ESTADÍSTICA I (ADE): TEORÍA Y EJERCICIOS

Isabel Maqueda de Anta

Carme Muñoz Vaquer

Elizabeth Torrelles Puig

Núria Viladomiu Canela

PRESENTACIÓN

Esta publicación electrónica va dirigida a los alumnos de Estadística I del grado de Administración y Dirección de Empresas de la Universidad de Barcelona.

Su principal objetivo es proporcionar un material de trabajo que facilite la dinámica de las clases presenciales y ayude al alumno en su planificación del estudio.

Cada capítulo de la publicación corresponde a un tema del programa. Se inicia con un breve resumen de los conceptos básicos e imprescindibles. A continuación se proporciona un conjunto de ejercicios que permitirán trabajar los conceptos de la asignatura.

Estos ejercicios están ordenados siguiendo el orden utilizado en la programación de las sesiones presenciales y también el grado de dificultad de los ejercicios.

Este material es fruto de años de experiencia práctica de las autoras en la asignatura Estadística y se ha constatado que proporciona a los alumnos un instrumento útil para el seguimiento de la asignatura.

Barcelona, julio 2012

Tema 1. CONCEPTO Y CONTENIDO DE LA ESTADÍSTICA

1.1 Objecto de la estadística
1.2 Estadística descriptiva e inferencia estadística
1.3 Población y muestra
1.4 Datos. Clasificación y escalas de medida
1.5 Instalación del programa R-Commander

Tema 2. DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA

2.1 Tabla de frecuencias simple: variable discreta
2.2 Diagrama de barras y de frecuencias acumuladas
2.3 Tabla de frecuencias con valores agrupados: variable continua
2.4 Histograma y polígonos de frecuencias
2.5 Análisis exploratorio de datos: diagrama de tallo y hojas (Stem and Leaf)

TIPO DE DATOS

Los DATOS son el conjunto de observaciones de una o más características obtenidas de una población o de una muestra.
Es importante distinguir entre los distintos tipos de datos con los que podemos tratar. Sus diferencias determinan la selección y aplicación de las técnicas estadísticas.

Según la naturaleza del fenómeno	CUALITATIVOS Atributos	NOMINALES
		ORDINALES
	CUANTITATIVOS Variables	DISCRETAS
		CONTINUAS
Según la periodicidad	Corte Transversal	
	Series Temporales	
	Datos de Panel	
Según el número de características observadas conjuntamente	Unidimensionales	
	Multidimensionales (Bidimensionales)	

REPRESENTACIONES GRÁFICAS

TABULACIÓN Y REPRESENTACIÓN GRÁFICA

CUANTITATIVOS

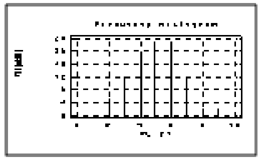
- EXPLORATORIO
  - STEM AND LEAF

```
X Stem-and-Leaf Plot

Frequency    Stem &  Leaf

    6,00        4 .  777999
    7,00        5 .  2244444
    9,00        5 .  666668888
    6,00        6 .  000222
    9,00        6 .  555558888
    3,00        7 .  001

Stem width:         10
Each leaf:          1 case(s)
```

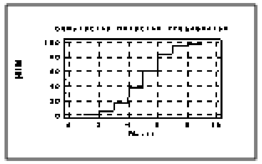
- DISTR. SIMPLES

Xi	ni	fi	Ni	Fi
x1	n1	f1	N1	F1
x2	n2	f2	N2	F2
...	...	...	...	...
xi	ni	fi	Ni	Fi
...	...	...	...	...
xk	nk	fk	N	1
	N	1		

  - BARRAS



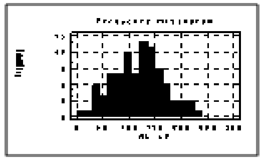
  - DIAGRAMA EN ESCALERA



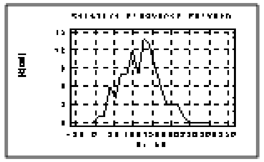
- DISTR. AGRUPADAS

Li-1-Li	ci	ni	fi	Ni	Fi
L0-L1	c1	n1	f1	N1	F1
L1-L2	c2	n2	f2	N2	F2
...	...	...	...	...	...
Li-1-Li	ci	ni	fi	Ni	Fi
...	...	...	...	...	...
Lk-1-Lk	ck	nk	fk	N	1
		N	1		

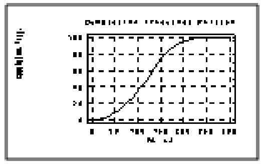
  - HISTOGRAMA



  - POLÍGONO DE FREC SIMPLE

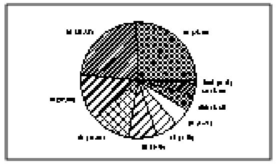


  - POLÍGONO DE FREC ACUMULADAS

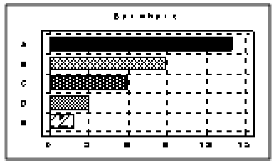


CUALITATIVOS

- DIAG. DE SECTORES



- DIAG. COLUMNAS O BARRAS



- CARTOGRAMAS
- PICTOGRAMAS

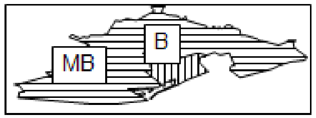


TABLA DE FRECUENCIAS

Recoge de forma resumida el conjunto de datos resultantes de la observación de una variable en un colectivo o muestra de n individuos.

ELEMENTOS DE UNA TABLA DE FRECUENCIAS

1. Tabla de frecuencias con los valores de la variable sin agrupar:

Xi	ni	fi	Ni	Fi
x1	n1	f1	N1	F1
x2	n2	f2	N2	F2
….	…	…	…	…
xi	ni	fi	Ni	Fi
…	…	…	…	…
xk	nk	fk	Nk=n	Fk=1
	n	1		

Interpretación de las columnas de la tabla

- Xi: recoge cada uno de los valores observados de X, ordenados de menor a mayor.
- ni: frecuencias absolutas; siendo ni el número de elementos de la muestra para los que X = xi.
  La suma de todas las frecuencias absolutas es igual a n.
- fi: frecuencias relativas; siendo fi = ni/n la proporción en tanto por uno de elementos para los que X = xi.
  La suma de todas las frecuencias relativas es igual a 1. Si se multiplican las frecuencias relativas por 100 se obtienen los correspondientes porcentajes.
- Ni: frecuencias absolutas acumuladas. Ni es el número de elementos para los que X ≤ xi.
  Ni = n1 + n2 + … + ni-1 + ni.
- Fi: frecuencias relativas acumuladas. Fi es la proporción de elementos para los que X ≤ xi.
  Fi = f1 + f2 + … + fi-1 + fi.
  Si las frecuencias relativas acumuladas se multiplican por 100 se obtiene los porcentajes acumulados.

2. Tabla de frecuencias con los valores de la variable agrupados en intervalos.

Li-1-Li	Xi (ci)	ni	fi	Ni	Fi
L0-L1 L1-L2 …. Li-1-Li Lk-1-Lk	x1 x2 …. xi … xk	ni n2 …. ni .... nk	f1 f2 .... fi .... fk	N1 N2 …. Ni .... Nk=n	F1 F2 …. Fi .... Fk=1
		n	1		

Interpretación de las columnas de la tabla

- Li-1-Li: recoge todos los intervalos o clases en los que se agrupan los valores de la variable; Li-1 y Li son los límites inferior y superior del intervalo i-esimo. Los intervalos, por omisión, se establecen abiertos en el límite inferior y cerrados en el superior.
  La amplitud del intervalo i-ésimo es ai = Li – Li-i.
  La marca de clase o punto medio del intervalo es el valor que representa al intervalo en el análisis descriptivo, siendo ci = (Li-1+Li)/2.
- ni: frecuencias absolutas; siendo ni el número total de elementos para los que el valor de X está dentro del intervalo i-ésimo.
- fi: frecuencias relativas; siendo fi = ni/n la proporción en tanto por uno de elementos para los que X está dentro del intervalo i-ésimo.
  La suma de todas las frecuencias relativas es igual a 1. Si se multiplican las frecuencias relativas por 100 se obtienen los correspondientes porcentajes.
- Ni: frecuencias absolutas acumuladas. Ni es el número de elementos para los que X ≤ Li.
  Ni = n1 + n2 + … + ni-1 + ni.
- Fi: frecuencias relativas acumuladas. Fi es la proporción de elementos para los que X ≤ Li
  Fi = f1 + f2 + … + fi-1 + fi.
  Si las frecuencias relativas acumuladas se multiplican por 100 se obtiene los porcentajes acumulados

EJERCICIOS TEMA 2

Ejercicio 1. La distribución del número de trabajadores en una muestra de gestorías es:

Determine:

a) Número de gestorías encuestadas.
b) Porcentaje de gestorías con 5 trabajadores.
c) Nº de gestorías con un máximo de 3 trabajadores.
d) Nº máximo y mínimo de trabajadores.
e) Nº más frecuente de trabajadores.
f) Nº máximo de trabajadores que tienen las 30 gestorías con menos personal.
g) Si una empresa de software únicamente está interesada en enviar propaganda a las gestorías con más de 6 empleados, ¿a qué porcentaje de las gestorías muestreadas se dirigirá?
h) Si la empresa de software está interesada en enviar propaganda al 25% de las gestorías con mayor empleo, ¿cuál es el número mínimo de empleados que debe tener una gestoría para poder estar incluida en este grupo?
i) Si el INEM se propone ayudar al 25% de las gestorías con menor empleo enviando un trabajador en prácticas, ¿cuántos empleados como máximo deberán tener para poderse beneficiar de dicha ayuda?
j) Represente gráficamente la distribución de frecuencias y compruebe las respuetas anteriores en los gráficos.

Ejercicio 2. Los siguientes datos corresponden al número de bibliotecas públicas por 1000 habitantes en las 41 comarcas de Catalunya:

25, 9, 12, 31, 57, 13, 16, 14, 22, 13, 6, 11, 12, 15, 27, 42, 9, 36, 22, 21,
7, 13, 25, 19, 16, 64, 33, 16, 43, 37, 23, 37, 19, 11, 43, 14, 49, 28, 51, 8, 9

a) Obtenga la distribución de frecuencias agrupando los datos en intervalos de amplitud 10, fijando el extremo inferior de la primera clase en 5.
b) Represente gráficamente la distribución obtenida en el apartado anterior.

Ejercicio 3. La tabla siguiente recoge la distribución de la variable X='Superficie en m^2' del conjunto de viviendas registradas en el Barcelonès con un máximo de 210 m^2:

Superficie (m^2)	Nº viviendas
(0 -30]	4050
(30 - 60]	153 900
(60 - 90]	437 400
(90 - 120]	162 000
(120 - 150]	29 160
(150 - 180]	12 150
(180 - 210]	11 340
Total	**810 000**

Con esta información se pide:

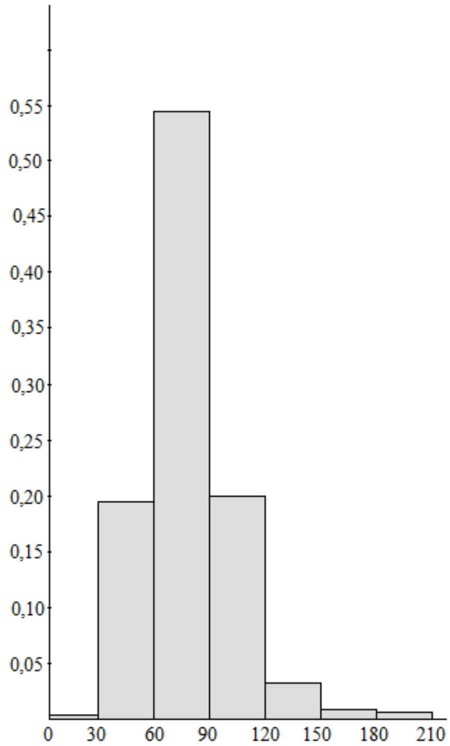
a) Complete la tabla de distribución de frecuencias.
b) Represente gráficamente la distribución de frecuencias (simples y acumuladas).
c) ¿Cuántas viviendas con un máximo de 210 m^2 hay censadas en el Barcelonès?
d) ¿Cuántas viviendas tienen superficies entre 60 y 90 m^2?
e) ¿Cuántas viviendas tienen como máximo 60 m^2?
f) ¿Cuántas viviendas superan los 150 m^2?
g) ¿Qué porcentaje de viviendas tiene entre 60 y 90 m^2? ¿Y como máximo 60 m^2?
h) ¿Puede decirse que más del 20% de estas viviendas tienen superficies entre 90 y 120 m^2?
i) ¿Cuál es la supercicie máxima del 73,5% de las viviendas con menor superficie?
j) ¿Puede decirse que algo más de la cuarta parte de estas viviendas superan los 90 m^2? Exactamente cuantas (en términos absolutos y en términos relativos).
k) ¿Cuál es el intervalo de superficie más frecuente?
l) Si una medida de tipo fiscal quiere abarcar al 93,5% de las viviendas con menor superficie, ¿cuál es la superficie máxima a considerar por dicha medida?
m) Si una inmobiliaria está interesada sólo en las viviendas de mayor superficie y quiere conectar con un 2,9% de las viviendas registradas, ¿cuál es la superficie mínima a considerar?
n) Una vivienda con 135 m^2 ¿está, aproximadamente, entre el 5% de las registradas con mayor superficie?

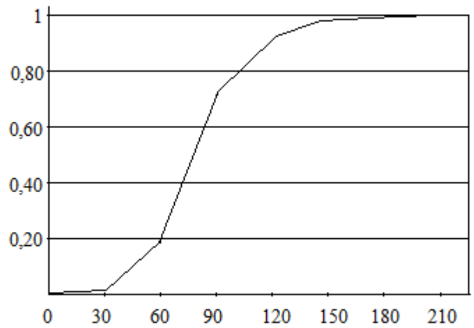
Resolución

a) Distribución de frecuencias absolutas, relativas y acumuladas.

Superficie (m^2)	n_i	N_i	f_i	F_i
(0-30]	4050	4050	0,005	0,005
(30-60]	153900	157950	0,190	0,195
(60-90]	437400	595350	0,540	0,735
(90-120]	162000	757350	0,200	0,935
(120-150]	29160	786510	0,036	0,971
(150-180]	12150	898660	0,015	0,986
(180-210]	11340	810000	0,014	1
Total	810000		1	

b) Histograma de frecuencias relativas y polígono de frecuencias relativas acumuladas





c) n = 810000 viviendas
d) N(60-90] = 437400
e) N(60) = 157950
f) n(X > 150) = 12150 + 11340 = 23490
g) f(60-90) = 0,540 el 54% F(60) = 0,195 el 19,5%
h) f(90-120) = 0,20 El porcentaje de viviendas con superficie entre 90 y 120m^2 no es superior al 20%
i) F(90) = 0,735 La superficie máxima del 73,5% de las viviendas con menor superficie es 90 m^2

j) $f(X > 90) = 1 - 0,735 = 0,265$ Más del 25% de las viviendas tienen más de 90 m^2, exactamente el 26,5%, lo que representa un total de 810000 (0,265) = 214650 viviendas
k) La frecuencia absoluta máxima es 437400; el intervalo más frecuente es (60-90]
l) $F(120) = 0,935$ La medida afectará a las viviendas que tienen como máximo 120 m^2
m) $f(X > 150) = 0,015 + 0,014 = 0,029$ La superficie mínima que le interesa a la inmobiliaria es 150 m^2
n) Bajo el supuesto de que la frecuencia se distribuye uniformemente dentro del intervalo, del 3,6% de las viviendas en el intervalo (120-150] el 1,8% corresponde al subintervalo (120-135) y el 1,8% al intervalo (135-180). El porcentaje total de viviendas con más de 135 m^2 es 1,8 + 1,5 + 1,4 = 4,7%

Ejercicio 4. Los siguientes resultados recogen la distribución de frecuencias de la puntuación obtenida en una determinada prueba:

```
$breaks
 [1]  0  2  4  6  8  10
$counts
 [1]   5  11  19  10  5
$mids
 [1]  1  3  5  7  9
```

Realice la tabla de frecuencias simples (absolutas y relativas) y acumuladas e indique:

a) ¿Cuántos alumnos han realizado esta prueba?
b) ¿Qué porcentaje de alumnos ha obtenido como mínimo 4 puntos?
c) ¿Cuál es la puntuación mínima del 30% de las mejores puntuaciones?
d) Si Pedro ha sacado un 3, ¿puede decirse que está entre el 10% de los alumnos con peores calificaciones?

Ejercicio 5. Los siguientes datos corresponden a la observación de la variable X='número de habitantes' observada en 50 municipios de una Comunidad Autónoma:

2727	1418	655	4582	764	6982	2509	1527	981	2945
6218	1745	1200	982	1418	3600	4036	5345	1091	1636
982	1527	1309	3927	2727	1445	2073	3055	1636	1964
1309	2400	1636	2400	2073	4691	1200	5564	4364	1309
3382	1418	2945	2291	1745	4036	4691	873	3491	2945

a) Realice el diagrama de tallo y hojas.
b) Con el programa R-Commander obtenga el diagrama de tallo y hojas (Stem & Leaf)
c) El 30% de los municipios con menos habitantes, aproximadamente, ¿cuántos tienen como máximo?
d) Aproximadamente, ¿cuántos habitantes como mínimo tiene que tener un municipio si se encuentra en el 16% con más habitantes?

Ejercicio 6. El siguiente diagrama Stem-and-Leaf recoge el *ÍNDICE de ALFABETIZACIÓN* (IA) correspondiente a un conjunto de países en vías de desarrollo:

```
  Stem-and-leaf unit = 1    1|2  represents 12
      2    1|27
      7    2|67799
     21    3|13333566667789
     (7)   4|1333358
     19    5|023444679
     10    6|0149
      6    7|4555
      2    8|12
```

a) ¿Cuál es el tamaño muestral?
b) Indique el índice de alfabetización máximo, mínimo y el más frecuente.
c) ¿Cuántos de estos países tiene un IA inferior a 48?
d) ¿En cuántos de estos países el IA no supera el valor 53?
e) ¿Cuál es el IA mínimo de los 10 países con mayor IA?

Tema 3. MEDIDAS DE POSICIÓN

3.1 Media aritmética
3.2 Moda
3.3 Mediana
3.4 Medidas de localización: cuantiles

MEDIDAS DE POSICIÓN CENTRAL

MEDIA ARITMÉTICA
Centro de gravedad

$$\overline{X} = \frac{\sum_{i=1}^{k} X_i n_i}{n}$$

MEDIANA
Valor central

$$Me = X_i \Leftrightarrow N_i \geq n/2 \text{ y } N_{i-1} < n/2$$

$$Me = L_{i-1} + \frac{n/2 - N_{i-1}}{n_i} a_i$$

MODA,
Valor más frecuente

$$Mo = X_i \Leftrightarrow \text{máx}\{n_i\}$$

MEDIA ARITMÉTICA

- Es la medida de tendencia central más adecuada cuando la característica observada es una variable.
- Se define como la suma de los valores de la variable observados en los elementos de la muestra dividida por el tamaño de ésta.
- Si la distribución de frecuencias se presenta con los valores de la variable agrupados en intervalos, al calcular la media utilizando las correspondientes marcas de clase se obtiene un resultado aproximado.

Propiedades:

- La media se expresa en las unidades de medida de la variable.
- La media siempre toma un valor comprendido entre los valores de X mínimo y máximo observados.
- La media de una constante es la misma constante.
- La media aritmética es el punto de equilibrio de la distribución, es decir, la suma de las desviaciones de todos los valores de la variable con respecto a la media es igual a cero.
- En el cálculo de la media se utiliza toda la información contenida en la distribución de frecuencias.

Inconvenientes

- Sólo se puede obtener si la característica observada es cuantitativa.
- La media es muy sensible a la presencia de observaciones extremas, tendiendo a desplazarse hacia éstas. Cuando esto ocurre la media no sintetiza adecuadamente el orden de magnitud de los valores de la variable.

MEDIANA

- Si se ordenan los elementos de la muestra desde el que tiene el menor valor de la variable hasta el que tiene el valor mayor, la mediana es el valor de la variable correspondiente al elemento que ocupa la posición central. La mediana, por tanto, divide la distribución de frecuencias en dos partes con igual número de elementos.
- Si el número de observaciones es par, la mediana se calcula como el promedio de los valores de la variable correspondiente a los dos elementos centrales.
- Si los valores de la variable se agrupan en intervalos, el intervalo que contiene a la mediana es el primero que presenta una frecuencia absoluta acumulada igual o superior a n/2.
  Una vez localizado el intervalo que la contiene, el valor de la mediana se aproxima mediante una expresión (ver cuadro) basada en el supuesto de que la frecuencia correspondiente a cada intervalo se distribuye uniformemente dentro de éste.

Propiedades

- La mediana se expresa en las mismas unidades de medida de la variable.
- La mediana puede ser una medida de tendencia central más representativa que la media cuando la variable presenta valores extremos.

Inconvenientes

- Sólo se puede obtener si la característica observada es ordinal.
- En el cálculo de la mediana no se tiene en cuenta toda la información contenida en la distribución de frecuencias.

MODA

- La moda es el valor de la variable que más veces se repite.
- Para localizar la moda se busca cual es la frecuencia (absoluta o relativa) máxima: el valor de la variable al que ésta corresponde es la moda.
- Si los valores de la variable se agrupan en intervalos, el intervalo modal es aquel al que le corresponde la frecuencia máxima. En tal caso puede tomarse la marca de clase del intervalo modal como valor aproximado de la moda.

Propiedades

- La moda se expresa en las unidades de medida de la variable.
- La moda es la única medida de posición que sintetiza la distribución de frecuencias de una característica categórica nominal.

Inconvenientes

- Una distribución de frecuencias puede tener más de una moda.
- Para determinar la moda no se tiene en cuenta toda la información contenida en la distribución de frecuencias.

MEDIDAS DE LOCALIZACIÓN

Cuantiles

- Si se ordenan los elementos de la muestra desde el que tiene el menor valor de la variable hasta el que tiene el valor mayor, los cuantiles son los valores de la variable que dividen a la distribución en un cierto número de partes con igual número de elementos.
- Los cuantiles se expresan en las mismas unidades de medida de la variable.
- Los cuantiles más utilizados son los cuartiles, los deciles y los centiles o percentiles.

Cuartiles

- Son los tres valores de la variable, Q_1, Q_2, Q_3 que dividen la distribución en cuatro partes con igual número de observaciones.
- El primer cuartil, Q_1, es el valor de la variable que deja por debajo el 25% del total de observaciones. El segundo cuartil, Q_2, es el valor de la variable que deja por debajo el 50% de las observaciones y, por tanto, coincide con la mediana. El tercer cuartil, Q_3, es el valor de la variable que deja por debajo el 75% del total de observaciones.
- Entre dos cuartiles consecutivos se encuentra el 25% del total de observaciones.
- El r-ésimo cuartil es el valor de la variable al que le corresponde la primera frecuencia absoluta acumulada mayor o igual que $rn/4$, dónde $r = 1, 2, 3$.
- Si los valores de la variable se agrupan en intervalos, el intervalo que contiene al r-ésimo cuartil es aquel cuya frecuencia absoluta acumulada es la primera mayor o igual que $rn/4$ para $r = 1, 2, 3$.

  Una vez localizado el intervalo que lo contiene, el valor de Q_r se aproxima mediante la siguiente fórmula basada en el supuesto de que la frecuencia correspondiente a cada intervalo se distribuye uniformemente dentro de éste.

$$Q_r = L_{i-1} + \frac{r\,n/4 - N_{i-1}}{n_i} a_i$$

Deciles

- Son los nueve valores de la variable, D_1, D_2, …., D_8, D_9 que dividen la distribución en diez partes con igual número de observaciones.
- El primer decil D_1, es el valor de la variable que deja por debajo el 10% del total de observaciones; el segundo decil, D_2, es el valor de la variable que deja por debajo el 20% de las observaciones y así sucesivamente. El quinto decil, D_5, coincide con la mediana.
- Entre dos deciles consecutivos se encuentra el 10% del total de observaciones.
- El r-ésimo decil es el valor de la variable al que le corresponde la primera frecuencia absoluta acumulada mayor o igual que rn/10, dónde r = 1, 2, 3, …, 8, 9.
- Si los valores de la variable se agrupan en intervalos, el intervalo que contiene al r-ésimo decil es aquel cuya frecuencia absoluta acumulada es la primera mayor o igual que rn/10 para r = 1, 2, 3, …, 8, 9.
  Una vez localizado el intervalo que lo contiene, el valor de D_r se aproxima mediante la siguiente expresión basada en el supuesto de que la frecuencia correspondiente a cada intervalo se distribuye uniformemente dentro de éste.

$$Dr = L_{i-1} + \frac{r\,n/10 - N_{i-1}}{n_i} a_i$$

Centiles o percentiles

- Son los noventa y nueve valores de la variable, C_1, C_2, …., C_{98}, C_{99} que dividen la distribución en cien partes con igual número de observaciones.
- El primer decil C_1, es el valor de la variable que deja por debajo el 1% del total de observaciones; el segundo centil, C_2, es el valor de la variable que deja por debajo el 2% de las observaciones y así sucesivamente. El quincuagésimo centil, C_{50}, coincide con la mediana.
- Entre dos centiles consecutivos se encuentra el 1% del total de observaciones.

- El r-ésimo centil es el valor de la variable al que le corresponde la primera frecuencia absoluta acumulada mayor o igual que rn/100, dónde r = 1, 2, 3, …, 98, 99.
- Si los valores de la variable se agrupan en intervalos, el intervalo que contiene al r-ésimo centil es aquel cuya frecuencia absoluta acumulada es la primera mayor o igual que que rn/100 para r = 1, 2, 3, …, 98, 99.

  Una vez localizado el intervalo que lo contiene, el valor de C_r se aproxima mediante la siguiente expresión basada en el supuesto de que la frecuencia correspondiente a cada intervalo se distribuye uniformemente dentro de éste.

$$Cr = L_{i-1} + \frac{rn/100 - N_{i-1}}{n_i} a_i$$

EJERCICIOS TEMA 3

Ejercicio 1. Dada la siguiente distribución de la variable X = "Número de créditos personales concedidos en una determinada oficina bancaria" observada en una muestra de n días:

X	(0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
n	6	15	39	40	8	6	6

a) Indique el tamaño de la muestra
b) Calcule el promedio de créditos concedidos por día.
c) Calcule la media recortada de X eliminando el 10% de las observaciones extremas.

Ejercicio 2. Una empresa dedicada a la venta a domicilio ha fijado en sus 5 sucursales las siguientes dietas por vendedor:

Sucursal	Número Vendedores	Dietas (€/Vendedor)
A	11	150
B	13	140
C	19	110
D	20	150
E	17	200

En el conjunto de las 5 sucursales, ¿cuál es la dieta media por vendedor?

Ejercicio 3. Se invierten 1.000 u.m. en dos carteras compuestas por diferentes títulos de renta variable. Las cantidades invertidas y la rentabilidad de los títulos han sido las siguientes:

CARTERA A			CARTERA B		
Tít	Cant. Invertida	Rentabilidad	Tít	Cant. Invertida	Rentabilidad
A	80	4%	E	350	6%
B	170	7%	F	100	5,5%
C	130	8%	G	50	7%
D	120	5,64%			

a) ¿En cuál de las dos carteras la rentabilidad media ha sido mayor?
b) ¿Cuál es la rentabilidad media obtenida del total invertido?

Ejercicio 4. Durante el último mes se han realizado las siguientes operaciones de cambio:

A) De Libras a Euros		B) De Euros a Libras	
Tipo de Cambio (Euros/Libra)	Importe (Libras)	Tipo de Cambio (Euros/Libra)	Importe (Euros)
1,50	120	1,50	300
1,45	180	1,45	348
1,40	260	1,40	700
1,42	340	1,42	1136

Para cada uno de los casos anteriores, indique el tipo de cambio medio en Euros/Libra.

Resolución

De Libras a Euros

Tipo de cambio €/£	Importe £	Contravalor €
1,50	120	180
1,45	180	261
1,40	260	364
1,42	340	482,8
Total	900	1287,8

De Euros a Libras

Tipo de cambio €/£	Importe €	Contravalor £
1,50	300	200
1,45	348	240
1,40	700	500
1,42	1136	800
Total	2484	1740

A) Tipo de cambio medio $= \frac{\text{Total Euros}}{\text{Total Libras}} = \frac{1287,8}{900} = 1,4309$ €/£

B) Tipo de cambio medio $= \frac{\text{Total Euros}}{\text{Total Libras}} = \frac{2484}{1740} = 1,4276$ €/£

Ejercicio 5. La cotización media mensual en Bolsa de un cierto valor durante los últimos 12 meses ha sido: 70,5; 67,9; 69,1;72,3; 73,1; 74,2; 75,1; 74,8; 72,1; 71,2; 69,5 y 67,1. Si por experiencia se sabe que su comportamiento es ajustarse al valor medio, qué haría ¿comprar o vender?

Ejercicio 6. El personal de una empresa está formado por operarios y técnicos. Se sabe que 45 son operarios y suponen 3/4 partes del total de la plantilla, y el resto son técnicos. El salario medio de los operarios es 12500 u.m. y la masa salarial de los técnicos asciende a 630000 u.m. ¿Cuál es el salario medio de los trabajadores de esta empresa?

Ejercicio 7. El salario medio mensual en una determinada empresa, cuyo personal está formado únicamente por técnicos y operarios, es 1820 €. Si el salario medio mensual de los técnicos es 2100 € y el de los operarios 1700 €, ¿cuál es el porcentaje de operarios empleados?

Ejercicio 8. Se sabe que un taller produce por término medio 10 piezas/hora en el turno de día, 8 en el turno de noche y 9,2 piezas/hora considerando los dos turnos. Si el taller funciona 20 horas al día, ¿cuántas horas se trabaja en el turno de día? ¿y en el de noche?

Ejercicio 9. Un examen de una determinada asignatura consta de dos partes tales que al calcular la nota final se da doble importancia al resultado de la primera parte. Un alumno que ha obtenido un 5 en la primera parte y su nota final es un 6, ¿qué nota tenía en la segunda parte del examen?

Ejercicio 10. Calcule las medidas de tendencia central (media, mediana y moda) de las distribuciones de frecuencias de los ejercicios propuestos 1, 2 y 3 del Tema 2.

Ejercicio 11. El importe (en Euros) de 150 tickets de la cafeteria de un hotel se ha tabulado y se ha obtenido:

Importe

Indique, aproximadamente:

a) Importe máximo y mínimo observado.
b) Importe máximo del 50% de los tickets con menor importe.
c) Número de tickets con un importe mínimo de 175 Euros.
d) Importe mínimo y máximo del 50% de los tickets centrales.
e) ¿Cuál es el importe mínimo de un ticket que está entre el 30% de los de mayor importe?
f) ¿Qué porcentaje de los tickets contabilizados en esta tabla tienen un importe superior a 250€?
g) Se quiere seleccionar los 9 tickets con menor importe. Aproximadamente, ¿cuál es el importe máximo de estos tickets?

Resolución

a) El gasto mínimo es 0 Euros y el máximo 300 Euros

b) Distribución de frecuencias absolutas acumuladas

Gastos extras		ni	Ni
Válidos	(0, 50]	10	10
	(50; 70]	30	40
	(70; 90]	25	65
	(90; 120]	27	92
	(120; 150]	20	112
	(150; 200]	18	130
	(200; 250]	15	145
	(250; 300]	5	150
Total		150	

150/2 = 75 La primera Ni > 75 es 92; Me $\in$ (90-120)

$$Me = 90 + \frac{75-65}{27}30 = 101,1 \text{ Euros}$$

c) La marca de clase del intervalo (150-200) es 175, cuya frecuencia absoluta es 18. Bajo el supuesto de que esta frecuencia se distribuye uniformemente dentro del intervalo, al subintervalo (175-200) le corresponden la frecuencia absoluta 9. El número de tickets con un importe mínimo de 175 Euros es, aproximadamente:

9 + 15 + 5 = 29 tickets

d) 150/4 = 37,5 La primera Ni > 37,5 es 40; Q1 $\in$ (50-70)

$$Q1 = 50 + \frac{37,5-10}{30}20 = 68,33 \text{ Euros}$$

3(150)/4 = 112,5 La primera Ni > 112,5 es 130; Q3 $\in$ (150-200)

$$Q3 = 150 + \frac{112,5-112}{18}50 = 151,39 \text{ Euros}$$

Los importes mínimo y máximo del 50% de las ticket centrales son, aproximadamente, 68 y 151 Euros.

e) 7(150)/10 = 105 La primera Ni > 105 es 112; D7 $\in$ (120-150)

$$D7 = 120 + \frac{105-92}{20}30 = 139,5 \text{ Euros}$$

El importe mínimo de un ticket que está entre el 30% de las de mayor importe es 139,5 Euros

f) Con importe superior a 250 € hay 5 tickets, el 3,3%

g) Al intervalo (0-50), de amplitud 50 le corresponde una frecuencia de 10 tickets, suponiendo que la frecuencia se reparte uniformemente dentro del intervalo:

Intervalo	Amplitud	Frecuencia
(0-x)	x	9
(x-50)	50-x	1

a = 50 n = 10
a = x n = 9

$x = 9(50)/10 = 45$

El importe máximo de los 9 tickets con menor importe es, aproximadamente, 45 Euros.

Ejercicio 12. El importe (en Euros) de los últimos 58 tickets de caja de un establecimiento es:

```
1 | 2: represents 1.2
 leaf unit: 0.1
            n: 58
    2      5 | 05
    6      6 | 0589
    7      7 | 9
    9      8 | 34
   14      9 | 12345
   21     10 | 0133458
   23     11 | 33
  (16)    12 | 0122235677777899
   19     13 | 12568899
   11     14 | 01225
    6     15 | 266689
```

Conteste las siguientes cuestiones e indique el nombre del estadístico correspondiente.

a) Importe máximo del 25% de los tickets con menor importe.
b) Importe mínimo del 25% de los tickets de mayor cuantía.
c) Diferencia entre el importe mayor y el menor del 50% de los tickets centrales.
d) Importe máximo del 40% de los tickets con menor importe.
e) Se ha aplicado un descuento a los tickets con mayor importe. Si sólo se han beneficiado un 15%, ¿a partir de qué importe se ha aplicado?
f) ¿Qué porcentaje de tickets presentan un importe superior a12,8€?

Ejercicio 13. En una Cooperativa el 20% del personal es administrativo y el resto técnico. Se ha recogido información sobre la antigüedad del personal y los resultados son:

Antigüedad (en años)	Administrativos %	Técnicos %
1	10	5
2	15	20
3	25	20
4	30	40
5	20	15

a) Halle la antigüedad media del total de la plantilla.
b) En la distribución de los Administrativos, ¿cuál es la antigüedad máxima del 20% de los trabajadores con menor antigüedad?
c) En la distribución de los Técnicos, ¿cuál es la antigüedad mínima del 30% de los trabajadores con mayor antigüedad?
d) En la distribución de toda la plantilla, ¿cuál es la antigüedad máxima del 50% de los trabajadores?

Tema 4. MEDIDAS DE DISPERSIÓN Y FORMA

4.1 Recorrido. Recorrido intercuartílico. Diagrama de caja (Box Plot)
4.2 Varianza y desviación estándar
4.3 Transformaciones lineales. Variable estandarizada
4.4 Medidas de forma

MEDIDAS DE DISPERSIÓN

VARIANZA, S^2

- Es una medida de dispersión respecto a la media aritmética.
- Se define como el promedio de las desviaciones, elevadas al cuadrado, de los valores observados respecto a la media.
- Complementa la información proporcionada por la media. Indica si $\overline{X}$ es representativa de la distribución.
- Se puede simplificar su cálculo utilizando la siguiente expresión:

$$S^2 = \frac{\sum_{i=1}^{k} x_i^2 n_i - n\overline{X}^2}{n-1}$$

Propiedades:

- La varianza siempre toma valores no negativos.
- Si todos los valores de la distribución son iguales la varianza es 0.
- La varianza no cambia cuando se suma una misma cantidad a todos los valores observados. Cuando se realiza un cambio de origen.
- La varianza queda modificada si se multiplican todas las observaciones de la distribución por la misma constante. Cuando se realiza un cambio de escala o cambio de unidades de medida.

Inconvenientes:

- No presenta la misma unidad de medida que la variable.
- Depende de los cambios de la unidad de medida.
- No está acotada.

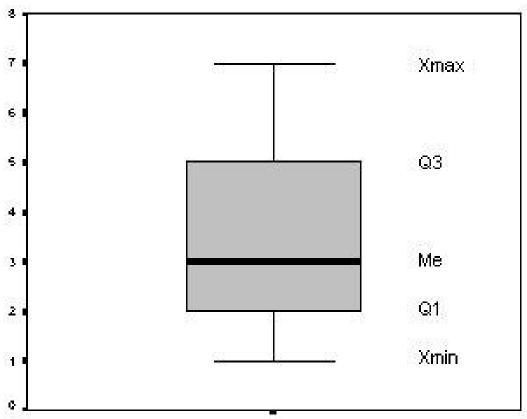
DESVIACIÓN ESTÁNDAR O TÍPICA, S

- Se define como la raíz cuadrada positiva de la varianza.
- Presenta la misma unidad de medida que la variable y que la media aritmética.
- Permite establecer el siguiente intervalo alrededor de la media:
  $(\bar{X} - 3\,s; \quad \bar{X} + 3\,s)$
  Prácticamente contiene todos los valores de la distribución. Sólo algunos valores extremos superarán estos límites; es decir, los valores de la variable sólo en algunos casos extremos diferirán de la media en más de 3 veces la desviación estándar.
- Presenta las mismas propiedades que la varianza.

COEFICIENTE DE VARIACIÓN, CV(x)

- Es el cociente entre la desviación estándar y el valor absoluto de la media.
- Indica el número de veces que la desviación estándar contiene a la media, $\bar{X}$.
- Es una medida de dispersión relativa (no presenta unidades de medida).
- Permite comparar la dispersión alrededor de la media de dos o más distribuciones aunque presenten distintas unidades de medida o medias aritméticas diferentes.
- No le afectan los cambios de unidades (cambios de escala).
- Cuando $\bar{X}$=0 el CV es indeterminado.

DIAGRAMA BOX-PLOT (diagrama de caja)



- Recoge de forma esquemática la dispersión y la forma de la distribución.
- Representa la distribución en base a 5 valores: X_{MIN}, Q_1, Me, Q_3, X_{MAX}.
- Se compone de una caja central de longitud igual al recorrido intercuartíllico y unos segmentos laterales que abarcan el rango o recorrido de la distribución.
- Los valores extremos u outliers se representan como puntos separados. Estos valores se clasifican en:
  - Atípicos, si distan del primer o tercer cuartíl en más de 1,5 veces el recorrido intercuartílico o superan los límites:
    $L_I = Q_1 - 1{,}5\ R_Q$ y $L_S = Q_3 + 1{,}5\ R_Q$
  - Extremos, si distan del primer o tercer cuartíl en más de 3 veces el recorrido intercuartílico o superan los límites:
    $L_I = Q_1 - 3\ R_Q$ y $L_S = Q_3 + 3\ R_Q$

TRANSFORMACIONES LINEALES DE UNA VARIABLE X

1. Cambio de origen: a todas las observaciones de la variable X se les suma una constante cualquiera (a). La nueva variable será X'=X+a
2. Cambio de escala: todas las observaciones se multiplican por una constante cualquiera (b). La nueva variable será X'= bX
3. Cambio de origen y de escala: todas las observaciones se multiplican por una constante (b) y se les suma otra constante (a). La nueva variable será X'= a+bX

Incidencia de estas transformaciones lineales en las medidas resumen:

Transformación	Variable transformada	Media aritmética	Mediana	Moda
Cambio de origen	X'=X+a	$\overline{X}' = \overline{X} + a$	Me(X')=Me(X) +a	Mo(X')=Mo(X) +a
Cambio de escala	X'=bX	$\overline{X}' = b\,\overline{X}$	Me(X')= b Me(X)	Mo(X')= b Mo(X)
Cambio de origen y escala	X'=a+bX	$\overline{X}' = a + b\,\overline{X}$	Me(X')= a+b Me(X)	Mo(X')= a+b Mo(X)

Transformación	Variable transformada	Varianza	Desviación estándar	Coeficiente Variación
Cambio de origen	X'=X+a	$S^2_{X'} = S^2_X$	$S_{X'} = S_X$	$CV_{X'} \neq CV_X$
Cambio de escala	X'=bX	$S^2_{X'} = b^2 S^2_X$	$S_{X'} = \lvert b \rvert S_X$	$CV_{X'} = CV_X$
Cambio de origen y escala	X'=a+bX	$S^2_{X'} = b^2 S^2_X$	$S_{X'} = \lvert b \rvert S_X$	$CV_{X'} \neq CV_X$

TIPIFICACIÓN O ESTANDARIZACIÓN

Es un caso particular de cambio de origen y escala.
Definición: si a todas las observaciones de X se les resta su media ($\overline{X}$) y la diferencia se divide por su desviación estándar (S_X) se obtiene una nueva variable Z_i que recibe el nombre de variable tipificada o estandarizada.

$$Z_i = \frac{X_i - \overline{X}}{S_X}$$

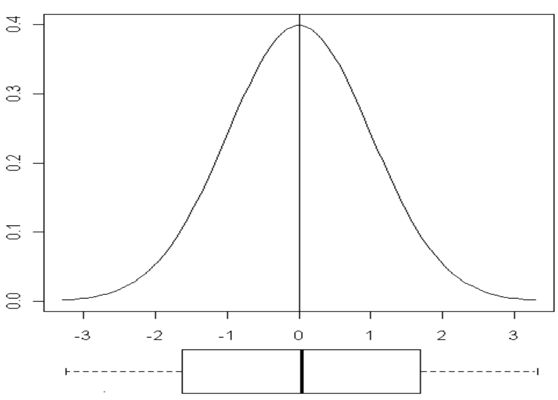
Sus valores (Z_i) permiten comparar en términos relativos la posición de datos en diferentes distribuciones.

Interpretación: sus valores indican el número de desviaciones estándar que el valor particular X_i está por encima (si Z es positivo) o por debajo (si Z es negativo) de su media: $Z_i\ S_X = X_i - \overline{X}$

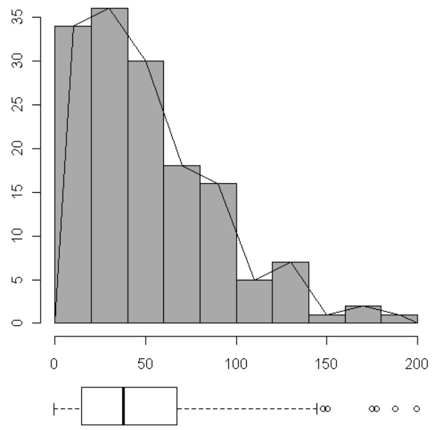
Propiedades:
- Su media aritmética es cero: $\overline{Z}$=0
- Su desviación estándar es uno: S_Z=1
- Es una variable sin unidades de medida (adimensional)

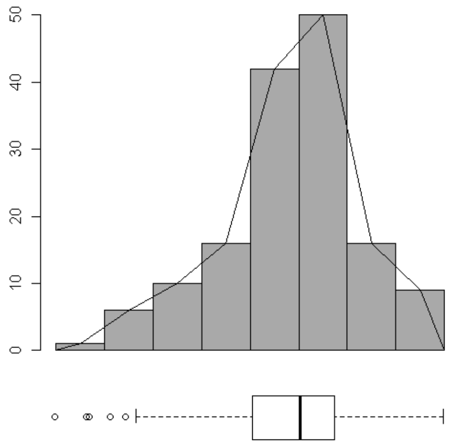
ASIMETRÍA

Una distribución de frecuencias es simétrica si su representación gráfica (diagrama de barras o histograma), tiene un eje de simetría perpendicular al eje de abcisas tal que la parte de la distribución que queda a un lado del eje es la imagen especular de la parte que queda la otro lado del eje. En caso contrario, la distribución es asimétrica.
La siguiente distribución es simétrica.



Las siguientes distribuciones son asimétricas:





Una distribución asimétrica con la cola hacia la derecha se dice que tiene asimetría positiva, si la cola es hacia la izquierda se dice que tiene asimetría negativa. De las distribuciones anteriores la primera tiene asimetría positiva y la segunda asimetría negativa.

MEDIDAS DE ASIMETRÍA

Coeficiente de asimetría de Fisher

Se define como:

$$g_1 = \frac{\dfrac{\sum_{i=1}^{k} \left(x_i - \bar{X}\right)^3 n_i}{n}}{S^3}$$

- Si la distribución es simétrica $g_1=0$
- Si hay asimetría positiva $g_1>0$.
- Si hay asimetría negativa $g_1<0$.

Coeficiente de asimetría de Pearson

Si la distribución tiene forma campanoide, con una sola moda y asimetría moderada, puede utilizarse como medida de asimetría el coeficiente de asimetría de Pearson:

$$A=\frac{\bar{X}-Me}{S}$$

- Si la distribución es simétrica la media es igual que la mediana y A=0.
- Si hay asimetría positiva la media es mayor que la mediana y A>0.
- Si hay asimetría negativa la media es menor que la mediana y A<0.

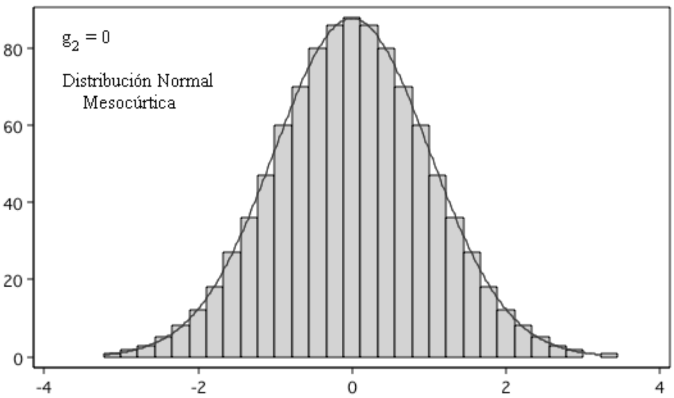
MEDIDAS DE CURTOSIS

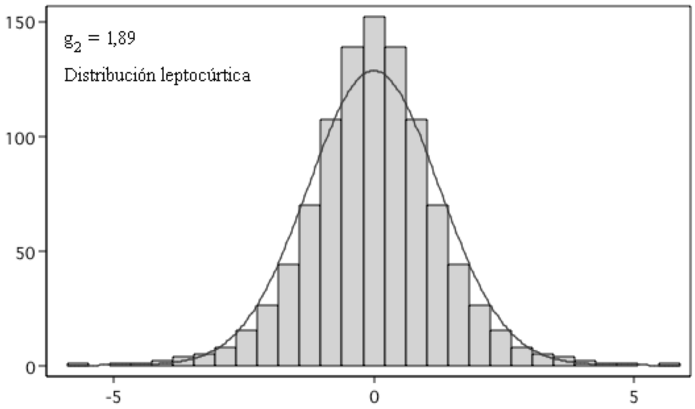
La curtosis mide el grado de apuntamiento de una distribución de frecuencias por comparación con una distribución teórica (distribución de probabilidad) de una variable continua, que recibe el nombre de distribución Normal, que se toma como modelo de referencia. Este modelo tiene forma de campana simétrica y unimodal.
La curtosis analiza la deformación, en sentido vertical, de una distribución con respecto a la Normal.

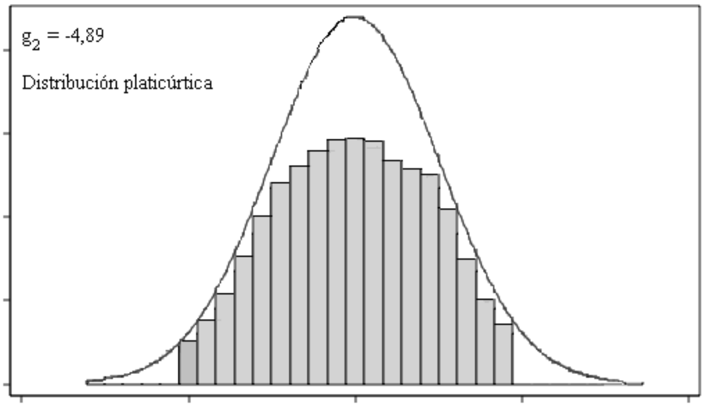
Coeficiente de Curtosis de Fisher

$$g_2=\frac{\dfrac{\sum_{i=1}^{k}(x_i-\bar{X})^4 n_i}{n}}{S^4} - 3$$

- Si g_2=0 la distribución tiene el mismo grado de apuntamiento que la curva Normal y se denomina mesocúrtica.
- Si g_2>0 la distribución es leptocúrtica, y tiene mayor grado de apuntamiento que la curva Normal
- Si g_2<0 la distribución es platicúrtica, y tiene menor grado de apuntamiento que la curva Normal.







La curtosis puede considerarse como una medida del peso relativo de las colas de la distribución dentro de la variación total. En una distribución leptocúrtica el peso relativo de las colas es mayor que en la distribución Normal; en una distribución platicúrtica el peso relativo de las colas es menor que en la distribución Normal.

EJERCICIOS TEMA 4

Ejercicio 1. Calcule las medidas de dispersión (absoluta y relativa) de las distribuciones de frecuencias de los ejercicios propuestos 1 y 2 del Tema 2 y ejercicio propuesto 1 del Tema 3. Interprete estos resultados.

Ejercicio 2. Dadas las siguientes distribuciones de frecuencias, calcule:

a) Número medio de visitas semanales en cada comunidad y en el conjunto de ambas comunidades.
b) Desviación estándar del número de visitas semanales en cada una de estas comunidades.
c) Indique en cuál de las dos distribuciones es más representativa la media.

Ejercicio 3. Los resultados de la prueba de acceso a la Universidad correspondientes a los alumnos de cuatro institutos de Barcelona (A, B, C y D) se resumen en la siguiente tabla:

	mean	sd	0%	25%	50%	75%	100%	n
A	6.62	1.85	4.01	5.28	6.97	8.47	10.00	130
B	6.48	1.68	4.31	5.28	6.13	8.71	10.00	50
C	6.05	1.46	4.02	5.60	6.93	8.22	9.97	110
D	5.99	1.95	4.01	6.65	8.22	9.32	10.00	140

Se pide:
a) Nota media del total de alumnos.
b) Indique en cuál de los cuatro institutos las calificaciones de los alumnos son más homogéneas.
c) ¿Qué transformación se debería realizar en las calificaciones de los alumnos del instituto D para que tengan la misma nota media que los del instituto A y varianza igual a 1.

Ejercicio 4. El salario medio semanal de los empleados de una empresa es 550 € con desviación típica 300 €. La empresa tiene 500 trabajadores y se

plantea realizar un incremento salarial para lo cual propone las siguientes alternativas:
A) Efectuar un aumento lineal de 80 €.
B) Incrementar un 6%
C) Un incremento del 4% más un aumento lineal de 60 €.
D) Aumentar 50 € más un 5% sobre el salario que resulte después de aumentar los 50 €.
Se pide:
a) Salarios medios tras aplicar las alternativas anteriores.
b) ¿Cuál es el aumento en € de la masa salarial resultante en cada alternativa?
c) Indique cuál de las cuatro propuestas de revisión salarial disminuye la dispersión relativa del salario.

Resolución

a) Situación actual de la distribución de salarios: $\bar{X} = 550$; $S = 300$

Con la alternativa A el nuevo salario semanal es $X_A = X + 80$. El salario medio y la desviación típica serán:

$\bar{X}_A = \bar{X} + 80 = 550 + 80 = 630$ €
$S_A = S = 300$ €

Con la alternativa B el nuevo salario semanal es $X_B = X + 0{,}06\ X = 1{,}06\ X$. El salario medio y la desviación típica serán:

$\bar{X}_B = 1{,}06\ \bar{X} = 1{,}06\ (550) = 583$ €
$S_B = 1{,}06\ S = 1{,}06\ (300) = 318$ €

Con la alternativa C el nuevo salario semanal es $X_C = 1{,}04\ X + 60 + 0{,}04\ X = 1{,}04\ X + 60$. El salario medio y la desviación típica serán:

$\bar{X}_C = 1{,}04\ \bar{X} + 60 = 1{,}04\ (550) + 60 = 632$ €
$S_C = 1{,}04\ S = 1{,}04\ (300) = 312$ €

Con la alternativa D el nuevo salario semanal es

$X_D = X + 50 + 0{,}05\ (X + 50) = X + 50 + 0{,}05\ X + 2{,}5 = 1{,}05\ X + 52{,}5$

El salario medio y la desviación típica serán:

$\bar{X}_D = 1{,}05\ \bar{X} + 52{,}5 = 1{,}05\ (550) + 52{,}5 = 630$ €
$S_D = 1{,}05\ S = 1{,}05\ (300) = 315$ €

b) En la situación actual la masa salarial es: $MS = 550\ (500) = 275000$

Con las diferentes alternativas la masa salarial sería:

MSA = 630 (500) = 315000 €, lo que supone un incremento de 315000 – 275000 = 40000 €

MSB = 583 (500) = 291500 €, lo que supone un incremento de 291500 – 275000 = 16500 €

MSC = 632 (500) = 316000 €, lo que supone un incremento de 316000 – 275000 = 41000 €

MSD = 630 (500) = 315000 €, lo que supone un incremento de 315000 – 275000 = 40000 €

c) Las dispersiones relativas de los salarios son:

Actual	CV = 300/550 = 0,5454	
Con A	CV_A = 300/630 = 0,4762	Dispersión relativa menor
Con B	CV_B = 318/583 = 0,5454	
Con C	CV_C = 312/632 = 0,4937	
Con D	CV_D = 315/630 = 0,5	

Ejercicio 5. Una empresa de telefonía controla el número de clientes captados en sus cuatro agencias. Con la información recogida en cada agencia se han obtenido los siguientes resultados de la variable X="Número de clientes captados por empleado":

	mean	sd	0%	25%	50%	75%	100%
Agencia 1	5.106	1.81	1	3	5	6	10
Agencia 2	6.712	2.63	0	5	6	8	16
Agencia 3	7.575	2.89	1	5	8	11	15
Agencia 4	8.787	3.02	4	8	9	12	16

a) Si el número de clientes captados por el último empleado contratado en cada una de estas agencias (1, 2, 3 y 4) ha sido 6, 7, 8 y 10, respectivamente, ¿cuál de estos empleados ocupa mejor posición relativa en su respectiva agencia?
b) Un empleado de la agencia 2 en la escala estandarizada del "número de clientes captados" presenta una puntuación de -1,38. ¿Cuántos clientes captó dicho empleado?

Ejercicio 6. Una cadena de televisión presenta un nivel medio de audiencia (en miles de espectadores) de 1.450 en la programación de tarde y de 2.350 en la programación de noche. Si en el conjunto de cadenas de televisión la media y la varianza de la audiencia en las franjas horarias anteriores son, respectivamente, 1.850 y 160.000 en la primera, y 2.600 y 40.000 en la segunda, ¿en qué franja ocupa esta cadena mejor posición relativa?

Ejercicio 7. En el año 2009 las bibliotecas de Cataluña disponían de un promedio por comarca de 2896 volúmenes por 1000 habitantes con una desviación típica de 1407. Sabiendo que las comarcas Garraf, Barcelonès, Tarragonès, Baix Llobregat y Val d'Aran tenían en la escala estandarizada puntuaciones de: −0,77, 0,78, 0,36, −1,3 y −1, respectivamente

a) Indique en qué comarca/s el número de volúmenes por 1000 habitantes estaba por encima de la media de Catalunya.
b) Indique si alguna de estas comarcas presenta un número de volúmenes por 1000 habitantes outlier o anómalo.
c) Calcule el número de volúmenes por 1000 habitantes de estas comarcas

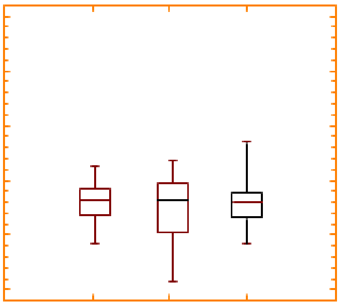
Ejercicio 8. El siguiente diagrama Stem-and-Leaf recoge el *ÍNDICE de ALFABETIZACIÓN* (IA) correspondiente a un conjunto de países en vías de desarrollo:

```
1 | 2: represents 12
 leaf unit: 1
             n: 44
   3     2* | 234
   7     2. | 5566
   8     3* | 3
   9     3. | 5
  10     4* | 3
  15     4. | 56889
  (6)    5* | 112223
  23     5. | 556677788899
  11     6* | 1122
   7     6. | 5677789
```

a) Obtenga el diagrama Box-Plot.
b) Comente las características más importantes de esta distribución.

Ejercicio 9. A partir de los resultados siguientes indique el Box-plot que corresponde a cada una de las siguientes variables:

Variable	X1	X2	X3
Mediana	20	20	20
Q_3-Q_1	6,2	2,8	2,3
Coef. Asimetría	-2,856	0	3,29
Coef. Curtosis	-1,873	-0,265	2,77



Ejercicio 10. De una compañía aérea A se sabe que por término medio el 20% de los vuelos no presentan retraso, mientras que los vuelos con retraso presentan la siguiente distribución de frecuencias:

Retraso (mn)	Núm. de Vuelos
5	1000
10	1000
15	500
20	300
30	200

Se pide:

a) ¿Cuál es el retraso medio de los vuelos con retraso? ¿y del total de vuelos de la compañía?

b) Si el último vuelo realizado se encuentra entre el 20% de los vuelos con más retraso de la distribución de los vuelos totales, ¿cuál es el mínimo retraso que puede haber presentado?

c) Si otra compañía aérea B presenta los siguientes resultados:

Retraso medio del total de vuelos: 10 mn

Varianza del total de vuelos: 60

Indique en qué compañía un vuelo con un retraso de 15 mn, presentará peor posición relativa.

d) ¿En qué compañía, A o B, los retrasos presentan más dispersión relativa?

e) Si tras una política de incentivos en la Cia. A el retraso de los vuelos ha experimentado una reducción del 5%, ¿cuál será la nueva media aritmética de los vuelos con retraso y su desviación estándar? y ¿cuál será la media del total de vuelos?

Ejercicio 11. La base de datos **Pasajeros.rda** contiene observaciones de la variable x = "Número de pasajeros en n vuelos de una compañía aérea". Explique el resultado que se obtiene ejecutando cada una de las siguientes instrucciones con el programa R-Commander (cargue previamente la base de datos **Pasajeros.rda).**

a)
```
M =cut (Pasajeros$x,
        breaks=c(100,150,200,250,300,350,400,450,500,550,600,650))
table(M)
```

b)
```
table(M)/length(Pasajeros$x)
```

c)
```
numSummary(Pasajeros[,"x"], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

d)
```
var(Pasajeros$x)
```

e)
```
IQR(Pasajeros$x)
```

f)
```
mean(Pasajeros$x, 0.2)
```

g)
```
library(fBasics)
skewness(Pasajeros$x)
```

h)
```
kurtosis(Pasajeros$x)
```

i) Realice los siguientes gráficos con las opciones del menú *Gráficas*:
Histograma; Gráfica de tallos y hojas; Diagrama de Caja

Resolución

a) cut(Pasajeros$x,breaks=c(100,150,200,250,300,350,400,450,500,550,600,650))
table(M)

Distribución de frecuencias absolutas con los valores de X agrupados en intervalos

(100,150]	(150,200]	(200,250]	(250,300]	(300,350]	(350,400]	(400,450]	(450,500]
24	24	21	13	21	13	13	8

(500,550]	(550,600]	(600,650]
4	1	2

b) table(M)/length(Pasajeros$x)

Distribución de frecuencias relativas

(100,150]	(150,200]	(200,250]	(250,300]	(300,350]	(350,400]
0.166666667	0.166666667	0.145833333	0.090277778	0.145833333	0.090277778

(400,450]	(450,500]	(500,550]	(550,600]	(600,650]
0.090277778	0.055555556	0.027777778	0.006944444	0.013888889

c) numSummary(Pasajeros[,"x"], statistics=c("mean", "sd", "quantiles"), quantiles=c(0, .25, .75,1))

La media, la desviación estándar los valores mínimo y máximo, los cuarteles y el tamaño de la muestra

mean	sd	0%	25%	75%	100%	n
280.2986	119.9663	104	180	360.5	622	144

d) var(Pasajeros$x)

La varianza de X
[1] 14391.92

e) IQR(Pasajeros$x)

El recorrido intercuartílico
[1] 180.5

f) mean(Pasajeros$x, 0.2)

La media recortada eliminando el 20% de las observaciones extremas
[1] 267.625

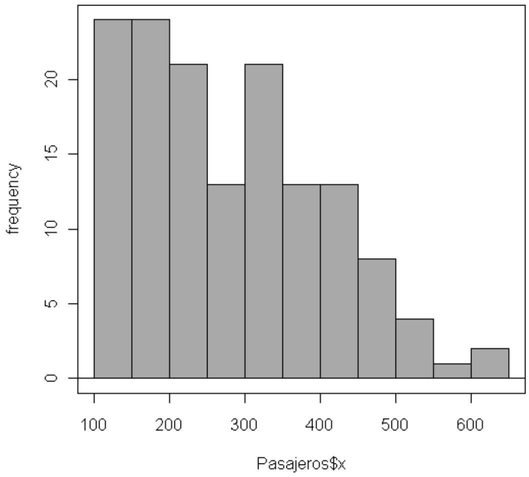
g) library(fbasic)
skewness(Pasajeros$x)

El coeficiente de asimetría g1
[1] 0.5710676

h) kurtosis(Pasajeros$x)

El coeficiente de curtosis g2
[1] -0.4298441

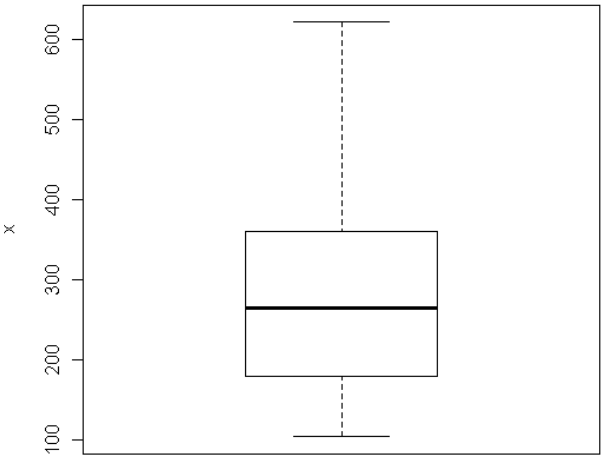
i)



GRÁFICA DE TALLOS Y HOJAS

```
1 | 2: represents 120
 leaf unit: 10
            n: 144
   23    1* | 01111112222333334444444
   48    1. | 5566677777778888889999999
   69    2* | 000011222233333333444
  (13)   2. | 5666677777789
   62    3* | 000001111111133444444
   41    3. | 5555666667999
   28    4* | 0000001112233
   15    4. | 66666779
    7    5* | 0034
    3    5. | 5
    2    6* | 02
```

DIAGRAMA DE CAJA



x
100
200
300
400
500
600

Tema 5. DISTRIBUCIÓN DE FRECUENCIAS BIDIMENSIONALES

5.1 Distribución de frecuencias conjuntas
5.2 Distribuciones marginales
5.3 Distribuciones condicionadas
5.4 Independencia estadística

DISTRIBUCIÓN DE FRECUENCIAS CONJUNTAS

Cuando sobre los n elementos de la muestra se observan simultáneamente dos variables, X e Y, cada elemento está representado por el par ordenado (x_i, y_j), donde x_i es el valor que toma la variable X e y_j es el valor que toma la variable Y.

ELEMENTOS DE UNA TABLA DE DOBLE ENTRADA

X \ Y	y_1	y_2	...	y_j	...	y_h	$n(x_i)$
x_1	n_{11} (f_{11})	n_{12} (f_{12})	...	n_{1j} (f_{1j})	...	n_{1h} (f_{1h})	$n(x_1)$ $f(x_1)$
x_2	n_{21} (f_{21})	n_{22} (f_{22})	...	n_{2j} (f_{2j})	...	n_{2h} (f_{2h})	$n(x_2)$ $f(x_2)$
...	...	...	...	...	...	...	...
x_i	n_{i1} (f_{i1})	n_{i2} (f_{i2})	...	n_{ij} (f_{ij})	...	n_{ih} (f_{ih})	$n(x_i)$ $f(x_i)$
...	...	...	...	...	...	...	...
x_k	n_{k1} (f_{k1})	n_{k2} (f_{k2})	...	...	...	n_{kh} (f_{kh})	$n(x_k)$ $f(x_k)$
n(y) f(y)	$n(y_1)$ $f(y_1)$	$n(y_2)$ $f(y_2)$	...	$n(y_j)$ $f(y_j)$	...	$n(y_h)$ $f(y_h)$	n 1

Interpretación de la tabla:

- x_i / y_j: en el margen izquierdo de la tabla se recogen los k valores de la variable X. En el margen superior de la tabla se recogen los h valores de la variable Y
- Si X o Y o ambas son variables contínuas, en los márgenes izquierdo y superior de la tabla se recogerán los intervalos en los que se agrupan los valores de cada variable.
- n_{ij}: frecuencias absolutas conjuntas; siendo n_{ij} el número de elementos de la muestra para los que $X = x_i$ e $Y = y_j$. La suma de todas las frecuencias absolutas conjuntas es igual a n:

$$\sum_{i=1}^{k}\sum_{j=1}^{h} n_{ij} = n$$

- f_{ij}: frecuencias relativas conjuntas; siendo $f_{ij} = n_{ij}/n$ la proprción en tanto por uno de elementos para los que $x = x_i$ e $Y = y_j$. La suma de todas las frecuencias relativas conjuntas es igual a 1:

$$\sum_{i=1}^{k}\sum_{j=1}^{h} f_{ij} = 1$$

  Si las frecuencias relativas conjuntas se multiplican por 100 se obtienen los correspondientes porcentajes.
- $n(x_i)$: frecuencia marginal de X. La suma de las frecuencias absolutas (relativas) conjuntas de la fila i-ésima es igual a la frecuencia absoluta (relativa) correspondiente al valor x_i:

$$\sum_{j=1}^{h} n_{ij} = n(x_i) \qquad \sum_{j=1}^{h} f_{ij} = f(x_i)$$

- $n(y_j)$: frecuencia marginal de Y. La suma de las frecuencias absolutas (relativas) conjuntas de la columna j-ésima es igual a la frecuencia absoluta (relativa) correspondiente al valor y_j:

$$\sum_{i=1}^{k} n_{ij} = n(y_j) \qquad \sum_{i=1}^{k} f_{ij} = f(y_j)$$

Distribuciones de frecuencias marginales

A partir de la distribución de frecuencias conjuntas se pueden establecer las distribuciones de frecuencias unidimensionales de X y de Y, que se llaman "marginales" porque los valores de las variables y las correspondientes frecuencias absolutas o relativas se encuentran en los márgenes de la tabla.
Distribuciones de frecuencias marginales de X y de Y

X	n(X)	f(X)
x_1	$n(x_1)$	$f(x_1)$
x_2	$n(x_2)$	$f(x_2)$
...	...	...
x_i	$n(x_i)$	$f(x_i)$
...	...	...
x_k	$n(x_k)$	$f(x_k)$
Total	n	1

Y	n(Y)	f(Y)
y_1	$n(y_1)$	$f(y_1)$
y_2	$n(y_2)$	$f(y_2)$
...	...	...
y_j	$n(y_j)$	$f(y_j)$
...	...	...
y_h	$n(y_h)$	$f(y_h)$
Total	n	1

Distribuciones de frecuencias condicionadas

A partir de la distribución de frecuencias conjuntas puede establecerse el comportamiento de una de las variables, por ejemplo X, cuando la otra, Y, cumple determinada condición. Por ejemplo, la distribución de X condicionada a $Y = y_j$ es:

$X/Y = y_j$	Frec. absolutas	Frec. relativas
x_1	n_{1j}	$f_{1/j} = n_{1j}/n(y_j)$
x_2	n_{2j}	$f_{2/j} = n_{2j}/n(y_j)$
...	...	...
x_i	n_{ij}	$f_{i/j} = n_{ij}/n(y_j)$
...	...	...
x_k	n_{kj}	$f_{k/j} = n_{kj}/n(y_j)$
Total	$n(y_j)$	1

Análogamente, la distribución de Y condicionada a X = x_i es:

$Y/X = x_i$	Frec. absolutas	Frec. relativas
y_1	n_{i1}	$f_{1/i}= n_{i1}/n(x_i)$
y_2	n_{i2}	$f_{2/i} = n_{i2}/n(x_i)$
...	...	...
y_j	n_{ij}	$f_{j/i} = n_{ij}/n(x_i)$
...	...	...
Y_h	n_{ih}	$f_{h/i} = n_{ih}/n(x_i)$
Total	$n(x_i)$	1

INDEPENDENCIA ESTADÍSTICA

Dos variables X e Y son estadísticamente independientes si y sólo si cada frecuencia relativa conjunta es igual al producto de las correspondientes frecuencias relativas marginales:

$$f_{ij} = f(x_i).f(y_j) \qquad \forall i, j$$

Teniendo en cuenta la relación entre frecuencias absolutas y relativas, la condición de independencia puede expresarse como:

$$n_{ij} = \frac{n(x_i)\,n(y_j)}{n} \qquad \forall i, j$$

Si X e Y son independientes:

- Todas las distribuciones de frecuencias relativas de X condicionada a cualquier valor de Y son iguales a la distribución de frecuencias relativas marginal de X.
- Todas las distribuciones de frecuencias relativas de Y condicionada a cualquier valor de X son iguales a la distribución de frecuencias relativas marginal de Y.

EJERCICIOS TEMA 5

Ejercicio 1. Los siguientes datos corresponden a una muestra de 30 personas en paro siendo X=Edad e Y=Género:

Edad	Género	Edad	Género	Edad	Género	Edad	Género	Edad	Género
18	H	26	H	34	M	45	M	51	H
19	M	27	H	36	H	47	M	52	H
21	H	29	M	37	M	48	M	54	H
22	M	30	H	39	H	48	H	57	M
24	H	32	M	41	H	50	M	58	H
25	H	33	H	44	M	50	H	60	H

Tabule estos datos agrupando los valores de la variable Edad en las siguientes categorías:
Edad = De 16 a 19, de 20 a 24, de 25 a 54 y más de 54

Ejercicio 2. La distribución de frecuencias relativas conjunta de las variables X= 'Actividad realizada durante el tiempo libre en un fin de semana' e Y= 'Edad' observada en un colectivo formado por 6000 personas es:

Activitats realitzades	[15-30)	[30-45)	[45-65)	≥65	Total
passejar	2.6	4.4	5.5	7.6	20.1
mirar TV o vídeo	2.9	3.2	4.0	4.2	14.3
reunions família/amics	3.7	3.1	2.9	3.7	13.4
platja/piscina	3.4	3.2	2.7	2.2	11.5
llegir	1.9	3.0	2.7	2.3	9.9
feines llar/cuinar	0.8	2.1	2.9	3.7	9.5
dormir/reposar	1.6	2.1	2.2	2.0	7.9
fer esport	2.3	1.7	1.4	0.9	6.3
sortida a bars	2.5	0.6	0.4	0.2	3.7
estudiar/classes	2.7	0.4	0.2	0.1	3.4
Total	24.4	23.8	24.9	26.9	100

Indique:

a) ¿Qué porcentaje del colectivo observado son mayores de 65 años y la actividad realizada ha sido "leer"?
b) ¿Qué porcentaje del colectivo de mayores de 65 años ha dedicado su tiempo libre a "leer"?
c) ¿Qué porcentaje del colectivo observado tiene una edad inferior a 30 años y la actividad realizada ha sido "deporte"?
d) ¿Qué porcentaje del colectivo que ha realizado deporte tiene menos de 30 años?
e) ¿Qué porcentaje del colectivo observado tiene menos de 45 años y la actividad realizada ha sido "pasear " o "leer"?
f) Del colectivo de edad inferior a 45 ¿qué porcentaje ha "paseado" o "leido"?
g) ¿Cuál es la actividad más frecuente?

h) ¿Qué grupo de edad predomina en el colectivo observado?
i) ¿Cuál es la edad mediana?
j) Entre los que tienen edad inferior a 45, ¿cuál es la actividad menos preferida?
k) Las personas con menos de 30 años, ¿representan un mayor porcentaje en el colectivo de "playa-piscina" o en el colectivo de "deportes"?
l) Hacer deporte o pasear, ¿es más frecuente en el colectivo de edad inferior a 30 años o en el de edad superior a 65?

Ejercicio 3. La siguiente tabla de contingencia recoge información sobre la estructura de 100 familias elegidas al azar en cinco países de la UE.

	España	Francia	Italia	Portugal	Grecia
UP	16	33	25	17	22
ASN	57	44	52	53	54
UAN	9	4	3	4	6
DAN	18	19	20	26	18

UP: una persona; ASN: adultos sin niños; UAN: un adulto y niños; DAN: dos adultos y niños.

a) Analice si hay independencia entre estas dos categorías.
b) ¿Cuál es la proporción de familias UAN en el conjunto formado por España y Portugal?
c) ¿Qué porcentaje representan las familias ASN portuguesas sobre la muestra?
d) Indique cuales serían las frecuencias conjuntas (teóricas) bajo el supuesto de independencia y manteniéndose las frecuencias marginales.

Ejercicio 4. Se sabe que las variables M = "Máquina utilizada" y X = "Número de piezas defectuosas producidas en un día" son estadísticamente independientes.
Se ha observado un total de 400 días y se han obtenido las siguientes distribuciones marginales:

M	M1	M2	M3	M4
f(M)	0.15	0.15	0.30	0.40

X	0	1	2	3	4
f(X)	0.25	0.30	0.25	0.15	0.05

Indique:
a) Proporción de días en los que se han observado 3 piezas defectuosas.
b) Proporción de días en que se ha observado la máquina 2.
c) Número de días en los que se han observado 0 piezas defectuosas.
d) Número de días en que se ha observado la máquina 1.
e) Número de días en que se ha observado la máquina 4 y 3 piezas defectuosas.
f) Número de días en que se han observado 2 piezas defectuosas de la máquina 1.
g) Número de piezas defectuosas observado con mayor frecuencia en la máquina 3.

h) Número máximo de piezas defectuosas que presenta el 50% de los días con menor número de defectuosas observados en la máquina 3.
i) Proporción de días en que se ha observado la máquina 2 y 0 piezas defectuosas.
j) Del total de días en que se ha observado la máquina 2, ¿qué proporción ha habido con 0 piezas defectuosas?
k) Del total de días con 2 piezas defectuosas, ¿qué proporción corresponden a la máquina 3?
l) ¿En qué máquina se han observado 4 piezas defectuosas en un mayor número de días?

Resolución

Distribución de frecuencias conjuntas de M y X

	0	1	2	3	4	Total
M1	15 (0,0375)	18 (0,045)	15 (0,0375)	9 (0,0225)	3 (0,0075)	60 (0,15)
M2	15 (0,0375)	18 (0,045)	15 (0,0375)	9 (0,0225)	3 (0,0075)	60 (0,15)
M3	30 (0,075)	36 (0,09)	30 (0,075)	18 (0,045)	6 (0,015)	120 (0,30)
M4	40 (0,10)	48 (0,12)	40 (0,10)	24 (0,06)	8 (0,02)	160 (0,40)
Total	100 (0,25)	120 (0,30)	100 (0,25)	60 (0,15)	20 (0,05)	400 (1)

a) f(X=3) = 0,15
b) f(M=M2) =0,15
c) n(X=0) = 100 días
d) n(M= M1) = 60 días
e) n(M= M4, Y=3) = 24 días
f) n(X=2/M1) = 15 días
g) El número de piezas defectuosas observado con mayor frecuencia en M3 es 1
h) Para M3, el número máximo de piezas defectuosas del 50% de los días con menos piezas defectuosas es 1
i) f(M=M2, Y=0) =0,0375
j) f(X=0/M=M2) = 15/60 = 0,25
k) f(M=M3/Y=2) = 30/100 = 0,30
l) En la M4

Ejercicio 5. Para analizar la aceptación de dos nuevos modelos de motocicleta se ha observado durante los 25 días laborables del último mes las unidades vendidas en un concesionario:
X= Unidades vendidas del modelo A
Y= Unidades vendidas del modelo B

X (Modelo A)	Y (Modelo B)	Núm. días
0	3	1
1	1	5
2	1	10
3	2	9

Se pide:

a) Tabla de doble entrada.

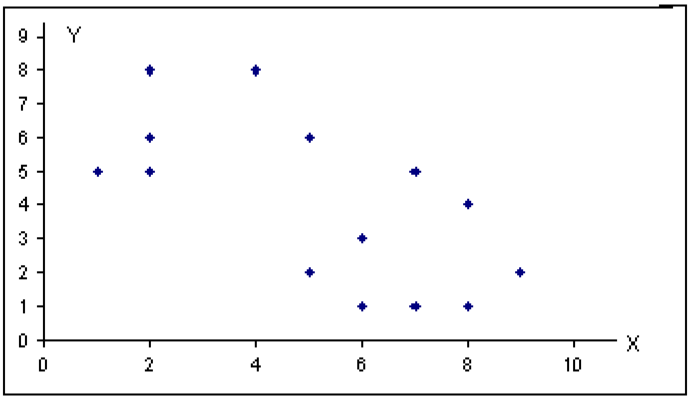
b) Total de unidades vendidas de cada modelo en el período observado.

Tema 6. ASOCIACIÓN ENTRE VARIABLES

6.1 Diagrama de dispersión
6.2 Asociación lineal. Covarianza
6.3 Coeficiente de correlación de Pearson
6.4 Regresión lineal

MEDIDAS DE ASOCIACIÓN LINEAL

Análisis Gràfico. Para analizar si existe alguna relación entre las variables (X, Y) es aconsejable realizar, en primer lugar, la representación gráfica o **Diagrama de Dispersión** o Nube de Puntos. Se construye representando cada elemento (xi, yi) por un punto en el plano de manera que sus coordenadas son los valores que toman las dos variables.



Análisis Cuantitativo: Las principales medidas de asociación lineal para datos cuantitativos son: **Covarianza** y **Coeficiente de Correlación Lineal de Pearson**.

COVARIANZA, S_{XY}

Indica si existe asociación lineal y su signo.

Si se calcula con distribución de frecuencias conjuntas no unitarias:

$$S_{XY} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{h} (X_i - \overline{X}).(Y_j - \overline{Y})\, n_{ij}}{n - 1}$$

Si se calcula con distribución de frecuencias conjuntas unitarias:

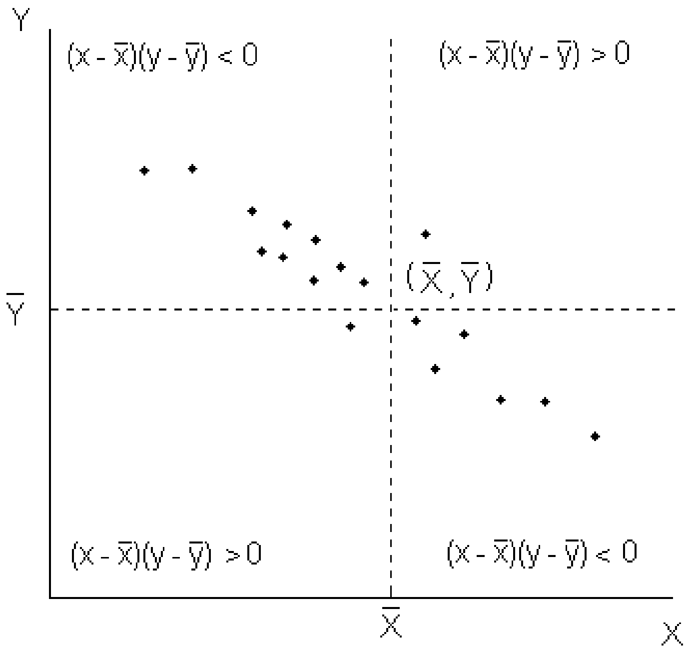
$$S_{XY} = \frac{\sum_{i=1}^{n} (X_i - \overline{X}).(Y_i - \overline{Y})}{n - 1}$$

Fórmula abreviada: $$S_{XY} = \frac{\sum_{i=1}^{n} X_i . Y_i - n\bar{X}\,\bar{Y}}{n - 1}$$

Propiedades:

- Puede tomar cualquier valor real ($-\infty \leq S_{XY} \leq +\infty$).
- Indica la presencia de asociación lineal y su signo:

$$S_{XY} = \begin{cases} < 0 & \text{Asociación lineal negativa} \\ = 0 & \text{No existe asociación lineal} \\ > 0 & \text{Asociación lineal positiva} \end{cases}$$



- Si dos variables X e Y son estadísticamente independientes, la covarianza es 0. Pero si $S_{XY} = 0$ **no** implica que las variables sean independientes.
- La covarianza queda afectada por los cambios de escala, pero no por los cambios de origen:

  $X' = a + bX \quad Y' = c + dY \quad \Rightarrow \quad S_{X'Y'} = bdS_{XY}$

  En consecuencia le afectan los cambios de unidades de medida.

Inconvenientes:

- No está acotada.
- No es adimensional. Tiene unidades de medida.

COEFICIENTE DE CORRELACIÓN LINEAL, r_{XY}

Su definición solventa los inconvenientes de la covarianza.

$$r_{XY} = \frac{S_{XY}}{S_X\, S_Y}$$

Indica el grado de asociación lineal que existe entre variables cuantitativas.

Propiedades:

- Está acotado: siempre toma valores comprendidos entre -1 y +1 ($-1 \leq r_{XY} \leq +1$).

$$r_{XY} = \begin{cases} = -1 & \text{Asociación lineal negativa perfecta} \\ = 0 & \text{No existe asociación lineal} \\ = +1 & \text{Asociación lineal positiva perfecta} \end{cases}$$

- Indica el grado de asociación lineal y su signo.

$$r_{XY} = \begin{cases} r_{XY} \cong \pm 1 & \text{Alto grado de asociación lineal} \\ r_{XY} \cong 0 & \text{Débil grado de asociación linea} \end{cases}$$

- Las transformaciones lineales sólo le afectan si hay cambio de signo en el cambio de escala.
- El coeficiente de correlación lineal coincide con la covarianza de las variables estandarizadas.

RESUMEN DEL ANÁLISIS DESCRIPTIVO BIDIMENSIONAL

- Vector de Medias: $\left[\bar{X}, \bar{Y}\right]$

  Centro de gravedad de la nube de puntos.

- Matriz de Varianzas y Covarianzas: $S^2 = \begin{bmatrix} S_X^2 & S_{XY} \\ S_{XY} & S_Y^2 \end{bmatrix}$

  Dispersión de la nube de puntos.

- Matriz de Coeficientes de Correlación: $r = \begin{bmatrix} 1 & r_{XY} \\ r_{XY} & 1 \end{bmatrix}$

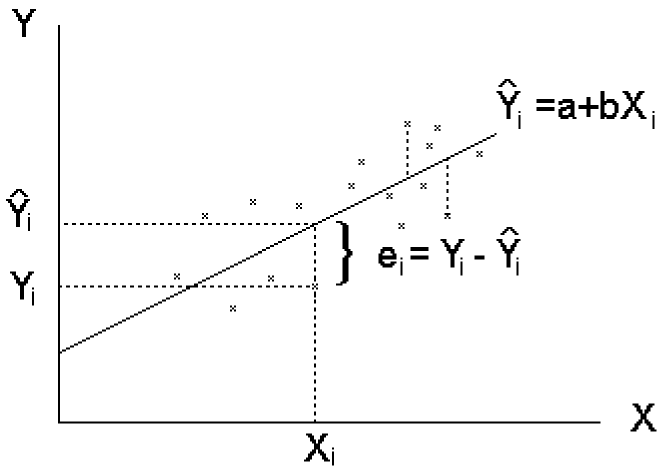
  Grado de asociación lineal.

REGRESIÓN LINEAL

La RECTA DE REGRESIÓN LINEAL (MRLS) modeliza la relación de causalidad de una variable Y con respecto a otra X, de forma que el comportamiento de una viene explicado de forma **lineal** por la otra.

X es la variable independiente o exógena que explica el comportamiento de Y, variable dependiente o endógena.

La recta de regresión lineal $\hat{Y}_i = a+bX_i$ será aquella que mejor se ajuste a la nube de puntos.



a: ordenada en el origen
b: pendiente de la recta
e_i: error de predicción o residuo

A estos efectos, utilizando el método de Mínimos Cuadrados Ordinarios (MCO) se obtienen los valores **a** y **b**, tales que determinan la recta que minimiza la variación o dispersión de las observaciones a su alrededor

$$\min \sum_{i=1}^{n} e_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$b = \frac{S_{XY}}{S_X^2}$$

$$a = \overline{Y} - b\overline{X}$$

Los valores $\hat{Y}_i$ recogidos en la recta son estimaciones de los promedios de Y para valores concretos de X.

El valor **b**, pendiente de la recta, recoge una estimación de la variación de la variable Y por cada incremento unitario de X.

El valor **a**, ordenada en el origen, recoge el valor ajustado de Y (estimación) suponiendo nulo el valor de X.

Características:

- la recta de regresión siempre pasa por el punto ($\bar{X}, \bar{Y}$).

- la pendiente (**b**) presenta el mismo signo que la covarianza entre X e Y (S_{XY}).

- el coeficiente de correlación lineal, rxy, está directamente relacionado con **b**: $r_{XY} = (S_X/S_Y)b$ y siempre presentan el mismo signo.

- la recta de regresión de Y sobre X (**Y = a+bX**), en general, no presenta la misma solución que la regresión de X sobre Y (**X = c+dY**).

COEFICIENTE DE DETERMINACIÓN

- Mide el grado de Bondad del Ajuste indicando si el modelo (la recta de regresión) es bueno para explicar la relación de causalidad entre las dos variables, es decir, si la variable independiente X explica el comportamiento de la variable dependiente Y.

- Se define como el cociente entre la variación de Y explicada por X (la recogida por la recta de regresión) y la variación total observada en Y: R^2 = VE/VT. Por lo tanto, cuantifica el tanto por uno de la variación observada en Y que queda explicada por la recta ajustada

- Toma valores acotados entre 0 y 1: $0 \leq R^2 \leq 1$.
  ▫ $R^2 = 1$ significa que el ajuste es perfecto (la nube de puntos está sobre la recta),
  ▫ $R^2 = 0$ entonces es que no existe relación lineal entre las dos variables. Es decir, X no explica de forma lineal el comportamiento de Y, por lo tanto el modelo especificado no es el adecuado.

- Se demuestra que $R^2 = r_{XY}^2$.

EJERCICIOS TEMA 6

Ejercicio1. A partir de los siguientes datos correspondientes a 8 empresas de USA sobre las ventas y los beneficios obtenidos en billones de dólares:

Empresa	ventas	beneficios
General Motors	78	1,25
Exxon	69	1,10
Ford	62	0,77
IBM	51	0,56
Mobil	44	0,45
General Electric	35	0,44
AT&T	34	0,44
Texaco	31	0,42

Se pide:

a) Obtenga las medidas de asociación.

b) Si las ventas de todas estas empresas se incrementan en un 5% y los beneficios en 0,5 billones de dólares, ¿cuál será la covarianza de las variables transformadas? ¿y el coeficiente de correlación lineal?

Ejercicio 2. Un concesionario, para analizar la aceptación de dos nuevos modelos de motocicleta ha observado durante los 25 días laborables del último mes las unidades vendidas:

X= Unidades vendidas del modelo A

Y= Unidades vendidas del modelo B

X Modelo A	Y Modelo B	Núm. días
0	3	1
1	1	5
2	1	10
3	2	9

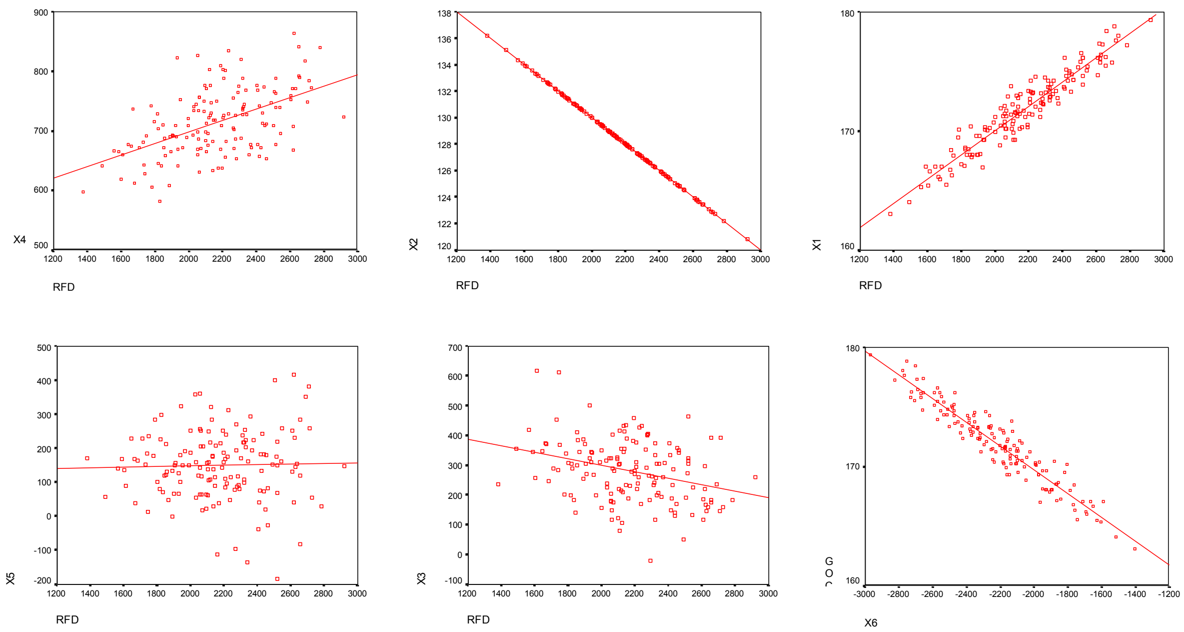
Se pide:

a) Vector de medias y matriz de varianzas y covarianza.

b) Coeficiente de correlación lineal y su interpretación.

c) Si el número de motos vendidas de cada modelo en cada uno de los días observados fuera el doble cuál sería la covarianza y el coeficiente de correlación entre las unidades vendidas de los dos modelos.

Ejercicio3. Los siguientes diagramas de dispersión corresponden a observaciones conjuntas de 6 pares de variables (X,Y):



Asigne a cada uno de los diagramas el valor del coeficiente de correlación que le parezca más adecuado: r=0,98; r=0,03; r=-0,42; r=-0,95; r=-1; r=0,32.

Ejercicio 4. A partir de una muestra de 100 observaciones referente a las variables:

Y = Saldo de las Imposiciones en Cajas de Ahorros

X = Renta Familiar Disponible

se han obtenido los siguientes resultados:

$$\overline{X} = 4{,}65 \quad \overline{Y} = 1{,}55 \qquad S_X^2 = 5{,}48 \qquad S_Y^2 = 1{,}04 \qquad S_{XY} = 2{,}13$$

Se pide:

a) Determine el grado de asociación lineal entre estas variables.

b) Obtenga la ecuación de regresión lineal que explica el Saldo de las Imposiciones en función de la Renta Familiar.

c) Indique el porcentaje de variación observado en Y explicado por el ajuste anterior

Ejercicio 5. La siguiente tabla recoge la edad (X) y la presión sanguínea máxima (Y) de un grupo de 10 mujeres:

Edad	56	42	72	36	63	47	55	49	38	42
Presión	14,8	12,6	15,9	11,8	14,9	13	15,1	14,2	11,4	14,1

a) Calcule el coeficiente de correlación lineal entre las variables anteriores y comente el resultado obtenido.

b) Determine la recta de regresión de Y sobre X justificando la adecuación de un ajuste lineal. Interprete los coeficientes.
c) Valore la bondad del ajuste.
d) Obtenga las siguientes predicciones, únicamente en los casos que tenga sentido hacerlo:
    Presión sanguínea de una mujer de 51 años.
    Presión sanguínea de una niña de 10 años.
    Presión sanguínea de un hombre de 54 años.

Ejercicio 6. A fin de analizar si la lluvia caída puede ser explicativa de la calidad del vino, se han observado en 60 muestras la calificación del vino en grados (Y) y la lluvia anual registrada (X), obteniendo los siguientes resultados:

$$\sum_{i=1}^{60} X_i = 32400 \quad \sum_{i=1}^{60} Y_i = 640 \quad \sum_{i=1}^{60} X_i^2 = 21280000$$

$$\sum_{i=1}^{60} Y_i^2 = 7088 \quad \sum_{i=1}^{60} X_i Y_i = 321600$$

a) Obtenga la regresión lineal que recoge el comportamiento de la calidad del vino en función de la lluvia.
b) Indique el porcentaje de variación observada en la graduación del vino que viene explicada por la recta de regresión obtenida.
c) ¿En cuánto se puede estimar la variación de la graduación del vino si la lluvia caída aumenta 10 unidades?
d) ¿Qué predicción haría de la graduación del vino obtenido de una cosecha en un año en que la lluvia registrada fuera de 950?

Ejercicio 7. El vector de medias y la matriz de varianzas y covarianzas de las variables X_1 = Ingreso semanal y X_2 = Gasto semanal (en u.m.) observadas sobre un conjunto de 100 familias son:

$$\bar{X} = \begin{bmatrix} 1250 & 1050 \end{bmatrix}' \qquad S^2 = \begin{bmatrix} 1200 & 900 \\ & 4000 \end{bmatrix}$$

a) Calcule el coeficiente de correlación lineal.
b) Obtenga el ajuste lineal del Gasto semanal en función del Ingreso.
c) ¿Qué porcentaje de variación del Gasto no queda explicado por el ajuste anterior?
d) ¿Qué predicción del Gasto semanal haría para una familia con 1100 u.m. de Ingreso semanal? Comente la validez del resultado anterior.

Resolución

Coeficiente de correlación lineal $r_{12} = \dfrac{90000}{\sqrt{(120000)(400000)}} = 0{,}41$

Ajuste lineal de X2 = Gasto sobre X1 = Ingreso

$$b = \frac{S_{12}}{S_1^2} = \frac{90000}{120000} = 0{,}75 \qquad a = \bar{X}_2 - b\bar{X}_1 = 10500 - (0{,}75)(12500) = 1125$$

$\hat{X}_2 = 1125 + 0{,}75X1$

Bondad del ajuste: R2 = 0,412 = 0,1681

El porcentaje no explicado de la variación total del Gasto es del 83,2%

Predicción para X1 = 11000

$\hat{X}_2 = 1125 + 0{,}75(11000) = 9375$ u.m.

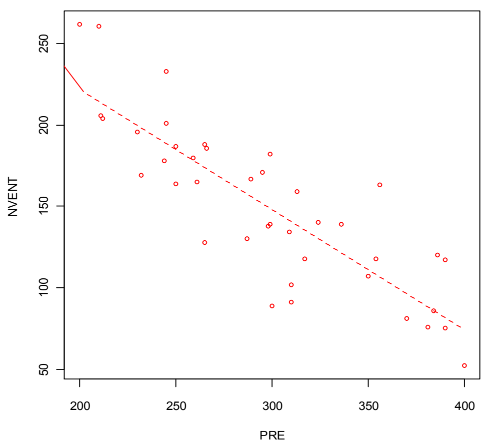
Ejercicio 8. Suponga que ha observado simultáneamente el precio X y la demanda Y de un determinado producto, obteniendo:

X	1	2	3	4	5	6	7	10	10	12
Y	12	10	10	8	6	5	7	6	4	2

$$\sum_{i=1}^{10}(X_i - \bar{X})^2 = 124 \qquad \sum_{i=1}^{10}(Y_i - \bar{Y})^2 = 84 \qquad \sum_{i=1}^{10}(X_i - \bar{X})(Y_i - \bar{Y}) = -93$$

a) Obtenga el ajuste lineal de la demanda en función del precio.
b) Determine la bondad del ajuste.
c) Compruebe los resultados con los obtenidos con el R-Commander.
d) Si el precio se incrementa en 1 unidad, ¿cuál es el incremento esperado en la demanda?
e) Obtenga la predicción de la demanda esperada si el precio es 9.

Ejercicio 9. El siguiente diagrama de dispersión corresponde a la distribución de frecuencias conjuntas de las variables NVENT = "Nº de unidades vendidas mensualmente de determinado artículo" y PRE = "precio del artículo en Euros·".



Razone que signo y magnitud aproximada presentarán la pendiente y el coeficiente de determinación de la recta de regresión ajustada.

Ejercicio 10. Sobre una muestra de 61 pisos vendidos en el área metropolitana de Barcelona durante el último trimestre del año 2011 se han observado las variables:
Preu.de.venda = precio pagado por el comprador (Euros)
Preu.inicial = primer precio ofrecido al comprador (Euros)
Metres.quadrats = superficie en m^2

Con el programa R-Commander se han obtenido los siguientes resultados:
Análisis descriptivo unidimensional:

```
                        mean          sd  n
Metres.quadrats     65.06721    12.13026 61
Preu.de.venda   170109.67803 35628.50863 61
Preu.inicial    181766.75115 36727.83416 61
```

Matriz de correlación:

```
                Metres.quadrats Preu.de.venda Preu.inicial
Metres.quadrats       1.0000000     0.9510897    0.9668259
Preu.de.venda         0.9510897     1.0000000    0.9828272
Preu.inicial          0.9668259     0.9828272    1.0000000
```

Para explicar el comportamiento de la variable Preu.de.venda se proponen las siguientes rectas de regresión lineal:

Recta I

```
Coefficients:
                Estimate
(Intercept)     -11655.7
Metres.quadrats   2793.5
```

Recta II

```
Coefficients:
              Estimate
(Intercept) -3188.4840
Preu.inicial    0.9534
```

Se pide:

a) Razone porqué la recta II tiene mayor capacidad explicativa del comportamiento de Preu.de.venda que la recta I e indique qué porcentaje de la variación total de Preu.de.venda queda explicado por esta recta.
b) Con respecto a la recta II, indique en cuánto repercute un aumento de 1000 Euros en el Preu.inicial sobre el Preu.de.venda esperado.
c) Estime cuál será el Preu.de.venda de un piso cuyo precio inicial es de 150000 Euros.
d) Estime cuál será el Preu.de.venda de un piso cuyo precio inicial es de 270000 Euros.
e) ¿Cuál de las dos predicciones es más fiable? Razone la respuesta.

Tema 7. INTRODUCCIÓN A LA TEORÍA DE LA PROBABILIDAD

7.1 Experimento aleatorio. Probabilidad: axiomática y propiedades
7.2 Probabilidad condicionada
7.3 Teorema de la intersección. Independencia de sucesos
7.4 Teorema de la probabilidad total. Teorema de Bayes

Espacio muestral

El espacio muestral o espacio referencial, E, es el conjunto de todos los resultados posibles de un experimento aleatorio.

Suceso aleatorio

Dado un espacio referencial, se define como suceso aleatorio cualquier subconjunto de dicho espacio referencial. El suceso que contiene un solo resultado de E se llama suceso elemental o elemento muestral.

Ejemplo.- Una caja contiene 1 bola blanca (B), 1 bola roja (R) y 1 bola negra (N). El experimento consiste en extraer dos bolas con devolución y observar la secuencia de colores obtenida. El espacio muestral o referencial es:

E = {BB, BR, BN, RB, RR, RN, NB, NR, NN}
Sobre este espacio pueden definirse, entre otros, los siguientes sucesos:
A_1 = 'Dos bolas del mismo color' = {BB, RR, NN}
A_2 = 'Dos bolas de distinto color' = {BR, BN, RB, RN, NB, NR}
A_3 = 'Por lo menos una bola blanca' = {BB, BR, BN, RB, NB}
A_4 = 'Exactamente una bola blanca' = {BR, BN, RB, NB}
A_5 = 'Ninguna bola blanca' = {RR, RN, NR, NN}
A_6 = 'Dos bolas blancas' = {BB}
A_7 = 'O dos bolas blancas o ninguna blanca' = {BB, RR, RN, NR, NN}
A_8 = 'Como máximo una bola blanca' = {BR, BN, RB, RR, RN, NB, NR, NN}
A_9 = 'Por lo menos una blanca o dos del mismo color' = {BB, BR, BN, RB, RR, NB, NN}

Relaciones entre sucesos

I) **Suceso complementario**- Dado un suceso A, el suceso complementario de A, $\overline{A}$, es aquel que contiene todos los resultados del experimento que no están contenidos en A; es decir, $\overline{A}$ es el suceso que ocurre cuando no ocurre A.

Por ejemplo:

$\overline{A}_1$ = {BR, BN, RB, RN, NB, NR} = A_2 $\qquad$ $\overline{A}_2 = A_1$
$\overline{A}_3$ = {RR, RN, NR, NN} = A_5 $\qquad$ $\overline{A}_5 = A_3$
$\overline{A}_4$ = {BB, RR, RN, NR, NN} = A_7 $\qquad$ $\overline{A}_7 = A_4$
$\overline{A}_6$ = {BR, BN, RB, RR, RN, NB, NR, NN}= A_8 $\qquad$ $\overline{A}_8 = A_6$

II) **Suceso unión**- Dados dos sucesos A_i y A_j el suceso unión $(A_i \cup A_j)$ es aquel que contiene todos los resultados que pertenecen a A_i, a A_j o a ambos.

Por ejemplo:
$(A_1 \cup A_3)$= {BB, BR, BN, RB, RR, NB, NN} = A_9
$(A_1 \cup A_5)$= {BB, RR, RN, NR, NN} = A_7
$(A_4 \cup A_6)$= {BB, BR, BN, RB, NB} = A_3

- Si $(A_i \cup A_j)$= A_i se dice que A_j está contenido en A_i; es decir todos los resultados que pertenecen a A_j pertenecen también a A_i de forma que si ocurre A_j ocurre también A_i.

Por ejemplo: $(A_1 \cup A_6)$= {BB, RR, NN} = A_1 A_6 está contenido en A_1 $(A_6 \subset A_1)$

- La unión de un suceso cualquiera A y su complementario $\bar{A}$ es E.

Por ejemplo: $(A_1 \cup A_2)$= {BB, RR, NN, BR, BN, RB, RN, NB, NR}= E

- La unión puede generalizarse para más de dos sucesos;

Por ejemplo: $(A_1 \cup A_5 \cup A_7)$= {BB, RR, RN, NR, NN}

III) **Suceso intersección**- Dados dos sucesos A_i y A_j el suceso intersección $(A_i \cap A_j)$ es aquel que contiene todos los resultados que pertenecen a A_i y a A_j simultáneamente.

Por ejemplo:
$(A_1 \cap A_3)$ = {BB}
$(A_1 \cap A_5)$ = {RR, NN}
$(A_1 \cap A_7)$ = {BB, RR, NN}
$(A_5 \cap A_7)$ = {RR, RN, NR, NN}

- Si un suceso A_j está contenido en A_i, $(A_i \cap A_j) = A_j$.

Por ejemplo:
A_5 está contenido en A_8 $(A_5 \subset A_8) \Rightarrow (A_5 \cap A_8)$ = {RR, RN, NR, NN} = A_5

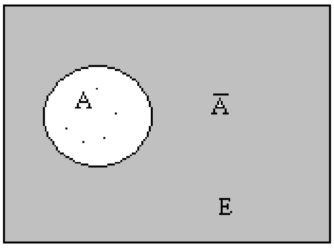
- Cuando dos sucesos A_i y A_j son tales que $(A_i \cap A_j) = \phi$ se dice que A_i y A_j son sucesos incompatibles o **mutuamente excluyentes**.

Por ejemplo:
$(A_1 \cap A_4) = \phi$ A_1 y A_4 son incompatibles
$(A_4 \cap A_6) = \phi$ A_4 y A_6 son incompatibles

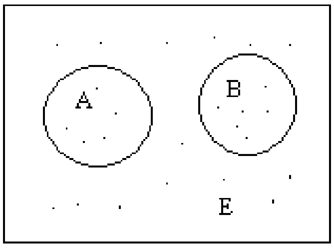
- La intersección de un suceso A y su complementario $\bar{A}$ es igual al conjunto vacío, por tanto A y $\bar{A}$ son sucesos incompatibles.
- La intersección puede generalizarse para más de dos sucesos.

Por ejemplo: $(A_1 \cap A_5 \cap A_7)$ = {RR, NN}

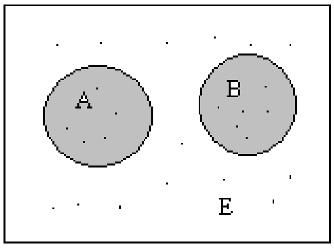
Diagramas de Venn



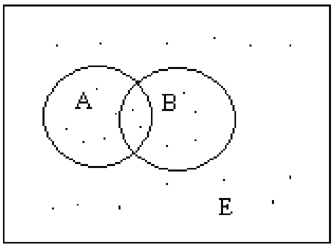
El suceso A y su complementario



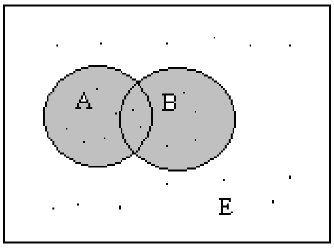
A y B son sucesos incompatibles



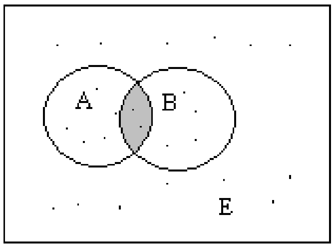
Unión de sucesos incompatibles



A y B son sucesos compatibles



Unión de sucesos compatibles



Intersección de A y B

Probabilidad. Axiomática y propiedades

Dado un espacio referencial E se dice que P es una función de probabilidad definida en E si:

- a cualquier suceso, A, de E le hace corresponder un número real, P(A), que mide el grado de posibilidad de la ocurrencia de A, y
- verifica los siguientes axiomas:

Axioma I $P(A) \geq 0$

Axioma II $P(E) = 1$

Axioma III Si A_1, A_2, A_3, ... es una sucesión numerable de sucesos mutuamente excluyentes, la probabilidad de su unión es:
$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots.$

De los axiomas se deducen las siguientes propiedades:

1. $P(\phi) = 0$

2. $0 \leq P(A) \leq 1 \qquad \forall A \in E$

3. $P(\bar{A}) = 1 - P(A)$

4. Si $A \subset B$, entonces $P(A) \leq P(B)$

5. Ley aditiva: si A y B son dos sucesos cualesquiera, la probabilidad de su unión es: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
   Esta propiedad puede generalizarse a tres o más sucesos, por ejemplo:
   $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Asignación de probabilidad a un suceso.

1. Teoría Clásica: Regla de Laplace

Si un espacio referencial, E, contiene un número finito de resultados y éstos son igualmente probables, la probabilidad de un suceso cualquiera A es:

$$P(A) = \frac{\text{Resultados favorables a A}}{\text{Total de resultados posibles}}$$

2. Teoría Frecuencialista

La probabilidad de un suceso A es el límite de la frecuencia relativa de A (n_A/n) cuando el número de experimentos tiende a infinito.

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

3. Teoría Subjetivista

Esta teoría interpreta la probabilidad como el grado de convencimiento subjetivo que cada individuo puede tener en relación a la ocurrencia de un determinado suceso.

Ejemplo: Determinación de la probabilidad de un suceso. Regla de Laplace. Se han definido los siguientes sucesos sobre el espacio muestral
E= {BB, BR, BN, RB, RR, RN, NB, NR, NN}

A_1 = {BB, RR, NN}	$P(A_1) = 3/9$
A_2 = {BR, BN, RB, RN, NB, NR}	$P(A_2) = 6/9$
A_3 = {BB, BR, BN, RB, NB}	$P(A_3) = 5/9$
A_4 = {BR, BN, RB, NB}	$P(A_4) = 4/9$
A_5 = {RR, RN, NR, NN}	$P(A_5) = 4/9$
A_6 = {BB}	$P(A_6) = 1/9$
A_7 = {BB, RR, RN, NR, NN}	$P(A_7) = 5/9$
A_8 = {BR, BN, RB, RR, RN, NB, NR, NN}	$P(A_8) = 8/9$
A_9 = {BB, BR, BN, RB, RR, NB, NN}	$P(A_9) = 7/9$
$(A_1 \cap A_3)$ = {BB}	$P(A_1 \cap A_3) = 1/9$
$(A_1 \cap A_5)$ = {RR, NN}	$P(A_1 \cap A_5) = 2/9$

$(A_1 \cap A_7) = \{BB, RR, NN\}$ $\qquad P(A_7 \cap A_8) = 3/9$

$(A_5 \cap A_7) = \{RR, RN, NR, NN\}$ $\qquad P(A_5 \cap A_7) = 4/9$

$(A_1 \cap A_4) = \phi$ $\qquad P(A_1 \cap A_4) = 0$

$(A_4 \cap A_6) = \phi$ $\qquad P(A_4 \cap A_6) = 0$

$(A_1 \cap A_5 \cap A_7) = \{RR, NN\}$ $\qquad P(A_1 \cap A_5 \cap A_7) = 2/9$

Probabilidad de la unión de sucesos

$(A_1 \cup A_3) = \{BB, BR, BN, RB, RR, NB, NN\}$ $\qquad P(A_1 \cup A_3) = 7/9$

A_1 y A_3 no son incompatibles. Por la propiedad 3:

$$P(A_1 \cup A_3) = P(A_1) + P(A_3) - P(A_1 \cap A_3) = \frac{3}{9} + \frac{5}{9} - \frac{1}{9} = \frac{7}{9}$$

$(A_1 \cup A_5) = \{BB, RR, RN, NR, NN\}$ $\qquad P(A_1 \cup A_5) = 5/9$

A_1 y A_5 no son incompatibles. Por la propiedad 3:

$$P(A_1 \cup A_5) = P(A_1) + P(A_5) - P(A_1 \cap A_5) = \frac{3}{9} + \frac{4}{9} - \frac{2}{9} = \frac{5}{9}$$

$(A_4 \cup A_6) = \{BB, BR, BN, RB, NB\}$ $\qquad P(A_4 \cup A_6) = 5/9$

A_4 y A_6 son incompatibles. Por el axioma 3:

$$P(A_4 \cup A_6) = P(A_4) + P(A_6) = \frac{4}{9} + \frac{1}{9} = \frac{5}{9}$$

$(A_1 \cup A_5 \cup A_7) = \{BB, RR, RN, NR, NN\}$ $\qquad P(A_1 \cup A_5 \cup A_7) = 5/9$

Por la propiedad 3:

$$P(A_1 \cup A_5 \cup A_7) = P(A_1)+P(A_5)+P(A_7)-P(A_1 \cap A_5)-P(A_1 \cap A_7)-$$
$$P(A_5 \cap A_7)+P(A_1 \cap A_5 \cap A_7) = \frac{3}{9} + \frac{4}{9} + \frac{5}{9} - \frac{2}{9} - \frac{3}{9} - \frac{4}{9} + \frac{2}{9} = \frac{5}{9}$$

Probabilidad del suceso complementario

$\overline{A}_1 = \{BR, BN, RB, RN, NB, NR\}$ $\qquad P(\overline{A}_1) = 6/9$

$\overline{A}_3 = \{RR, RN, NR, NN\}$ $\qquad P(\overline{A}_3) = 4/9$

$\overline{A}_4 = \{BB, RR, RN, NR, NN\}$ $\qquad P(\overline{A}_4) = 5/9$

$\overline{A}_6 = \{BR, BN, RB, RR, RN, NB, NR, NN\}$ $\qquad P(\overline{A}_6) = 8/9$

Por la propiedad 2:

$P(\overline{A}_1) = 1 - P(A_1) = 1 - \frac{3}{9} = \frac{6}{9}$ $\qquad P(\overline{A}_3) = 1 - P(A_3) = 1 - \frac{5}{9} = \frac{4}{9}$

$P(\overline{A}_4) = 1 - P(A_4) = 1 - \frac{4}{9} = \frac{5}{9}$ $\qquad P(\overline{A}_6) = 1 - P(A_6) = 1 - \frac{1}{9} = \frac{8}{9}$

Probabilidad condicionada

Dados dos sucesos, A y B, de un mismo espacio muestral, con probabilidades no nulas, P(A) ≠ 0 y P(B) ≠ 0, la probabilidad de que ocurra el suceso A sabiendo que ha ocurrido el suceso B recibe el nombre de probabilidad condicionada de A respecto a B y se define como:

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

Análogamente, la probabilidad condicionada de B respecto a A es:

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

Teorema de la intersección

De la definición de probabilidad condicionada se deduce que la probabilidad del suceso $(A \cap B)$ es:
$P(A \cap B) = P(A)\ P(B/A)$
o
$P(B \cap A) = P(B)\ P(A/B)$

El teorema se puede generalizar para n sucesos:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\ P(A_2/A_1)\ P(A_3/A_1 \cap A_2)\ \ldots .\ P(A_n/\bigcap_{i=1}^{n-1} A_i)$$

Independencia de sucesos

Dos sucesos, A y B, son estocásticamente independientes si la ocurrencia de uno de ellos no modifica la probabilidad de ocurrencia del otro; es decir, si

$P(A/B) = P(A)$
o
$P(B/A) = P(B)$

De donde se deduce que:
si A y B son independientes $\Rightarrow$ $P(A \cap B) = P(A)\ P(B)$

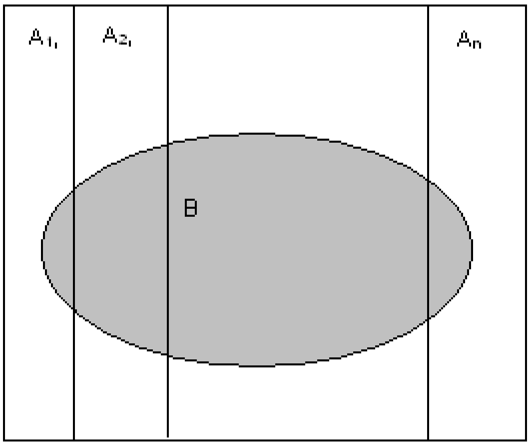
Propiedades:

- Si dos sucesos A y B son independientes, **no** son incompatibles.
- Si dos sucesos A y B son independientes también son independientes $\overline{A}$ y B; A y $\overline{B}$; $\overline{A}$ y $\overline{B}$.

Teorema de la probabilidad total

Si $A_1, A_2, ..., A_n$ son sucesos de E con probabilidades no nulas ($P(A_i) \neq 0 \; \forall i$), tales que forman una partición, es decir:

- $A_1 \cup A_2 \cup ... \cup A_n = E$ (exhaustivos)
- $A_i \cap A_j = \phi \quad \forall i \neq j$ (incompatibles)



y B es un suceso de E que ocurre siempre acompañado con uno de los A_i, la probabilidad de B es:

$P(B) = P[(A_1 \cap B) \cup (A_2 \cap B) \cup ... \cup (A_n \cap B)] = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) +$

$+ .... + P(A_n)P(B/A_n)$

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$$

Teorema de Bayes

Sean los sucesos $A_1, A_2, ..., A_n$ una partición de E. Asignadas unas probabilidades iniciales $P(A_1), P(A_2), ..., P(A_n)$, o probabilidades a priori que reflejan el grado de creencia sobre la ocurrencia de $A_1, A_2, ..., A_n$, la realización de un experimento en dicho espacio referencial puede modificar las probabilidades asignadas a priori.

Si el resultado del experimento es el suceso B y se conoce la ocurrencia de B en los sucesos A_i de la partición, es decir, $P(B/A_i)$, esta evidencia experimental permite obtener las nuevas probabilidades de los A_k condicionadas al resultado B, es decir, $P(A_k/B)$ que se denominan probabilidades a posteriori.

$$P(A_k/B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B/A_k)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

El teorema de Bayes formula que:

$$P(A_k/B) = P(A_k)\frac{P(B/A_k)}{P(B)}$$

la probabilidad a posteriori es igual a la probabilidad a priori multiplicada por un factor modificativo que depende del resultado del experimento.

EJERCICIOS TEMA 7

Ejercicio 1. Indique cuál es el espacio muestral asociado a los siguientes experimentos aleatorios:

a) De una población con N=7 elementos se extrae una muestra de tamaño n=2 sin devolución.
b) Se observa el número de clientes atendidos por un vendedor hasta que realiza la primera venta.
c) Se observa cada hora la temperatura de una cámara frigorífica que se mantiene a una temperatura de entre 0º y 3º
d) Se elige al azar un aparato eléctrico y se observa el tiempo que transcurre hasta la primera avería.
e) Se echan 5 bolas en 2 cajas, de modo que cada bola tenga la misma probabilidad de caer en cualquiera de las cajas, y se observa el número de bolas que caen en cada caja.

Ejercicio 2. Determine el valor de la probabilidad en cada una de las siguientes situaciones e indique el criterio de asignación utilizado en cada caso:

a) Probabilidad de que un negocio de alimentación tenga éxito si, en términos generales, se estima que por cada negocio de este tipo que fracasa, tres tienen éxito.
b) Probabilidad de obtener una carta de copas al extraer al azar una carta de una baraja española de 48 cartas.
c) Probabilidad de obtener una puntuación total superior a 7 al tirar dos dados.
d) Probabilidad de accidente laboral en un sector industrial sabiendo que en una muestra de 8000 trabajadores, 40 han sufrido un accidente.
e) Probabilidad de que llueva el próximo fin de semana sabiendo que sobre las Azores hay una borrasca.
f) Probabilidad de que llegue con retraso un vuelo del puente aéreo si se sabe que de los 80 vuelos semanales en promedio sólo llegan 2 con retraso.
g) Probabilidad de obtener un número primo al lanzar un dado.
h) Se echan 5 bolas en 2 cajas, de modo que cada bola tenga la misma probabilidad de caer en cualquiera de las cajas, la probabilidad de que en la primera caja caigan exactamente 3 bolas.

Ejercicio 3. A partir de una encuesta sobre hábitos de lectura de la prensa diaria realizada sobre una muestra de 1000 personas se obtiene información sobre los siguientes sucesos:

A = ‘lee el diario A’
B = ‘lee el diario B’
C = ‘lee el diario C’

Resulta que del total de entrevistados:

El 62% lee A
El 92% lee B
El 11% lee C
El 60% lee A y B
El 6% lee A y C
El 11% lee B y C
El 6% lee los tres diarios

a) ¿Son incompatibles los sucesos 'leer el diario A' y 'leer el diario B'?

Si se elige al azar un entrevistado, halle:

b) Probabilidad de que no lea A.
c) Probabilidad de que lea A o B.
d) Probabilidad de que lea B o C.
e) Probabilidad de que no lea ni A ni B.
f) Probabilidad de que lea por lo menos uno de los tres diarios.
g) Probabilidad de que no lea ninguno de los tres diarios.
h) Probabilidad de que lea B y no lea C.

Ejercicio 4. Al lanzar dos veces un dado:

a) ¿Cuál es la probabilidad de que la puntuación total sea 8 si se sabe que los dos resultados son diferentes?
b) ¿Cuál es la probabilidad de obtener como máximo 4 puntos si se sabe que ha salido un 2?
c) ¿Cuál es la probabilidad de que haya salido un 2 si se sabe que se ha obtenido más de 4 puntos?

Ejercicio 5. Se sabe que la probabilidad de que se atasque el papel en unas determinadas máquinas fotocopiadoras depende del color del papel. La probabilidad de que la fotocopia se haga en papel blanco y se atasque es 0,05; y la de hacerla en papel de color y atascarse es 0,10. Si el 80% de les fotocopies se hacen con papel blanco:

a) Determine todas las probabilidades conjuntas y marginales de los sucesos 'atascar (si/no)' y 'papel (blanco/color)'.
b) ¿Cuál es la probabilidad de que se atasque una fotocopia que se ha hecho con papel blanco?
c) ¿Cuál es la probabilidad de que el papel sea de color si la fotocopia se ha atascado?
d) Si el papel es de color, ¿cuál es la probabilidad de que la fotocopia no se atasque?

Ejercicio 6. Sobre determinado colectivo se sabe que:

- La edad del 25% es menor de 30 años.
- El 60% de los que tienen menos de 30 años practica algún deporte.
- El 48% de los que tienen 30 o más años no practica ningún deporte.

Se elige al azar a una persona de este colectivo. Se pide la probabilidad de que:

a) Sea joven (menos de 30 años) y practique algún deporte.
b) Sea joven y no practique ningún deporte.

c) Tenga 30 o más años y practique algún deporte.
d) Tenga 30 o más años y no practique ningún deporte.
e) Practique algún deporte.
f) No practique ningún deporte.
g) Si la persona elegida practica algún deporte, ¿cuál es la probabilidad de que sea joven?
h) Si la persona elegida no practica ningún deporte, ¿cuál es la probabilidad de que sea joven?

Resolución

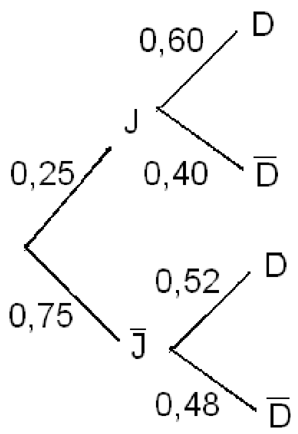
Sean los sucesos J = Menos de 30 años $\overline{J}$ = 30 años o más
D = Practica deporte $\overline{D}$ = No practica deporte

Se tiene:
P(J) = 0,25 P(D/J) = 0,60 y $P(\overline{D}/\overline{J}) = 0{,}48$

Entonces:

P(J) = 0,25 P(D/J) = 0,60 $P(\overline{D}/J) = 0{,}40$

$P(\overline{J}) = 0{,}75$ $P(\overline{D}/\overline{J}) = 0{,}48$ $P(D/\overline{J}) = 0{,}52$



a) $P(J \cap D) = P(J)\,P(D/J) = (0{,}25)(0{,}60) = 0{,}15$

b) $P(J \cap \overline{D}) = P(J)\,P(\overline{D}/J) = (0{,}25)(0{,}40) = 0{,}10$

c) $P(\overline{J} \cap D) = P(\overline{J})P(D/\overline{J}) = (0{,}75)(0{,}52) = 0{,}39$

d) $P(\overline{J} \cap \overline{D}) = P(\overline{J})P(\overline{D}/\overline{J}) = (0{,}75)(0{,}52) = 0{,}36$

e) $P(D) = P(J \cap D) + P(\overline{J} \cap D) = P(J)PD/J) + P(\overline{J})P(D/\overline{J}) = (0{,}25)(0{,}60) + (0{,}75)(0{,}52) = 0{,}54$

f) $P(\overline{D}) = 1 - P(D) = 1 - 0{,}54 = 0{,}46$

g) $P(J/D) = \dfrac{P(J \cap D)}{P(D)} = \dfrac{P(J)P(D/J)}{P(D)} = \dfrac{(0{,}25)(0{,}60)}{0{,}54} = 0{,}2778$

h) $P(J/\bar{D}) = \dfrac{P(J \cap \bar{D})}{P(\bar{D})} = \dfrac{P(J)P(\bar{D}/J)}{P(\bar{D})} = \dfrac{(0{,}25)(0{,}40)}{0{,}46} = 0{,}2174$

Ejercicio 7. Una empresa compra el 80% de ciertas piezas a un proveedor que le entrega el género con retraso el 10% de les veces. A su vez, por motivos de calidad, la empresa devuelve el 20% de las partidas de este proveedor que llegan con retraso. ¿Cuál es la probabilidad de que una partida de estas piezas haya sido comprada a este proveedor, haya llegado con retraso y se haya tenido que devolver?

Ejercicio 8. Una prueba de selección de personal consta de dos partes: la primera consiste en un test psicotécnico y la segunda de una entrevista personal. Si de 100 persones el 60% supera el test y de estas últimas el 30% supera la entrevista:

a) ¿Cuál es la probabilidad de que una de estas 100 persones, escogida al azar, haya superado les dos pruebas?
b) ¿Cuál es la probabilidad de que no haya superado la entrevista pero sí el test?

Ejercicio 9. En cierta localidad la probabilidad de que una persona compre un diario es 0,4; la probabilidad de que compre una revista 0,2 y la probabilidad de que compre ambos 0,08.

a) ¿Cuál es la probabilidad de que compre alguna de estas 2 publicaciones?
b) Comprar un diario y comprar una revista, ¿son sucesos mutuamente excluyentes?
c) Comprar un diario y comprar una revista, ¿son sucesos independientes?

Ejercicio 10. Con el enunciado del ejercicio 3 determine:

a) ¿Son independientes los sucesos leer el diario B y leer el diario C?
b) Si sabemos que una persona lee el diario A, ¿la probabilidad de que lea B queda modificada?
c) ¿Cuál es la probabilidad de que un lector de B y C lea también A?
d) Sabiendo que un entrevistado lee B, ¿cuál es la probabilidad de que lea A?
e) Sabiendo que un entrevistado lee por lo menos uno de los tres diarios, ¿cuál es la probabilidad de que lea A?

Ejercicio 11. Las probabilidades de que dos personas, que actúan independientemente, lleguen a tiempo a coger el tren se estiman en 0,9 y 0,8, respectivamente. Halle la probabilidad de que:

a) Lleguen ambas.
b) No llegue ninguna.
c) Sólo llegue una.

Ejercicio 12. Un recién graduado que ha solicitado empleo en dos compañías, A y B, estima que tiene doble probabilidad de ser contratado por A que por B y que la probabilidad de que no lo contraten ninguna de las dos es 0,28. Si las decisiones de las dos compañías son independientes, ¿cuál es la probabilidad de que lo contrate A?

Ejercicio 13. Un comercio vende 3 modelos de lector de DVD: V1, V2 y V3. Las probabilidades de venta de cada uno de ellos son: P(V1)=0,3 P(V2)=0,2 y P(V3)=0,5. La probabilidad de avería durante el período de garantía para cada uno de ellos son: 0,10; 0,15 y 0,04 respectivamente. Si un cliente devuelve un aparato averiado en el período de garantía, ¿de qué modelo (V1, V2 o V3) hay más probabilidad de que sea dicho aparato?

Ejercicio 14. Se tienen cuatro cajas idénticas que contienen cuatro bolígrafos cada una. La caja C_1 contiene 4 bolígrafos negros; la caja C_2 contiene 3 negros y 1 rojo; la caja C_3 contiene 2 negros y 2 rojos y la caja C_4 contiene 1 negro y 3 rojos. Se elige una caja al azar y se extrae al azar un bolígrafo.

a) Probabilidad de que el bolígrafo extraído sea negro.
b) Probabilidad de que el bolígrafo extraído sea rojo.
c) Halle las probabilidades *a posteriori* de cada una de las cuatro cajas si se sabe que el bolígrafo extraído es negro.
d) Halle las probabilidades *a posteriori* de cada una de las cuatro cajas si se sabe que el bolígrafo extraído es rojo.

Ejercicio 15. Tres contratistas, A, B y C, licitan por un contrato para construir un polideportivo. Se cree que la probabilidad de obtener el contrato es la misma para cada uno de ellos. La probabilidad de que finalicen las obras en la fecha prevista es 0,85 si el contrato lo consigue A, 0,55 si lo consigue B y 0,30 si lo consigue C. Sabiendo que las obras no han finalizado en la fecha prevista, ¿cuál es la probabilidad de que el contrato lo haya conseguido B?

Ejercicio 16. Respecto al medio de transporte se sabe que el 50% de los alumnos de la Universidad de Barcelona que residen fuera de Barcelona utiliza el tren para ir a clase; mientras que de los que residen en Barcelona no hay ninguno que lo utilice. Sabiendo que el 40% de los alumnos de la UB reside fuera de Barcelona, ¿cuál es la probabilidad de que un alumno elegido al azar entre los que no utilizan el tren resida fuera de Barcelona?

Ejercicio 17. Una entidad financiera ha comprobado que el 5% de los clientes con suficiente saldo se equivocan en la fecha del cheque y el 100% de los que no tienen suficiente saldo cometen el mismo error. Si un 85% de los clientes de esta entidad tienen saldo suficiente, ¿cuál es la probabilidad de que al recibir un cheque con fecha equivocada corresponda a un cliente sin saldo?

Ejercicio 18. Una encuesta contiene las preguntas:
1. ¿Es par el último dígito de su DNI? y
2. ¿Cree que los coches que circulan por la ciudad deben pagar un peaje?

Las respuestas posibles son "sí" o "no".
Se les pide a los encuestados que contesten sólo la primera pregunta si al lanzar una moneda sale cara y sólo a la segunda si sale cruz. Si el 40% de los encuestados ha respondido "sí", ¿cuál es la probabilidad de que un encuestado que ha contestado la segunda pregunta haya respondido "sí"?

Ejercicio 19. En un laboratorio se sabe por experiencia que 1 de cada 25 frascos de cierto medicamento se deterioran al cabo de un mes, por lo que las existencias se someten a controles mensuales. El test utilizado para determinar si un frasco está deteriorado da positivo el 99% de las veces si el frasco está deteriorado y el 2% de las veces cuando el frasco no está deteriorado. Hallar:

a) Probabilidad de que el test dé positivo.
b) Probabilidad de que el test dé negativo.
c) Si el test resulta positivo, ¿cuál es la probabilidad de que la muestra esté deteriorada?
d) Si el test resulta negativo, ¿cuál es la probabilidad de que la muestra no esté deteriorada?
e) ¿Cuál es la probabilidad de que el test dé el resultado correcto?

Resolución

Sean los sucesos D = muestra deteriorada y A = resultado del test positivo

$P(D) = 1/25 = 0,04$ $P(\bar{D}) = 0,96$
$P(A/D) = 0,99$ $P(\bar{A}/D) = 0,01$
$P(A/\bar{D}) = 0,02$ $P(\bar{A}/\bar{D}) = 0,98$

Probabilidad de que el test dé positivo

$P(A) = P(D)\, P(A/D) + P(\bar{D})\, P(A/\bar{D}) = (0,04)(0,99) + (0,96)(0,02) = 0,0588$

Probabilidad de que el test dé negativo

$P(\bar{A}) = 1 - P(A) = 1 - 0,0588 = 0,9412$

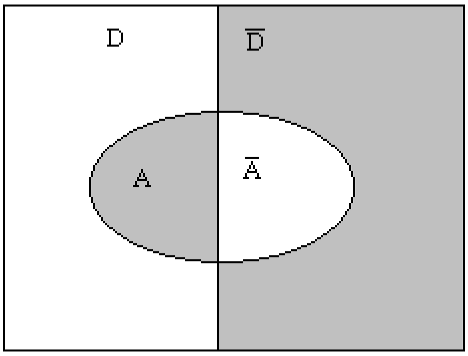
c) Si el test dá positivo, ¿cuál es la probabilidad de que la muestra esté deteriorada?

$$P(D/A) = \frac{P(D \cap A)}{P(A)} = \frac{P(D)P(A/D)}{P(A)} = \frac{(0,04)(0,99)}{0,0588} = 0,6735$$

d) Si el test da negativo, ¿cuál es la probabilidad de que la muestra no esté deteriorada?

$$P(\bar{D}/\bar{A}) = \frac{P(\bar{D} \cap \bar{A})}{P(\bar{A})} = \frac{P(\bar{D})P(\bar{A}/\bar{D})}{P(\bar{A})} = \frac{(0,96)(0,98)}{0,9412} = 0,9995$$

e) Probabilidad de que el test dé el resultado correcto



Resultado correcto = $(D \cap A) \cup (\bar{D} \cap \bar{A})$

$P[(D \cap A) \cup (\bar{D} \cap \bar{A})]$= P(D) P(A/D) + $P(\bar{D})$ $P(\bar{A}/\bar{D})$ = (0,04) (0,99) + (0,96) (0,98) = 0,9804

Ejercicio 20. Un banco dispone de dos sistemas de alarma A y B que funcionan independientemente entre sí. La alarma A tiene una probabilidad de que funcione correctamente del 80%, mientras que B sólo el 65%. ¿Cuál es la probabilidad de que sólo funcione correctamente una de las dos alarmas?

Ejercicio 21. Una empresa del sector eléctrico tiene un consejo directivo formado por 10 personas, 2 de las cuales son mujeres. Para evaluar las repercusiones que tendría para la empresa la aplicación de un nuevo sistema de tarifas, se designa una comisión de estudio formada por cuatro personas pertenecientes al consejo directivo. ¿Cuál es la probabilidad de que, si se eligen al azar, todos los miembros de dicha comisión sean hombres?

Ejercicio 22. En un multicine, la experiencia indica que el 75% de los asistentes compran palomitas y el 50% compran alguna bebida, siendo la compra de los dos productos independiente. Si sabemos que un espectador ha comprado sólo uno de los productos, ¿cuál es la probabilidad de que haya comprado sólo palomitas?

Ejercicio 23. La probabilidad de que un individuo sea usuario habitual de Internet es 0,3; la probabilidad que sea usuario de la biblioteca municipal es 0,15 y la de que sea usuario habitual de por lo menos uno de estos dos servicios es del 0,4. Se elige al azar un individuo, si es usuario de la biblioteca municipal, ¿cuál es la probabilidad aproximada de que sea usuario de Internet?

Tema 8. VARIABLE ALEATORIA UNIDIMENSIONAL

8.1 Variable aleatoria: discreta y continua
8.2 Distribución de probabilidad: función de cuantia y función de densidad
8.3 Función de distribución
8.4 Esperanza matemática y varianza. Variable estandarizada

VARIABLE ALEATORIA

DEFINICIÓN

Una **variable aleatoria** es una función definida entre el espacio muestral y el conjunto de los números reales.

Ejemplo:
Experimento: se lanzan 2 monedas
Variable aleatoria: X= {nº de cruces}

$$
\begin{array}{ccc}
X: E & \longrightarrow & R \\
(c,c) & \longrightarrow & 0 \\
(c,+) & \longrightarrow & 1 \\
(+,c) & \longrightarrow & 1 \\
(+,+) & \longrightarrow & 2
\end{array}
$$

La definición de una variable aleatoria permite asignar un valor numérico a cada uno de los resultados de un experimento aleatorio.

La variable aleatoria puede ser: $\begin{cases} \text{Discreta} \\ \text{Continua} \end{cases}$

DISTRIBUCIÓN DE PROBABILIDAD

DISCRETA: La distribución de probabilidad de un variable discreta recibe el nombre de **Función de Cuantía, P(X)**.

X	P(x)
x_1	$P(x_1)$
x_2	$P(x_2)$
…	…
X_i	$P(x_i)$
…	…
X_n	$P(x_n)$

Asocia a cada valor x_i su probabilidad P(xi) = P(X=xi).

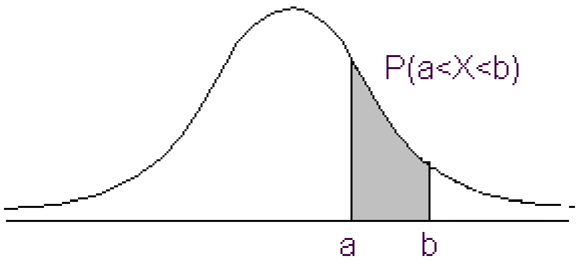
La función de cuantía siempre verifica:

- $P(x) \geq 0 \ \forall x$. Siempre toma valores no negativos.

- $\sum_{\forall x} P(x) = 1$. La probabilidad total es igual a 1.

CONTINUA: La distribución de probabilidad de un variable continua queda descrita por una función continua que recibe el nombre de **Función de Densidad, f(x),** tal que:

- $f(x) \geq 0 \ \forall x$. Toma valores no negativos.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$. El área definida por esta función es igual a 1 y, por lo tanto, representa la probabilidad total.



Características:

- $P(x=a) = 0$ La probabilidad de un valor concreto es siempre 0.
- $P(a<X\leq b) = \int_a^b f(x)\,dx$ La probabilidad de un intervalo es el área definida por la función de densidad.
- $P(a\leq X\leq b) = P(a<X\leq b) = P(a\leq X<b) = P(a<X<b)$

FUNCIÓN DE DISTRIBUCIÓN, F(X)

Recoge la probabilidad acumulada hasta un valor x:

$F(x) = P(X\leq x) = P[-\infty, x]$.

$$F(x) = P(X\leq x) = \begin{cases} \sum_{x\leq xi} P(x_i) & \text{Discreta} \\ \int_{-\infty}^{x} f(x)dx & \text{Continua} \end{cases}$$

Propiedades:

- La función es no negativa. $0\leq F(x)\leq 1$.
- La función es no decreciente. Dados dos valores $a < b$ entonces $F(a)\leq F(b)$.

- La función converge a cero por la izquieda. $F(-\infty) = 0$.
- La función converge a 1 por la derecha. $F(+\infty) =1$.
- $P(a < X \leq b) = F(b) - F(a)$.
- $P(X>a) = 1- F(a)$.

ESPERANZA MATEMÁTICA Y VARIANZA

Permiten resumir la distribución de probabilidad de una variable aleatoria. Reciben el nombre de Parámetros.

Esperanza Matemática o Valor Esperado de X, E(X)=μ

Es el valor medio teórico de la distribución de probabilidad.

Se obtiene:

$$\mu = E(X) = \begin{cases} \sum_{\forall x} x\, P(x) & \text{Discreta} \\ \int_{-\infty}^{+\infty} x\, f(x)\, dx & \text{Continua} \end{cases}$$

Se interpreta como la media de los valores de la variable aleatoria que obtendríamos si el experimento se realizara infinitas veces.

Propiedades:

- La esperanza matemática es el centro de gravedad de la distribución de probabilidad. $E(X- \mu)=0$
- $E(a) = a$. La esperanza matemática de una constante es la constante.
- $E(X+a) = E(X)+a$. La esperanza matemática queda afectada por los cambios de origen.
- $E(bX) = b\, E(X)$. También le afectan los cambios de escala.
- $E(a+b X) = a+b\, E(X)$.
- $E(aX+bY) = aE(X) + bE(Y)$ siendo X e Y dos variables aleatorias y, a y b, dos constantes.

Varianza, $V(X)=\sigma^2$

Mide la dispersión de la distribución de probabilidad alrededor de μ.

$$V(X) = \sigma^2 = E(X - \mu)^2 = \begin{cases} \sum_{\forall x} (x-\mu)^2 P(x) = \sum_{\forall x} x^2 P(x) - \mu^2 & \text{Discreta} \\ \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \mu^2 & \text{Continua} \end{cases}$$

Propiedades:

- $V(X) \geq 0$.
- $V(a) = 0$.
- $V(b X) = b^2 \cdot V(X)$
- $V(a + b X) = b^2 \cdot V(X)$

Desviación estándar, D(x) = σ

$D(X) = \sigma = +\sqrt{\sigma^2}$

Presenta las mismas propiedades que la varianza: es no negativa y sólo queda afectada por los cambios de escala.

Ejercicios Tema 8

Ejercicio 1. De un colectivo compuesto por un 60% de hombres se eligen al azar con devolución 3 personas. Establezca la distribución de probabilidad de la variable aleatoria X = "Número de mujeres elegidas".

Ejercicio 2. Un llavero tiene 5 llaves, de las cuales sólo 2 abren una determinada puerta, y se van probando, una a una, hasta conseguir abrir. Se define la variable aleatoria X = 'Número de llaves probadas". Determine la distribución de probabilidad de X bajo los siguientes supuestos:

a) se separan las llaves probadas;

b) las llaves probadas quedan de nuevo mezcladas con las restantes y es imposible diferenciarlas.

Resolución

a) Se separan las llaves probadas. Sea Ai = 'la i-ésima llave abre'. El espacio muestral del experimento es:

$E = \{A_1,\ \bar{A}_1 A_2,\ \bar{A}_1\bar{A}_2 A_3,\ \bar{A}_1\bar{A}_2\bar{A}_3 A_4\}$

La variable X toma los siguientes valores:

Si el resultado es A_1 $X = 1$
Si el resultado es $\bar{A}_1 A_2$ $X = 2$
Si el resultado es $\bar{A}_1\bar{A}_2 A_3$ $X = 3$
Si el resultado es $\bar{A}_1\bar{A}_2\bar{A}_3 A_4$ $X = 4$

Las probabilidades de los resultados del experimento son:

$P(A_1) = \frac{2}{5} = 0{,}4$ $P(X = 1) = 0{,}4$

$P(\bar{A}_1 A_2) = \frac{3}{5}\frac{2}{4} = 0{,}3$ $P(X = 2) = 0{,}3$

$P(\bar{A}_1\bar{A}_2 A_3) = \frac{3}{5}\frac{2}{4}\frac{2}{3} = 0{,}2$ $P(X = 3) = 0{,}2$

$P(\bar{A}_1\bar{A}_2\bar{A}_3 A_4) = \frac{3}{5}\frac{2}{4}\frac{1}{3}\frac{2}{2} = 0{,}1$ $P(X = 4) = 0{,}1$

X	1	2	3	4
P(X)	0,4	0,3	0,2	0,1

b) No se separan las llaves probadas

El espacio muestral del experimento es:

$E = \{A_1,\ \bar{A}_1 A_2,\ \bar{A}_1\bar{A}_2 A_3,\ \bar{A}_1\bar{A}_2\bar{A}_3 A_4,\ \bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4 A_5,\ \bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5 A_6,$
$,\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5\bar{A}_6 A_7 \ldots\ldots\}$

E contiene un número infinito numerable de resultados posibles; por tanto X será una variable discreta que toma un número infinito de valores:

Si el resultado es A_1	$X = 1$
Si el resultado es $\bar{A}_1 A_2$	$X = 2$
Si el resultado es $\bar{A}_1\bar{A}_2 A_3$	$X = 3$
Si el resultado es $\bar{A}_1\bar{A}_2\bar{A}_3 A_4$	$X = 4$
Si el resultado es $\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4 A_5$	$X = 5$
Si el resultado es $\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5 A_6$	$X = 6$
Si el resultado es $\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5\bar{A}_6 A_7$	$X = 7$

Teniendo en cuenta que las probabilidades de los resultados son:

$P(A_1) = \frac{2}{5} = 0{,}4$

$P(\bar{A}_1 A_2) = \frac{3}{5}\frac{2}{5} = 0{,}24$

$P(\bar{A}_1\bar{A}_2 A_3) = \frac{3}{5}\frac{3}{5}\frac{2}{5} = 0{,}144$

$P(\bar{A}_1\bar{A}_2\bar{A}_3 A_4) = \frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{2}{5} = 0{,}086$

$P(\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4 A_5) = \frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{2}{5} = 0{,}052$

$P(\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5 A_6) = \frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{2}{5} = 0{,}031$

$P(\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5\bar{A}_6 A_7) = \frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{3}{5}\frac{2}{5} = 0{,}018$

La distribución de probabilidad de X es:

X	1	2	3	4	5	6	7	......
P(X)	0,4	0,24	0,144	0,086	0,052	0,031	0,018	....

Esta distribución puede sintetizarse mediante la función

$P(x) = (3/5)^{x-1}\ (2/5)$

o también

$P(x) = 0{,}6^{x-1}\ 0{,}4$

Ejercicio 3. Se lanza una moneda hasta que sale cara con un máximo de 5 lanzamientos. Determine la distribución de probabilidad de la variable aleatoria X="Número de lanzamientos realizados".

Ejercicio 4. Un servicio de atención al cliente estima que la distribución de probabilidad de la variable X= "Número de clientes atendidos por hora" es la siguiente:

X_i	0	1	2	3	4	5	6	7
$P(x_i)$	0,01	0,12	0,22	0,32	0,20	0,08	0,03	0,02

a) Halle la función de Distribución de la variable X.
b) ¿Cuál es la probabilidad de que en una hora utilice este servicio algún cliente?
c) ¿Cuál es la probabilidad de que en una hora utilice este servicio un mínimo de cinco clientes?
d) ¿Cuál es la probabilidad de que en una hora sean atendidos más de dos y como máximo cinco clientes?
e) Si se sabe que en una hora determinada algún cliente ha utilizado el servicio, ¿cuál es la probabilidad de que hayan sido más de tres?

Resolución

a) Función de distribución

$$F(x) = \begin{cases} 0 & x<0 \\ 0{,}01 & 0 \le x<1 \\ 0{,}13 & 1 \le x<2 \\ 0{,}35 & 2 \le x<3 \\ 0{,}67 & 3 \le x<4 \\ 0{,}87 & 4 \le x<5 \\ 0{,}95 & 5 \le x<6 \\ 0{,}98 & 6 \le x<7 \\ 1 & x \ge 7 \end{cases}$$

b) P(Por lo menos uno) = 1 – P(0) = 1 – 0,01 = 0,99

c) $P(X \ge 5) = 1 - P(X \le 4) = 1 - F(4) = 1 - 0{,}87 = 0{,}13$

d) $P(2 < X \le 5) = F(5) - F(2) = 0{,}95 - 0{,}35 = 0{,}60$

e) $P(X > 3/X > 0) = \dfrac{P[(X<3)\cap(X>0)]}{P(X>0)} = \dfrac{P(X \ge 4)}{P(X>0)} = \dfrac{1-P(X \le 3)}{1-P(0)} = \dfrac{1-F(3)}{1-P(0)} = \dfrac{1-0{,}67}{1-0{,}01} =$

$=0{,}333$

Ejercicio 5. Una variable aleatoria X queda caracterizada por la función: P(x)=k/x si x=1, 2, 3 y 4.

a) Determine el valor de k.

b) Obtenga la función de distribución.
c) Calcule $P(1 \leq X<3)$, $P(2<X \leq 4)$, $P(X>2)$, $P(X<2/X<4)$.

Ejercicio 6. La función de cuantía de una variable aleatoria X es:

$$P(x)=\begin{cases} 0{,}05 & \text{para} \quad x=1 \quad y \quad x=8 \\ 0{,}10 & \text{para} \quad x=2 \quad y \quad x=7 \\ 0{,}15 & \text{para} \quad x=3 \quad y \quad x=6 \\ 0{,}20 & \text{para} \quad x=4 \quad y \quad x=5 \\ 0 & \text{Otros casos} \end{cases}$$

a) Compruebe que es función de cuantía.
b) Determine la función de distribución.
c) Obtenga las siguientes probabilidades: $P(X<4)$, $P(1<X \leq 4)$, $P(X>2)$, $P(3 \leq X<6)$, $P(X \geq 3/X<7)$.

Ejercicio 7. Se sabe que en promedio uno de cada 10 clientes que entran en cierto establecimiento realiza una compra. La variable X = 'Nº de clientes que entran hasta que compra uno (incluido éste)' tiene la siguiente función de distribución:

$F(x) = 1\text{-}0{,}90^x$ si $x = 1, 2, 3, ...$

Se pide:

a) Probabilidad de que el décimo cliente que entre sea el primero que compre.
b) Probabilidad de que hayan entrado como máximo 4 clientes hasta que compra uno.
c) Probabilidad de que hayan entrado por lo menos 3 clientes hasta que compra uno.
d) Si se sabe que ya han entrado más de 4 clientes, ¿cuál es la probabilidad de que entren como máximo 7 considerando el que compra?

Ejercicio 8. La función de distribución de una variable aleatoria es:

$$F(x)=\begin{cases} 0 & x<1/8 \\ 0{,}2 & 1/8 \leq x<1/4 \\ 0{,}9 & 1/4 \leq x<3/8 \\ 1 & x \geq 3/8 \end{cases}$$

a) Indique si la variable es continua o discreta.
b) Obtenga la función de probabilidad.
c) Represente gráficamente las dos funciones.

Ejercicio 9. Sea la función,

$$f(x)=\begin{cases} 3x^2/2 & -1 \leq x \leq +1 \\ 0 & \text{Otros casos} \end{cases}$$

a) Compruebe si es función de densidad.
b) Calcule la probabilidad de que X sea superior a -0,5 e inferior a 0,25.
c) Obtenga la función de distribución.

Ejercicio 10. La función de densidad que caracteriza a la variable aleatoria X es:

$$f(x)=\begin{cases} 1/k & 1 \le x \le 5 \\ 0 & \text{Otros casos} \end{cases}$$

a) Determine el valor de k.
b) Obtenga la función de distribución.
c) Calcule las siguientes probabilidades: $P(X\le 3)$, $P(1,5\le X<4)$, $P(X>2)$, $P(X>3/X>2)$.

Resolución

a) $$\int_1^5 \frac{1}{k}dx=\frac{1}{k}\int_1^5 dx=\frac{1}{4}\left[x\right]_1^5=\frac{1}{k}\left[5-1\right]=\frac{4}{k}=1 \qquad k=4$$

b) Para cualquier valor de X, xi:

$$F(xi) = P(X\le xi) = \int_1^{x_{ii}} \frac{1}{4}dx=\frac{1}{4}\int_1^{x_i} dx=\frac{1}{4}\left[x\right]_1^{x_i}=\frac{1}{4}\left[x_i-1\right]=\frac{x_1-1}{4}$$

La función de distribución es:

$$F(x) = \begin{cases} 0 & x<1 \\ \dfrac{x-1}{4} & 1\le x\le 5 \\ 1 & x>5 \end{cases}$$

c) $P(X \le 3) = F(3) = \dfrac{3-1}{4} = 0,5$ $\qquad P(1,5 \le X < 4) = F(4) - F(1,5) = \dfrac{4-1}{4}-\dfrac{1,5-1}{4} =$ $= 0,625$

$$P(X > 2) = 1 - P(X \le 2) = 1 - F(2) = 1 - \frac{2-1}{4} = 0,75$$

$$P(X > 3/X > 2) = \frac{P[(x<3)\cap(x<2)}{P(X<2)}=\frac{P(X<3)}{p(x<2)}=\frac{1-F(3)}{1-F(2)}=\frac{1-\frac{3-1}{4}}{0,75} = 0,667$$

Ejercicio 11. Sea la función,

$$f(x)=\begin{cases} x^2/9 & 0 \le x \le k \\ 0 & \text{Otros casos} \end{cases}$$

a) Determine k para que la función sea función de densidad.
b) Calcule la probabilidad de que X sea superior a 0,5 sabiendo que es inferior a 2,5.

Ejercicio 12. Dadas las siguientes funciones de distribución:

A) $F(x)=\begin{cases} 0 & x<0 \\ 0,1 & 0 \le x<2 \\ 0,3 & 2 \le x<3 \\ 0,6 & 3 \le x<4 \\ 1 & x \ge 4 \end{cases}$ B) $F(x)=\begin{cases} 0 & x<0 \\ \dfrac{x^3}{64} & 0 \le x \le 4 \\ 1 & x>4 \end{cases}$

a) Indique, en cada caso, si la variable es continua o discreta.
b) Calcule, para estas variables, $P(X\le 2)$, $P(1\le X<3)$, $P(X>2)$, $P(X>1/X<2,5)$.

Ejercicio 13. En determinada estación de servicio la variable aleatoria X = 'Demanda semanal de gasolina en miles de litros' tiene la siguiente función de distribución:

$$F(x)=\begin{cases} 0 & x<0 \\ 1,2x-0,2x^2 & 0 \le x \le 1 \\ 1 & x>1 \end{cases}$$

Determine:
a) La probabilidad de que la demanda sea inferior a 0,5 (miles de litros).
b) La probabilidad de que la demanda sea inferior a 0,5 y superior a 0,3.
c) La probabilidad de que la demanda sea exactamente 0,2.
d) La probabilidad de que la demanda sea superior a 0,3 sabiendo que es inferior a 0,5.
e) La probabilidad de que la demanda supere los 0,4 (miles de litros).

Ejercicio 14. En los paquetes de harina de 10 kg se comete un error aleatorio en el peso, X (en kg), cuyo comportamiento queda recogido por la función de distribución:

$$F(x)=\begin{cases} 0 & x<-1 \\ \dfrac{x^3}{2}+k & -1 \le x \le 1 \\ 1 & x>1 \end{cases}$$

a) Determine el valor de k.
b) Calcule la probabilidad de que un paquete de harina supere 10,5 kg.
c) En una partida de 1000 paquetes ¿qué proporción de paquetes se espera que pesen más de 9 kg y cuarto?
d) Sabiendo que un paquete supera los 9 kg y medio, ¿cuál es la probabilidad de que no alcance los 10 kg y cuarto?
e) ¿Cuál es el peso mínimo de un paquete para poder decir que está entre el 30% de los que más pesan?

Resolución

a) La variable X toma valores en el intervalo [-1; 1], por tanto $P(X \le 1) = 1$. Por otra parte, por definición de F(x) se tiene que:

$P(X \le 1) = F(1) = \dfrac{1^3}{2}+k=1 \qquad 1 + 2k = 2 \qquad 2k = 1 \qquad k = 1/2 = 0,5$

b) P(Peso > 10,5) = P(X > 0,5) = 1 – P(X ≤0,5) =1 - F(0,5) = 1 - $\left[\frac{0,5^3}{2}+0,5\right]$=

= 1 – 0,5625 = 0,4375

c) P(Peso>9,25) = P(X > -0,75) = 1 – P(X < -0,75) = 1 – F(-0,75) =

= 1 - $\left[\frac{-0,75^3}{2}+0,5\right]$=1 – 0,2890 = 0,7109

El 71,09% de los paquetes de la partida, o sea 710,9, pesan más de 9,25 kg.

d) P(X < 0,25/X > -0,5) = $\frac{P\left[(X<0,25)\cap(X>-0,5)\right]}{P(X>-0,5)}$=

$$=\frac{P(-0,5<X<0,25)}{1-P(X<-0,5)}=\frac{F(0,25)-F(-0,5)}{1-F(-0,5)}=\frac{\left[\frac{0,25^3}{2}+0,5\right]-\left[\frac{-0,5^3}{2}+0,5\right]}{1-\left[\frac{-0,5^3}{2}+0,5\right]}= 0,1250$$

d) Se trata de hallar el séptimo decil del peso, o, lo que es equivalente, el séptimo decil del error X. Sea D7 el valor de X tal que P(X > D7) = 0,30 y, por tanto,

P(X ≤ D7) = 0,70

P(X ≤ D7) = F(D7) = 0,70

F(D7) = $\frac{D_7^3}{2}$+0,5=0,70 $\quad$ D_7^3+ 2(0,5) = 2(0,70) $\quad$ D_7^3+ 1 = 1,40 $\quad$ D_7^3= 0,40

D7 = 0,7368

El paquete tendrá que pesar como mínimo 10,7368 kg.

Ejercicio 15. Un taller estima que la distribución de probabilidad de la variable X= “Tiempo empleado en el empaquetado de un determinado tipo de piezas” es la siguiente:

$$f(x)=\begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{Otros casos} \end{cases}$$

a) ¿Cuál es el tiempo medio del empaquetado?
b) ¿Considera que este valor esperado es una buena medida de síntesis del comportamiento de X?
c) Se reajusta el proceso y el nuevo tiempo del empaquetado es Y= 1,5X-0,5 manteniendo la misma distribución de probabilidad. Calcule el nuevo valor esperado del tiempo.
d) ¿El ajuste introducido ha conseguido mejorar la regularidad del proceso?

Ejercicio 16. Una variable aleatoria discreta, X, tiene la siguiente función de distribución:

$$F(x)=\begin{cases} 0 & x<3 \\ 1/12 & 3 \leq x<5 \\ 1/3 & 5 \leq x<7 \\ 3/4 & 7 \leq x<8 \\ 1 & x \geq 8 \end{cases}$$

a) Calcule $P(3{,}5 < X \leq 5)$.
b) Calcule la probabilidad de que X sea inferior a 7 sabiendo que ha tomado un valor superior a 4.
c) Determine la función de cuantía.
d) Calcule la mediana, el valor esperado y la desviación estándar de X.

Ejercicio 17. Un inversor está considerando tres alternativas para invertir 1000€. Se estiman los siguientes resultados:

Alternativa 1: un beneficio de 10.000€. con probabilidad de 0,15 y una pérdida de 1.000€ con probabilidad de 0,85

Alternativa 2: un beneficio de 1.000€. con probabilidad de 0,5 y una pérdida de 500€ con probabilidad de 0,5

Alternativa 3: un beneficio seguro de 400€

a) ¿En cuál de las tres opciones su beneficio esperado es mayor?
b) ¿Aconsejaría sin dudar al inversor que escogiera esta opción?

Ejercicio 18. El gerente de una fábrica está considerando cambiar una máquina cuyo número de averías semanales presenta la siguiente distribución de probabilidad.

Nº de averías	0	1	2	3	4
$P(x_i)$	0,1	0,26	0,42	0,16	0,06

La decisión será cambiarla si el coste esperado de las reparaciones semanales supera los 200 Euros.
¿Qué decisión tomará si la reparación de cada avería cuesta 150 Euros?

Ejercicio 19. Una compañía fabrica clips y los comercializa en paquetes de aproximadamente 50 unidades. El número de clips por paquete varía según la siguiente distribución de probabilidad:

Nº de clips	47	48	49	50	51	52
Probabilidad	0,05	0,10	0,25	0,30	0,20	0,10

a) ¿Cuál es la probabilidad de que un paquete contenga más de 48 clips y como máximo 51?
b) Sabiendo que un paquete contiene menos de 50 clips, ¿cuál es la probabilidad de que contenga más de 48?
c) ¿Cuál es el número esperado de clips por paquete y su varianza?
d) Si el coste de un paquete es 15+2X, donde X es el nº de clips, ¿cuál es el coste esperado y la varianza del coste de estos paquetes?

Ejercicio 20. Calcule el valor esperado, la varianza y la desviación estándar de las variables aleatorias de los ejercicios 9, 11 y 12.

Ejercicio 21. X es una variable aleatoria que toma los valores:
-2, 1, 2 y 4
con probabilidades:
(2k-3)/10, (k-2)/10, (k-1)/10 y (k+1)/10, respectivamente.
Se pide:
a) Valor de k.
b) Valor esperado, mediana y moda de X.
c) Varianza y desviación estándar de X.
d) Distribución estandarizada y comprobar que su valor esperado es 0 y la varianza 1.

Ejercicio 22. Las piezas producidas se empaquetan al azar de tres en tres. Se sabe que el 25% de las piezas son de color negro.
a) Establezca la distribución de probabilidad de la variable aleatoria X="Número de piezas negras por paquete".
b) Calcule el valor esperado de X y compruebe que coincide con el producto de n y p siendo n el número de piezas por paquete y p la probabilidad de que la pieza sea negra.
c) Calcule la varianza de X y compruebe que coincide con el producto np(1-p)

Ejercicio 23. Una lotería con 100 boletos da 1 premio de 500 €, 2 premios de 100 € y 6 premios de 50 €. Si cada boleto cuesta 10 €, ¿cuál es el valor esperado de este juego por boleto?

Ejercicio 24. Un jugador puede lanzar 2 veces una moneda equilibrada. Gana 3 € si salen 2 caras y 1 € si sólo sale una cara. Para que el juego sea justo ¿cuánto tiene que perder si no salen caras?

Ejercicio 25. La función de densidad de la variable aleatoria X ="Importe semanal facturado en miles de €" en un establecimiento es:

$$f(x)=\begin{cases} \dfrac{1+x}{12} & x \in [0, 4] \\ 0 & x \notin [0, 4] \end{cases}$$

a) Calcule el importe que en promedio se espera facturar por semana.
b) ¿Qué importe máximo espera conseguir el 40% de las semanas con menos facturación?
c) ¿Cuál será, aproximadamente, el mínimo importe que facturará el 25% de las semanas con mayor facturación?
d) ¿Cuál es la varianza y la desviación estándar de X?
e) Se estima que la facturación se ha transformado y ha pasado a ser 0,75X+0,25. ¿Es más regular la facturación que se espera obtener en esta nueva situación?

Tema 9. DISTRIBUCIONES BINOMIAL Y NORMAL

9.1 Distribución Binomial
9.2 Distribución Normal.

Distribución Dicotómica o de Bernoulli

Si en un experimento aleatorio sólo consideramos dos resultados, 'éxito' o 'no éxito', y la probabilidad de éxito es p y la probabilidad de no éxito es q=1-p, la variable:

X = "Número de éxitos obtenidos en una realización del experimento"

presenta una distribución Dicotómica con función de cuantía:

$$P(x) = p^x\, q^{1-x} \qquad \text{para } x = 0, 1$$

X	P(X)
0	1-p=q
1	p

- Para identificar una distribución dicotómica concreta dentro de la familia de distribuciones dicotómicas basta conocer el valor del parámetro p, probabilidad de éxito:
- La distribución de X se abrevia $X \sim D(p)$
- El valor esperado de X es $E(X) = p$
- La varianza de X es $V(X) = p\,q$

Distribución Binomial

Consideremos un experimento aleatorio tal que:

- Cada vez que se realiza el experimento ocurre uno y sólo uno de los siguientes resultados: 'éxito' o 'no éxito'.
- Cada vez que se repite el experimento el resultado es independiente del obtenido en las realizaciones anteriores, por lo que la probabilidad de éxito, p, es constante en cada prueba.
- Se realiza el experimento n veces.

Se define la variable:

X = "Número de éxitos obtenidos en las n realizaciones del experimento"

La variable X presenta distribución Binomial de parámetros n y p, y su función de cuantía es:

$$P(x) = \binom{n}{x} p^x q^{n-x} \quad \text{para } x = 0, 1, 2, ..., n \text{ y } q = 1\text{-}p$$

- Para identificar una distribución binomial concreta dentro de la familia de distribuciones binomiales basta con conocer los valores de los parámetros, n y p, que la caracterizan.
- La distribución de X se abrevia X ~ B(n; p).
- La distribución dicotómica es un caso particular de distribución binomial con parámetros n=1 y p: X ~ B(1; p) = D(p).
- La distribución B(n; p) se obtiene como suma de n distribuciones dicotómicas independientes de parámetro p.
- El valor esperado de de X es E(X) = n p.
- La varianza de X es V(X) = n p q.
- La distribución binomial es reproductiva en el parámetro p: si se suman dos o más variables binomiales independientes con el mismo parámetro p, la variable resultante tiene distribución: $B(\sum_{i=1}^{k} n_i, p)$.
- Para cualquier valor de n, la distribución de X es simétrica si p=0,5; presenta asimetría positiva si p < 0,5 y asimetría negativa si p > 0,5. La asimetría se reduce a medida que p se aproxima a 0,5. Asimismo, para cualquier valor de p la asimetría disminuye cuando aumenta el valor de n.

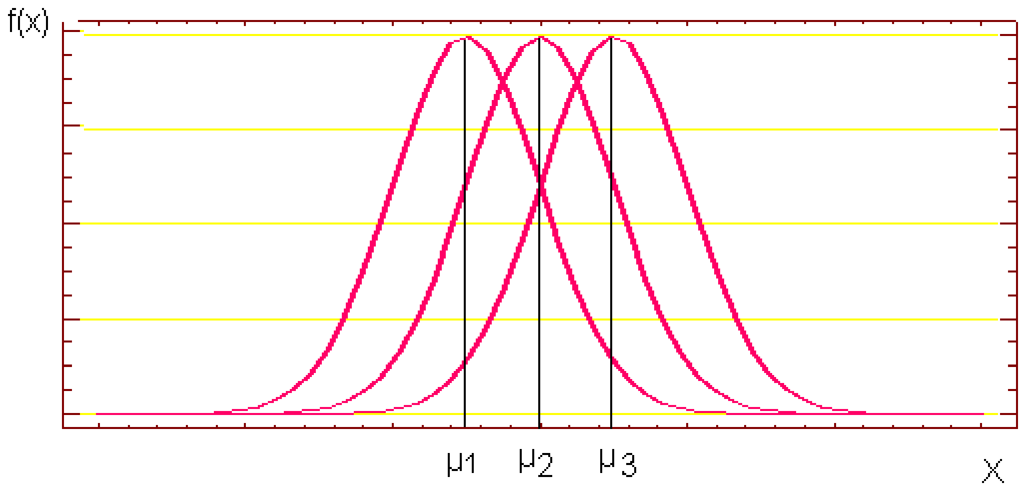
Distribución Normal

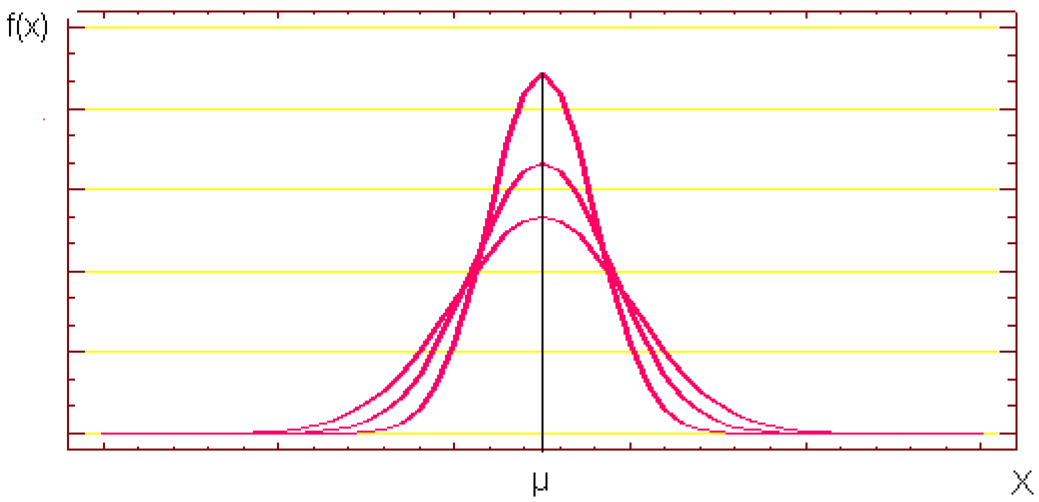
Una variable aleatoria continua X presenta una distribución normal de parámetros μ y σ si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \qquad \forall x \in R$$

donde: $\mu \in R$, $\sigma > 0$, $\pi = 3{,}14$ y $e = 2{,}71$

- Existe una familia de infinitas distribuciones normales:





- La distribución normal queda identificada por dos parámetros: su valor esperado, μ y su desviación estándar, σ.
- La distribución de X se abrevia: $X \sim N(\mu, \sigma)$.
- La variable X puede tomar cualquier valor Real de $-\infty$ a $+\infty$.
- La distribución de X es campanoide y simétrica:

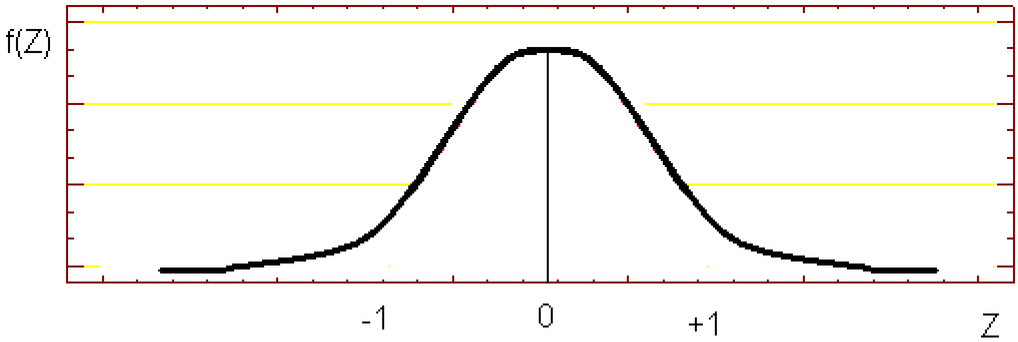
- El coeficiente de asimetría es 0.
- Esperanza matemática, mediana y moda coinciden.
- $P(X<\mu)=P(X>\mu)=0{,}5$.

- La distribución de X es mesocúrtica y su coeficiente de curtosis es 0.
- La distribución normal presenta dos puntos de inflexión en $\mu-\sigma$ y $\mu+\sigma$.
- Es asintótica respecte al eje de abscisas.
- La distribución normal es reproductiva; al sumar o restar dos o más variables normales independientes se obtiene una nueva variable normal de parámetros $N(\Sigma\mu, \sqrt{\Sigma\sigma^2})$.

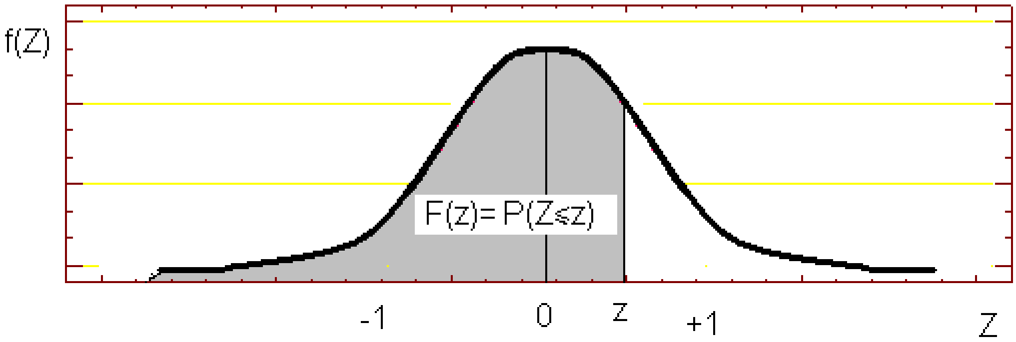
Distribución Normal Estandarizada o Tipificada

De entre las infinitas curvas normales la correspondiente a los parámetros $\mu=0$ y $\sigma=1$ recibe el nombre de distribución normal estandarizada o tipificada y presenta una especial importancia. Se simboliza $Z\sim N(0, 1)$.

Su función de densidad es $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\quad \forall z\in R$.



- Es simétrica: $P(Z<0) = P(Z>0) = 0{,}5$.
- Presenta un máximo en $z=0$.
- Tiene dos puntos de inflexión -1 y +1.
- Cualquier otra variable normal X de parámetros μ y σ se puede transformar en una normal estandarizada simplemente mediante la transformación lineal $Z=\frac{X-\mu}{\sigma}$.
- La función de distribución de esta variable, F(z), está tabulada.



La tabla permite obtener probabilidades de sucesos referidos a cualquier variable normal

X~N(μ, σ), ya que: $P(a \leq X \leq b) = P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})$

Utilización de la tabla de la distribución Normal

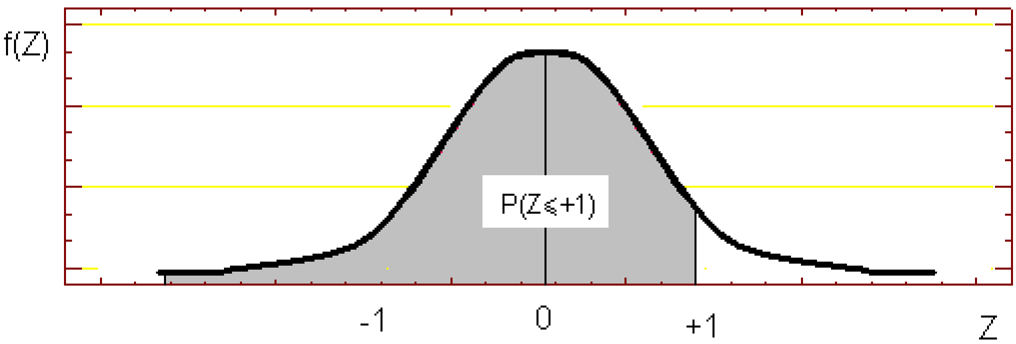
La tabla contiene los valores de la función de distribución de la variable Z~N(0, 1) o normal tipificada para valores positivos de la variable.

Para cualquier otra variable normal X~N(μ, σ) es necesario transformar X en Z:

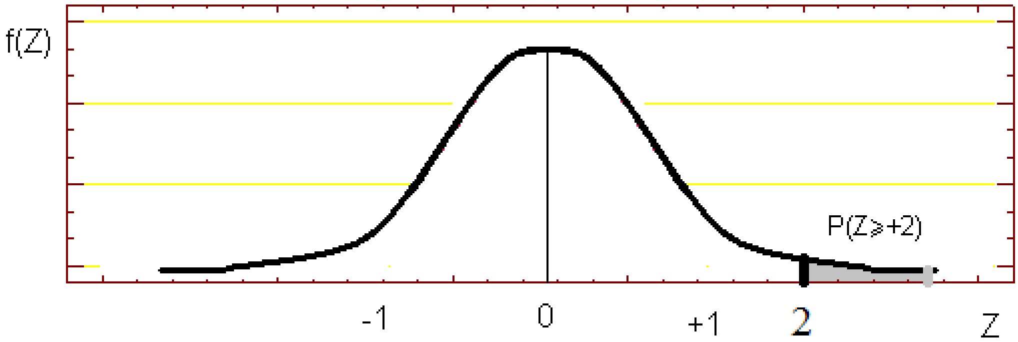
$$\mathbf{Z = \frac{X - \mu}{\sigma}}$$

Ejemplo: Dada una variable aleatoria X~N(40, σ=10), halle las siguientes probabilidades:

1. $\mathbf{P(X \leq 50) = P(Z \leq \frac{50-40}{10}) = P(Z \leq 1) = F(1) = 0{,}8413}$

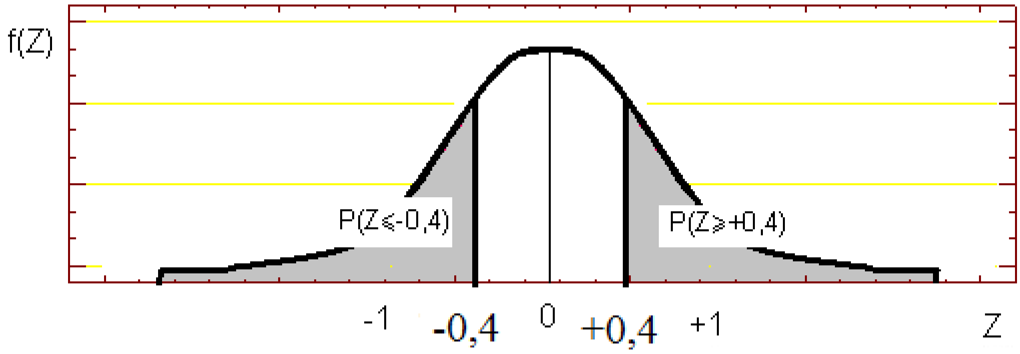


2. $\mathbf{P(X \geq 60) = P(Z \geq \frac{60-40}{10}) = P(Z \geq 2) = 1\text{-}P(Z \leq 2) = 1\text{-}F(2) = 1\text{-}0{,}9772 = 0{,}0228}$

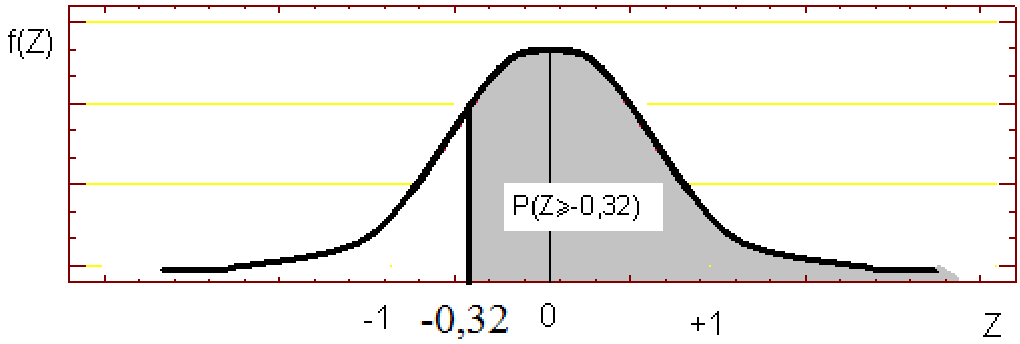


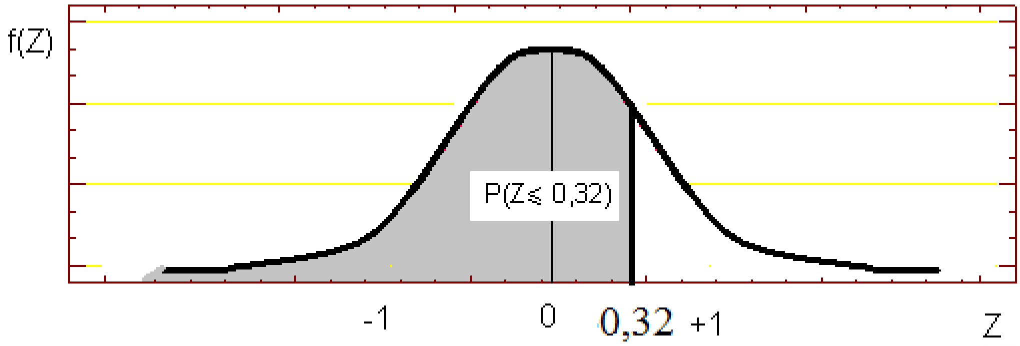
3. $\mathbf{P(X \leq 36) = P(Z \leq \frac{36-40}{10}) = P(Z \leq -0{,}4) = P(Z > 0{,}4) = 1\text{-}P(Z \leq 0{,}4) =}$

$\mathbf{= 1\text{-}F(0{,}4) = 1\text{-}0{,}6554 = 0{,}3446}$



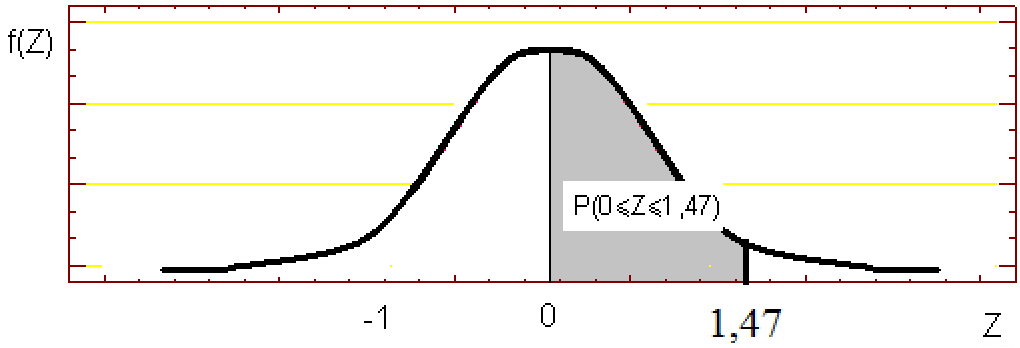
4. $\mathbf{P(X \geq 36{,}8) = P(Z \geq \frac{36{,}8-40}{10}) = P(Z \geq -0{,}32) = P(Z \leq 0{,}32) = F(0{,}32) = 0{,}6255}$





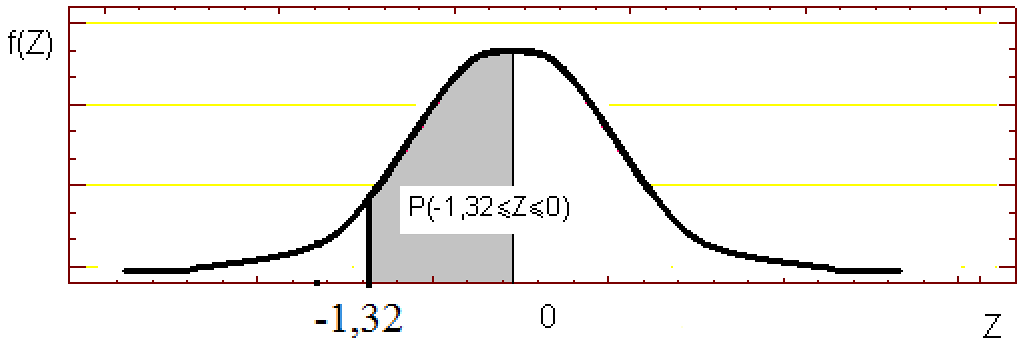
5. $P(40 \le X \le 54{,}7) = P(\frac{40-40}{10} \le Z \le \frac{54{,}7-40}{10}) = P(0 \le Z \le 1{,}47) =$

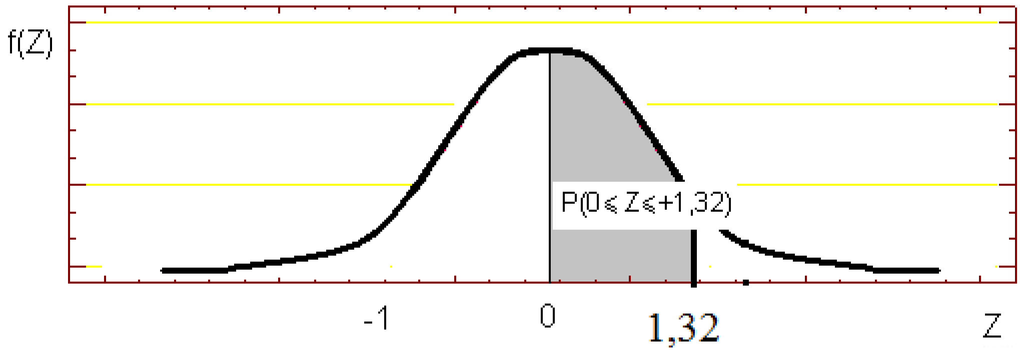
$= F(1{,}47) - F(0) = 0{,}9292 - 0{,}5 = 0{,}4292$



6. $P(26{,}8 \le X \le 40) = P(\frac{26{,}8-40}{10} \le Z \le \frac{40-40}{10}) = P(-1{,}32 \le Z \le 0) = P(0 \le Z \le 1{,}32) =$

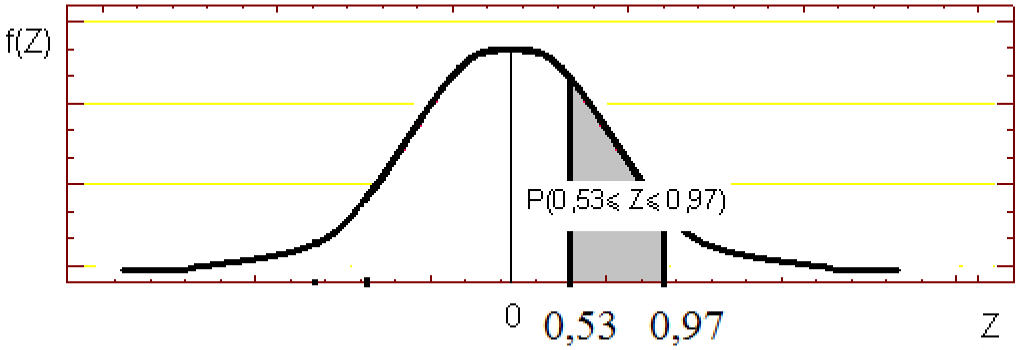
$= F(1{,}32) - F(0) = 0{,}9065 - 0{,}5 = 0{,}4065$





7. $P(45{,}3 \leq X \leq 49{,}7) = P(\frac{45{,}3-40}{10} \leq Z \leq \frac{49{,}7-40}{10}) = P(0{,}53 \leq Z \leq 0{,}97) =$

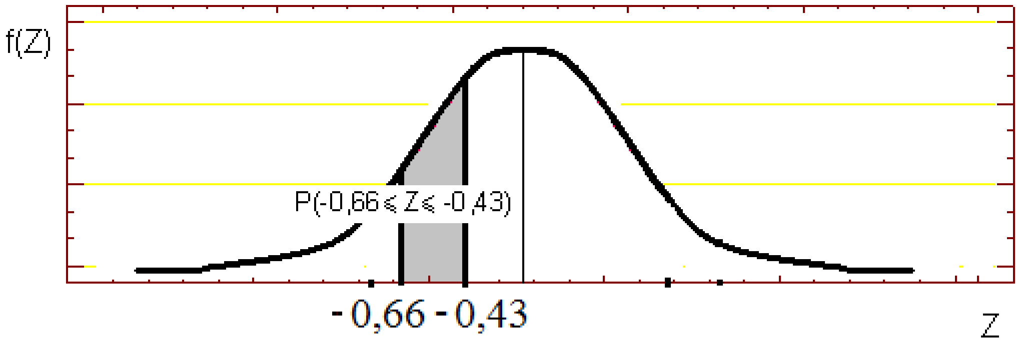
$= F(0{,}97) - F(0{,}53) = 0{,}8339 - 0{,}7019 = 0{,}132$

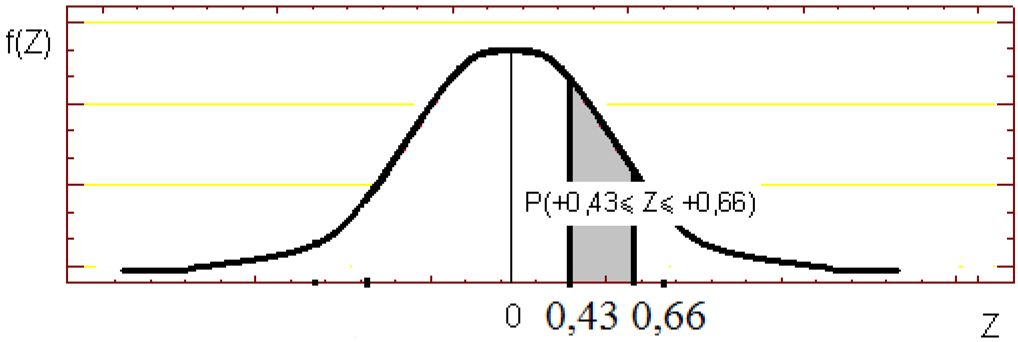


8.

$P(33{,}4 \leq X \leq 35{,}7) = P(\frac{33{,}4-40}{10} \leq Z \leq \frac{35{,}7-40}{10}) = P(-0{,}66 \leq Z \leq -0{,}43) = P(0{,}43 \leq Z \leq 0{,}66) =$

$= F(0{,}66) - F(0{,}43) = 0{,}7453 - 0{,}6664 = 0{,}0789$

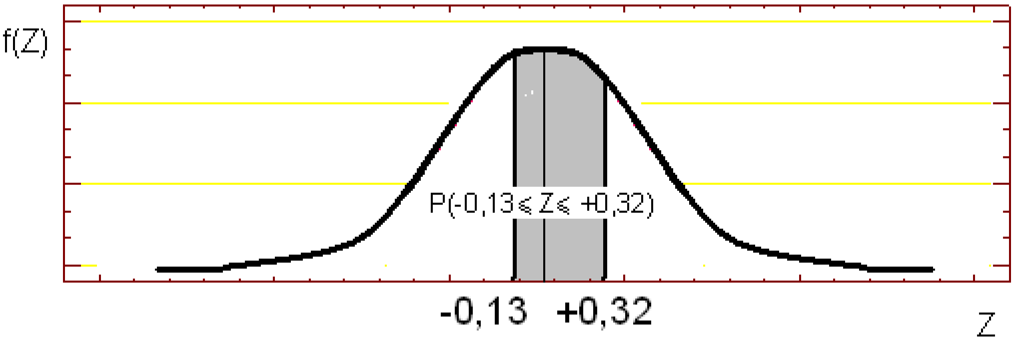




9. $P(38{,}7 \leq X \leq 43{,}2) = P(\frac{38{,}7-40}{10} \leq Z \leq \frac{43{,}2-40}{10}) = P(-0{,}13 \leq Z \leq 0{,}32) =$

$= P(Z \leq 0{,}32) - P(Z \leq -0{,}13) = P(Z \leq 0{,}32) - P(Z \geq 0{,}13) = P(Z \leq 0{,}32) - [1 - P(Z \leq 0{,}13)] =$

$= F(0{,}32) - 1 + F(0{,}13) = 0{,}6255 - 1 + 0{,}5517 = 0{,}1772$



f(Z)
P(-0,13 ≤ Z ≤ +0,32)
-0,13
+0,32
Z

Ejercicios capítulo 9

Ejercicio 1. Obtenga las siguientes probabilidades indicando, en cada caso, la definición y la distribución de la variable:

a) Probabilidad de que exactamente 2 clientes, entre 6 elegidos al azar, paguen con tarjeta de crédito si en general el 40% pagan utilizando este sistema.
b) Probabilidad de que obtengamos como máximo 2 caras al lanzar 5 veces una moneda.
c) Probabilidad de que una familia con 3 hijos elegida al azar tenga 2 o más niñas si consideramos que la probabilidad de tener niño o niña es la misma.
d) Probabilidad de que en un día elegido al azar más de una máquina se estropee si en el taller hay 8 máquinas idénticas con comportamientos independientes y presentan una probabilidad de estropearse de 0,25.
e) Probabilidad de obtener como mínimo 4 resultados impares en 5 lanzamientos de un dado.
f) Probabilidad de obtener 3 o más facturas impagadas si se extraen con reposición 10 de un archivador que contiene 600 facturas pagadas y 200 impagadas.

Ejercicio 2. La tasa de paro de un determinado país es del 10% de la población activa. Si se seleccionan al azar 7 personas determinar la probabilidad de que:

a) Exactamente 2 estén en paro.
b) Como máximo 2.
c) Como mínimo 2.

Ejercicio 3. El departamento de control de calidad sabe que la producción de cierto artículo presenta una proporción de defectuosos del 5%. Si estos artículos se venden en cajas de 10 unidades, calcule la probabilidad de que una caja contenga:

a) Ninguna unidad defectuosa.
b) Entre 2 y 4 unidades defectuosas.
c) Como máximo 1 unidad defectuosa.
d) Como mínimo un 80% de unidades de defectuosas.
e) Como máximo 1 sabiendo que contiene por lo menos 1 unidad defectuosa.
f) Ninguna defectuosa sabiendo que como máximo contiene 3 defectuosas.
g) ¿Cuál es el número más probable de unidades defectuosas por caja?
h) Si un pedido consta de 5 cajas, ¿cuál es la probabilidad de que exactamente el 80% de las cajas no contenga ninguna unidad defectuosa?

Ejercicio 4. Un test consta de 20 preguntas con 4 respuestas cada una de las cuales sólo una es correcta. A cada pregunta correcta se le asigna 1 punto y se restan 0,25 puntos por respuesta incorrecta. Si un estudiante responde a todas las preguntas al azar:

a) ¿Cuál es el número esperado de respuestas correctas y su desviación típica?
b) ¿Cuál es la puntuación esperada y su desviación típica?

Ejercicio 5. De una lista de 1000 números de teléfono se tiene la siguiente información: 350 corresponden a móviles y 250 corresponden al distrito de Gràcia. Se extraen al azar con reposición 5 números de esta lista. Halle la probabilidad de que:
a) Ninguno corresponda a un teléfono móvil.
b) Ninguno corresponda al distrito de Gràcia.
c) Ninguno corresponda ni a móvil ni al distrito de Gràcia (suponga independencia entre clase de teléfono y distrito).

Ejercicio 6. Una empresa que compra determinados componentes en lotes grandes sólo los acepta si contiene como máximo un 10% de defectuosos. A estos efectos se inspeccionan al azar 20 unidades de cada lote, y éste se acepta si como máximo la proporción de unidades defectuosas es del 10%. Si el proveedor estima que la proporción real de componentes defectuosas es del 5%:
a) ¿Cuál es la probabilidad de rechazar un lote?
b) ¿Cuál es la varianza y la desviación estándar de la variable 'número de piezas defectuosas en la muestra'?
c) Si el pedido consta de 10 lotes, ¿cuál es la probabilidad de rechazar como máximo 1?
d) Si el número de piezas inspeccionadas fuera de 10, ¿varía la probabilidad de rechazar el lote?

Ejercicio 7. Un vendedor ha comprobado que la distribución de probabilidad de sus ventas semanales es:

X	1	2	3	4	5	6
P(X)	0,25	0,35	0,15	0,10	0,05	0,10

(Siendo X=nº de coches vendidos)
La empresa le paga una prima semanal de 500 Euros si consigue vender más de 2 coches.
Si las ventas semanales son independientes, calcule la probabilidad de que en 1 mes (4 semanas),
a) Consiga como máximo 500 Euros de prima,
b) Consiga más de 1000 Euros de prima.
c) ¿Cuál es el valor esperado y la desviación estándar de la prima mensual que puede obtener este vendedor?

Resolución

a) Sea X = 'Número de semanas al mes con más de 2 coches vendidos'
   $X \sim \text{Bin}\ (4;\ 0{,}40)$

M = 'Prima mensual'

Si X = 0 M = 0 P(0) = 0,1296
Si X = 1 M = 500 P(1) = 0,3456
Si X = 2 M = 1000 P(2) = 0,3456
Si X = 3 M = 1500 P(3) = 0,1536
Si X = 4 M = 2000 P(4) = 0,0256

P(M ≤ 500) = PX = 0) + P(X = 1) = 0,1296 + 0,3456 = 0,4752

b) P(M > 1000) = P(X = 3) + P(X = 4) = 0,1536 + 0,0256 = 0,1792

c) La distribución de probabilidad de M es:

M	P(M)	MP(M)	M2P(M)
0	0,1296	0	0
500	0,3456	172,8	86400
1000	0,3456	345,6	345600
1500	0,1536	230,4	345600
2000	0,0256	51,2	102400
Total	1	800	880000

μ_M = E(M) = 800 €

σ^2_M = V(M) = E(M2) - μ^2_M = 880000 - 8002 = 240000

σ_M = D(M)= 489,9 €

Ejercicio 8. Una compañía tiene dos proveedores de un determinado artículo. El 70% de los pedidos procede del proveedor cuyos envíos suelen contener un 10% de artículos defectuosos y el resto proceden del otro proveedor cuyos envíos suelen contener un 20% de defectuosos. Se recibe un pedido, pero se desconoce su procedencia, se analiza una muestra de 20 artículos de este pedido y se encuentra 1 defectuoso. ¿Cuál es la probabilidad de que el pedido proceda del proveedor con menor tasa de artículos defectuosos?

Ejercicio 9. Dada una variable Normal de parámetros μ=10 σ=2 obtenga:
a) P(X≤ 14)
b) P(X>11,75)
c) P(X< 9)
d) P(5<X≤12)
e) P(6≤X≤14)
f) P(X>11/X>9,5)

Ejercicio 10. Determine el valor a que verifique:
a) P(X<a)= 0,846 siendo X ~ N(50; 15)
b) P(X>a)= 0,42 siendo X ~ N(10; 5)
c) P(X<a)= 0,25 siendo X ~ N(250; 50)
d) P(X>a)= 0,95 siendo X ~ N(20; 8)

Ejercicio 11. Por experiencia se sabe que la puntuación obtenida en cierto examen es una variable aleatoria Normal con esperanza matemática 11,5 puntos y desviación estándar 5.

a) Si la puntuación mínima para aprobar se fija en 10 puntos ¿qué porcentaje de alumnos aprobará?
b) Si se quiere aprobar al 69,5% de los presentados ¿qué nota mínima se debe exigir?
c) Si las puntuaciones mínima y máxima para obtener un notable son 15,5 y 18, respectivamente ¿cuántos notables se espera contabilizar en una prueba con 200 presentados?
d) ¿Qué puntuación mínima debe tener un examen para estar entre los 5 mejores en una prueba de 500 presentados?

Ejercicio 12. El peso neto (en gr.) de los paquetes de galletas de la marca A puede modelizarse mediante la variable aleatoria $X \sim N(500; 50)$. Se pide:

a) ¿Qué porcentaje de paquetes supera los 550 gr?
b) ¿Cuál es la probabilidad de que un paquete pese entre 480 y 520 gr?
c) En una partida de 2000 paquetes ¿cuántos se espera que no alcancen los 475 gr?
d) Se considera que un paquete es apto para la venta si pesa más de 440 gr. ¿Cuál es la probabilidad de que un paquete sea apto para la venta?
e) Bajo el supuesto que sólo un 15% de los paquetes se va a considerar **No** aptos para la venta ¿qué peso debe presentar un paquete para considerarlo no apto?
f) Los pedidos se sirven al minorista en cajas de 10 paquetes. ¿Cuál es la probabilidad de que una caja contenga por lo menos 8 paquetes aptos para la venta, es decir con peso superior a 440 gr?

Resolución

a) $$P(X \quad 550) = P\left(Z > \frac{550-500}{50}\right) = P(Z > 1) = 1 - P(Z > 1) = 1 - 0{,}84134 = 0{,}15866$$

El 15,86% de los paquetes superan los 550 gr.

b) $$P(480 \le x \le 520) = P\left(\frac{480-500}{50} \le Z \le \frac{520-500}{50}\right) = P(-0{,}4 \le Z \le 0{,}4) =$$

$$= P(Z < 0{,}4) - P(Z < -0{,}4) = P(Z < 0{,}4) - P(Z > 0{,}4) =$$

$$= P(Z < 0{,}4) - [1 - P(Z < 0{,}4] = 2P(Z < 0{,}4) -1 = 2\,(0{,}65542) - 1 = 0{,}31084$$

c) $$P(X < 475) = P\left(Z < \frac{475-500}{50}\right) = P(Z < -0{,}5) = P(Z > 0{,}5) = 1 - P(Z < 0{,}5) =$$

$$= 1 - 0{,}69146 = 0{,}30854$$

En una partida de 2000 paquetes se espera que 2000 (0,30854) = 617 no superen los 475 gr.

d) $P(X > 440) = P\left(Z > \frac{440-500}{50}\right) = P(Z > -1,2) = P(Z < 1,2) = 0,88493 \approx 0,885$

e) Si se quiere que sólo un 15% de los paquetes no sean aptos para la venta, el peso máximo de este 15% es C15, tal que $P(X \leq C15) = 0,15$ y $P(X > 0,15) = 0,85$

$$P(X > C15) = P\left(Z > \frac{C_{15}-500}{50}\right) = 0,85$$

En las tablas de la N(0; 1) se halla que $P(Z < 1,04) = 0,85$, por tanto

$$\left(\frac{C_{15}-500}{50}\right) = -1,04 \quad C15 = 448 \text{ gr.}$$

f) Y = 'número de paquetes con más de 440 gr. en una caja de 10 paquetes'

$Y \sim Bin\ (10; 0,885)$

$$P(Y > 8) = P(Y=9) + P(Y=10) = \binom{10}{9} 0,885^{9}(1-0,885)^{1} + \binom{10}{10} 0,885^{10}(1-0,885)^{0} =$$

$=0,6777$

Ejercicio 13. En una estación de servicio se sabe que la demanda mensual de gasolina (en litros) puede modelizarse mediante una Normal de media 150000 litros y desviación típica 10000 litros. Determine la cantidad que hay que tener disponible cada mes si se desea que la probabilidad de satisfacer la demanda sea igual a 0,95

Ejercicio 14. Un mecanógrafo sabe que el tiempo que tarda en mecanografiar una página de un manuscrito es una variable aleatoria aproximadamente normal de media 6 minutos y varianza 1,44.

a) ¿Cuál es la probabilidad de que necesite más de 8 minutos para mecanografiar una página?
b) De un manuscrito de 100 páginas, ¿qué tiempo como máximo necesitará para mecanografiar una de las 20 páginas más cortas?

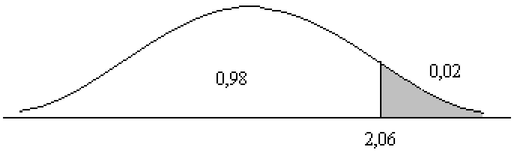
Ejercicio 15. Las retribuciones del personal de una empresa sigue una distribución normal. Se sabe que el 2% son superiores a 42000 Euros y el 10% inferiores a 15000 Euros. ¿Qué proporción son inferiores a 30000 Euros?

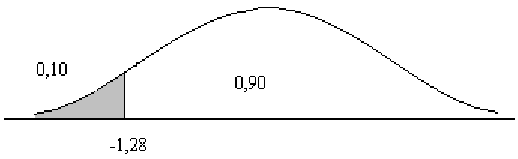
Resolución

X = 'Retribución (em €)' $\qquad X \sim N(\mu; \sigma)$
Se sabe que:

$$P(X > 42000) = P\left(Z > \frac{42000-\mu}{\sigma}\right) = 0{,}02$$

$$P(X < 15000) = P\left(Z < \frac{15000-\mu}{\sigma}\right) = 0{,}10$$





$$\frac{42000-\mu}{\sigma} = 2{,}06$$

$$\frac{15000-\mu}{\omega} = -1{,}28$$

Resolviendo el sistema para σ por la regla de Cramér:

$$\sigma = \frac{\begin{vmatrix} 1 & 42000 \\ 1 & 15000 \end{vmatrix}}{\begin{vmatrix} 1 & 2{,}06 \\ 1 & -1028 \end{vmatrix}} = \frac{15000-42000}{-1028-2{,}06} = \frac{-27000}{-3034} = 8083{,}83$$

$$\frac{42000-\mu}{8083{,}83} = 2{,}06 \quad \blacktriangleright \quad 42000 - \mu = 16652{,}69$$

$$\mu = 42000 - 16652{,}69 = 25347{,}31$$

La distribución de X es N(25347,31; 8083,83)

$$P(X < 30000) = \left(P(Z < \frac{30000-25347{,}31}{8083{,}83}\right) = P(Z < 0{,}57) = 0{,}7156$$

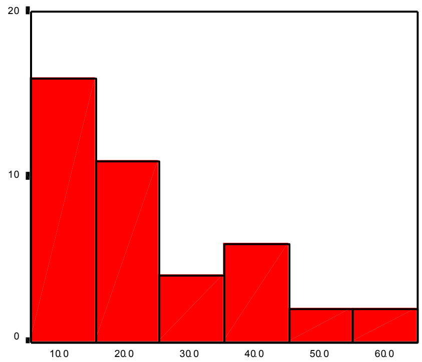
El 71,56% de las retribuciones son inferiores a 3000 Euros

SOLUCIÓN EJERCICIOS

TEMA 1 Y 2

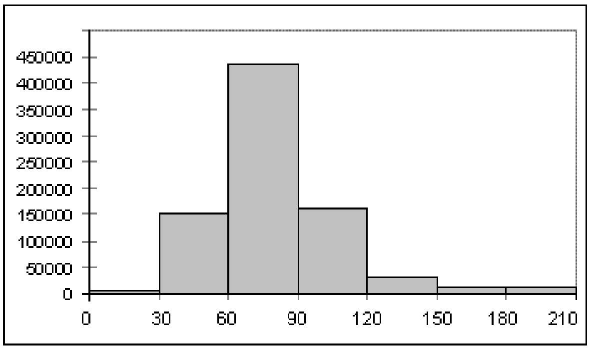
1. a) 100 b) 23%
   c) 18 d) 9 y 1
   e) 5 y 6 f) 4
   g) 16% h) 6
   i) 4

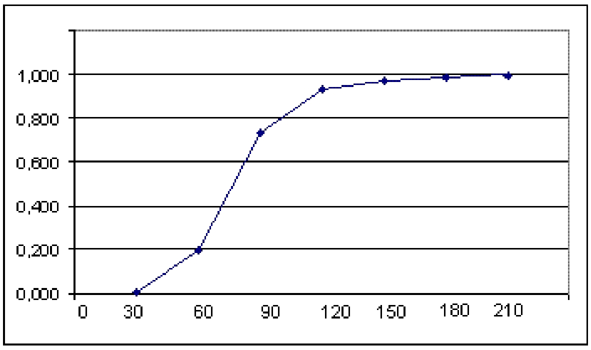
2.



3.

Superfície (m^2)	ni	Ni	fi	Fi
(0 -30]	4050	4050	0,005	0,005
(30 - 60]	153900	157950	0,190	0,195
(60 - 90]	437400	595350	0,540	0,735
(90 - 120]	162000	757350	0,200	0,935
(120 - 150]	29160	786510	0,036	0,971
(150 - 180]	12150	798660	0,015	0,986
(180 - 210]	11340	810000	0,014	1,000
Total	**810000**		**1**	





c) 810000
d) 437400
e) 157950
f) 810000-786510 = 23490
g) 54% y 19,5%
h) No
i) 90
j) Si, 214650 y 26,5%
k) (60,90]
l) 120 m^2
m) 150 m^2
n) Puede ser.

4.

L_{i-1}-L_i	X_i	n_i	f_i	N_i	F_i
0-2	1	5	0,1	5	0,1
2-4	3	11	0,22	16	0,32
4-6	5	19	0,38	35	0,7
6-8	7	10	0,2	45	0,9
8-10	9	5	0,1	50	1
		50	1		

a) 50
b) 68%
c) 6
d) No

5.

a) Xmin=655 Xmax=6982 n=50 número de tallos (o intervalos) como mínimo $\sqrt{50}\approx 7$

Unidades de los tallos: 1000
Unidades de la hoja: 100
0 | 6: representa 600

6	0	678999
24	1	022333444455666779
(11)	2	00244577999
15	3	03469
10	4	003566
4	5	35
2	6	29

b)

```
1 | 2: represents 1200
 leaf unit: 100
            n: 50
   6    0. | 678999
  16    1* | 0223334444
  24    1. | 55666779
  (5)   2* | 00244
  21    2. | 577999
  15    3* | 034
  12    3. | 69
  10    4* | 003
   7    4. | 566
   4    5* | 3
   3    5. | 5
   2    6* | 2
   1    6. | 9
```

c) 1400
d) 4300

6. a) 47
   b) Mín=12 Máx=82. Los valores más frecuentes (Mo) son: 33, 36 y 43.
   c) 27 d) 31 e) 60

TEMA 3

1. a) 120 días
   b) $\overline{X}$=15,458
   c) $\overline{X}$(10%)=15,231

2. 149,5 €/vendedor

3. a) $\overline{X}_A$= 6,4536% $\overline{X}_B$= 6%   b) 6,2%

4. A) 1,4309 Euros/Libra B) 1,4276 Euros/Libras

5. $\overline{X}$=71,408

6. 19875 u.m.

7. 70%

8. 12 horas en el turno de día y 8 en el de noche

9. 8 puntos

10.

	Ej 1	Ej 2	Ej 3
$\overline{X}$	4,98	23,41	80,19
Me	5	19	76,94
Mo	5 y 6	9 y 13	(60;90)

11. a) 0 y 300 €
    b) 101 €
    c) 29
    d) 68 y 151 €
    e) 139,5 €
    f) 3,3%
    g) 45 €

12. a) Q_1=10 €
    b) Q_3 = 13,8 €
    c) RIQ = 3,8 €
    d) D_4 = 12 €
    e) C_{85} = 14,2 €
    f) Frecuencia relativa 36,2%

13. a) $\overline{X}_A$=3,35 $\overline{X}_T$= 3,40 $\overline{X}_G$= 3,39
    b) 2 años
    c) 4 años
    d) 4 años

TEMA 4

1.

	Ej 1	Ej 2	Ej 3
$\overline{X}$	4,98	23,41	15,46
S	1,6	14,93	6,78
CV	0,32	0,64	0,44

2.
a) $\overline{X}_A$= 74,13 $\overline{X}_B$= 113,38 $\overline{X}_G$= 93,75
b) S_A= 34,17 S_B= 34,95
c) CV_A = 0,46 CV_B = 0,308 En B

3. a) $\overline{X}$ = 6,253
b) CV_A=0,279 CV_B =0,259 CV_C=0,241 CV_D=0,325. El instituto C
c) X'_D=3,5482+0,5128X_D

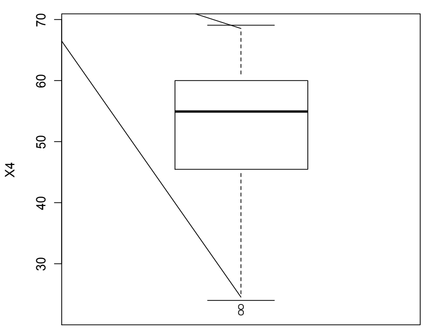
4. a) $\overline{X}_A$= 630 $\overline{X}_B$= 583 $\overline{X}_C$= 632 $\overline{X}_D$= 630
b) Δ M_A = 40.000 € Δ M_B= 16.500 € Δ M_C = 41.000 € Δ M_D = 40.000 €
c) CV_A= 47,62% CV_B= 54,54% CV_C= 49,36% CV_D =50%

5. a) El empleado de la agencia 1
b) Aproximadamente 3 clientes

6. Tarde

7. a) Barcelones, Tarragones
b) No
c) X_G= 1812,61 X_B= 3993,46 X_T= 3402,52 X_{BLl}= 1066,9 X_{VA}= 1489

8.



9. X2, X1 y X3

10. a) $\overline{X}_R$= 11,5 mn $\overline{X}_T$= 9,2 mn
b) 15 mn
c) Z_A= 0,757 Z_B= 0,645 En la compañía A
d) CV_A = 0,83 CV_B = 0,77 En la compañía A
e) $\overline{X}_R$= 10,925 mn S= 6,51 $\overline{X}_T$= 8,74 mn

TEMA 5

1.

	H	M	Total
16-19	1	1	2
20-24	2	1	3
25-54	13	9	22
más de 55	2	1	3
Total	18	12	30

2.

a) 2,3%
b) 8,55%
c) 2,3%
d) 36,5%
e) 11,9%
f) 24,7%
g) Pasear
h) Mayores de 65
i) Aproximadamente 46 años
j) Tareas del hogar/cocinar
k) Deportes.
l) Mayores de 65

3.

a) No son independientes ya que, por ejemplo, $f(UP, Esp) \neq f(UP)\ f(Esp)$
b) 6,5%
c)10,6%

	E	F	I	P	G	n(X)
UP	22.6	22.6	22.6	22.6	22.6	113
ASN	52.0	52.0	52.0	52.0	52.0	260
UAN	5.2	5.2	5.2	5.2	5.2	26
DAN	20.2	20.2	20.2	20.2	20.2	101
n(Y)	100	100	100	100	100	500

4.

a) 0,15
b) 0,15
c) 100
d) 60
e) 24
f) 15
g) 1
h) 1
i) 0,0375
j) 0,25
k) 0,30
l) M4

5.

a)

X \ Y	1	2	3	Total
0	0	0	1	1
1	5	0	0	5
2	10	0	0	10
3	0	9	0	9
Total	15	9	1	25

b) 52 unidades de A y 36 unidades de B

TEMA 6

1. a) S_{XY} = 5,52 r= 0,953
   b) S_{XY} = 5,80 r= 0,953

2. a) $\begin{bmatrix} 2{,}08 & 1{,}44 \end{bmatrix}'$, $\begin{bmatrix} 0{,}7433 & 0{,}1717 \\ & 0{,}34 \end{bmatrix}$
   b) r_{XY} = 0,341
   c) S_{XY} = 0,6868 r_{XY} = 0,341

3. A r=0,32 B r= -1 C r=0,98
   D r=0,03 E r=-0,42 F r=-0,95

4. a) 0,892
   b) Y*= -0,2574 + 0,3887 X
   c) 79,6%

5. a) 0,89
   b) Presión*= 7,949+0,1166·Edad
   c) R^2 = 0,795
   d) Presión (m 51) = 13,896

6. a) Y*= 14,092 - 0,00634 X
   b) 58,25%
   c) Se reduce 0,0634 grados
   d) 8,069 grados

7. a) 0,411
   b) X_2* = 112,5+0,75X_1
   c) 83%
   d) 937,5

8. a) Y*= 11,5-0,75 X
   b) 0,83
   d) Un decremento de 0,75 unidades
   e) 4,75 unidades

10. a) $r_{(\text{Preu.de.venda; Metres.quadrats})}$= 0,951 < $r_{(\text{Preu.de.venda; Preu.inicial})}$= 0,9828.
    R^2_{II} = 0,966; el 96,6%
    b) 953,4€
    c) 139821,5€
    d) 254229,5€
    e) la del apartado c)

TEMA 7

1. a) E={(1,2) (1,3) (1,4) (1,5) (1,6) (1,7) (2,1) (2,3) (2,4) .... (7,1) (7,2) (7,3) (7,4) (7,5) (7,6)} En total 42 resultados elementales si se tiene en cuenta el orden de aparición.; o 21 resultados elementales si no se tiene en cuenta el orden.
   b) E = { 1, 2, 3, 4, ...} Infinitos resultados numerables
   c) E = $\{x \in \Re \ / \ 0 \le x \le 3\}$ Reales entre 0 y 3. Infinitos resultados continuos.
   d) E = $\{x \in \Re \ / \ 0 \le x \le \infty\}$ Reales no negativos
   e) E = { (0,5), (1,4), (2,3), (3,2), (4,1), (5,0)}

2. a) P(Éxito)=3/4
   b) P(copa) =1/4
   c) P(Superior a 7) = 15/36
   d) P(Accidente) = 1/200
   e) P(llueva) = (subjetiva)
   f) P(retraso) = 1/40
   g) P(primo) = 2/3
   h) P(3,2) =1/6

3. a) No. $P(A \cap B)=0,6 \neq 0$
   b) $P(\bar{A}) = 1 - PA) = 1 - 0,62 = 0,38$
   c) $P(A \cup B)=P(A)+P(B)-P(A \cap B) = 0,62 + 0,92 - 0,60 = 0,94$
   d) $P(B \cup C)=P(B)+P(C)-P(B \cap C) = 0,92 + 0,11 - 0,11 = 0,92$
   e) $P(\bar{A} \cap \bar{B})=1-P(A \cup B) = 1 - 0,94 = 0,06$
   f) $P(A \cup B \cup C)=P(A)+P(B)+P(C)-P(A \cap B)-P(A \cap C)-P(B \cap C)+P(A \cap B \cap C) =$
   $= 0,62+ 0,92 + 0,11 - 0,60 - 0,06 - 0,11 + 0,06 = 0,94$
   g) $P(\bar{A} \cap \bar{B} \cap \bar{C})=1-P(A \cup B \cup C) = 1 - 0,94 = 0,06$
   h) $P(B \cap \bar{C})=P(B)-P(B \cap C) = 0,92 - 0,11=0,81$

4. a) 2/15
   b) 3/11
   c) 4/15

5. a)

	Blanco (B)	Color (C)	
Si (A)	0,05	0,10	0,15
No (A^C)	0,75	0,10	0,85
	0,80	0,20	1

   b) P(A/B)=0,05/0,80 = 0,0625
   c) P(C/A)=0,10/0,15 = 0,667
   d) $P(A^C/C)$=0,10/0,20 = 0,5

6. a) 0,15
   b) 0,10
   c) 0,39
   d) 0,36
   e) 0,54
   f) 0,46

g) 0,2778
h) 0,2174

7. 0,016

8. a) 0,18
b) 0,42

9. a) 0.52
b) No $P(A \cap B) \neq 0$
c) Si $P(A \cap B) = P(A)\ P(B)$

10. a) No. $P(B \cap C)=0,11$ $P(B)P(C)=0,1012$
b) Sí. $P(B) = 0,92$ y $P(B/A) = 0,9677$
c) $P(A/\ B \cap C) = 0,5454$
d) $P(A/B) = 0,652$
e) $P(A/\ A \cup B \cup C) = 0,6595$

11. a) 0,72
b) 0,02
c) 0,26

12. 0,6

13. Del tipo V1 o V2 con probabilidad 0,375

14. a) 0,625
b) 0,375
c) 0,4; 0,3; 0,2; 0,1
d) 0; 0,167; 0,333; 0,5

15. 0,346

16. 0,25

17. 0,779

18. 0,3

19. a) 0,0588
b) 0,9412
c) 0,6734
d) 0,9996
e) 0,9804

20. 0,41
21. 1/3
22. 0,75
23. 1/3

TEMA 8

1.

X	0	1	2	3
P(X)	0,216	0,432	0,288	0,064

2. a)

X	1	2	3	4
P(X)	0,4	0,3	0,2	0,1

b)

X	1	2	3	4	5	6	7	8	...
P(X)	0,4	0,24	0,144	0,086	0,052	0,031	0,018	0,011	...

$P(x)=0,6^{x-1}0,4 \quad x=1,2,3,4...$

3.

X	1	2	3	4	5
P(X)	0,5	$0,5^2$	$0,5^3$	$0,5^4$	$0,5^4$

4. b) 0,99
   c) 0,13
   d) 0,6
   e) 0,333

5. a) k=12/25 =0,48

b) $F(x)=\begin{cases} 0 & x<1 \\ 0,48 & 1 \leq x<2 \\ 0,72 & 2 \leq x<3 \\ 0,88 & 3 \leq x<4 \\ 1 & x \geq 4 \end{cases}$

c) $P(1\leq X<3)=0,72 \quad P(2<X\leq 4)=0,28 \quad P(X>2)=0,28 \quad P(X<2/X<4)=0,545$

6. a) $\sum_{\forall x} P(x)=1$

b) $F(x)=\begin{cases} 0 & x<1 \\ 0,05 & 1 \leq x<2 \\ 0,15 & 2 \leq x<3 \\ 0,30 & 3 \leq x<4 \\ 0,50 & 4 \leq x<5 \\ 0,70 & 5 \leq x<6 \\ 0,85 & 6 \leq x<7 \\ 0,95 & 7 \leq x<8 \\ 1 & x \geq 8 \end{cases}$

c) $P(X<4)=0,3 \quad P(1<X\leq 4)=0,45 \quad P(X>2)=0,85 \quad P(3\leq X<6)=0,55$
$P(X\geq 3/X<7)=0,82$

7. a) 0,0387
   b) 0,3439

c) 0,81
d) 0,27

8. a) Discreta
b)

X	1/8	1/4	3/8
P(X)	0,2	0,7	0,1

9. b) 0,0703

c) $F(x)=\begin{cases} 0 & x<-1 \\ \dfrac{x^3+1}{2} & -1 \le x \le 1 \\ 1 & x>1 \end{cases}$

10. a) 4

b) $F(x)=\begin{cases} 0 & x<1 \\ \dfrac{x-1}{4} & 1 \le x \le 5 \\ 1 & x>5 \end{cases}$

c) $P(X\leq 3)=0,5$, $P(1,5\leq X<4)=0,625$, $P(X>2)=0,75$, $P(X>3/X>2)=0,667$

11. a) 3
b) 0,992

12. a) A- Discreta, B-Continua
b) A- $P(X\leq 2)=0,3$ $P(1\leq X<3)=0,2$ $P(X>2)=0,7$ $P(X>1/X<2,5)=0,667$
B- $P(X\leq 2)=0,125$ $P(1\leq X<3)=0,406$, $P(X>2)=0,875$
$P(X>1/X<2,5)=0,936$

13. a) 0,55
b) 0,208
c) 0
d) 0,378
e) 0,552

14. a) k=1/2
b) 0,4375
c) 71,09%
d) 0,125
e) 10,7368 Kg

15. a) 0,5
b) 0,05
c) 0,25
d) No. CV(X) < CV(Y)

16. a) 0,25
b) 0,2727

c)

X	3	5	7	8
P(X)	1/12	3/12	5/12	3/12

d) Me=7, E(X)=6,417 D(X)=1,5

17. E(X1) = 650   E(X2)=250   E(X3)=400
V(X1) = 15427500 V(X2)=562500 V(X3)=0
CV(X1) = 604% CV(X2)=300%

18. E(Coste)=273

19. a) 0,75
b) 0,625
c) E(X)=49,8 V(X)=1,66
d) E(C)=114,6 V(C)= 6,64

20. E(X) = 0 V(X) = 0,6 D(X) = 0,774 (ejercicio 9)
E(X) = 2,25 V(X) = 0,3375 D(X) = 0,58 (ejercicio 11)
$E(X_A)$ = 2,9 V(X) = 1,49 D(X) = 1,22 (ejercicio 12)
$E(X_B)$ = 3 V(X) = 0,6 D(X) = 0,774 (ejercicio 12)

21. a) k= 3
b) E(X) = 1,5 Me = 2 Mo = 4
c) V(X) = 6,25 D(X) = 2,5
d)

Z	-1,4	-0,2	0,2	1
P(Z)	0,3	0,1	0,2	0,4

22. a)

X	0	1	2	3
P(X)	0,42	0,42	0,14	0,016

b) E(X) = 0,75
c) V(X) = 0,56

23. 0€

24. 5€

25. a) 2,44
b) 2,26
c) 3,36
d) V(X) = 1,13 D(X) = 1,06
e) Si

TEMA 9

1. a) 0,311
   b) 0,5
   c) 0,5
   d) 0,6329
   e) 0,1875
   f) 0,4744

2. a) 0,1240
   b) 0,9743
   c) 0,1497

3. a) 0,5987
   b) 0,086
   c) 0,9139
   d) 0
   e) 0,7852
   f) 0,5993
   g) 0
   h) 0,2578

4. a) E(X)=5 D(X)=1,94
   b) E(Y)=1,25 D(Y)=2,42

5. a) 0,1160
   b) 0,2373
   c) 0,0275

6. a) 0,0755
   b) V(X) =0,95 D(X)=0,975
   c) 0,8286
   d) 0,08614

7. a) 0,4752
   b) 0,1792
   c) E(X)=800 D(X)=489,9

8. 0,9162

9. a) 0,9772
   b) 0,1922
   c) 0,3086
   d) 0,835
   e) 0,9544
   f) 0,5154

10. a) 65,3
    b) 11
    c) 216,5
    d) 6,8

11. a) 61,79%
    b) 8,95
    c) 23
    d) 23,1 puntos

12. a) 0,15866
    b) 0,31084
    c) 617
    d) 0,88493
    e) Inferior a 448 gr.
    f) 0,6777

13. 166500 litros

14. a) 0,04746
    b) 4,992 mn

15. 71,6%

BIBLIOGRAFIA

ALEA, V., MAQUEDA, I., MUÑOZ, MC., TORRELLES; E., VILADOMIU, N. (2001) Estadística para las ciencias Sociales: Cuestiones tipo test Madrid, Thomson

ALEA, V., MAQUEDA, I., MUÑOZ, MC., TORRELLES; E., VILADOMIU, N. (2009) Introducción al análisis estadístico con R-Commander. Barcelona, UB

LIND, A.D. et. al. (2008) Estadística aplicada a los negocios y la economía. Madrid, Mc Graw-Hill

MARTÍN-GUZMÁN, P (2006) Manual de Estadística: Descriptiva. Madrid, Thomson

MONTIEL, A.M. et. al. (1997) Elementos básicos de Estadística Económica y Empresarial. Madrid, Prentice-Hall

NEWBOLD, P. et. al. (1997) Estadística para los negocios y la economía. Madrid, Prentice Hall

PEÑA, D., ROMO, J. (1997) Introducción a la estadística para las ciencias sociales. Madrid, McGraw-Hill