

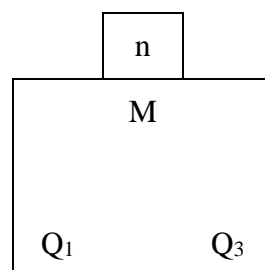
# Práctica 5 de Estadística

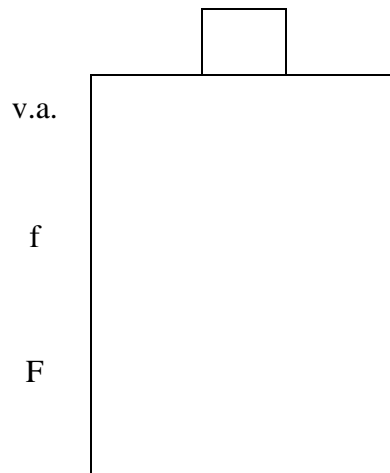
## Análisis de una variable medible. Gráficos caja.

El gráfico caja es una representación gráfica que proporciona información sobre una distribución muestral complementaria a la dada por los histogramas. Este gráfico presta especial interés sobre las colas de la distribución, indicando qué valores pueden considerarse atípicos. Pero también se representan en este gráfico la mediana y los cuartiles, con lo que el gráfico proporciona información tanto de localización como de dispersión. El método está basado en comparar la distribución de frecuencias muestral con aquella que debería corresponder a una distribución normal con recorrido intercuartílico igual al de la muestra. Está basado en el “Principio de Windsor”, el cual establece que, en la práctica, todas las distribuciones de frecuencia son normales en el centro.

### 5.1 Construcción del gráfico caja

La información necesaria para construir el gráfico caja de una muestra puede organizarse de forma conveniente utilizando una caja resumen, que contiene información sobre la mediana y los cuartiles, y añadiendo una segunda caja como se muestra a continuación:

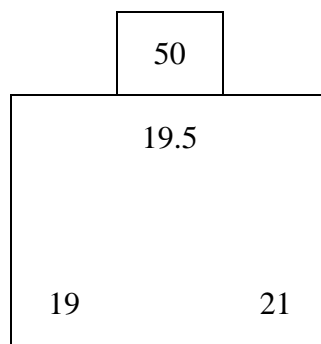




Así, si consideramos la variable edad de los alumnos de una clase de 50:

<b>Edad</b>	17	18	19	20	21	22	23	24	29
<b>Alumnos</b>	2	5	18	9	5	5	3	2	1

Necesitamos conocer la mediana  $M = 19.5$  y los cuartiles  $Q_1 = 19$  y  $Q_3 = 21$ .  
El intervalo intercuartílico es  $IQR = 2$ .



Ahora procederemos a llenar la caja inferior. La caja superior contendrá un factor de escala  $1.5 \times IQR$ . Así, para este ejemplo:

$$1.5 \times 2 = 3$$

Este valor es restado de  $Q_1 = 19$  para obtener la valla inferior,  $f = 19 - 3 = 16$ , y de nuevo restamos de este valor (f) para obtener la valla exterior inferior,  $F = 16 - 3 = 13$ . Las vallas interior y exterior superiores se obtienen añadiendo el factor de escala a  $Q_3 = 21$  obteniéndose  $f = 24$  y  $F = 27$ . Estos valores se colocan en el diagrama como sigue:

	3	
v.a.		
f	16	24
F	13	27

Una vez establecidas estas fronteras, el resto de la información procede de los datos de la variable (es preferible tenerlos ordenados de mayor a menor). Los valores adyacentes (v.a.) son los valores más extremos de la región comprendida entre los cuartiles respectivos y las vallas interiores (incluidas). Así, 17 es el valor mínimo entre 16 y 19, mientras que el valor máximo entre 21 y 24 es 24.

	3	
v.a.	17	24
f	16	24
F	13	27

Todas las observaciones pertenecientes a las regiones comprendidas entre la valla interior y exterior serán llamadas atípicas o medianamente atípicas. El hecho de que no haya observaciones entre 13 (incluido) y 16 se indica en el diagrama siguiente. Tampoco hay observaciones mayores que 24 y menores o iguales que 27.

	3	
v.a.	17	24
f	16	24
	ninguno	ninguno
F	13	27

Las observaciones fuera de las vallas exteriores F son llamadas extremas o extremadamente atípicas. No hay datos menores que 13 y tenemos un dato mayor que 27, en concreto 29.

	3	
v.a.	17	24
f	16	24
	ninguno	ninguno
F	13	27
	ninguno	uno

29

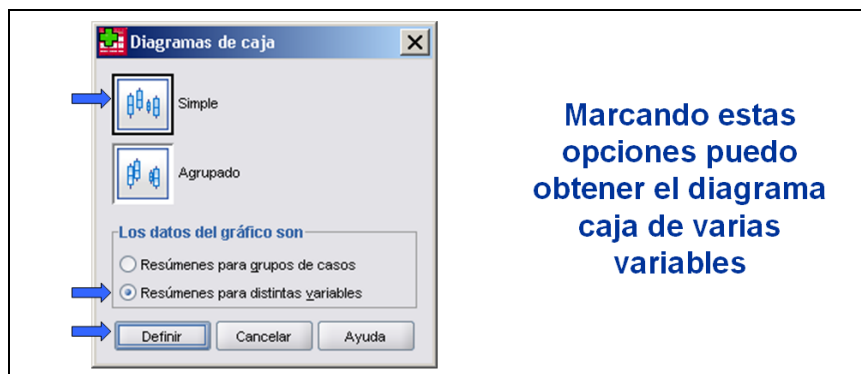
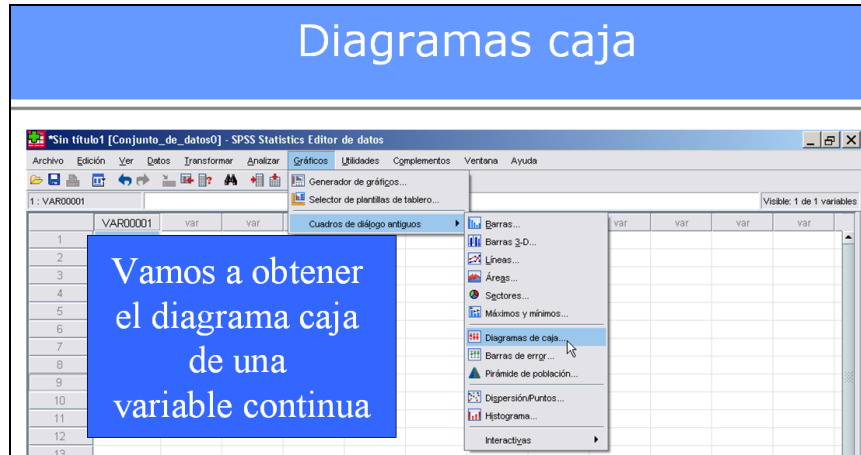
A partir de aquí se construye el gráfico caja. Se requiere para ello una escala que englobe el rango total de los datos. La caja termina en  $Q_1$  y  $Q_3$  con una recta interior trazada en la mediana. A continuación se unen los valores adyacentes a la caja mediante rectas. Finalmente se marcan los valores atípicos y extremos.

## 5.2 Interpretación del gráfico caja

El centro (posición central) de la distribución se representa por la recta mediana. La dispersión central de la distribución viene medida por la longitud de la caja, la cual contiene el 50% de las observaciones, 25% a cada lado de la mediana, la cual informa por tanto de la forma de la parte central. La longitud de los bigotes informan sobre la simetría de las colas; los valores atípicos alargan esas colas.

### 5.3 Manejo de SPSS

Para calcular el gráfico caja escogeremos la opción 'Gráficos>cuadro de diálogo antiguos>Diagramas de caja' y seleccionaremos la variable.



## Diagramas cajas



## 5.4 Ejercicios

1. En tres computadores de un determinado centro de cálculo, se ha medido durante ocho días el tiempo (en minutos) de realización de la copia de seguridad, obteniéndose los siguientes resultados:

COMPUTADOR UNO	COMPUTADOR DOS	COMPUTADOR TRES
16,8	13,3	15,8
15,8	17,8	17,2
17,3	17,1	17,8
10,8	19,1	17,1
16,5	14,1	17,5
15,5	15,7	19,7
17,1	15,8	20,2
17,7	20,8	13,3
18,5	16,8	18,2
19,6	15,5	23,8

- Obtener los gráficos caja para los resultados de los tres computadores.
- ¿Cuál de ellos tiene mayor dispersión? ¿Son simétricos? Indica los valores extremos y atípicos si los hubiera.
- ¿Qué computador es más fiable? ¿Cuál proporciona una mayor velocidad media de realización de copia?

2. Realizada una encuesta por muestreo entre usuarios de Internet, se ha obtenido los siguientes datos:

Minutos de conexión/día	Individuos de la muestra
34.5 – 39.5	10
39.5 – 44.5	15
44.5 – 49.5	25
49.5 – 54.5	18
54.5 – 59.5	22

- Construir el gráfico caja y comentar los resultados (dispersión, valores extremos, simetría...)
- Obtén los parámetros de centralización. ¿Cuál será más adecuado para esta muestra la media o la mediana? ¿Por qué?

3. La sección de control de calidad de una empresa ha medido la cantidad promedio en Kg. de llenado de recipientes de un determinado producto de una línea de producción:

12.00, 11.97, 12.10, 12.03, 12.01, 11.81, 11.91, 11.98, 12.30, 11.89, 12.05, 11.38, 12.89, 11.78, 12.72, 11.08, 11.78, 11.90, 11.44, 12.50, 13.75, 11.78, 11.91, 12.68, 11.39, 11.94, 13.45, 11.98, 11.27, 10.72, 11.93, 11.59, 12.79, 11.39, 13.05, 12.50, 11.58, 11.87, 11.78, 11.05, 12.02, 12.43

- a) Calcular la media, mediana desviación típica y tercer cuartil.
- b) Construir el gráfico caja de los datos anteriores.
- c) Explicar los resultados obtenidos.