ESTADÍSTICA

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

Descriptiva

Probabilidad

Inferencia

Escuela Técnica Superior de Ingenieros Industriales
Universidad Politécnica de Madrid

ESTADÍSTICA **CURSO 2011/2012**

GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES (GITI)

GRADO EN INGENIERÍA QUÍMICA (GIQ)

DEPARTAMENTO DE INGENIERÍA DE ORGANIZACIÓN,
ADMINISTRACIÓN DE EMPRESAS Y ESTADÍSTICA.

PROGRAMA

1. **ESTADÍSTICA DESCRIPTIVA**. Descripción de una variable: datos y distribución de frecuencias. Representaciones gráficas: histogramas. Medidas de centralización y dispersión. Medidas de asimetría y curtosis. Diagramas de cajas y diagrama de tallos y hojas. Transformaciones lineales y no lineales. Gráficos de dispersión. Covarianza y correlación. Matriz de varianzas y covarianzas. Transformaciones lineales de varias variables.

2. **PROBABILIDAD Y VARIABLE ALEATORIA**. Probabilidad. Definición y propiedades. Probabilidad condicionada. Independencia de sucesos. Teorema de la probabilidad total y teorema de Bayes. Variables aleatorias discretas y continuas. Funciones de probabilidad, densidad y distribución. Esperanza, varianza y momentos. Transformaciones lineales y no lineales. Generación de números aleatorios (Método de Monte Carlo).

3. **MODELOS UNIVARIANTES DE PROBABILIDAD**. Distribución binomial y distribución geométrica. Distribución de Poisson y distribución exponencial. Distribución normal.

4. **MODELOS MULTIVARIANTES**. Distribución conjunta. Distribuciones marginales y condicionadas. Independencia de variables aleatorias. Esperanza de vectores aleatorios. Covarianza y correlación. Matriz de varianzas. Esperanzas y varianzas condicionadas. Suma de variables aleatorias. Teorema central del límite. Relación entre binomial, Poisson y normal. Aplicación a control de recepción. Planes de muestreo simple por atributos. Riesgo del comprador y riesgo del vendedor. La distribución normal n-dimensional.

5. **ESTIMACIÓN PUNTUAL**. Muestra y población. Muestreo aleatorio simple. La identificación del modelo. Método de los momentos. Método de máxima verosimilitud. Propiedades de los estimadores. Distribución de medias, varianzas y proporciones muestrales. Distribución χ^2.

6. **INTERVALOS DE CONFIANZA**. Intervalo para una proporción. Intervalo para la media de la distribución de Poisson. Intervalos para medias de distribuciones normales: varianza conocida y varianza desconocida. Distribución t de Student. Intervalo para varianzas de distribuciones normales. Intervalo general (asintótico) para la media.

7. **CONTRASTES DE HIPÓTESIS**. Contraste para la proporción, contraste para la media y la varianza de distribuciones normales. Contraste para la media de la distribución de Poisson. Comparación de dos medias y comparación de dos varianzas. Distribución F. Contrastes de bondad de ajustes de χ^2 y Kolmogorov-Smirnov.

LIBROS RECOMENDADOS

• *Fundamentos de Estadística.* Daniel Peña, Alianza Editorial (2010).

• *Problemas resueltos de Estadística.* J. Juan, J.G. Palomo, M.J. Sánchez, e I. Sánchez. Síntesis (2000).

1. Estadística descriptiva

Curso 2011-2012

Estadística

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

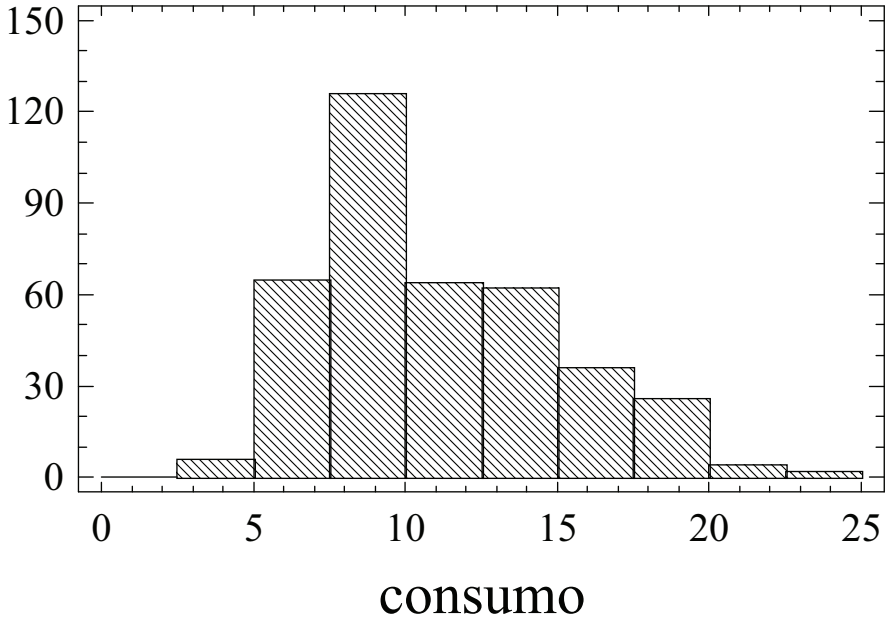
Tipos de datos

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

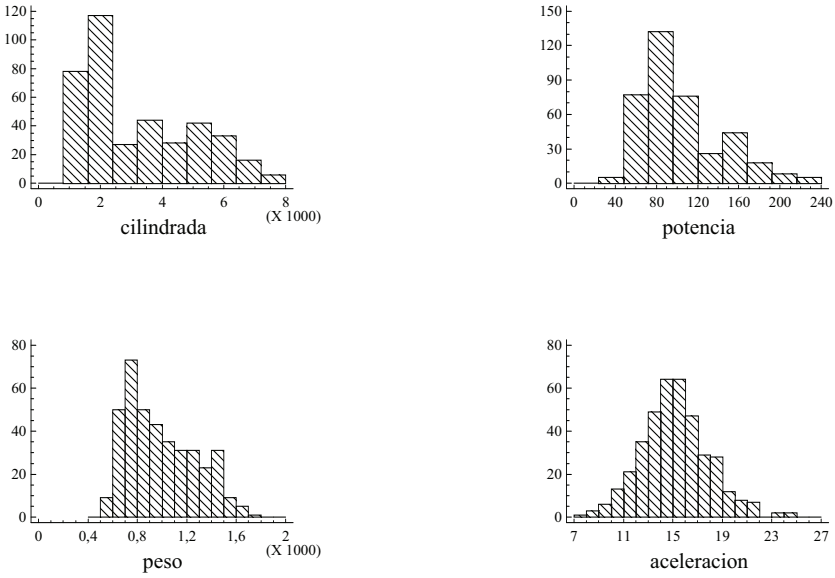
Distribución de frecuencias: consumo *l/100 km*

Clase	Limite Inferior	Limite Superior	Punto Medio	Frecuencia Absoluta	Frecuencia Relativa
1	0,0	2,5	1,25	0	0,0000
2	2,5	5,0	3,75	6	0,0153
3	5,0	7,5	6,25	65	0,1662
4	7,5	10,0	8,75	126	0,3223
5	10,0	12,5	11,25	64	0,1637
6	12,5	15,0	13,75	62	0,1586
7	15,0	17,5	16,25	36	0,0921
8	17,5	20,0	18,75	26	0,0665
9	20,0	22,5	21,25	4	0,0102
10	22,5	25,0	23,75	2	0,0051
Total				391	1,0000

Histograma



Histogramas para coches



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)

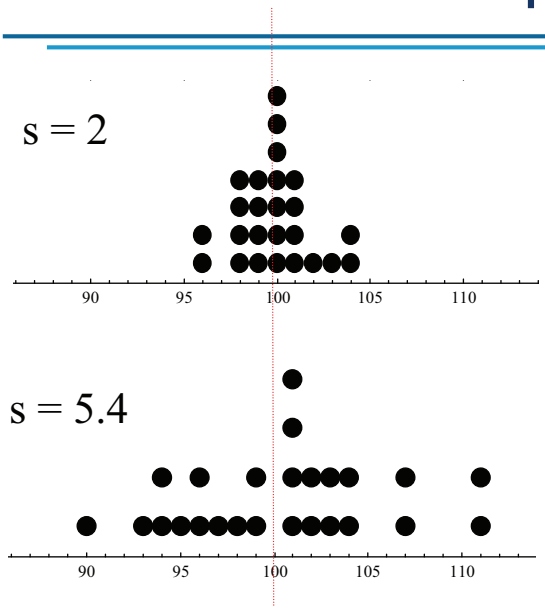
$$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n}$$

Media armónica (si $x_i > 0$ para todo i)

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

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

Medidas de dispersión



Media 100

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

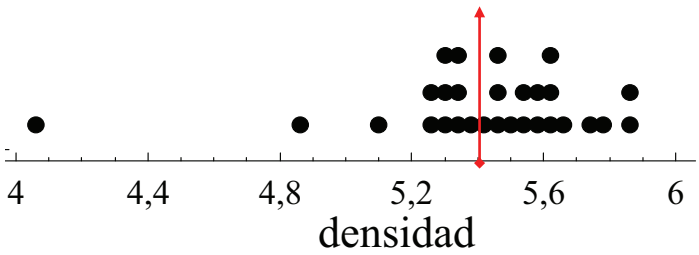
Desviación Típica

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

Varianza : s^2

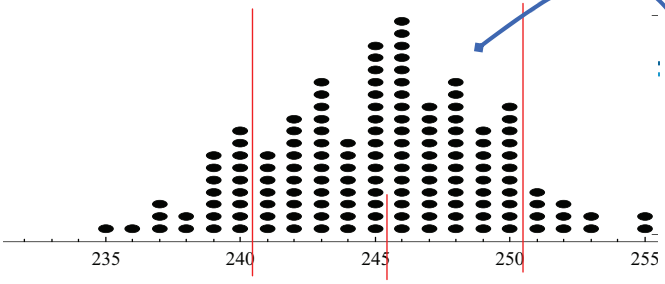
Densidad de la tierra (Cavendish, 1798)

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



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

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

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

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

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

Mediana y Cuartiles

$x_1, x_2, \ldots, x_n$

Datos ordenados

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

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

Cuartiles

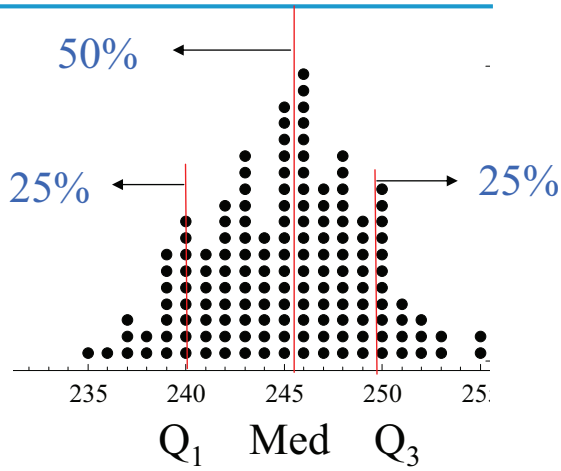
$$Q_1 = x_{(r)} \qquad Q_3 = x_{(s)}$$

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

Mediana y Cuartiles

$x_1, x_2, \ldots, x_n$

$$\boxed{\begin{array}{l} \text{Mediana} : (Med) \\ fr(x_i \leq Med) = 0.50 \end{array}}$$



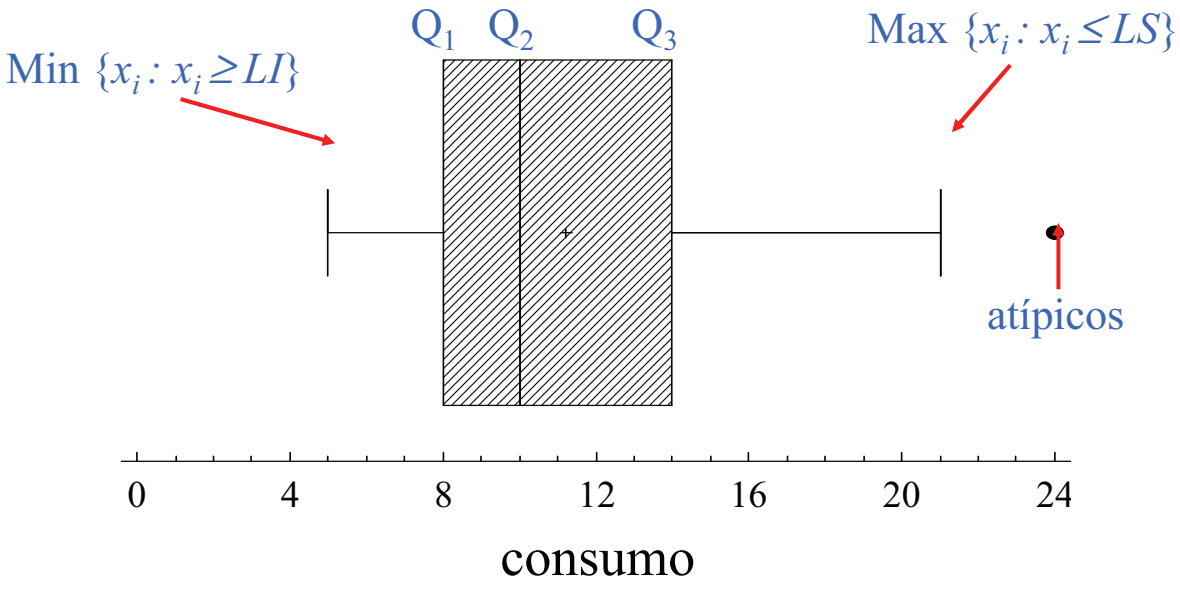
Cuartiles

$$\boxed{\begin{array}{c} Q_1 \\ fr(x_i \leq Q_1) = 0.25 \end{array}} \qquad \boxed{\begin{array}{c} Q_3 \\ fr(x_i \leq Q_3) = 0.75 \end{array}}$$

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

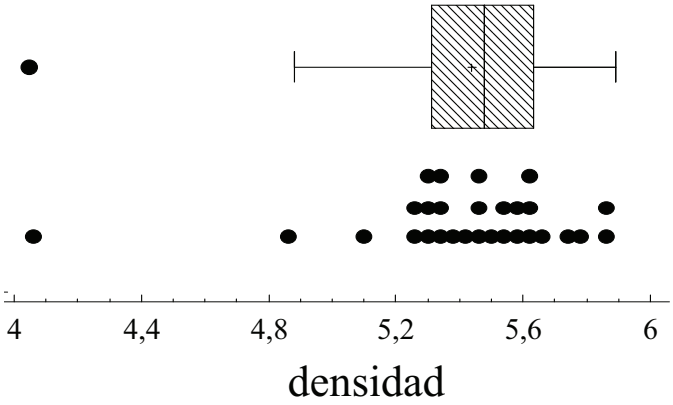
Diagrama de caja



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

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

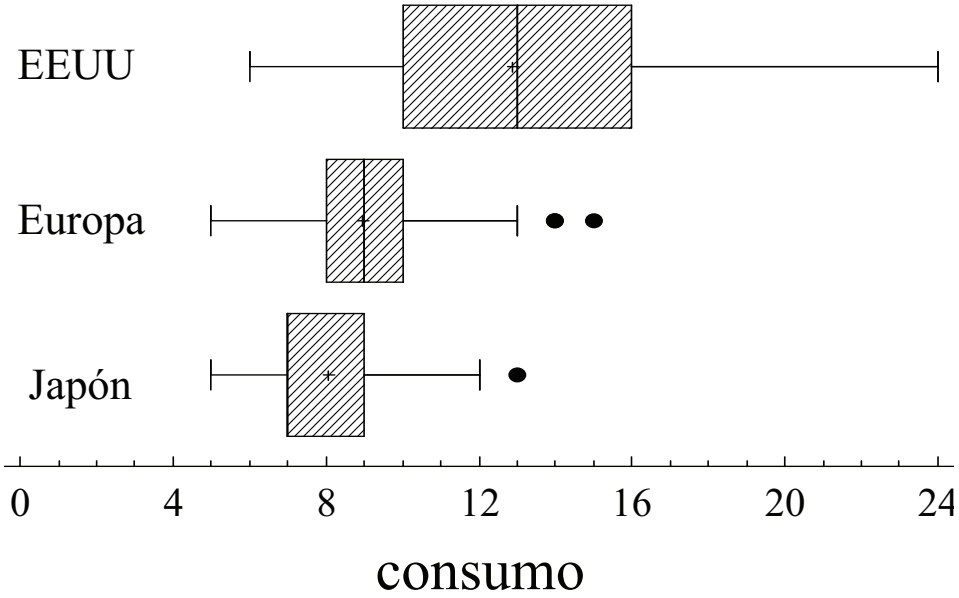
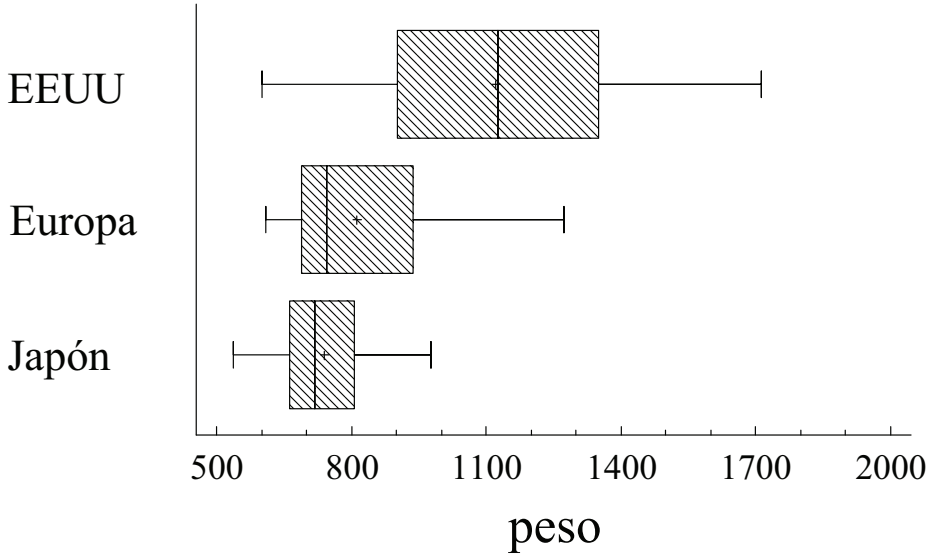


Diagrama de caja múltiple



Consumo según año de fabricación

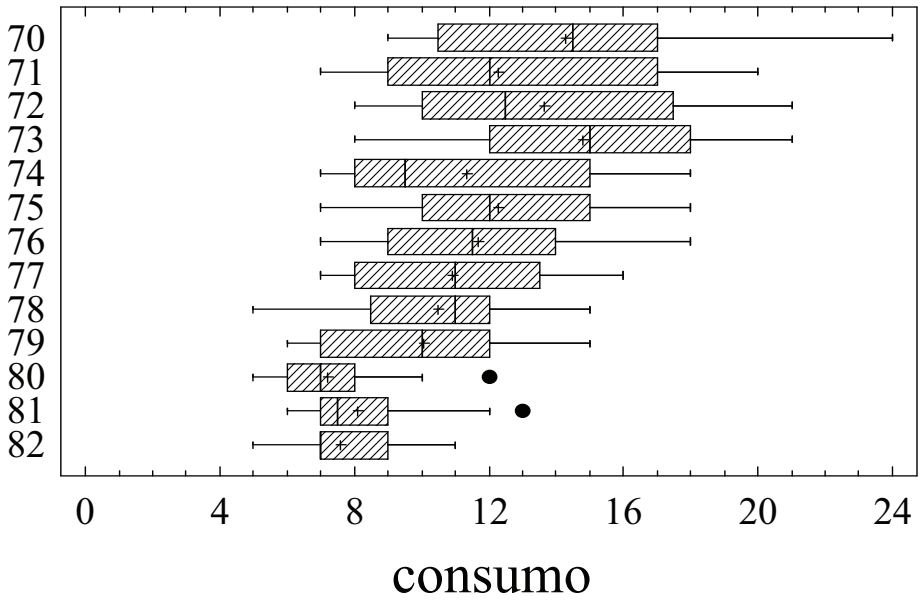


Diagrama de Caja Múltiple

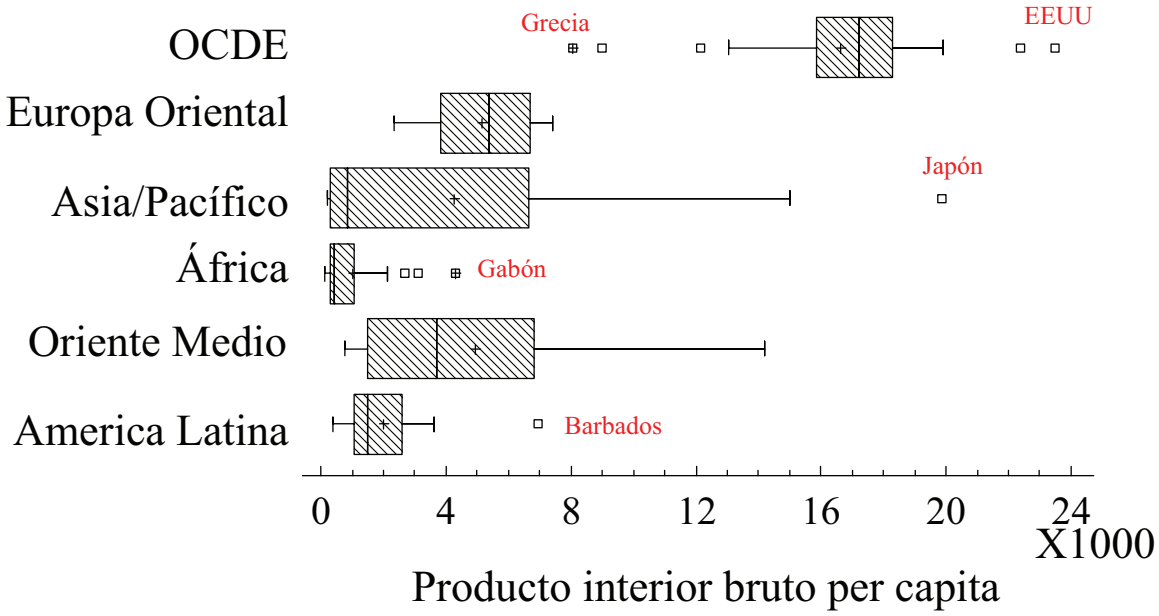


Diagrama de tallos y hojas

```
        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

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

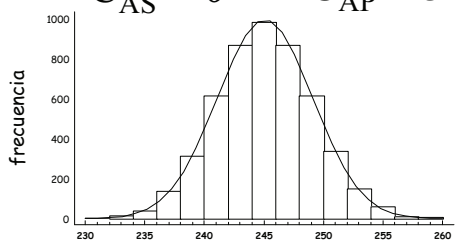
$$a_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$

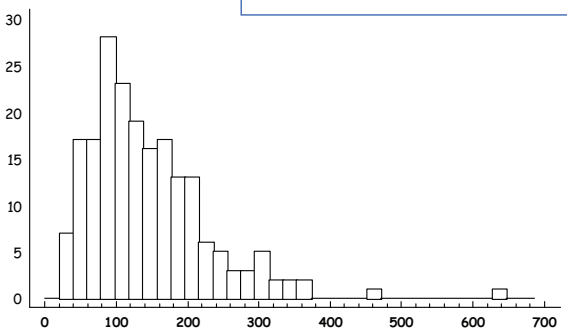
Momentos respecto a la media

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

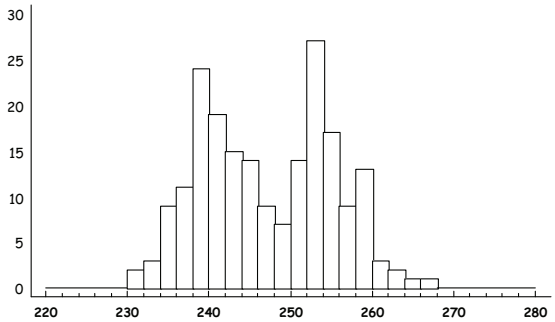
Modelo ideal

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





$C_{AS} > 0$



$C_{AP} < 3$

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)

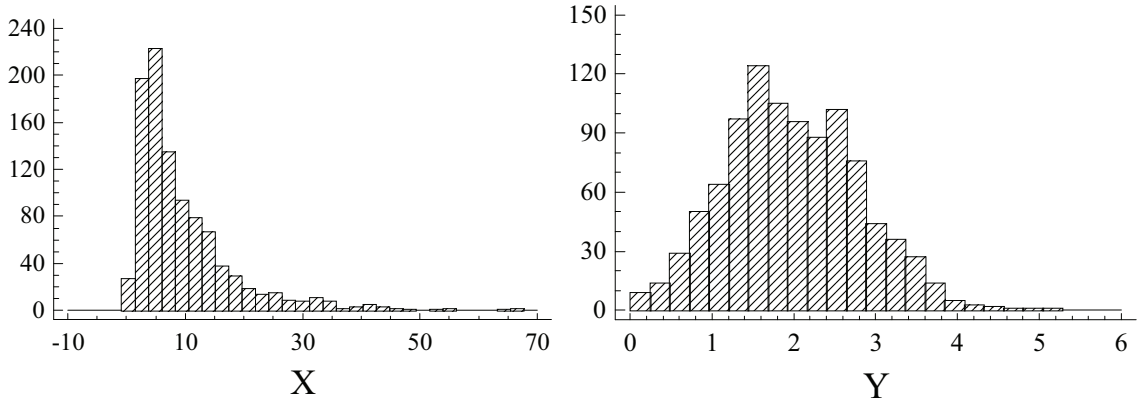
- Transformaciones no-lineales

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

Cambia la " *forma*" de la distribución

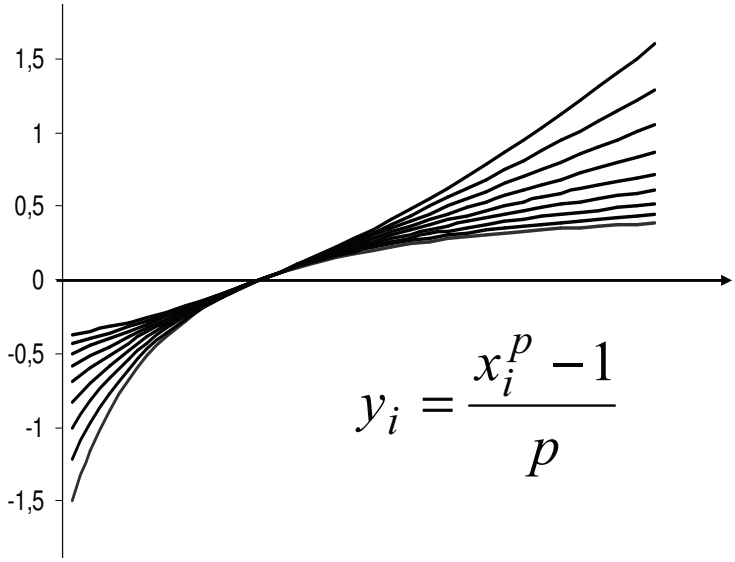
(coeficientes de asimetría y curtosis cambian)

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

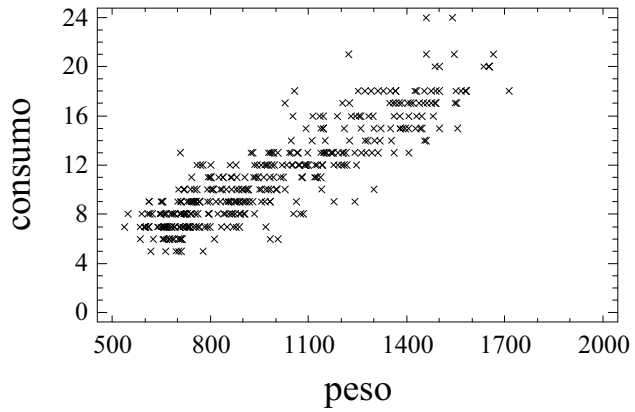
Variables

Observaciones	Y_1	Y_2	$\cdots$	Y_k		
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}	$\longrightarrow$	$\mathbf{x_n}$

Vector de Medias

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

Covarianza



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

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

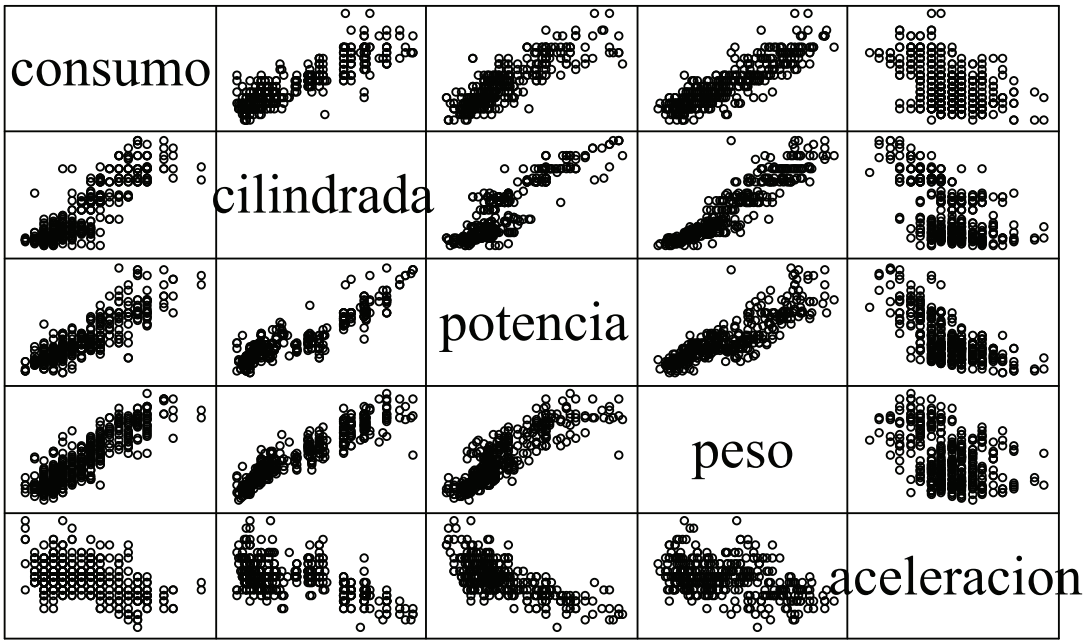
Matriz de Varianzas

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

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

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

Gráficos de dispersión: ejemplo coches



Matriz de varianzas: ejemplo coches

$$
\begin{array}{ccccc}
\text{consumo} & \text{c.c.} & \text{pot.} & \text{peso} & \text{acel.}
\end{array}
$$

$$
\mathbf{S} = \begin{pmatrix}
15{,}2 & 5.824{,}4 & 127{,}3 & 971{,}5 & -5{,}0 \\
5.824{,}4 & 2{,}94E6 & 58.965{,}4 & 451.461{,}0 & -2.597{,}4 \\
127{,}3 & 58.965{,}4 & 1.465{,}2 & 9.312{,}8 & -73{,}5 \\
971{,}5 & 451.461{,}0 & 9.312{,}8 & 7.949{,}5 & -328{,}0 \\
-5{,}0 & -2.597{,}4 & -73{,}5 & -328{,}0 & 7{,}6
\end{pmatrix}
$$

Propiedades de S^2

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

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

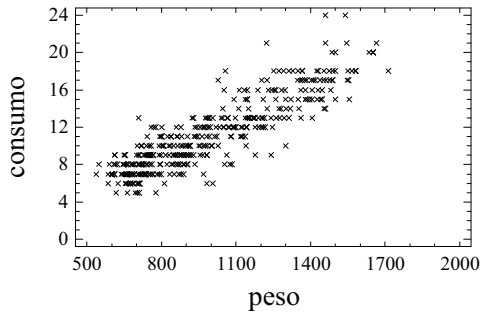
$\mathbf{S^2}$ es *semidefinida positiva*:

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

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

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

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

Matriz de correlaciones ejemplo coches

$$\begin{array}{ccccc} \text{consumo} & \text{c.c.} & \text{pot.} & \text{peso} & \text{Acel.} \end{array}$$

$$\mathbf{R} = \begin{pmatrix} 1 & 0{,}873 & 0{,}854 & 0{,}885 & -0{,}466 \\ 0{,}873 & 1 & 0{,}898 & \underline{0{,}934} & -0{,}549 \\ 0{,}854 & 0{,}898 & 1 & 0{,}863 & -0{,}696 \\ 0{,}885 & 0{,}934 & 0{,}863 & 1 & -0{,}422 \\ -0{,}466 & -0{,}549 & -0{,}696 & \underline{-0{,}422} & 1 \end{pmatrix}$$

Las variables están muy correlacionadas

Transformaciones Lineales

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

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

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

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

Transformaciones lineales II

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

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

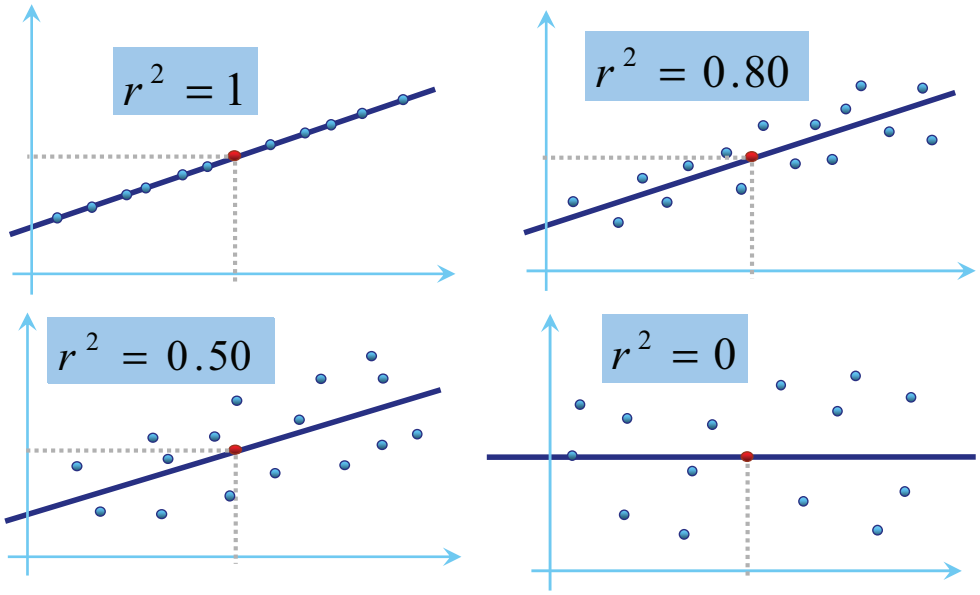
$$\bar{\mathbf{y}} = \frac{\sum_{i=1}^{n}\mathbf{y_i}}{n} = \frac{\sum_{i=1}^{n}\mathbf{A}\mathbf{x_i}}{n} = \frac{\mathbf{A}(\sum_{i=1}^{n}\mathbf{x_i})}{n} = \mathbf{A}\bar{\mathbf{x}}$$

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

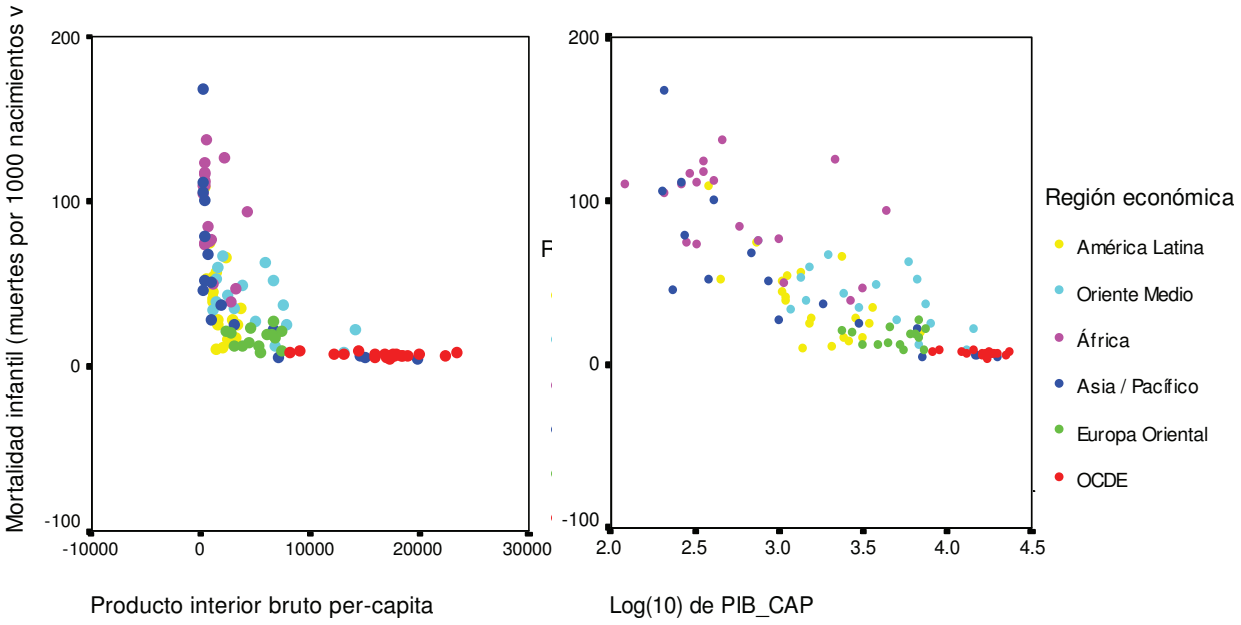
$$= \frac{\sum_{i=1}^{n}(\mathbf{A}\mathbf{x_i} - \mathbf{A}\bar{\mathbf{x}})(\mathbf{x_i}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}} - \bar{\mathbf{x}}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}})}{n} = \mathbf{A}\underbrace{\left(\frac{\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T}{n}\right)}_{\mathbf{S_X}^2}\mathbf{A}^{\mathbf{T}}$$

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

Coef. correlación



Falta de linealidad



Ejercicios propuestos

Capítulo 1. Descriptiva

1.1 Calcule la media, mediana y desviación típica de los datos siguientes: 28, 22, 35, 42, 44, 53, 58, 41, 40, 32, 31, 38, 37, 61, 25, 35.

(a) Directamente. *(b)*Agrupando en 5 clases de longitud 10 cm (20-30; 30-40; etc.) y utilizando las fórmulas para datos agrupados. *(c)* Construya un diagrama de tallo y hojas y un histograma de estos datos.

1.2 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?

1.3 Si $y = \dfrac{1}{x}$, obtener expresiones aproximadas para la media y la varianza de y en función de las de x.

1.4 Si $y = \log x$, obtener expresiones aproximadas para la media y la varianza de y en función de las de x.

1.5 Sea la variable x y su distribución de frecuencias relativas

x	fr
20-40	0,2
40-60	0,4
60-80	0,25
80-100	0,10
100-120	0,05

Construir el histograma y el histograma de la variable $y = \log x$.

1.6 En un taller mecánico una pieza pasa primero por una máquina 1 y luego por la máquina 2. La tabla muestra los tiempos X e Y (en minutos) empleados por diez piezas en cada máquina.

X	5	15	10	10	10	5	10	10	15	10
Y	10	15	15	10	10	15	20	15	20	20

Se pide:

1. (a) Obtener la distribución de frecuencias conjunta de los tiempos de ambas máquinas.
   (b) Obtener la distribución marginal de la máquina 1, dando su media y su varianza.
   (c) Calcular la covarianza entre ambas variables, interpretando el resultado.
   (d) Si el coste de cada pieza en pesetas es $0.8X + 0.6Y$, obtener la distribución de frecuencias del coste, calculando su media y su varianza.

1.7 ¿Es posible que la varianza de una variable x sea 4, la de y sea 9 y la de $z = x + y$ sea igual a 2? Justificar la respuesta.

1.8 Demostrar que al multiplicar x por k_1 e y por k_2, el coeficiente de correlación entre ambas no varía (k_1 y k_2 deben tener el mismo signo).

1.9 Demostrar que si entre dos variables existe una relación exacta $y = a + bx$, con $b > 0$, el coeficiente de correlación es uno.

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

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

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

¿Podemos afirmar que hay un error en los cálculos? ¿Por qué?

1.12 La matriz de varianzas de tres variables expresadas en centímetros es:

$$\begin{pmatrix} 20 & 8 & 7 \\ 8 & 10 & 4 \\ 7 & 4 & 15 \end{pmatrix}$$

Calcular la varianza generalizada de estas variables cuando se miden en milímetros.

1.13 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?

1.14 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

2. Probabilidad y variable aleatoria

Curso 2011-2012

Estadística

2. 1 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

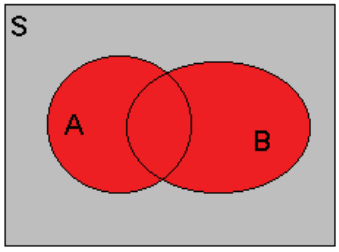
$$A \cup B = \{x : (x \in A) \textbf{ o } (x \in B)\}$$

- Intersección

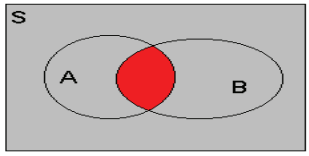
$$A \cap B = \{x : (x \in A) \textbf{ y } (x \in B)\}$$

- Complementario

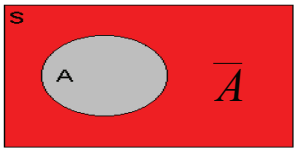
$$\overline{A} = \{x : x \notin A\}$$



$A \cup B$



$A \cap B$



Propiedades

Dados tres sucesos A, B y C de un espacio muestral S

$$\text{Conmutativa}: \begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap A \end{cases}$$

$$\text{Asociativa}: \begin{cases} A \cup (B \cup C) = (A \cup B) \cup C \\ A \cap (B \cap C) = (A \cap B) \cap C \end{cases}$$

$$\text{Distributiva}: \begin{cases} A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \end{cases}$$

$$\text{Leyes de De Morgan}: \begin{cases} \overline{\mathrm{A} \cup \mathrm{B}} = \overline{A} \cap \overline{B} \\ \overline{\mathrm{A} \cap \mathrm{B}} = \overline{A} \cup \overline{B} \end{cases}$$

Axiomas de Probabilidad

Dado un espacio muestral S, una función de probabilidad asigna valores P(A) a cada suceso $\mathrm{A} \subset \mathrm{S}$ 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 $A_1, A_2, \ldots, A_n$, 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.

¿ Es una asignación correcta?

Propiedades elementales

1. $P(\varnothing)=0.$
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

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

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}

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

- 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

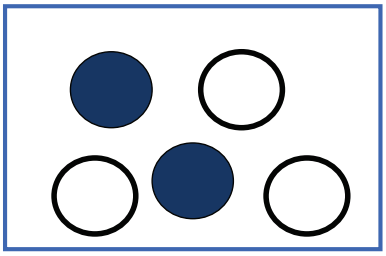
1er Dado

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

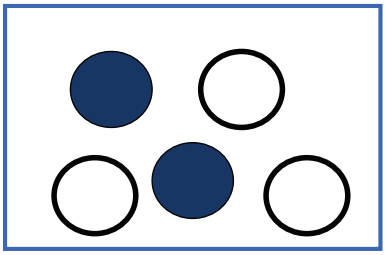
2ª Bola

	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

2ª Bola

	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

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

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

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

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

1 - 6 - 21 - 29 - 33 -43

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.

Primitiva

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

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

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

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

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

- 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?

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

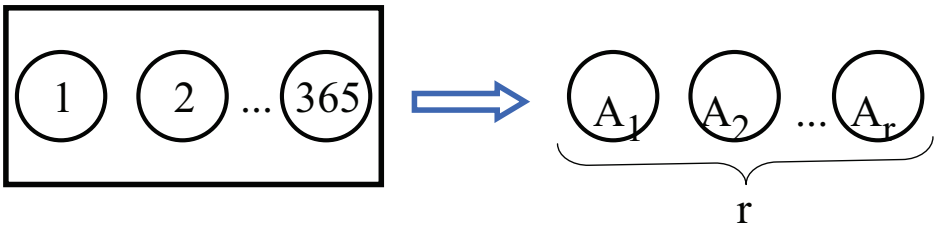
- De un lote con 100 piezas se toman al azar 10, si todas las piezas elegidas son buenas se acepta el lote y se rechaza en caso contrario. ¿Cuál es la probabilidad de aceptar un lote con 10 piezas defectuosas?

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

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

Cumpleaños

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



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

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

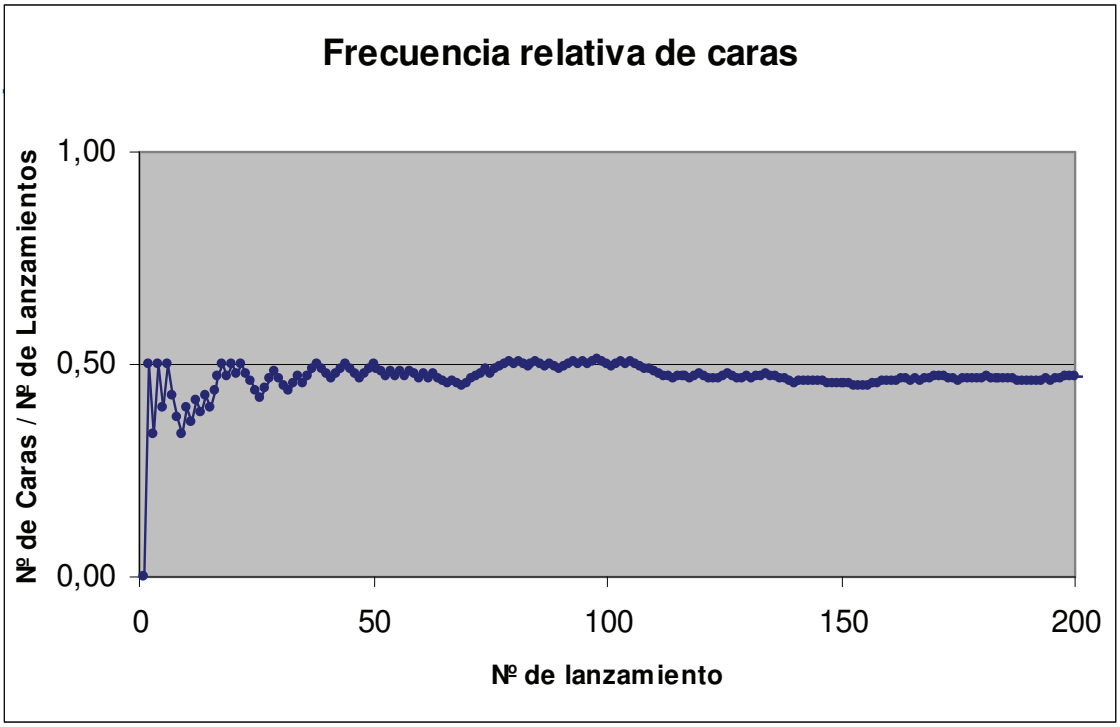
$$P(\overline{A}) = 1 - P(A), \qquad r = 25 \rightarrow P(\overline{A}) = 0.578$$

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 el experimento *n* veces en idénticas condiciones.



La Primitiva — APARICIONES DE LOS NUMEROS EN LA COMBINACIÓN GANADORA

Nº	TOTAL	2000	2001	Nº	TOTAL	2000	2001
1	171	9	6	2	154	11	11
3	173	13	7	4	164	17	8
5	169	12	4	6	191	10	12
7	166	11	6	8	159	9	6
9	173	20	10	10	166	19	11
11	164	17	6	12	166	8	12
13	165	9	12	14	182	14	8
15	179	15	16	16	167	7	1
17	170	11	15	18	161	14	14
19	168	14	10	20	152	10	9
21	169	13	6	22	176	16	7
23	188	11	10	24	149	13	7

Total sorteos **1.380** (2000 - 104; 2001 - 74)

La Primitiva — APARICIONES DE LOS NUMEROS EN LA COMBINACIÓN GANADORA

Nº	TOTAL	2000	2001	Nº	TOTAL	2000	2001
25	177	14	7	26	165	10	3
27	169	14	16	28	156	14	10
29	169	10	10	30	173	13	12
31	161	12	8	32	160	15	7
33	162	6	11	34	173	16	16
35	174	12	7	36	178	18	7
37	165	13	8	38	193	12	10
39	190	16	9	40	168	8	9
41	175	11	8	42	163	12	10
43	158	9	10	44	163	15	11
45	182	13	9	46	157	15	7
47	190	19	6	48	166	15	8
49	151	9	11				

NO INCLUYE LA APARICION DEL NUMERO COMPLEMENTARIO

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**

$$P(F) = 0{,}30 \quad \begin{cases} P(F \mid H) = \dfrac{0{,}18}{0{,}45} = 0{,}40 \\ \\ P(F \mid M) = \dfrac{0{,}12}{0{,}55} = 0{,}22 \end{cases}$$

Probabilidad Condicionada

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

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

Utilidad

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

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

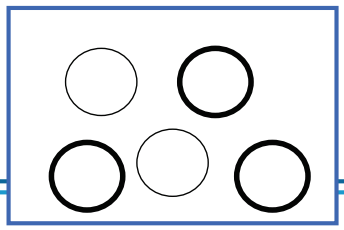
- Cálculo de la intersección de sucesos

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

- 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**:

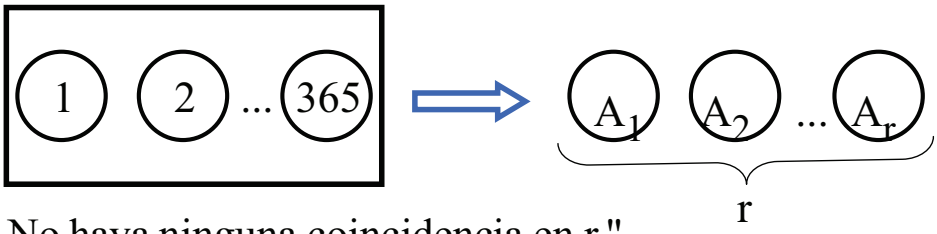
$$P(B1 \cap N2) = P(B1)\ P(N2|\ B1)$$
$$= (3/5)(2/4) = 3/10$$

- **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.



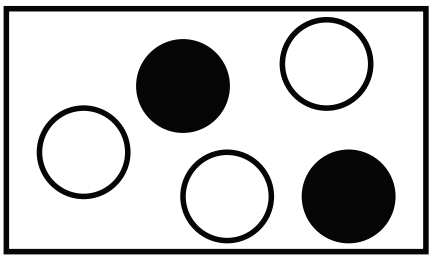
$B_r =$ " No haya ninguna coincidencia en r "

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

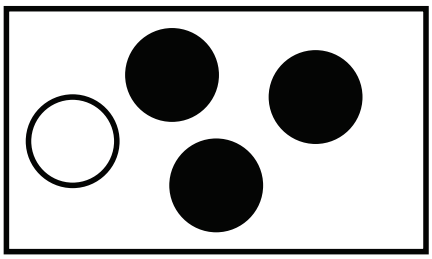
$$= 1 \times \frac{365-1}{365} \times \frac{365-2}{365} \times \cdots \times \frac{365-r+1}{365}$$

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

Ejemplo



Urna **U1**



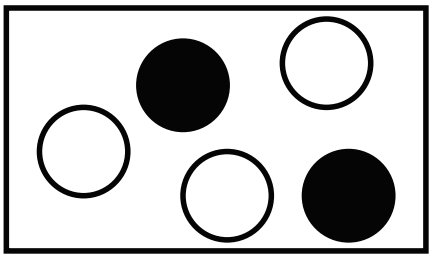
Urna **U2**

Se elige una urna al azar y se extrae una bola: ¿ <u>*P*(Blanca)</u> ?

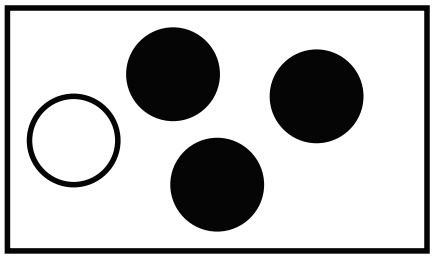
$$P(B) = P(B \mid U1)P(U1) + P(B \mid U2)P(U2)$$

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

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: ¿ *P*(Blanca) ?

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

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

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

A **y** *B* **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)×(1/2) = 1/4

P (CX) = P(C) P(X) = (1/2)×(1/2) = 1/4

P (XC) = P(X) P(C) = (1/2)×(1/2) = 1/4

P (XX) = P(X) P(X) = (1/2)×(1/2) = 1/4

Independencia (3 o más sucesos)

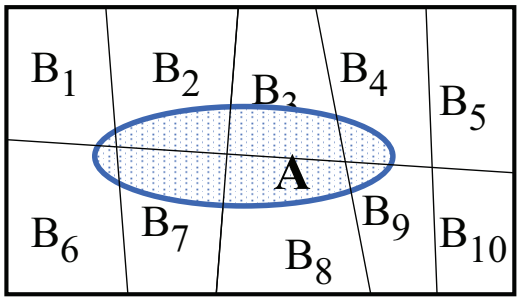
- Tres sucesos *A, B* y *C* son independientes si

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

- Los sucesos $A_1, A_2, \ldots, A_n$ son independientes si cualquier subconjunto $A_{i1}, A_{i2}, \ldots, A_{ik}$, cumple

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}) = P(A_{i1})P(A_{i2})\ldots P(A_{ik})$$

Probabilidad Total



Partición.

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

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

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

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

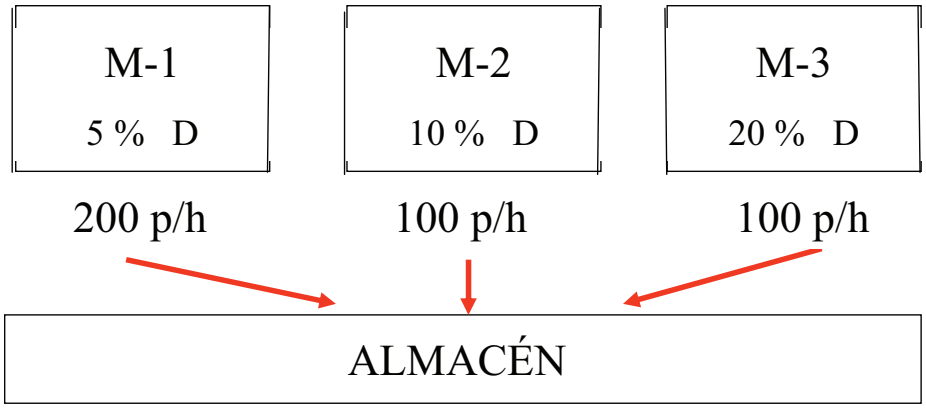
$$\boxed{P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots P(A \mid B_n)P(B_n)}$$

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)



El porcentaje de piezas defectuosas fabricadas por tres máquinas es 5%, 10% y 20%. 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?

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

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

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

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

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?

$V =$ "Tener el Virus" $\qquad$ S = "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

2. 2 Variable aleatoria

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."

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

$X(s) \equiv$ 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, ...}

Distribución de probabilidad

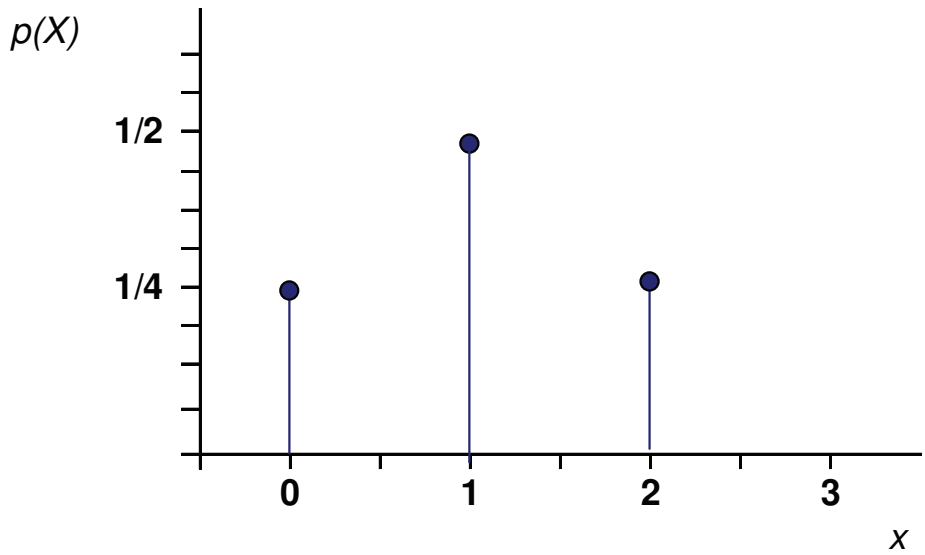
Sea { x_1 , x_2 , ..., x_n } los valores que puede tomar la variable aleatoria X. Se denomina **distribución de probabilidad** de la variable aleatoria a $P(X=x_i)$ que cumple:

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

Nº de Caras al lanzar 2 monedas

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

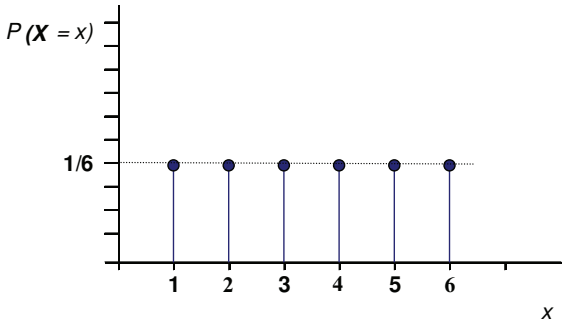
Distribución de probabilidad



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



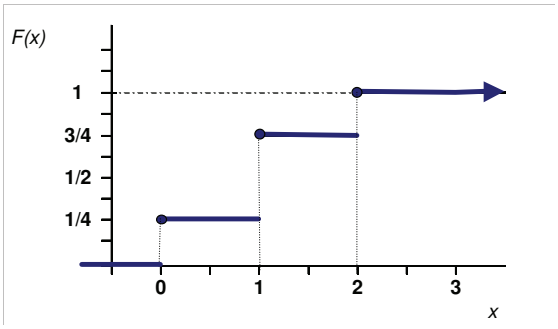
Función de distribución

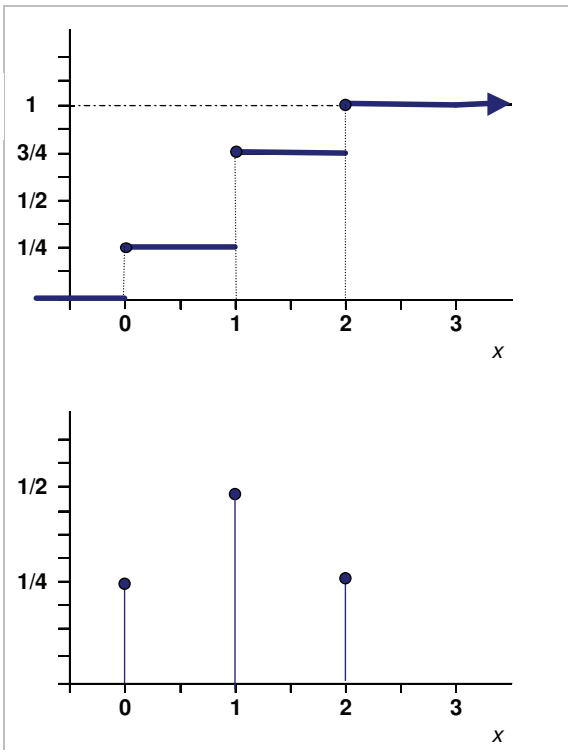
La función de distribución $F_X(x)$ de una variable aleatoria X se define para todo número real x como:

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

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

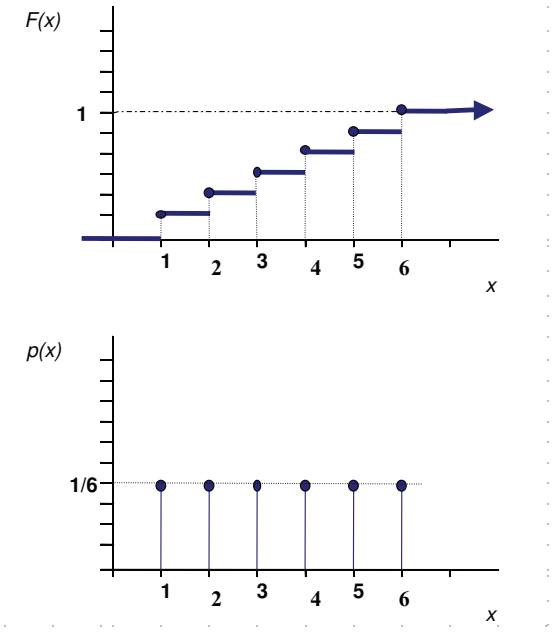




Función de Distribución

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

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 $\lim_{x \to +\infty} F(x) = 1.$

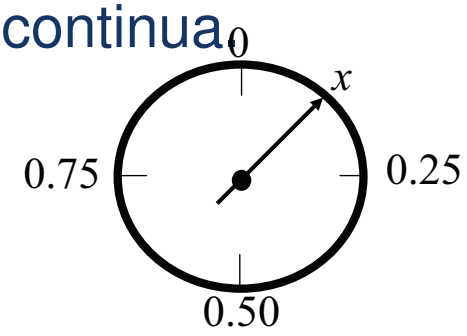
b. $F(x)$ es una función no decreciente.

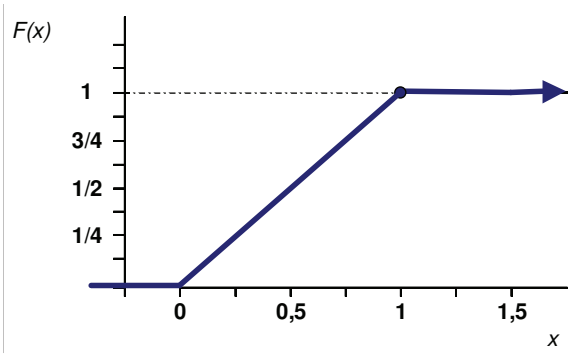
c. $F(x)$ es continua por la derecha:

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

Variable aleatoria continua

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





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

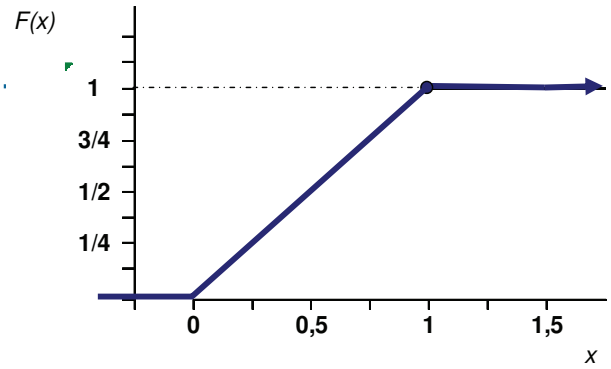
Función de densidad

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

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

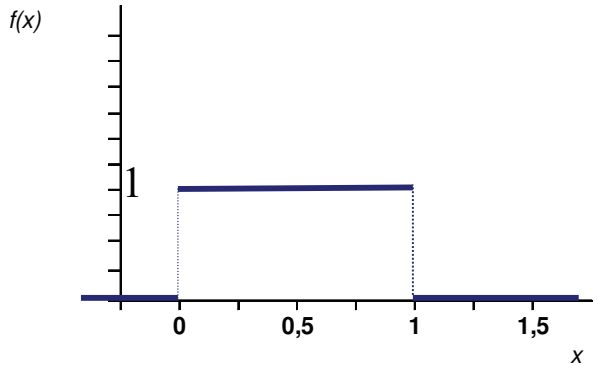
Si $F_X(x)$ es derivable, además

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



Función de distribución

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



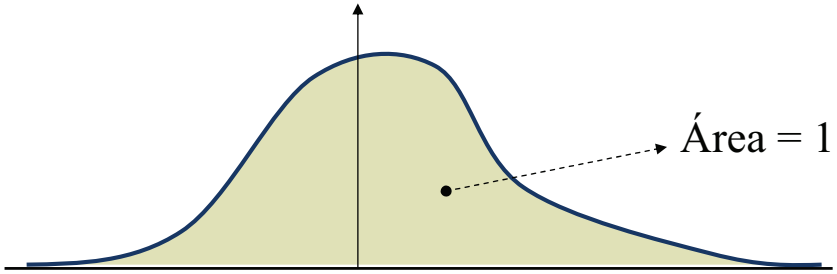
Función de densidad

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

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



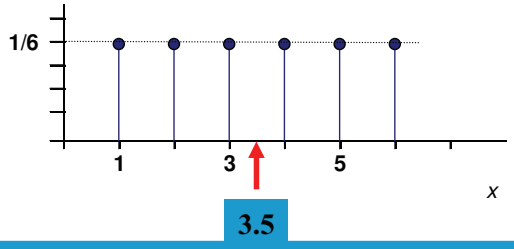
Esperanza

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

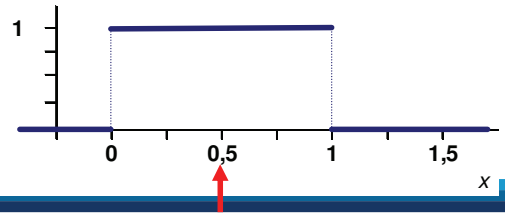
Esperanza

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)

$$E[aX + b] = aE[X] + b$$

Varianza

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

$$\mathrm{Var}(X) = E[(X - \mu)^2].$$

- Variable aleatoria discreta

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

- 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

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

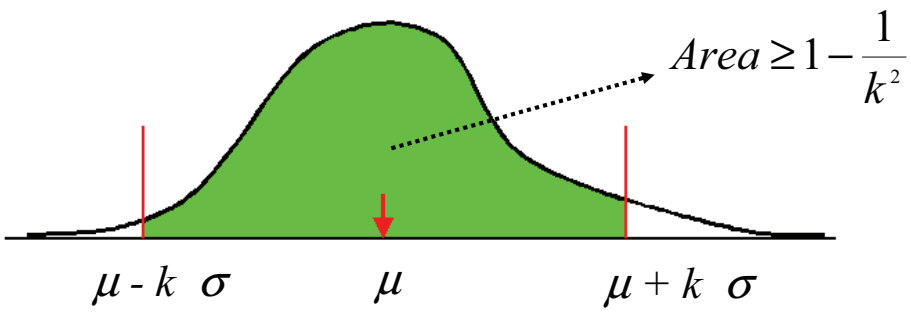
$$= \frac{35}{12}.$$

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 V.A.

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 no lineales

z=h(y)

Desarrollo de Taylor para $z = h(y)$ en $\mu = \mathrm{E[y]}$

$$z \approx h(\mu) + h'(\mu)(y-\mu) + \frac{1}{2}h''(\mu)(y-\mu)^2$$

La media y varianzas de z son aprox.

$$E[z] \approx h(\mu) + \frac{1}{2}h''(\mu)\mathrm{Var}(y)$$

$$\mathrm{Var}[z] \approx [h'(\mu)]^2 \mathrm{Var}[y]$$

Transformaciones

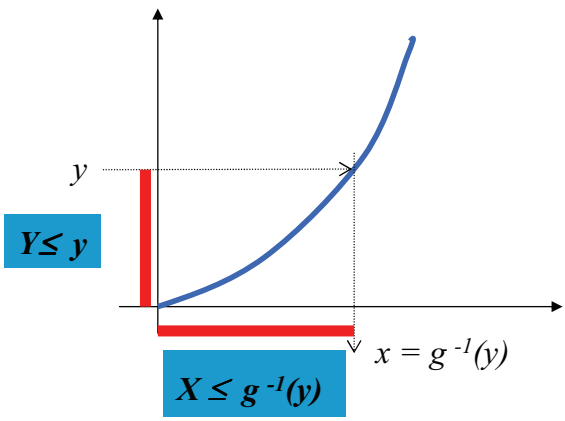
Dada una variable aleatoria X con función de densidad $f_X(x)$ vamos a ver como se obtiene la función de densidad $f_Y(y)$ de la variable aleatoria Y, definida como

$$Y = g(X)$$

Casos a considerar:

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

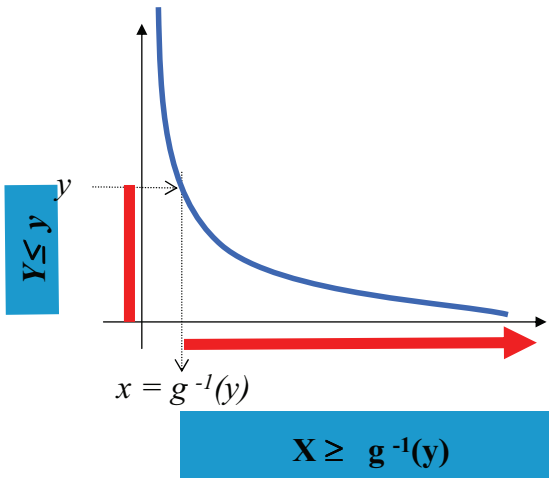
Función *g* creciente



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

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

Función *g* decreciente



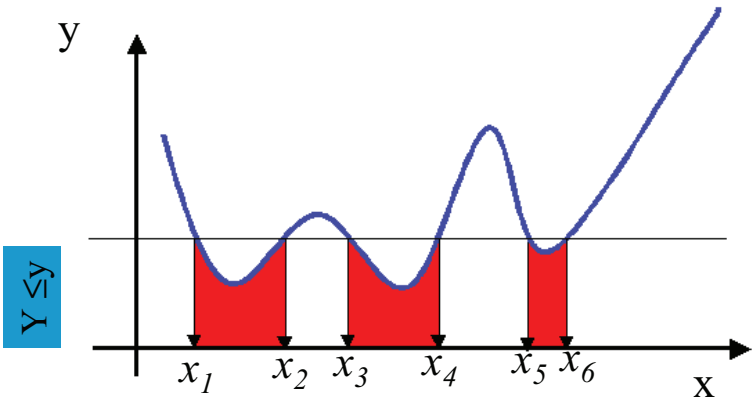
$$F_Y(y) = P(Y \le y)$$
$$= P(g(X) \le y)$$
$$= P(\, X \ge g^{-1}(y)\,)$$
$$= 1 - F_X(g^{-1}(y))$$

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

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

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

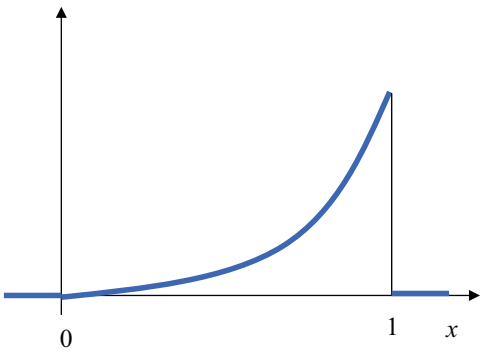
¿Cuál es la función de densidad del volumen?

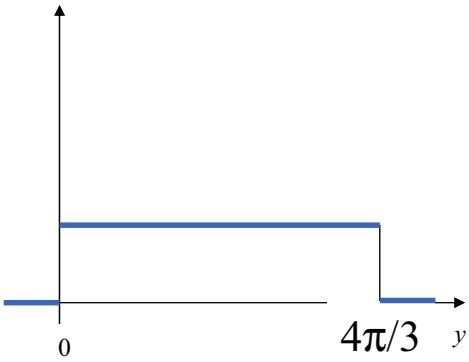
$$Y = \text{Volumen de la esfera}$$

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

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

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| \times f_X(g^{-1}(y))$$
$$= \frac{3}{4\pi}, \quad 0 \le y \le \frac{4\pi}{3}.$$





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

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

$$y = \frac{4}{3}\pi x^3$$

Un caso muy relevante del problema inverso

Dadas:

1) Una variable aleatoria *x* con distribución uniforme *(0,1)*

2) Una variable aleatoria *y* con función de densidad $f_Y(y)$

Encontrar la transformación *y=g(x)* que convierte *x* en *y*

Solución del problema

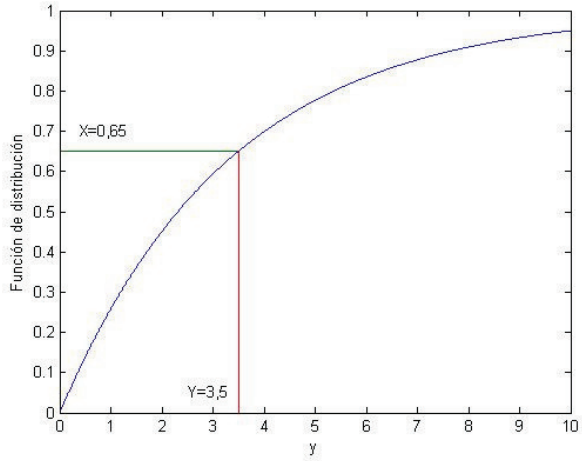
$$\frac{d\ g^{-1}(y)}{dy} = f_Y(y)$$

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

Representación gráfica con ejemplo

$$f_Y(y) = \lambda e^{-\lambda y}, \lambda = 0{,}3$$

$$X = 0{,}65, y = -(1/\lambda)\log(1-x) = 3{,}5$$



Aplicación: Simulación de Monte Carlo

Simulación con ordenador de procesos aleatorios; tiene dos vertientes:

1) Pedagógica: como herramienta para entender mejor los modelos de probabilidad

2) Computacional: herramienta muy potente para resolver problemas no abordables por métodos convencionales

Ejemplo de problema resuelto por simulación de Monte Carlo

Problema de los cumpleaños:

1) Se generan al azar con distribución uniforme discreta en 1,2,..,365, r valores (fechas de nacimiento de las r personas)

2) Se cuenta el número de coincidencias

3) Repitiendo N veces los pasos 1 y 2, se obtiene una aproximación de la distribución del número de coincidencias

Ejercicios propuestos

Capítulo 2. Probabilidad y variable aleatoria

2.1 Si la probabilidad de que un interruptor cualquiera de la figura se encuentre cerrado (es decir, pase corriente) es p, calcular la probabilidad de que pase corriente de A a B para las dos configuraciones siguientes:

2.2 Teniendo en cuenta que para cualquier par de sucesos, A_1, A_2 se cumple que

$$\Pr(A_1 A_2) \geq \Pr(A_1) + \Pr(A_2) - 1,$$

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

2.3 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 averís 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.

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

2.5 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.

2.6 Dos pueblos de alta montaña están comunicados por cuatro tramos de carretera según se muestra en la figura (a). 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 (b)?

2.7 Un concursante debe elegir entre tres puertas, detrás de una de las cuales se encuentra un premio. Hecha la elección y antes de abrir la puerta, el presentador le muestra que en una de las dos puertas no escogidas está el premio y le da la posibilidad de reconsiderar su decisión. ¿Qué debe hacer el concursante?

2.8 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.

2.9 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?

2.10 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?

2.11 Cuatro fichas están marcadas con las letras A, B, C, ABC;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.

2.12 Tres personas comparten una oficina con un teléfono. De las llamadas que llegan, 2/5 son para A, 2/5 para B y 1/5 para C. El trabajo de estos hombres les obliga a frecuentes salidas, de manera que A está fuera el 50% de su tiempo, y B y C el 25%. Calcular la probabilidad de que :

(a) No esté ninguno para responder al teléfono.
(b) Esté la persona a la que se llama.
(c) Haya tres llamadas seguidas para una persona.
(d) Haya tres llamadas seguidas para tres personas diferentes.

2.13 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?

2.14 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?

2.15 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?

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

2.17 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.

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

2.19 La función de densidad de la variable aleatoria X es

$$f(x) = \begin{cases} 1/(kx), & \text{si } 25 \leq x \leq 50 \\ 0, & \text{en el resto.} \end{cases}$$

Obtener k, la media y la varianza de X.

2.20 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:

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

donde $\alpha = \sqrt{2kT/m}$, siendo k la constante de Boltzmann.
Además, $E(V) = 2\alpha/\sqrt{\pi}$ y $\mathrm{Var}(V) = (3/2 - 4/\pi)\alpha^2$.

(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.

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

2.22 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, ¿cuál es la probabilidad de que al lanzar 10 proyectiles, ninguno haya impactado a una distacia menor de 5 cm del centro del blanco?

2.23 Adaptar la demostración de la desigualdad de Chebychev y demostrar la desigualdad de Markov

$$P(X > a) \leq \frac{1}{a} E[X]$$

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

2.24 La variable aleatoria Z tiene como función de densidad

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

Obtener la función de densidad de

$$Y = |Z|$$

y su media.

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

$P(X = x) = kx$ con $x = 1, 2, ..., 20$

Calcular el valor de la constante y $E(X|X > 10)$.

2.26 Dada la variable aleatoria X, cuya función de densidad es

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

Obtener k, así como la media y la varianza de la variable $Y = 3X - 1$.

2.27 El diámetro D de las bolas de acero para un determinado tipo de rodamiento sigue una distribución de media μ y desviación típica σ. Obtener, de forma aproximada, la media y la varianza de la distribución de los volúmenes de estas bolas.

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

2.29 Sea X una variable aleatoria discreta que toma los valores 1 y 2 con probabilidades $0,4$ y $0,6$ respectivamente. Si se extrae con reposición una muestra aleatoria simple de tamaño 2, obtener la función de probabilidad del promedio de las dos extracciones y relacionarla con resultados teóricos conocidos.

3. Modelos Univariantes de Probabilidad

Curso 2011-2012

Estadística

Modelos Univariantes

Proceso de Bernoulli

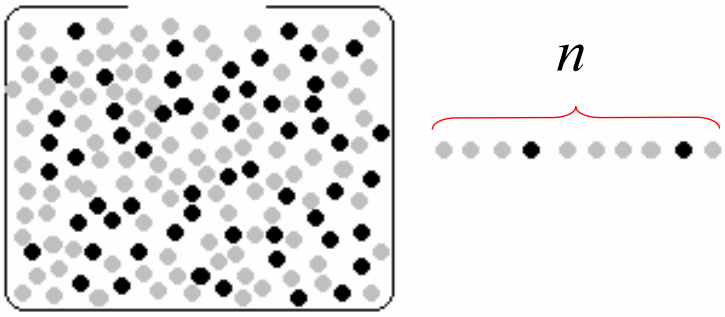
El resultado de un experimento admite dos categorías: "*Aceptable*" y "*Defectuoso*".

- 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*)

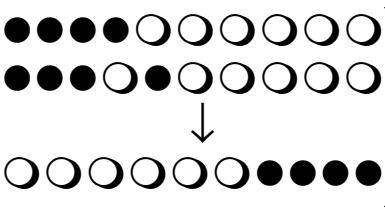


Proporción defectuosas = p

X = " Nº de defectuosas al extraer n piezas"

Distribución de probabilidad binomial *(n,p)* *n=10*

$$p^4(1-p)^6$$



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

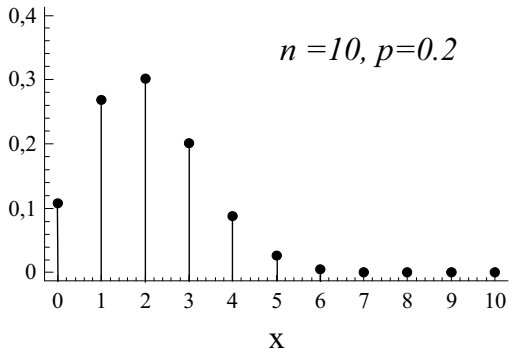
$$P(X=4)=\binom{10}{4}p^4(1-p)^6$$

Distribución de probabilidad binomial *(n,p)*

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

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

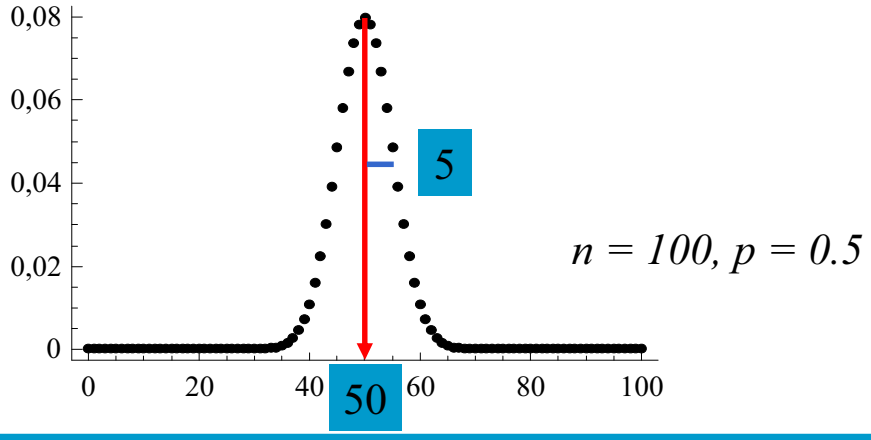
(2) $\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=1.$



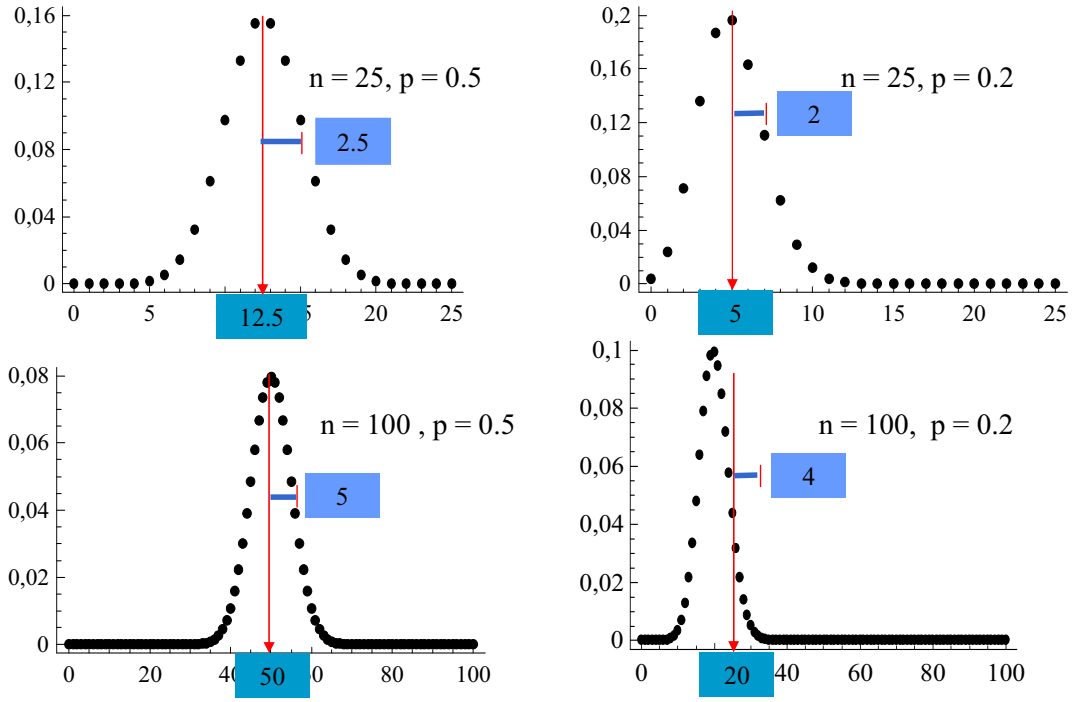
Propiedades de la dist. binomial

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

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



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,

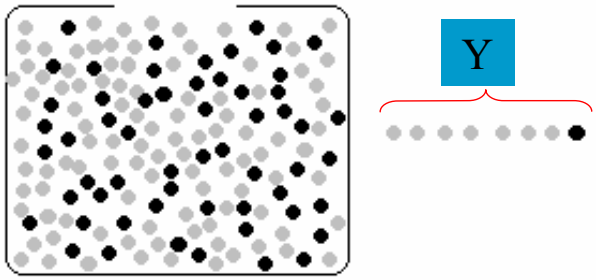
$$X \equiv \text{Número de defectuosas en la muestra}$$

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

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

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

Distribución Geométrica (*p*)



Proporción defectuosas = p

Y = "Piezas extraídas hasta que aparezca una defectuosa"

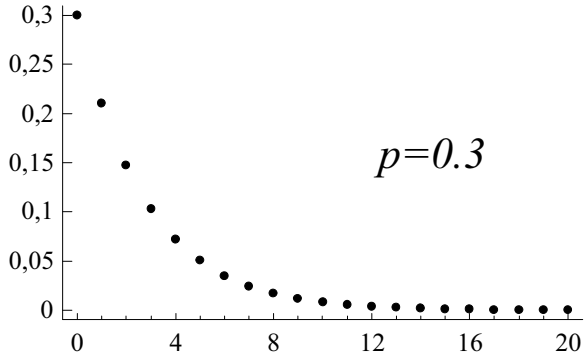
Distribución de probabilidad geométrica *(p)*

●	1	→	p
○●	2	→	$(1-p)p$
○○●	3	→	$(1-p)^2p$
↓			↓
○○○...●	k	→	$(1-p)^{k-1}p$

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

Propiedades de la v.a. geométrica

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



Distribución 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

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

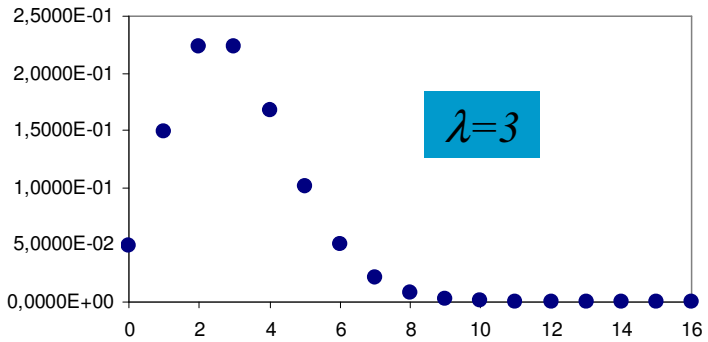
$$P_X(x) = \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

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

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

Distribución de Poisson

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



Media y Varianza

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

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

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

e^{λ}

$$E[X]=\lambda \qquad Var[X]=\lambda$$

Ejemplo

Una fuente radiactiva emite partículas según la distribución 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. $P(X=0)=e^{-10}=4.54E-15.$
3. $P(X>5)=1-P(X\le 5)$

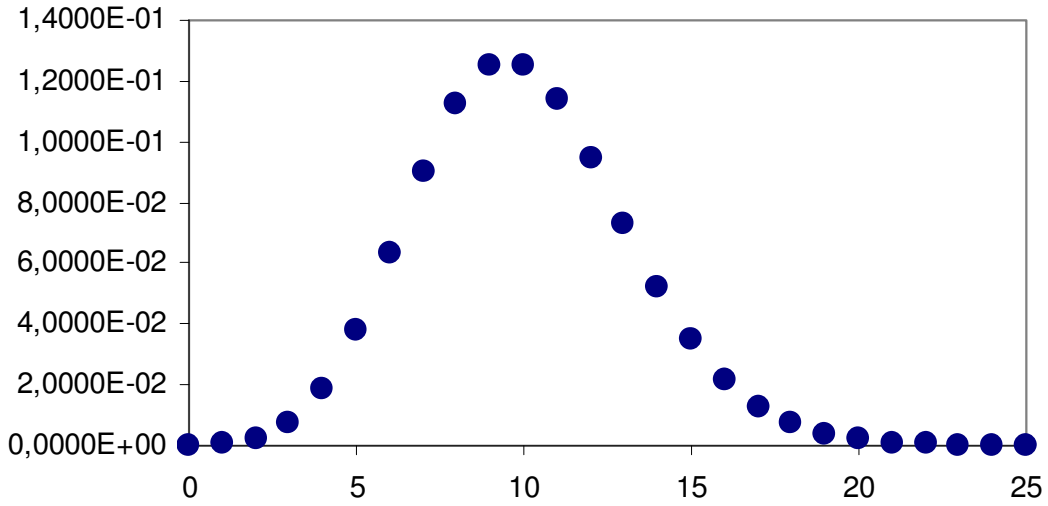
$$=1-e^{-10}\sum_{x=0}^{5}\frac{10^x}{x!}=0.933.$$

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

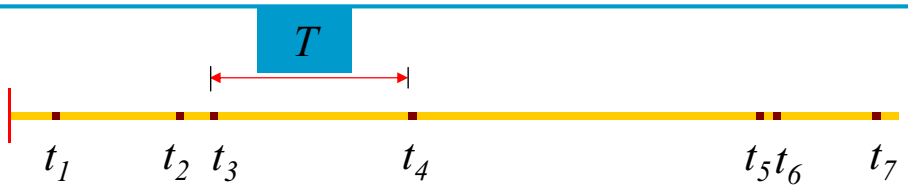
$\lambda'=5\times 10=50$

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

Poisson de media 10



Distribución Exponencial

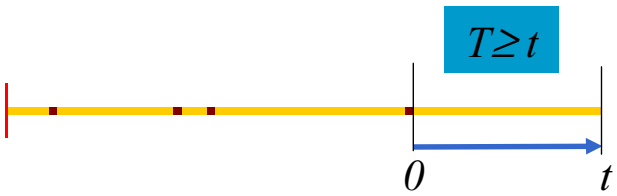


Ejemplo: Fabricación continua de conductor de cobre.

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

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

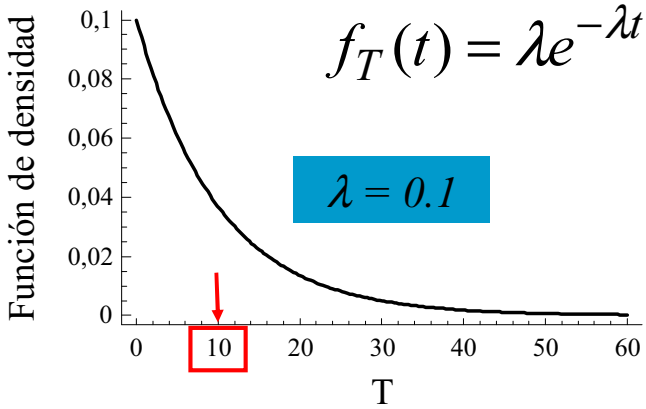
Distribución Exponencial



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

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

Propiedades (Exponencial)



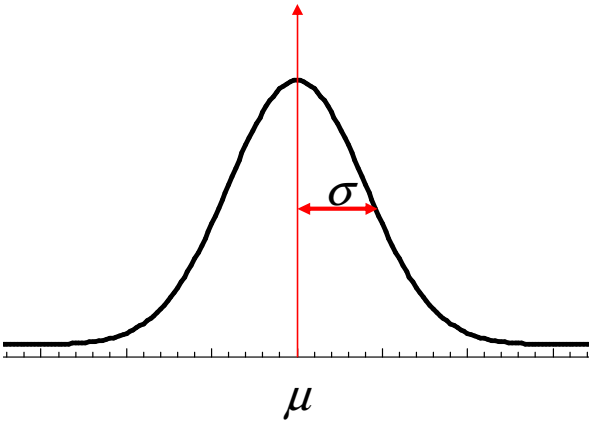
$$f_T(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

$\lambda = 0.1$

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

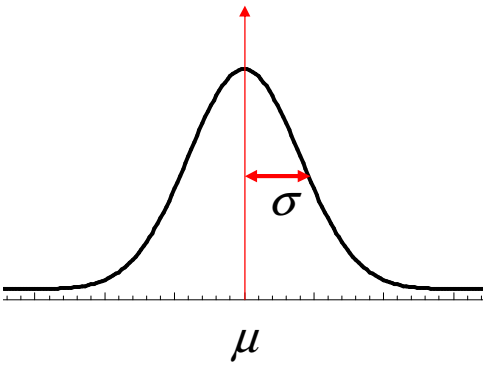
$$Var[T] = E[T^2] - E[T]^2$$
$$= \int_0^{\infty} \lambda t^2 e^{-\lambda t} dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

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



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

$E[X] = \mu$

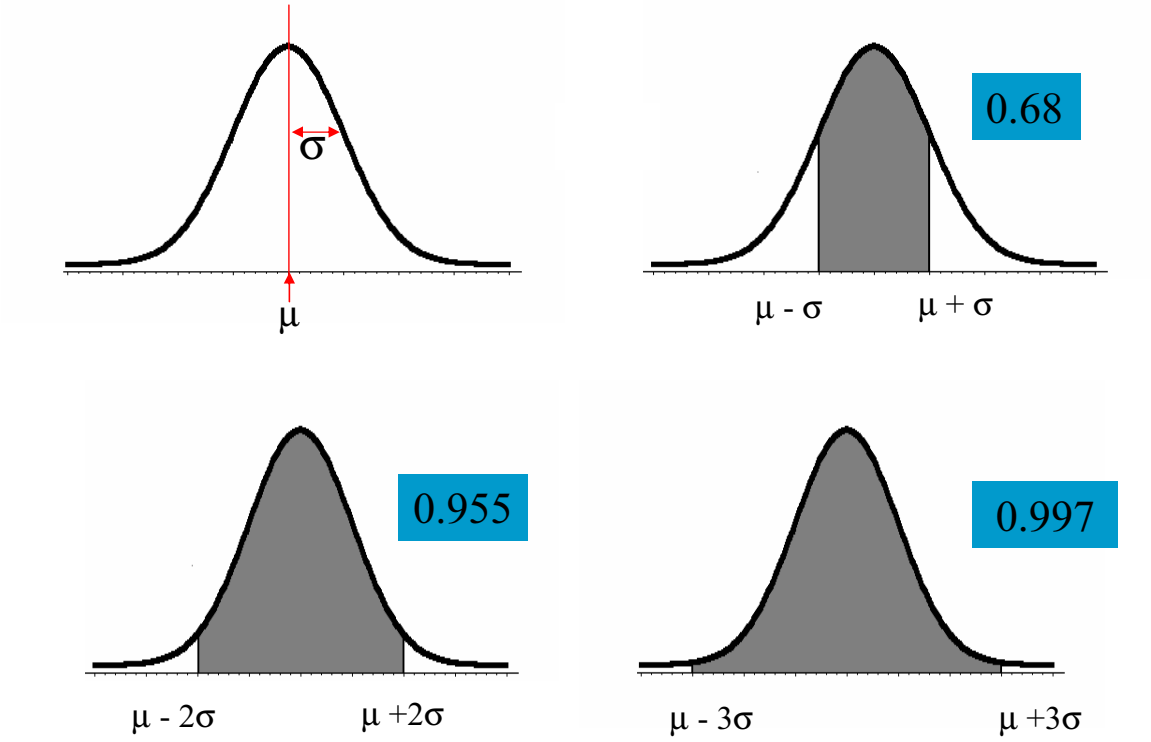
$Var[X] = \sigma^2$

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

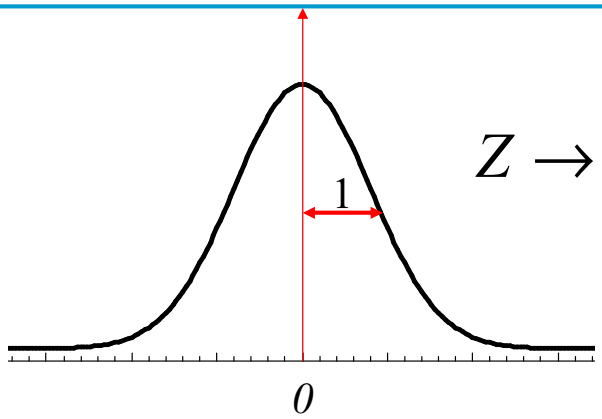
$E[(X-\mu)^4] = 3\sigma^4$

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

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



Normal Estándar



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

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

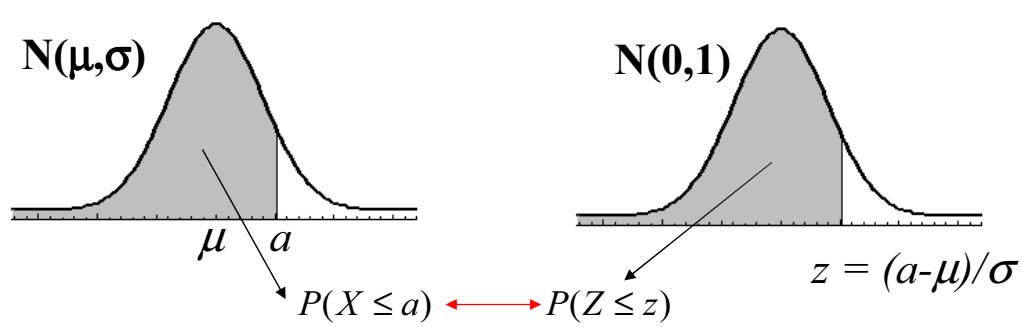
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

TABLAS

Estandarización

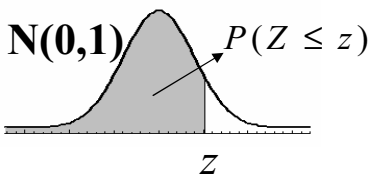
$$X \to N(\mu,\sigma) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \to N(0,1)$$

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



TABLA

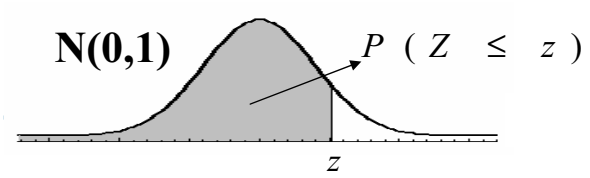
Normal Estándar



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



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)

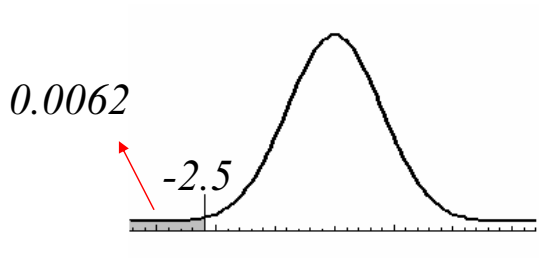
La longitud X de ciertos tornillos es una variable aleatoria con distribución normal de media 30 mm y desviación típica 0.2 mm. Se aceptan como válidos aquellos que cumplen 29.5 < X < 30.4.

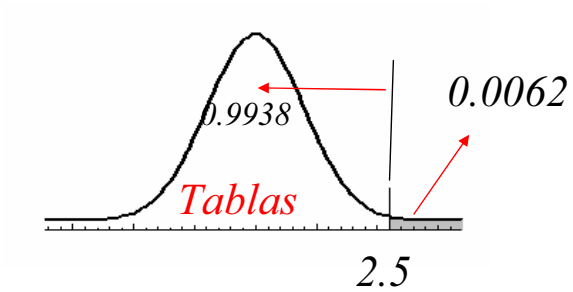
- Proporción de tornillos no aceptables por cortos.
- Proporción de tornillos no aceptables por largos.
- Proporción de tornillos válidos.

Ejemplo (Solución)

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

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





2. $$P(X \ge 30.4) = P\left(\frac{X-30}{0.2} \ge \frac{30.4-30}{0.2}\right)$$
$$= P(Z \ge 2.0) = 1 - \Phi(2) = 0.0228$$

3. $$P(29.5 < X < 30.4) = 1 - 0.0228 - 0.0062 = 0.971$$

Ejercicios propuestos

Capítulo 3. Modelos univariantes

3.1 Si las llamadas telefónicas a una centralita siguen una distribución de Poisson de parámetro $\lambda =$ 3llamadas/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.

3.2 La variable aleatoria X tiene distribución exponencial con media 1. Obtener la función de distribución y la función de densidad de

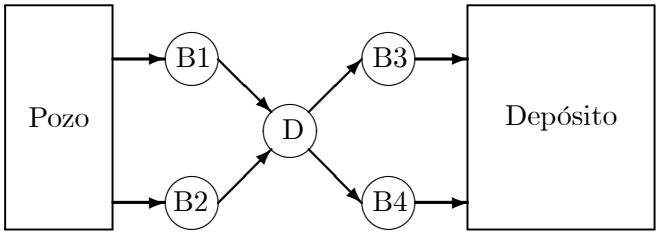
$$W = aX^{1/b}, \quad a > 0, b > 0$$

3.3 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.

3.4 A un puesto de servicio llegan de manera inedependiente, 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.

3.5 En una planta industrial 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.



(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?

3.6 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?

3.7 De un lote con una proporción de piezas defectuosas p, se extraen piexas 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.*

3.8 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?

3.9 Una pareja decide tener hijos hasta el nacimiento de la primera niña. Calcular la probabilidad de que tengan más de 4 hijos. (Supóngase $P(niño) = P(niña) = 0.5$)

3.10 La distancia D entre dos vehículos consecutivos es 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?

3.11 La función de densidad del tiempo T de funcionamiento de un componente hasta que falla es

$$f(t) = k\beta t^{\beta-1} \exp(-kt^{\beta}), \quad t > 0, k > 0, \beta > 0.$$

Cuando un componente falla se puede reparar y queda igual que otro que no hubiera fallado nunca y tuviera la misma edad. Además, el tiempo necesario para reparar el componente se considera despreciable. Si un componente tiene su primer fallo en el instante t_1, calcular la probabilidad de que *el segundo fallo* se produzca después de t_2 con $t_2 > t_1$.

3.12 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?

3.13 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 interpolar el resultado.

3.14 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).

3.15 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.)

3.16 Dada una variable aleatoria X, cuya distribución es $N(0, \sigma^2)$, calcular la mediana de la variable $Y = |X|$.

3.17 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.

3.18 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).

3.19 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.

3.20 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?

3.21 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?

3.22 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?

3.23 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)

3.25 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.

3.26 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).

3.26 "*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 € y para los de Madrid 50€. 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.

4. Modelos Multivariantes

Curso 2011-2012

Estadística

Distribución conjunta de variables aleatorias

Definiciones (v. a. discretas)

- Distribución de probabilidad <u>conjunta</u> 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*

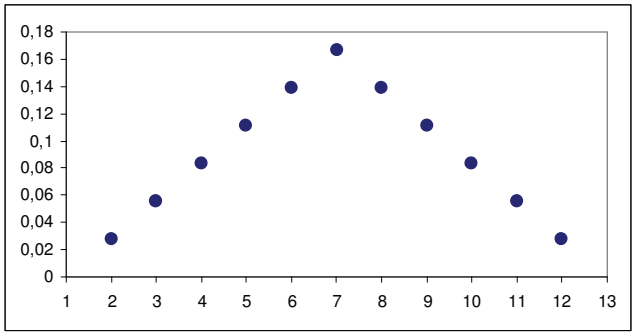
Marginal de *D*

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	

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

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

$$P(D=y \mid S=8) = P(D=y, S=8) / P(S=8)$$

Independencia

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

$$P(X=i,\ Y=j) = P(\ X=i\ ) \times P(\ Y=j\ )$$

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

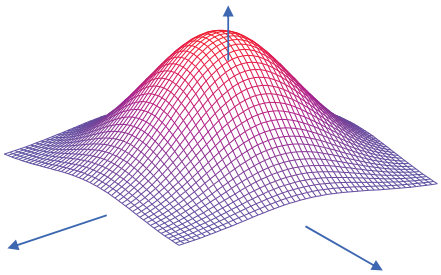
Variables aleatorias continuas

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

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

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

- Funciones de densidad marginales

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

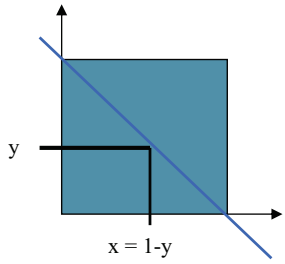
$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

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. $P(X + Y \le 1) = \iint_{x+y\le 1} 6xy^2\,dx\,dy$

   $= \int_0^1 \int_0^{1-y} 6xy^2\,dx\,dy = \frac{1}{5}$



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

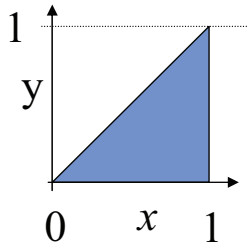
Las variables aleatorias X, Y son independientes si y sólo si

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

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

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}, \quad \text{cuando } f_Y(y) > 0$$

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}, \quad \text{cuando } f_X(x) > 0$$

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

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

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

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

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

No independientes

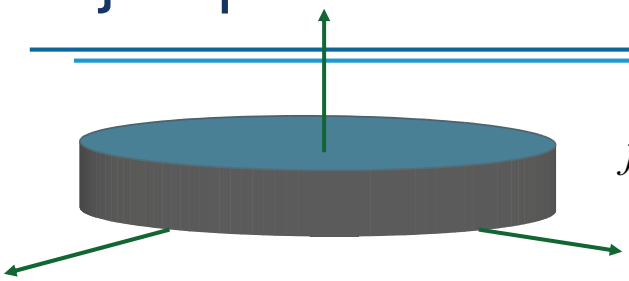
Independencia -III

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

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

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

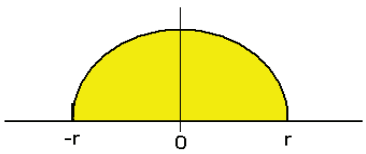
Ejemplo (cont.)

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

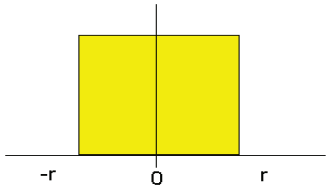
$$=\frac{1}{\pi r^2}\int_{-\sqrt{r^2-y^2}}^{+\sqrt{r^2-y^2}} dx$$

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

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



$f_Y(y)$



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

Independencia

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

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

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

Independientes

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

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 $f(x,y)$

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

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

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

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 hacer cálculo de probabilidades de un suceso en el que intervenga las variables aleatorias $X_1, X_2, \ldots, X_n$ es preciso conocer la distribución de probabilidad conjunta.

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

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

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

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

Vector de variables aleatorias

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

Su distribución de probabilidad conjunta viene caracterizada por su función de densidad conjunta : $f_{\mathbf{X}}(x_1, x_2, ..., x_n)$

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

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

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 $X_1, X_2, \ldots, X_n$ son independientes si y sólo si

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

Transformaciones Lineales

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

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

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

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

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

Transformaciones Lineales Caso General

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

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

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

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

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

Transformaciones Lineales (Independencia)

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

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

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

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

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

Ejemplo:

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

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

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

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

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

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?

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

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

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

$$= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = 1.$$

Ejemplo

- Un proceso fabrica una proporción *p* de tornillos defectuosos. Se define X como la variable *"número de tornillos extraídos del proceso hasta que aparecen r defectuosos"*. Se pide E[X] y Var[X].

$X_1 \equiv$ Número de tornillos extraídos hasta que aparece el primer defectuoso

$X_2 \equiv$ Número de tornillos **nuevos** extraídos hasta que aparece el 2º defectuoso

$X_i \equiv$ Número de tornillos **nuevos** extraídos hasta que aparece el i - ésimo defectuoso

$$X = X_1 + X_2 + \cdots + X_r$$

$$X_i : \text{variable aleatoria geométrica} \begin{cases} E[X_i] = 1/p \\ Var[X_i] = (1-p)/p^2 \end{cases}$$

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_r] = \frac{r}{p}$$

$$Var[X] = Var[X_1] + Var[X_2] + \cdots + Var[X_r] = \frac{r(1-p)}{p^2}.$$

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

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

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

Teorema Central del Límite

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

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

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

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

es la función de distribución de la normal estándar.

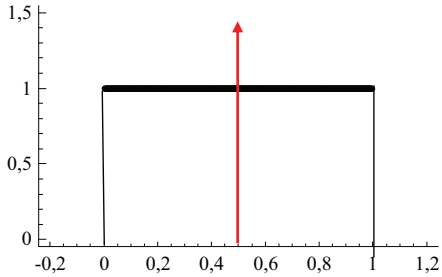
Teorema Central del Límite

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

Entonces (aprox.):

$$\overline{X} \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

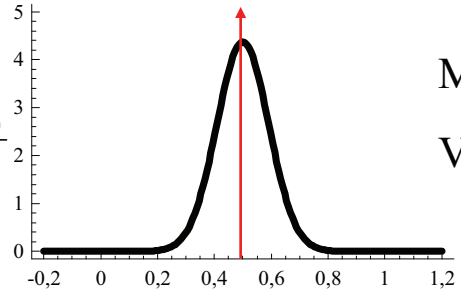
X_i



Media = 0.5

Var = 0.08

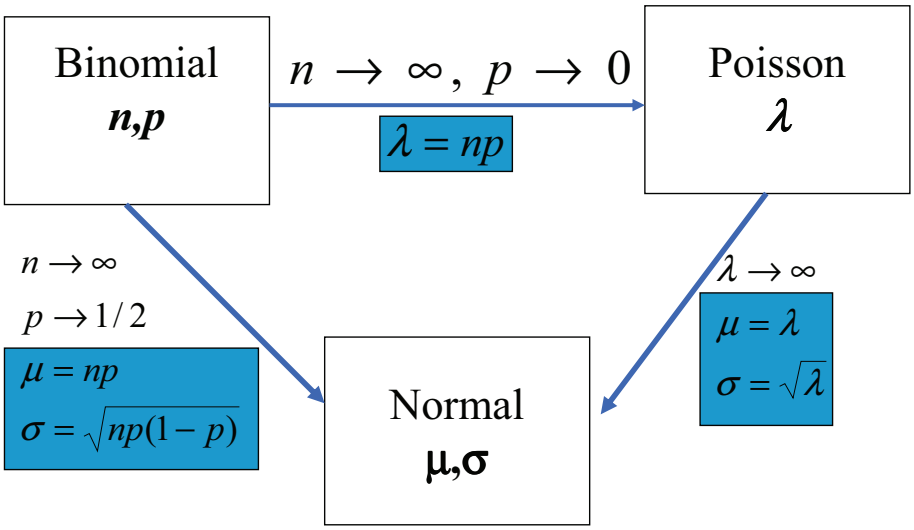
$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10}$$



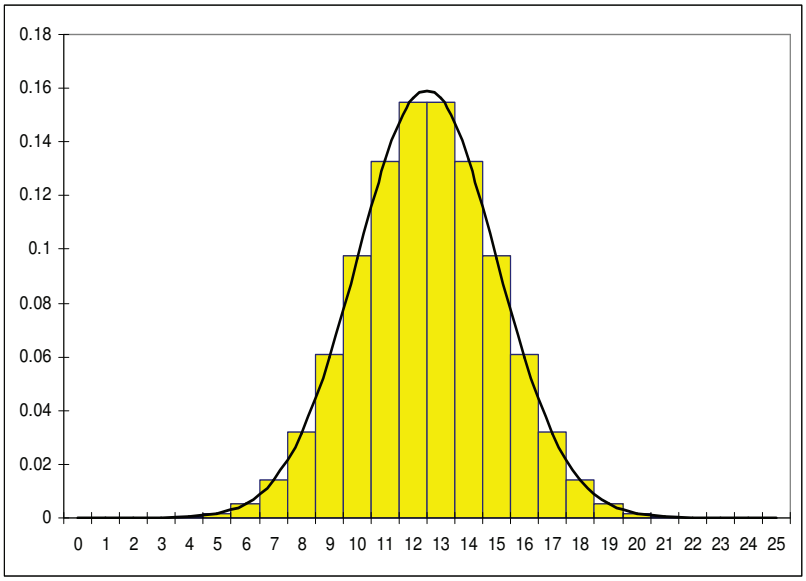
Media = 0.5

Var = 0.08/10

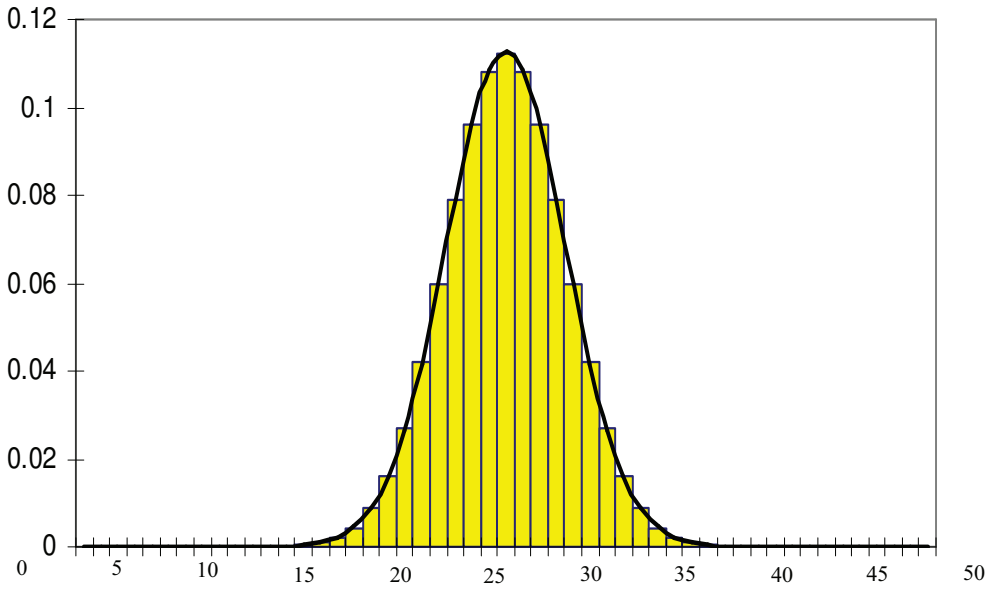
Binomial-Poisson-Normal



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

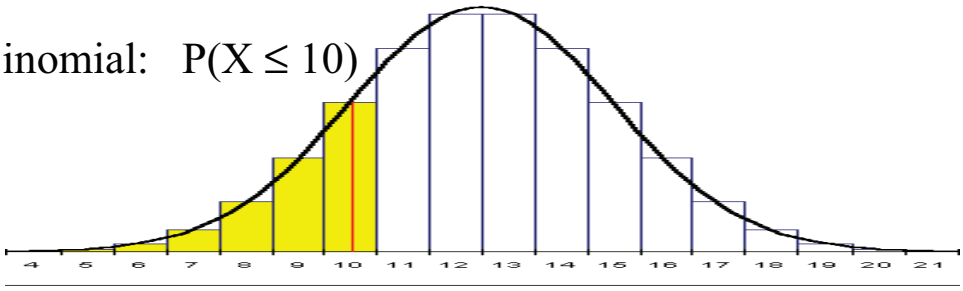


Aproximación Binomial-Normal n = 50, p=0.5

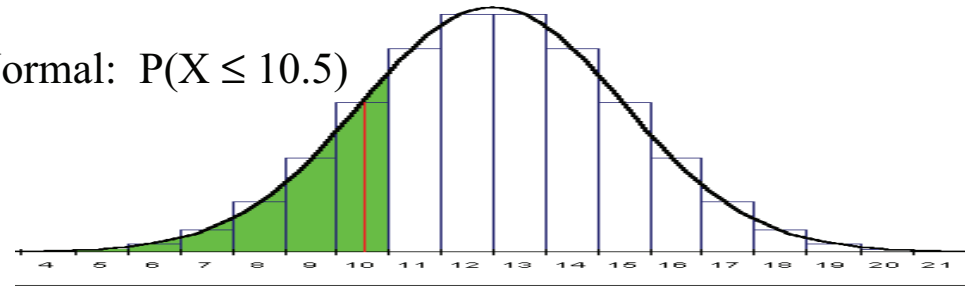


Corrección por continuidad

Binomial: $P(X \leq 10)$



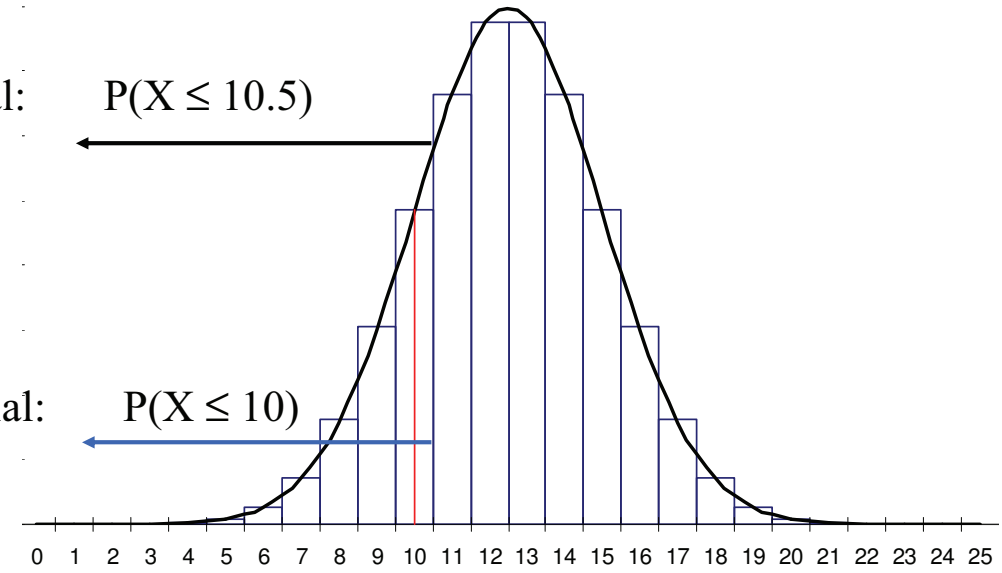
Normal: $P(X \leq 10.5)$



Corrección por continuidad

Normal: $P(X \leq 10.5)$

Binomial: $P(X \leq 10)$



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: $X \rightarrow \text{Binomial}(n = 45,\ p = 0.25)$

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

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?

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

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

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

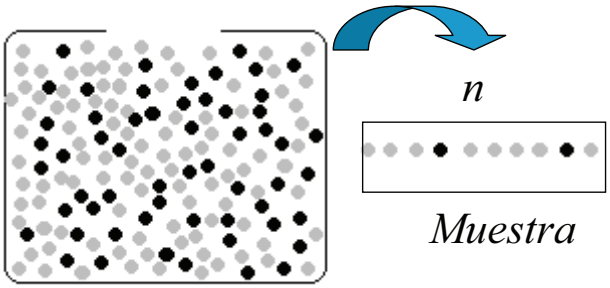
- **Riesgo del vendedor**: Probabilidad de rechazar un lote bueno (con porcentaje de defectuosas igual al p_A *(AQL))*

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

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

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

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

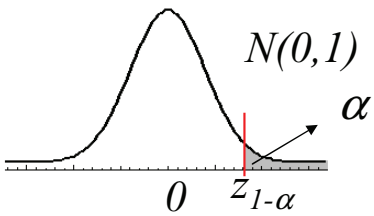
$X \rightarrow \text{Binomial}\ (n, p) \approx N(np, \sqrt{np(1-p)})$

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



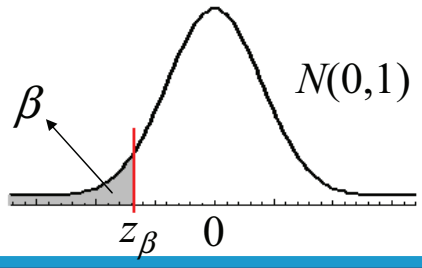
Ecuación del comprador

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



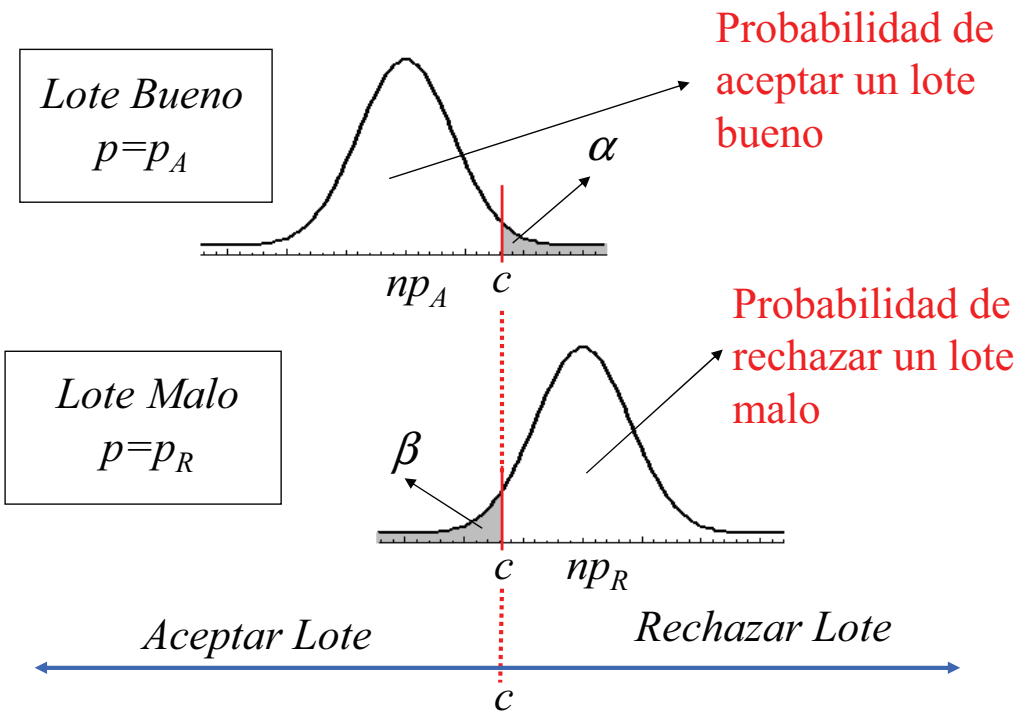
Valores de *n* y *c*

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

$$\downarrow$$

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

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



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.

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

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

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

Distribución normal multivariante

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

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

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

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

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

Distribución normal bivariante

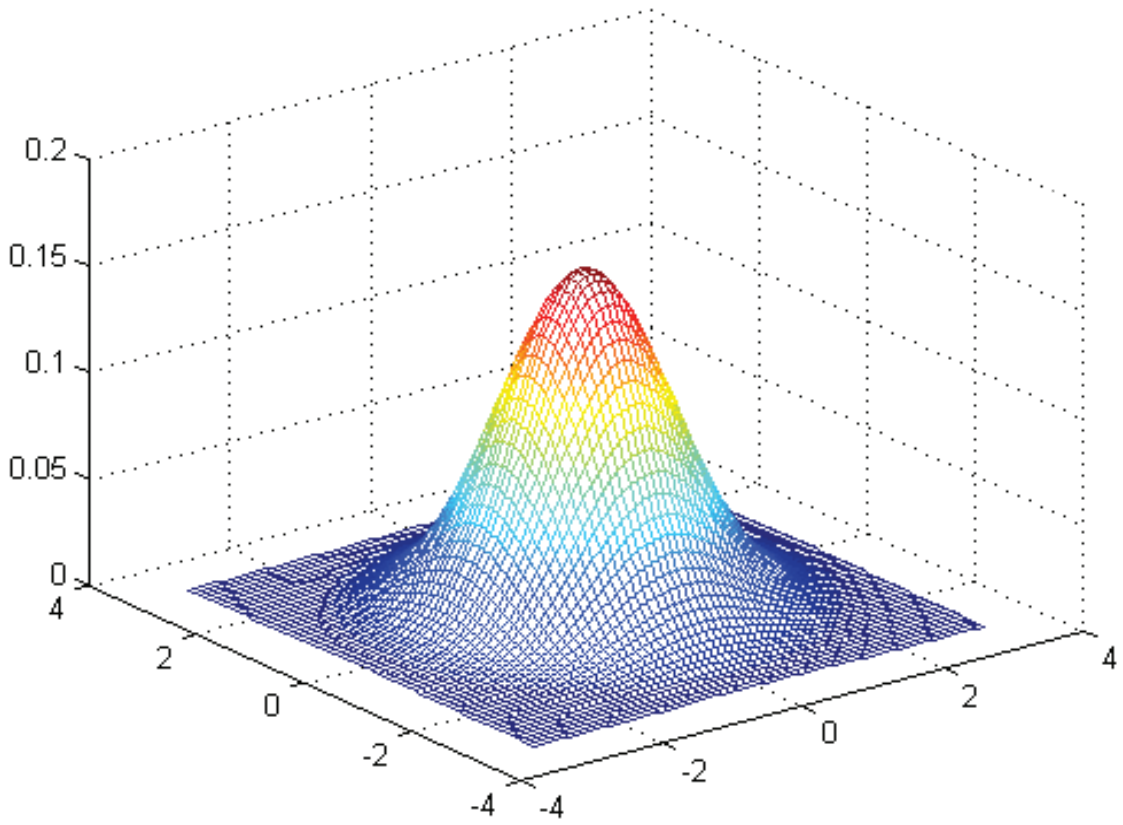
$$f(x_1, x_2) = \frac{1}{2\pi|M|^{1/2}} \exp\left\{-\frac{1}{2}(x_1-\mu_1 \; x_2-\mu_2)\mathbf{M}^{-1}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right\}$$

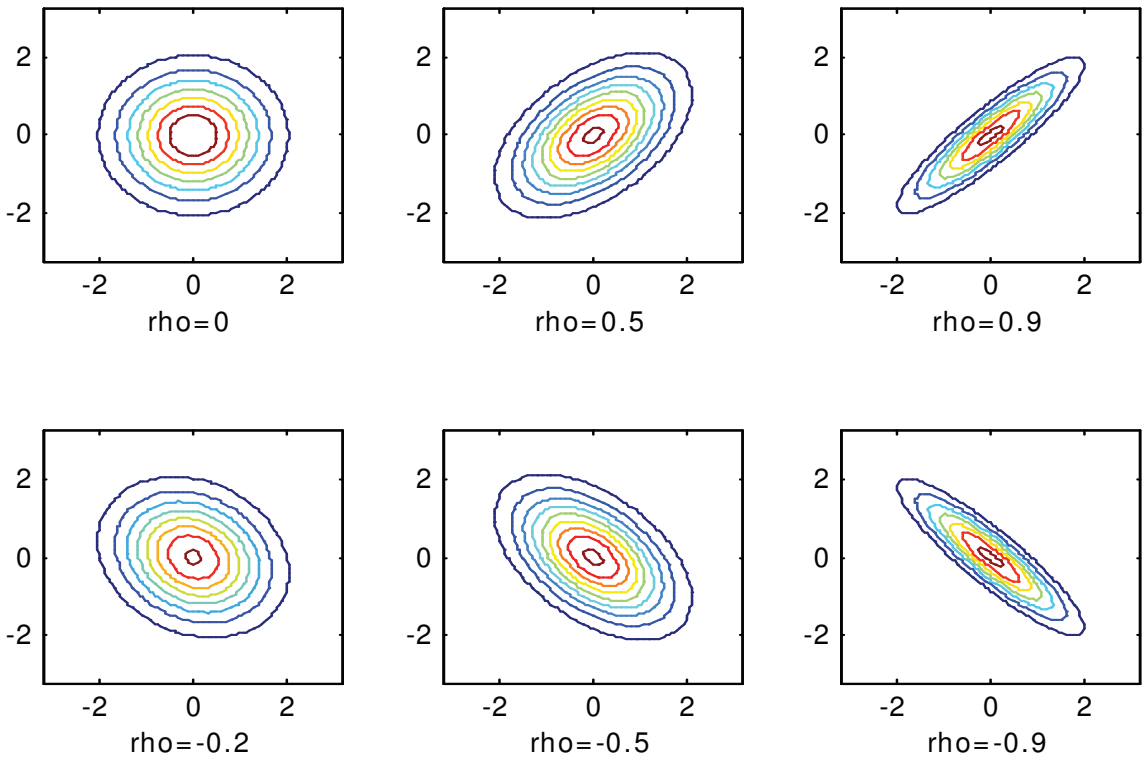
$$\mathbf{x} = (x_1, x_2)^T \in \Re^2 \qquad \mathbf{M} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix};$$

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

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

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


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

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

¿ $P(X_1 + X_2 \geq X_3 + 1)$?

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

$$\begin{aligned} P(X_1 + X_2 \geq X_3 + 1) &= P(X_1 + X_2 - X_3 \geq 1) \\ &= P(Y \geq 1) \\ &= P\left(\frac{Y-0}{\sqrt{6}} \geq 1/\sqrt{6}\right) \\ &= 1 - \Phi\left(1/\sqrt{6}\right) \end{aligned}$$

Ejercicios propuestos

Capítulo 4. Distribución conjunta de varias variables

4.1 Supóngase una diana circular con centro en el origen de coordenadas y 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 $f_{XY}(x, y)$ y $f_X(x)$.

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

$$f(p) = kp(1-p) \text{ con } 0 \leq p \leq 1$$

(a) Calcular el valor de k, la media y la varianza de la variable aleatoria p.

(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% ?

4.3 X e Y son dos variables aleatorias independientes con la misma función de distribución F. Calcular la función de densidad de

$$U = \max(\mathrm{X}, \mathrm{Y}).$$

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

4.5 Una empresa de celulosa tiene dos líneas para fabricar pasta de papel en planchas de 1m $\times$ 1m. 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.

4.6 La función de densidad de una variable aleatoria bidemensional viene dada por la expresión:

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

¿Son independientes las variables aleatorias X e Y?

4.7 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

$$\Pr\{Z = n\} = \sum_{k=0}^{n} \Pr\{X = k\}\Pr\{Y = n - k\}$$

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?

4.8 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

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

Calcular la probabilidad de que la cantidad del primer componentes sea menor que 0.3.3 miligramos cuando la del segundo es 0.8 miligramos.

4.9 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?

4.10 La función de densidad de la variable aleatoria bidemendional (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.

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

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

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

4.12 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)

4.13 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).

4.14 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.

4.15 La función de densidad conjunta de X e Y viene dada por

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

1. Obtener las funciones de densidad marginales y decir si X e Y son independientes.
   (a) Calcular $\mathrm{P}(X + Y < 1)$.

4.16 Un ordenador tarda un total de T_2 segundos en procesar un mensaje de correo electrónico, esta cantidad incluye el tiempo T_1 durante el cual el mensaje está en la cola esperando a ser procesado ($T_2 \geq T_1$). La función de densidad conjunta de las variables aleatorias T_1, T_2 es

$$f_{T_1T_2}(t_1, t_2) = e^{-t_2}, \quad 0 \leq t_1 \leq t_2 < \infty$$

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.

4.17 Sea X un valor elegido al azar de la distribución uniforme en el intervalo [0,1]. A continuación se toma al azar otro valor Y de la distribución uniforme $[X, 1]$. Calcular la función de densidad marginal de Y.

4.18 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?

4.19 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.

4.20 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 $180cm$?

4.21 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.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?

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

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

4.23 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 $\Pr(Y \leq X + 1)$.

4.24 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 $Y \to N(10, 1)$, $corr(X, Y) = -0.5$).

4.25 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 $N(100, 5)$.

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?

(a) 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.

(b) 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?

(c) Si la ganancia durante las dos horas ha sido 170, ¿cuál es el valor medio de la ganancia en la primera hora?

4.26 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.

1. Para un día, calcular la función de distribución de la variable aleatoria $Y =$*Potencia disponible* en la comarca.
2. 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 que24000 MW. (Utilizar la aproximación normal).
3. 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?

4.27 En 1958 George Edward Pelham Box y Mervin Edgar Müller desarrollaron un método para generar variables aleatorias a partir de una fuente de números aleatorios uniformemente distribuidos, que eran fáciles de obtener de forma computacional. Sobre ese método versa el siguiente problema. http://es.wikipedia.org/wiki/Método_de_Box-Muller

4.28 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?

4.28 Se tienen dos variables aleatorias U y V que se distribuyen según un modelo de probabilidad uniforme continuo entre $(0, 1)$ independientes. A partir de las anteriores se definen 2 nuevas variables aleatorias Z y W como función de las anteriores:

$$\begin{cases} Z = \sqrt{-2\log U} \\ W = 2\pi V \end{cases}$$

1. 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.

2. 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.

3. Con la información del apartado 1 y 2 obtener el valor numérico de las siguientes probabilidades:

   (a) $P(-1 < X < 1, -1 < Y < 1)$
   (b) $P(X^2 + Y^2 < 1)$
   (c) $P(Y < X)$

4. Finalmente, calcular los coeficientes de la siguiente tansformación lineal

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c & 0 \\ d & e \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

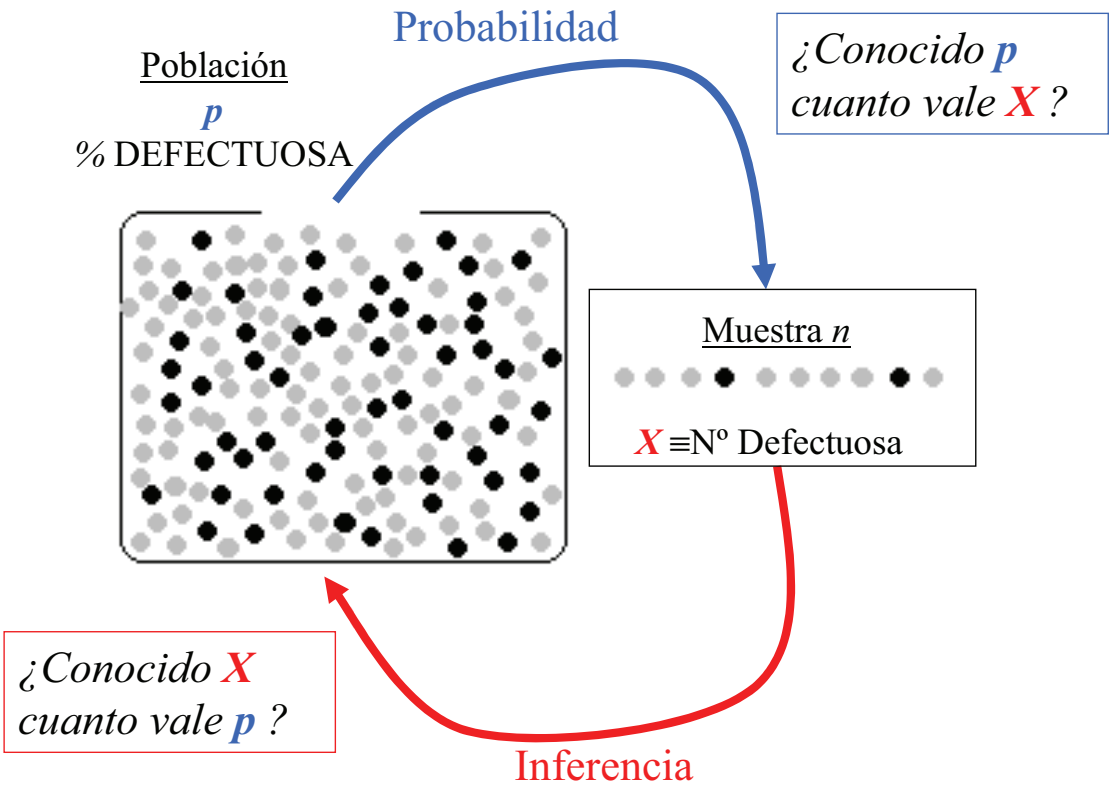
con $c, e > 0$, para que las nuevas variables aleatorias $(X', Y')^T$ tengan una distribución de probrobabilidad con vector de medias y matriz de varianzas dadas por:

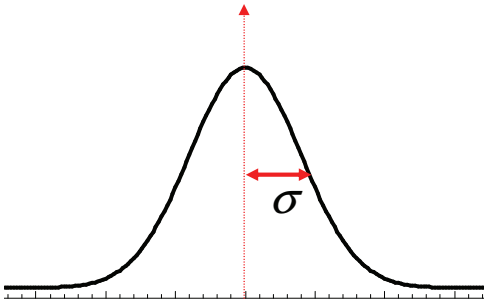
$$\mu = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \qquad S = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}.$$

5. Estimación puntual

Curso 2011-2012

Estadística



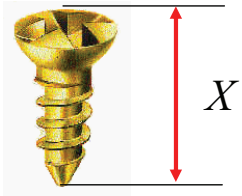


μ

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

Parámetros

$$¿\mu,\sigma?$$



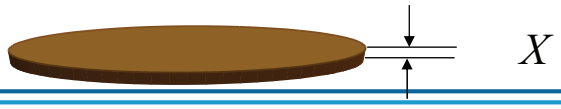
MUESTRA n

$$X_1, X_2, \ldots, X_n$$

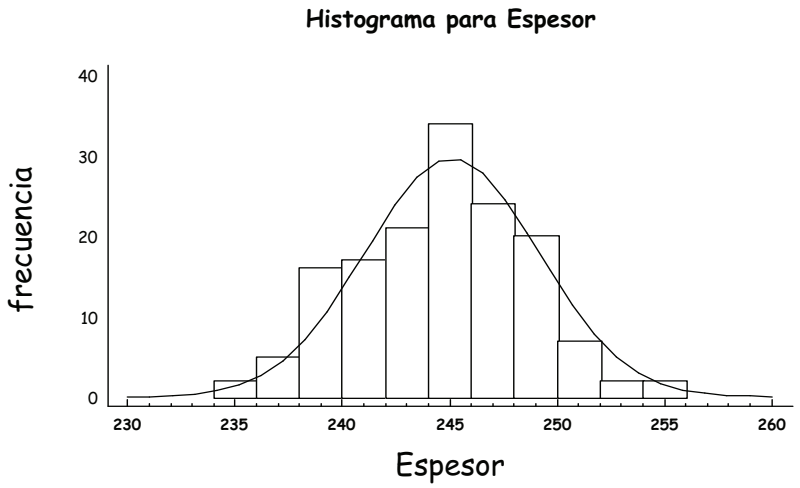
Datos Conocidos

?

Espesores de 150 obleas de Silicio (micras)



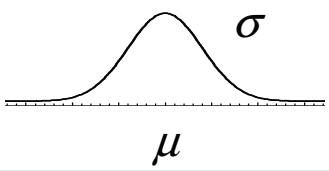
240	235	240	240	247	237	243	242	236	239
243	237	243	242	245	239	245	245	239	240
250	246	244	246	255	242	248	248	241	242
253	249	249	249	250	247	251	251	246	243
248	246	246	248	249	245	250	249	242	244
238	240	245	240	237	242	244	242	243	239
242	241	250	243	239	244	246	245	246	240
245	246	250	246	243	246	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



Distintos problemas de inferencia

Dado un modelo para los datos:

$$X_1, X_2, \ldots, X_n$$

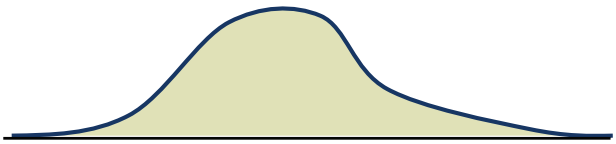


- *Estimar* μ y σ
- Dar un *intervalo de confianza* para μ y σ
- Elegir entre (*contraste de hipótesis*):

$$\mu \geq 250 \quad \text{o} \quad \mu < 250$$

- Comprobar la validez del modelo (*contraste de bondad de ajuste*).

Métodos de Estimación



$$X \to f_X(x,\theta_1,\theta_2,\ldots,\theta_r)$$

f_X conocida

Parámetros desconocidos

$\theta_1,\theta_2,\ldots,\theta_r$

Dada una muestra aleatoria simple de X

$$x_1,x_2,\ldots,x_n \quad \to \quad ¿\theta_1,\theta_2,\ldots,\theta_r\,?$$

1. Método de los momentos
2. Método de máxima verosimilitud

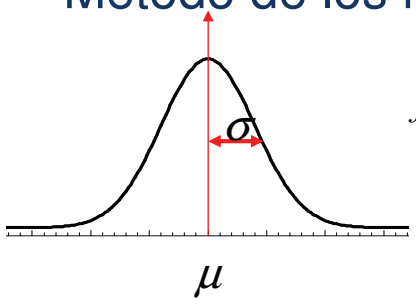
Métodos de los momentos

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 $\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_r$

$$\to \begin{cases} g_1(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_r)=a_1 \\ g_2(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_r)=a_2 \\ \vdots \\ g_r(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_r)=a_r \end{cases}$$

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?

$x_1, x_2, \ldots, x_n$

$$a_1 = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$a_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

$$f_X(x, \mu, \sigma)$$

$$\alpha_1 = E[X] = \mu$$

$$\alpha_2 = E[X^2] = \sigma^2 + \mu^2$$

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = s^2$$

Método de máxima verosimilitud (Introducción)

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:

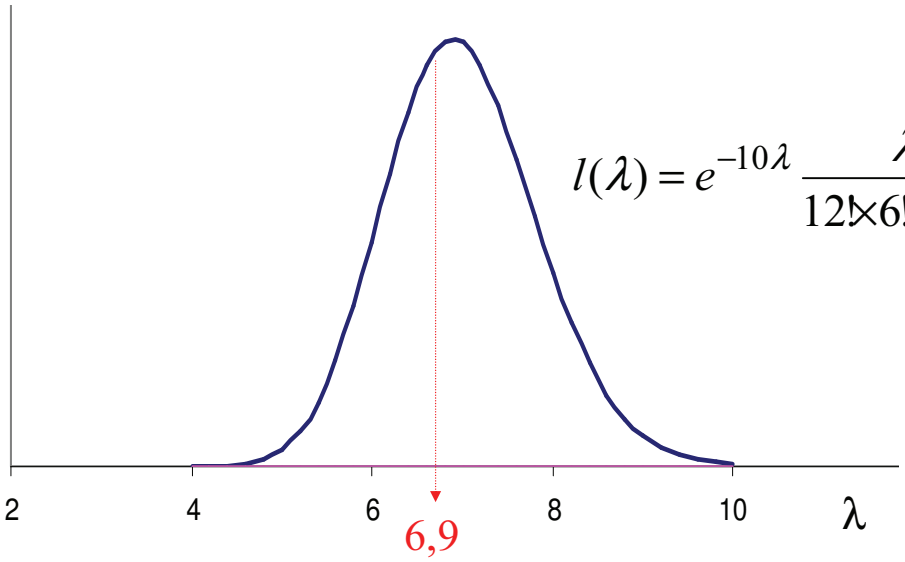
12, 6, 11, 3, 8, 5, 3, 9, 7, 5

$$P(X_1 = 12, X_2 = 6, \ldots, X_{10} = 5) = e^{-\lambda}\frac{\lambda^{12}}{12!} \times e^{-\lambda}\frac{\lambda^{6}}{6!} \times \cdots \times e^{-\lambda}\frac{\lambda^{5}}{5!}$$

$$= e^{-10\lambda}\frac{\lambda^{12+6+\cdots+5}}{12! \times 6! \times \cdots \times 5!} = e^{-10\lambda}\frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}$$

$$l(\lambda) = e^{-10\lambda}\frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}$$

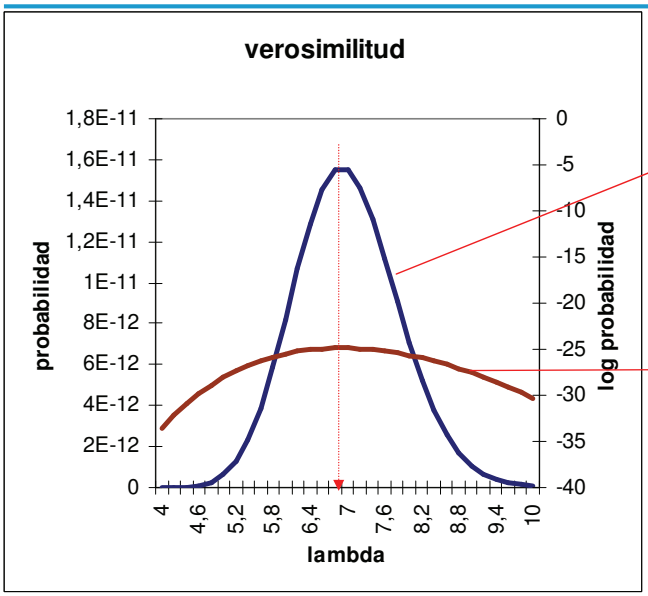
Función de verosimilitud



$$l(\lambda) = e^{-10\lambda} \frac{\lambda^{69}}{12! \times 6! \times \cdots \times 5!}$$

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

$$X \to f_X(x,\theta_1,\theta_2,...,\theta_r)$$

$$f_X \quad \text{conocida}$$

$$\text{Parámetros desconocidos}: \theta_1,\theta_2,...,\theta_r$$

$$\text{Muestra aleatoria simple}: X_1, X_2,..., X_n$$

Distribución conjunta :

$$f(x_1,x_2,...,x_n;\theta_1,...,\theta_r) =$$

$$= f_X(x_1,\theta_1,\theta_2,...,\theta_r) f_X(x_2,\theta_1,\theta_2,...,\theta_r) \cdots f_X(x_n,\theta_1,\theta_2,...,\theta_r)$$

$$\log f(x_1,x_2,...,x_n) = \sum_{i=1}^{n} \log f_X(x_i,\theta_1,\theta_2,...,\theta_r) = L(\theta_1,\theta_2,...,\theta_r)$$

$$\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_r \quad \Rightarrow \quad L(\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_r) = \max L(\theta_1,\theta_2,...,\theta_r)$$

Máx. verosimilitud: Distribución normal

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

$$X_1, X_2,..., X_n : \text{muestra aleatoria simple}$$

$$f(x_1,x_2,...,x_n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$L(\mu,\sigma^2) = \log f(x_1,x_2,...,x_n)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\max L: \left\{ \begin{array}{l} \dfrac{\partial L}{\partial \mu} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0 \\ \\ \dfrac{\partial L}{\partial \sigma^2} = -\dfrac{n}{2}\dfrac{1}{\hat{\sigma}^2} + \dfrac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = s^2 \end{array} \right.$$

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: $x_1, x_2, \ldots, x_n$

$$p(x_1,x_2,\ldots,x_n)=P(X_1=x_1, X_2=x_2,\cdots,X_n=x_n)$$

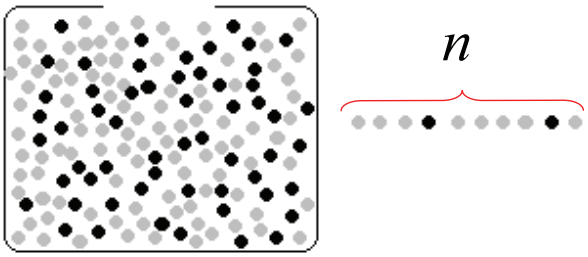
$$=e^{-\lambda}\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}\cdots e^{-\lambda}\frac{\lambda^{x_n}}{x_n!}=e^{-\lambda n}\frac{\lambda^{\Sigma x_i}}{x_1!x_2!\cdots x_n!}$$

$$L(\lambda)=-\lambda n+(\Sigma x_i)\log\lambda-\sum\log x_i!$$

$$\max L(\lambda): \frac{dL(\lambda)}{d\lambda}=-n+\frac{\Sigma x_i}{\hat{\lambda}}=0 \Rightarrow \hat{\lambda}=\bar{x}$$

Máx. verosimilitud: Binomial(*n*,*p*)

Proporción defectuosas = ¿ p ?



Muestra (Bernoulli):

$x_1, x_2, \ldots, x_n$

$$x_i=\begin{cases}0 & \text{si es } \textbf{Aceptable}\\ 1 & \text{si es } \textbf{Defectuosa}\end{cases}$$

$$p(x_1,x_2,\ldots,x_n)=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)$$

$$=P(X_1=x_1)P(X_2=x_2)\cdots P(X_n=x_n)$$

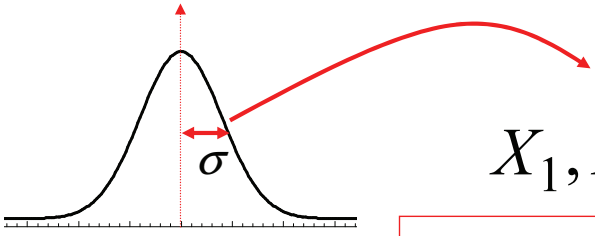
$$=p^{\sum x_i}(1-p)^{n-\sum x_i}$$

$$=p^r(1-p)^{n-r} \quad \text{donde } r=\sum x_i \text{ es el nº de defectuosas}$$

$$L(p)=r\log p+(n-r)\log(1-p)$$

$$\frac{dL(p)}{dp}=\frac{r}{\hat{p}}-\frac{n-r}{1-\hat{p}}=0 \quad \rightarrow \quad \boxed{\hat{p}=\frac{r}{n}}$$

Distribución de media (normal)



$$X_1, X_2, \ldots, X_n$$

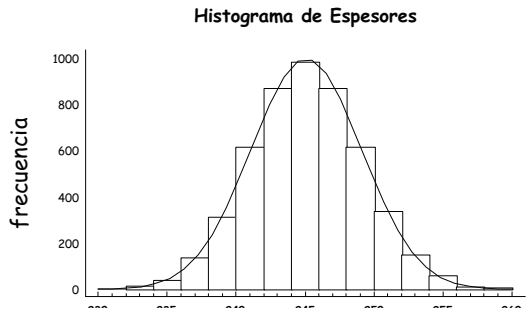
$$X \to N(\mu, \sigma)$$

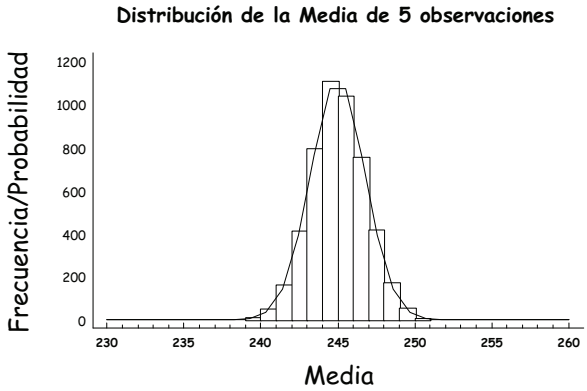
$$\boxed{\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}}$$

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

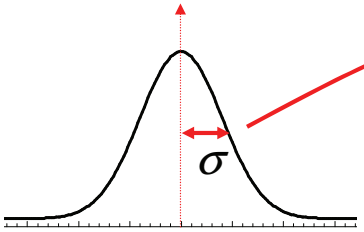
$$\mathrm{Var}[\overline{X}] = \frac{\mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots + \mathrm{Var}[X_n]}{n^2} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\boxed{\overline{X} \to N(\mu, \frac{\sigma}{\sqrt{n}})}$$





Distribución de S^2 (Normal)



$$X_1, X_2, \ldots, X_n$$

$$\mu$$

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

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

$$S^2 = \frac{1}{n}\sum_1^n (X_i - \bar{X})^2 = \frac{1}{n}\sum_1^n X_i^2 - \bar{X}^2$$

$$E[S^2] = \frac{1}{n}\Sigma E[X_i^2] - E[\bar{X}^2] = \frac{1}{n}\Sigma(\sigma^2 + \mu^2) - (\frac{\sigma^2}{n} + \mu^2) = \frac{n-1}{n}\sigma^2$$

$$\hat{S}^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

$$\hat{S}^2 = \frac{n}{n-1} S^2 \quad \Rightarrow \quad E[\hat{S}^2] = \sigma^2$$

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

- ⊗ $E[\chi_n^2] = n$
- ⊗ $\mathrm{Var}[\chi_n^2] = 2n$
- ⊗ $\chi_n^2 + \chi_m^2 = \chi_{n+m}^2 \quad (\chi_n^2 \text{ y } \chi_m^2 \text{ indep.})$

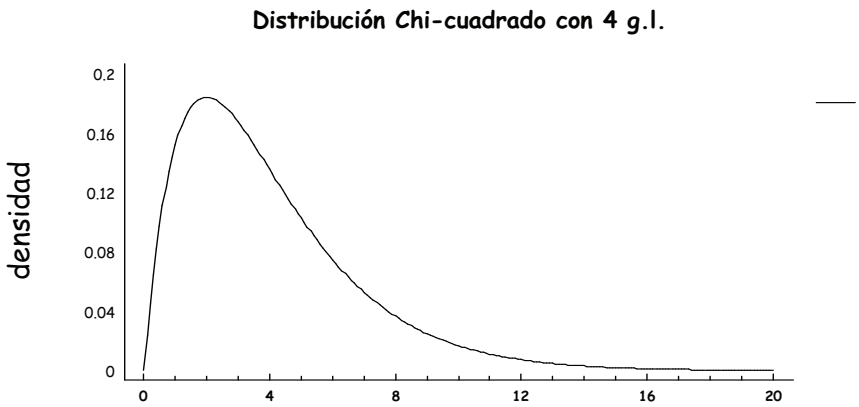
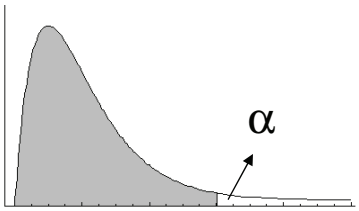


Tabla χ^2

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



$\chi_{\nu,1-\alpha}$

ν: grados de libertad (g.l.)

EJEMPLO

$P(\chi_9 \geq 19{,}02) = 0{,}025$

Distribución de S^2 (Normal)

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \to \chi_n^2$$

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\mu)^2$$

$$= \sum_{i=1}^{n}(X_i-\overline{X})^2 + \sum_{i=1}^{n}(\overline{X}-\mu)^2 + 2\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)$$

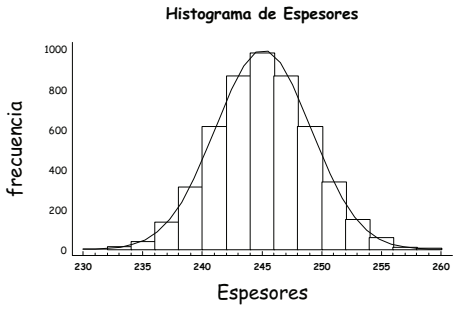
$$= \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$$

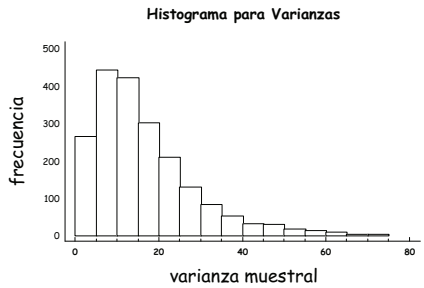
$$\underbrace{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}}_{\chi_n^2} = \underbrace{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}}_{\chi_{n-1}^2} + \underbrace{\frac{(\overline{X}-\mu)^2}{\sigma^2/n}}_{\chi_1^2}$$

Distribución de S^2 (Normal)

$$\frac{nS^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \xrightarrow{dist} \chi_{n-1}^2$$

$$\frac{(n-1)\hat{S}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \xrightarrow{dist} \chi_{n-1}^2$$





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

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

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \Rightarrow \begin{cases} E[\overline{X}] = \mu \\ Var[\overline{X}] = \dfrac{\sigma^2}{n} \end{cases} \Rightarrow \overline{X} \xrightarrow{n \to \infty} N(\mu, \frac{\sigma}{\sqrt{n}})$$

Binomial

$$\text{Binomial}: X_1, X_2, \ldots, X_n$$

$$X_i = \begin{cases} 0 & \text{si es } \textbf{Aceptable} \\ 1 & \text{si es } \textbf{Defectuosa} \end{cases} \Rightarrow \begin{array}{c} E[X_i] = p \\ Var[X_i] = p(1-p) \end{array}$$

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \begin{cases} E[\hat{p}] = p \\ Var[\hat{p}] = \dfrac{p(1-p)}{n} \end{cases}$$

$$\boxed{\hat{p} \underset{n\to\infty}{\rightarrow} N(p, \sqrt{\frac{p(1-p)}{n}})}$$

Poisson

$$\text{Poisson}: X_1, X_2, \ldots, X_n$$

$$P(X_i = k) = e^{-\lambda} \frac{\lambda^k}{k!} \Rightarrow \begin{cases} E[X_i] = \lambda \\ Var[X_i] = \lambda \end{cases}$$

$$\hat{\lambda} = \frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \begin{cases} E[\hat{\lambda}] = \lambda \\ Var[\hat{\lambda}] = \dfrac{\lambda}{n} \end{cases}$$

$$\boxed{\hat{\lambda} \underset{n\to\infty}{\rightarrow} N(\lambda, \sqrt{\frac{\lambda}{n}})}$$

Propiedades de los estimadores

$$X_1, X_2, \ldots, X_n \text{ m.a.s de } f(x,\theta) : \hat{\theta} \equiv \hat{\theta}(X_1, X_2, \ldots, X_n)$$

- Centrados: $E[\hat{\theta}] = \theta \qquad (\text{Sesgo}[\hat{\theta}] = E[\hat{\theta}] - \theta)$
- 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

$$X_1, X_2, \ldots, X_n \text{ m.a.s de } N(\mu, \sigma) : \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

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

$$X_1, X_2, \ldots, X_n \text{ m.a.s de } N(\mu,\sigma): S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

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 :

$$\lim_{n\to\infty} \text{E}[S^2] = \sigma^2 \quad \text{y} \quad \lim_{n\to\infty} \text{Var}[S^2] = 0$$

Ejercicios propuestos

Capítulo 5. Estimación Puntual

5.1 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.

5.2 Los taxis en servicio de una ciudad están numerados del 1 al N. Se observa una muestra de 10 taxis y se apuntan sus números. Obtener un estimador de N por el método de los momentos.

5.3 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. 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	17	5	3

Estimar θ por máxima verosimilitud.

5.4 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 $F(t) = 1-exp(-\alpha t)$, calcular la probabilidad estimada de esperar al autobús más de 10 minutos.

5.5 La función de distribución de una variable aleatoria es

$$F(x) = \begin{cases} 0 & x < 0, \\ (x/\beta)^\alpha & 0 \leq x \leq \beta, \\ 1 & x > \beta. \end{cases}$$

donde los parámetros α y β son positivos. Estimar los parámetros de la distribución por el método de máxima verosimilitud.

5.6 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 13 mejores tiradores.

Sabiendo que la función de densidad de la variable aleatoria presentada es

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

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

y la distancia de los otros tres fue mayor que la distancia máxima permitida en su categoría que es de $11cm$.

5.7 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

5.8 Se supone que el tiempo que tarda un paciente en recuperarse totalmente de una dolencia cuando se le aplica el tratamiento A es una variable aleatoria 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 hasta la recuperación total..

5.9 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.

5.10 Sea $X_1, X_2, ..., X_n$ una muestra aleatoria simple de la función de densidad

$$f(x) = 2(\theta - x)/\theta^2, \quad 0 \leq x \leq \theta.$$

Obtener por el método de los momentos un estimador insesgado de θ y calcular su varianza.

5.11 Sea $\overline{X}$ la media aritmética de una muestra aleatoria simple de una distribución $\mathrm{N}(\mu, \sigma)$. Se define $\hat{X} = c\overline{X}$ 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 $\sigma/\mu = 2$.

5.12 $X_1, X_2, ..., X_n$ es una muestra aleatoria simple de una distribución normal con parámetros desconocidos. Para estimar la varianza se propone el siguiente estimador

$$S^2 = k \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (X_i - X_j)^2.$$

Determinar k para que el estimador sea centrado.

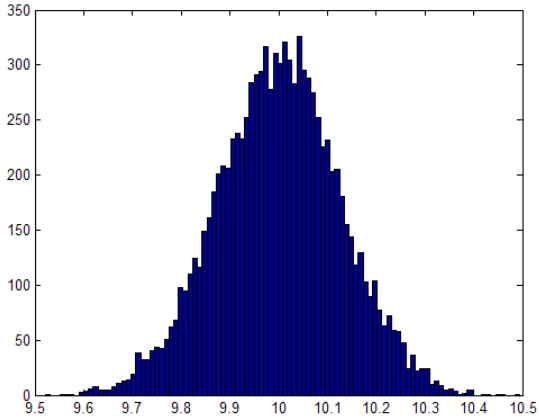
5.13 Para estimar la media σ^2 de una población normal se utiliza el estimador $\widehat{\sigma}^2 = k\widehat{s}^2$, siendo $\widehat{s}^2$ la varianza muestral corregida y k una constante. Calcular el valor de k que minimiza el error cuadrático medio. (Utilizar $\mathrm{V}ar[\chi^2_g] = 2g$, siendo g el número de grados de libertad).

5.14 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.

5.15 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.

5.16 Un estudiante ha realizado el siguiente ejercicio de simulación. Ha tomado generado con STATGRAPHICS 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.



A partir de esta simulación, explica qué propiedades tiene la mediana como estimador de la media de una distribución normal, sus ventajas e inconvenientes.

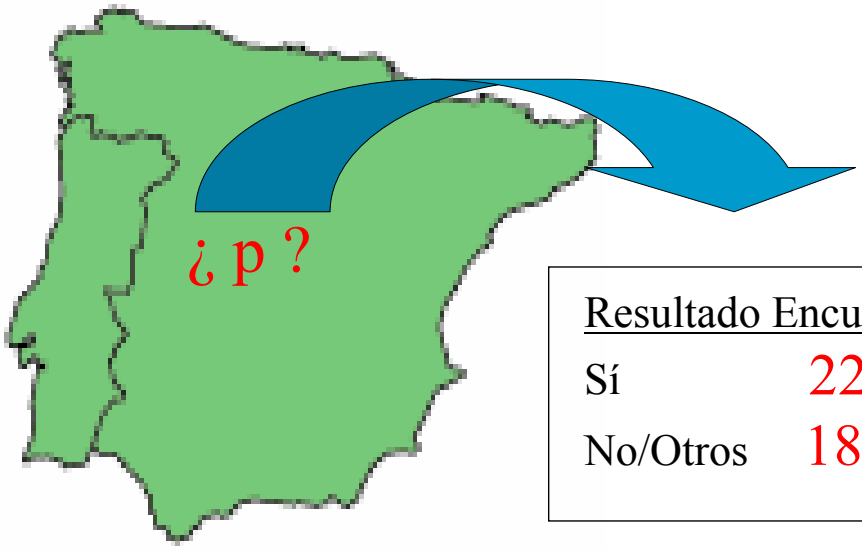
6. Intervalos de confianza

Curso 2011-2012

Estadística

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.



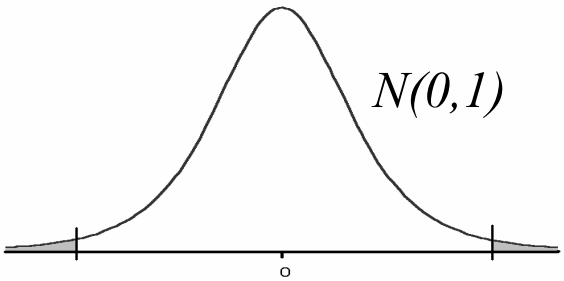
Resultado Encuesta	
Sí	220
No/Otros	180

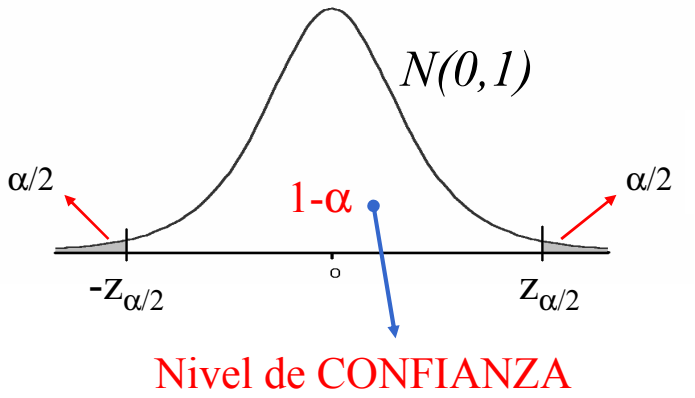
Introducción

$$X \rightarrow B(n,p) \qquad X \overset{aprox.}{\rightarrow} N(np, \sqrt{np(1-p)})$$

$$\hat{p} = \frac{X}{n} \rightarrow N(p, \sqrt{\frac{p(1-p)}{n}})$$

$$\boxed{\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0,1)}$$





$$P(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}) = 1-\alpha$$

Despejando p de: $\boxed{-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}}$

$$-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}$$

$$\Downarrow$$

$$-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le z_{\alpha/2}$$

$$\Downarrow$$

$$\boxed{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le \quad p \quad \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

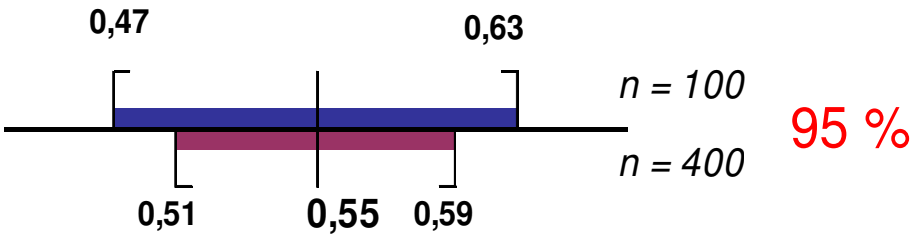
Nivel de confianza: (1-α)

Tamaño Muestral n

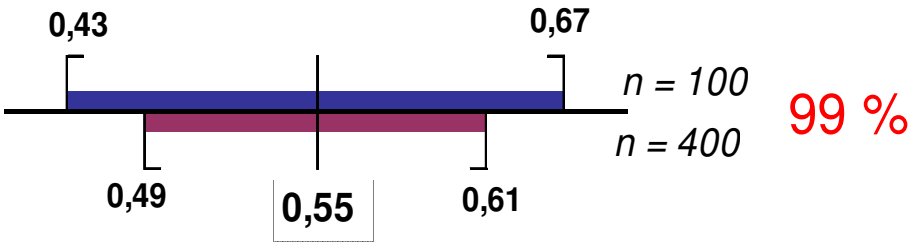
Ejemplo

$$\hat{p} = \frac{220}{400} = 0{,}55$$

$$p \in 0{,}55 \pm 1{,}96\sqrt{\frac{0{,}55 \times 0{,}45}{400}}$$



95 %



99 %

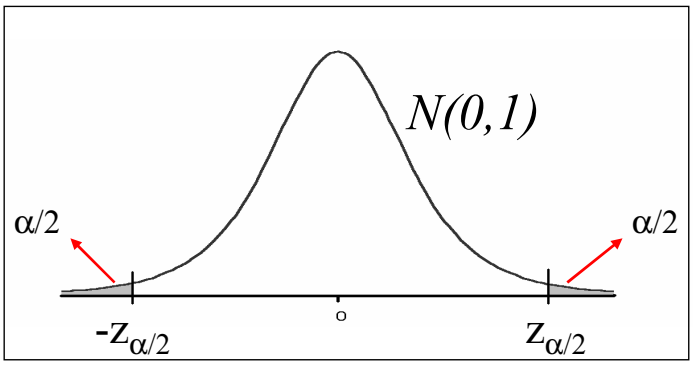
1. Normal: Intervalo para μ con σ conocido

$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$\overline{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \rightarrow N(0,1)$$

$$-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \leq z_{\alpha/2}$$



$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\boxed{\mu \in \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

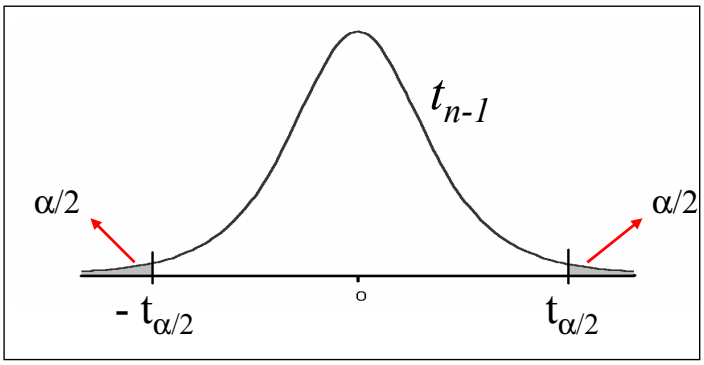
2. Normal:Intervalo para μ con σ desconocido

$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$\overline{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \rightarrow N(0,1)$$

$$\frac{\overline{X} - \mu}{\hat{S} / \sqrt{n}} \rightarrow t_{n-1}$$



$$-t_{n-1,\alpha/2} \leq \frac{\overline{X} - \mu}{\hat{S} / \sqrt{n}} \leq t_{n-1,\alpha/2}$$

$$\boxed{\bar{x} - t_{n-1,\alpha/2} \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,\alpha/2} \frac{\hat{s}}{\sqrt{n}}}$$

$$\boxed{\mu \in \bar{x} \pm t_{n-1,\alpha/2} \frac{\hat{s}}{\sqrt{n}}}$$

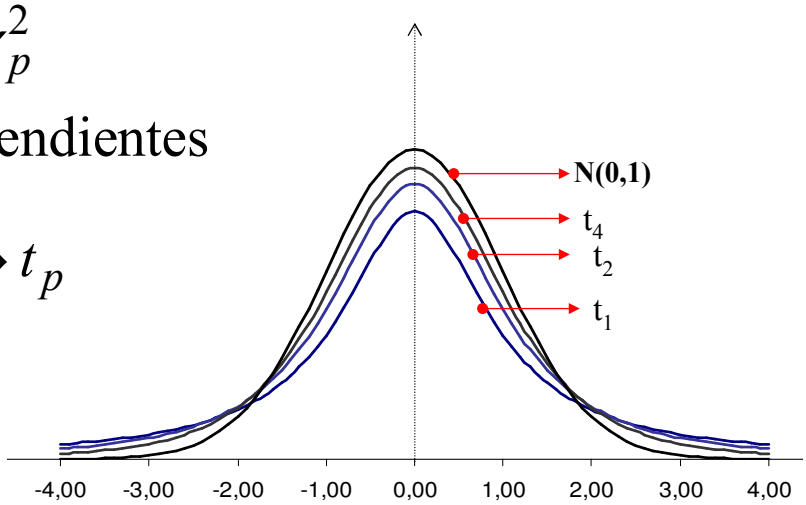
Distribución **t** de *Student*

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

$$V \to \chi^2_p$$

Z, V son independientes

$$\frac{Z}{\sqrt{V/p}} \to t_p$$



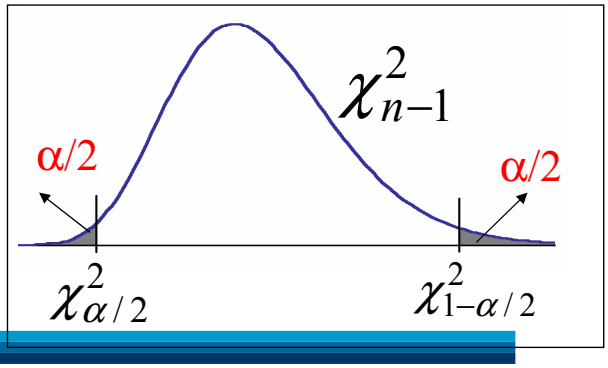
3. Normal: Intervalo para σ^2

$$X_1, X_2, \ldots, X_n \to N(\mu, \sigma)$$

$$\hat{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \longrightarrow \frac{(n-1)\hat{S}^2}{\sigma^2} \to \chi^2_{n-1}$$

$$P\left(\chi^2_{\alpha/2} \le \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2_{1-\alpha/2}\right) = 1-\alpha$$

$$\boxed{\frac{(n-1)\hat{s}^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)\hat{s}^2}{\chi^2_{\alpha/2}}}$$

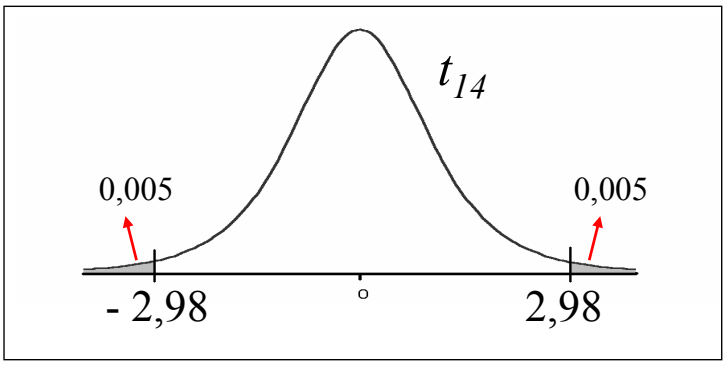


EJEMPLO 1. La resistencia a la compresión de 15 probetas de acero elegidas al azar es:

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

$$\frac{\overline{X}-\mu}{\hat{S}/\sqrt{15}} \rightarrow t_{14}$$

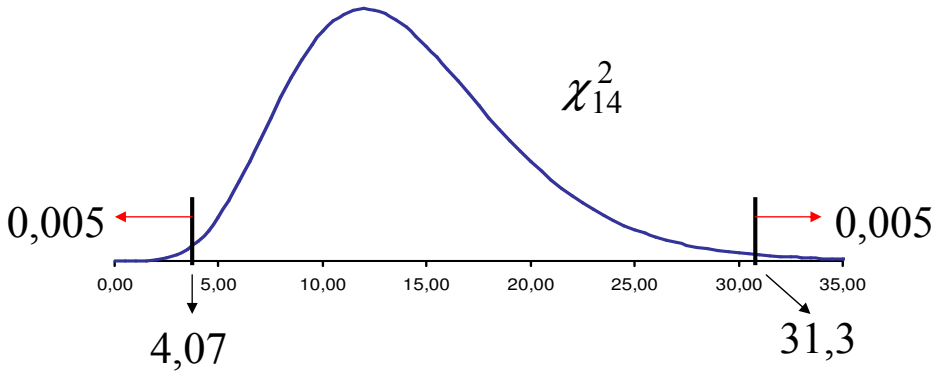
$$-2{,}98 \le \frac{\overline{X}-\mu}{\hat{S}/\sqrt{15}} \le 2{,}98$$



$$\overline{x} = 45{,}75 \qquad \hat{s} = 14{,}2$$

$$45{,}75 - 2{,}98\frac{14{,}2}{\sqrt{15}} \le \mu \le 45{,}75 + 2{,}98\frac{14{,}2}{\sqrt{15}}$$

99 % confianza: $34{,}8 \le \mu \le 56{,}7$



$$\frac{(n-1)\hat{S}^2}{\sigma^2} \rightarrow \chi^2_{n-1} \qquad \frac{14\hat{S}^2}{\sigma^2} \rightarrow \chi^2_{14}$$

$$P\left(4{,}07 \le \frac{14\hat{S}^2}{\sigma^2} \le 31{,}3\right) = 0{,}99$$

$$\hat{s}^2 = 201{,}6 \longrightarrow \frac{14\times 201{,}6}{31{,}3} \le \sigma^2 \le \frac{14\times 201{,}6}{4{,}07}$$

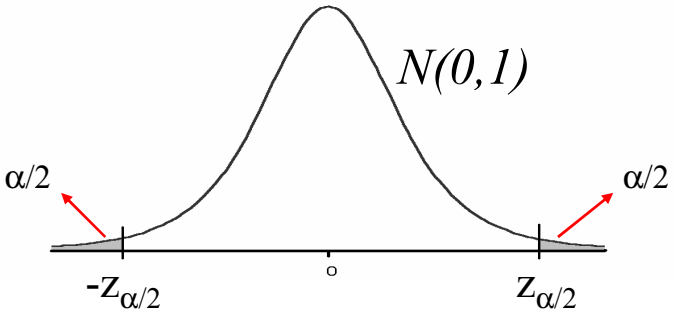
99% confianza: $90{,}2 \le \sigma^2 \le 693{,}6$

4. Poisson: Intervalo para λ

$$X_1, X_2, \ldots, X_n \rightarrow \text{Poisson}\,(\lambda)$$

$$\hat{\lambda} = \overline{X} \overset{aprox}{\rightarrow} \text{N}(\lambda, \sqrt{\frac{\lambda}{n}})$$

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} \rightarrow \text{N}(0,1)$$



$$P(-z_{\alpha/2} \leq \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda / n}} \leq z_{\alpha/2}) = 1 - \alpha$$

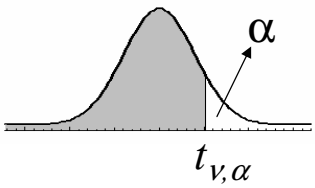
$$\boxed{\hat{\lambda} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}}$$

Tabla

t-Student

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

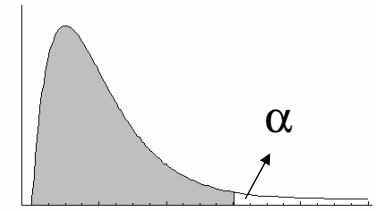


ν: grados de libertad (g.l.)

EJEMPLO

$P(t_9 \geq 2{,}262) = 0{,}025$

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

Ejercicios propuestos

Capítulo 6. Intervalos de confianza

6.1 Se han tomado 12 valores de una variable física X, que se supone 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.

6.2 En la lista adjunta se indica la edad y el área científica en que trece importantes científicos de diversas áreas descubrieron la teoría que les ha dado la fama. Construir con estos datos un intervalo de confianza para la edad a la que los científicos realizan su contribución más importante: Galileo (34, astronomía), Franklin (40, electricidad), Lavoisier (31, química), Lyell (33, geología), Darwin (49, biología), Maxwell (33, ecuaciones de la luz), Curie (34, radiactividad), Plank (43, teoría cuántica), Marx (30, socialismo científico), Freud (31, psicoanálisis), Bohr (26, modelo del átomo), Einstein (26, relatividad), Keynes (36, macroeconomía).

6.3 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.

6.4 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%?.

6.5 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. Obtener un intervalo de confianza ($\alpha = 0.05$) para la proporción p de compradores potenciales de este nuevo producto. ¿Cúal debería ser el tamaño muestral si se quisiera reducir la longitud del intervalo a la mitad.

6.6 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?.

6.7 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 $P(Z \leq 1.96) = 0.975$).

6.8 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, $f(x) = \frac{1}{\lambda}e^{-x/\lambda}, \quad x \geq 0,\ \lambda > 0$; 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.

6.9 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.

6.10 La velocidad de una molécula según el modelo de Maxwell, es una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \dfrac{4}{\sqrt{\pi}} \times \dfrac{1}{\alpha^3} x^2 \exp -(x/\alpha)^2, \ x \geq 0 \\ 0, \ x \leq 0. \end{cases}$$

donde $\alpha > 0$, es el parámetro de la distribución y se verifica que

$$E(X) = \frac{2\alpha}{\sqrt{\pi}} \text{ y } Var(X) = \frac{3}{2} - \frac{4}{\pi}\alpha^2.$$

(a) Calcular el estimador máximo verosímil de α y su varianza asintótica.

(b) Calcular el estimador por momentos de α y la varianza de dicho estimador.

(c) Para una muestra de tamaño n=100, para la que se verifica que $\sum_{i=1}^{100} x_i = 342$ y que $\sum_{i=1}^{100} x_i^2 = 1339$,hallar un intervalo de confianza de α con el 95% de confianza utilizando ambos estimadores.

6.11 Los núcleos (radionucleidos) del elemento radiactivo Carbono 14 (C^{14}) se desintegran aleatoriamente. El tiempo que tarda en desintegrarse cada radionucleido es una variable aleatoria con distribución exponencial de media $8,27 \times 10^3$ años.

(a) Si inicialmente había 10^{12} radionucleidos, obtener el número esperado de los radionucleidos sin desintegrar al cabo de los 20.000 años.

(b) Obtener, para la variable aleatoria *número de radionucleidos sin desintegrar al cabo de 20.000 años,* un intervalo que contenga al valor de esa variable con probabilidad $0,95$ e interpretar el resultado.

(c) Una pieza arqueológica ha estado enterrada durante 20.000 años al cabo de los cuales se han observado 10^{10} radionucleidos de C^{14}. Estimar por el método de los momentos el número inicial de radionucleidos N y calcular la media y la varianza del estimador obtenido.

(d) Determinar el tiempo que debe transcurrir para que el número de radionucleidos iniciales se reduzca a la mitad.

7. Contrastes de Hipótesis

Curso 2011-2012

Estadística

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

$$H_0: \quad p \le 0.5$$
$$H_1: \quad p > 0.5$$

$$\boxed{\begin{aligned} H_0: &\quad p = 0.5 \\ H_1: &\quad p > 0.5 \end{aligned}}$$

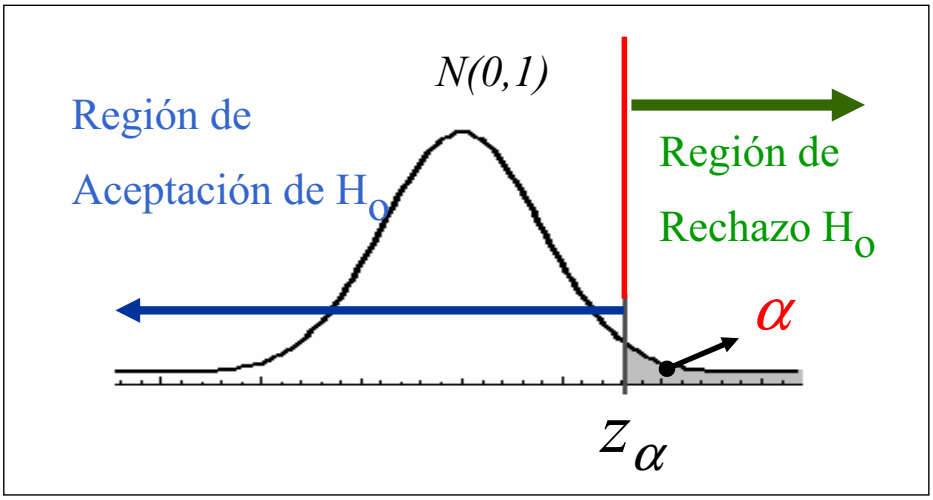
$$\hat{P} \to N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Si H_0 es cierto $p = 0.5$

$$\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} \to N(0,1)$$

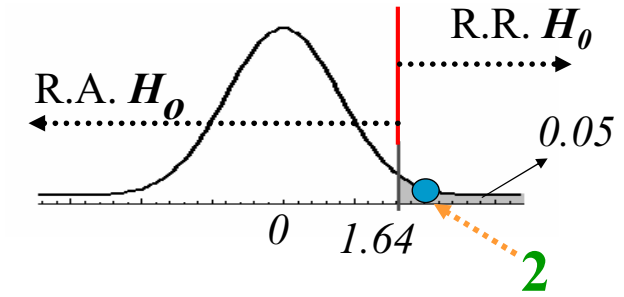
Nivel de significación α

$$\alpha = P(\text{ rechazar } H_0 \mid H_0 \text{ es cierta})$$



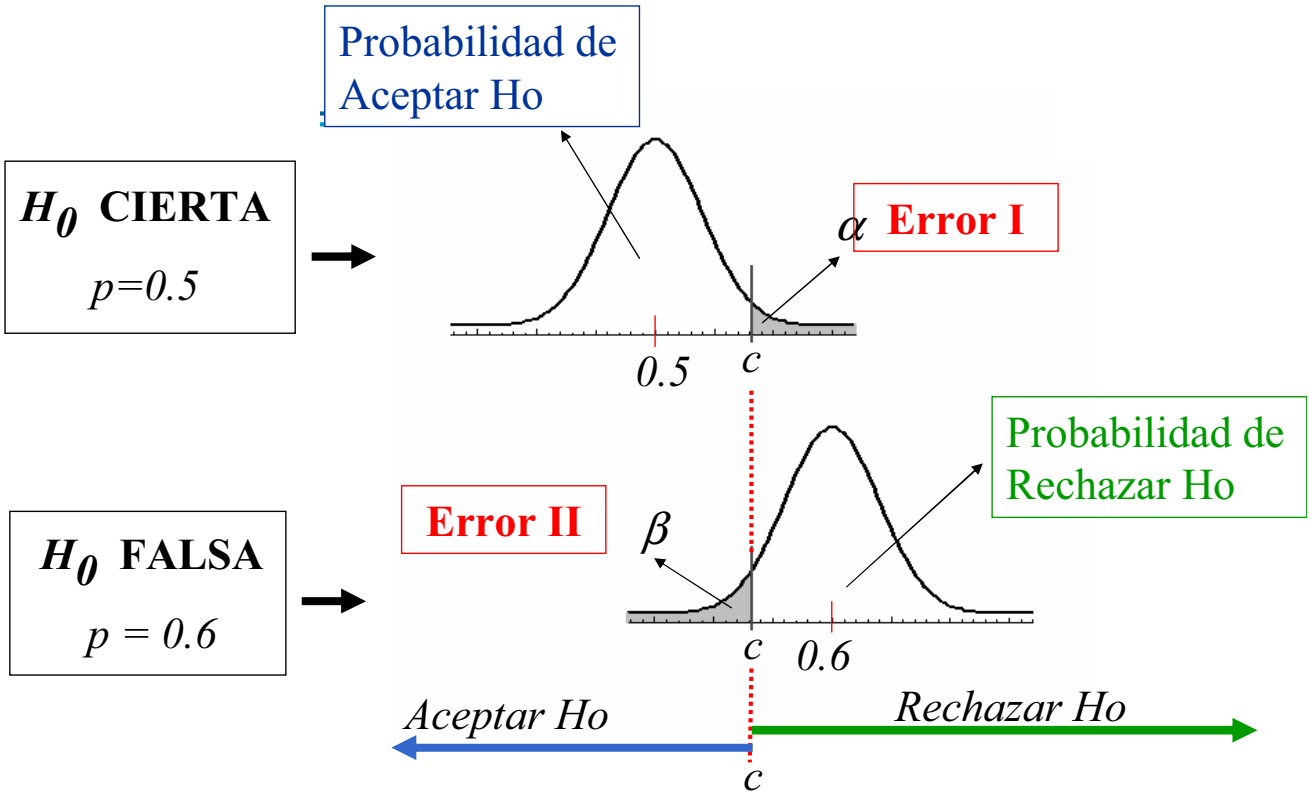
$$\text{Muestra } \hat{p} = 0.55 \Rightarrow z_0 = \frac{0.55 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = 2$$

Elegido el nivel de significación $\alpha = 0.05$



$z_0 = 2 > z_\alpha = 1.64 \quad \Rightarrow \quad$ Se rechaza H_0

Con un nivel de significación de $\alpha = 0.05 \quad \Rightarrow \quad p > 0.5$



Tipos de errores

REALIDAD	RESULTADO CONTRASTE: Se **Acepta** H_o	RESULTADO CONTRASTE: Se **Rechaza** H_o
H_o **CIERTA**	**Ok**	**Error tipo I** α
H_o **FALSA**	**Error tipo II** β	**Ok**

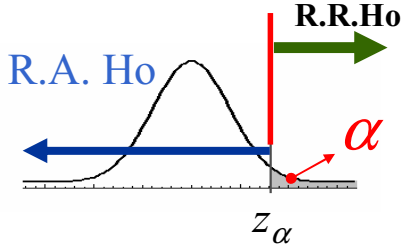
Tipos de Contrastes

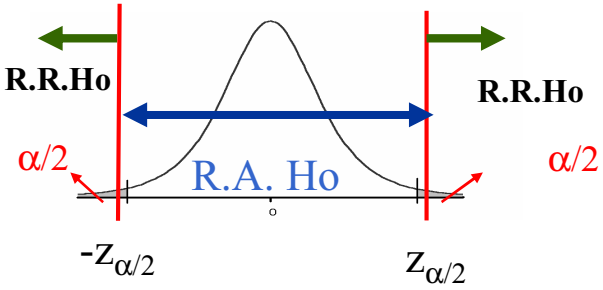
$$H_0: \quad p = 0.5$$
$$H_1: \quad p > 0.5$$

UNILATERAL

$$H_0: \quad p = 0.5$$
$$H_1: \quad p \neq 0.5$$

BILATERAL





Nivel de significación α

Nivel crítico o *p*-valor

$$\boxed{\begin{array}{ll} H_0: & p = p_0 \\ H_1: & p > p_0 \end{array}} \qquad \hat{P} \to N(p, \sqrt{\frac{p(1-p)}{n}})$$

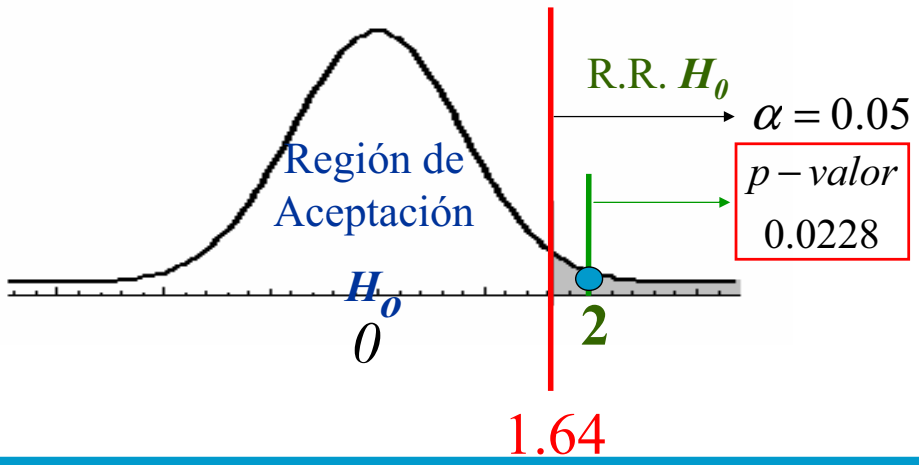
$$p - valor = \Pr(\hat{P} > \hat{p} \mid H_0 \text{ es cierto})$$

$$= \Pr\left(\frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}} > \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right)$$

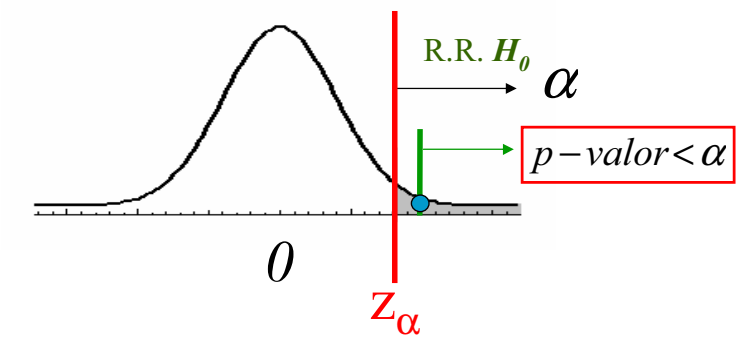
$$= \Pr(Z > z_0)$$

Ejemplo

Datos: $\hat{p} = \frac{220}{400} = 0.55$



Relación entre ***p-valor*** y α

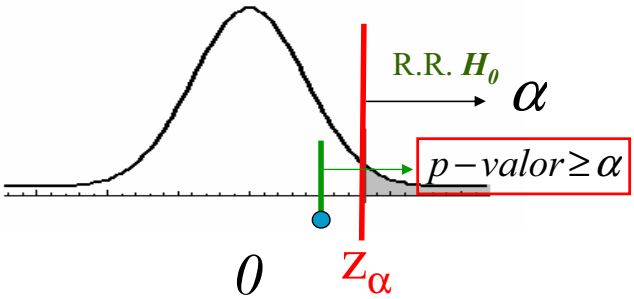


p-valor < α

Se rechaza H_0

p-valor ≥ α

No se rechaza H_0



1. Normal: Contraste para μ con σ conocido

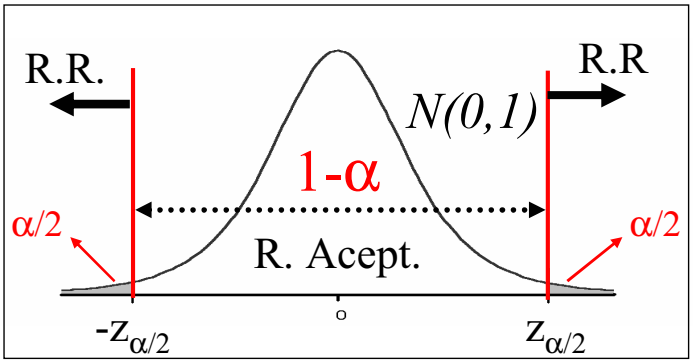
$$X_1, X_2, \ldots, X_n \rightarrow N(\mu, \sigma)$$

$$H_0: \quad \mu = \mu_0$$

$$H_1: \quad \mu \neq \mu_0$$

$$\overline{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \rightarrow N(0,1)$$



$$z_0 = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \longrightarrow$$

$$|z_0| \leq z_{\alpha/2} \Rightarrow \text{No se rechaza } H_0$$

$$|z_0| > z_{\alpha/2} \Rightarrow \text{Se rechaza } H_0$$

2. Normal: Contraste para μ con σ desconocido

$$X_1, X_2, \ldots, X_n \to N(\mu, \sigma)$$

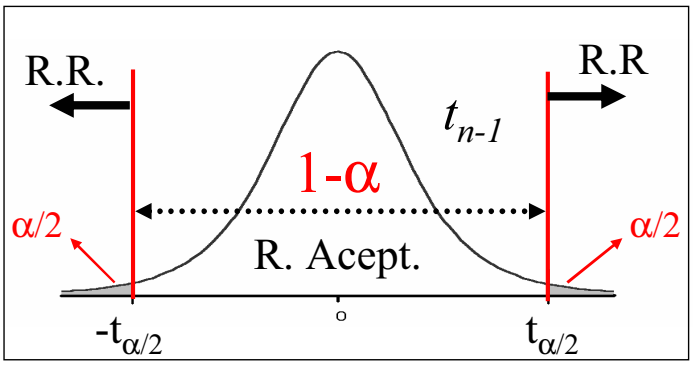
$$H_0: \quad \mu = \mu_0$$

$$H_1: \quad \mu \neq \mu_0$$

$$\overline{X} \to N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\overline{X} - \mu}{\hat{S} / \sqrt{n}} \to t_{n-1}$$

$$t_0 = \frac{\overline{x} - \mu_0}{\hat{s} / \sqrt{n}} \longrightarrow \boxed{\begin{array}{l} |t_0| \leq t_{\alpha/2} \Rightarrow \text{No se rechaza } H_0 \\ |t_0| > t_{\alpha/2} \Rightarrow \text{Se rechaza } H_0 \end{array}}$$

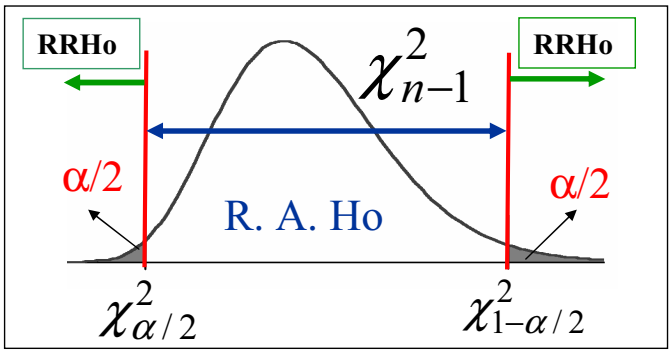


3. Normal: Contraste para σ²

$$X_1, X_2, \ldots, X_n \to N(\mu, \sigma)$$

$$\boxed{\begin{array}{l} H_0: \quad \sigma^2 = \sigma_0^2 \\ H_0: \quad \sigma^2 \neq \sigma_0^2 \end{array}}$$

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$



$$\frac{(n-1)\hat{S}^2}{\sigma^2} \to \chi^2_{n-1} \longrightarrow P(\chi^2_{\alpha/2} \leq \chi^2_{n-1} \leq \chi^2_{1-\alpha/2}) = 1 - \alpha$$

Datos: $\hat{s}^2$

$$\chi_0^2 = \frac{(n-1)\hat{s}^2}{\sigma_0^2} \longrightarrow \begin{cases} \chi_0^2 \in [\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}] & \Rightarrow \text{No rechazo } H_0 \\ \chi_0^2 \notin [\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}] & \Rightarrow \text{Rechazo } H_0 \end{cases}$$

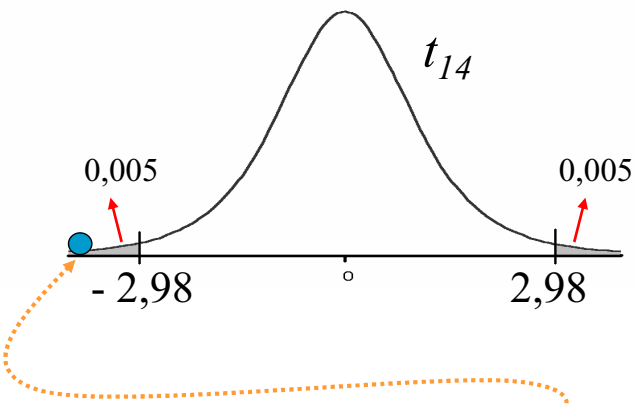
EJEMPLO 1. La resistencia a la compresión de 15 probetas de acero elegidas al azar es:

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

$$\left.\begin{array}{l} H_0 : \mu = 60 \\ H_1 : \mu \neq 60 \end{array}\right\} \quad \alpha = 0.01$$

$$\frac{\overline{X} - \mu}{\hat{S} / \sqrt{15}} \rightarrow t_{14}$$

$$P(-2{,}98 \leq t_{14} \leq 2{,}98) = 0.99$$



Datos:

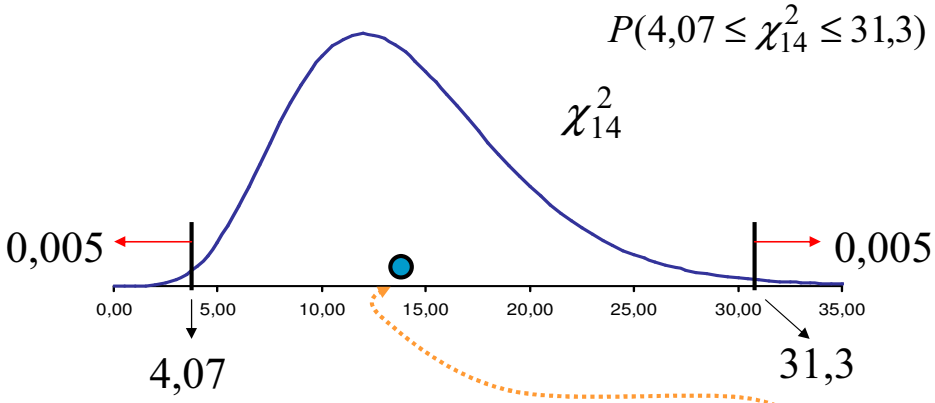
$$\overline{x} = 45{,}75 \quad \hat{s} = 14{,}2 \quad \Rightarrow \qquad t_0 = \frac{45{,}75 - 60}{14{,}2 / \sqrt{15}} = -3.88$$

Como 3.88 > 2.98 ⇒ Se **rechaza** H_0 con **α=0.01**

$$H_0 : \sigma^2 = 200 \qquad \frac{(n-1)\hat{S}^2}{\sigma^2} \rightarrow \chi^2_{n-1} \qquad \frac{14\hat{S}^2}{\sigma^2} \rightarrow \chi^2_{14}$$

$$H_1 : \sigma^2 \neq 200$$

$$P(4{,}07 \leq \chi^2_{14} \leq 31{,}3) = 0{,}99$$



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

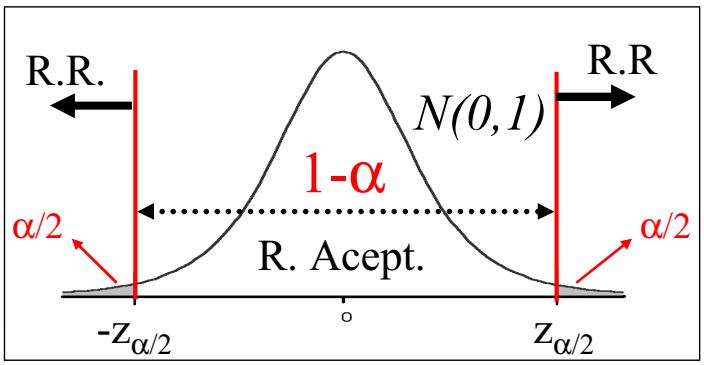
$$X_1, X_2, \ldots, X_n \rightarrow \text{Poisson}(\lambda)$$

$$\left.\begin{array}{ll} H_0: & \lambda = \lambda_0 \\ H_1: & \lambda \neq \lambda_0 \end{array}\right\rangle \quad \text{nivel de significación } \alpha$$

$$\hat{\lambda} = \overline{X} \overset{aprox}{\rightarrow} \mathrm{N}(\lambda, \sqrt{\frac{\lambda}{n}})$$

$$Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} \rightarrow \mathrm{N}(0,1)$$

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



$$z_0 = \frac{\hat{\lambda} - \lambda_0}{\sqrt{\frac{\lambda_0}{n}}} \rightarrow \boxed{\begin{array}{l} |z_0| \leq z_{\alpha/2} \Rightarrow \text{No se rechaza } H_0 \\ |z_0| > z_{\alpha/2} \Rightarrow \text{Se rechaza } H_0 \end{array}}$$

Dos tratamientos

Comparación de dos tratamientos

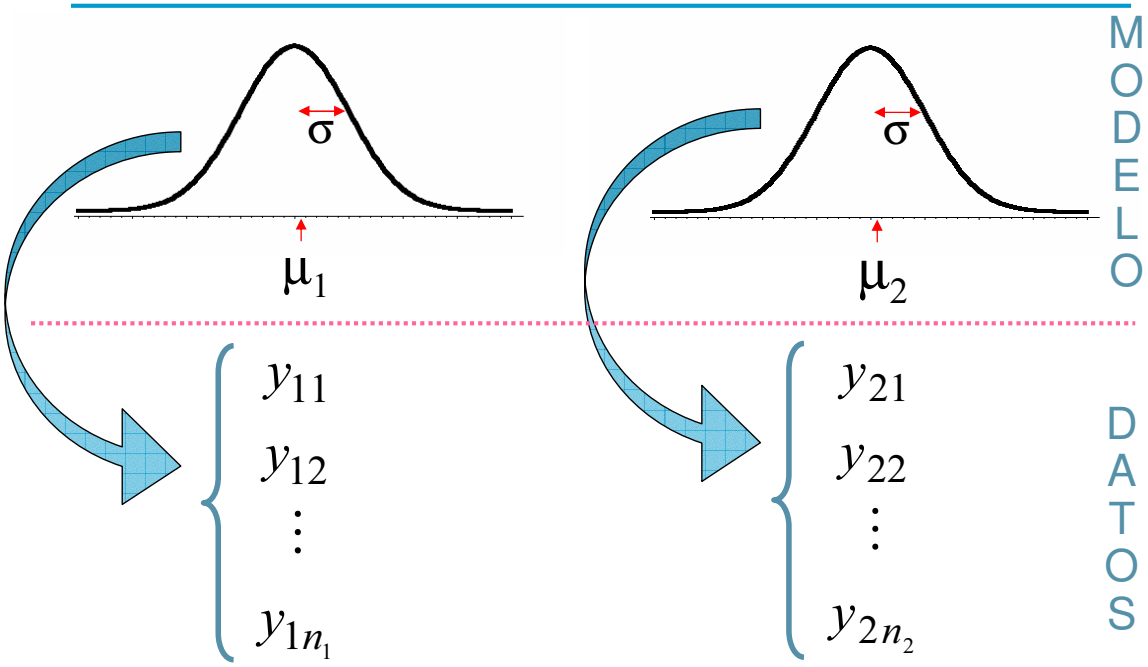
A	B
51,3	29,6
39,4	47,0
26,3	25,9
39,0	13,0
48,1	33,1
34,2	22,1
69,8	34,1
31,3	19,5
45,2	43,8
46,4	24,9

- 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**

  $Var\,[y_{ij}] = \sigma^2$

- **Independencia**

  $Cov\,[y_{ij}, y_{kl}] = 0$

Parámetros

$$\mu_1$$

$$\mu_2$$

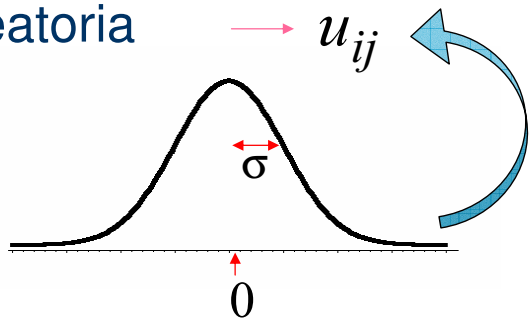
$$\sigma^2$$

Modelo

$$y_{ij} = \mu_i + u_{ij}, \qquad u_{ij} \rightarrow N(0, \sigma^2)$$

Las observaciones se descomponen en:

- Parte predecible $\rightarrow$ μ_i
- Parte aleatoria $\rightarrow$ u_{ij}



Estimación medias:

$$\mu_1 : \quad \rightarrow \quad \bar{y}_{1\bullet} = \frac{\sum_{j=1}^{n_1} y_{1j}}{n_1}$$

$$\mu_2 : \quad \rightarrow \quad \bar{y}_{2\bullet} = \frac{\sum_{j=1}^{n_2} y_{2j}}{n_2}$$

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)

$$y_{ij} = \mu_i + u_{ij}, \quad u_{ij} \to N(0, \sigma^2)$$

$$u_{ij} = y_{ij} - \mu_i$$

$$e_{ij} = y_{ij} - \bar{y}_{i\bullet}$$

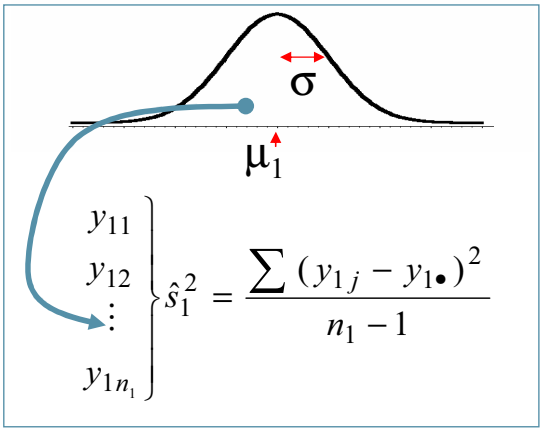
e_{ij} : **RESIDUO**

$$\sigma^2 : \to \hat{s}_R^2 = \frac{\sum_{i=1}^{2} \sum_{j=1}^{n_i} e_{ij}^2}{n-2}$$

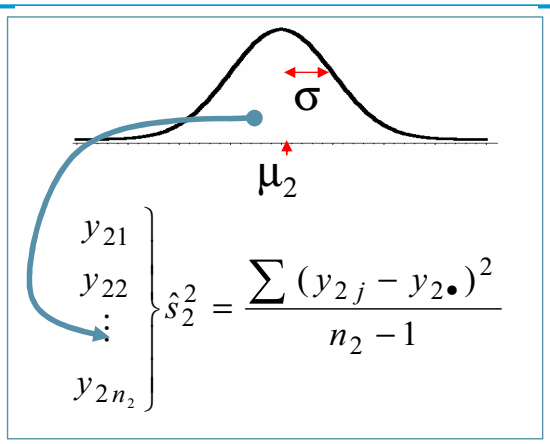
Residuos	
A	**B**
8,2	0,3
-3,7	17,7
-16,8	-3,4
-4,1	-16,3
5,0	3,8
-8,9	-7,2
26,7	4,8
-11,8	-9,8
2,1	14,5
3,3	-4,4
0,0	**0,0**

$$\sum_{j=1}^{n_i} e_{ij} = 0; \hat{s}_R^2 = 130.95$$

Varianza residual: $\hat{s}_R^2$



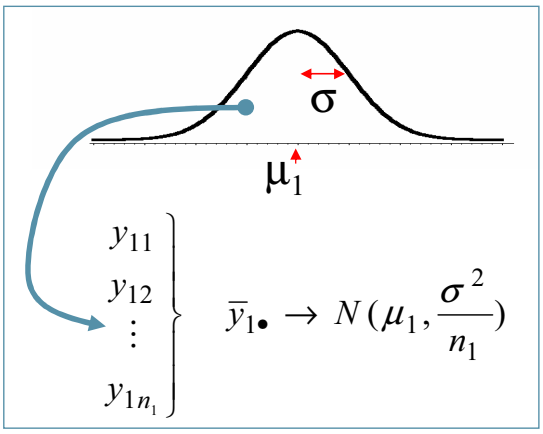
$$\hat{s}_1^2 = \frac{\sum (y_{1j} - y_{1\bullet})^2}{n_1 - 1}$$



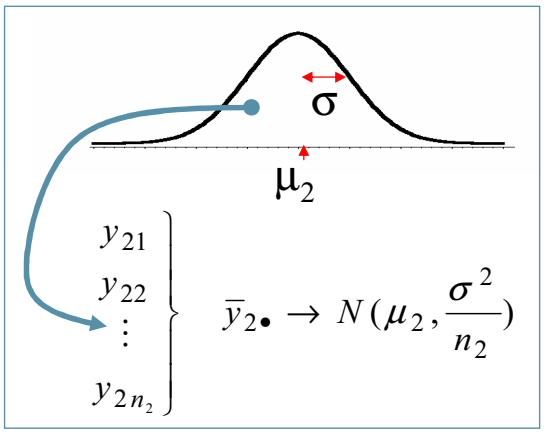
$$\hat{s}_2^2 = \frac{\sum (y_{2j} - y_{2\bullet})^2}{n_2 - 1}$$

$$\hat{s}_R^2 = \frac{\sum_{i=1}^{2} \sum_{j=1}^{n_i} e_{ij}^2}{n-2} = \frac{n_1 - 1}{n-1} \hat{s}_1^2 + \frac{n_2 - 1}{n-1} \hat{s}_2^2$$

Diferencia de medias: $\bar{y}_{1\bullet} - \bar{y}_{2\bullet}$



$$\bar{y}_{1\bullet} \to N(\mu_1, \frac{\sigma^2}{n_1})$$



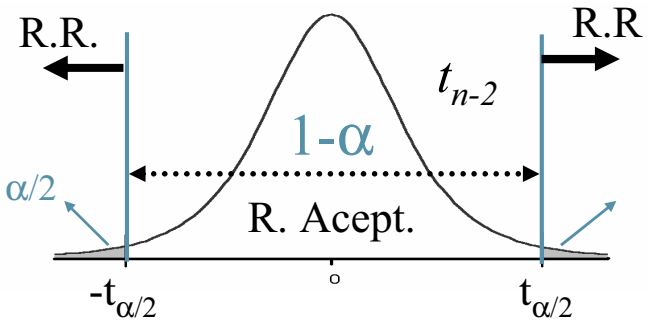
$$\bar{y}_{2\bullet} \to N(\mu_2, \frac{\sigma^2}{n_2})$$

$$\left.\begin{array}{l} \bar{y}_{1\bullet} - \bar{y}_{2\bullet} \to N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}) \\ \dfrac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to N(0,1) \end{array}\right\} \Rightarrow \boxed{\dfrac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to t_{n-2}}$$

Contraste de igualdad de medias

$$\boxed{\begin{array}{ll} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{array}}$$

$$t_0 = \frac{\bar{y}_{1\bullet} - \bar{y}_{2\bullet}}{\hat{s}_R\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to t_{n-2}$$



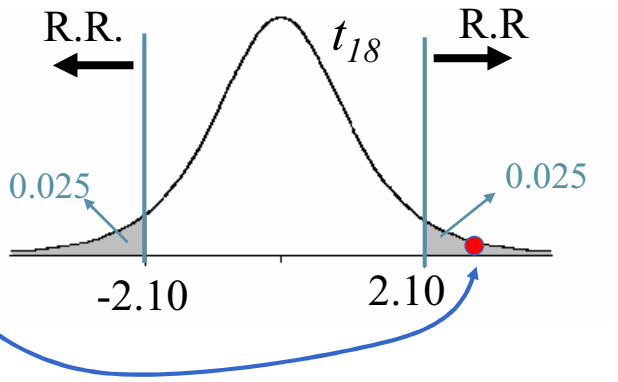
$$|t_0| \leq t_{\alpha/2} \Rightarrow \text{No se rechaza } H_0$$

$$|t_0| > t_{\alpha/2} \Rightarrow \text{Se rechaza } H_0$$

Ejemplo: α = 0.05

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{43.1 - 29.3}{11.44\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.69$$

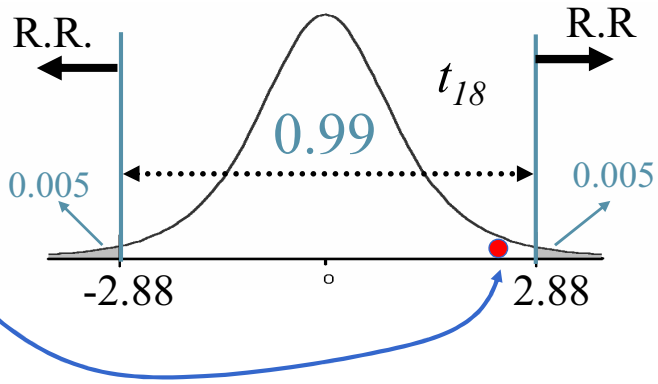


$$2.69 > 2.10 \Rightarrow Se \text{ rechaza } H_0$$

Ejemplo: α = 0.01

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{43.1 - 29.3}{11.44\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.69$$

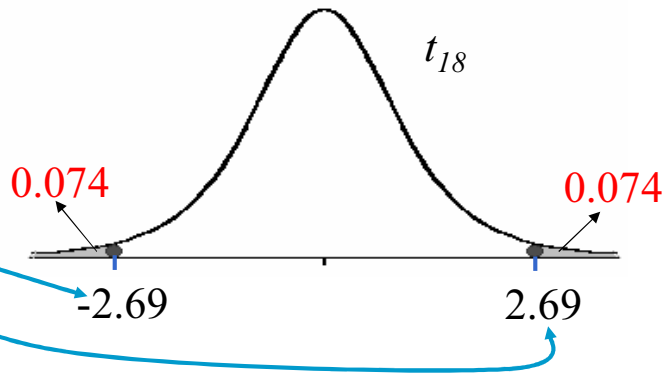


$$2.69 \leq 2.88 \Rightarrow \text{No se rechaza } H_0$$

Nivel crítico (bilateral)

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

$$t_0 = \frac{43.1 - 29.3}{11.44\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.69$$



$$p - valor = \Pr(|t_{18}| > 2.69) = 0.0147$$

- $\alpha = 0.05 > p\text{-}valor \Rightarrow$ 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.

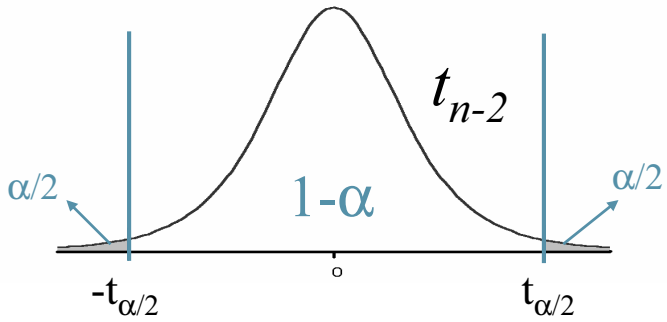
No rechazar H_o, no implica que H_o sea cierta

- 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: $\mu_1 - \mu_2$

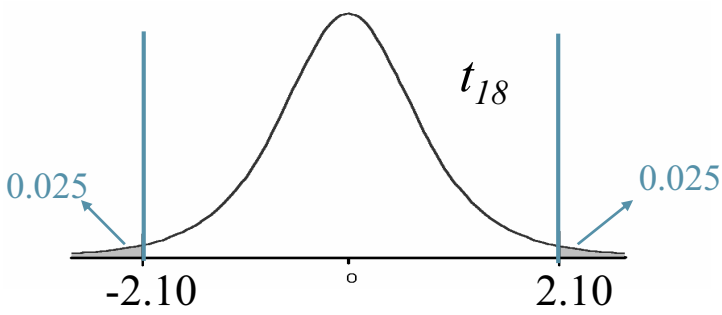
$$\frac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\hat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow t_{n-2}$$



$$\Pr\{-t_{\alpha/2} \leq \frac{(\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) - (\mu_1 - \mu_2)}{\hat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{\alpha/2}\} = 1 - \alpha$$

$$\mu_1 - \mu_2 \in (\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) \pm t_{\alpha/2} \hat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Ejemplo: intervalo de confianza $\mu_1 - \mu_2$

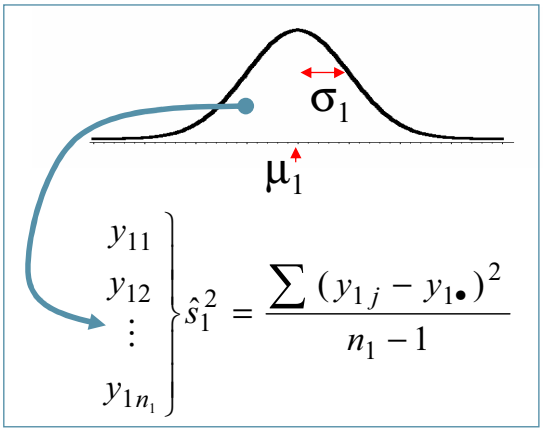


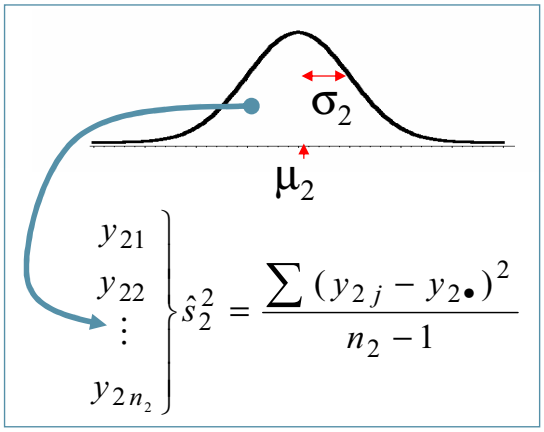
$$\mu_1 - \mu_2 \in (\bar{y}_{1\bullet} - \bar{y}_{2\bullet}) \pm t_{\alpha/2}\hat{s}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$\mu_1 - \mu_2 \in (43.1 - 29.3) \pm 2.10 \times 11.44 \times \sqrt{\frac{1}{10}+\frac{1}{10}}$$

$$\mu_1 - \mu_2 \in 13.8 \pm 10.74$$

Hipótesis de homocedasticidad





$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

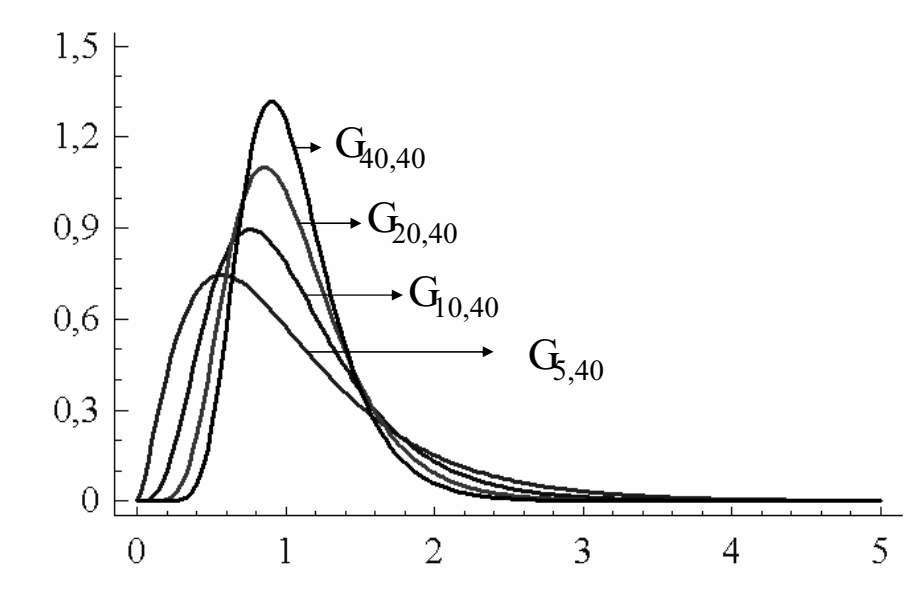
Distribución F

$$\left.\begin{matrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \end{matrix}\right\} \hat{s}_1^2 = \frac{\sum (y_{1j} - y_{1\bullet})^2}{n_1 - 1} \qquad \left.\begin{matrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{matrix}\right\} \hat{s}_2^2 = \frac{\sum (y_{2j} - y_{2\bullet})^2}{n_2 - 1}$$

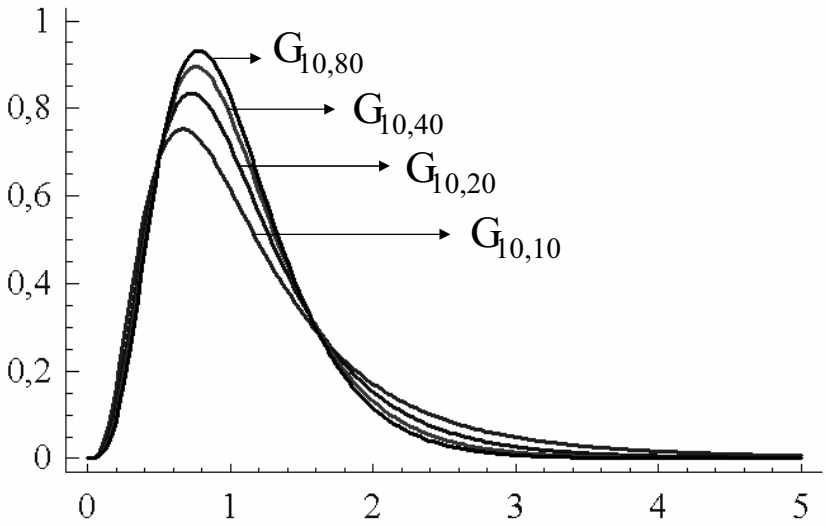
$$\frac{(n_1 - 1)\hat{s}_1^2}{\sigma_1^2} \to \chi^2_{n_1 - 1} \qquad \frac{(n_2 - 1)\hat{s}_2^2}{\sigma_2^2} \to \chi^2_{n_2 - 1}$$

$$\boxed{F = \frac{\chi^2_{n_1-1} \big/ (n_1 - 1)}{\chi^2_{n_2-1} \big/ (n_2 - 1)} = \frac{\dfrac{\hat{s}_1^2}{\sigma_1^2}}{\dfrac{\hat{s}_2^2}{\sigma_2^2}} \to F_{n_1-1, n_2-1}}$$

Distribución F



Algunas distribuciones F



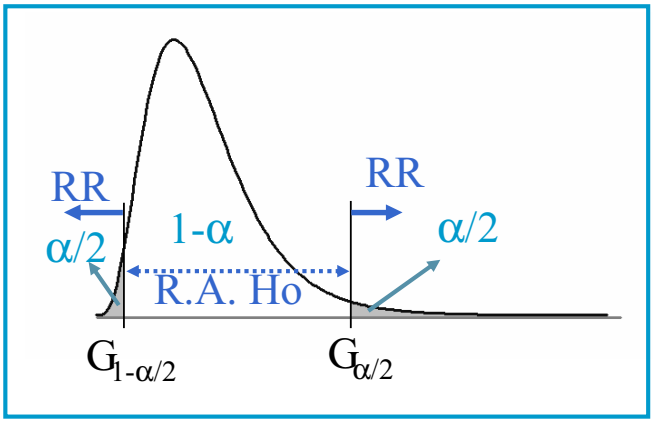
Contraste de igualdad de varianzas

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Si H_0 es cierto $\sigma_1^2 = \sigma_2^2$,

$$F_0 = \frac{\hat{s}_1^2}{\hat{s}_1^2} \rightarrow F_{n_1-1,n_2-1}$$



Si $F_0 \in [F_{1-\alpha/2}, F_{\alpha/2}] \Rightarrow$ No se rechaza H_0

Si $F_0 \notin [F_{1-\alpha/2}, F_{\alpha/2}] \Rightarrow$ Se rechaza H_0

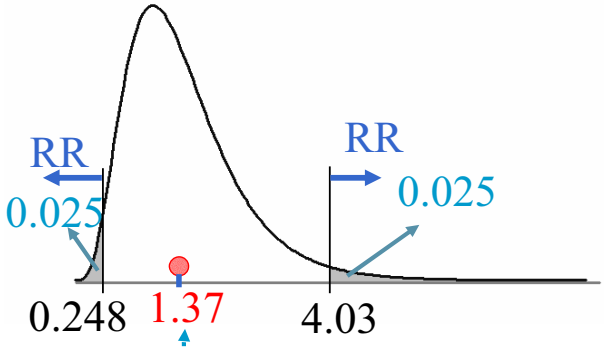
Ejemplo: Contraste de igualdad de varianzas

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$\hat{s}_1^2 = 154.02 \quad \hat{s}_2^2 = 111.7$$

$$F_0 = \frac{154.02}{111.7} = 1.37$$



$$1.37 \in [0.248, 4.03] \Rightarrow \text{No se rechaza } H_0$$

Tabla G

$$F_{\nu_1,\nu_2,\alpha} \quad \Rightarrow \quad P(F_{\nu_1,\nu_2} \geq F_{\nu_1,\nu_2,\alpha}) = \alpha$$

α=0.05

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.	

$$Ejemplo: \quad P(F_{7,8} \geq 3.50) = 0.05$$

Tabla G

α=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 G

α=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

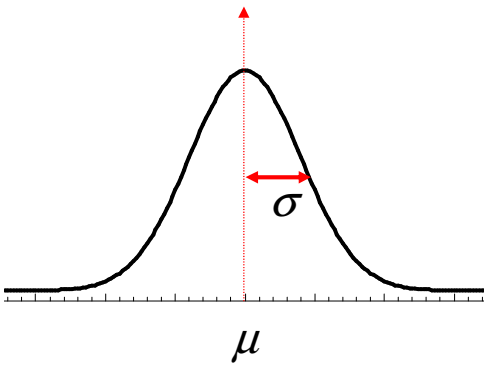
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	100	120	Inf.	
1	4052,2	4999,3	5403,5	5624,3	5764,0	5859,0	5928,3	5981,0	6022,4	6055,9	6106,7	6157,0	6208,7	6234,3	6260,4	6286,4	6313,0	6333,9	6339,5	6365,6	**1**
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,48	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.**	

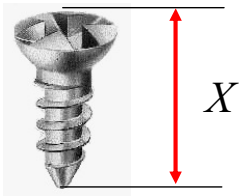
Ejemplo: $P(F_{7,8} \geq 6.18) = 0.01$

Contrastes de bondad de ajuste

Contraste χ^2 de bondad de ajuste

POBLACIÓN





MUESTRA n

$$X_1, X_2, \ldots, X_n$$

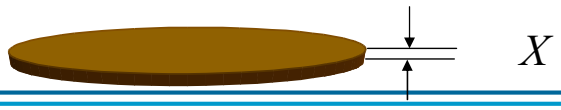
$$¿\, X \rightarrow N(\mu, \sigma)\, ?$$

$$\hat{\mu} = \bar{x}$$
$$\hat{\sigma} = \hat{s}$$

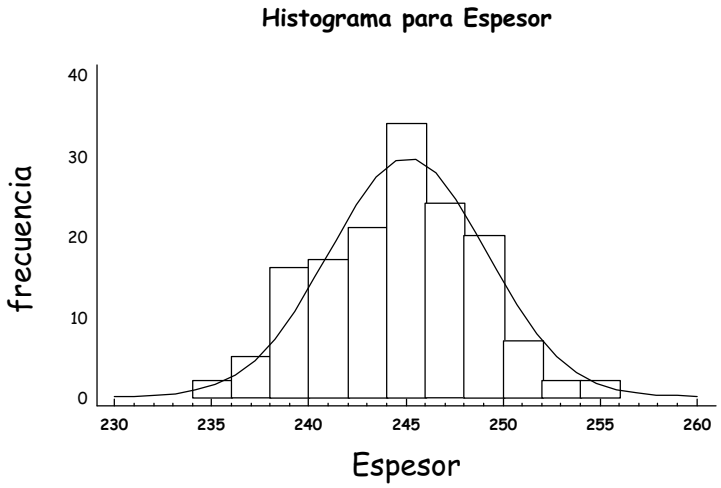
Datos Conocidos

¿Tienen los datos distribución Normal?

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



Contraste χ^2 de bondad de ajuste

$$H_0 : X_i, \forall i = 1,2,\ldots,n \rightarrow f_X$$
$$H_1 : X_i, \forall i = 1,2,\ldots,n \not\rightarrow f_X$$

Si H_0 es cierto:

Clases	Gr.Observada	Gr.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

$$\boxed{\begin{aligned} E_k &= np_k \\ p_k &= \Pr(c_{k-1} \le X < c_k) \end{aligned}}$$

$$\sum_{k=1}^{K} O_k = \sum_{k=1}^{K} E_k = n$$

$$\boxed{\sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \rightarrow \chi^2_{K-r-1}}$$

Justificación del contraste χ^2

$$O_k \rightarrow B(n, p_k) \xrightarrow{n \rightarrow \infty} \boxed{O_k \rightarrow N(np_k, \sqrt{np_k(1-p_k)})}$$

$$E_k = np_k$$

$$\frac{O_k - np_k}{\sqrt{np_k \underbrace{(1-p_k)}_{\cong 1}}} \rightarrow N(0,1)$$

$$\frac{O_k - E_k}{\sqrt{E_k}} \rightarrow N(0,1) \Rightarrow \sum_{k=1}^{K} \left(\frac{O_k - E_k}{\sqrt{E_k}} \right)^2 \xrightarrow{aprox} \chi^2_{K-r-1}$$

K ≡ Nº de CLASES ***r*** ≡ 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

$\bar{x} = 245.1$ micras

$\hat{s} = 4$ 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

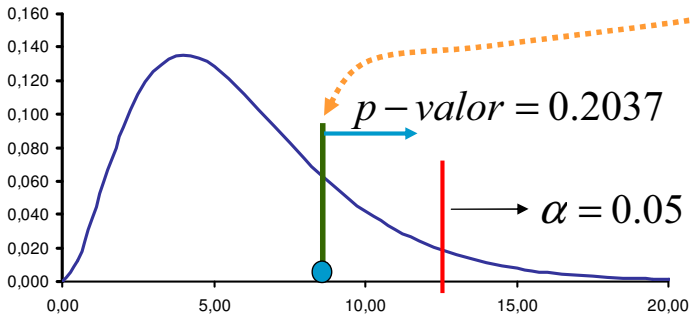
$$p_4 = \Pr(242 \le X \le 244)$$
$$= \Pr\left(\frac{242-245.1}{4} \le \frac{X-245.1}{4} \le \frac{244-245.1}{4}\right)$$
$$= P(-0.275 \le Z \le -0.775)$$
$$= 0.1714$$
$$E_4 = np_4 = 150 \times 0.1714 = 25.71$$

Contraste de Normalidad

$$H_0 : X_i \rightarrow \text{Normal}$$

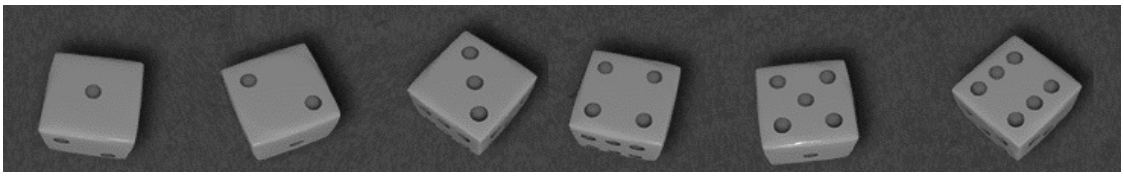
$$H_1 : X_i \not\rightarrow \text{Normal}$$

$$\chi_0^2 = \frac{(7-5.88)^2}{5.88} + \frac{(16-9{,}58)^2}{9.58} + \cdots + \frac{(4-6.54)^2}{6.54} = 8.5$$



$$\Pr(\chi_6^2 \leq 12.59) = 0.95$$

$$\chi_0^2 = 8.5 < 12.59 \Rightarrow \text{No se rechaza la hipótesis de normalidad}$$



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 α=0.05 que el dado está desequilibrado?

$$\Pr(X=i)=1/6, \quad \forall i=1,2,\ldots,6$$

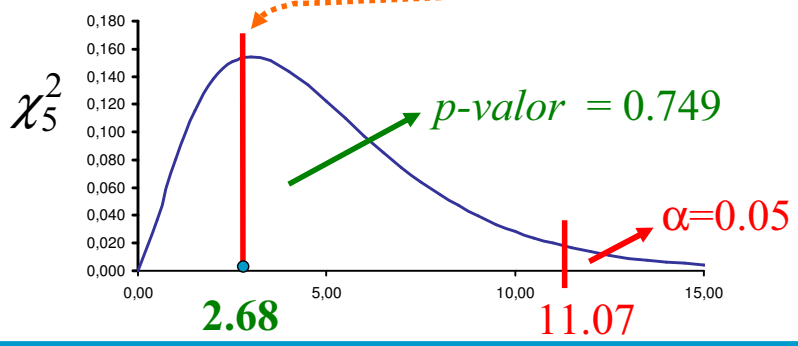
$$H_o : X \rightarrow \text{Equiprobable}$$

$$\mathrm{H}_1 : X \rightarrow \text{No equiprobable}$$

Resultados	Observados	Esperados
1	49	50
2	59	50
3	49	50
4	51	50
5	43	50
6	49	50
Total	300	300

$$\chi_0^2 = \frac{(49-50)^2}{50} + \frac{(59-50)^2}{50} + \cdots + \frac{(49-50)^2}{50} = 2.68$$

$$\Pr(\chi_5^2 \leq 11.07) = 0.95 \Rightarrow \quad \chi_0^2 = 2.68 < 11.07 \Rightarrow \text{No se Rechaza } \mathrm{H}_0$$



Contraste *K-S* de bondad de ajuste

- Válido únicamente para variables continuas
- Compara la Función de distribución teórica F(x) con la Función de distribución empírica $F_n(x)$.
- No requiere agrupar los datos
- El contraste es aproximado si es necesario estimar los parámetros del modelo

Contraste *K-S* de bondad de ajuste

- Pasos:
  - Ordenar la muestra $x(1) \leq x(2) \leq x(3) \leq \ldots \leq x(n)$
  - Calcular la distribución empírica de la muestra $F_n(x)$

$$F_n(x) = \begin{cases} 0 & x < x(1) \\ \frac{r}{n} & x(r) \leq x < x(r+1) \\ 1 & x \geq x(n) \end{cases}$$

  - Calcular la discrepancia máxima

$$max\{Dn(x)\} = max\{|F_n(x) - F(x)|\}$$

  - Comparar con el valor obtenido en las tablas para un *n* y *α* determinados

Contraste *K-S* de bondad de ajuste

Se han tomado 12 valores de una variable física X, que se supone Normal, resultando
30,2 30,8 29,3 29,0 30,9 30,8 29,7 28,9 30,5 31,2 31,3 28,5
Contraste la hipótesis de normalidad mediante el contraste de K-S.

$$H_0 : X_i \rightarrow \text{Normal}$$

$$H_1 : X_i \not\rightarrow \text{Normal}$$

De la muestra $\Rightarrow$ $\overline{x} = 30,09$, $\hat{s} = 0,97$

Contraste *K-S* de bondad de ajuste

X	Fn(x)	F(x)	Fn(xh-1)-F(xh)	Fn(xh)-F(xh)	Dn(xh)
28,5	1/12=0,08	0,05	0,05	0,03	0,05
28,9	2/12=0,17	0,11	0,03	0,06	0,06
29	3/12=0,25	0,13	0,04	0,12	0,12
29,3	4/12=0,33	0,21	0,04	0,12	0,12
29,7	5/12=0,42	0,34	0,01	0,08	0,08
30,2	6/12=0,5	0,54	0,12	0,04	0,12
30,5	7/12=0,58	0,66	0,16	0,08	0,16
30,8					
30,8	9/12=0,75	0,77	0,19	0,02	**0,19**
30,9	10/12=0,83	0,8	0,05	0,03	0,05
31,2	11/12=0,92	0,87	0,04	0,05	0,05

$$\max\{Dn(x_h)\} = 0,19$$

$$F(29,7) = P(X \leq 29,7 | X \to N(30,09;0,97)) =$$
$$= P\left(\frac{X-30,09}{0,97} \leq \frac{29,7-30,09}{0,97}\right) =$$
$$= P(Z \leq -0,40) = 0,34.$$

Valores Críticos de *D=|Fn(x)-F(x)|* donde *Fn(x)* es la distribución muestral de tamaño *n* y *F(x)* la distribución teórica

Ejemplo

Tabla Contraste de Kolmogorov-Smirnov

Tamaño muestral	Nivel de significación				
n	**0,20**	**0,15**	**0,10**	**0,05**	**0,01**
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,362	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
>35	1,07/n½	1,14/n½	1,22/n½	1,36/n½	1,63/n½

$$\max\{Dn(x_h)\} = 0,19 \leq D_{12;0,05} = 0,375$$

No se puede rechazar la hipótesis de normalidad con un nivel de significación de 0,05

Ejercicios propuestos

Capítulo 7. Contraste de Hipótesis

7.1 Un proceso industrial fabrica piezas cuya longitud en mm se distribuye según una $N(190, 10)$. Una muestra de 5 piezas proporciona los resultados siguientes:

187, 212, 195, 208, 192

(a) Contrastar la hipótesis de que la media del proceso μ es efectivamente 190.

(b) Contrastar la hipótesis de que la varianza del proceso σ^2 es 100

(c) Admitiendo que la varianza del proceso es 100, construir la curva de potencia para un contraste de la media con 5 observaciones. Tómese $\alpha = 0.05$ en todos los contrastes.

7.2 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?.

7.3 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 $\alpha = 0.05$).

7.4 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. ¿Se puede concluir que la proporción de cojinetes defectuosos ha aumentado? 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 $\alpha = 0,05$.).

7.5 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 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, $H_0 : \lambda = 5 \times 10^3$ horas, frente a $H_1 : \lambda > 5 \times 10^3$ horas; indicando: (a) el valor crítico, y (b) la probabilidad de error tipo II cuando $\lambda = 7 \times 10^3$ horas. (Es suficiente con proporcionar el valor más proximo obtenido en las tablas del libro de texto).

(b) Se va a realizar un ensayo con 15 equipos fabricados por una segunda empresa. Si el tiempo de funcionamiento de estos tiene también distribución exponencial. ¿ Cuál es el valor máximo de la media muestral de estos quince equipos que permitiría concluir con $\alpha = 0.05$ que son peores que los de la primera empresa? Después de 6000 horas de ensayo han fallado 6 equipos, siendo el promedio de estos seis valores igual a 2350 horas. ¿Es necesario seguir el ensayo para tomar una decisión ?

7.6 En la tabla se reproducen los valores para la velocidad de la luz en el aire que obtuvieron Michelson y Newcomb en 1879 y 1992 respectivamente, utilizando métodos distintos . A los datos originales en km/s se les ha restado 299000 para facilitar su tratamiento matemático.

1.

	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	119,47
New.	578	796	774	820	772	763,2	111,91

(a) Contraste la hipótesis de que la variabilidad es la misma en los experimentos de Michelson y Newcomb. ($\alpha = 0,05$)

(b) Sabiendo que la velocidad de la luz en el vacio es 299.792,5 km/s (792,5 si se le resta 299.000), y que en aire debe ser igual o menor que en el vacio, contraste la hipótesis de que los experimentos no tienen error sistemático, es decir la medida obtenida es igual al verdadero valor más un error aleatorio de media cero.($\alpha = 0,05$)

7.7 Se estudian dos tipos de neumáticos con los resultados siguientes:

Tipo	n_i	$\overline{x}_i(Km)$	$\widehat{s}_i(Km)$
A	121	27465	2500
B	121	27572	3000

1. Calcular, con $\alpha = 0.01$,
2. Un intervalo de confianza para $\dfrac{\sigma_1^2}{\sigma_2^2}$.
3. Un intervalo de confianza para $\mu_1 - \mu_2$.

7.8 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 reultados para la B son: 103.3; 121.7; 98.4; 161.5; 167.8 y 67.3. ¿Son las máquinas iguales? (Suponer que los rendimientos de ambas máquinas siguen distribuciones normales).

7.9 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\%

7.10 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 ($\alpha = 0.05$) ?

7.11 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

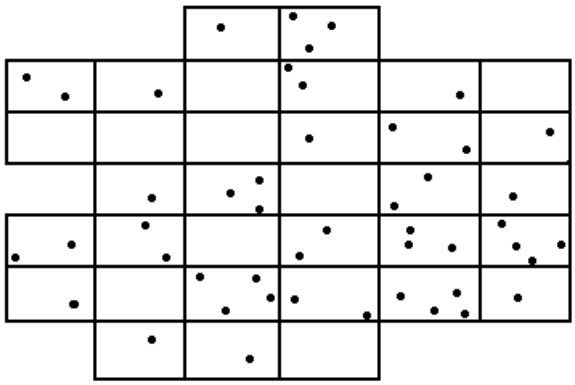
Contrastar la hipótesis de que el dado está equilibrado y que, por tanto, sus caras son equiprobables. (Tómese $\alpha = 0.05$).

7.12 Se han tomado 12 valores de una variable física X, que se supone 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.

Analizar la hipótesis de normalidad mediante el contraste de Kolmogorov-Smirnov.

7.13 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 nivel crítico aproximado del contraste.



7.14 El Ministerio de defensa está considerando un nuevo sistema de apoyo para el lanzamineto de misiles de corto alcance. El sistema existente tiene errores en el 7% de los lanzamientos y se desea comprobar si el nuevo sistema tiene una probabilidad de fallo menor. El ensayo va a consistir en realizar 20 lanzamientos y se concluirá que el nuevo sistema es mejor si no se produce ningún fallo. Llamando p a la probabilidad de fallo del sistema nuevo y aceptando independencia entre los resultados del lanzamiento, obtenga y represente gráficamente la probabilidad de error de tipo II del contraste

$$\begin{cases} H_0 : p = 0.07 \\ H_1 : p < 0.07 \end{cases}$$

Obtenga la probabilidad de error tipo I. Interprete el resultado y valore si el método de decisión es adecuado.

7.15 El tiempo de duración T de un componente electrodinámico es una variable aleatoria con distribución exponencial de media μ. Veinte componentes han sido sometidos a un ensayo y el número de horas que han durado ha sido:

10.99	15.79	24.14	34.43	43.72	51.72	56.12	60.27	77.20	88.47
91.07	117.58	130.40	133.12	152.90	159.00	193.62	208.71	308.82	316.07

Teniendo en cuenta que $2T/\mu$ tiene distribución χ^2 con dos grados de libertad, realiza el siguiente contraste

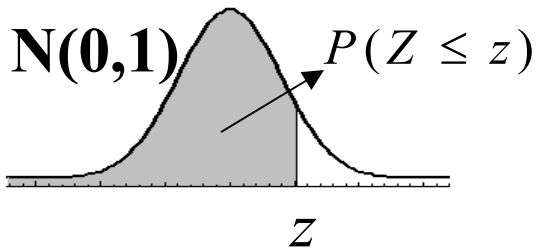
$$\begin{aligned} H_0 &: \mu = 200 \text{ horas}, \\ H_1 &: \mu < 200 \text{ horas}, \end{aligned}$$

con $\alpha = 0.05$.

Tablas

TABLA

Normal Estandar

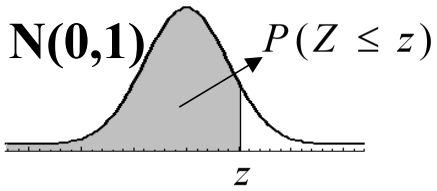


Ejemplo.

$P(Z \leq 1.96) = 0.9750$

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

Distribución *normal estándar* (continuación)



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

Relación entre Binomial, Poisson y Normal

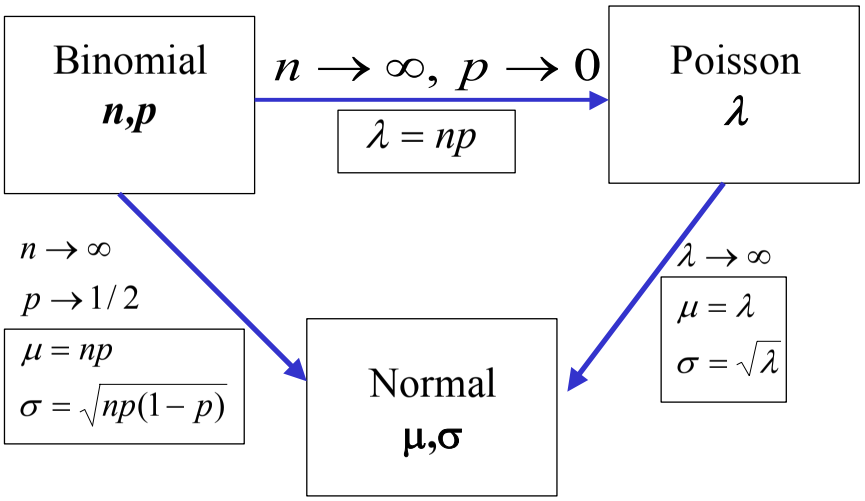
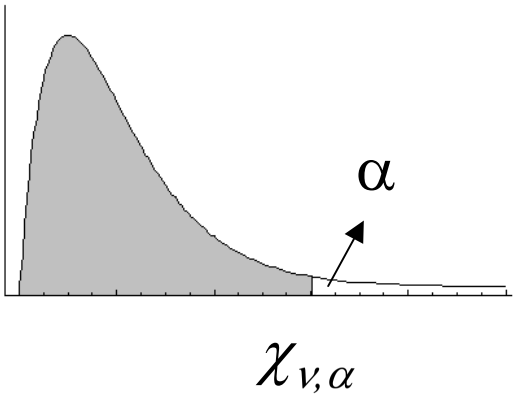


Tabla χ^2



ν: grados de libertad (g.l.)

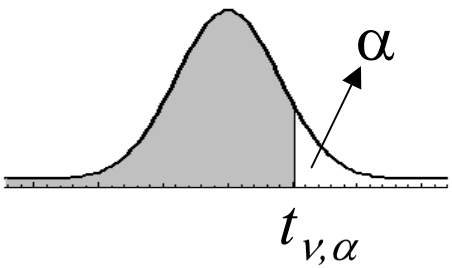
EJEMPLO

$P(\chi_9 \geq 19{,}02) = 0{,}025$

	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

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 F

α=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

α=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

α=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

	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.**	

Ejemplo : $P(F_{7,8} \geq 6.18) = 0.01$