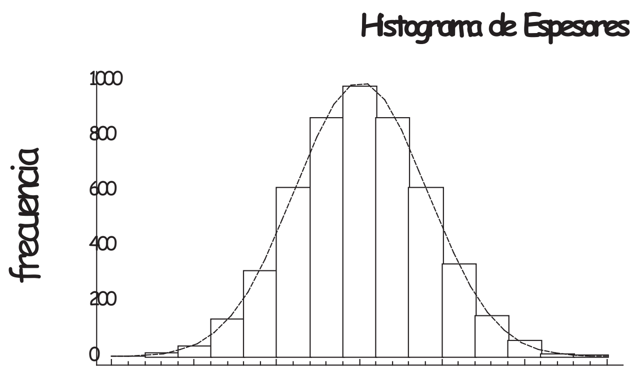
$$X_1, X_2, \ldots, X_n$$

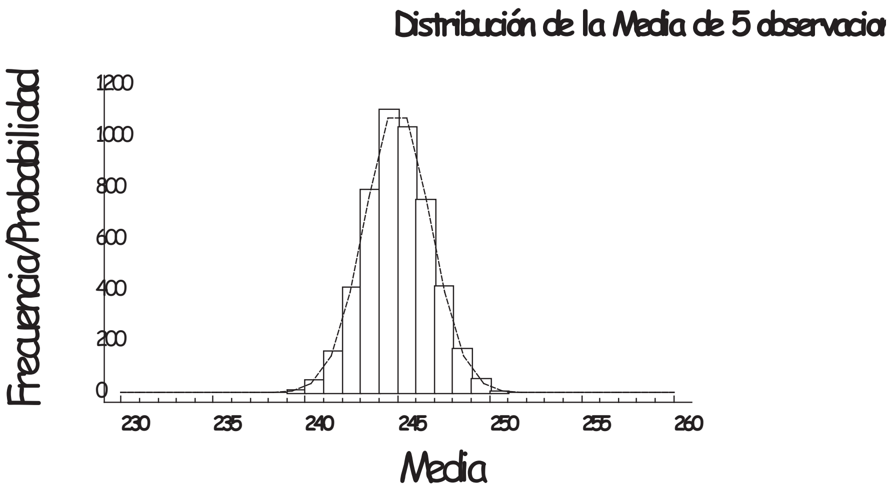
$$\boxed{\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}}$$

$$E[\bar{X}] = \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

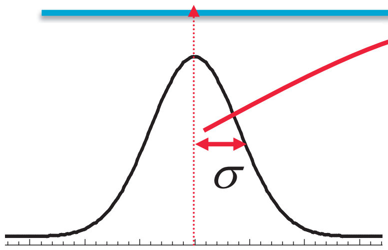
$$\mathrm{Var}[\bar{X}] = \frac{\mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots + \mathrm{Var}[X_n]}{n^2} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\boxed{\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$





Distribución de S^2 (Normal)



$$X_1, X_2, \ldots, X_n$$

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X}}{n}$$

$$X \to N(\mu, \sigma)$$

$$S^2 = \frac{1}{n}\sum_1^n (X_i - \overline{X})^2 = \frac{1}{n}\sum_1^n X_i^2 - \overline{X}^2$$

$$E[S^2] = \frac{1}{n}\Sigma E[X_i^2] - E[\overline{X}^2] = \frac{1}{n}\Sigma(\sigma^2 + \mu^2) - (\frac{\sigma^2}{n} + \mu^2) = \frac{n-1}{n}\sigma^2$$

$$\hat{S}^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}$$

$$\hat{S}^2 = \frac{n}{n-1}S^2 \quad \Rightarrow \quad E[\hat{S}^2] = \sigma^2$$

Distribución χ^2

$$Z_1, Z_2, \ldots, Z_n \to N(0,1) \quad \text{independientes}$$

$$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \to \chi_n^2$$

Propiedades

- $E[\chi_n^2] = n$
- $\mathrm{Var}[\chi_n^2] = 2n$
- $\chi_n^2 + \chi_m^2 = \chi_{n+m}^2 \quad (\chi_n^2 \text{ y } \chi_m^2 \text{ indep.})$

Chi-Cuadrado con 4 gl

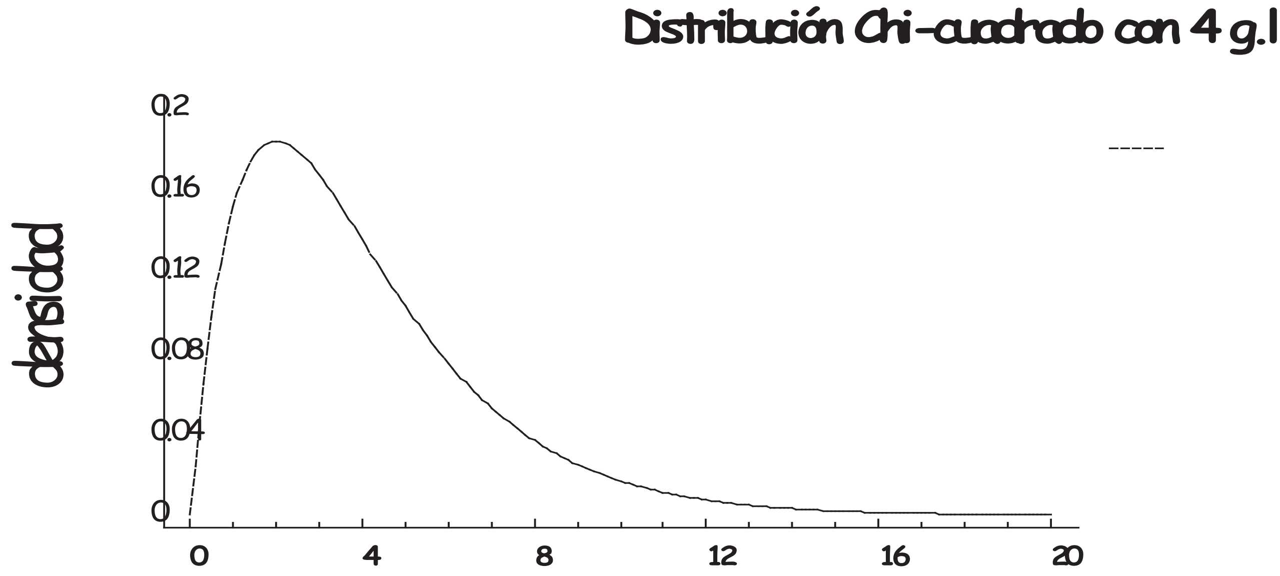
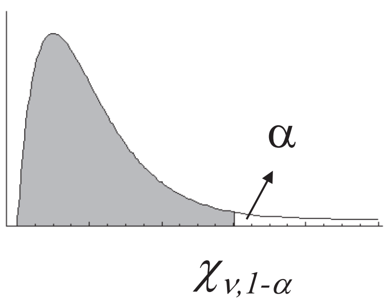


Tabla χ^2



ν: grados de libertad (g.l.)

EJEMPLO

$P(\chi_9 \geq 19{,}02) = 0{,}025$

g.l.	0,995	0,990	0,975	0,950	0,500	0,050	0,025	0,010	0,005
1	,00004	,00016	,00098	,00393	0,455	3,841	5,024	6,635	7,879
2	,01002	,0201	0,051	0,103	1,386	5,991	7,378	9,210	10,60
3	,0717	0,115	0,216	0,352	2,366	7,815	9,348	11,34	12,84
4	0,207	0,297	0,484	0,711	3,357	9,488	11,14	13,28	14,86
5	0,412	0,554	0,831	1,145	4,351	11,07	12,83	15,09	16,75
6	0,676	0,872	1,237	1,635	5,348	12,59	14,45	16,81	18,55
7	0,989	1,239	1,690	2,167	6,346	14,07	16,01	18,48	20,28
8	1,344	1,647	2,180	2,733	7,344	15,51	17,53	20,09	21,95
9	1,735	2,088	2,700	3,325	8,343	16,92	19,02	21,67	23,59
10	2,156	2,558	3,247	3,940	9,342	18,31	20,48	23,21	25,19
11	2,603	3,053	3,816	4,575	10,341	19,68	21,92	24,73	26,76
12	3,074	3,571	4,404	5,226	11,340	21,03	23,34	26,22	28,30
13	3,565	4,107	5,009	5,892	12,340	22,36	24,74	27,69	29,82
14	4,075	4,660	5,629	6,571	13,339	23,68	26,12	29,14	31,32
15	4,601	5,229	6,262	7,261	14,339	25,00	27,49	30,58	32,80
16	5,142	5,812	6,908	7,962	15,338	26,30	28,85	32,00	34,27
17	5,697	6,408	7,564	8,672	16,338	27,59	30,19	33,41	35,72
18	6,265	7,015	8,231	9,390	17,338	28,87	31,53	34,81	37,16
19	6,844	7,633	8,907	10,117	18,338	30,14	32,85	36,19	38,58
20	7,434	8,260	9,591	10,851	19,337	31,41	34,17	37,57	40,00
21	8,034	8,897	10,283	11,591	20,337	32,67	35,48	38,93	41,40
22	8,643	9,542	10,982	12,338	21,337	33,92	36,78	40,29	42,80
23	9,260	10,196	11,689	13,091	22,337	35,17	38,08	41,64	44,18
24	9,886	10,856	12,401	13,848	23,337	36,42	39,36	42,98	45,56
25	10,520	11,524	13,120	14,611	24,337	37,65	40,65	44,31	46,93
26	11,160	12,198	13,844	15,379	25,336	38,89	41,92	45,64	48,29
27	11,808	12,878	14,573	16,151	26,336	40,11	43,19	46,96	49,65
28	12,461	13,565	15,308	16,928	27,336	41,34	44,46	48,28	50,99
29	13,121	14,256	16,047	17,708	28,336	42,56	45,72	49,59	52,34
30	13,787	14,953	16,791	18,493	29,336	43,77	46,98	50,89	53,67
40	20,707	22,164	24,433	26,509	39,335	55,76	59,34	63,69	66,77
50	27,991	29,707	32,357	34,764	49,335	67,50	71,42	76,15	79,49
60	35,534	37,485	40,482	43,188	59,335	79,08	83,30	88,38	91,95
70	43,275	45,442	48,758	51,739	69,334	90,53	95,02	100,43	104,21
80	51,172	53,540	57,153	60,391	79,334	101,88	106,63	112,33	116,32
90	59,196	61,754	65,647	69,126	89,334	113,15	118,14	124,12	128,30
100	67,328	70,065	74,222	77,929	99,334	124,34	129,56	135,81	140,17
120	83,852	86,923	91,573	95,705	119,334	146,57	152,21	158,95	163,65

Distribución de S^2 (Normal)

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \to \chi_n^2$$

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\mu)^2$$

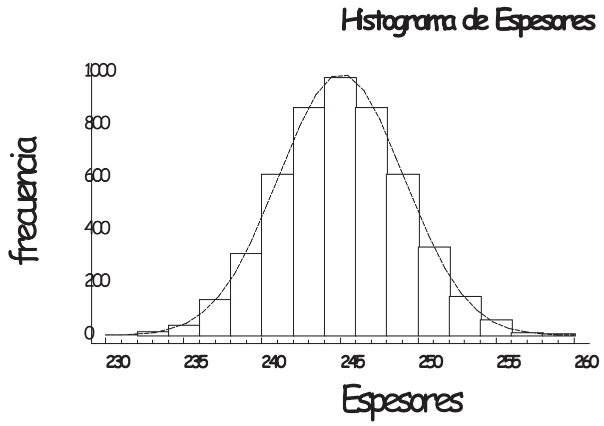
$$= \sum_{i=1}^{n}(X_i-\overline{X})^2 + \sum_{i=1}^{n}(\overline{X}-\mu)^2 + 2\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)$$

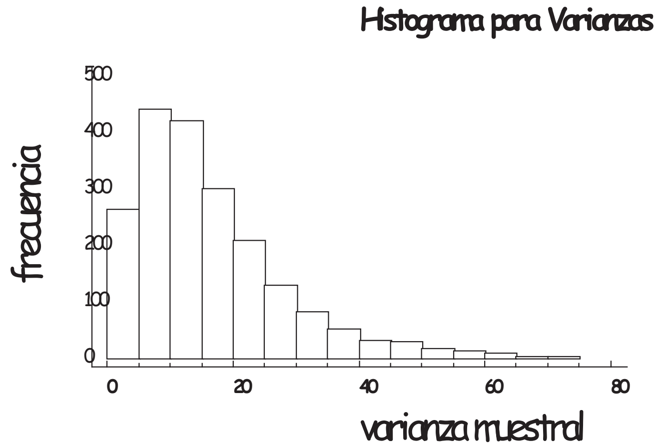
$$= \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$$

$$\underbrace{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}}_{\chi_n^2} = \underbrace{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}}_{\chi_{n-1}^2} + \underbrace{\frac{(\overline{X}-\mu)^2}{\sigma^2/n}}_{\chi_1^2}$$

Distribución de S^2 (Normal)

$$\boxed{\begin{aligned} \frac{nS^2}{\sigma^2} &= \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \xrightarrow{dist} \chi_{n-1}^2 \\ \frac{(n-1)\hat{S}^2}{\sigma^2} &= \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \xrightarrow{dist} \chi_{n-1}^2 \end{aligned}}$$





Distribución de la media (general)

$$X_i \rightarrow f(x,\theta) : \mathrm{Var}[X_i] < \infty$$

$$\mathbf{X} = (X_1, X_2, ..., X_n)^T, \rightarrow \text{Vector de } n \text{ variables aleatorias independientes}$$

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$E[\overline{X}] = \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n}$$

$$Var[\overline{X}] = \frac{Var[X_1] + Var[X_2] + \cdots + Var[X_n]}{n^2}$$

Si las variables tienen la misma media y varianza

$$\mu = E[X_i] \quad \forall i \qquad \sigma^2 = Var[X_i] \quad \forall i$$

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \Rightarrow \begin{cases} E[\overline{X}] = \mu \\ Var[\overline{X}] = \dfrac{\sigma^2}{n} \end{cases} \Rightarrow \overline{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}}) \quad n \rightarrow \infty$$

Binomial

$$\text{Binomial}: X_1, X_2, \dots, X_n$$

$$X_i = \begin{cases} 0 & \text{si es } \textbf{Aceptable} \\ 1 & \text{si es } \textbf{Defectuosa} \end{cases} \Rightarrow \begin{array}{c} E[X_i] = p \\ Var[X_i] = p(1-p) \end{array}$$

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \begin{cases} E[\hat{p}] = p \\ Var[\hat{p}] = \dfrac{p(1-p)}{n} \end{cases}$$

$$\hat{p} \underset{n\to\infty}{\rightarrow} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Poisson

$$\text{Poisson}: X_1, X_2, \dots, X_n$$

$$P(X_i = k) = e^{-\lambda} \frac{\lambda^k}{k!} \Rightarrow \begin{cases} E[X_i] = \lambda \\ Var[X_i] = \lambda \end{cases}$$

$$\hat{\lambda} = \frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \begin{cases} E[\hat{\lambda}] = \lambda \\ Var[\hat{\lambda}] = \dfrac{\lambda}{n} \end{cases}$$

$$\hat{\lambda} \underset{n\to\infty}{\rightarrow} N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$$

Propiedades de los estimadores

$$X_1, X_2, ..., X_n \text{ m.a.s de } f(x,\theta): \hat{\theta} \equiv \hat{\theta}(X_1, X_2, ..., X_n)$$

- Centrados: $E[\hat{\theta}] = \theta$ $(\text{Sesgo}[\hat{\theta}] = E[\hat{\theta}] - \theta)$
- Varianza mínima: $\forall \hat{\theta}': \text{Var}[\hat{\theta}] \leq \text{Var}[\hat{\theta}']$
- *Error cuadrático medio* mínimo

$$\text{ECM}[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = \text{Sesgo}^2(\hat{\theta}) + \text{Var}(\hat{\theta})$$

- Consistentes

$$\lim_{n\to\infty} E[\hat{\theta}] = \theta \quad \text{y} \quad \lim_{n\to\infty} \text{Var}[\hat{\theta}] = 0$$

Ejemplo 1

$$X_1, X_2, ..., X_n \text{ m.a.s de } N(\mu,\sigma): \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Es centrado : $E[\overline{X}] = \mu$

Es de varianza mínima

Es consistente :

$$\lim_{n\to\infty} \text{E}[\overline{X}] = \mu \quad \text{y} \quad \lim_{n\to\infty} \text{Var}[\overline{X}] = 0$$

Ejemplo 2

$$X_1, X_2, \ldots, X_n \text{ m.a.s de } N(\mu,\sigma): S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\text{No es centrado}: E[S^2] = \frac{n-1}{n}\sigma^2$$

$$\text{Sesgo}(S^2) = -\frac{\sigma^2}{n}$$

Varianza:

$$\text{Var}\left[\frac{nS^2}{\sigma^2}\right] = Var\left[\chi^2_{n-1}\right] = 2(n-1) \Rightarrow Var[S^2] = \frac{2(n-1)}{n^2}\sigma^4$$

Es consistente:

$$\lim_{n\to\infty} \text{E}[S^2] = \sigma^2 \quad \text{y} \quad \lim_{n\to\infty} \text{Var}[S^2] = 0$$

Concepto de intervalo de confianza

Se ha realizado una encuesta a 400 personas elegidas al azar para estimar la proporción p de votantes de un partido político.

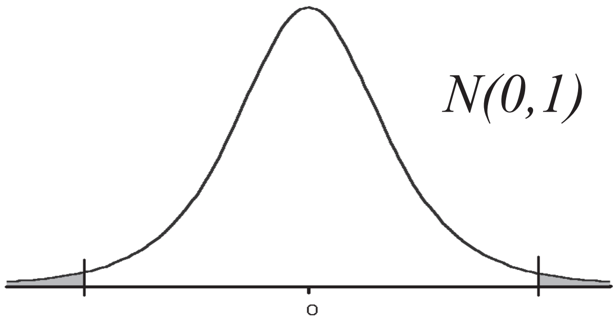
Resultado Encuesta	
Sí	220
No/Otros	180

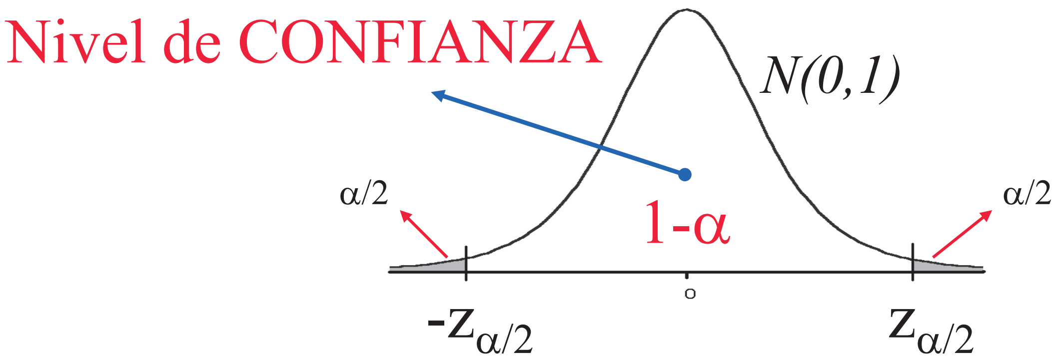
Introducción

$$X \to B(n,p) \qquad\qquad X \overset{aprox.}{\longrightarrow} N(np, \sqrt{np(1-p)})$$

$$\hat{p} = \frac{X}{n} \to N(p, \sqrt{\frac{p(1-p)}{n}})$$

$$\boxed{\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \to N(0,1)}$$





$$P(-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}) = 1-\alpha$$

Despejando p de: $$-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}$$

$$-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \leq z_{\alpha/2}$$

$$\Downarrow$$

$$-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} \leq z_{\alpha/2}$$

$$\Downarrow$$

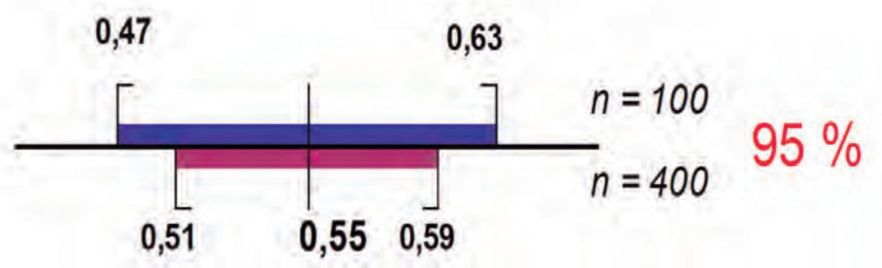
$$\boxed{\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \quad p \quad \leq \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Nivel de confianza: (1-α)

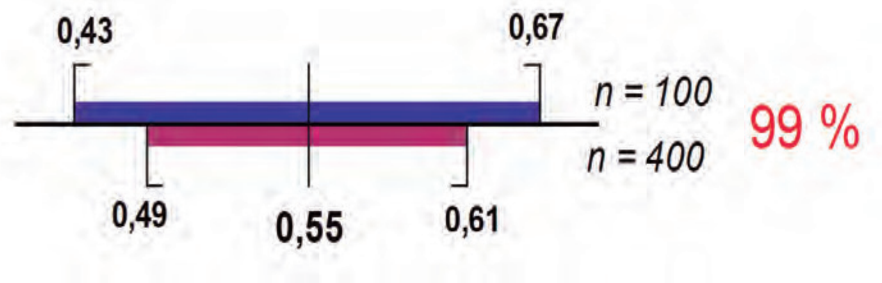
Tamaño Muestral n

Ejemplo $\hat{p} = \frac{220}{400} = 0{,}55$

$$p \in 0{,}55 \pm 1{,}96\sqrt{\frac{0{,}55\times 0{,}45}{400}}$$



$$p \in 0.55 \pm 2.58\sqrt{\frac{0.55\times 0.45}{400}}$$



1. Normal: Intervalo para μ con σ conocido

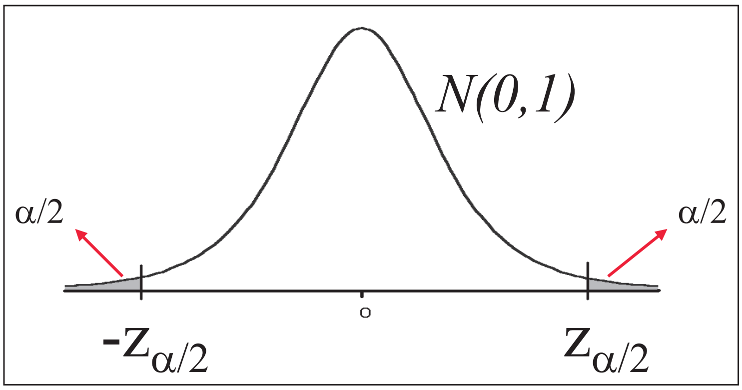
$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$\bar{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$$

$$-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$



$$\boxed{\mu \in \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

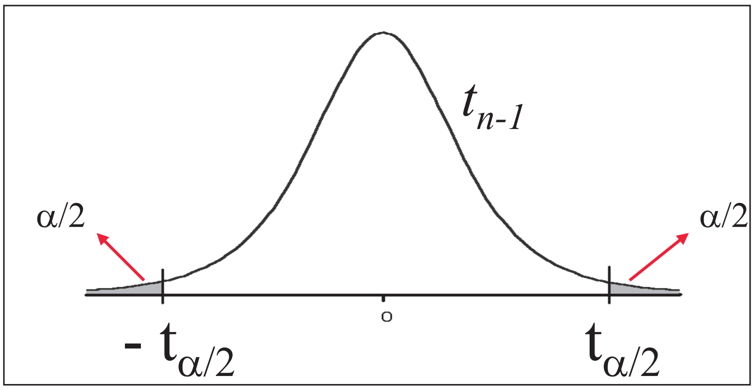
2. Normal:Intervalo para μ con σ desconocido

$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$\bar{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$$

$$\frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \rightarrow t_{n-1}$$



$$-t_{n-1,\alpha/2} \leq \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \leq t_{n-1,\alpha/2}$$

$$\boxed{\bar{x} - t_{n-1,\alpha/2}\frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,\alpha/2}\frac{\hat{s}}{\sqrt{n}}}$$

$$\boxed{\mu \in \bar{x} \pm t_{n-1,\alpha/2}\frac{\hat{s}}{\sqrt{n}}}$$

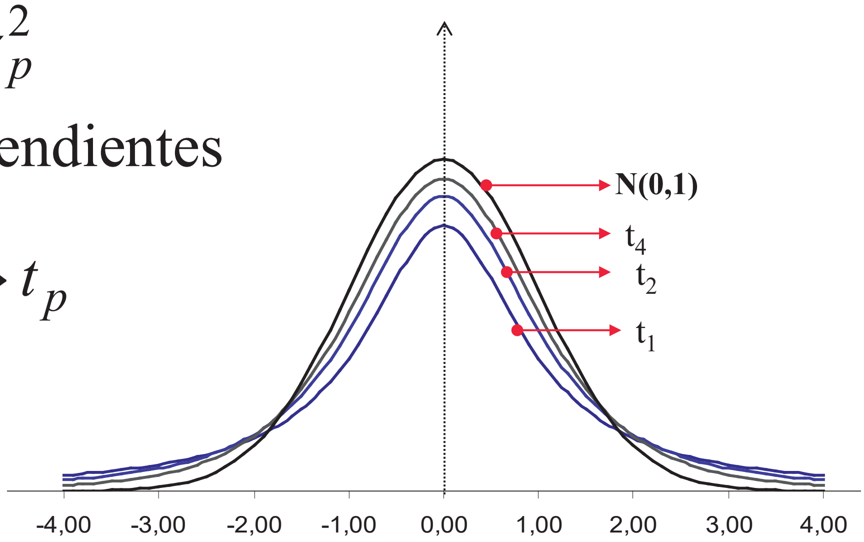
Distribución **t** de *Student*

$$Z \to N(0,1)$$

$$V \to \chi^2_p$$

$$Z, V \text{ son independientes}$$

$$\frac{Z}{\sqrt{V/p}} \to t_p$$



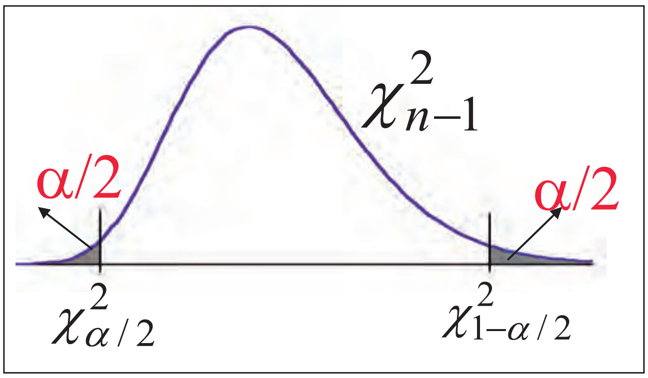
3. Normal: Intervalo para σ^2

$$X_1, X_2, \ldots, X_n \to N(\mu, \sigma)$$

$$\hat{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \quad \cdots\cdots\to \quad \frac{(n-1)\hat{S}^2}{\sigma^2} \to \chi^2_{n-1}$$

$$P\left(\chi^2_{\alpha/2} \le \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2_{1-\alpha/2}\right) = 1 - \alpha$$

$$\boxed{\frac{(n-1)\hat{s}^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)\hat{s}^2}{\chi^2_{\alpha/2}}}$$



EJEMPLO 1. La resistencia a la compresión de 15 probetas de acero elegidas al azar es:

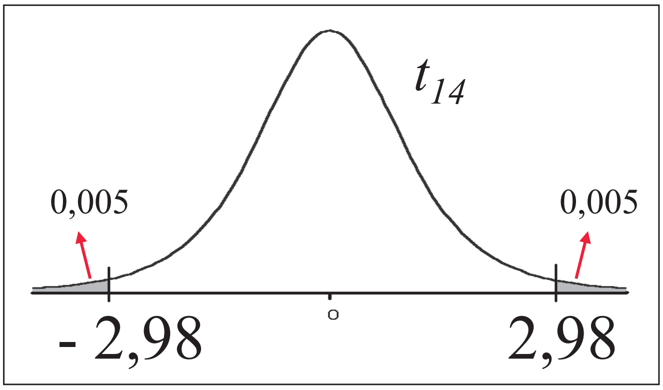
40,15	65,10	49,50	22,40	38,20
60,40	43,40	26,35	31,20	55,60
47,25	73,20	35,90	45,25	52,40

$$\frac{\overline{X}-\mu}{\hat{S}/\sqrt{15}} \to t_{14}$$

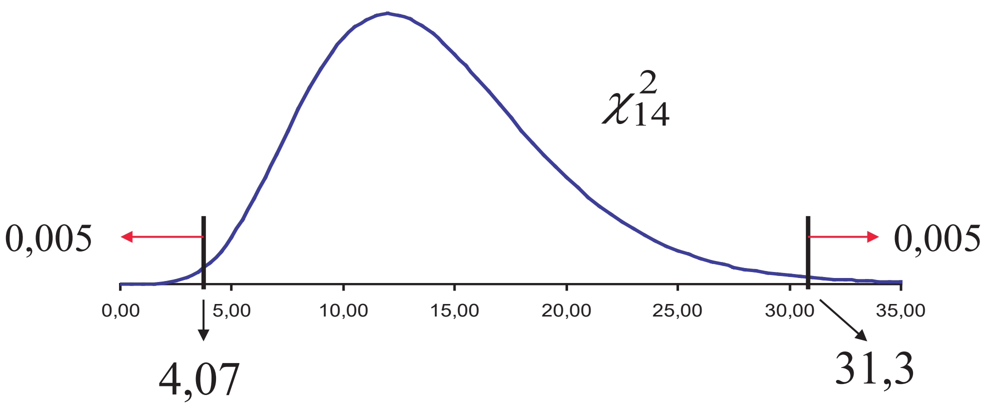
$$-2{,}98 \le \frac{\overline{X}-\mu}{\hat{S}/\sqrt{15}} \le 2{,}98$$

$$\bar{x} = 45{,}75 \quad \hat{s} = 14{,}2$$

$$45{,}75 - 2{,}98\frac{14{,}2}{\sqrt{15}} \le \mu \le 45{,}75 + 2{,}98\frac{14{,}2}{\sqrt{15}}$$



99 % confianza: $34{,}8 \le \mu \le 56{,}7$



$$\frac{(n-1)\hat{S}^2}{\sigma^2} \to \chi^2_{n-1} \qquad \frac{14\hat{S}^2}{\sigma^2} \to \chi^2_{14}$$

$$P(4{,}07 \le \frac{14\hat{S}^2}{\sigma^2} \le 31{,}3) = 0{,}99$$

$$\hat{s}^2 = 201{,}6 \longrightarrow \frac{14\times 201{,}6}{31{,}3} \le \sigma^2 \le \frac{14\times 201{,}6}{4{,}07}$$

99% confianza: $90{,}2 \le \sigma^2 \le 693{,}6$

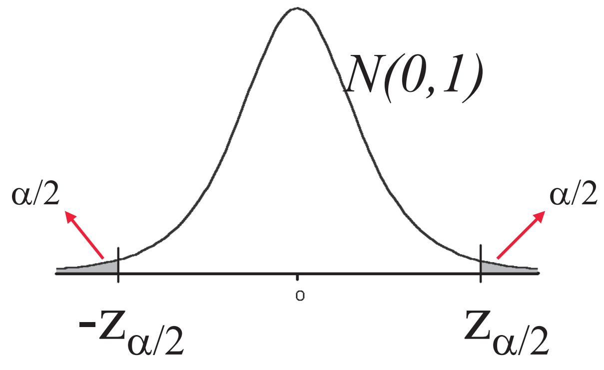
4. Poisson: Intervalo para λ

$$X_1, X_2, \dots, X_n \rightarrow \text{Poisson}\ (\lambda)$$

$$\hat{\lambda} = \overline{X} \overset{aprox}{\rightarrow} \text{N}(\lambda, \sqrt{\frac{\lambda}{n}})$$

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda / n}} \rightarrow \text{N}(0,1)$$

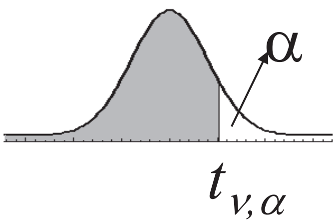
$$P(-z_{\alpha/2} \leq \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda / n}} \leq z_{\alpha/2}) = 1 - \alpha$$



$$\boxed{\hat{\lambda} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \ \leq \ \lambda \ \leq \ \hat{\lambda} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}}$$

Tabla

t-Student



ν: grados de libertad (g.l.)

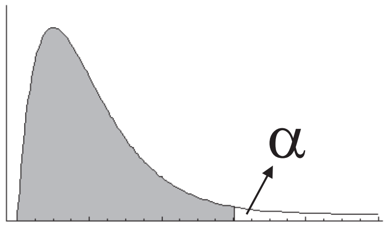
EJEMPLO

$P(t_9 \geq 2{,}262) = 0{,}025$

α

g.l	0,20	0,15	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,376	1,963	3,078	6,314	12,706	31,821	63,656	127,321	318,289	636,578
2	1,061	1,386	1,886	2,920	4,303	6,965	9,925	14,089	22,328	31,600
3	0,978	1,250	1,638	2,353	3,182	4,541	5,841	7,453	10,214	12,924
4	0,941	1,190	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,920	1,156	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,906	1,134	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,889	1,108	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,883	1,100	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,879	1,093	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,876	1,088	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,870	1,079	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,868	1,076	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,866	1,074	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,865	1,071	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,660
30	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
40	0,851	1,050	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
50	0,849	1,047	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
60	0,848	1,045	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460
70	0,847	1,044	1,294	1,667	1,994	2,381	2,648	2,899	3,211	3,435
80	0,846	1,043	1,292	1,664	1,990	2,374	2,639	2,887	3,195	3,416
90	0,846	1,042	1,291	1,662	1,987	2,368	2,632	2,878	3,183	3,402
100	0,845	1,042	1,290	1,660	1,984	2,364	2,626	2,871	3,174	3,390
infinito	0,842	1,036	1,282	1,645	1,960	2,327	2,576	2,808	3,091	3,291
	0,20	0,15	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005

Tabla χ^2



$\chi_{\nu,1-\alpha}$

ν: grados de libertad (g.l.)

EJEMPLO

$P(\chi_9 \geq 19{,}02) = 0{,}025$

	α								
g.l.	**0,995**	**0,990**	**0,975**	**0,950**	**0,500**	**0,050**	**0,025**	**0,010**	**0,005**
1	,00004	,00016	,00098	,00393	0,455	3,841	5,024	6,635	7,879
2	,01002	,0201	0,051	0,103	1,386	5,991	7,378	9,210	10,60
3	,0717	0,115	0,216	0,352	2,366	7,815	9,348	11,34	12,84
4	0,207	0,297	0,484	0,711	3,357	9,488	11,14	13,28	14,86
5	0,412	0,554	0,831	1,145	4,351	11,07	12,83	15,09	16,75
6	0,676	0,872	1,237	1,635	5,348	12,59	14,45	16,81	18,55
7	0,989	1,239	1,690	2,167	6,346	14,07	16,01	18,48	20,28
8	1,344	1,647	2,180	2,733	7,344	15,51	17,53	20,09	21,95
9	1,735	2,088	2,700	3,325	8,343	16,92	19,02	21,67	23,59
10	2,156	2,558	3,247	3,940	9,342	18,31	20,48	23,21	25,19
11	2,603	3,053	3,816	4,575	10,341	19,68	21,92	24,73	26,76
12	3,074	3,571	4,404	5,226	11,340	21,03	23,34	26,22	28,30
13	3,565	4,107	5,009	5,892	12,340	22,36	24,74	27,69	29,82
14	4,075	4,660	5,629	6,571	13,339	23,68	26,12	29,14	31,32
15	4,601	5,229	6,262	7,261	14,339	25,00	27,49	30,58	32,80
16	5,142	5,812	6,908	7,962	15,338	26,30	28,85	32,00	34,27
17	5,697	6,408	7,564	8,672	16,338	27,59	30,19	33,41	35,72
18	6,265	7,015	8,231	9,390	17,338	28,87	31,53	34,81	37,16
19	6,844	7,633	8,907	10,117	18,338	30,14	32,85	36,19	38,58
20	7,434	8,260	9,591	10,851	19,337	31,41	34,17	37,57	40,00
21	8,034	8,897	10,283	11,591	20,337	32,67	35,48	38,93	41,40
22	8,643	9,542	10,982	12,338	21,337	33,92	36,78	40,29	42,80
23	9,260	10,196	11,689	13,091	22,337	35,17	38,08	41,64	44,18
24	9,886	10,856	12,401	13,848	23,337	36,42	39,36	42,98	45,56
25	10,520	11,524	13,120	14,611	24,337	37,65	40,65	44,31	46,93
26	11,160	12,198	13,844	15,379	25,336	38,89	41,92	45,64	48,29
27	11,808	12,878	14,573	16,151	26,336	40,11	43,19	46,96	49,65
28	12,461	13,565	15,308	16,928	27,336	41,34	44,46	48,28	50,99
29	13,121	14,256	16,047	17,708	28,336	42,56	45,72	49,59	52,34
30	13,787	14,953	16,791	18,493	29,336	43,77	46,98	50,89	53,67
40	20,707	22,164	24,433	26,509	39,335	55,76	59,34	63,69	66,77
50	27,991	29,707	32,357	34,764	49,335	67,50	71,42	76,15	79,49
60	35,534	37,485	40,482	43,188	59,335	79,08	83,30	88,38	91,95
70	43,275	45,442	48,758	51,739	69,334	90,53	95,02	100,43	104,21
80	51,172	53,540	57,153	60,391	79,334	101,88	106,63	112,33	116,32
90	59,196	61,754	65,647	69,126	89,334	113,15	118,14	124,12	128,30
100	67,328	70,065	74,222	77,929	99,334	124,34	129,56	135,81	140,17
120	83,852	86,923	91,573	95,705	119,334	146,57	152,21	158,95	163,65

Capítulo 5: Ejercicios de Inferencia

1. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de una variable aleatoria exponencial con función de densidad, $f_X(x) = \frac{1}{\mu}e^{-x/\mu}, \quad x \geq 0, \ \mu > 0$;. Obtener el estimador de μ por el método de los momentos y por el método de máxima verosimilitud.

2. La función de densidad de Rayleigh de parámetro b es

$$f(x) = \frac{2x}{b^2} \exp\left(-\frac{x^2}{b^2}\right), \qquad x \geq 0, b \geq 0,$$

   a) Sabiendo que $E[X] = \frac{\sqrt{\pi}}{2}b$, estimar b por el método de los momentos.

   b) Estimar b por máxima verosimilitud.

3. La variable aleatoria X tiene distribución binomial con parámetros n y p, ambos desconocidos. Si $\{16,18,22,25,27\}$ es una muestra aleatoria simple de la distribución anterior, estimar por el método de los momentos n y p.

4. Los taxis en servicio de una ciudad están numerados del 1 al N. Se observa una muestra de 5 taxis y se apuntan sus números. Obtener un estimador de N por el método de los momentos. Imagina que los números obtenidos son 7,23,34,36,84; a la vista de los resultados explica si es un buen estimador. (El ejercicio 5 sugiere un estimador más adecuado en un problema similar).

5. La variable aleatoria X sigue una distribución uniforme en el intervalo $[0, \theta]$. Dada una muestra aleatoria simple $x_1, x_2, ..., x_n$ se pide:

   a) Obtener el estimador de θ por el método de los momentos. Calcular su sesgo y su varianza

   b) Obtener el estimador máximo verosímil. Calcular su sesgo y su varianza.

   c) Compara el Error Cuadrático Medio de los dos estimadores.

   d) A partir del estimador máximo verosímil propón un estimador centrado, calcula su varianza y su ECM.

6. Una variable aleatoria discreta puede tomar los valores 0, 1 y 2 con probabilidades $1.5/\theta$, $2.5/\theta$ y $(\theta - 4)/\theta$ respectivamente, con $\theta > 4$. Se toma una muestra de tamaño 25 con los resultados siguientes (la segunda fila corresponde a la fracción observada O_i para 0, 1 y 2).

x	0	1	2
O_i	5	17	3

   Estimar θ por máxima verosimilitud.

7. Se ha tomado una muestra de tamaño 10 del tiempo, en minutos, entre el paso de dos autobuses T en una parada con los siguientes resultados: 9,10,6,4,15,6,1,5,4,10. Si la función de distribución del tiempo de paso es $F_T(t) = 1 - \exp(-\alpha t)$, estima el parámetro α por máxima verosimilitud y con este valor, la probabilidad estimada de esperar al autobús más de 8 minutos.

8. El club de tiro de una determinada ciudad está estudiando la distancia X del punto de impacto del proyectil al centro de la diana de sus 10 mejores tiradores.

   Sabiendo que la función de densidad de la variable aleatoria presentada es

$$f_X(x) = \frac{2x}{\theta^2} \exp[-\frac{x^2}{\theta^2}], \qquad x \geq 0, \theta \geq 0,$$

   Estimar θ si la distancia en cm al blanco de 10 tiradores fue

   2,1 3,2 6,3 5,4 2,2 6,9 7,1 6,6 2,5 9,1

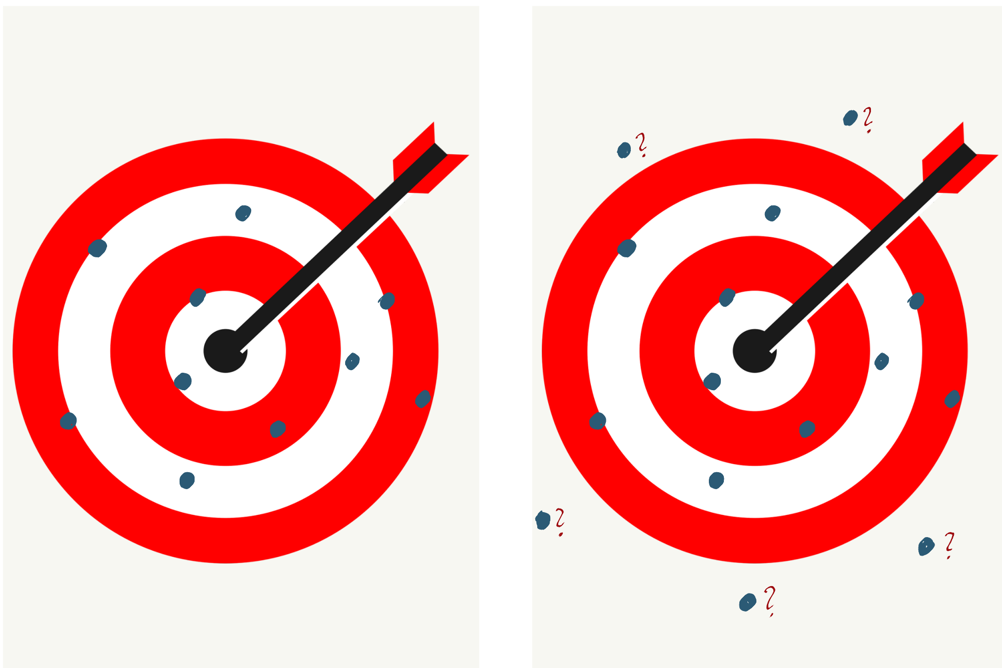


Figura 1: Resultados de disparos a una diana de los ejercicios 8 y 9

9. Repite el ejercicio 8 del club de tiro, suponiendo que en el test han participado 15 tiradores, 10 de ellos obtuvieron los valores indicados y los otros 5 no impactaron en la diana, por tanto sus valores son desconocidos y cumplen $x_i > 10$.

10. La velocidad de una molécula en un gas ideal según el modelo de Maxwell-Boltzmann es una variable aleatoria con función de densidad

$$f(x) = \frac{4}{\alpha^3\sqrt{\pi}} x^2 \, e^{-(x/\alpha)^2}, \quad x \geq 0$$

   donde $\alpha > 0$, es un parámetro de la distribución. Teniendo en cuenta que

$$E[X] = \frac{2\alpha}{\sqrt{\pi}} \quad \text{y} \quad \mathrm{Var}[X] = (\frac{3}{2} - \frac{4}{\pi})\alpha^2.$$

   a) Calcular el estimador máximo verosímil de α y obtén la varianza asíntótica del estimador.

b) Calcular el estimador por momentos de α y la varianza de dicho estimador.

11. La función de distribución de una variable aleatoria es

$$F_X(x) = \begin{cases} 0 & x < 0, \\ (x/\beta)^\alpha & 0 \leq x \leq \beta, \\ 1 & x > \beta. \end{cases}$$

donde los parámetros α y β son positivos.

a) Estimar los parámetros de la distribución por el método de máxima verosimilitud.

b) La longitud en milímetros de los huevos de una especie de pájaros se modelna con dicha distribución. Obtener los estimadores de α y β si se ha obtenido la siguiente muestra: 22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1,23.1, 23.5, 23.0, 23.0

12. Los tiempos de funcionamiento de dos componentes electrónicos distintos siguen distribuciones exponenciales con esperanzas μ y 2μ. Se han obtenido los tiempos de fallo de una muestra de cada tipo de componente, en ambos casos de tamaño n. Obtener el estimador de máxima verosimilitud de μ, calcular su media y su varianza.

13. Una compañía, para determinar el número de consumidores de un determinado producto en Madrid, ha encuestado a personas elegidas al azar hasta encontrar a 20 que utilicen el producto. Estimar por máxima verosimilitud la proporción de consumidores en la ciudad si el número total de entrevistados ha sido 115

14. Sea $\overline{X}$ la media aritmética de una muestra aleatoria simple de una distribución $\mathrm{N}(\mu, \sigma)$. Se define $\tilde{X} = c\overline{X}$ como nuevo estimador para μ. Determinar c (en función de μ y σ) para que el nuevo estimador tenga Error Cuadrático Medio (ECM) mínimo. Calcular c si se sabe que el coeficiente de variación $\sigma/\mu = 2$.

15. Para estimar la varianza σ^2 de una distribución normal $N(\mu, \sigma)$ se suelen utilizan tres estimadores

$$\hat{S}_1^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}, \quad \hat{S}_2^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n}, \quad \hat{S}_3^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n+1}$$

Compara los tres estimadores utilizando sesgo, varianza y ECM.

16. Para estimar la varianza σ^2 de una población normal se utiliza el estimador $\widehat{\sigma}^2 = k\widehat{S}^2$, siendo $\widehat{S}^2$ la varianza muestral corregida y k una constante. Calcular el valor de k que minimiza el error cuadrático medio. (Utilizar $Var[\chi_g^2] = 2g$, siendo g el número de grados de libertad).

17. Un sistema de lectura telemática de consumo de energía eléctrica emplea un mensaje de 128-bit. Ocasionalmente las interferencias aleatorias provocan que un bit se invierta produciéndose un error de transmisión. Se acepta que la probabilidad de que cada bit cambie en una transmisión es constante e igual a p, y que los cambios son independientes. Estima el valor de p si se ha comprobado que de las últimas 10000 lecturas efectuadas (todas de 128-bit) 340 eran erróneas.

18. Un estudiante ha realizado el siguiente ejercicio de simulación. Ha generado con R $n = 100$ observaciones de una normal de media 10 y desviación típica 1. Para estos 100 datos ha obtenido la mediana y la ha denominado M_1. Ha repetido el mismo proceso $N = 10000$ veces. De forma que ha obtenido $M_1, M_2, ..., M_N$. En el gráfico se muestra el histograma de estos 10000 valores, que tienen media 10.002 y desviación típica 0.1266.

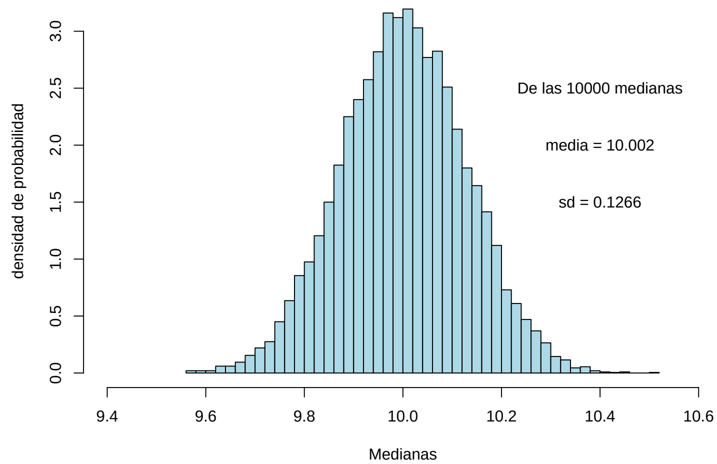


Figura 2: Histograma de las 10000 medianas de muestras de tamaño 100 de una distribución norma N(10,1)

A partir de esta simulación, explica qué propiedades tiene la mediana como estimador de la media μ de una distribución normal, sus ventajas e inconvenientes respecto a utilizar la media aritmética como estimador.

19. El tiempo que tarda un paciente en recuperarse de una dolencia es una variable aleatoria con distribución exponencial. Si al observar a 20 pacientes, al cabo de 20 días 15 no se han recuperado, mientras que los 5 restantes tardaron $17, 15, 19, 18$ y 17 días en hacerlo, estime por máxima verosimilitud el tiempo medio de recuperación.

20. El tiempo de duración de ciertos componentes electrónicos es una variable aleatoria con distribución exponencial. Se ha realizado un ensayo con 10 componentes cuyos tiempos de duración han sido: 37,45,92,104,109,200,295. Despues de 400 horas, tres componentes seguían funcionando. Con esta información, estimar por máxima verosimilitud el parámetro de la distribución exponencial.

21. Se han tomado 12 valores de una variable física X, que se supone sigue una distribución Normal, resultando

    30.2, 30.8, 29.3, 29, 30.9, 30.8, 29.7, 28.9, 30.5, 31.2, 31.3, 28.5.

    a) Construir un intervalo de confianza para la media de la población al 95 % de confianza.

    b) Construir un intervalo de confianza para la varianza de la población con el mismo nivel de confianza del apartado anterior.

22. Una muestra de 12 estaciones de servicio de una cadena de gasolineras proporciona un ingreso medio por persona al mes de 2340 euros con una desviación típica de 815 euros. Calcular un intervalo de confianza para el ingreso medio por trabajador en esta empresa. Calcular el número de estaciones que debemos estudiar para que el intervalo tenga una amplitud máxima de 500 euros.

23. Se han escogido al azar 15 probetas de un determinado acero, cuya resistencia a la compresión se supone que se distribuye normalmente, y se ha medido ésta en las unidades adecuadas, habiéndose observado los resultados siguientes

$$40.15,\ 65.10,\ 49.5,\ 22.4,\ 38.2,\ 60.4,\ 43.4,\ 26.35,\ 31.2,\ 55.6,\ 47.25,\ 73.2,\ 35.9,\ 45.25,\ 52.4$$

   a) Estimar la resistencia media del acero y su varianza.

   b) Hallar un intervalo de confianza del 99 % para la resistencia media.

   c) Hallar un intervalo de confianza del 99 % para la varianza.

   d) Cuántas probetas deberían haberse utilizado en el estudio si se quisiera estimar la resistencia media del acero con una precisión de ± 6 unidades y una confianza del 95 %?.

24. Una compañía de comida precocinada desea lanzar al mercado un nuevo producto. Para conocer la aceptación del mismo realiza previamente una encuesta entre 200 personas elegidas al azar, de las que 37 manifiestan su disposición a comprarlo.

   a) Obtener un intervalo de confianza ($\alpha = 0.05$) para la proporción p de compradores potenciales de este nuevo producto.

   b) ¿Cúal debería ser el tamaño muestral si se quisiera reducir la longitud del intervalo a la mitad.

25. Se desea estimar la proporción de niños entre 0 y 14 años que se encuentran adecuadamente vacunados contra la poliomielitis. Si se quiere que la diferencia en valor absoluto entre la estimación final y el verdadero valor de la proporción sea menor que 0.05 con probabilidad 0.95, ¿ Cúal es el tamaño muestral mínimo requerido?.

26. Una roca lunar es enviada a un laboratorio para determinar su nivel de radiactividad θ, nivel que se mide por el número medio de partículas emitidas por hora. Después de 15 horas, el equipo Geiger ha contabilizado un total de 3.547 partículas emitidas. Aceptando que el número de partículas emitidas sigue una distribución de Poisson, dar un intervalo con 95 % de confianza para el nivel de radiactividad de la roca. (Nota.- Utilizar que si Z tiene distribución N(0,1), entonces $P(Z \leq 1.96) = 0.975$).

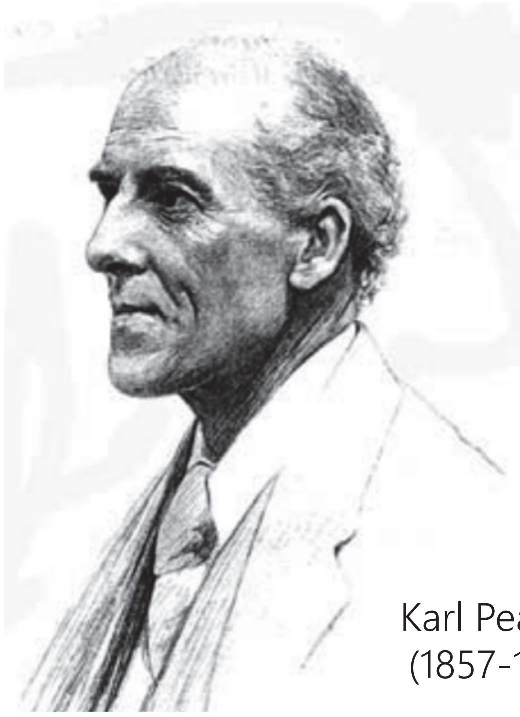
27. Teniendo en cuenta que si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una variable aleatoria exponencial con función de densidad, $f(x) = \frac{1}{\lambda}e^{-x/\lambda}, \quad x \geq 0,\ \lambda > 0$; el estadístico $U = 2n\overline{X}/\lambda$ tiene distribución χ^2_{2n}, donde $\overline{X} = (X_1+X_2+\cdots+X_n)/n$; resolver la cuestión siguiente: El tiempo de funcionamiento de un equipo electrónico es una variable aleatoria con distribución exponencial. Se han tomado los tiempos de funcionamiento hasta el fallo de 30 equipos elegidos al azar, obteniéndose 6.2×10^3 horas de media. Calcular un intervalo con 95 % de confianza para la vida media de un equipo.

28. En una centralita ha habido 180 llamadas durante las últimas dos horas. Obtenga un intervalo de confianza para la esperanza del número de llamadas por hora suponiendo que el número de llamadas durante un periodo de duración T cualquiera sigue una distribución de Poisson.

Estadística

Parte II.

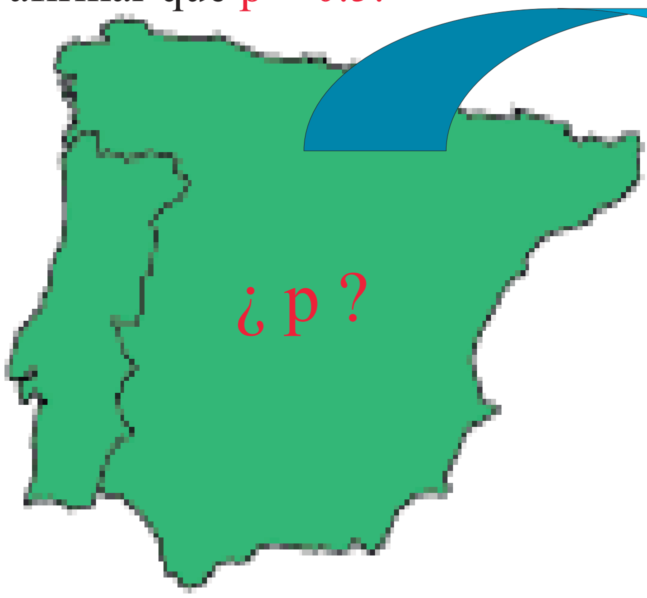
Tema 6. Contraste de Hipótesis

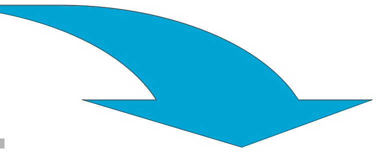


Karl Pearson
(1857-1936)

Contraste de Hipótesis

Se ha realizado una encuesta a 400 personas elegidas al azar
Llamando *p* a la proporción de votantes del partido político A.
Podemos afirmar que p > 0.5.





Resultado Encuesta	
Sí	220
No/Otros	180

Contraste de hipótesis

$$H_0: \quad p \leq 0.5$$
$$H_1: \quad p > 0.5$$

$$\boxed{\begin{array}{l} H_0: \quad p = 0.5 \\ H_1: \quad p > 0.5 \end{array}}$$

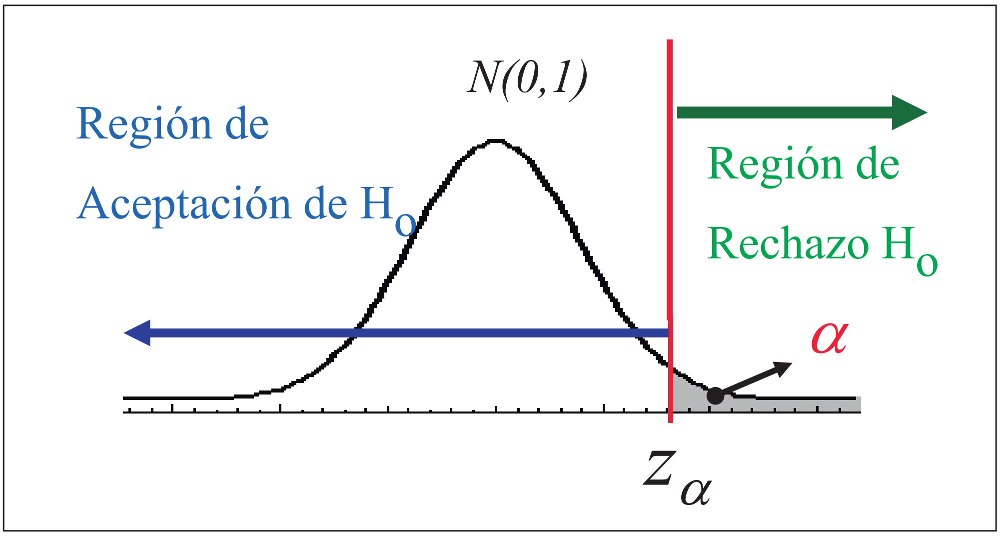
$$\hat{P} \to N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Si H_0 es cierto $p = 0.5$

$$\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} \to N(0,1)$$

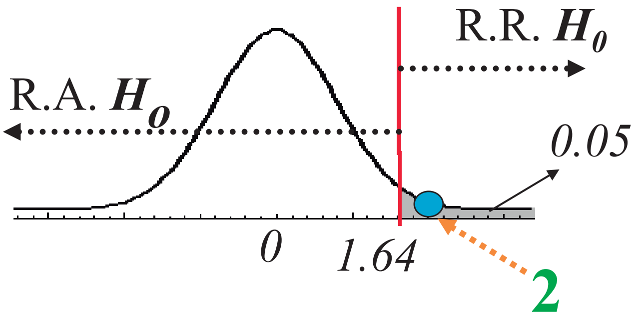
Nivel de significación α

α = P(rechazar H_0 | H_0 es cierta)



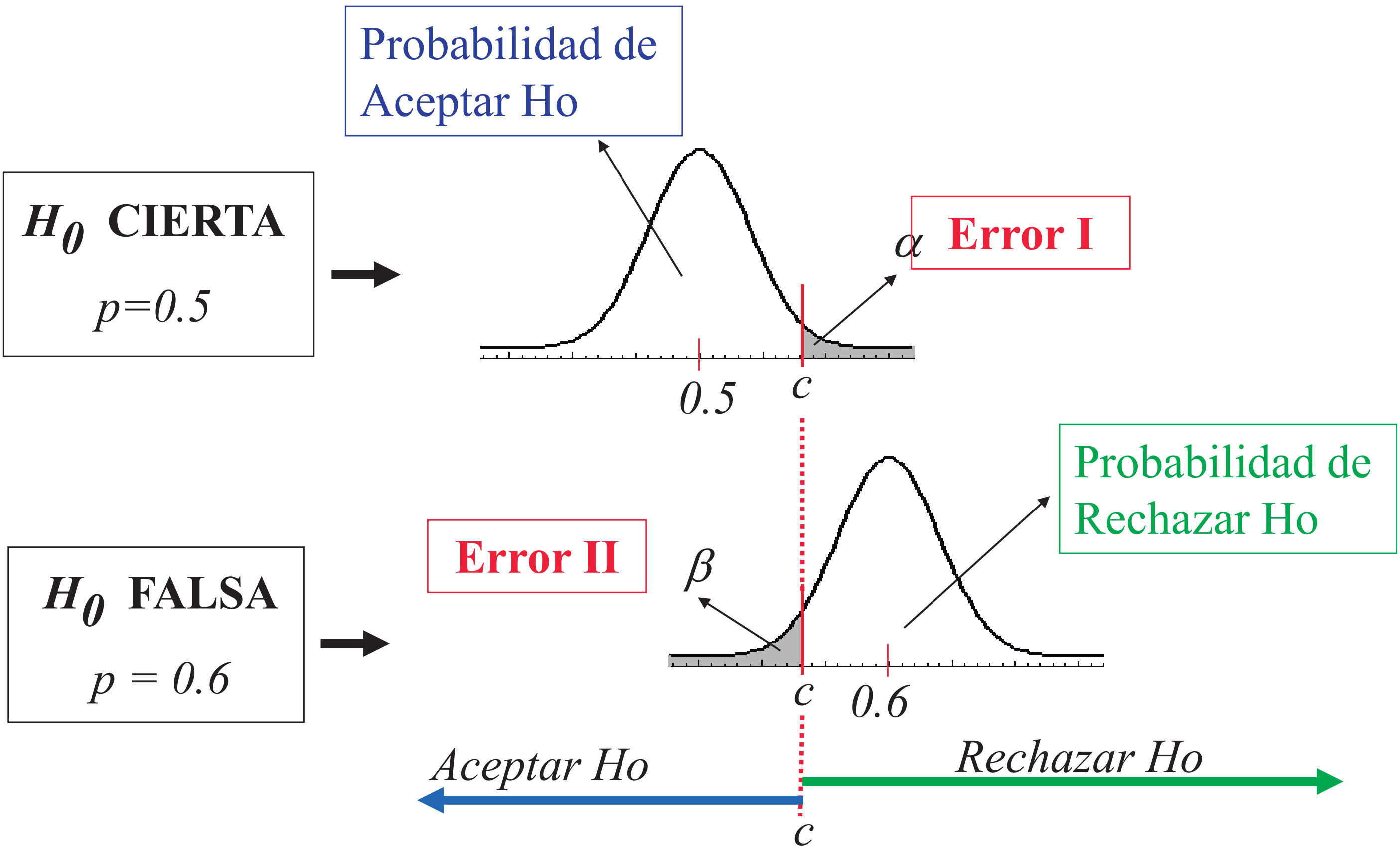
$$\text{Muestra } \hat{p} = 0.55 \Rightarrow z_0 = \frac{0.55 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{400}}} = 2$$

$$\text{Elegido el nivel de significación } \alpha = 0.05$$



$$z_0 = 2 > z_\alpha = 1.64 \quad \Rightarrow \quad \text{Se rechaza } \mathrm{H}_0$$

Con un nivel de significación de $\alpha = 0.05 \Rightarrow p > 0.5$



Tipos de errores

RESULTADO CONTRASTE

REALIDAD	Se **Acepta** H_o	Se **Rechaza** H_o
H_o **CIERTA**	**Ok**	**Error tipo I** α
H_o **FALSA**	**Error tipo II** β	**Ok**

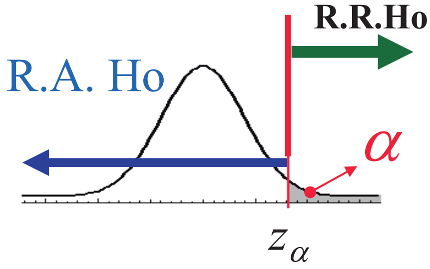
Tipos de Contrastes

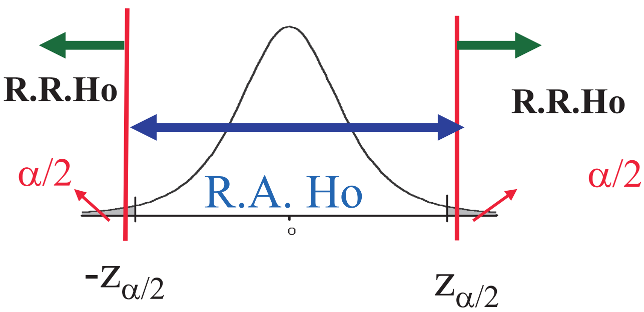
$$H_0: \quad p = 0.5$$
$$H_1: \quad p > 0.5$$

UNILATERAL

$$H_0: \quad p = 0.5$$
$$H_1: \quad p \neq 0.5$$

BILATERAL





Nivel de significación α

Nivel crítico o *p*-valor

$$H_0: \quad p = p_0$$
$$H_1: \quad p > p_0$$

$$\hat{P} \to N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

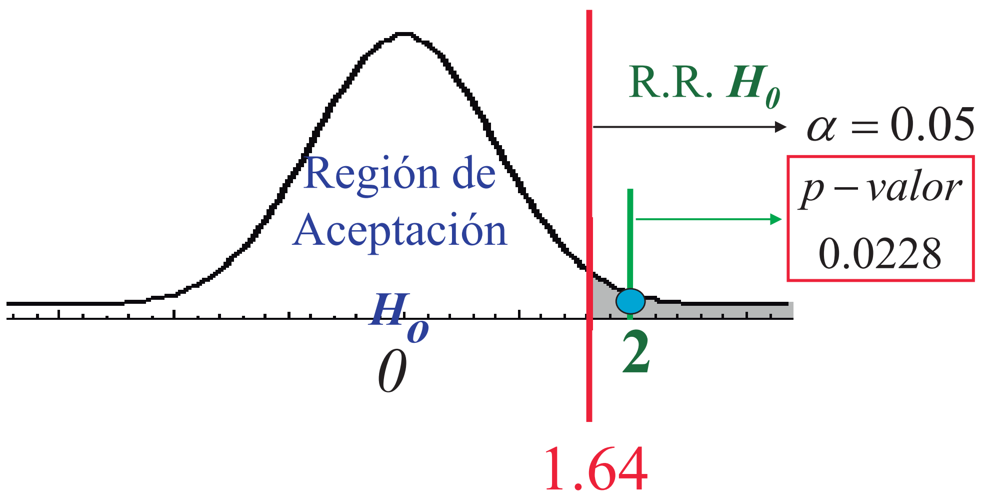
$$p - valor = \Pr(\hat{P} > \hat{p} \mid H_0 \text{ es cierto})$$

$$= \Pr\left(\frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}} > \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right)$$

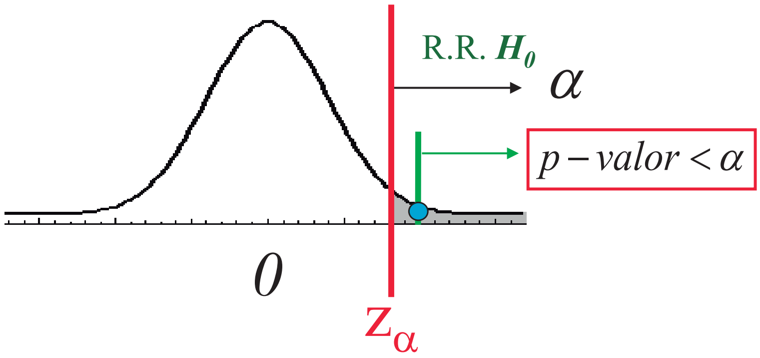
$$= \Pr(Z > z_0)$$

Ejemplo

Datos: $\hat{p} = \frac{220}{400} = 0.55$



Relación entre *p-valor* y α

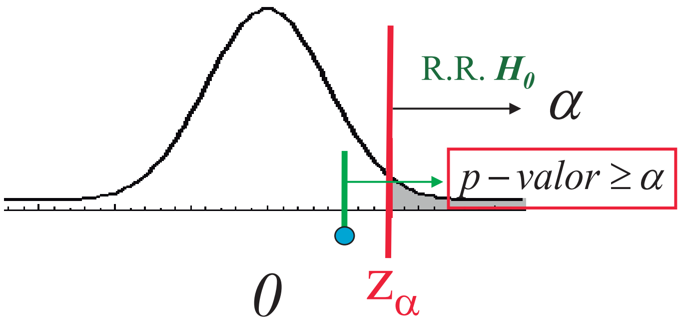


p-valor < α

Se rechaza H_0

p-valor ≥ α

No se rechaza H_0



1. Normal: Contraste para μ con σ conocido

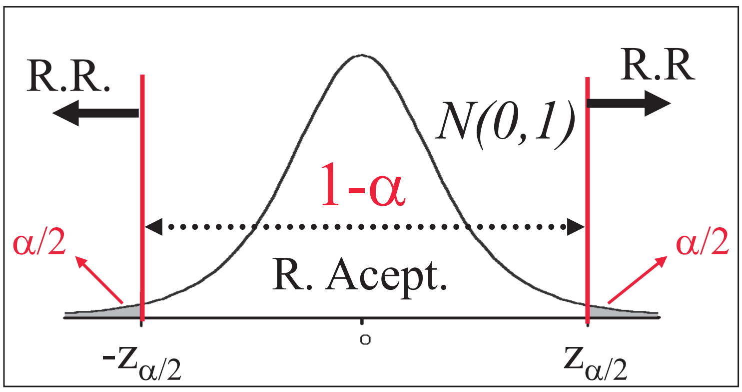
$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$H_0: \quad \mu = \mu_0$$

$$H_1: \quad \mu \neq \mu_0$$

$$\overline{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \rightarrow N(0,1)$$



$$z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \longrightarrow$$

$$|z_0| \leq z_{\alpha/2} \Rightarrow \text{No se rechaza } H_0$$

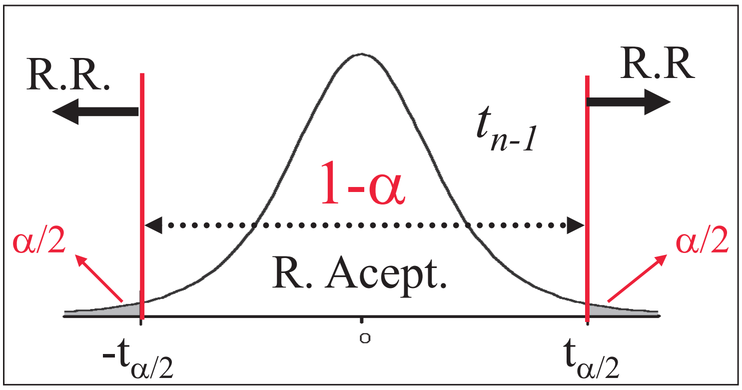
$$|z_0| > z_{\alpha/2} \Rightarrow \text{Se rechaza } H_0$$

2. Normal: Contraste para μ con σ desconocido

$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$H_0: \quad \mu = \mu_0$$

$$H_1: \quad \mu \neq \mu_0$$

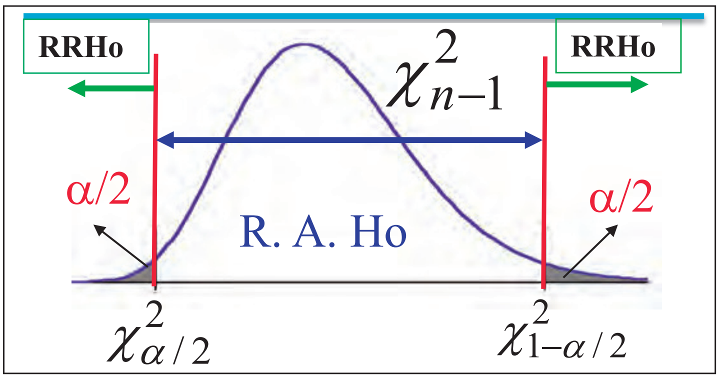


$$\overline{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\frac{\overline{X} - \mu}{\hat{S} / \sqrt{n}} \rightarrow t_{n-1}$$

$$t_0 = \frac{\overline{x} - \mu_0}{\hat{s} / \sqrt{n}} \longrightarrow \boxed{\begin{array}{l} |t_0| \leq t_{\alpha/2} \Rightarrow \text{No se rechaza } H_0 \\ |t_0| > t_{\alpha/2} \Rightarrow \text{Se rechaza } H_0 \end{array}}$$

3. Normal: Contraste para σ^2



$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$\boxed{\begin{array}{l} H_0: \quad \sigma^2 = \sigma_0^2 \\ H_0: \quad \sigma^2 \neq \sigma_0^2 \end{array}}$$

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$\frac{(n-1)\hat{S}^2}{\sigma^2} \rightarrow \chi^2_{n-1} \longrightarrow P(\chi^2_{\alpha/2} \leq \chi^2_{n-1} \leq \chi^2_{1-\alpha/2}) = 1 - \alpha$$

Datos: $\hat{s}^2$

$$\chi_0^2 = \frac{(n-1)\hat{s}^2}{\sigma_0^2} \longrightarrow \left\{ \begin{array}{ll} \chi_0^2 \in \left[\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}\right] & \Rightarrow \text{No rechazo } H_0 \\ \chi_0^2 \notin \left[\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}\right] & \Rightarrow \text{Rechazo } H_0 \end{array} \right.$$

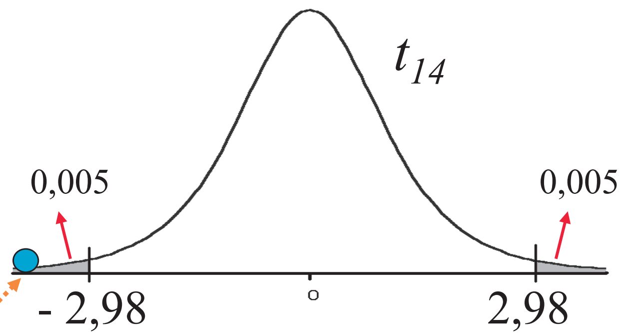
EJEMPLO 1. La resistencia a la compresión de 15 probetas de acero elegidas al azar es:

40,15	65,10	49,50	22,40	38,20
60,40	43,40	26,35	31,20	55,60
47,25	73,20	35,90	45,25	52,40

$$\left.\begin{array}{l} H_0 : \mu = 60 \\ H_1 : \mu \neq 60 \end{array}\right\} \quad \alpha = 0.01$$

$$\frac{\overline{X} - \mu}{\hat{S} / \sqrt{15}} \to t_{14}$$

$$P(-2{,}98 \leq t_{14} \leq 2{,}98) = 0.99$$



Datos:

$$\bar{x} = 45{,}75 \quad \hat{s} = 14{,}2 \quad \Rightarrow \quad t_0 = \frac{45{,}75 - 60}{14{,}2 / \sqrt{15}} = -3.88$$

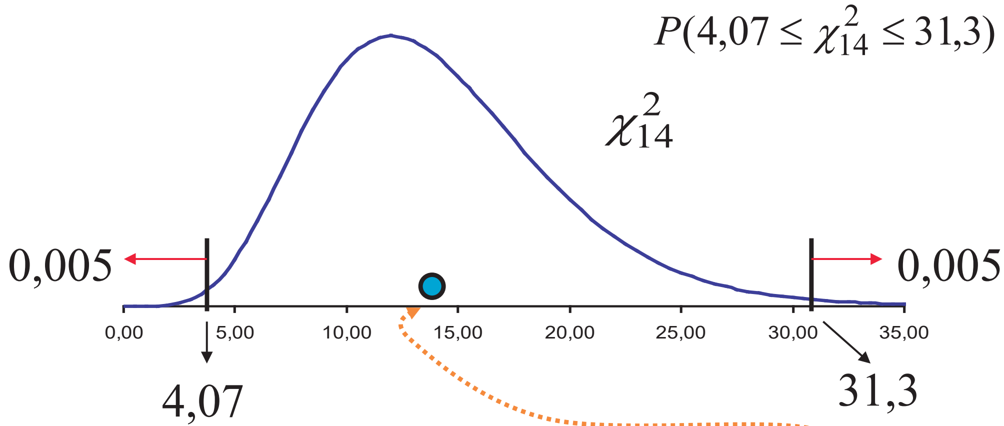
Como 3.88 > 2.98 $\Rightarrow$ Se **rechaza** H_0 con **α=0.01**

$$H_0 : \sigma^2 = 200$$

$$H_1 : \sigma^2 \neq 200$$

$$\frac{(n-1)\hat{S}^2}{\sigma^2} \to \chi^2_{n-1} \qquad \frac{14\hat{S}^2}{\sigma^2} \to \chi^2_{14}$$

$$P(4{,}07 \leq \chi^2_{14} \leq 31{,}3) = 0{,}99$$



Datos:

$$\hat{s}^2 = 201{,}6 \longrightarrow \chi^2_0 = \frac{14 \times 201{,}6}{200} = 14.1$$

Como $4{,}07 \leq 14{,}1 \leq 31{,}3 \Rightarrow$ **No se rechaza H_0**

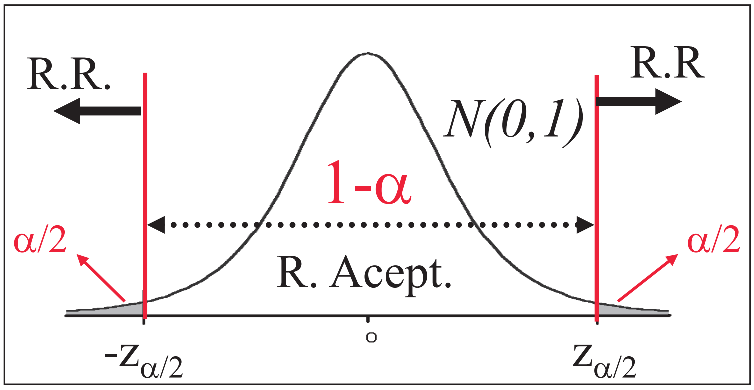
4. Poisson: Contraste para λ

$$X_1, X_2, \ldots, X_n \rightarrow \text{Poisson}\ (\lambda) \qquad \left.\begin{array}{ll} H_0: & \lambda = \lambda_0 \\ H_1: & \lambda \neq \lambda_0 \end{array}\right\rangle \quad \text{nivel de significación } \alpha$$

$$\hat{\lambda} = \overline{X} \overset{aprox}{\rightarrow} \mathrm{N}\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$$

$$Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} \rightarrow \mathrm{N}(0,1)$$

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



$$z_0 = \frac{\hat{\lambda} - \lambda_0}{\sqrt{\frac{\lambda_0}{n}}} \longrightarrow \boxed{\begin{array}{l} |z_0| \leq z_{\alpha/2} \Rightarrow \text{No se rechaza } H_0 \\ |z_0| > z_{\alpha/2} \Rightarrow \text{Se rechaza } H_0 \end{array}}$$

Comparación de dos tratamientos

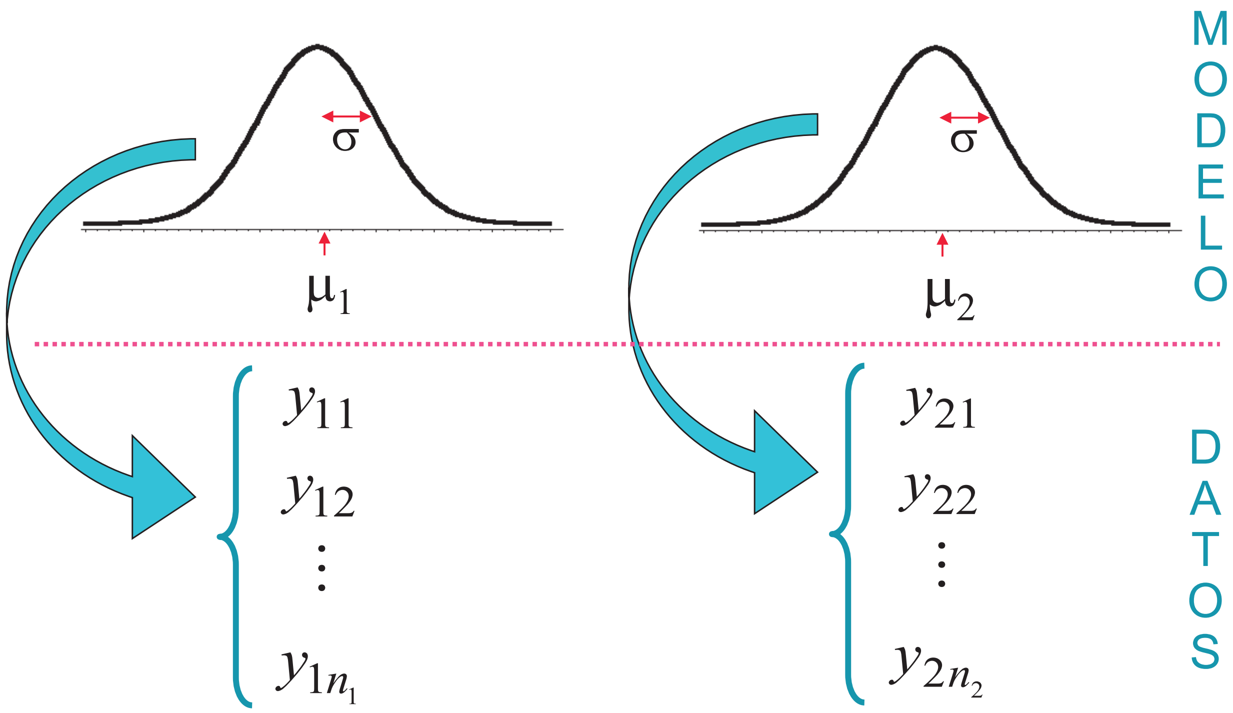
A	B
51,3	29,6
39,4	47,0
26,3	25,9
39,0	13,0
48,1	33,1
34,2	22,1
69,8	34,1
31,3	19,5
45,2	43,8
46,4	24,9

Sea desea comparar dos tratamientos para reducir el nivel de colesterol en la sangre. Se seleccionan 20 individuos y se asignan al azar a dos tipos de dietas A y B. La tabla muestra la reducción conseguida después de dos meses.

Método: 4 pasos

- Definición del modelo de distribución de probabilidad:
  - Hipótesis
  - Parámetros
- Estimación de los parámetros
- Diagnosis de las hipótesis
- Aplicación

Modelo



Modelo: Hipótesis y Parámetros

Hipótesis básicas:

- **Normalidad**

  $y_{ij} \Rightarrow N(\mu_i, \sigma^2)$

- **Homocedasticidad**

  $\text{Var}\,[y_{ij}] = \sigma^2$

- **Independencia**

  $\text{Cov}\,[y_{ij}, y_{kl}] = 0$

Parámetros

μ_1

μ_2

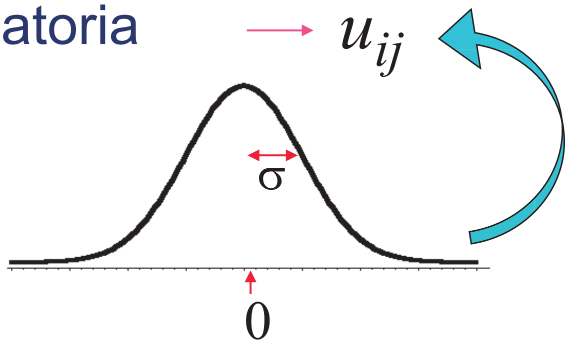
σ^2

Modelo

$$\boxed{y_{ij} = \mu_i + u_{ij}, \qquad u_{ij} \to N(0, \sigma^2)}$$

Las observaciones se descomponen en:

- Parte predecible $\rightarrow$ μ_i
- Parte aleatoria $\rightarrow$ u_{ij}



Estimación medias:

$$\mu_1 : \rightarrow \quad \bar{y}_{1\bullet} = \frac{\sum_{j=1}^{n_1} y_{1j}}{n_1}$$

$$\mu_2 : \rightarrow \quad \bar{y}_{2\bullet} = \frac{\sum_{j=1}^{n_2} y_{2j}}{n_2}$$

A	B
51,3	29,6
39,4	47,0
26,3	25,9
39,0	13,0
48,1	33,1
34,2	22,1
69,8	34,1
31,3	19,5
45,2	43,8
46,4	24,9
43,1	**29,3**

Estimación varianza (residuos)

$$y_{ij} = \mu_i + u_{ij}, \quad u_{ij} \rightarrow N(0, \sigma^2)$$

$$u_{ij} = y_{ij} - \mu_i$$

$$e_{ij} = y_{ij} - \bar{y}_{i\bullet}$$

$$e_{ij} : \quad \textbf{RESIDUO}$$

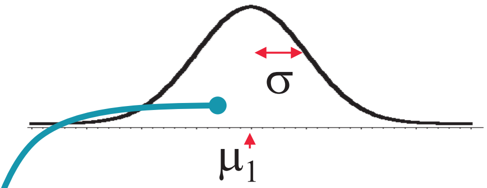
$$\sigma^2 : \rightarrow \hat{s}_R^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i} e_{ij}^2}{n-2}$$

Residuos

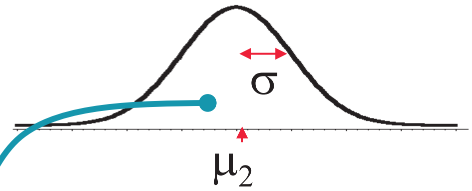
A	B
8,2	0,3
-3,7	17,7
-16,8	-3,4
-4,1	-16,3
5,0	3,8
-8,9	-7,2
26,7	4,8
-11,8	-9,8
2,1	14,5
3,3	-4,4
0,0	**0,0**

$$\sum_{j=1}^{n_i} e_{ij} = 0; \hat{s}_R^2 = 130.95$$

Varianza residual: $\hat{s}_R^2$



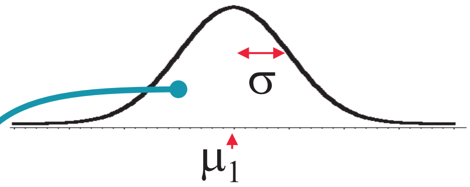
$$\left.\begin{array}{l} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \end{array}\right\} \hat{s}_1^2 = \frac{\sum (y_{1j} - y_{1\bullet})^2}{n_1 - 1}$$



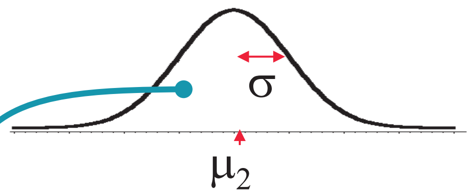
$$\left.\begin{array}{l} y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{array}\right\} \hat{s}_2^2 = \frac{\sum (y_{2j} - y_{2\bullet})^2}{n_2 - 1}$$

$$\hat{s}_R^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i} e_{ij}^2}{n-2} = \frac{n_1 - 1}{n-2}\hat{s}_1^2 + \frac{n_2 - 1}{n-2}\hat{s}_2^2$$

Diferencia de medias: $\bar{y}_{1\bullet} - \bar{y}_{2\bullet}$



$$\left.\begin{array}{l} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \end{array}\right\} \bar{y}_{1\bullet} \to N(\mu_1, \frac{\sigma^2}{n_1})$$



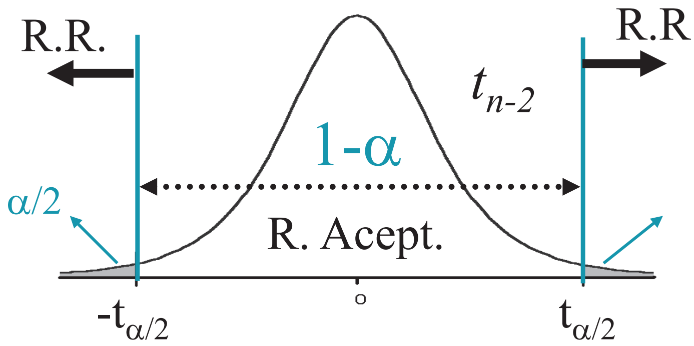
$$\left.\begin{array}{l} y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{array}\right\} \bar{y}_{2\bullet} \to N(\mu_2, \frac{\sigma^2}{n_2})$$

$$\left.\begin{array}{c} \bar{y}_{1\bullet} - \bar{y}_{2\bullet} \to N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}) \\ \dfrac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to N(0,1) \end{array}\right\} \Rightarrow \boxed{\frac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to t_{n-2}}$$

Contraste de igualdad de medias

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{\bar{y}_{1\bullet} - \bar{y}_{2\bullet}}{\hat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow t_{n-2}$$

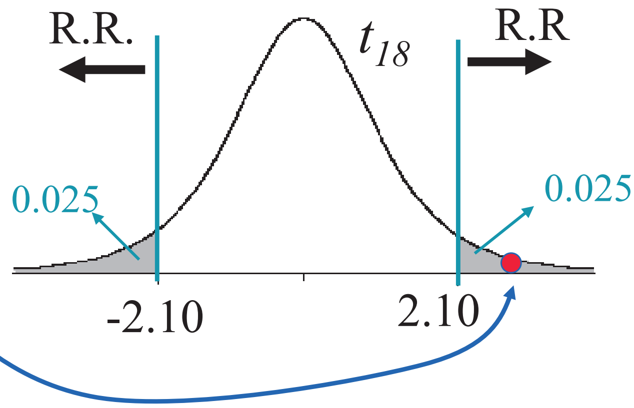


$$|t_0| \leq t_{\alpha/2} \Rightarrow \text{No se rechaza } H_0$$
$$|t_0| > t_{\alpha/2} \Rightarrow \text{Se rechaza } H_0$$

Ejemplo: α = 0.05

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{43.1 - 29.3}{11.44\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.69$$

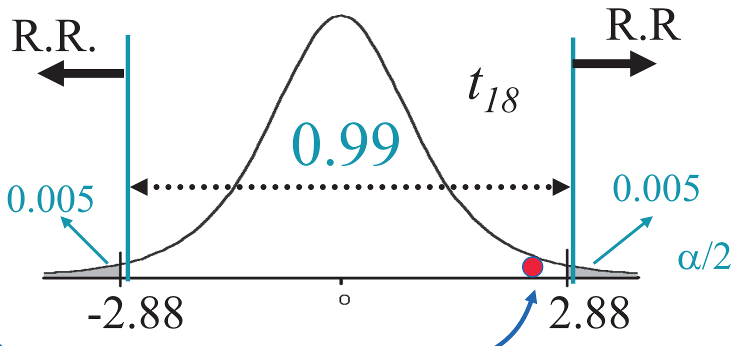


$$\boxed{2.69 > 2.10 \Rightarrow \text{Se rechaza } H_0}$$

Ejemplo: α = 0.01

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{43.1 - 29.3}{11.44\sqrt{\frac{1}{10}+\frac{1}{10}}} = 2.69$$

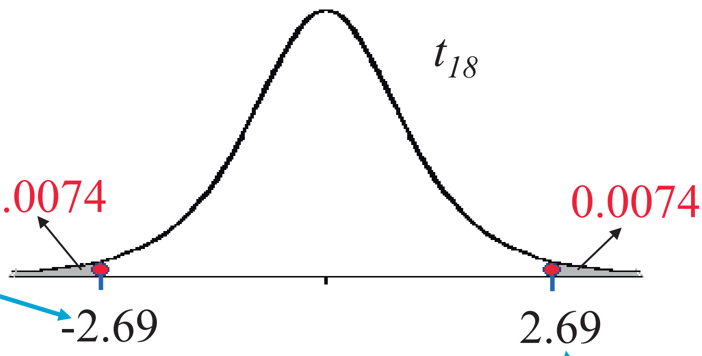


$$\boxed{2.69 \leq 2.88 \Rightarrow \text{No se rechaza } H_0}$$

Nivel crítico (bilateral)

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{43.1 - 29.3}{11.44\sqrt{\frac{1}{10}+\frac{1}{10}}} = 2.69$$



$$\boxed{p - valor = \Pr(|t_{18}| > 2.69) = 0.0147}$$

- $\alpha = 0.05 > p\text{-}valor \Rightarrow$ Se rechaza H_0
- $\alpha = 0.01 < p\text{-}valor \Rightarrow$ No se rechaza H_0

Conclusiones (fijado α)

- Si $|t_o| > t_{\alpha/2}$ se dice que **la diferencia de medias es significativa**. O simplemente que los tratamientos son distintos (tienen medias distintas)

- Si $|t_o| \leq t_{\alpha/2}$ se dice que **la diferencia de medias no es significativa**. No hay evidencia suficiente para afirmar que las medias de los tratamientos sean diferentes.

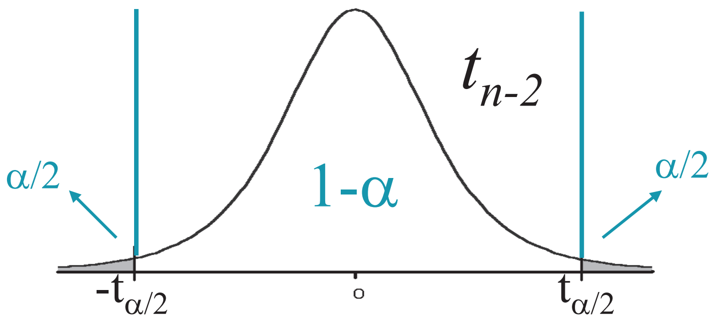
No rechazar H_o, <u>no implica</u> que H_o sea cierta

El resultado $|t_o| \leq t_{\alpha/2}$, (no se rechaza H_o) **no debe interpretarse** como que "*se ha demostrado que las dos medias son iguales*".

No-rechazar la hipótesis nula implica que la diferencia entre las medias $\mu_1 - \mu_2$ no es lo suficientemente grande como para ser detectada con el tamaño muestral dado.

Intervalo de confianza para la diferencia de medias: $\mu_1 - \mu_2$

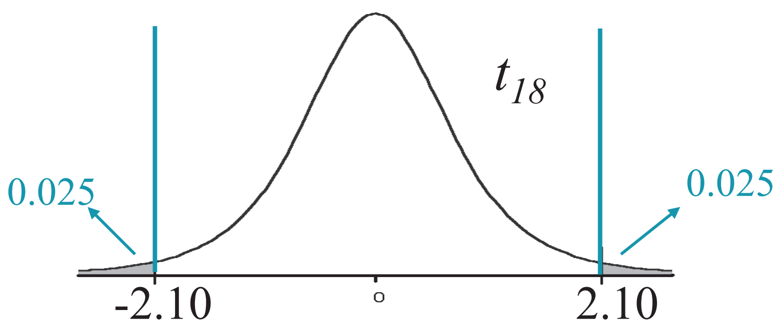
$$\frac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\hat{s}_R\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \rightarrow t_{n-2}$$



$$\Pr\{-t_{\alpha/2} \le \frac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\hat{s}_R\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \le t_{\alpha/2}\} = 1 - \alpha$$

$$\mu_1 - \mu_2 \in (\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) \ \pm \ t_{\alpha/2}\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Ejemplo: intervalo de confianza $\mu_1 - \mu_2$

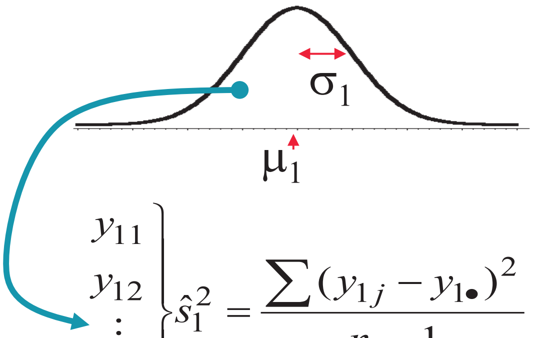


$$\mu_1 - \mu_2 \in (y_{1\bullet} - y_{2\bullet}) \ \pm \ t_{\alpha/2}\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

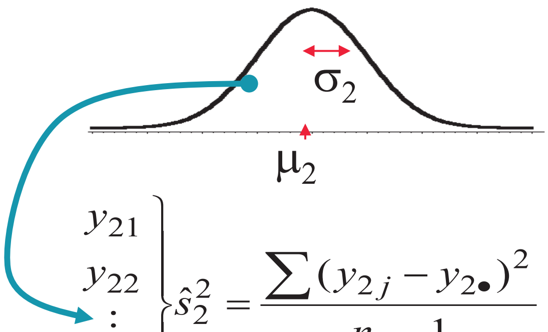
$$\mu_1 - \mu_2 \in (43.1 - 29.3) \pm 2.10 \times 11.44 \times \sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$\mu_1 - \mu_2 \in 13.8 \ \pm \ 10.74$$

Hipótesis de homocedasticidad



$$\left.\begin{array}{c} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \end{array}\right\} \hat{s}_1^2 = \frac{\sum (y_{1j} - y_{1\bullet})^2}{n_1 - 1}$$



$$\left.\begin{array}{c} y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{array}\right\} \hat{s}_2^2 = \frac{\sum (y_{2j} - y_{2\bullet})^2}{n_2 - 1}$$

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

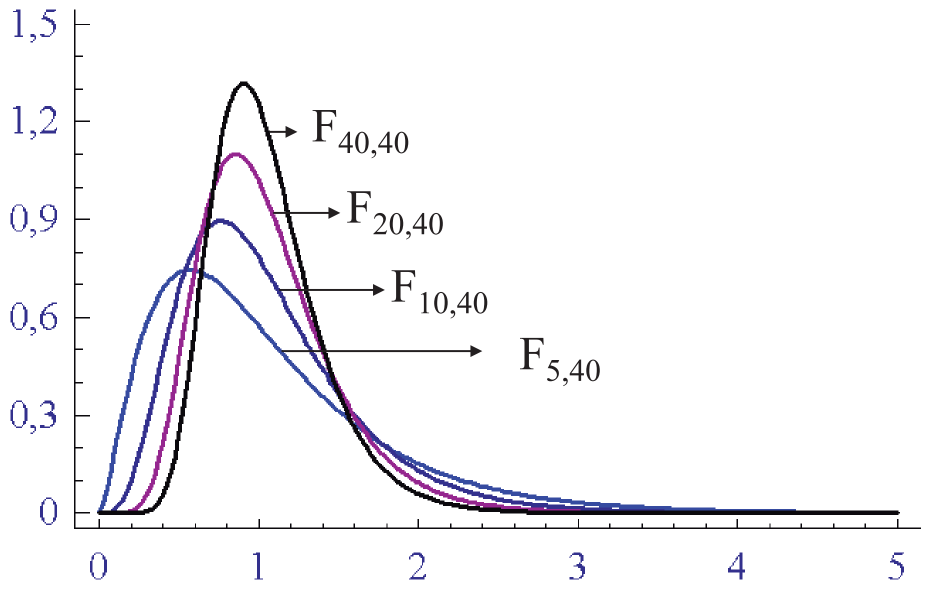
Distribución F

$$\left.\begin{array}{c} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \end{array}\right\} \hat{s}_1^2 = \frac{\sum (y_{1j} - y_{1\bullet})^2}{n_1 - 1} \qquad \left.\begin{array}{c} y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{array}\right\} \hat{s}_2^2 = \frac{\sum (y_{2j} - y_{2\bullet})^2}{n_2 - 1}$$

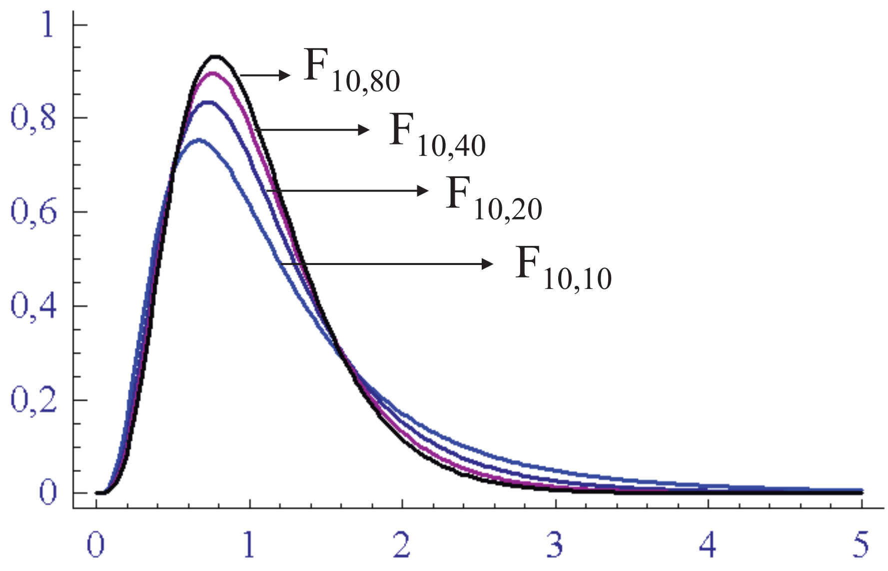
$$\frac{(n_1 - 1)\hat{s}_1^2}{\sigma_1^2} \to \chi^2_{n_1 - 1} \qquad \frac{(n_2 - 1)\hat{s}_2^2}{\sigma_2^2} \to \chi^2_{n_2 - 1}$$

$$\boxed{F = \frac{\chi^2_{n_1-1} / (n_1 - 1)}{\chi^2_{n_2-1} / (n_2 - 1)} = \frac{\dfrac{\hat{s}_1^2}{\sigma_1^2}}{\dfrac{\hat{s}_2^2}{\sigma_2^2}} \to F_{n_1-1, n_2-1}}$$

Distribución F



Algunas distribuciones F



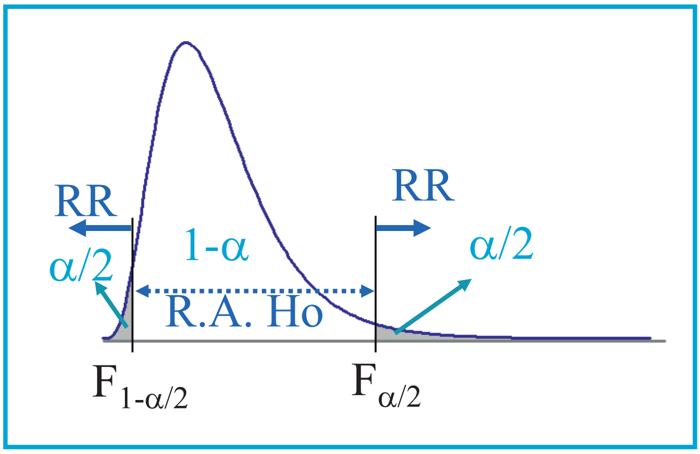
Contraste de igualdad de varianzas

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Si H_0 es cierto $\sigma_1^2 = \sigma_2^2$,

$$F_0 = \frac{\hat{s}_1^2}{\hat{s}_1^2} \rightarrow F_{n_1-1,n_2-1}$$



$$\text{Si } F_0 \in \left[F_{1-\alpha/2}, F_{\alpha/2}\right] \Rightarrow \text{No se rechaza } H_0$$

$$\text{Si } F_0 \notin \left[F_{1-\alpha/2}, F_{\alpha/2}\right] \Rightarrow \text{Se rechaza } H_0$$

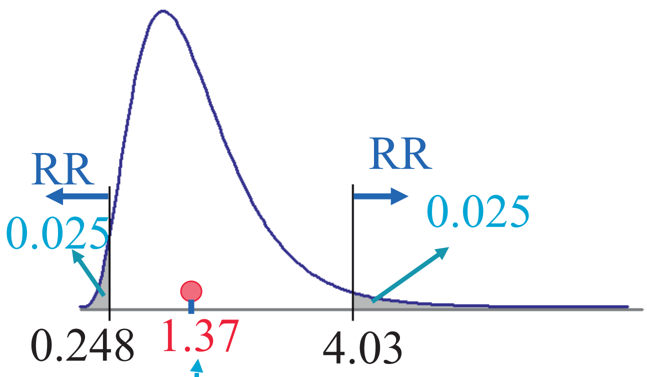
Ejemplo: Contraste de igualdad de varianzas

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$\hat{s}_1^2 = 154.02 \quad \hat{s}_2^2 = 111.7$$

$$F_0 = \frac{154.02}{111.7} = 1.37$$



$$1.37 \in \left[0.248, 4.03\right] \Rightarrow \text{No se rechaza } H_0$$

Tabla F

$\alpha=0.05$

$$F_{\nu_1,\nu_2,\alpha} \quad \Rightarrow \quad P(F_{\nu_1,\nu_2} \geq F_{\nu_1,\nu_2,\alpha}) = \alpha$$

Grados de libertad del numerador: ν_1

Grados de libertad del denominador: ν_2

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	100	120	Inf.	
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,0	253,3	254,3	**1**
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,49	19,50	**2**
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,55	8,53	**3**
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,66	5,63	**4**
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,41	4,40	4,37	**5**
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,71	3,70	3,67	**6**
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,27	3,23	**7**
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,97	2,93	**8**
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,76	2,75	2,71	**9**
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,59	2,58	2,54	**10**
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,46	2,45	2,40	**11**
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,35	2,34	2,30	**12**
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,26	2,25	2,21	**13**
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,19	2,18	2,13	**14**
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,12	2,11	2,07	**15**
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,07	2,06	2,01	**16**
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,02	2,01	1,96	**17**
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,98	1,97	1,92	**18**
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,94	1,93	1,88	**19**
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,91	1,90	1,84	**20**
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,88	1,87	1,81	**21**
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,85	1,84	1,78	**22**
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,82	1,81	1,76	**23**
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,80	1,79	1,73	**24**
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,78	1,77	1,71	**25**
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,76	1,75	1,69	**26**
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,74	1,73	1,67	**27**
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,73	1,71	1,65	**28**
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,71	1,70	1,64	**29**
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,70	1,68	1,62	**30**
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,59	1,58	1,51	**40**
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,95	1,87	1,78	1,74	1,69	1,63	1,58	1,52	1,51	1,44	**50**
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,48	1,47	1,39	**60**
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,89	1,81	1,72	1,67	1,62	1,57	1,50	1,45	1,44	1,35	**70**
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,79	1,70	1,65	1,60	1,54	1,48	1,43	1,41	1,32	**80**
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,86	1,78	1,69	1,64	1,59	1,53	1,46	1,41	1,39	1,30	**90**
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,77	1,68	1,63	1,57	1,52	1,45	1,39	1,38	1,28	**100**
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,37	1,35	1,25	**120**
Inf	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,24	1,22	1,00	**Inf**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**100**	**120**	**Inf.**	

Ejemplo: $P(F_{7,8} \geq 3.50) = 0.05$

Tabla F

$\alpha=0.025$

$$F_{\nu_1,\nu_2,\alpha} \quad \Rightarrow \quad P(F_{\nu_1,\nu_2} \geq F_{\nu_1,\nu_2,\alpha}) = \alpha$$

Grados de libertad del numerador: ν_1

Grados de libertad del denominador: ν_2

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	100	120	Inf.	
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,6	963,3	968,6	976,7	984,9	993,1	997,3	1001,4	1005,6	1009,8	1013,2	1014,0	1018,3	**1**
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,49	39,50	**2**
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,96	13,95	13,90	**3**
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,32	8,31	8,26	**4**
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,08	6,07	6,02	**5**
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,92	4,90	4,85	**6**
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,41	4,36	4,31	4,25	4,21	4,20	4,14	**7**
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,74	3,73	3,67	**8**
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,40	3,39	3,33	**9**
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,15	3,14	3,08	**10**
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,96	2,94	2,88	**11**
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,80	2,79	2,72	**12**
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,67	2,66	2,60	**13**
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,56	2,55	2,49	**14**
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,47	2,46	2,40	**15**
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,40	2,38	2,32	**16**
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,33	2,32	2,25	**17**
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,27	2,26	2,19	**18**
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,22	2,20	2,13	**19**
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,17	2,16	2,09	**20**
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,13	2,11	2,04	**21**
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,09	2,08	2,00	**22**
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,06	2,04	1,97	**23**
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,02	2,01	1,94	**24**
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	2,00	1,98	1,91	**25**
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,97	1,95	1,88	**26**
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,94	1,93	1,85	**27**
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,92	1,91	1,83	**28**
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,90	1,89	1,81	**29**
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,88	1,87	1,79	**30**
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,74	1,72	1,64	**40**
50	5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46	2,38	2,32	2,22	2,11	1,99	1,93	1,87	1,80	1,72	1,66	1,64	1,55	**50**
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,60	1,58	1,48	**60**
70	5,25	3,89	3,31	2,97	2,75	2,59	2,47	2,38	2,30	2,24	2,14	2,03	1,91	1,85	1,78	1,71	1,63	1,56	1,54	1,44	**70**
80	5,22	3,86	3,28	2,95	2,73	2,57	2,45	2,35	2,28	2,21	2,11	2,00	1,88	1,82	1,75	1,68	1,60	1,53	1,51	1,40	**80**
90	5,20	3,84	3,26	2,93	2,71	2,55	2,43	2,34	2,26	2,19	2,09	1,98	1,86	1,80	1,73	1,66	1,58	1,50	1,48	1,37	**90**
100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,24	2,18	2,08	1,97	1,85	1,78	1,71	1,64	1,56	1,48	1,46	1,35	**100**
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,45	1,43	1,31	**120**
Inf	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,30	1,27	1,00	**Inf**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**100**	**120**	**Inf.**	

Ejemplo: $P(F_{7,8} \geq 4.53) = 0.025$

Tabla F

$\alpha=0.01$

$$F_{\nu_1,\nu_2,\alpha} \quad \Rightarrow \quad P(F_{\nu_1,\nu_2} \geq F_{\nu_1,\nu_2,\alpha}) = \alpha$$

Grados de libertad del numerador: ν_1

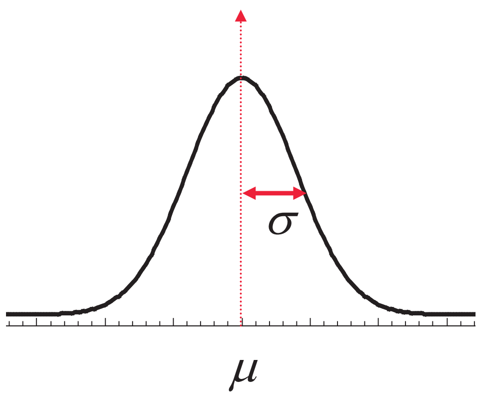
Grados de libertad del denominador: ν_2

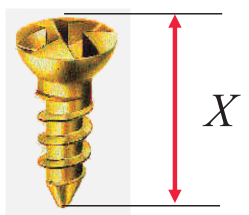
ν_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	100	120	Inf.	
1	4052,2	4999,3	5403,5	5624,3	5764,0	5859,0	5928,3	5981,0	6022,4	6055,9	6106,7	6157,0	6208,7	6234,3	6260,4	6286,4	6313,0	6333,9	6339,5	6365,6	1
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,50	2
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,24	26,22	26,13	3
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,58	13,56	13,46	4
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,13	9,11	9,02	5
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,99	6,97	6,88	6
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,75	5,74	5,65	7
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,96	4,95	4,86	8
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,41	4,40	4,31	9
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,01	4,00	3,91	10
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,71	3,69	3,60	11
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,47	3,45	3,36	12
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,27	3,25	3,17	13
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,11	3,09	3,00	14
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,98	2,96	2,87	15
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,86	2,84	2,75	16
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,76	2,75	2,65	17
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,68	2,66	2,57	18
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,60	2,58	2,49	19
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,54	2,52	2,42	20
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,48	2,46	2,36	21
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,42	2,40	2,31	22
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,37	2,35	2,26	23
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,33	2,31	2,21	24
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,29	2,27	2,17	25
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,25	2,23	2,13	26
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,22	2,20	2,10	27
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,19	2,17	2,06	28
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,16	2,14	2,03	29
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,13	2,11	2,01	30
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,94	1,92	1,80	40
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,56	2,42	2,27	2,18	2,10	2,01	1,91	1,82	1,80	1,68	50
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,75	1,73	1,60	60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,45	2,31	2,15	2,07	1,98	1,89	1,78	1,70	1,67	1,54	70
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,42	2,27	2,12	2,03	1,94	1,85	1,75	1,65	1,63	1,49	80
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,39	2,24	2,09	2,00	1,92	1,82	1,72	1,62	1,60	1,46	90
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,37	2,22	2,07	1,98	1,89	1,80	1,69	1,60	1,57	1,43	100
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,56	1,53	1,38	120
Inf	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,36	1,32	1,00	Inf
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	100	120	Inf.	

$$\textit{Ejemplo}: \quad P(F_{7,8} \geq 6.18) = 0.01$$

Contraste χ^2 de bondad de ajuste

POBLACIÓN





¿Tienen los datos distribución Normal?

MUESTRA n

$$x_1, x_2, \ldots, x_n$$

$$¿\, X \rightarrow N(\mu, \sigma)\,?$$

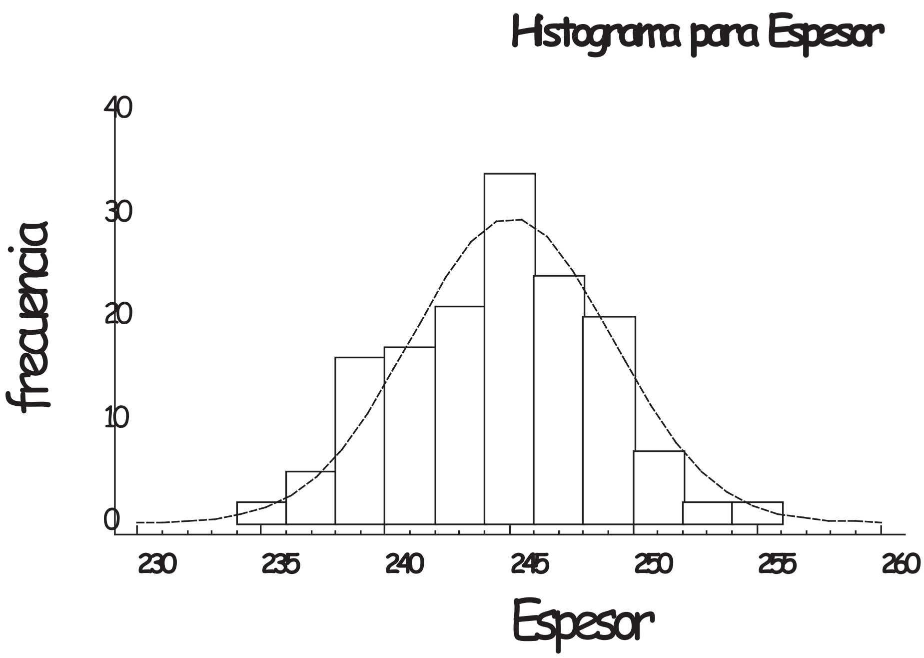
Datos Conocidos

$$\hat{\mu} = \bar{x}$$
$$\hat{\sigma} = \hat{s}$$

Espesores de 150 obleas de Silicio (micras)

240	235	240	240	247	237	243	242	236	239
243	237	243	242	245	239	245	245	239	240
250	246	244	246	255	242	248	248	241	242
253	249	249	249	250	247	251	251	246	243
248	246	246	248	249	245	250	249	242	244
238	240	245	240	237	242	244	242	243	239
242	241	250	243	239	244	246	245	246	240
245	246	250	246	243	246	246	248	247	250
251	247	247	250	247	251	250	243	252	252
247	249	248	248	246	248	246	246	247	250
239	240	238	241	242	243	241	241	241	241
242	243	240	245	244	245	239	244	243	243
246	244	245	243	245	247	244	245	245	249
250	248	248	247	248	252	250	249	248	255
248	245	246	245	245	249	246	247	246	253

¿Tiene distribución normal?



Contraste χ^2 de bondad de ajuste

$$H_0 : X_i, \forall i = 1,2,\ldots,n \to f_X$$
$$H_1 : X_i, \forall i = 1,2,\ldots,n \not\to f_X$$

Si H_0 es cierto:

Clases	Fr. Observada	Fr. Esperada
$c_o \le x_i < c_1$	O_1	E_1
$c_1 \le x_i < c_2$	O_2	E_2
$\vdots$	$\vdots$	$\vdots$
$c_{k-1} \le x_i < c_k$	O_k	E_k
$\vdots$	$\vdots$	$\vdots$
$c_{K-1} \le x_i < c_K$	O_K	E_K

$$\boxed{\begin{aligned} E_k &= np_k \\ p_k &= \Pr(c_{k-1} \le X < c_k) \end{aligned}}$$

$$\sum_{k=1}^{K} O_k = \sum_{k=1}^{K} E_k = n$$

$$\boxed{\sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \to \chi^2_{K-r-1}}$$

Justificación del contraste χ^2

$$O_k \to B(n, p_k) \xrightarrow{n\to\infty} \boxed{O_k \to N\left(np_k, \sqrt{np_k(1-p_k)}\right)}$$

$$E_k = np_k$$

$$\frac{O_k - np_k}{\sqrt{np_k \underbrace{(1-p_k)}_{\cong 1}}} \to N(0,1)$$

$$\frac{O_k - E_k}{\sqrt{E_k}} \to N(0,1) \Rightarrow \sum_{k=1}^{K} \left(\frac{O_k - E_k}{\sqrt{E_k}} \right)^2 \xrightarrow{aprox} \chi^2_{K-r-1}$$

K $\equiv$ Nº de CLASES ***r*** $\equiv$ Nº de parámetro estimados

Obleas: Frecuencias Observadas

Clase		Fr. Observada
		Oi
-inf	238	7
238	240	16
240	242	17
242	244	21
244	246	34
246	248	24
248	250	20
250	252	7
252	+inf	4
Total		150

$$\bar{x} = 245.1 \quad \text{micras}$$

$$\hat{s} = 4 \quad \text{micras}$$

Frecuencias esperadas N(245.1;4)

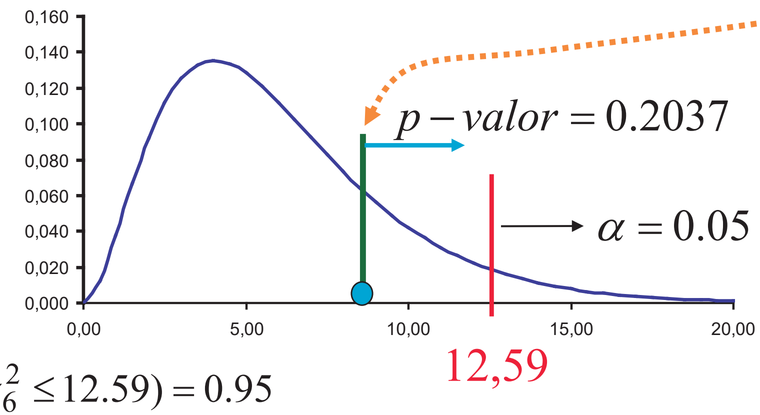
Clase		Fr. Observada	Fr. Esperada
		Oi	Ei
-inf	238	7	5,88
238	240	16	9,58
240	242	17	17,71
242	244	21	25,71
244	246	34	29,35
246	248	24	26,34
248	250	20	18,58
250	252	7	10,3
252	+inf	4	6,54
Total		150	150

$$\left.\begin{aligned}
p_4 &= \Pr(242 \le X \le 244) \\
&= \Pr\left(\frac{242-245.1}{4} \le \frac{X-245.1}{4} \le \frac{244-245.1}{4}\right) \\
&= P(-0.275 \le Z \le -0.775) \\
&= 0.1714 \\
E_4 &= np_4 = 150 \times 0.1714 = 25.71
\end{aligned}\right\}$$

Contraste de Normalidad

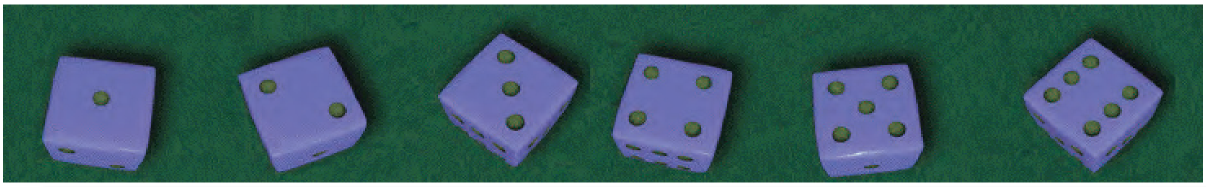
$$H_0 : x_i \rightarrow \text{Normal}$$
$$H_1 : x_i \nrightarrow \text{Normal}$$

$$\chi_0^2 = \frac{(7-5.88)^2}{5.88} + \frac{(16-9,58)^2}{9.58} + \cdots + \frac{(4-6.54)^2}{6.54} = 8.5$$



$$\Pr(\chi_6^2 \leq 12.59) = 0.95$$

$\chi_0^2 = 8.5 < 12.59 \Rightarrow$ No se rechaza la hipótesis de normalidad



Se ha lanzado 300 veces un dado y se han obtenido los resultados:

Resultados	Observadas
1	49
2	59
3	49
4	51
5	43
6	49
Total	300

¿Se puede afirmar con $\alpha=0.05$ que el dado está desequilibrado?

$$\Pr(X=i)=1/6, \quad \forall i=1,2,\ldots,6$$

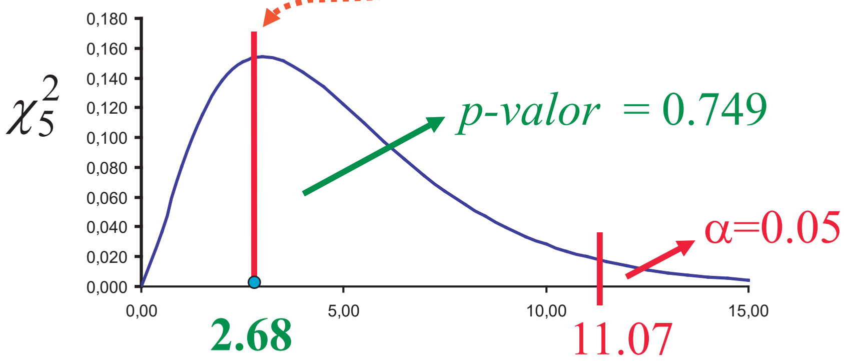
$$H_o : x_i \rightarrow \text{Equiprobable}$$

$$\mathrm{H}_1 : x_i \rightarrow \text{No equiprobable}$$

Resultados	Observados	Esperados
1	49	50
2	59	50
3	49	50
4	51	50
5	43	50
6	49	50
Total	300	300

$$\chi_0^2 = \frac{(49-50)^2}{50} + \frac{(59-50)^2}{50} + \cdots + \frac{(49-50)^2}{50} = 2.68$$

$$\Pr(\chi_5^2 \le 11.07) = 0.95 \Rightarrow \quad \chi_0^2 = 2.68 < 11.07 \Rightarrow \text{No se Rechaza } \mathrm{H}_0$$



Capítulo 6: Contraste de hipótesis

1. Un proceso industrial fabrica piezas cuya longitud en mm se distribuye según una distribución normal de media $\mu = 190$. Para controlar el proceso y se toman muestras de cinco piezas cada hora. En la última hora se han obtenido los siguientes valores:

   $$183.34, 211.68, 197.77, 205.89, 197.32$$

   La desviación típica de la distribución es conocida y supondremos que no cambia, $\sigma = 10$ mm. Contrasta la hipótesis de que la media del proceso μ es efectivamente 190 con $\alpha = 0.05$. Calcula el pvalor del contraste.

2. Repetir el ejercicio 1 con σ desconocida: Un proceso industrial fabrica piezas cuya longitud en mm se distribuye según una distribución normal de media $\mu = 190$. Para controlar el proceso se toman cada hora una muestra de cinco piezas. En la última hora se han obtenido los siguientes valores:

   $$183.34, 211.68, 197.77, 205.89, 197.32$$

   Contrasta la hipótesis de que la media del proceso μ es efectivamente 190 con $\alpha = 0.05$. Calcula el pvalor del contraste.

3. Un proceso de fabricación cuando está bajo control produce piezas cuya longitud tiene distribución normal de media $\mu = 190$ y varianza $\sigma^2 = 100$. Para controlar que la media no cambia, se toman muestras de tamaño $n = 5$ y se realiza el contraste de hipótesis:

   $$H_0 : \mu = 190$$

   $$H_1 : \mu \neq 190$$

   con $\alpha = 0.05$. Si se acepta H_0 el proceso continua fabricando. Si se rechaza H_0 el operario detiene el proceso y comprueba si todo está funcionando correctamente.

   a) ¿Cuál es la probabilidad de que si el proceso funciona correctamente, al tomar una muestra de tamaño 5, se rechace H_0 y se revise el proceso?

   b) ¿Cuál es la probabilidad de que cuando el proceso se desajuste y fabrique piezas con media $\mu = 195$, al tomar una muestra de tamaño 5 y realizar el contraste, se acepte H_0 y se concluya que la media del proceso no ha cambiado?

   c) Finalmente, se pide calcular la curva de potencia del contraste de media anterior. En todo el ejercicio suponer que la varianza es igual a 100 y no cambia.

4. Un proceso de fabricación en condiciones ideales produce piezas cuya longitud tiene distribución normal de media $\mu = 190$ y varianza $\sigma^2 = 100$. Para controlar que la varianza no cambia, se toman muestras de tamaño $n = 5$ y se realiza el contraste de hipótesis:

   $$H_0 : \sigma^2 = 100$$

$$H_1 : \sigma^2 \neq 100$$

con $\alpha = 0.05$.

a) Se ha tomado una muestra obteniendo los siguientes valores:

$$183.34,\ 211.68,\ 197.77,\ 205.89,\ 197.32$$

Realizar el contraste con $\alpha = 0.05$.

b) Calcula el pvalor del contraste en el caso de que la hipótesis alternativa sea (A) $H_1 : \sigma^2 > 100$, (B) $H_1 : \sigma^2 > 100$ o (C) $H_1 : \sigma^2 \neq 100$.

c) ¿Cuál es la probabilidad de que si el proceso funciona correctamente, al tomar una muestra de tamaño 5, se rechace H_0 ?

d) ¿Cuál es la probabilidad de que cuando el proceso se desajuste y $\sigma^2 = 200$, al tomar una muestra de tamaño 5 y realizar el contraste, se acepte H_0 y se concluya que la varianza del proceso no ha cambiado?

5. Para contrastar unilateralmente que la esperanza μ de una variable aleatoria normal es 10, se toma una muestra de tamaño 16 y se rechaza la hipótesis en el caso en que la media muestral sea mayor que 11, aceptándose en el caso contrario. Sabiendo que la desviación típica de la población es $\sigma = 2$, ¿cúal es la probabilidad de error de tipo I de este contraste?. ¿Cúal sería la probabilidad de error de tipo II del contraste si el valor verdadero de la esperanza fuese 12?.

6. Una medicina estándar es efectiva en el 75 % de los casos en los que se aplica. Se ha comprobado un nuevo medicamento en 100 pacientes, observándose su efectividad en 85 de ellos. ¿ Es la nueva medicina más efectiva que la estándar ? (Contrastar con $\alpha = 0.05$).

7. Un empresario quiere comprar una empresa que fabrica cojinetes. Durante los 5 últimos años la proporción de cojinetes defectuosos se ha mantenido en un 3 %. Para verificar esto, se toma una muestra de 200 cojinetes y obtiene que 9 son defectuosos.

   a) ¿ Se puede concluir que la proporción de cojinetes defectuosos ha aumentado?

   b) Calcular la potencia del contraste planteado anteriormente en función de p. Calcular la probabilidad de error de tipo II cuando la hipótesis alternativa es $p = 0.06$, siendo p la proporción de defectuosos.(Nota: Utilícese la aproximación normal y $\alpha = 0,05$.).

   c) Utilizando R calcula y dibuja las curvas de potencia para tamaños muestrales $n = 200$, $n = 300$, $n = 500$ y $n = 1000$. Interpreta el resultado.

8. Teniendo en cuenta que si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una variable aleatoria exponencial con función de densidad,

$$f(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0,\ \mu > 0;$$

el estadístico $U = 2n\overline{X}/\mu$ tiene distribución χ^2_{2n}, donde $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$; resolver las cuestiones siguientes:

a) El tiempo de funcionamiento de un equipo electrónico es una variable aleatoria con distribución exponencial. Se han tomado los tiempos de funcionamiento hasta el fallo de 30 equipos elegidos al azar, obteniéndose 6.2×10^3 horas de media. Contrastar con nivel de significación igual a 0.05, $H_0 : \mu = 5 \times 10^3$ horas, frente a $H_1 : \mu > 5 \times 10^3$ horas; indicando el pvalor.

b) Calcula la probabilidad de error tipo II cuando $\mu = 7 \times 10^3$ horas.

9. En la tabla se reproducen los valores para la velocidad de la luz en el aire que obtuvieron Michelson y Newcomb en 1879 y 1882 respectivamente, utilizando métodos distintos . A los datos originales en km/s se les ha restado 299000 para facilitar su tratamiento matemático.

	1	2	3	4	5	6	7	8	9	10
Mic.	850	740	900	1070	930	850	950	980	980	880
New.	883	816	778	796	682	711	611	599	1051	781

	11	12	13	14	15	$\overline{y}$	$\widehat{s}$
Mic.	1000	980	930	650	760	896.6	111.91
New.	578	796	774	820	772	763.2	119.47

a) Realizar el contraste de igualdad de medias utilizando el modelo de comparación de dos tratamientos con $\alpha = 0.05$.

b) Contrastar la hipótesis de que la variabilidad es la misma en los experimentos de Michelson y Newcomb. ($\alpha = 0.05$)

c) El valor utilizado actualmente para la velocidad de la luz en el vacío es 299.792,5 km/s (792,5 si se le resta 299.000). Michelson y Newcomb midieron en sus experimentos la velocidad de la luz en el aire. Contrastar para cada grupo por separado si la media de las observaciones es igual a 792.5 con $\alpha = 0.05$.

10. Se estudian dos tipos de neumáticos con los resultados siguientes:

Tipo	n_i	$\overline{y}_i(Km)$	$\widehat{s}_i(Km)$
A	121	27465	2500
B	121	27572	3000

a) Calcular un intervalo de confianza al 99 % para $\frac{\sigma_1^2}{\sigma_2^2}$.

b) Obtener un intervalo de confianza al 95 % para $\mu_1 - \mu_2$.

11. Se dispone de rendimientos de dos máquinas. Los resultados de la máquina A son 137.5; 140.7; 106.9; 175.1; 177.3; 120.4; 77.9 y 104.2, mientras que los resultados para la B son: 103.3; 121.7; 98.4; 161.5; 167.8 y 67.3. Suponiendo que los rendimientos de cada máquina tienen distribuciones normales, contrastar si las dos máquinas tienen el mismo rendimiento medio? Resolverlo con R, comprobando previamente la hipótesis de homocedasticidad (nivel de significación 0.05 para todos los contrastes).

12. Un fabricante de automóviles debe elegir entre un determinado tipo de piezas de acero suministradas por un proveedor A y otras suministradas por otro proveedor B. Para proceder a la elección se ha analizado la resistencia a la tracción de las piezas suministradas por ambos proveedores, tomando una muestra de tamaño 10 de las piezas del primero, y otra de tamaño 12 del segundo. La resistencia media de la muestra de A es de 54000 unidades y la de la muestra de B es de 49000 unidades, siendo las desviaciones típicas muestrales corregidas $\widehat{s}_A = 2100$ y $\widehat{s}_B = 1900$. Las resistencias de las piezas de ambos proveedores se distribuyen normalmente. Las piezas del proveedor B son más baratas que las del proveedor A, por lo que estas últimas sólo son rentables si tienen una resistencia media al menos 2000 unidades mayor que las de B, y la misma variabilidad.

    a) ¿A qué proveedor habría que comprar las piezas a la vista de los resultados muestrales?

    b) Obtener un intervalo de confianza del 90\ %

13. La estatura de 60 niños de una escuela infantil se resume en la siguiente tabla de frecuencias, dónde la última columna muestra la frecuencia esperada bajo la hipótesis de normalidad.

Intervalo	Frecuencia Observada	Frecuencia Esperada
41.5-43.5	4	4.08
43.5-45.5	7	5.58
45.5-47.5	12	9.06
47.5-49.5	8	11.27
49.5-51.5	6	11.27
51.5-53.5	11	9.08
53.5-55.5	9	5.58
55.5-57.5	3	4.08
Total	60	60

    ¿Se puede aceptar la hipótesis de normalidad de los datos ($\alpha = 0.05$) ?

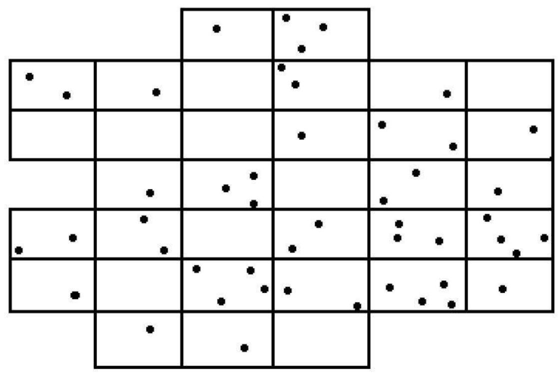
14. Se tira 120 veces un dado y se obtienen los resultados de la tabla

VALOR	1	2	3	4	5	6
FRECUENCIA	20	14	23	12	26	25

    Contrastar la hipótesis de que el dado está equilibrado y que, por tanto, sus caras son equiprobables. (Tómese $\alpha = 0.05$).

15. Un modelo sísmico indica que la distribución de los epicentros de sismos en una región debería seguir una distribución de Poisson en el plano. Un grupo de expertos pretende

contrastar si ese modelo se cumple, para ello ha representado un mapa de la región dividido en cuadrículas de tamaño 100 km^2, y ha señalado con puntos las posiciones de los epicentros (véase figura adjunta). Realizar el contraste χ^2 de bondad de ajuste con nivel de significación $\alpha = 0,05$ proporcionando el pvalor del contraste.



Universidad Politécnica de Madrid
Octubre, 2018

Estadística: Formulario y Tablas

Escuela Técnica Superior de Ingenieros Industriales
Laboratorio de Estadística.
C/ José Gutiérrez Abascal, 2, 28006 Madrid

1. Distribuciones de Probabilidad Univariantes

Bernoulli	$P(X=x)=p^x(1-p)^{1-x}$	$x=0,1$ $0\leq p\leq 1$	$\mathrm{E}[X]=p$ $\mathrm{Var}[X]=p(1-p)$
Binomial	$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$	$x=0,1,\ldots,n$ $0\leq p\leq 1,\ n\in\mathbb{N}$	$\mathrm{E}[X]=np$ $\mathrm{Var}[X]=np(1-p)$
Geométrica	$P(X=x)=(1-p)^{x-1}p$	$x=1,2,\ldots,\infty$ $0\leq p\leq 1$	$\mathrm{E}[X]=\frac{1}{p}$ $\mathrm{Var}[X]=\frac{1-p}{p^2}$
Hipergeométrica	$P(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$x=0,1,\ldots,n$ $N,M,n\in\mathbb{N}$	$\mathrm{E}[X]=n\frac{M}{N}$ $\mathrm{Var}[X]=n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$
Binomial Negativa	$P(X=x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$	$x=r,r+1,\ldots$ $0\leq p\leq 1,\ r\in\mathbb{N}$	$\mathrm{E}[X]=\frac{r}{p}$ $\mathrm{Var}[X]=\frac{r(1-p)}{p^2}$
Poisson	$P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}$	$x=0,1,2,\ldots$ $\lambda>0$	$\mathrm{E}[X]=\lambda$ $\mathrm{Var}[X]=\lambda$
Uniforme Discreta	$P(X=x)=\frac{1}{n}$	$x=1,2,\ldots,n$ $n\geq 1$	$\mathrm{E}[X]=\frac{n+1}{2}$ $\mathrm{Var}[X]=\frac{n^2-1}{12}$
Exponencial	$f_X(x)=\lambda e^{-\lambda x}$	$x\geq 0$ $\lambda\geq 0$	$\mathrm{E}[X]=\frac{1}{\lambda}$ $\mathrm{Var}[X]=\frac{1}{\lambda^2}$
Normal o Gausiana $X\sim N(\mu,\sigma^2)$	$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty<x<\infty$ $\mu\in\mathbb{R},\ \sigma>0$	$\mathrm{E}[X]=\mu$ $\mathrm{Var}(X)=\sigma^2$
Uniforme Continua	$f_X(x)=\frac{1}{b-a}$	$a\leq x\leq b$ $-\infty<a<b<\infty$	$E[X]=\frac{a+b}{2}$ $\mathrm{Var}[X]=\frac{(b-a)^2}{12}$
Chi-cuadrado $X\sim\chi^2_n$	$f_X(x)=\frac{x^{(n/2)-1}e^{-x/2}}{2^{(n/2)}\Gamma(n/2)}$	$x\geq 0$ $n\in N$	$\mathrm{E}[X]=n$ $\mathrm{Var}[X]=2n$
t de Student $X\sim t_n$	$f_X(x)=\frac{1}{\sqrt{\pi n}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$	$-\infty<x<\infty$ $n\in\mathbb{N}$	$E[X]=0,\ n\geq 2$ $\mathrm{Var}[X]=\frac{n}{n-2},\ n\geq 3$
F $X\sim F_{m,n}$	$f_X(x)=\frac{\Gamma\left(\frac{m+n}{2}\right)m^{\frac{m}{2}}n^{\frac{n}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}x^{\frac{m}{2}-1}(mx+n)^{\frac{m+n}{2}}$	$x>0$ $m,n\in\mathbb{N}$	$\mathrm{E}[X]=\frac{n}{m-2},n\geq 3$ $\mathrm{Var}[X]=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)},n\geq 5$

(Nota: $\Gamma(z)=\int_0^\infty t^{z-1}e^{-z}dt$)

2. Distribuciones de Probabilidad Multivariantes

1. Normal n-dimensional

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{|M|}} e^{-\frac{1}{2}(x-\mu)^T M^{-1}(x-\mu)} \quad x^T = (x_1, x_2, ..., x_n)$$

Parámetros: $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad M = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_n^2 \end{pmatrix}$

2. Normal bi-dimensional

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right)$$

Parámetros: $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$

Propiedades:

a) $E[X] = \mu_X, E[Y] = \mu_Y, \mathrm{Var}[X] = \sigma_X^2, \mathrm{Var}[Y] = \sigma_Y^2, \mathrm{Cov}(X,Y) = \rho\sigma_X\sigma_Y$

b) La distribución de X dado $Y = y$ es normal, con las siguientes media y varianza:

$$E[X|Y=y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)$$

$$\mathrm{Var}[X|Y=y] = \sigma_X^2(1-\rho^2)$$

La distribución de Y dado $X = x$ es normal, con las siguientes media y varianza:

$$E[Y|X=x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_x)$$

$$\mathrm{Var}[Y|X=x] = \sigma_Y^2(1-\rho^2)$$

c) X en Y son independientes si y solo si $\rho = 0$.

3. Distribución Multinomial (discreta)

$$p(x_1, x_2, ..., x_k) = \frac{n!}{x_1!\, x_2!\, \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

con $x_i = 0, 1, ..., n, \sum_{i=1}^k x_i = n, \sum_{i=1}^k p_i = 1$

Parámetros: $n, p_1, p_2, ..., p_k$

Propiedades:

a) $E[X_i] = np_i$

b) $\mathrm{Var}[X_i] = np_1(1-p_i)$

c) $\mathrm{Cov}[X_i, X_j] = -np_ip_j$

3. Combinatoria

1. Número de disposiciones de n elementos tomados de k en k (**sí** importa el orden y **sí** se puede repetir)

$$n^k$$

   Ejemplo: Del conjunto $\{a, b, c\}$ se pueden formar $3^2 = 9$ disposiciones con dos elementos $aa, ab, ac, ba, bb, bc, ca, cb, cc$.

2. Número de permutaciones de n elementos tomados de k en k (**sí** importa el orden y **no** se puede repetir)

$$\frac{n!}{(n-k)!}$$

   Ejemplo: Del conjunto $\{a, b, c\}$ se pueden formar $3 \times 2 = 6$ disposiciones con dos elementos sin repetir ab, ac, ba, bc, ca, cb.

3. Número de combinaciones de n elementos tomados de k en k (**no** importa el orden y **no** se puede repetir)

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

   Ejemplo: Del conjunto $\{a, b, c\}$ se pueden formar $\binom{3}{2} = 3$ combinaciones con dos elementos sin repetir ab, ac, bc.

4. Número de disposiciones de n elementos tomados de k en k (**no** importa el orden y **sí** se puede repetir)

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

   Ejemplo: Del conjunto $\{a, b, c\}$ se pueden formar $\binom{4}{2} = 6$ disposiciones de dos elementos con repetición sin importar orden aa, ab, ac, bb, bc, cc.

4. Intervalos de Confianza

Se proporcionan intervalos con confianza $1-\alpha$ para los parámetros, siendo

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad \text{y} \quad \hat{s}^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1}$$

los estimadores de los parámetros de la distribución normal y $\hat{p}$ y $\hat{\lambda}$ los estimadores de p y λ en las distribuciones Binomial y Poisson, respectivamente.

1. Media de Normal $N(\mu,\sigma)$ con σ conocido

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

siendo $z_{\alpha/2}$ el valor de $Z \sim N(0,1)$ tal que $P(Z > z_{\alpha/2}) = \alpha/2$

2. Media de Normal $N(\mu,\sigma)$ con σ desconocido

$$\bar{x} - t_{\alpha/2}\frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2}\frac{\hat{s}}{\sqrt{n}}$$

siendo $t_{\alpha/2}$ el valor de T de Student con $n-1$ grados de libertad tal que $P(T > t_{\alpha/2}) = \alpha/2$

3. Varianza de Normal $N(\mu,\sigma)$

$$\frac{(n-1)\hat{s}^2}{\chi^2_b} \leq \sigma^2 \leq \frac{(n-1)\hat{s}^2}{\chi^2_a}$$

siendo χ^2_a y χ^2_b los valores de la distribibición chi-cuadrado χ^2_{n-1} con $n-1$ grados de libertad que cumple

$$P(\chi^2_{n-1} \leq \chi^2_a) = P(\chi^2_{n-1} \geq \chi^2_b) = \alpha/2$$

4. Diferencia de medias de dos distribuciones normales $(N(\mu_1,\sigma),\ N(\mu_2,\sigma))$ con la misma varianza con muestras independientes de tamaño n_1 y n_2

$$(\bar{y}_1 - \bar{y}_2) - t_{\alpha/2}\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{y}_1 - \bar{y}_2) + t_{\alpha/2}\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

siendo $t_{\alpha/2}$ el valor de T de Student con $n_1 + n_2 - 2$ grados de libertad tal que $P(T > t_{\alpha/2}) = \alpha/2$ y

$$\hat{s}^2_R = \frac{(n_1-1)\hat{s}^2_1 + (n_2-1)\hat{s}^2_2}{n_1+n_2-2}$$

5. p en la distribución binomial $B(n,p)$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

siendo $z_{\alpha/2}$ el valor de $Z \sim N(0,1)$ tal que $P(Z > z_{\alpha/2}) = \alpha/2$

6. λ en la distribución Poisson

$$\hat{\lambda} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}$$

siendo $z_{\alpha/2}$ el valor de $Z \sim N(0,1)$ tal que $P(Z > z_{\alpha/2}) = \alpha/2$

5. Contraste de hipótesis

Se muestran los contrastes bilaterales con nivel de significación α. Se utiliza la misma notación que en la sección de Intervalos de Confianza.

1. Contraste de media de la distibución normal con varianza conocida

$$H_0:\ \mu=\mu_0; \qquad H_1:\ \mu\neq\mu_0$$

$$\text{se rechaza } H_0 \text{ si } \quad \frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

siendo $z_{\alpha/2}$ el valor de $Z\sim N(0,1)$ tal que $P(Z>z_{\alpha/2})=\alpha/2$

2. Contraste de varianza de la distribución normal

$$H_0:\ \sigma^2=\sigma_0^2; \quad H_1:\ \sigma^2\neq\sigma_0^2$$

$$\text{se rechaza } H_0 \text{ si } \quad \frac{(n-1)\hat{s^2}}{\sigma_0^2} \notin [\chi_a^2,\chi_b^2]$$

siendo χ_a^2 y χ_b^2 los valores de la distribición chi-cuadrado χ_{n-1}^2 con $n-1$ grados de libertad que cumple

$$P(\chi_{n-1}^2\leq\chi_a^2)=P(\chi_{n-1}^2\geq\chi_b^2)=\alpha/2$$

3. Contraste igualdad de medias de dos distribuciones normales $(N(\mu_1,\sigma),\ N(\mu_2,\sigma))$ con la misma varianza con muestras independientes de tamaño n_1 y n_2

$$H_0:\ \mu_1=\mu_2; \quad H_1:\ \mu_1\neq\mu_2$$

$$\text{se rechaza } H_0 \text{ si } \quad \frac{|\bar{y}_1-\bar{y}_2|}{\hat{s}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{\alpha/2}$$

siendo $t_{\alpha/2}$ el valor de T de Student con n_1+n_2-2 grados de libertad tal que $P(T>t_{\alpha/2})=\alpha/2$ y

$$\hat{s}_R^2=\frac{(n_1-1)\hat{s}_1^2+(n_2-1)\hat{s}_2^2}{n_1+n_2-2}$$

4. Contraste de igualdad de varianzas de dos distribuciones normales $(N(\mu_1,\sigma_1),\ N(\mu_2,\sigma_2))$ con muestras independientes de tamaño n_1 y n_2

$$H_0:\ \sigma_1^2=\sigma_2^2; \quad H_1:\ \sigma_1^2\neq\sigma_2^2$$

$$\text{se rechaza } H_0 \text{ si } \quad \frac{\hat{s}_1^2}{\hat{s}_2^2} \notin [F_a,F_b]$$

siendo F_a y F_b los valores de la distribición F de n_1-1, n_2-1 de grados de libertad que cumple

$$P(F_{n_1-1,n_2-1}\leq F_a)=P(F_{n_1-1,n_2-1}\geq F_b)=\alpha/2$$

5. Contraste sobre p de la distribución binomial $B(n,p)$

$$H_0:\ p=p_0 \quad H_1:\ p\neq p_0$$

$$\text{se rechaza } H_0 \text{ si } \quad \frac{|\hat{p}-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{\alpha/2}$$

siendo $z_{\alpha/2}$ el valor de $Z\sim N(0,1)$ tal que $P(Z>z_{\alpha/2})=\alpha/2$

6. Contraste sobre λ de la distribución de Poisson

$$H_0:\ \lambda=\lambda_0; \quad H_1:\ \lambda\neq\lambda_0$$

$$\text{se rechaza } H_0 \text{ si } \quad \frac{|\hat{\lambda}-\lambda_0|}{\sqrt{\frac{\lambda_0}{n}}} > z_{\alpha/2}$$

siendo $z_{\alpha/2}$ el valor de $Z\sim N(0,1)$ tal que $P(Z>z_{\alpha/2})=\alpha/2$

6. Tablas

6.1. Distribución Normal Estandar

La tabla muestra los valores z tales que $P(Z \leq z)$.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998
4.1	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999

Ejm: $P(Z \leq 1{,}96) = 0{,}97500$

6.2. Distribución χ^2

La tabla muestra los valores x tales que $P(\chi^2_n \geq x) = \alpha$

	α								
n	0.995	0.99	0.975	0.95	0.5	0.05	0.025	0.01	0.005
1	0.00004	0.0002	0.001	0.004	0.455	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	1.386	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	2.366	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	3.357	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	4.351	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	5.348	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	6.346	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	7.344	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	8.343	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	9.342	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	10.341	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	11.340	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	12.340	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	13.339	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	14.339	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	15.338	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	16.338	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	17.338	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	18.338	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	19.337	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	20.337	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	21.337	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	22.337	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	23.337	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	24.337	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	25.336	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	26.336	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	27.336	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	28.336	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	29.336	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	39.335	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	49.335	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	59.335	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	69.334	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	79.334	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	89.334	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	99.334	124.342	129.561	135.807	140.169
110	75.550	78.458	82.867	86.792	109.334	135.480	140.917	147.414	151.948
120	83.852	86.923	91.573	95.705	119.334	146.567	152.211	158.950	163.648

Ejm: $P(\chi^2_9 \geq 19{,}02) = 0{,}025$

6.3. Distribución t-Student

La tabla muestra los valores x tales que $P(t_n \geq x) = \alpha$.

					α					
n	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	0.847	1.044	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	0.846	1.042	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
Inf	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Ejm: $P(t_9 \geq 2{,}262) = 0{,}025$

6.4. Distribución F($\alpha = 0{,}05$)

La tabla muestra los valores x tales que $P(F_{m,n} \geq x) = 0{,}05$.

m

n	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951
90	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043	1.986	1.938
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927
Inf	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831

Ejm: $P(F_{7,8} \geq 3{,}50) = 0{,}05$

Distribución F($\alpha = 0{,}05$) (continuación)

La tabla muestra los valores x tales que $P(F_{m,n} \geq x) = 0{,}05$.

m

n	12	15	20	24	30	40	60	100	120	Inf
1	243.906	245.950	248.013	249.052	250.095	251.143	252.196	253.041	253.253	254.314
2	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.486	19.487	19.496
3	8.745	8.703	8.660	8.639	8.617	8.594	8.572	8.554	8.549	8.526
4	5.912	5.858	5.803	5.774	5.746	5.717	5.688	5.664	5.658	5.628
5	4.678	4.619	4.558	4.527	4.496	4.464	4.431	4.405	4.398	4.365
6	4.000	3.938	3.874	3.841	3.808	3.774	3.740	3.712	3.705	3.669
7	3.575	3.511	3.445	3.410	3.376	3.340	3.304	3.275	3.267	3.230
8	3.284	3.218	3.150	3.115	3.079	3.043	3.005	2.975	2.967	2.928
9	3.073	3.006	2.936	2.900	2.864	2.826	2.787	2.756	2.748	2.707
10	2.913	2.845	2.774	2.737	2.700	2.661	2.621	2.588	2.580	2.538
11	2.788	2.719	2.646	2.609	2.570	2.531	2.490	2.457	2.448	2.404
12	2.687	2.617	2.544	2.505	2.466	2.426	2.384	2.350	2.341	2.296
13	2.604	2.533	2.459	2.420	2.380	2.339	2.297	2.261	2.252	2.206
14	2.534	2.463	2.388	2.349	2.308	2.266	2.223	2.187	2.178	2.131
15	2.475	2.403	2.328	2.288	2.247	2.204	2.160	2.123	2.114	2.066
16	2.425	2.352	2.276	2.235	2.194	2.151	2.106	2.068	2.059	2.010
17	2.381	2.308	2.230	2.190	2.148	2.104	2.058	2.020	2.011	1.960
18	2.342	2.269	2.191	2.150	2.107	2.063	2.017	1.978	1.968	1.917
19	2.308	2.234	2.155	2.114	2.071	2.026	1.980	1.940	1.930	1.878
20	2.278	2.203	2.124	2.082	2.039	1.994	1.946	1.907	1.896	1.843
21	2.250	2.176	2.096	2.054	2.010	1.965	1.916	1.876	1.866	1.812
22	2.226	2.151	2.071	2.028	1.984	1.938	1.889	1.849	1.838	1.783
23	2.204	2.128	2.048	2.005	1.961	1.914	1.865	1.823	1.813	1.757
24	2.183	2.108	2.027	1.984	1.939	1.892	1.842	1.800	1.790	1.733
25	2.165	2.089	2.007	1.964	1.919	1.872	1.822	1.779	1.768	1.711
26	2.148	2.072	1.990	1.946	1.901	1.853	1.803	1.760	1.749	1.691
27	2.132	2.056	1.974	1.930	1.884	1.836	1.785	1.742	1.731	1.672
28	2.118	2.041	1.959	1.915	1.869	1.820	1.769	1.725	1.714	1.654
29	2.104	2.027	1.945	1.901	1.854	1.806	1.754	1.710	1.698	1.638
30	2.092	2.015	1.932	1.887	1.841	1.792	1.740	1.695	1.683	1.622
40	2.003	1.924	1.839	1.793	1.744	1.693	1.637	1.589	1.577	1.509
50	1.952	1.871	1.784	1.737	1.687	1.634	1.576	1.525	1.511	1.438
60	1.917	1.836	1.748	1.700	1.649	1.594	1.534	1.481	1.467	1.389
70	1.893	1.812	1.722	1.674	1.622	1.566	1.505	1.450	1.435	1.353
80	1.875	1.793	1.703	1.654	1.602	1.545	1.482	1.426	1.411	1.325
90	1.861	1.779	1.688	1.639	1.586	1.528	1.465	1.407	1.391	1.302
100	1.850	1.768	1.676	1.627	1.573	1.515	1.450	1.392	1.376	1.283
Inf	1.752	1.666	1.571	1.517	1.459	1.394	1.318	1.243	1.221	1.000

6.5. Distribución F($\alpha = 0{,}025$)

La tabla muestra los valores x tales que $P(F_{m,n} \geq x) = 0{,}025$.

m

n	1	2	3	4	5	6	7	8	9	10
1	647.789	799.500	864.163	899.583	921.848	937.111	948.217	956.656	963.285	968.627
2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419
4	12.218	10.649	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774
21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700
23	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613
26	5.659	4.265	3.670	3.329	3.105	2.945	2.824	2.729	2.653	2.590
27	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568
28	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.547
29	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.669	2.592	2.529
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270
70	5.247	3.890	3.309	2.975	2.754	2.595	2.474	2.379	2.302	2.237
80	5.218	3.864	3.284	2.950	2.730	2.571	2.450	2.355	2.277	2.213
90	5.196	3.844	3.265	2.932	2.711	2.552	2.432	2.336	2.259	2.194
100	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179
Inf	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114	2.048

Ejm: $P(F_{7,8} \geq 4{,}53) = 0{,}025$

Distribución F($\alpha = 0{,}025$) (continuación)

La tabla muestra los valores x tales que $P(F_{m,n} \geq x) = 0{,}025$

m

n	12	15	20	24	30	40	60	100	120	Inf
1	976.708	984.867	993.103	997.249	1001.414	1005.598	1009.800	1013.175	1014.020	1018.258
2	39.415	39.431	39.448	39.456	39.465	39.473	39.481	39.488	39.490	39.498
3	14.337	14.253	14.167	14.124	14.081	14.037	13.992	13.956	13.947	13.902
4	8.751	8.657	8.560	8.511	8.461	8.411	8.360	8.319	8.309	8.257
5	6.525	6.428	6.329	6.278	6.227	6.175	6.123	6.080	6.069	6.015
6	5.366	5.269	5.168	5.117	5.065	5.012	4.959	4.915	4.904	4.849
7	4.666	4.568	4.467	4.415	4.362	4.309	4.254	4.210	4.199	4.142
8	4.200	4.101	3.999	3.947	3.894	3.840	3.784	3.739	3.728	3.670
9	3.868	3.769	3.667	3.614	3.560	3.505	3.449	3.403	3.392	3.333
10	3.621	3.522	3.419	3.365	3.311	3.255	3.198	3.152	3.140	3.080
11	3.430	3.330	3.226	3.173	3.118	3.061	3.004	2.956	2.944	2.883
12	3.277	3.177	3.073	3.019	2.963	2.906	2.848	2.800	2.787	2.725
13	3.153	3.053	2.948	2.893	2.837	2.780	2.720	2.671	2.659	2.595
14	3.050	2.949	2.844	2.789	2.732	2.674	2.614	2.565	2.552	2.487
15	2.963	2.862	2.756	2.701	2.644	2.585	2.524	2.474	2.461	2.395
16	2.889	2.788	2.681	2.625	2.568	2.509	2.447	2.396	2.383	2.316
17	2.825	2.723	2.616	2.560	2.502	2.442	2.380	2.329	2.315	2.247
18	2.769	2.667	2.559	2.503	2.445	2.384	2.321	2.269	2.256	2.187
19	2.720	2.617	2.509	2.452	2.394	2.333	2.270	2.217	2.203	2.133
20	2.676	2.573	2.464	2.408	2.349	2.287	2.223	2.170	2.156	2.085
21	2.637	2.534	2.425	2.368	2.308	2.246	2.182	2.128	2.114	2.042
22	2.602	2.498	2.389	2.331	2.272	2.210	2.145	2.090	2.076	2.003
23	2.570	2.466	2.357	2.299	2.239	2.176	2.111	2.056	2.041	1.968
24	2.541	2.437	2.327	2.269	2.209	2.146	2.080	2.024	2.010	1.935
25	2.515	2.411	2.300	2.242	2.182	2.118	2.052	1.996	1.981	1.906
26	2.491	2.387	2.276	2.217	2.157	2.093	2.026	1.969	1.954	1.878
27	2.469	2.364	2.253	2.195	2.133	2.069	2.002	1.945	1.930	1.853
28	2.448	2.344	2.232	2.174	2.112	2.048	1.980	1.922	1.907	1.829
29	2.430	2.325	2.213	2.154	2.092	2.028	1.959	1.901	1.886	1.807
30	2.412	2.307	2.195	2.136	2.074	2.009	1.940	1.882	1.866	1.787
40	2.288	2.182	2.068	2.007	1.943	1.875	1.803	1.741	1.724	1.637
50	2.216	2.109	1.993	1.931	1.866	1.796	1.721	1.656	1.639	1.545
60	2.169	2.061	1.944	1.882	1.815	1.744	1.667	1.599	1.581	1.482
70	2.136	2.028	1.910	1.847	1.779	1.707	1.628	1.558	1.539	1.436
80	2.111	2.003	1.884	1.820	1.752	1.679	1.599	1.527	1.508	1.400
90	2.092	1.983	1.864	1.800	1.731	1.657	1.576	1.503	1.483	1.371
100	2.077	1.968	1.849	1.784	1.715	1.640	1.558	1.483	1.463	1.347
Inf	1.945	1.833	1.708	1.640	1.566	1.484	1.388	1.296	1.268	1.000

6.6. Distribución F($\alpha = 0{,}01$)

La tabla muestra los valores x tales que $P(F_{m,n} \geq x) = 0{,}01$

m

n	1	2	3	4	5	6	7	8	9	10
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632
70	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.585
80	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551
90	6.925	4.849	4.007	3.535	3.228	3.009	2.845	2.715	2.611	2.524
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503
Inf	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321

Ejm: $P(F_{7,8} \geq 6{,}18) = 0{,}01$

Distribución F($\alpha = 0{,}01$) (continuación)

La tabla muestra los valores x tales que $P(F_{m,n} \geq x) = 0{,}01$

m

n	12	15	20	24	30	40	60	100	120	Inf
1	6106.321	6157.285	6208.730	6234.631	6260.649	6286.782	6313.030	6334.110	6339.391	6365.864
2	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.489	99.491	99.499
3	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.240	26.221	26.125
4	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.577	13.558	13.463
5	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.130	9.112	9.020
6	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.987	6.969	6.880
7	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.755	5.737	5.650
8	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.963	4.946	4.859
9	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.415	4.398	4.311
10	4.706	4.558	4.405	4.327	4.247	4.165	4.082	4.014	3.996	3.909
11	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.708	3.690	3.602
12	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.467	3.449	3.361
13	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.272	3.255	3.165
14	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.112	3.094	3.004
15	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.977	2.959	2.868
16	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.863	2.845	2.753
17	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.764	2.746	2.653
18	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.678	2.660	2.566
19	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.602	2.584	2.489
20	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.535	2.517	2.421
21	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.475	2.457	2.360
22	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.422	2.403	2.305
23	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.373	2.354	2.256
24	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.329	2.310	2.211
25	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.289	2.270	2.169
26	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.252	2.233	2.131
27	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.218	2.198	2.097
28	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.187	2.167	2.064
29	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.158	2.138	2.034
30	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.131	2.111	2.006
40	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.938	1.917	1.805
50	2.562	2.419	2.265	2.183	2.098	2.007	1.909	1.825	1.803	1.683
60	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.749	1.726	1.601
70	2.450	2.306	2.150	2.067	1.980	1.886	1.785	1.695	1.672	1.540
80	2.415	2.271	2.115	2.032	1.944	1.849	1.746	1.655	1.630	1.494
90	2.389	2.244	2.088	2.004	1.916	1.820	1.716	1.623	1.598	1.457
100	2.368	2.223	2.067	1.983	1.893	1.797	1.692	1.598	1.572	1.427
Inf	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.358	1.325	1.000