

DIAGRAMAS DE CAJAS Y BIGOTES

Abel MARTÍN. Profesor de Matemáticas del IES de Llanera (Asturias)

El presente artículo toma como referencia el libro “Enseñar ESTADÍSTICA con CALCULADORAS GRÁFICAS” de “Ediciones TREA S. L.”, de próxima aparición y escrito por el autor del presente artículo.

NIVEL: Primero de bachillerato; Matemáticas Aplicadas a las Ciencias Sociales.

OBJETIVOS: Realizar un repaso rápido de la estadística unidimensional vista en la ESO, añadir algún procedimiento nuevo y moderno que nos podemos encontrar en la bibliografía actual e incorporar nuevas tecnologías como soporte visual.

VARIABLE ESTADÍSTICA CUANTITATIVA DISCRETA

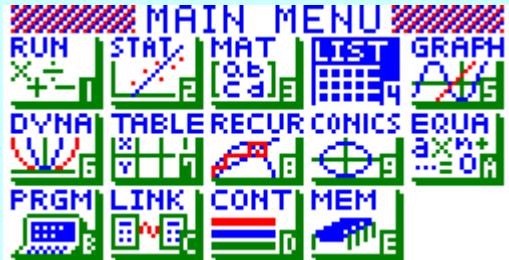
Agrupando los jugadores de la plantilla del Trea C.F. según la edad obtenemos la siguiente tabla estadística de frecuencias:

Edad en años	19	21	23	25	28	30	32	45
f_i	2	2	4	7	4	3	2	1

Para realizar el estudio de esta población utilizaremos como herramienta auxiliar la calculadora gráfica, en nuestro caso la CFX-9850 de CASIO, que nos permitirá eliminar los interminables y farragosos cálculos aritméticos, dando prioridad al razonamiento, permitiéndonos más tiempo para pensar y analizar lo que hacemos y los resultados que obtenemos.

AC


Presiona los cursores que consideres adecuados en el MENÚ INICIAL de presentación para seleccionar el modo LIST (o bien presiona directamente la tecla 4)



En el momento que presiones **EXE** entraremos en la alternativa de **LISTAS**.

El cursor aparecerá en la celda del primer elemento de la “List1”. Para ir introduciendo los diferentes valores de la variable bastará con ingresar los datos en una lista y las frecuencias en otra; Anotamos primero los valores x_i

1	9	EXE	2	1	EXE
2	3	EXE	2	5	EXE
2	8	EXE	3	0	EXE
3	2	EXE	4	5	EXE



A continuación, colocaremos en la “List2” la **frecuencia absoluta** (f_i) de cada valor de la variable (x_i) que es el número total de individuos que poseen ese valor.

<table border="1" style="margin: auto; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">2</td><td style="padding: 2px 10px;">EXE</td> <td style="padding: 2px 10px;">2</td><td style="padding: 2px 10px;">EXE</td> <td style="padding: 2px 10px;">4</td><td style="padding: 2px 10px;">EXE</td> <td style="padding: 2px 10px;">7</td><td style="padding: 2px 10px;">EXE</td> </tr> <tr> <td style="padding: 2px 10px;">4</td><td style="padding: 2px 10px;">EXE</td> <td style="padding: 2px 10px;">3</td><td style="padding: 2px 10px;">EXE</td> <td style="padding: 2px 10px;">2</td><td style="padding: 2px 10px;">EXE</td> <td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">EXE</td> </tr> <tr> <td colspan="8" style="text-align: center; padding: 10px 0;">▶</td> </tr> </table> <p>Por ejemplo, la fila 1 indica que el dato 19 aparece con una frecuencia absoluta de 2, es decir, hay 2 jugadores del R. Trea C.F. que tienen 19 años: $x_i = 19$; $f_i = 2$</p>	2	EXE	2	EXE	4	EXE	7	EXE	4	EXE	3	EXE	2	EXE	1	EXE	▶								<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th></th><th>List 1</th><th>List 2</th><th>List 3</th><th>List 4</th></tr> </thead> <tbody> <tr><td>1</td><td>19</td><td>2</td><td></td><td></td></tr> <tr><td>2</td><td>21</td><td>2</td><td></td><td></td></tr> <tr><td>3</td><td>23</td><td>4</td><td></td><td></td></tr> <tr><td>4</td><td>25</td><td>7</td><td></td><td></td></tr> <tr><td>5</td><td>28</td><td>4</td><td></td><td></td></tr> </tbody> </table> <p style="text-align: center;">SRTA SRTD DEL DELA INS</p>		List 1	List 2	List 3	List 4	1	19	2			2	21	2			3	23	4			4	25	7			5	28	4		
2	EXE	2	EXE	4	EXE	7	EXE																																																
4	EXE	3	EXE	2	EXE	1	EXE																																																
▶																																																							
	List 1	List 2	List 3	List 4																																																			
1	19	2																																																					
2	21	2																																																					
3	23	4																																																					
4	25	7																																																					
5	28	4																																																					

Si queremos obtener los valores de diferentes medidas de centralización y de dispersión tenemos que informar a la calculadora que los valores de los datos x_i los tenemos señalados en la **List1** y que las frecuencias se encuentran en la **List2**. Para ello bastará con realizar los siguientes pasos:

1.- Entrar en el MODO ESTADÍSTICO

<p>MENU ◀ ◀</p> <p>o bien presiona directamente la tecla (2)</p>	
--	--

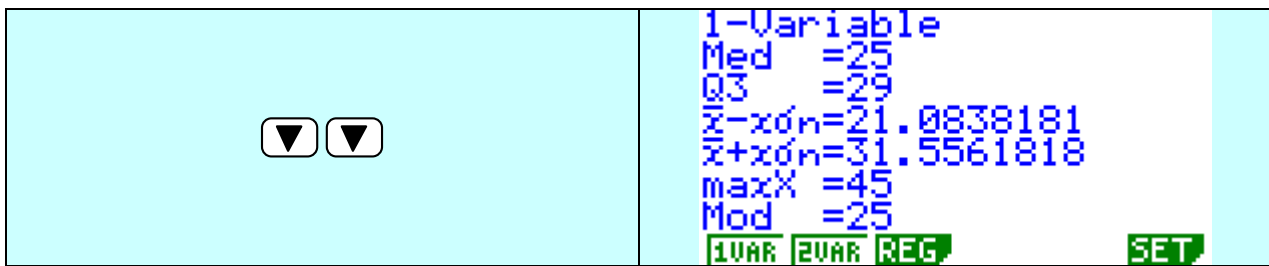
En el momento que presiones **EXE** entraremos en la función de ESTADÍSTICA.

2.- Adecuar las listas a las variables:

<p>CALC (F2) SET (F6)</p> <p>La variable x_i será la Lista 1 → List1 (F1) ▼</p> <p>La frecuencia será la Lista 2 → List2 (F3)</p>	<pre>1Var XList :List1 1Var Freq :List2 2Var XList :List1 2Var YList :List3 2Var Freq :1</pre> <p style="text-align: center;">1 List1 List2 List3 List4 ▸</p>
<p>EXE</p> <p>Al tratarse de una variable unidimensional</p> <p>1VAR (F1)</p>	<pre>1-Variable Σx =26.32 Σx² =658 Σx² =18004 xón =5.23618181 xón-1 =5.34415568 n =25</pre> <p style="text-align: center;">1VAR 2VAR REG SET</p>

Nos desplazamos hacia abajo con los cursores para seguir viendo otros cálculos realizados por la máquina que se encuentran en la siguiente pantalla:

<p>▼ ▼ ▼ ▼ ▼ ▼</p>	<pre>1-Variable minX =19 Q1 =23 Med =25 Q3 =29 Q3-Q1 =6 x-xón =21.0838181 x+xón =31.5561818</pre> <p style="text-align: center;">1VAR 2VAR REG SET</p>
--------------------	--



De esta forma se puede trabajar con el alumno el significado de cada uno de ellos, la concienciación de que no se trata de una simple relación de números, sino que están vivos y reflejan cosas, situaciones, (como se explica en el artículo de la Estadística con la calculadora científica de este mismo número de la revista), persiguiendo fundamentalmente el ANÁLISIS y la toma de DECISIONES:

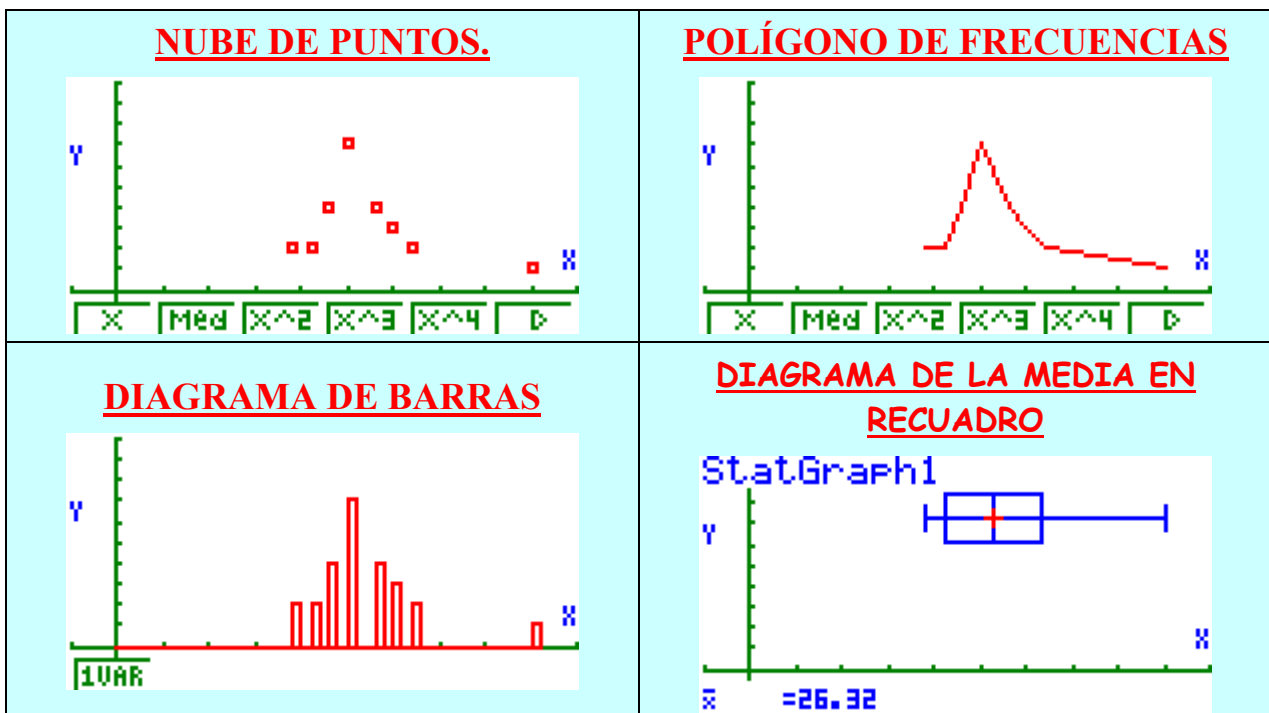
— ¿Cuál es la medida de centralización más adecuada para representar la distribución?—

—¿Cuál es la edad más frecuente?—

—¿Qué significa que $\bar{x} - S = 21.08$ y $\bar{x} + S = 31.55$?—

Preguntas de esta naturaleza presentan un nuevo espíritu, un nuevo enfoque de enseñar, diferente a aquellas de “Calcula la media, mediana y moda de la distribución”.

También pueden visualizarse los resultados a través de diversas representaciones gráficas. Presentamos a continuación la simple reseña y presentación de algunas de las representaciones gráficas que podemos generar fácilmente con una calculadora:



En lo que sí vamos a detenernos es en otro tipo de visualización gráfica, cada día más utilizada y que las calculadoras gráficas, en su práctica totalidad, incorporan entre sus utilidades y prestaciones. Nos referimos a los diagramas de cajas y bigotes, muy utilizados cuando nos encontramos casos en los que la presencia de “outliers” (*puntos fuera de línea*), puntos que se salen de un margen “razonable” y que modifican notablemente parámetros como la media y la desviación típica, haciendo más representativo este otro tipo de representación:

DIAGRAMA DE CAJAS Y BIGOTES

También denominados “Box and Whiskers” o “Boxplots” o “diagramas de la mediana en recuadro”, basados en la mediana, que resisten mejor las modificaciones ocasionadas por estos “outliers” que perturban la media.

Actualmente son mucho más utilizados que los diagramas de la media en recuadro y se podrían describir como representaciones gráficas de una distribución estadística unidimensional que reflejan directamente 5 parámetros (límite inferior, primer cuartil, mediana, tercer cuartil y límite superior) e indirectamente el rango y el rango intercuartílico; También dan una idea de la simetría, el sesgo y la dispersión de los datos de la distribución, permitiendo contrastar conjuntos de datos diferentes de una misma variable.

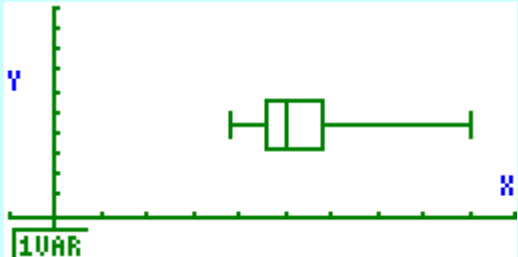
Diseñemos **GPH2** con unas características para guardar este tipo de gráfico:

<div style="text-align: center;"> EXIT SET GPH2 F6 F2 ▼ ☒ Box F6 F2 (Box: Diagrama de cajas con mediana en recuadro) List1 List2 Grn ▼ F1 ▼ F3 ▼ F3 </div>	<pre style="font-family: monospace; color: blue;">StatGraph2 Graph Type : MedBox XList : List1 Frequency : List2 Graph Color : Green</pre> <p style="text-align: center; color: green;"> Blue Orn3 Grn</p>
--	---

Sólo nos falta indicarle que queremos ser nosotros los que le introduzcamos los parámetros de escala (Comando **MANUAL**).

<div style="text-align: center;"> SHIFT SET UP Situados en la opción “Stat Wind” Man F2 </div>	<pre style="font-family: monospace; color: blue;">Stat Wind : Manual Graph Func : On Background : None Plot/Line : Blue Angle : Deg Coord : On Grid : Off</pre> <p style="text-align: center; color: green;"> Auto Man</p>
---	--




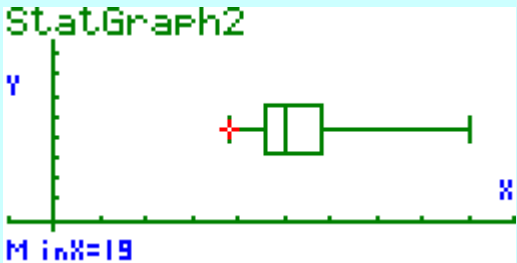

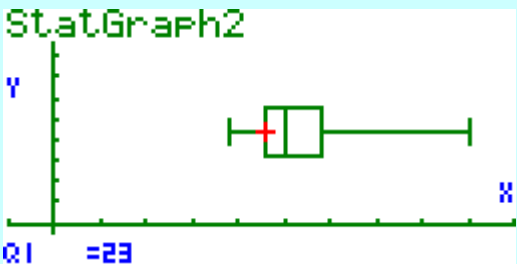

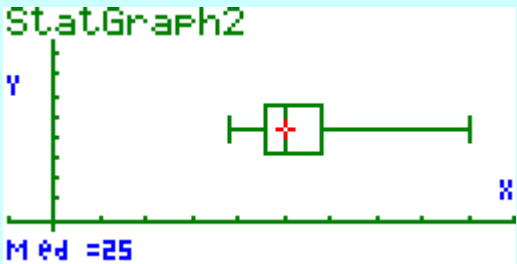

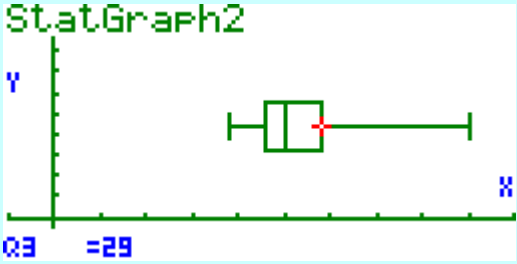

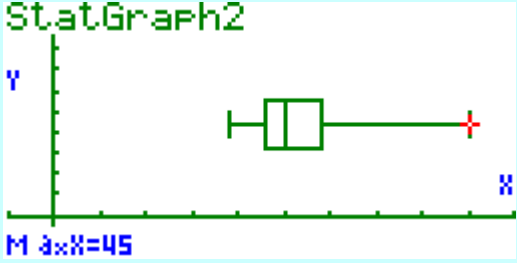
... y señalarle la escala adecuada para una correcta visualización:

<div style="text-align: center;"> EXE SHIFT V-Window F3 Podría ser la que se indica de forma adjunta </div>	<pre style="font-family: monospace; color: blue;">View Window Xmin : -5 max : 50 scale: 5 Ymin : -2 max : 10 scale: 1</pre> <p style="text-align: center; color: green;"> INIT TRIG STD STO RCL</p>
<div style="text-align: center;"> EXE GPH2 F2 </div>	

Un recuadro encierra todos los datos que se encuentran entre el primer cuartil (Q_1) y el tercer cuartil (Q_3), indicando mediante una línea vertical donde se encuentra la mediana; los

bigotes, filamentos o líneas van desde cualquier extremo hasta el mínimo o máximo dato de la distribución.

Si queremos realizar una exploración de todos estos parámetros en la gráfica procederemos de la siguiente forma, recorriendo el diagrama con el cursor y obteniendo de forma sucesiva los diferentes valores:

<p style="text-align: center;">    </p> <p style="text-align: center;">El valor mínimo de la distribución es 19</p>	
<p style="text-align: center;">  </p> <p style="text-align: center;">El primer cuartil es 23, es decir, deja un 25% de la distribución a su izquierda.</p>	
<p style="text-align: center;">  </p> <p style="text-align: center;">La mediana es 25, es decir, deja a cada lado el mismo número de datos.</p>	
<p style="text-align: center;">  </p> <p style="text-align: center;">El tercer cuartil es 29, es decir, deja el 75% de la distribución a su izquierda.</p>	
<p style="text-align: center;">  </p> <p style="text-align: center;">El valor máximo de la distribución es 45</p>	



COMENTARIOS:

1.- El bigote de la izquierda es mucho más corto que el de la derecha, es decir, que las edades de los individuos de la cuarta parte más corta están mucho más concentrada que la cuarta parte de los de mayor edad.

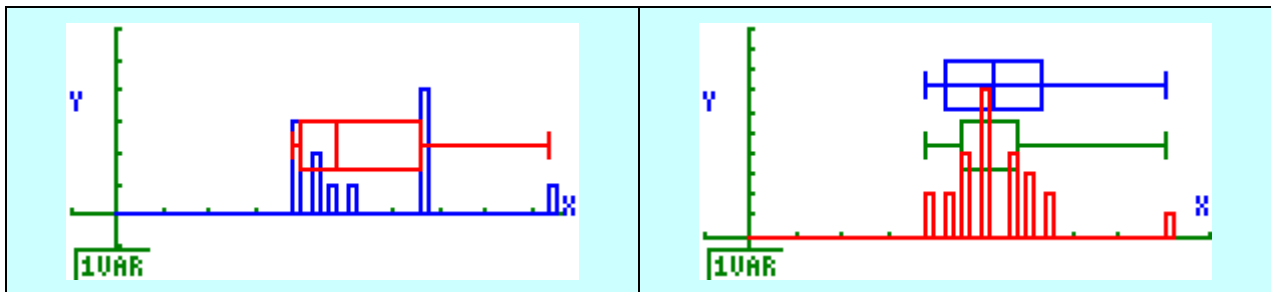
2.- La parte izquierda de la caja, comprendida por con las edades entre el 25% y 50% es menor que la derecha por lo que las edades de estos últimos están más dispersas.

3.- El rango es $Ls - Li = 45 - 19 = 26$

4.- El rango intercuartílico es $Q_3 - Q_1 = 29 - 23 = 6$. El 50% de la muestra se encuentra en un intervalo de 6 años, por lo que presenta una caja bastante estrecha.

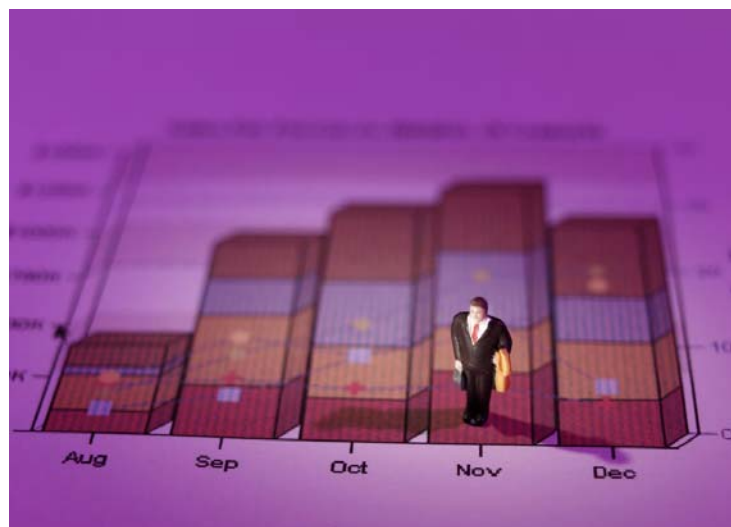
5.- La distribución es asimétrica y sesgada hacia la derecha.

Una vez visualizado el presente diagrama estaríamos en disposición de intentar relacionar diagramas de cajas; una vez que se hayan discutido las posibles soluciones podríamos ratificarlo con la imagen simultanea de ambas a través de la opción de gráficos múltiples, tal y como mostramos a continuación, pero que profundizaremos en sucesivas comunicaciones:



Este tipo de gráficos generarían actividades que permitiesen relacionar diferentes gráficos y ahondar en la comprensión de determinados conceptos. Entre ellas podríamos destacar alguna como la que aparece en la página siguiente.

Hemos propuesto una forma de afrontar un repaso rápido con la ayuda de la calculadora y hemos visto diferentes formas de visualizar la distribución; ahora faltaría una fase en la que la calculadora no puede hacer nada, que es la organización y sistematización de este tratamiento para la realización de un informe, la búsqueda de conclusiones y la toma de decisiones, objetivo fundamental de la estadística.



ACTIVIDAD 1

Dados los siguientes diagramas de cajas y bigotes y los siguientes diagramas de barras, asignar a cada diagrama de cajas su correspondiente diagrama de barras:

DIAGRAMA DE CAJAS Y BIGOTES

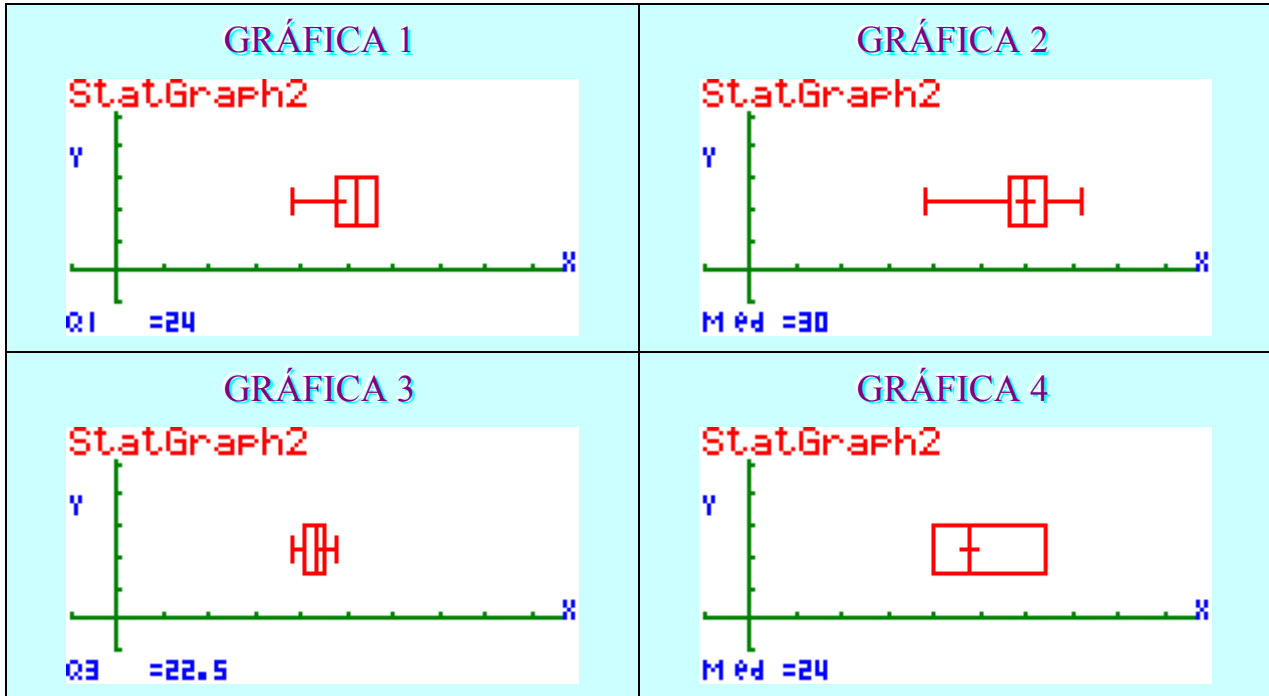
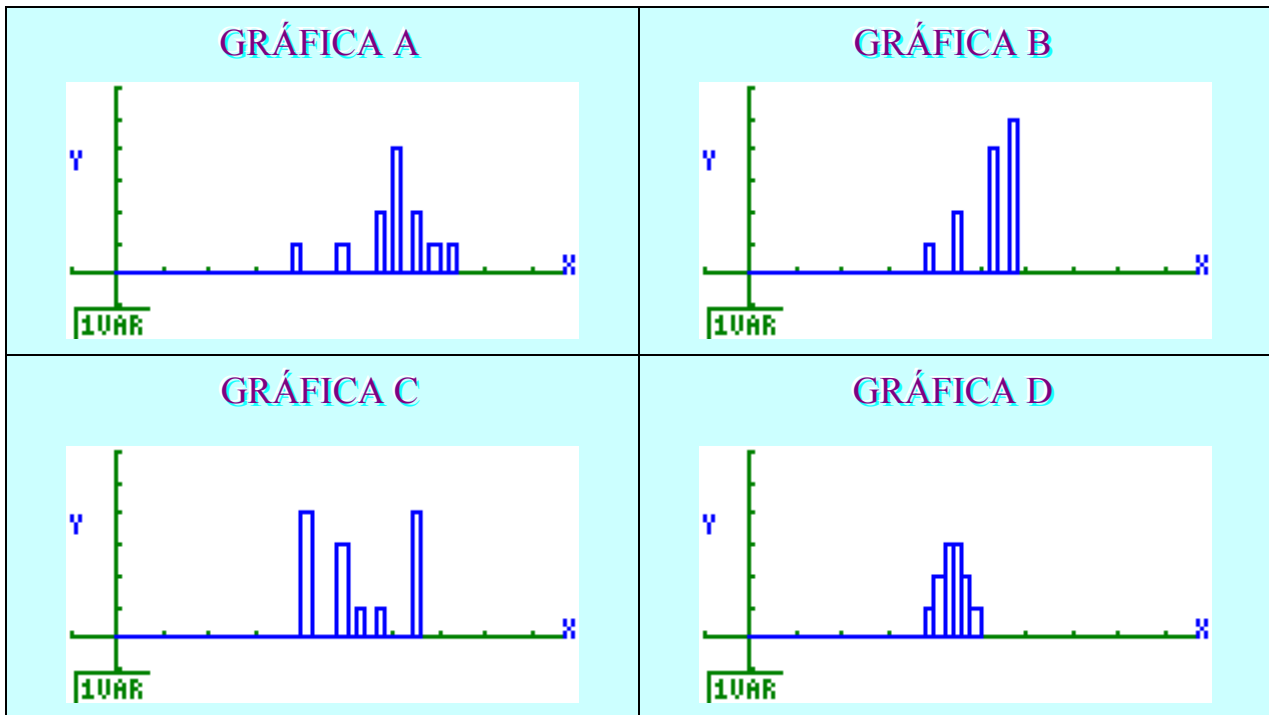


DIAGRAMA DE BARRAS



Todo esto y más lo puedes encontrar en www.aulamatematica.com