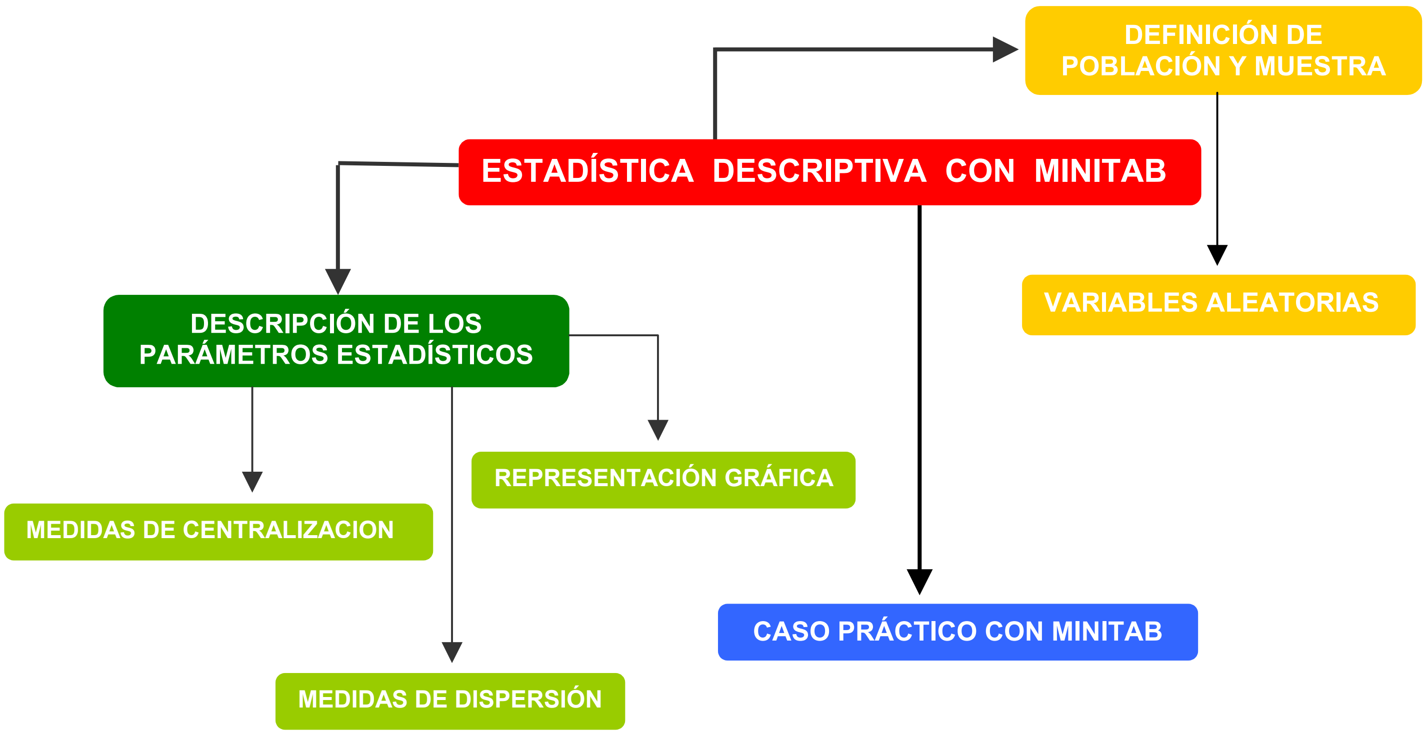
ESTADÍSTICA DESCRIPTIVA CON MINITAB

Autores: Angel Alejandro Juan (ajuanp@uoc.edu) , Maximo Sedano (msedanoh@uoc.edu) , Alicia Vila (avilag@uoc.edu) .

MAPA CONCEPTUAL



INTRODUCCIÓN

La estadística se encuentra frecuentemente en nuestro lenguaje cotidiano. Por ejemplo, cuando hacemos referencia a "la media del salario de los empleados de una determinada empresa" o "las variaciones de las cifras del Dow Jones en la última semana".

Así, podríamos definir la Estadística como la ciencia encargada de reunir, organizar, presentar, analizar e interpretar datos con el fin de obtener unas determinadas conclusiones y tomar unas determinadas decisiones[1].

En general, la estadística se divide en dos categorías:

- *Estadística descriptiva,* que es la parte de la estadística encargada de extraer y organizar los datos procedentes de un determinado conjunto de observaciones.

- *Estadística inferencial,* que pretende predecir una información acerca de un conjunto de datos, a partir de los resultados extraídos de un subconjunto de ellos.

Los pasos a seguir para realizar una investigación estadística serían los siguientes:

Problema → Recogida y Organización de datos → Análisis e interpretación → Conclusiones y decisiones

En este apartado nos encargaremos únicamente de definir los parámetros correspondientes a la estadística descriptiva para organizar, describir y analizar una colección de datos, así como las posibles representaciones gráficas de éstos.

OBJETIVOS

- Cálculo e interpretación de los parámetros de centralización: media aritmética, mediana y moda.

- Cálculo e interpretación de los parámetros de dispersión: rango, varianza y desviación estándar.

- Cálculo e interpretación de los cuartiles, rango intercuartílico y coeficiente de variación.

- Representación gráfica de los datos.

CONOCIMIENTOS PREVIOS

Población: Conjunto de objetos, individuos o sucesos cuyas propiedades queremos analizar.

Muestra: Subconjunto de la población objeto de estudio.

El conjunto de los datos recogidos para llevar a cabo un estudio estadístico, recibirán el nombre de **variable aleatoria** que normalmente se denotará X. Los datos correspondientes pueden ser, básicamente, de dos tipos:

- **Variable cualitativa**: los datos que forman la variable no son numéricos. Por ejemplo, religión a la que se pertenece, tipo de automóvil, color de los ojos,...

- **Variable cuantitativa**: los datos que forman la variable sí son numéricos. Por ejemplo, saldo de una cuenta corriente, velocidad de los coches,...
  Dentro de este grupo de variables podemos distinguir otras dos categorías. Por una parte, tendríamos las **variables discretas** que serían aquellas que sólo pueden asumir ciertos valores (por ejemplo, número de automóviles/h. en una autopista, número de estudiantes en la asignatura de estadística,...) y por otra parte, las **variables continuas** que pueden tomar cualquier valor dentro de un rango específico (por ejemplo, alturas de los alumnos de una clase, tiempo transcurrido en el vuelo de Barcelona a Madrid,...)

Por último, una vez tenemos recogidos los datos, agrupamos éstos de forma excluyente dando a cada uno de ellos el número de observaciones, es decir, el número de veces que se repite cada valor, este número tomará el nombre de **frecuencia**.

CONCEPTOS FUNDAMENTALES

Las técnicas utilizadas para la descripción de datos se dividen, básicamente, en dos bloques:

- **Parámetros de centralización**: Son aquellos cuyo objetivo es explicar mediante un valor numérico, cual es la tendencia mayoritaria de las observaciones de la colección de datos que se analizan. Dichos parámetros serán, entre otros, la media, la mediana y la moda.

- **Parámetros de dispersión:** Corresponden a aquellos parámetros cuyo objetivo es detectar el grado de proximidad de los datos respecto a los valores centrales. Dichos parámetros serán, entre otros, el rango, los cuartiles, la varianza y la desviación estándar.

❑ Medidas de centralización

Las medidas de centralización nos sirven para representar el valor medio de los datos, es decir, el valor que refleja el tamaño del dato más esperado. Ello nos indica la posición en la que se encuentra en el centro de los datos. [2]

Las medidas de centralización más utilizadas son:

Media : Es la suma de un conjunto de observaciones dividido por el número total de observaciones realizadas.

Si calculamos la media poblacional, la expresión será la siguiente:

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

En cambio, si lo que estamos calculando es la media muestral, la notación será la que sigue:

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

siendo x_i cada uno de los valores de la distribución, y n el número de observaciones.

Ejemplo:

La biblioteca virtual de la UOC quiere conocer el tiempo medio que tardan los usuarios en devolver los préstamos. Se ha tomado una muestra de 15 usuarios, obteniendo los siguientes tiempos, en días, de devolución:

U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
10	20	12	14	16	18	22	10	16	13	21	15	12	20	18

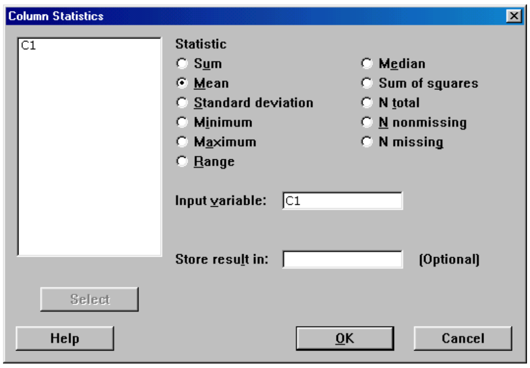
Para calcular la media aritmética de esta distribución,lo haremos de dos formas:

<u>Manualmente:</u>

$$\bar{X} = \frac{1}{15}\sum_{i=1}^{15} x_i = \frac{1}{15}(10 + 20 + ... + 18) = 15.8$$

<u>Con Minitab:</u>

Introducimos los valores en el espacio de trabajo y seleccionamos, *Calc > Column Statistics* y activamos la opción *Mean*:




```
Column Mean

Mean of C1 = 15.800
```

Mediana : Es un número tal que, si ordenamos los datos de forma creciente o decreciente, cumple la condición de ser mayor que una mitad y menor que la otra. Es decir, divide a la distribución en dos partes iguales.

Si el número de observaciones es impar la mediana es el valor central. En caso de que el número de observaciones sea par la mediana será la media de los dos valores centrales.

Ejemplo:

Siguiendo con el ejemplo anterior, ordenamos los datos en orden creciente:

10 10 12 12 13 14 15 16 16 16 18 20 20 21 22

Para calcular la mediana lo haremos también de dos formas:

Manualmente:

Como el número de observaciones es impar, la mediana será el valor que ocupa el lugar central, en este caso, la posición octava. Por tanto, la mediana será 16.

Con Minitab:

Seleccionamos, *Calc* > *Column Statistics* y activamos la opción *Median*:

```
Column Median

Median of C1 = 16.000
```

Moda : Es el valor que más veces se repite en la distribución. Si los datos de la distribución están agrupados en intervalos, la moda es el punto medio del intervalo que contiene el mayor número de frecuencias.
Una distribución de observaciones puede no tener moda, es decir, puede que no haya ningún valor de la distribución que aparezca con más frecuencia.

Ejemplo:

Siguiendo con el ejemplo anterior, agrupamos los datos según su frecuencia:

10 -> 2
12 -> 2
13 -> 1
14 -> 1
15 -> 1
16 -> 3
18 -> 1
20 -> 2
21 -> 1
22 -> 1

Por tanto, observamos que la moda será el 16 que es valor que más veces se repite.

Relación entre el valor de la media y la mediana de una distribución

Tanto la media aritmética como la mediana miden el centro de la distribución, pero lo hacen de formas diferentes. En el caso en que la distribución sea simétrica ambas medidas son iguales. Si la distribución es asimétrica, la media aritmética se desplaza hacia la cola de la distribución.
Observar que para calcular la media aritmética, utilizamos todos los datos, sin embargo, no ocurre así con la mediana. Así, si hay valores extremos, la media se verá mucho más afectada que la mediana.

Medidas de dispersión

Para conocer con detalle un conjunto de datos, no basta con conocer las medidas de tendencia central, necesitamos conocer también la dispersión que presentan los datos en su distribución, con objeto de tener una visión de los mismos más acorde con la realidad a la hora de describirlos e interpretarlos.

Recorrido o amplitud: Es la diferencia entre el valor máximo de la distribución y el valor mínimo.

Ejemplo:

Del ejemplo anterior, observamos que el valor máximo es 22 y el valor mínimo es 10, por tanto el recorrido de la distribución será 12.

Cuartiles: Así como la mediana divide a la distribución en mitades, los cuartiles de una distribución son los valores que dividen la distribución en cuartos.
El *primer cuartil* (Q1) deja un cuarto de las observaciones por debajo del mismo, y tres cuartos por encima. El *segundo cuartil* (Q2) tiene dos cuartos por debajo y dos por encima (el segundo cuartil coincide con la mediana). El *tercer cuartil* (Q3) tiene tres cuartos de las observaciones por debajo y un cuarto por encima.

Ejemplo:

Utilizamos los datos ordenados del ejemplo para calcular los cuartiles:

10 10 12 12 13 14 15 16 16 16 18 20 20 21 22

Manualmente:

Partimos en dos la distribución, la mitad superior:

10 10 12 12 13 14 15 → Q1=12

Y la mitad inferior:

16 16 18 20 20 21 22 → Q3=20

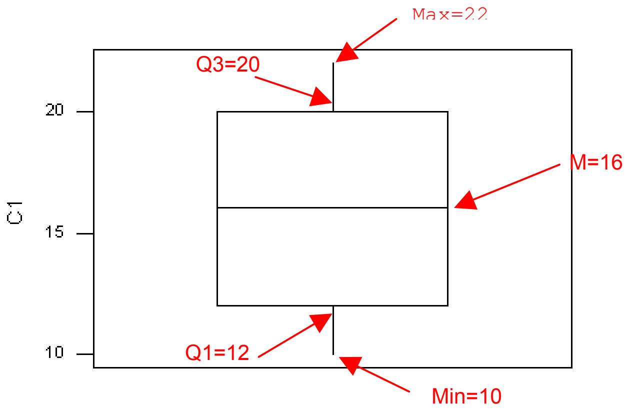
Para calcular los cuartiles calculamos las medianas de ambas mitades, la de la primera mitad corresponderá al primer cuartil y la de la segunda corresponderá al tercer cuartil.

Diagramas de caja

El diagrama de caja es un gráfico simple donde vienen representados los anteriores 5 valores anteriores (mínimo, primer cuartil, mediana (segundo cuartil), tercer cuartil y máximo)

Ejemplo:

Seleccionamos *Graph* > *Boxplot*:



Varianza: Se define como la media aritmética de las desviaciones de la media elevadas al cuadrado.

En el caso de la varianza, la expresión de la varianza poblacional y la muestral difieren ligeramente.

Varianza poblacional $\rightarrow \sigma^2 = \dfrac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$

Varianza muestral $\rightarrow s^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

Desviación Estándar: Es la raíz cuadrada positiva de la varianza.

Ejemplo:

En este caso, tras haber definido los parámetros estadísticos más importantes, veamos cómo, haciendo uso del Minitab, obtenemos un resumen práctico y sencillo de todos estos valores.

Seleccionar *Stat* > *Basic Statistics* > *Display Descriptive Statistics*:

```
Descriptive Statistics

Variable          N      Mean    Median    TrMean     StDev   SE Mean
C1               15     15.80     16.00     15.77      3.95      1.02

Variable    Minimum   Maximum        Q1        Q3
C1            10.00     22.00     12.00     20.00
```

En resumen,

- **N** es el número de observaciones que contiene la columna.
- **Mean** es la media de la muestra: $\bar{x} = \frac{\sum x_i}{n}$. La media es muy sensible a los valores extremos, por lo que también es interesante fijarse en la TrMean.
- **Median** es la mediana o cuartil segundo Q2 (aquel valor que deja a su izquierda el mismo número de observaciones que a su derecha). Este parámetro no se ve gravemente afectado por los valores extremos.
- **TrMean** es la media “recortada”: las observaciones se ordenan de menor a mayor y se descartan los valores extremos (un 5% por cada lado). De los valores restantes se calcula la media.
- **StDev** es la desviación estándar de la muestra: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$.
- **SE Mean** es el error estándar de la media, i.e.: $SEMean = StDev / \sqrt{N}$.
- **Minimum** y **Maximum** son los valores mínimo y máximo de los datos.
- **Q1** es el cuartil primero: aquel que deja a su izquierda un 25% de los datos.
- **Q3** es el cuartil tercero: aquel que deja a su izquierda un 75% de los datos.

Coeficiente de variación: Es la relación entre la desviación estándar y la media.

$$CV = \frac{s}{\bar{X}} 100$$

El coeficiente de variación es muy útil cuando queremos comparar dos o más medidas de dispersión y éstas están en unidades diferentes o bien están en las mismas unidades pero sus medias son muy distintas.

Ejemplo:

Pretendemos comparar la variación entre dos pruebas (una de aptitudes mecánicas y otra de destreza mental) realizadas a un grupo de aprendices pertenecientes al cuerpo de bomberos. La media aritmética de las puntuaciones obtenidas en la prueba de aptitudes mecánicas fue 200, con una desviación estándar de 10. En la segunda prueba los resultados fueron de media 30 y desviación estándar 6.

Para realizar dicha comparación calculamos el coeficiente de variación de ambas pruebas:

CV= 10/200*(100) = 5
CV= 6/30*(100) = 20

Por tanto, de los datos anteriores, deducimos que existe mayor dispersión en la prueba de destreza mental.

Representación gráfica

Dos de las técnicas estadísticas más importantes para representar un conjunto de datos son las siguientes:

Diagrama de tallo y hojas: Cada valor numérico se divide en dos partes. Los dígitos principales forman el tallo y los dígitos secundarios las hojas. Los tallos están colocados a lo largo del eje vertical, y las hojas de cada observación a lo largo del eje horizontal.

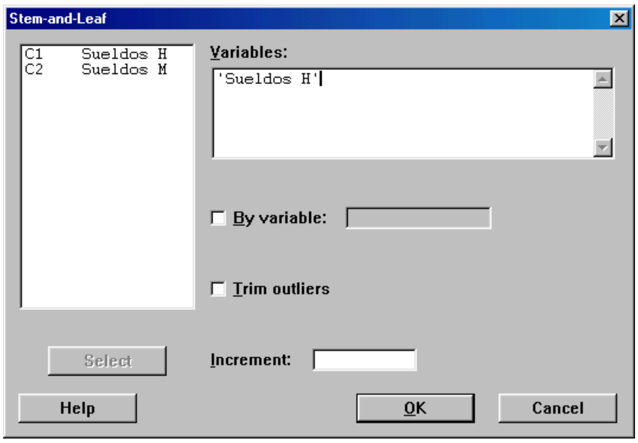
Ejemplo:

A continuación, se muestran los salarios anuales (en €) de los 11 trabajadores del departamento de marketing de una empresa de material informático:

Sueldos H	Sueldos M
38985	28938
29548	32920
41889	24749
31528	39828
38791	28985
32782	

Construir un diagrama de tallo y hojas para cada variable.

Para ello, tras introducir los datos en el espacio de trabajo del Minitab, seleccionamos *Graph* > *Stem-and-Leaf* y rellenamos los campos de la siguiente manera:




```
Character Stem-and-Leaf Display

Stem-and-leaf of Sueldos   N  = 6
Leaf Unit = 1000

    1    2 9
    2    3 1
    3    3 2
    3    3
    3    3
    3    3 88
    1    4 1
```

Análogamente resolveríamos para los sueldos de las mujeres:

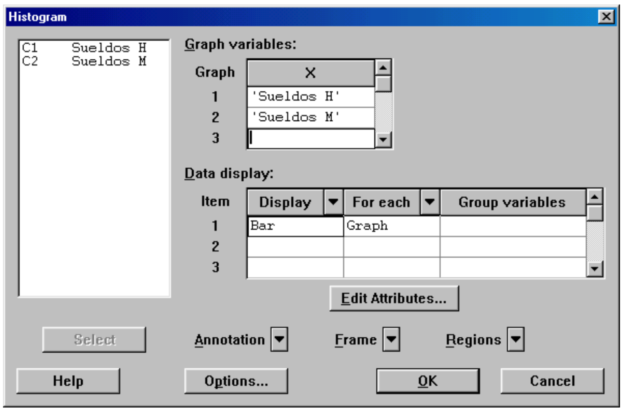
```
Character Stem-and-Leaf Display

Stem-and-leaf of Sueldos   N  = 5
Leaf Unit = 1000

 1    2 4
(2)   2 88
 2    3 2
 1    3 9
```

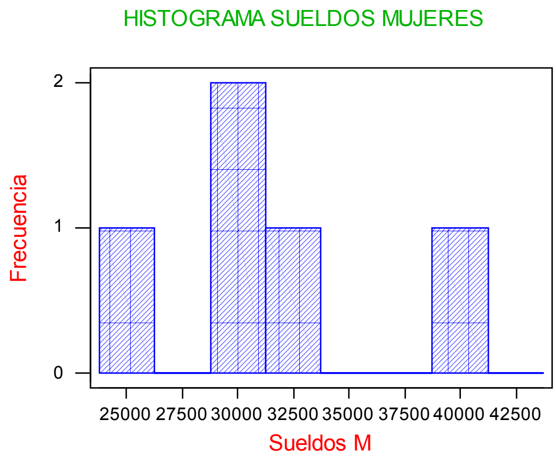
Histograma: Describe una distribución de frecuencias usando una serie de rectángulos adyacentes, en los que la altura de cada rectángulo es proporciona la frecuencia que representa el valor de la variable.

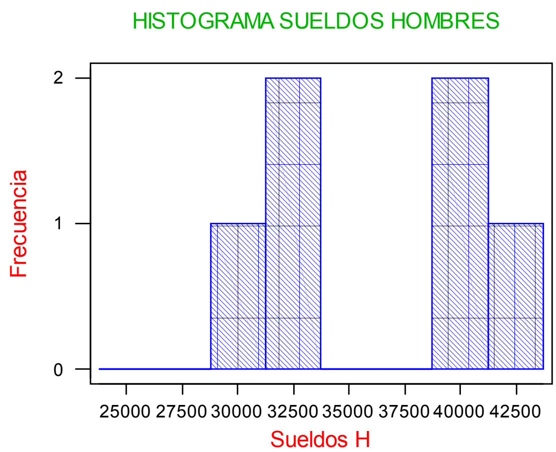
Ejemplo:

Siguiendo el ejemplo anterior, construiremos dos histogramas para comparar gráficamente la diferencia entre los sueldos de ambos sexos.

Para ello, seleccionamos *Graph > Histogram*, completando los campos como sigue:





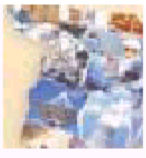


De ambos histogramas deducimos que, en este departamento, los sueldos de los hombres, son ligeramente más altos que los de las mujeres.

CASO PRÁCTICO CON SOFTWARE

1. Teniendo en cuenta la importancia que los beneficios económicos del turismo tienen en nuestro país, realicemos un breve análisis acerca de cuál ha sido la ocupación hotelera durante el año 2000.

   Para ello, vamos a www.ine.es y seleccionamos la opción *España en cifras 2001*. Posteriormente, escogemos la opción *Turismo*. Comprobaremos que la información que obtenemos es la siguiente:



Turismo

Ocupación hotelera. 2000

	Viajeros			Estancia media (días)		Grado de ocupación
	Total	Españoles	Extranjeros	Españoles	Extranjeros	
TOTAL	59.282.522	32.132.992	27.149.531	2,59	5,30	58,86
Enero	2.808.737	1.738.456	1.070.281	2,37	5,34	40,76
Febrero	3.390.877	2.098.381	1.292.496	2,28	5,17	48,22
Marzo	4.244.207	2.405.046	1.839.161	2,38	4,79	54,10
Abril	5.335.325	2.941.255	2.394.070	2,53	4,41	58,30
Mayo	5.592.100	2.656.249	2.935.851	2,38	5,00	56,34
Junio	5.806.706	2.853.802	2.952.904	2,54	5,70	63,05
Julio	6.374.389	3.187.268	3.187.122	2,98	5,86	68,86
Agosto	6.994.683	3.805.934	3.188.749	3,29	5,89	75,73
Septiembre	6.327.590	3.188.546	3.139.043	2,75	5,55	68,55
Octubre	5.368.961	2.798.565	2.570.396	2,44	5,28	58,30
Noviembre	3.691.092	2.283.268	1.407.825	2,28	4,73	49,09
Diciembre	3.347.855	2.176.222	1.171.633	2,27	4,60	42,45

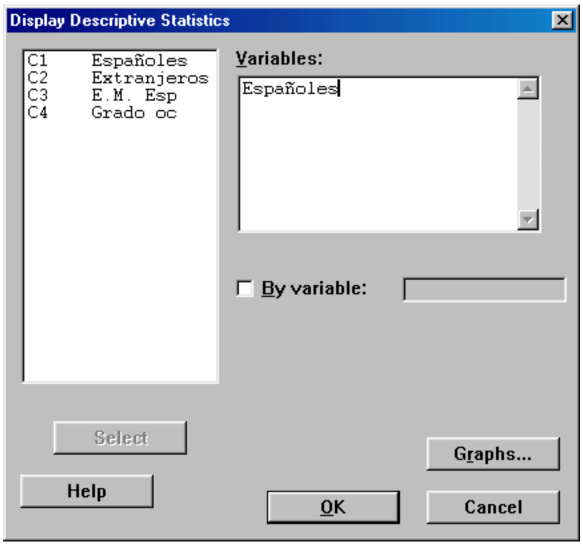
   a) Haciendo uso del Minitab, calcular todos los parámetros estadísticos correspondientes a la ocupación hotelera de *Españoles* a lo largo del año 2000.

      Una vez introducidos todos los datos correspondientes en el worksheet:

Worksheet 1 ***

	C1	C2	C3	C4	C5
↓	Españoles	Extranjeros	E.M.Esp	Grau Oc	
1	1738456	1070281	2.37	40.76	
2	2098381	1292496	2.28	48.22	
3	2405046	1839161	2.38	54.10	
4	2941255	2394070	2.53	58.30	
5	2656249	2935851	2.38	56.34	
6	2853802	2952904	2.54	63.05	
7	3187268	3187122	2.98	68.86	
8	3805934	3188749	3.29	75.73	
9	3188546	3139043	2.75	68.55	
10	2798565	2570396	2.44	58.30	
11	2283268	1407825	2.28	49.09	
12	2176222	1171633	2.27	42.45	

Seleccionamos *Stat* > *Basic Statistics* > *Display Descriptive Statistics*:

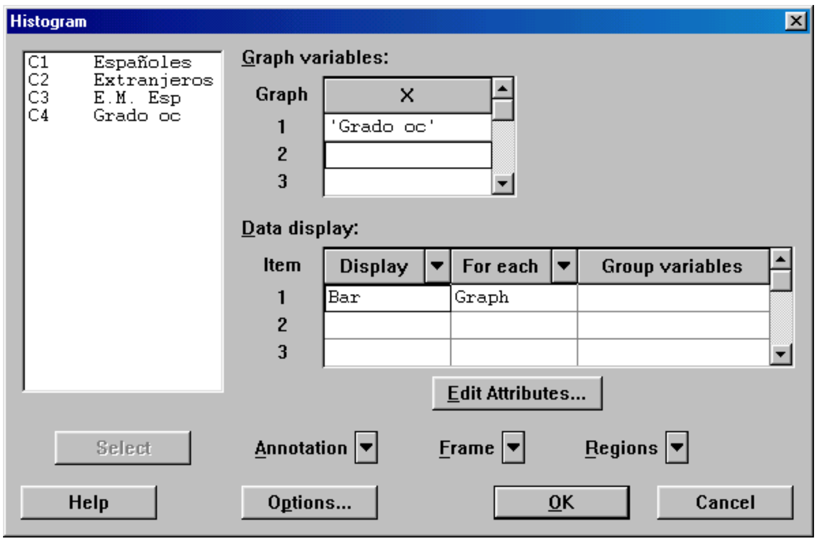


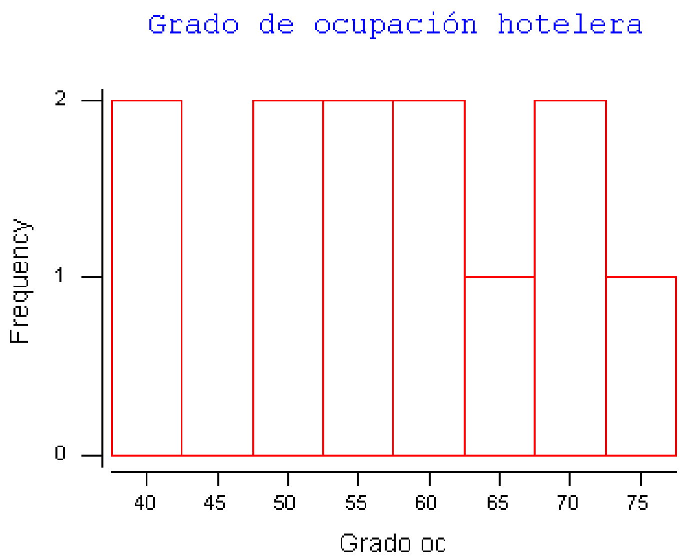

```
        Descriptive Statistics

        Variable             N       Mean     Median     TrMean      StDev    SE Mean
        Españole            12    2677749    2727407    2658860     572734     165334

        Variable       Minimum    Maximum         Q1         Q3
        Españole       1738456    3805934    2202983    3125765
```

Del anterior resultado, podemos observar cuáles son los parámetros estadísticos más importantes, por ejemplo, la media de españoles que residieron en algún hotel fue de 2677749, el valor central (mediana) es 2727407, así como 2202983 y 3125765 los valores de los cuartiles primero y tercero, respectivamente.

b) Construir un histograma del *Grado de ocupación* hotelera.

Para realizar un histograma de una colección de datos, seleccionamos la opción *Graph* > *Histogram*:



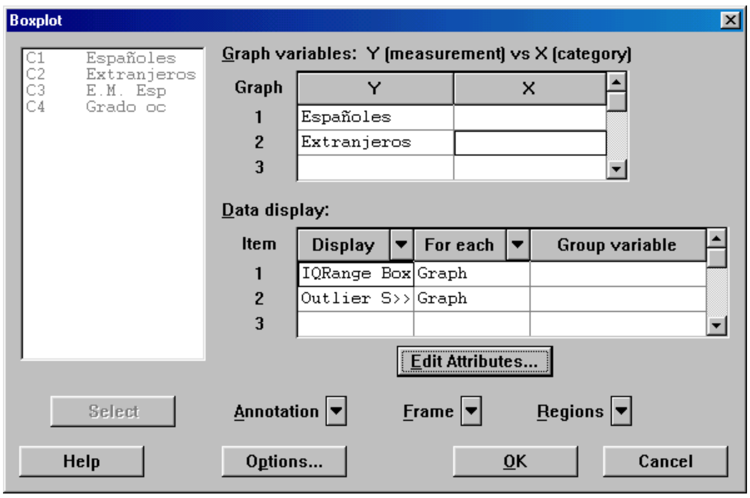


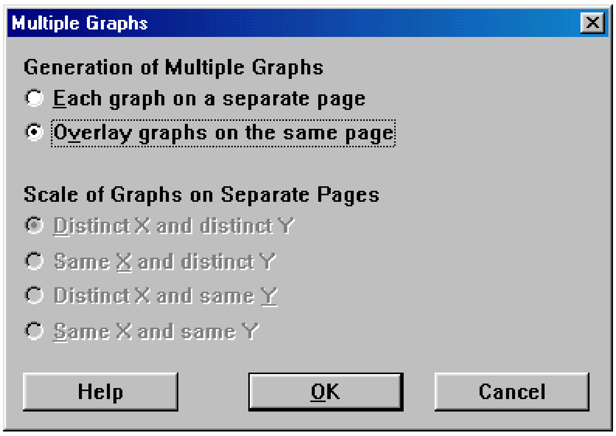
Nota.- Para introducir colores, texto, etc. en el gráfico seleccionaremos la opción de *Edit Attributes*, o bien, una vez realizado el gráfico clicando dos veces sobre la parte del gráfico que queremos modificar.

Del anterior gráfico podemos deducir, básicamente, que la ocupación hotelera se encuentra siempre por encima del 40% y ni siquiera en los meses de verano supera aproximadamente el 80%. Cabe destacar que son datos generales de todo el país, evidentemente, la ocupación hotelera en los meses de verano y vacaciones, los hoteles de la costa alcanzarán mayor cota de ocupación.

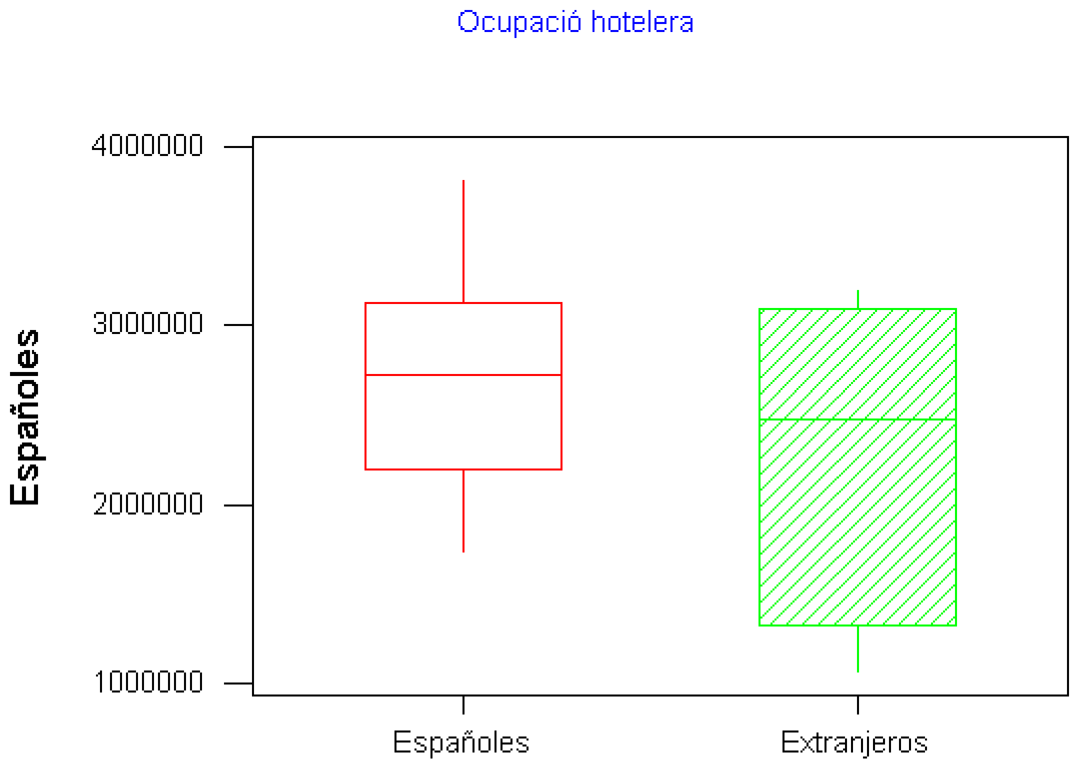
c) Dibujar los diagramas de cajas (Boxplot) correspondientes a la ocupación hotelera de Españoles y de Extranjeros, comentando las diferencias básicas entre ellos.

Seleccionamos *Graph* > *Boxplot*:





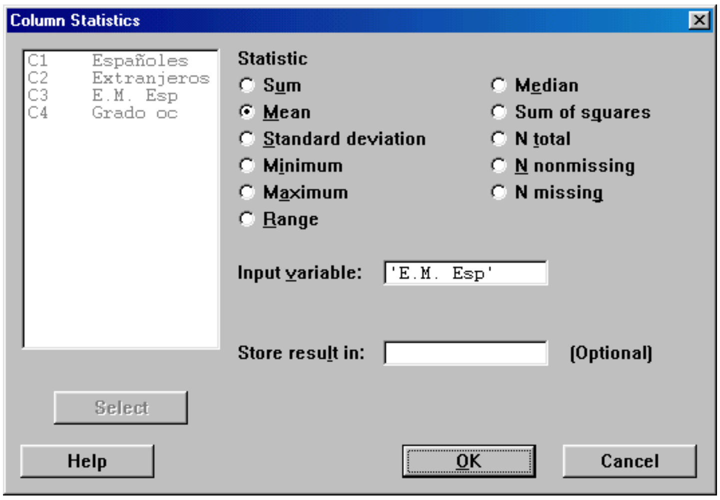
El resultado obtenido es el siguiente:



Del anterior gráfico cabe destacar que el valor de la mediana de turistas españoles es ligeramente superior a la de extranjeros, así como el resto de valores. Observamos que, por ejemplo, el valor del tercer cuartil de “Extranjeros” se acerca al valor del tercer cuartil de “Españoles”.
Observemos también que en ninguna de las dos observaciones aparecen valores extremos (“outliers”), si fuera así, aparecerían con un asterisco en la posición adecuada.

d) Calcular únicamente la media y la mediana correspondientes a la Estancia media en dias de la columna Españoles.

Cuando queremos calcular algun parámetro estadístico concreto, tenemos la opción de hacerlo seleccionando *Calc > Column Statitstics*, activando la opción elegida:



Nota.- Observar que, tras seleccionar la columna de la cual queremos calcular el parámetro estadístico, podemos guardar los resultados en una determinada columna o bien, dejar el espacio en blanco, obteniendo el resultado en la pantalla de *Session*.

```
Column Mean

Mean of E.M. Esp = 2.5408
```

Análogamente, procederemos para calcular el valor de la mediana, obteniendo el siguiente resultado:

```
Column Median

Median of E.M. Esp = 2.4100
```

Por tanto, deducimos que el número medio de estancia en días anual de turistas españoles es de aproximadamente 2.54, valor que difiere ligeramente de la mediana cuyo valor es 2.41.

2. Un tema de interesante análisis es la investigación en el área de educación en nuestro país, veamos cuál fue la cantidad de dinero invertida en el año 1999 en España y, comparativamente, en una determinada Comunidad Autónoma, por ejemplo, la de Valencia.

   Para ello, vamos a la página web del *Instituto Valenciano de Estadística*: http://ive.infocentre.gva.es/ . En el margen izquierdo de la página encontraremos un menú, del cual seleccionaremos la opción *"Información Estadística"* y *"Temas".* Ahora escogemos la opción *"Educación e investigación*" y seleccionamos "Investigación". Por último, elegimos *"Indicadores económicos de la actividad de investigación y desarrollo"*. El resultado que obtendréis, para el año 1999, será el siguiente:

Instituto Valenciano de Estadística

Indicadores económicos de la actividad de investigación y desarrollo

1999	C. Valenciana	España
Empresas		
Gastos internos (miles de PTA)	18.484.964	432.120.646
Investigadores	770,5	15.177,9
Técnicos en I+D	931,9	15.103,1
Auxiliares	610,4	8.042,1
Administración Pública		
Gastos internos (miles de PTA)	5.547.830	140.306.744
Investigadores	557,2	11.934,6
Técnicos en I+D	223,5	4.879,0
Auxiliares	209,1	5.469,7
Enseñanzas superiores		
Gastos internos (miles de PTA)	30.383.645	250.344.514
Investigadores	2.637,9	33.839,6
Técnicos en I+D	328,6	3.390,3
Auxiliares	628,2	3.396,2
Instituciones privadas sin lucro		
Gastos internos (miles de PTA)	854.752	8.385.773
Investigadores	104,7	615,6
Técnicos en I+D	25,6	221,3
Auxiliares	21,4	168,3

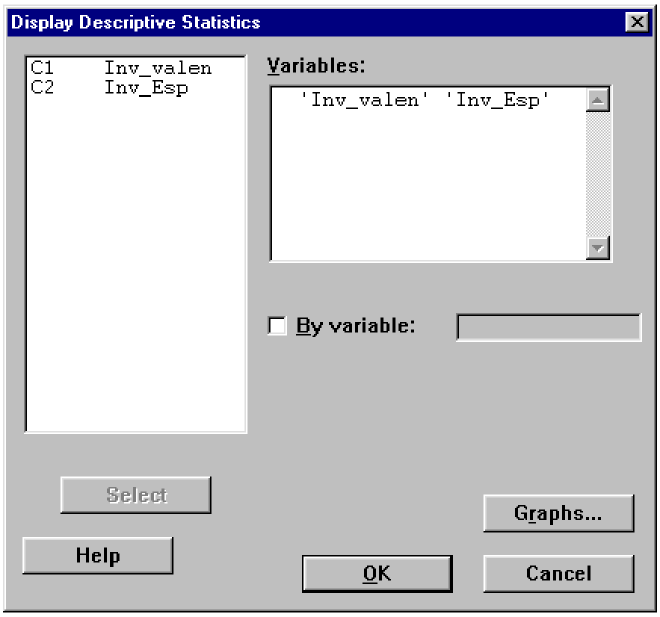
a) Calcula los estadísticos descriptivos correspondientes a la variable *Investigadores* en cada uno de los ámbitos, tanto en la Comunidad Valenciana como en España, y comenta los resultados.

   El primer paso será introducir los datos correspondientes en el espacio de trabajo de Minitab:

Worksheet 1 ***

	C1	C2
↓	Inv_valen	Inv_Esp
1	770,5	15177,9
2	557,2	11934,6
3	2637,9	33839,6
4	104,7	615,6

Seleccionamos *Stat* > *Basic Statistics* > *Display Descriptive Statistics*:



Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Inv_vale	4	1018	664	1018	1115	558
Inv_Esp	4	15392	13556	15392	13792	6896

Variable	Minimum	Maximum	Q1	Q3
Inv_vale	105	2638	218	2171
Inv_Esp	616	33840	3445	29174

En rojo, observamos los valores obtenidos más importante como son la media, la mediana, los cuartiles y los máximos y mínimos.

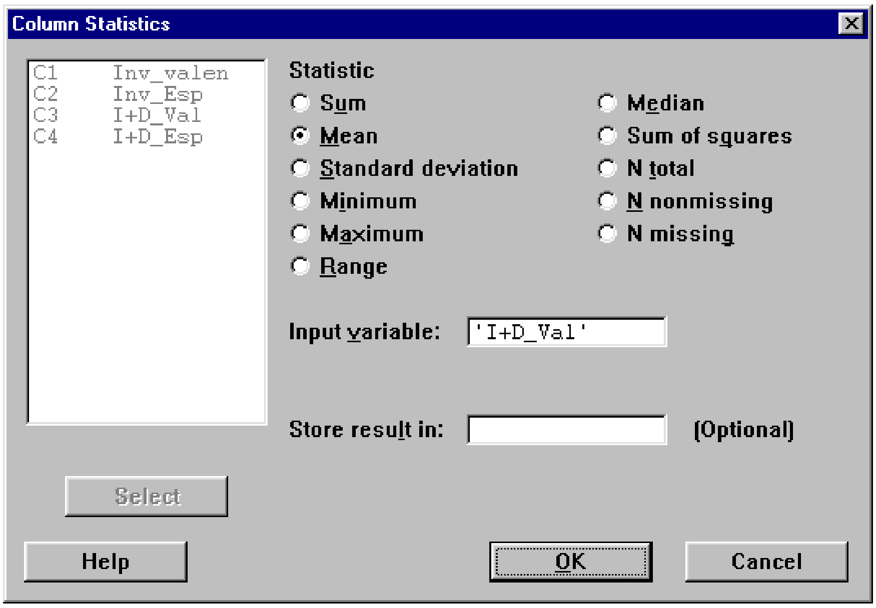
Cabe destacar la mínima inversión tanto en la Comunidad Valenciana como en España, se produce en *Instituciones privadas sin lucro*, correspondiendo la máxima inversión a las *Enseñanzas superiores*.

b) Como sabemos, el sesgo de una distribución depende de la simetría de ésta. Razona, si la distribución correspondiente a *Técnicos en I+D*, tanto en la Comunidad Valenciana como en España, es positivamente sesgada o negativamente sesgada. Recuerda que, para ello, tendrás que calcular la media y la mediana de ambas distribuciones.

Nuevamente, introducimos los valores correspondientes en el espacio de trabajo de Minitab:

C3	C4
I+D_Val	I+D_Esp
931,9	15103,1
223,5	4879,0
328,6	3390,3
25,6	221,3

Seleccionamos Calc > Column Statistics y rellenamos de la siguiente manera, para calcular la media de los *Técnicos I+D de Valencia*:




```
Column Mean

Mean of I+D_Val = 377,40
```

Análogamente, calculamos la mediana:

```
Column Median

Median of I+D_Val = 276,05
```

Por tanto, como la media aritmética es superior a la mediana, podríamos afirmar que en la distribución correspondiente a la variable “Técnicos en I+D” de la Comunidad Valenciana es positivamente sesgada.

Análogamente, resolvemos para la variable “Técnicos en I+D” de España, obteniendo los siguientes resultados:

```
Column Mean

Mean of I+D_Esp = 5898,4
```

```
Column Median

Median of I+D_Esp = 4134,6
```

Por tanto, en España, la distribución correspondiente a la variable “Técnicos en I+D” también es positivamente sesgada.

Desde la misma página web que estamos trabajando, y siguiendo los pasos anteriores hasta la opción “Educación e investigación”, seleccionamos esta vez la opción “Enseñanzas universitarias públicas” y “Profesores universitarios”, obteniendo el siguiente resultado:

Instituto Valenciano de Estadística

Profesores universitaris

Curso	Universidad de Alicante	U. Miguel Hernández	U. Jaume I	U. Politécnica de Valencia	Universidad de Valencia	Total
1996-97	1.379	-	653	1.951	2.935	6.918
1997-98	1.256	443	758	1.896	2.990	7.343
1998-99	1.430	611	789	2.057	3.028	7.915
1999-00	1.794	772	808	2.248	3.071	8.693
2000-01	1.900	767	814	2.387	3.106	8.974

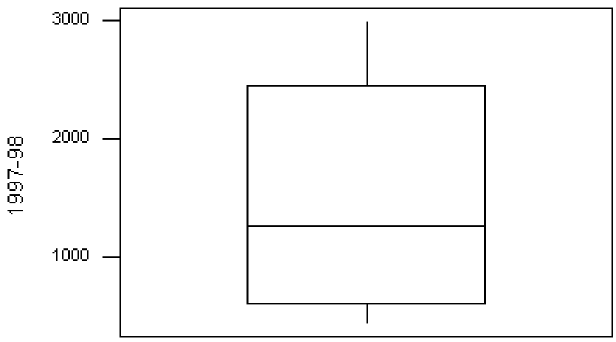
Fuente: Conselleria de Cultura i Educació. DG d'Ensenyaments Universitaris.

c) Representa en un diagrama de cajas, el número de profesores universitarios en cada una de las distintas universidades, correspondientes a los años 1997/98 y 2000/01.

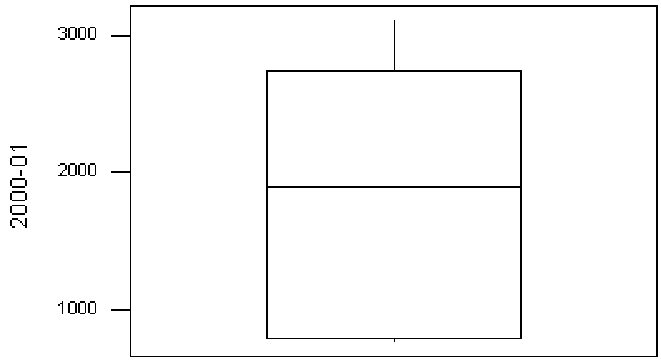
Introducimos los datos en el espacio de trabajo de Minitab:

C5	C6
1997-98	2000-01
1256	1900
443	767
758	814
1896	2387
2990	3106

Seleccionamos *Graph* > *Boxplot* y en la casilla X introducimos la variable 1997-98:



Análogamente, resolvemos para la variable 2000-01:



En los anteriores diagramas, podemos observar que la mediana del año 2000-01 es bastante superior a la del 1997-98, así como el máximo de ambas distribuciones.
Además, cabe destacar que ambas distribuciones estarán claramente sesgadas hacia la derecha ya que la línea superior es bastante más larga que la inferior. En la segunda variable, prácticamente coinciden el primer cuartil y el valor mínimo.
Finalmente, comentar que los valores mínimos tanto en una variable como otra se dan en la Universitad Miguel Hernández y los máximos en la Universidad de Valencia.

d) Calcula el coeficiente de variación para los Profesores de la universidad Jaume I.

Nuevamente, introducimos los datos en el espacio de trabajo de Minitab y calculamos la media y la desviación estándar:

```
Column Mean

Mean of C7 = 764,40
```

```
Column Standard Deviation

Standard deviation of C7 = 65,987
```

Por tanto, como $CV = \frac{s}{\overline{X}} = 0.086$, lo cual indica que existe una dispersión de los datos de aproximadamente el 9%.

BIBLIOGRAFÍA

[1] D.A. Lind, R.D. Mason, W.G. Marchal (2001): "Estadística para Administración y Economía". Ed. Irwin McGraw-Hill.F.

[2] F. Moya Anegón, J. López Gijón, C. García Caro (1996): "Técnicas cuantitativas aplicadas a la biblioteconomía y documentación". Ed. Síntesis.

[3] R. Johnson (1996): "Elementary Statistics". Ed. Duxbury

ENLACES

- http://www.mste.uiuc.edu/hill/dstat/dstat.html: Introduction to Descriptive Statistics
- www.ine.es: Instituto Nacional de Estadística
- www.fisterra.com/material/investiga/10descriptiva/10descriptiva.htm: Estadística descriptiva