CAPÍTULO X

ESTADÍSTICA UNIDIMENSIONAL y BIDIMENSIONAL

La enseñanza de la Estadística ha de cambiar inevitablemente su metodología y sus objetivos; no podemos perder el tiempo enseñando a rellenar largas, arduas y farragosas tablas, olvidándonos de la esencia de lo que se persigue:

¡TRATAR DATOS, BUSCAR CONCLUSIONES Y TOMAR DECISIONES!

¡Otras tablas más altas cayeron!, tal y como ocurrió con las tablas de logaritmos o las trigonométricas. Hay que huir de aquellos problemas en los que, simplemente, se decía: "dada la siguiente distribución estadística, calcula la media, la mediana, la moda y la desviación típica"; problemas vacíos, sin jugo, sin juego, en los que no hacía falta pensar, sólo operar. Es mucho más importante y reconfortante dedicar el tiempo a meditar sobre las soluciones obtenidas, analizar su coherencia... eliminando los interminables y farragosos cálculos aritméticos y dando prioridad al razonamiento.

CONTENIDOS MÍNIMOS

El tipo de ejercicios y problemas que se presentan pretende evaluar el conocimiento global teórico y práctico que los alumnos tienen del tema en algunos de los siguientes aspectos:

- **SABER** calcular e **INTERPRETAR** las medidas de centralización y dispersión de una muestra.

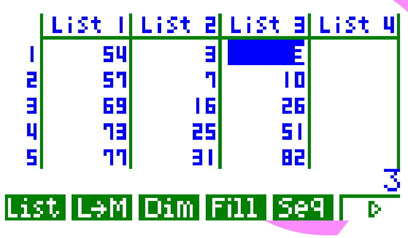
- **INTERPRETAR** el grado y tipo de relación existente entre dos variables y extraer las conclusiones apropiadas.

- **ASOCIAR**, identificar y relacionar, al margen de la investigación inicial, diferentes nubes de puntos con distintas situaciones.

- **CALCULAR** los parámetros estadísticos de una distribución bidimensional, coeficiente de correlación lineal y rectas de regresión.

- **ANALIZAR** el grado de relación entre las dos variables, conocido el coeficiente de correlación.

- **IDENTIFICAR**, ante varias distribuciones bidimensionales y un conjunto de parámetros estadísticos, los parámetros que corresponden a cada distribución.



LA ESTADÍSTICA UNIDIMENSIONAL

La Estadística es la parte de las Matemáticas que se encarga de los procedimientos que permiten el tratamiento sistemático de datos, la búsqueda de conclusiones acerca de los mismos y la toma de decisiones tras su análisis.

Es una de las disciplinas que mayor importancia y más auge está experimentando en nuestra sociedad, no sólo en cuanto a presentarnos y resumirnos gran cantidad de información de una forma simple y atractiva, sino ayudándonos a tomar decisiones, tratando de eliminar el componente aleatorio que muchas de dichas decisiones conlleva.

Inicialmente trataremos de exponer los conceptos y diferentes procedimientos que consideramos debe de conocer el alumnado de ESO para, a continuación, profundizar en otros correspondientes al curriculum de los bachilleratos de Ciencias de la Salud, Tecnológico y, sobre todo, el de Ciencias Sociales, conjugando el aspecto de usuario y ciudadano con el componente matemático que le pueda ser útil y necesario para el futuro.

Para ello presentamos unas actividades, especialmente diseñadas para comprender los conceptos fundamentales, utilizando la calculadora gráfica, autentica revolución didáctica, que nos permite olvidarnos parcialmente de las tablas estadísticas, pudiendo profundizar en su conocimiento y obviando los farragosos cálculos aritméticos.

ACTIVIDAD 1

Una compañía de seguros quiere realizar un estudio sobre la esperanza de vida de los españoles para ajustar sus cuotas de seguros. Para ello contrata los servicios de una empresa de investigación, dirigida por un matemático de gran prestigio, capaz de averiguarlo en pocos días o quizás horas.

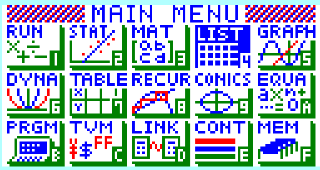
Para iniciar el estudio se parte de una muestra de 153 individuos; Agrupando la edad de los difuntos (x_i) obtenemos la siguiente TABLA ESTADÍSTICA de frecuencias:

años x_i	frecuencia f_i
54	3
57	7
69	16
73	25
77	31
79	38
83	21
85	12
	$\Sigma f_i = 153$

Para realizar el estudio de esta población utilizaremos como herramienta auxiliar la calculadora gráfica, en nuestro caso **la CFX–9850GB PLUS 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 presiona directamente la tecla [4]



Al presionar **EXE** entraremos en la alternativa de **LISTAS**.

Una de las opciones importantes que tenemos es que la máquina nos ofrece la posibilidad de almacenar todo el problema en un lugar determinado para luego, utilizar los datos cuando los necesitemos. Así, lo primero que haremos será entrar en:

SETUP
[SHIFT] [MENU]

Situados en la opción "**List File**"

File1
[F1]

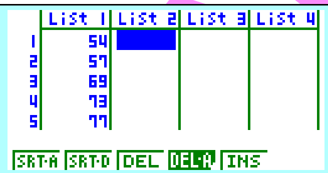
Así, todo lo que hagamos estará registrado como **File1**.

[EXE]

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.

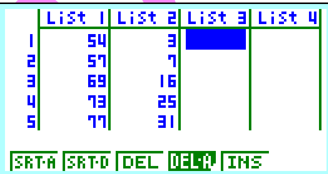
Anotamos primero los valores x_i

[5] [4] [EXE] [5] [7] [EXE]
[6] [9] [EXE] [7] [3] [EXE]
[7] [7] [EXE] [7] [9] [EXE]
[8] [3] [EXE] [8] [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.

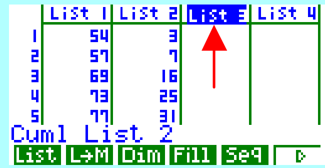
[3] [EXE] [7] [EXE] [1] [6] [EXE] [2] [5]
[EXE] [3] [1] [EXE] [3] [8] [EXE] [2] [1]
[EXE] [1] [2] [EXE] [▶]



Por ejemplo, la fila 1 indica que el dato 54 aparece con una frecuencia absoluta de 3, es decir, hay 3 personas que fallecieron a los 54 años ⇨ $x_i = 54$; $f_i = 3$

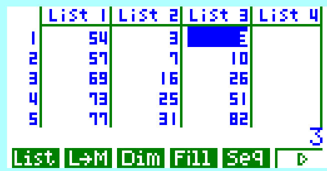
Una vez concluida la tarea de introducción de datos, va a ser la máquina la que realice el trabajo; nosotros nos limitaremos a reflexionar sobre lo que hay que hacer, lo que significa cada cosa y a dar órdenes para la obtención de las demás columnas. Bastará con situarse **ENCIMA DE LA LEYENDA "List 3"** y pensar en lo que vamos a hacer. En primer lugar vamos a generar la columna de las **frecuencias absolutas acumuladas** (**F_i**) de cada valor de la variable (x_i), que es el número total de individuos para los que la variable toma valores menores o iguales que x_i

[▲] [OPTN] LIST [F1] ▷ [F6] ▷ [F6] Cuml [F3] ▷ [F6]
List [F1] [2]



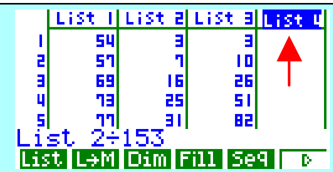
[EXE]

Nos coloca automáticamente las frecuencias absolutas acumuladas en la lista 3. Por ejemplo, la 3ª fila indica que el dato 69 aparece con una frecuencia absoluta acumulada de 26, es decir, hay 26 personas que fallecen a los 69 años o menos. $x_i = 69$; $F_i = 26$



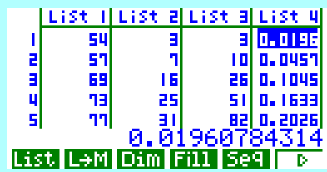
En la "**List4**" queremos crear la columna de las **frecuencias relativas** (**h_i**) de cada valor de la variable (x_i), que es el cociente que resulta de dividir su frecuencia absoluta por el número total de individuos. Nos indica el número de veces que aparece cada dato con respecto al número total de observaciones (N = 153). Recuerda que hay que situarse encima de la leyenda "List 4".

List
[▶] [▲] [F1] [2] [÷] [1] [5] [3]



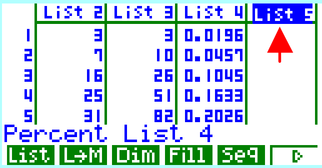
[EXE]

Nos genera automáticamente las frecuencias relativas. Por ejemplo, la primera fila indica que el dato 54 aparece con una frecuencia relativa de 0.0196, es decir, 0.0196 personas de cada 1 fallece a los 54 años. $x_i = 54$; $h_i = 0.0196$



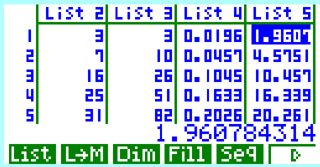
En la "**List5**" vamos a generar la columna de las **frecuencias relativas,** expresadas en **porcentaje,** de cada valor de la variable (x_i) ya que puede resultar más "real" que el que acabamos de exponer cuando decíamos 0.0196 de cada 1.

▷ ▷ % ▷ List
▶ ▲ F6 F6 F4 F6 F1 4



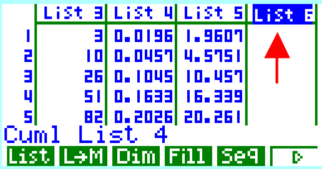
EXE

Nos coloca automáticamente las frecuencias relativas expresadas en porcentaje. Por ejemplo, en la fila 1, el dato 54 aparece con una frecuencia relativa del 1.96%; $x_i = 54$; $\%h_i = 1.96$



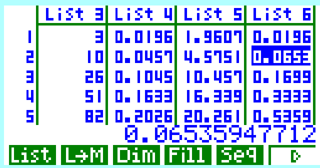
Y por último vamos a crear en la "**List6**" la columna de las **frecuencias relativas acumuladas**(H_i) de cada valor de la variable (x_i), que es el cociente que resulta de dividir la frecuencia absoluta acumulada por el número total de individuos. También se puede realizar de la siguiente forma:

▷ ▷ Cuml ▷ List
▶ ▲ F6 F6 F3 F6 F1 4



EXE ▼

Obtenemos automáticamente la columna de las frecuencias relativas acumuladas. Por ejemplo, en la segunda fila se indica que 57 aparece con una frecuencia relativa acumulada de 0.0653, es decir, 6.53 personas de cada 100 mueren a los 57 años o menos. $x_i = 57 \rightarrow H_i = 0.0653$



Como no cabe toda la tabla en pantalla, para tener una visión de todos los datos numéricos que necesitemos, sólo habría que moverse por ella con los cursores.

Recuerda que todo el problema lo tienes guardado en File1.

No obstante, y para asegurarse, toma una **muestra** de otra ciudad muy cercana, con más habitantes, pero con unas costumbres, un hábitat y una forma de vivir muy parecida; El inconveniente es que toma excesivos valores diferentes de las edades de defunción.

Debido a este hecho, en el que la variable estadística toma excesivos valores, hay que buscar otro método distinto al anterior para realizar el estudio, proponiéndose la siguiente tabla estadística de frecuencias, agrupando los datos en intervalos de edad de 5 años.

Intervalo	x_i	f_i
(55 – 60]	57.5	7
(60 – 65]	62.5	12
(65 – 70]	67.5	29
(70 – 75]	72.5	32
(75 – 80]	77.5	36
(80 – 85]	82.5	59
(85 – 90]	87.5	20
		$\Sigma f_i = 195$

Almacenemos todo el problema en un lugar determinado para luego utilizar los datos cuando los necesitemos:

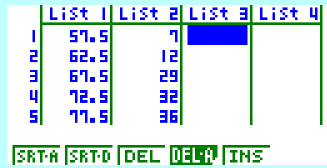
SETUP File2
SHIFT MENU F2

De esta forma todo lo que hagamos estará registrado como **File2**.

EXE

El cursor aparecerá en la celda del primer elemento de la "**List1**". A continuación ingresaremos todos los datos de las marcas de clase en "List1" y todas las frecuencias absolutas correspondientes en "List2", según hemos explicado anteriormente:

▶



1.- EJERCICIOS

1.1.- Coloca en "List3" la columna de las frecuencias absolutas acumuladas.

1.2.- Coloca en "List4" la columna de las frecuencias relativas.

1.3.- En "List5" la columna de las frecuencias relativas expresadas en porcentaje.

1.4.- En "List6" la columna de las frecuencias relativas acumuladas expresadas en porcentaje.

A continuación, y moviéndote con los cursores por la tabla, contesta directamente a las siguientes cuestiones:

1.5.- ¿Cuál es el carácter de la variable estadística?. ¿Es cualitativa o cuantitativa?. ¿Es discreta o continua?. Razona las respuestas.

1.6.- ¿Cuántas personas tardaron menos de 75 años en morirse?.

1.7.- ¿Qué significa que la frecuencia relativa del intervalo (70 – 75] es 32/195?.

1.8.- ¿Y que la frecuencia relativa acumulada de (65 – 70] es 48/195?.

1.9.- ¿Dónde se expresa, en la tabla, el número total de observaciones que realizamos?.

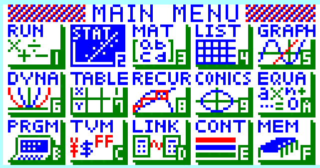
1.10.- ¿Qué quiere decir si la frecuencia relativa acumulada de una clase me da 3?.

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

MENU ◄ ◄

o presionar directamente la tecla 2



Al presionar **EXE** entramos en la función de ESTADÍSTICA.

2.- Adecuar las listas a las variables:

CALC SET
F2 F6

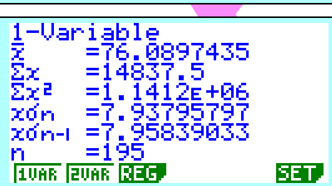
variable x_i ➩ Lista 1 ➡ List1 F1 ▼

Frecuencia ➩ Lista 2 ➡ List2 F3

EXE

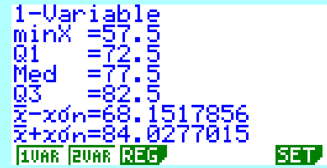
Al tratarse de una variable unidimensional

1VAR
F1

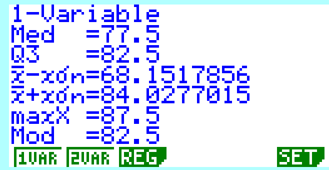


Nos desplazamos hacia abajo con los cursores para seguir viendo el valor de otros parámetros calculados por la máquina, que se encuentran a continuación, y que no caben en pantalla:

▼ ▼ ▼ ▼ ▼ ▼



▼ ▼



BREVE COMENTARIO DE LOS VALORES OBTENIDOS

✷ La media $\bar{x}$ de las edades de defunción de la población estudiada se aproxima a los 76.09 años.

✷ Σx ➡ Si sumamos los años de vida de toda la muestra alcanzaría los 14 837.5 años.

✷ Σx^2 ➡ La suma de los cuadrados de sus edades sería aproximadamente de 1 141 200 años cuadrados.

✷ $x\sigma n$ ➡ La desviación típica de las edades de la muestra, tomada como población, es 5.23618181.

✷ $x\sigma n\text{-}1$ ➡ La desviación típica de las edades de óbito de la población, tomada como muestra, se acerca a 7.94 años.

✷ n ➡ El número de individuos estudiado es 195.

✷ minX ➡ La edad a la que ha muerto la persona más joven se estima que es de es 57.5 años.

✷ Q_1 ➡ Una persona que ha perecido a los 72.5 años tiene el 25% de la distribución con una edad de defunción menor o igual que él.

✷ Med ➡ Una persona que ha fallecido a los 77.5 años deja a cada lado el mismo número de datos.

✷ Q_3 ➡ Una persona que ha perecido a los 82.5 años tiene el 75% de la distribución con una edad de defunción menor o igual que él.

✷ $\bar{x} - x\sigma n$ ➡ La mayor parte de la distribución se condensa en un intervalo que tiene como límite inferior 68.15.

✷ $\bar{x} + x\sigma n$ ➩ La mayor parte de la distribución se concentra en un intervalo que tiene como límite superior 84.03

✷ maxX ➩ La edad a la que ha muerto la persona más anciana se estima que es de es 87.5 años..

✷ ModX ➩ La edad más frecuente de defunción se aproxima a los 82.5 años.

Esos resultados también pueden visualizarse a través de diversas representaciones gráficas.

REPRESENTACIONES con la calculadora gráfica.

Existen unas representaciones que son más adecuadas para variables estadísticas discretas y que son las que vamos a ver a continuación, por lo que lo primero que vamos a hacer es recuperar el problema que teníamos guardado en "File1".

MENU 4 SHIFT MENU (SETUP) F1 (File1) EXE

Así recuperamos el problema completo de **File1**

MENU 2

	List 1	List 2	List 3	List 4
1	51	3	3	0.0196
2	57	7	10	0.0457
3	69	16	26	0.1045
4	73	25	51	0.1633
5	77	31	82	0.2026

54

GRPH CALC TEST INTR DIST ▷

NUBE DE PUNTOS.

Son representaciones de variables estadísticas discretas, en planos cartesianos, a través de puntos obtenidos al colocar en el eje OX los valores de la variable independiente y en el eje OY su correspondiente frecuencia absoluta.

F1 (GRPH)

	List 1	List 2	List 3	List 4
1	51	3	3	0.0196
2	57	7	10	0.0457
3	69	16	26	0.1045
4	73	25	51	0.1633
5	77	31	82	0.2026

54

GPH1 GPH2 GPH3 SEL SET

Los gráficos los podremos diseñar a nuestro gusto; Se pueden personalizar 3 tipos (GPH1, GPH2 y GPH3) a través del comando **SET,** para utilizarlos cuando queramos, simplemente acudiendo a ellos. Por ejemplo vamos a programar GPH1 como nube de puntos.

F6 (SET) F1 (GPH1) ▼ F1 (Scat)

Tipo de gráfico: **Nube de puntos (Scatter)**

StatGraph1
Graph Type :Scatter
XList :List1
YList :List2
Frequency :List2
Mark Type :▫
Graph Color :Blue
Scat xy NPP ▷

▼ F1 (List1) ▼ F2 (List2)

El eje **X** estará determinado por la List1 y el eje **Y** por la List2.

StatGraph1
Graph Type :Scatter
XList :List1
YList :List2
Frequency :List2
Mark Type :▫
Graph Color :Blue
List1 List2 List3 List4 List5 List6

▼ F1 (1)

La frecuencia con la que está sucediendo cada situación es 1. Esta opción es más propia de tablas de doble entrada, en bidimensionales.

StatGraph1
Graph Type :Scatter
XList :List1
YList :List2
Frequency :1
Mark Type :▫
Graph Color :Blue
1 List1 List2 List3 List4 ▷

▼ F1 (▫)

El tipo gráfico que representa a cada punto lo podemos elegir entre los que vienen señalados en el menú inferior de la pantalla. Por ejemplo hemos tomado el primer cuadrado:

StatGraph1
Graph Type :Scatter
XList :List1
YList :List2
Frequency :1
Mark Type :▫
Graph Color :Blue
▫ × ▪

▼ F2 (Orng)

Los puntos de la nube pueden tomar los colores Azul (Blue), naranja (Orange) y verde (Green). Por ejemplo podemos tomar:

StatGraph1
Graph Type :Scatter
XList :List1
YList :List2
Frequency :1
Mark Type :▫
Graph Color :Orange
Blue Orng Grn

Así pues, cada vez que presionemos **GPH1**, la representación que nos saldrá será una nube de puntos, con la lista 1 en el eje X, la lista 2 en el eje Y, en naranja y con la marca de puntos en cuadrado blanco.

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

EXE SHIFT MENU (SETUP) F2 (Man)

Stat Wind :Manual
Graph Func :On
Background :None
Plot/Line :Blue
Angle :Deg
Coord :On
Grid :Off ↓
Auto Man

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

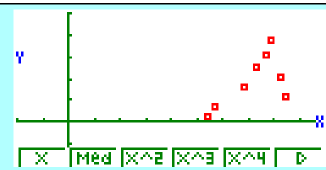
EXE SHIFT F3 (V-Window)

Podría ser la que se indica de forma adjunta:

View Window
Xmin :-20
max :100
scale:10
Ymin :-20
max :50
scale:10
INIT TRIG STD STO RCL

Recuerda que este tipo de gráfica lo tenemos diseñado en GPH1:

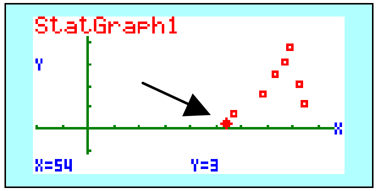
EXE F1 (GRPH) F1 (GPH1)



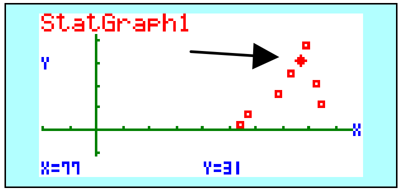
Si queremos observar las diferentes frecuencias de los distintos valores de la variable, directamente desde el dibujo, bastaría con utilizar el comando **TRACE ...**

SHIFT F1 (Trace)

... y moverse con los cursores, pudiendo obtenerse imágenes como estas:



▶ ▶ ▶ ▶



POLÍGONO DE FRECUENCIAS ABSOLUTAS

Son representaciones de variables estadísticas discretas, en planos cartesianos, uniendo con segmentos los puntos obtenidos al colocar en el eje OX los valores de la variable estadística y en el eje OY sus correspondientes frecuencias absolutas.

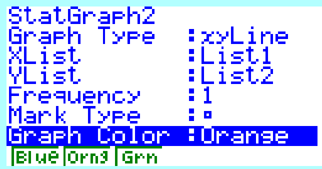
Vamos a programar **GPH2** con este tipo de gráfica para lo que tendremos que señalar las características siguientes:

EXIT | SET F6 | GPH2 F2 | ▼ | XY F2

(**xyLine**: Polígono de frec. absolutas)

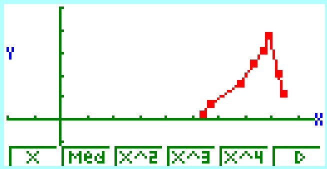
▼ | List1 F1 | ▼ | List2 F2 | ▼ | 1 F1 | ▼ | □ F1

▼ | Orng F2



Mantenemos la escala de los ejes y el ajuste MANUAL; Recuerda que este tipo de gráfica lo ESTAMOS DISEÑANDO en GPH2:

EXE | GPH2 F2



Con **TRACE** observamos las diferentes frecuencias de los distintos valores de la variable, directamente, desde el dibujo.

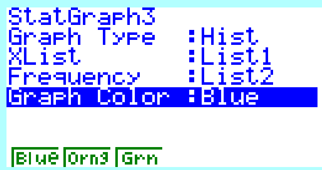
DIAGRAMA DE BARRAS

Son representaciones gráficas formadas por barras con anchura de trazo uniforme, cuya longitud viene determinada por la frecuencia absoluta, donde el eje OX determina los valores de la variable estadística.

EXIT | SET F6 | GPH3 F3 | ▼ | ▷ F6 | Hist F1

(**Hist**: Diagrama de barras, aunque también podríamos hacer un histograma, según indicaremos más adelante)

▼ | List1 F1 | ▼ | List2 F3 | ▼ | Blue F1

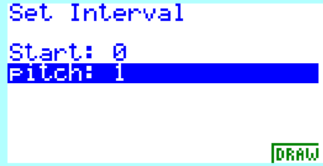


Mantenemos la escala de los ejes y el ajuste **MANUAL**; Recuerda que este tipo de gráfica lo estamos diseñando en GPH3:

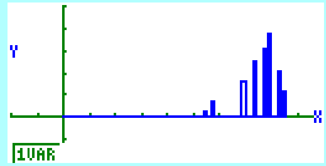
EXE | GPH3 F3

Nos pregunta cuál queremos que sea el comienzo (**Start**) y cual la amplitud de las barras (**pitch**); le introducimos los valores que aparecen al margen:

0 | EXE | 1 | EXE



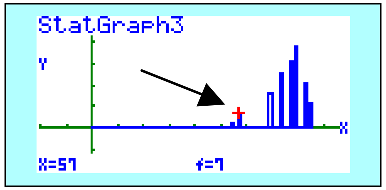
DRAW F6



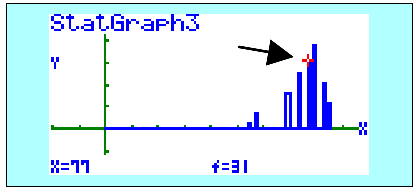
Si queremos observar las diferentes frecuencias en los distintos valores de la variable, directamente desde el dibujo, bastaría con utilizar el comando **TRACE ...**

SHIFT | Trace F1

... y moverse con los cursores, pudiendo obtenerse imágenes como estas:



▶ ▶ ▶ ▶



Recuerda que en estos momentos has personalizado GPH1, GPH2 y GPH3 con unas características determinadas, de forma que cualesquiera listas de datos que introduzcas a partir de ahora, a no ser que los modifiques, mantendrán estas configuraciones.

Vamos a estudiar la forma de realizar alguna representación del problema guardado en "File2" a través de la calculadora gráfica.

HISTOGRAMAS

Se utilizan para la representación de las frecuencias absolutas de v. e. continuas, donde las bases de los rectángulos son cada uno de los intervalos y donde la altura es tal que el área del rectángulo formado se corresponde con la frecuencia absoluta de dicho intervalo.

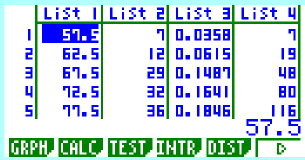
Por todo ello lo que primero que haremos será recuperar el problema que tenemos guardado en "File2"

SETUP File2

MENU 4 SHIFT MENU F2

EXE Así recuperamos el problema completo de **File2**

MENU 2



Diseñamos el Histograma en “GPH3”

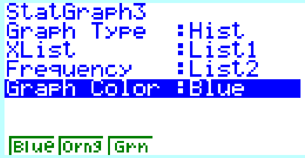
GRPH SET GPH3 ▷ Hist

F1 F6 F3 ▼ F6 F1

(**Hist**: histograma)

List1 List2 Blue

▼ F1 ▼ F3 ▼ F1



Mantenemos la escala de los ejes y el ajuste **MANUAL** y señalamos la escala adecuada para una correcta visualización:

V-Window

EXE SHIFT F3

Podría ser la que se indica de forma adjunta

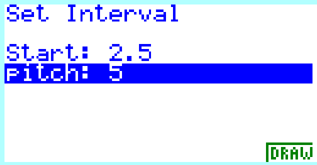
Recuerda que este nuevo diseño está en GPH3:

GRPH GPH3

EXE F1 F3

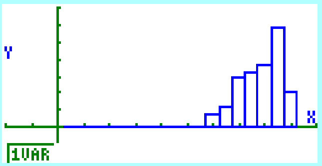
Nos pregunta cuál queremos que sea el comienzo (**Start**) y cuál la amplitud de los intervalos (**pitch**); le introducimos los valores que aparecen al margen para que, de esta forma, aparezcan los valores de las marcas de clase:

2 • 5 EXE 5 EXE



DRAW

F6

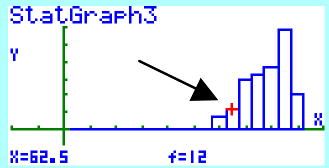


Si queremos observar las diferentes frecuencias en los distintos valores de la variable, directamente desde el dibujo, bastaría con utilizar el comando **TRACE ...**

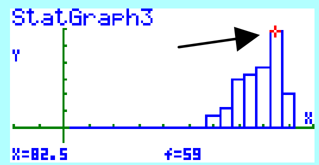
Trace

SHIFT F1

... y moverse con los cursores, pudiendo obtenerse imágenes como estas:



▶ ▶ ▶ ▶



Recuerda que en estos momentos has personalizado GPH1, GPH2 y GPH3 con unas características determinadas, de forma que cualesquiera listas de datos que introduzcas a partir de ahora, a no ser que los modifiques, mantendrán estas configuraciones.

Vamos a ver, a continuación, otro tipo de visualizaciones gráficas, cada día más utilizadas y que las calculadoras gráficas, en su práctica totalidad, incorporan entre sus utilidades y prestaciones.

Para ello, y para practicar un poco, comenzaremos introduciendo los datos de un nuevo problema, grabándolo previamente en **File3:**

ACTIVIDAD II

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

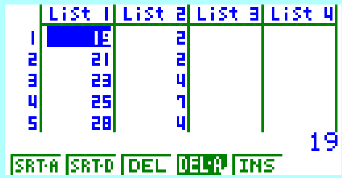
Edad en años	19	21	23	25	28	30	32	45
f_i	2	2	4	7	4	3	2	1

Si queremos realizar un estudio de la población, lo primero que vamos a hacer es grabar todo lo que hagamos en **File3:**

SETUP List3

MENU 4 SHIFT MENU F3 EXE

Introducimos todos los datos ➩



Entramos en el modo ESTADÍSTICO y adecuamos las listas a las variables:

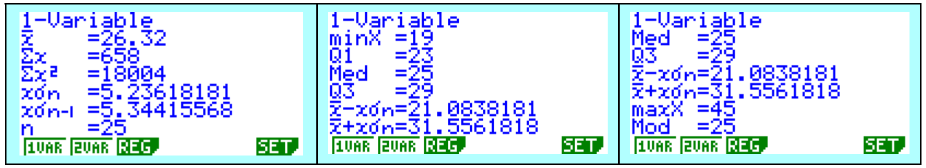
CALC SET List1 List2

MENU 2 F2 F6 F1 ▼ F3 EXE

1VAR

F1

Obteniéndose los siguientes resultados:



De esta forma se puede trabajar con el alumno el significado de cada uno de ellos, concienciándolo de que no se trata de una simple relación de números, sino que dichos números están vivos y reflejan cosas y situaciones, 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 $\bar{x}$ – 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".

Veamos pues, otro tipo de visualizaciones gráficas. En primer lugar, vamos a referirnos a los diagramas de la media en recuadro.

Antes de empezar la actividad, introduzcamos una escala adecuada, como puede ser:

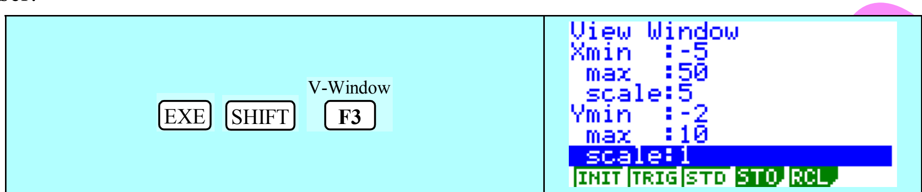
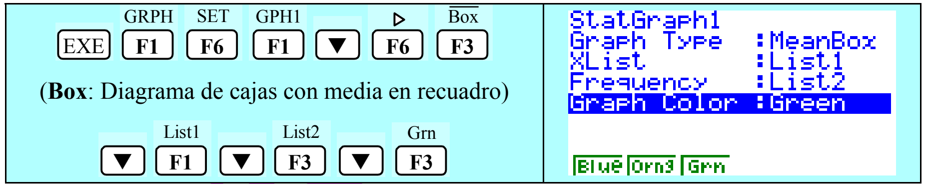


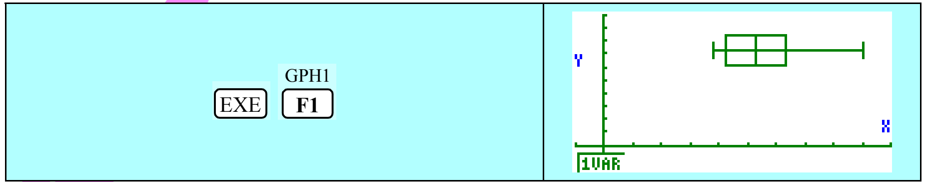
DIAGRAMA DE LA MEDIA EN RECUADRO

Son representaciones gráficas de una distribución estadística unidimensional, que reflejan directamente 5 parámetros (límite inferior, valor de la expresión $\bar{x}$ – S, media aritmética, valor de la expresión $\bar{x}$ + S y límite superior) e indirectamente el rango; También dan una idea de la simetría, del sesgo y la dispersión de los datos de la distribución, permitiendo contrastar conjuntos de datos diferentes de una misma variable.

Diseñemos **GPH1** con unas nuevas características:

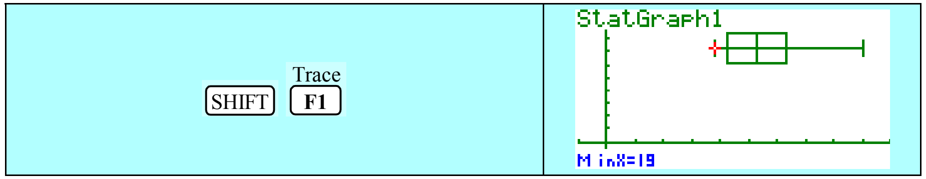


Mantenemos la escala de los ejes y el ajuste **MANUAL**; Recuerda que este tipo de gráfica lo acabamos de diseñar y guardar como **GPH1**, sustituyendo a la que estuviese anteriormente.

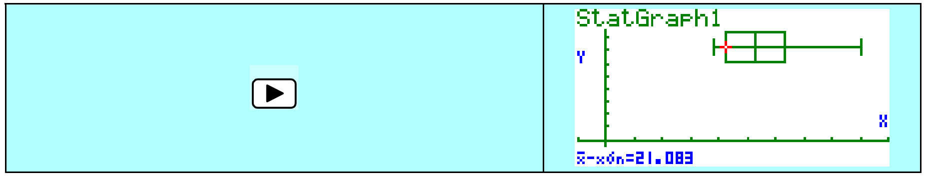


Un recuadro encierra todos los datos que se encuentran entre $\bar{x}$ – S y $\bar{x}$ + S, indicando mediante una línea vertical donde se encuentra la media; los bigotes, filamentos o líneas, van desde cualquier extremo hasta el mínimo o máximo dato de la distribución.

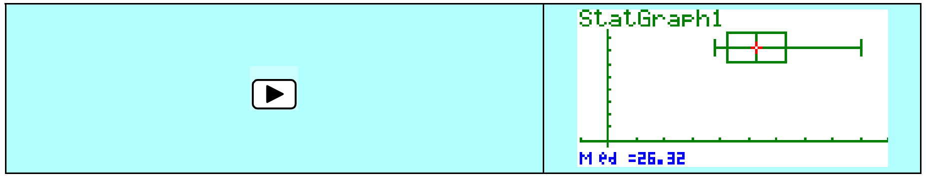
Si queremos realizar una exploración de todos estos parámetros en la gráfica procederemos de la siguiente forma, recorriendo el diagrama con el cursor ▶ ▶, obteniendo de forma sucesiva los diferentes valores:



El valor mínimo de la distribución es 19

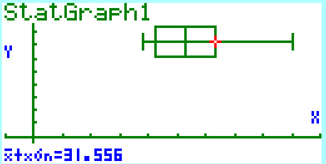


$\bar{x}$ – xσn ➡ La mayor parte de la distribución se condensa en un intervalo que tiene como límite inferior 21.083



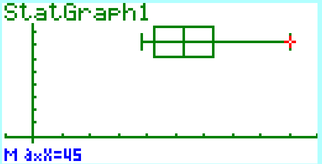
La media es 26.32, es decir, es el valor teórico en torno al cual se concentra la distribución

▶



$\bar{x}$ + xσn ➡ La mayor parte de la distribución se concentra en un intervalo que tiene como límite superior 31.556

▶



El valor máximo de la distribución es 45 años

COMENTARIO:

1.- La media $\bar{x}$ de las edades de la plantilla del Makuhari C.F. es de 26.32 años, concentrándose la mayor parte de los individuos de la distribución entre 21.083 y 31.556 años.

2.- El bigote de la izquierda es muy corto, es decir, la distribución de individuos que se encuentran entre el valor mínimo de la distribución y el límite inferior del intervalo [21.084, 31.556] está mucho más concentrada que los que se encuentran entre el valor máximo y el límite superior del mencionado intervalo, existiendo valores aislados.

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

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

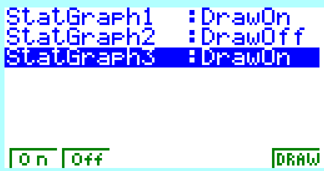
Una vez visualizado el presente diagrama estaríamos en disposición de intentar relacionar diagramas de cajas con diagramas de barras; una vez que se hayan discutido las posibles soluciones, podríamos ratificarlo con la imagen simultanea de ambas a través de la opción de GRÁFICOS MÚLTIPLES:

EXIT SEL F4

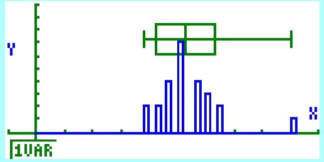
Queremos visualizar simultáneamente GPH1 y GPH3

(Diagrama de cajas y diagrama de barras)

On F1 ▼ Off F2 ▼ On F1



DRAW F6 DRAW F6

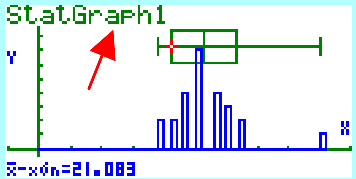


Si queremos rastrear con el TRACE por ambas gráficas bastará con activar esta opción...

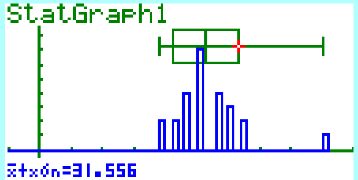
SHIFT Trace F1

... y mover los cursores ▶ ◀ para ir hacia adelante y hacia atrás, respectivamente, dentro de la gráfica en la que nos encontremos:

▶

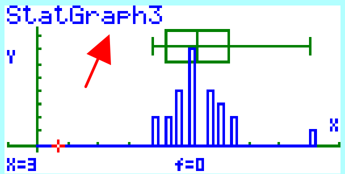


▶ ▶

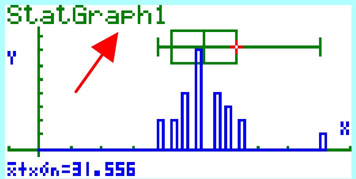


y los cursores ▼ ▲ para pasar de una gráfica a otra, señalada en el rótulo situado en la parte superior izquierda de la pantalla.

▼



▼



No obstante, cuando nos encontramos con estos casos en los que la presencia de ***"outliers" (puntos fuera de línea),*** que se salen de un margen "razonable" y modifican notablemente parámetros como la media y la desviación típica, se hace más gráfico otro tipo de representación que veremos a continuación:

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 a la media.

Actualmente son mucho más utilizados que los anteriores 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, del sesgo y la dispersión de los datos de la distribución, permitiendo contrastar conjuntos de datos diferentes de una misma variable.

Algunas máquinas, cuando hay elementos que se salen de unos límites razonables (outliers) lo representan mediante un asterisco.

Modifiquemos **GPH2** y démosle una nuevas características:

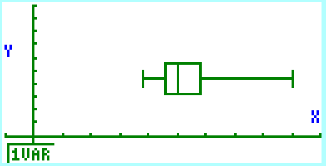
EXIT | GRPH F1 | SET F6 | GPH2 F2 | ▼ | ▷ F6 | Box F2

(**MedBox**: Diagrama de cajas con mediana en recuadro)

▼ | List1 F1 | ▼ | List2 F3 | ▼ | Grn F3

Mantenemos la escala de los ejes y el ajuste **MANUAL**; Ahora tenemos como **GPH2** la gráfica de CAJAS Y BIGOTES.

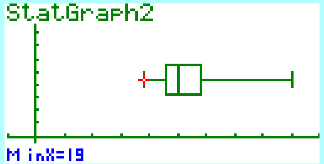
EXE | GPH2 F2



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.

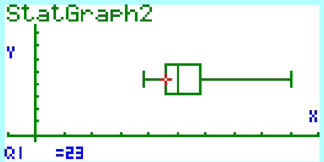
Si queremos realizar una exploración de todos estos parámetros en la gráfica, procederemos de la siguiente forma, recorriendo el diagrama con el cursor y obteniendo de forma sucesiva los diferentes valores:

SHIFT | Trace F1



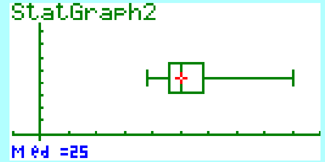
El valor mínimo de la distribución es 19.

▶



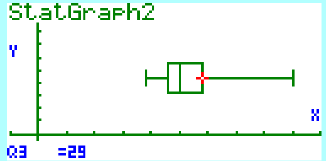
El primer cuartil es 23, es decir, deja un 25% de la distribución a su izquierda.

▶



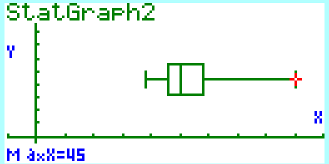
La mediana es 25, es decir, deja a cada lado el mismo número de datos.

▶



El tercer cuartil es 29, es decir, deja el 75% de la distribución a su izquierda.

▶



El valor máximo de la distribución es 45.

COMENTARIO:

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

2.- La parte izquierda de la caja, comprendida por las edades entre el 25% y 50% es menor que los de 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.

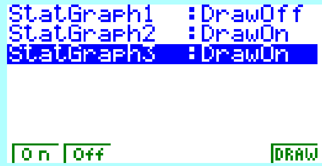
Una vez visualizado el presente diagrama estaríamos en disposición de intentar relacionar diagramas de cajas; Luego, podríamos ratificarlo con la imagen simultanea de ambas a través de la opción de gráficos múltiples:

EXIT | SEL F4

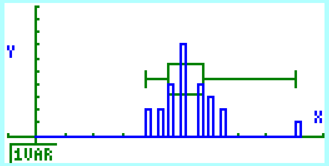
Vemos simultáneamente GPH2 y GPH3

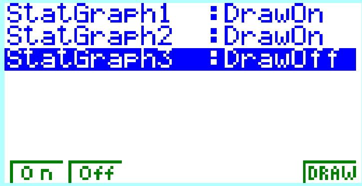
(Diagrama de cajas y diagrama de barras)

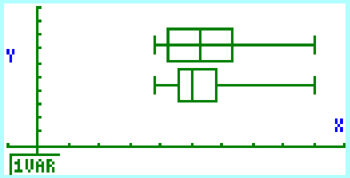
Off F2 | ▼ | On F1 | ▼ | On F1



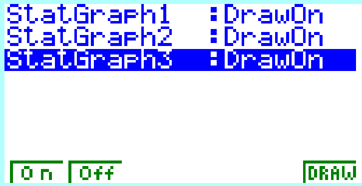
DRAW DRAW
F6 F6

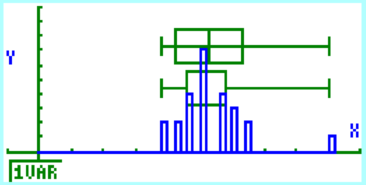






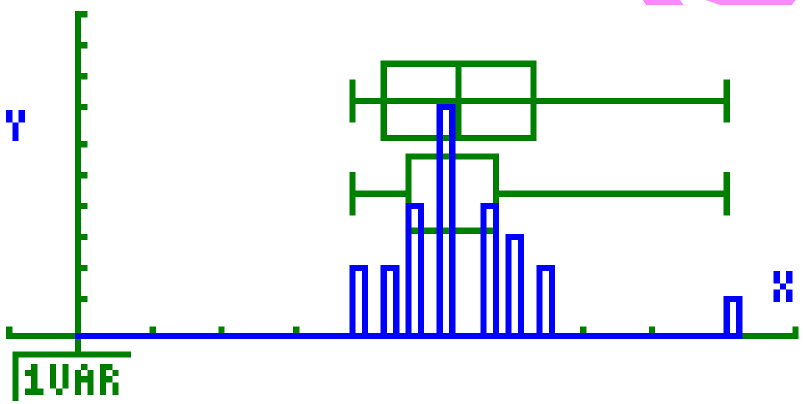
Como podemos ver, con la representación basada en la mediana, se intentan eliminar las posibles perturbaciones de las medidas de centralización y dispersión ocasionadas por los "outliers"





Si queremos rastrear con el **TRACE** por las diversas gráficas bastará mover los cursores según se ha explicado anteriormente:

Movemos los cursores ► ◄ para ir hacia adelante y hacia atrás, respectivamente, dentro de la gráfica en la que nos encontremos, y los cursores ▼ ▲ para pasar de una gráfica a otra, que será la que indique el rótulo situado en la parte superior izquierda de la pantalla.

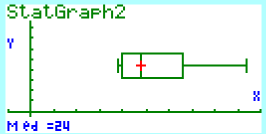


ACTIVIDAD 3

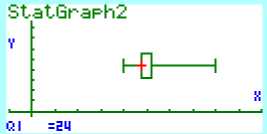
Dados los siguientes diagramas de cajas y bigotes y los siguientes diagramas de barras, asignar a cada diagrama de cajas su correspondiente diagrama de barras:

DIAGRAMAS DE CAJAS Y BIGOTES

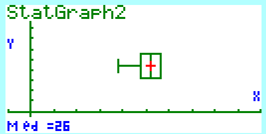
GRÁFICA 1



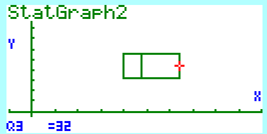
GRÁFICA 2



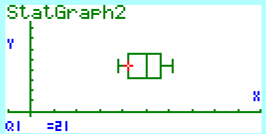
GRÁFICA 3



GRÁFICA 4



GRÁFICA 5



GRÁFICA 6

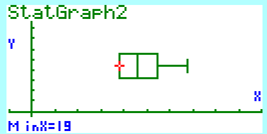
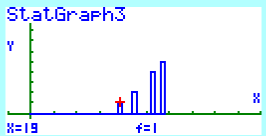
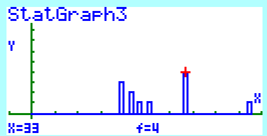


DIAGRAMA DE BARRAS

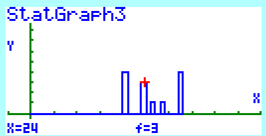
GRÁFICA A



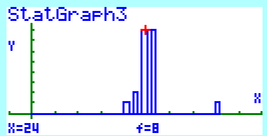
GRÁFICA B



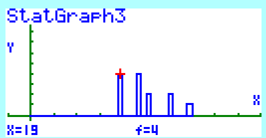
GRÁFICA C



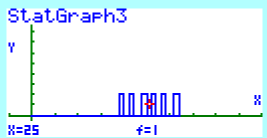
GRÁFICA D



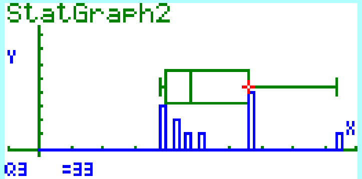
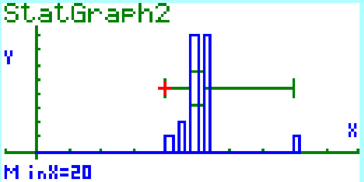
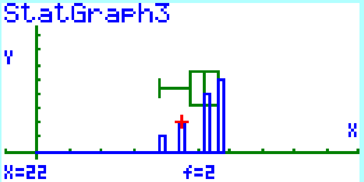
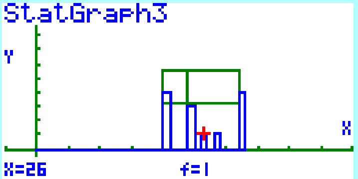
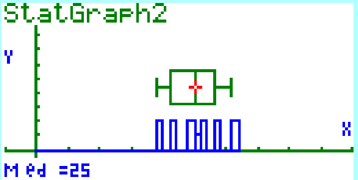
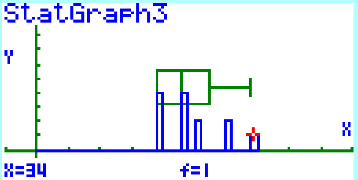
GRÁFICA E



GRÁFICA F



SOLUCIONES

GRÁFICA 1 – GRÁFICA B	GRÁFICA 2 – GRÁFICA D
	
GRÁFICA 3 – GRÁFICA A	GRÁFICA 4 – GRÁFICA C
	
GRÁFICA 5 – GRÁFICA F	GRÁFICA 6 – GRÁFICA C
	

Un ejercicio que podremos afrontar, rápidamente, con la ayuda de la calculadora gráfica, podría ser de este estilo:

ACTIVIDAD de repaso

Las notas obtenidas en Matemáticas por un alumno a lo largo del curso han sido las siguientes:

6	7.1	3	7	5.8	2.5	4.7	9	5.5	7.8	1	2	2

Contesta las cuestiones que más adelante se realizan especificando qué símbolo matemático representa a cada una de ellas y REDONDEANDO los resultados finales hasta las centésimas.

1.1.- ¿Cuántos exámenes ha realizado a lo largo del curso?.

1.2.- ¿Cuál fue la suma total de los puntos obtenidos?.

1.3.- Calcula la media aritmética de las notas.

1.4.- Calcula el valor que se encuentra en el centro de la distribución.

1.5.- Calcula el coeficiente de variación.

1.6.- ¿Es la media aritmética el parámetro que debe de utilizar el profesor para indicar si el alumno ha aprobado (nota ≥ 5 puntos) o suspendido (nota < 5 puntos) o crees que en este caso hay otro mejor?. Justifica tu respuesta.

PROBLEMAS PROPUESTOS
Estadística unidimensional

1.- Se ha aplicado un test de inteligencia a 40 alumnos, obteniéndose los siguientes resultados, agrupados en intervalos:

Puntuaciones	[15 – 20)	[20 – 25)	[25 – 30)	[30 – 35)	[35 – 40)	[40 – 45)
Nº de alumnos	3	8	13	7	6	3

Contesta las siguientes cuestiones especificando qué símbolo matemático representa a cada una de ellas.

¿Cuántos puntos obtuvieron en total los 40 alumnos?.

b) ¿Cuál es la cantidad CONCRETA de puntos más frecuentemente obtenida por los alumnos?

c) Un alumno que ha obtenido 29 puntos, ¿está por encima de la media?.

d) ¿Cuál es el valor CONCRETO que se encuentra en el centro de la distribución?.

e) ¿Cuál es el tipo de representación gráfica más adecuado para esta distribución?. Represéntalo de esta forma.

f) Realiza algún otro tipo de representación gráfica.

g) ¿Cuál es la desviación típica de las puntuaciones?. Interpreta su valor.

h) ¿Cuál es el coeficiente de variación?. A la vista de su valor, ¿cómo podrías decir qué es la distribución?.

i) ¿Cuál es la medida de centralización que mejor representa a la distribución?. Justifica la respuesta.

j) Para obtener una puntuación superior al 80% de los alumnos ¿qué nota hay que sacar?.

k) ¿En qué percentil está un alumno que obtiene 21.5 puntos?

2.- Las distribuciones de las edades de dos equipos de baloncesto, Equipo A y equipo B, son las siguientes:

Equipo A	
x_i	f_i
19	3
21	2
23	1
25	1
33	4
47	1

Equipo B	
x_i	f_i
25	1
26	2
27	5
28	3
30	1

Contesta las siguientes cuestiones especificando qué símbolo matemático representa cada una de ellas.

a) ¿Cuántos integrantes tienen las plantillas?.

b) ¿Cuánto suman las edades de cada una de las plantillas?

c) ¿Cuál es la media de edad de cada equipo?

d) ¿Cuál es la moda de edad de cada conjunto?

e) ¿Cuánto vale el coeficiente de variación de ambos equipos?

f) ¿Qué porcentaje, de jugadores del equipo A, se encuentra en el intervalo [$\bar{x}$ – S, $\bar{x}$ + S]?.

g) ¿Cuál es la medida de centralización que mejor representa al equipo?

h) ¿Cuál es la medida de centralización que mejor representa a cada uno de los equipos?

i) Interpreta y analiza los resultados, comparando ambos equipos.

3.- Las notas de un examen de Matemáticas de los alumnos de un grupo de 1º de Bachillerato han sido las siguientes:

2	3.4	4.5	5.6	2	8.1	3.5	5	5	8.7	5
9.4	4.5	3	5.7	6.6	3	2.7	3.9	4.9	5.1	8
6.7	9	5	7	7	6	5.6	5	4.7	5	5.7

Contesta las siguientes cuestiones especificando qué símbolo matemático representa cada una de ellas.

a) ¿Cuántos alumnos hicieron el examen?.

b) Calcula la media aritmética de las notas obtenidas

c) ¿Cuál fue la suma total de los puntos obtenidos?.

d) Calcula la desviación típica.

e) Calcula el coeficiente de variación.

f) Calcula el valor concreto de la mediana.

g) Calcula el valor concreto de la moda.

h) ¿Qué porcentaje de alumnos han obtenido notas que se encuentra en el intervalo [$\bar{x}$ – S, $\bar{x}$ + S]?

i) Analiza y comenta los resultados a la vista de este último valor y del coeficiente de variación:

j) ¿Qué porcentaje de alumnos han obtenido notas que se encuentra en el intervalo [$\bar{x}$ – 2S, $\bar{x}$ + 2S]?.

k) ¿Qué porcentaje de alumnos han obtenido notas que se encuentra en el intervalo [$\bar{x}$ – 3S, $\bar{x}$ + 3S]?.

l) Si el profesor decide subir un punto en todas las pruebas realizadas, ¿qué efecto tendrá sobre la media aritmética, la desviación típica y el coeficiente de variación?. Analiza y comenta los resultados.

ESTADÍSTICA BIDIMENSIONAL

OBJETIVOS: Realizar un repaso rápido de la estadística bidimensional vista en la ESO, analizando el significado de los diferentes parámetros estadísticos, añadiendo algún procedimiento nuevo y moderno para encontrar una recta que se ajuste a la nube de puntos e incorporar nuevas tecnologías como soporte visual.

INTRODUCCIÓN: A través de unas actividades y utilizando como herramienta habitual la calculadora gráfica, realizamos el análisis crítico y el estudio de los parámetros obtenidos en la observación simultánea de dos variables, con un enfoque más investigador e innovador, dando prioridad al razonamiento, permitiéndonos más tiempo para pensar y analizar lo que hacemos y los resultados que obtenemos.

ACTIVIDAD 1

La siguiente tabla recoge las puntuaciones obtenidas en un test sobre visión espacial (T) y sus correspondientes calificaciones en la asignatura de Dibujo (D):

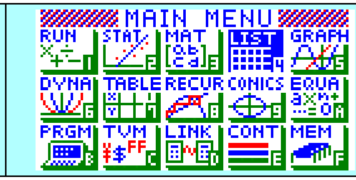
T	54	40	66	70	60	58	63	60
D	3	2	6	8	4	3	7	4

Con la calculadora gráfica podemos obtener rápidamente el valor de numerosos parámetros estadísticos para pasar, a continuación, a analizar su significado y hacer un breve comentario.

La forma de introducir los datos en la máquina es muy sencilla:

AC

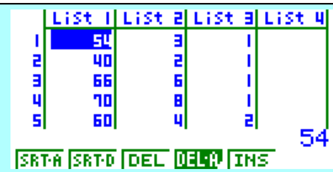
Selecciona el modo **LIST.**
Recuerda que puedes presionar directamente la tecla 4



El cursor aparecerá en la celda del primer elemento de la "**List1**"

En **List1** cargamos la variable T y en **List 2** la variable D, dejando para la **List 3** la frecuencia con la que aparece cada par (T, D).

Una pantalla podría ser ésta ⇨



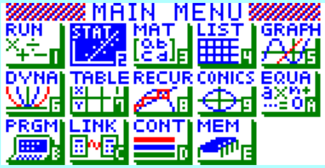
Por ejemplo, la fila 5 indica que 2 alumnos han obtenido 60 puntos en el Test sobre visión espacial y un 4 en Dibujo.

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 $\mathbf{x_i}$ los tenemos señalados en la **List1,** los valores de $\mathbf{y_i}$ los tenemos en la **List2** y que las frecuencias se encuentran en la **List3.** Para ello, bastará con realizar los siguientes pasos:

1.- Entrar en el MODO ESTADÍSTICO

MENU ◄ ◄

o bien, directamente, 2



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

2.- Adecuar las listas a las variables:

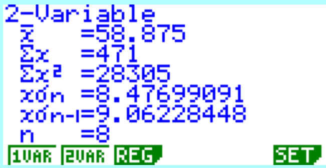
CALC SET
F2 F6

Estudiamos sólo las opciones que comienzan por **2VAR** (Variables bidimensionales) y le indicamos qué **Listas** serán las que se corresponden con cada una de las variables, quedando como se indica al margen ⇨

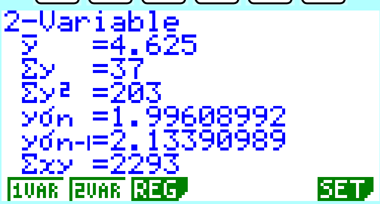
EXE

Solicitamos que nos calcule el valor de los parámetros con dichas 2 variables:

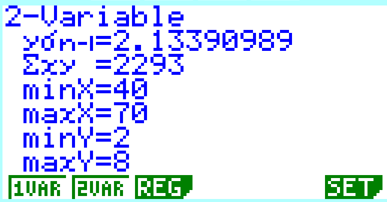
2VAR
F2



▼ ▼ ▼ ▼ ▼ ▼



▼ ▼ ▼ ▼



A continuación, ya estaríamos en disposición de interpretar, debatir y reflexionar acerca de los resultados obtenidos en el test sobre visión espacial (T) y las calificaciones correspondientes en la asignatura de Dibujo (D), evitando los tediosos, aburridos e interminables cálculos aritméticos, dedicando más tiempo al análisis de los diferentes conceptos y conclusiones que encierra cada uno de dichos valores.

TEST SOBRE VISIÓN ESPACIAL (T)

n ➡ El número de alumnos estudiados en un test sobre visión espacial es 8.

Σx ➡ La suma de todos los puntos obtenidos en el test sobre visión espacial por los 8 alumnos estudiados ha sido de 471 puntos.

Σx^2 ➡ La suma de los cuadrados de los puntos obtenidos en el test sobre visión espacial por los 8 alumnos ha sido de 28 305 puntos cuadrados.

Sx_{n-1} ➡ Si los 8 alumnos estudiados se consideran una muestra de la población, la desviación respecto de la media aritmética sería de 9.06228448 puntos.

La media $\bar{x}$ de las puntuaciones obtenidas en el test sobre visión espacial es de 58.875 puntos, estando el 75% (*) comprendidos entre 49.813 puntos y 67.937 puntos

(*) NOTA: ¿cómo se obtiene ese valor 75% ?
Hay 6 elementos de 8 posibles incluidos en ese intervalo:

$$\frac{6}{8}\cdot\frac{100}{100} = \frac{75}{100} = 75\%$$

Sx_n ➡ Si los alumnos estudiados se hubiesen considerado como "población", la desviación respecto de la media aritmética sería de 8.47699091 puntos.

CALIFICACIONES EN LA ASIGNATURA DE DIBUJO (D)

n ➡ El número de alumnos estudiados en las calificaciones de dibujo es 8.

Σy ➡ La suma de todas las notas de Dibujo ha sido de 37 puntos.

Σy^2 ➡ La suma de los cuadrados de los puntos obtenidos en las calificaciones de Dibujo ha sido de 203 puntos cuadrados.

Sy_{n-1} ➡Si los 8 alumnos estudiados se consideran una muestra de la población, la desviación respecto de la media aritmética en las calificaciones de Dibujo es de 2.13390989 puntos.

La media $\bar{y}$ de las puntuaciones obtenidas en la asignatura de Dibujo es de 4.625 puntos, estando el 62.5% (*) comprendidos entre 2.491 puntos y 6.759 puntos

(*) NOTA: ¿cómo se obtiene ese valor de 62.5% ?.
Hay 5 elementos de 8 posibles dentro de ese intervalo:

$$\frac{5}{8}\cdot\frac{100}{100} = \frac{62.5}{100} = 62.5\%$$

De esta forma se puede trabajar con el alumno el significado de cada uno de ellos, concienciándolo de que no se trata de una simple relación de números, sino que están vivos y reflejan cosas, situaciones, persiguiendo fundamentalmente el ANÁLISIS y la toma de DECISIONES

Un valor que no calcula directamente, pero que es muy fácil de averiguar, es la COVARIANZA:

$$\Rightarrow S_{xy} = \frac{\sum x_i \cdot y_i \cdot f_i}{N} - \bar{x}\cdot\bar{y}$$

El valor $\Sigma x_i \cdot y_i \cdot f_i$ viene representado en la calculadora por **Σxy**

$$\frac{2293}{8} - 58.875 \cdot 4.625$$

SOLUCIÓN: Covarianza = 14.33*

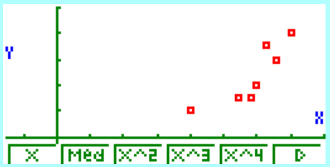
Otra de las ventajas de la calculadora gráfica es que nos permite representar con total sencillez el **DIAGRAMA DE DISPERSIÓN O NUBE DE PUNTOS**, para poder analizarlo rápidamente.

GRPH SET EXIT EXIT **F1** **F6** Así seleccionamos la forma de la gráfica, que podría ser la **GPH1** que aparece al margen ⇨	

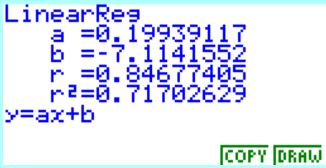
No hay que olvidar que hay que trabajar en **MANUAL** para introducirle el valor de los parámetros de visualización de pantalla que queramos:

SETUP Man EXE SHIFT MENU **F2** EXE SHIFT V-Window **F3** La escala adecuada para una correcta visualización de esta gráfica podría ser la señalada ⇨	

Recuerda que este tipo de gráfica lo tenemos diseñado en **GPH1**. Hay que tener en cuenta que, aún cuando las frecuencias de un par son mayores que la de otro, los puntos son trazados con igual grosor.

GRPH GPH1 EXE **F1** **F1** Podría ser la que se indica adjunta	

Si queremos dibujar la recta de regresión de **y** sobre **x**, conocer su ecuación y su coeficiente de correlación, bastará con realizar estos sencillos pasos:

X **F1**	

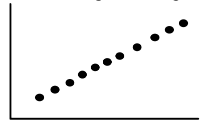
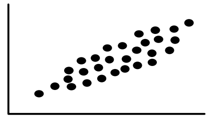
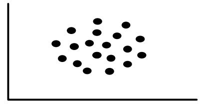
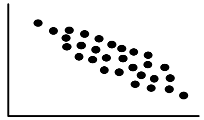
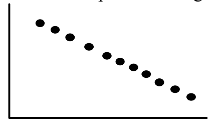
Obtenemos los valores teóricos de los parámetros **"a"** y **"b"** de la expresión y = ax + b

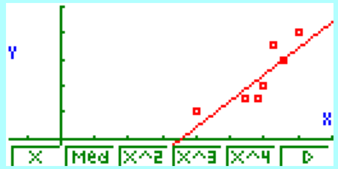
y = 0.19939117x – 7.1141552

Recordemos que "r" es el coeficiente de correlación lineal de Pearson.

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

El valor de "r" va a oscilar entre –1 y 1

r = 1 Todos los valores que toma la variable bidimensional se van a encontrar sobre una recta. Se dice que entre X e Y existe una dependencia funcional.	Es una recta de pendiente positiva 
0 < r < 1 La relación será directa y más fuerte cuanto más se aproxime su valor a 1 y más débil cuanto más se aproxime a 0.	
r = 0 No existe ninguna relación LINEAL entre ambas variables. Son independientes.	
– 1 < r < 0 La relación será inversa y más fuerte cuanto más se aproxime su valor a – 1 y más débil cuanto más se aproxime a 0.	
r = – 1 Todos los valores que toma la variable bidimensional se van a encontrar sobre una recta. Se dice que entre X e Y existe una dependencia funcional	Es una recta de pendiente negativa 

DRAW **F6** La opción **DRAW** dibujará la recta de regresión de y sobre x	

Si lo que queremos es guardar la ecuación de la recta de regresión como función, para utilizarla en cualquier momento o realizar algún estudio, sólo habrá que presionar:

COPY
F5 EXE

De esta forma procederemos a estudiar el grado y tipo de relación existente entre 2 variables, extrayendo las conclusiones apropiadas, analizando el grado de relación conocido el coeficiente de correlación y su nube de puntos:

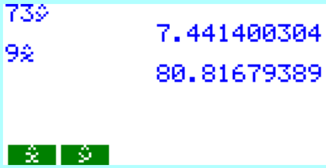
r = 0.85 ⇨ Se trata de una dependencia directa bastante fuerte, ya que se aproxima a 1

Asimismo las **ESTIMACIONES** son muy sencillas de realizar; por ejemplo, si un alumno ha obtenido un 73 en el test sobre visión espacial, ¿cuánto se espera que obtenga en dibujo?; y si ha logrado un 9 en dibujo, ¿cuánto se espera obtenga en el test?

*NOTA: Al estar "r" cercano a "**1**" los valores que alcancemos estarán próximos al valor estimado.*

Estos cálculos han de realizarse desde la opción **RUN** del **MENÚ INICIAL**.

MENU 1 OPTN STAT F5 7 3 $\hat{y}$ F2 EXE 9 $\hat{x}$ F1 EXE



ANÁLISIS CRITICO DE LOS RESULTADOS:

Se estima que un alumno que ha obtenido un 73 en el test sobre visión espacial obtendrá un 7.44 en dibujo [la recta de regresión explicaría el 84.68% de los casos, ya que r = 0.8468] y si ha logrado un 9 en dibujo se espera obtenga en el test 80.82 puntos.

ACTIVIDAD 2 (ERRORES HABITUALES)

Si la tabla que recoge las puntuaciones obtenidas en un test sobre visión espacial (T) y las calificaciones en la asignatura de Dibujo (D) tuviesen esta nueva distribución:

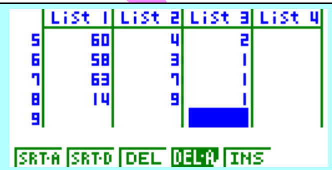
T	14	54	40	66	70	60	58	63	60
D	9	3	2	6	8	4	3	7	4

dibuja la nube de puntos correspondiente y estudia si la recta de regresión es la forma idónea de hacer una estimación. ¿Habrá alguna recta que se ajuste mejor?.

RESOLUCIÓN:

MENU 4

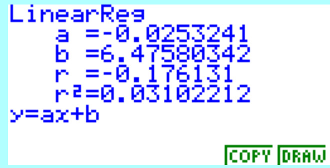
Llevamos el cursor al final de las columnas e introducimos el valor T, D (14, 9) con frecuencia 1



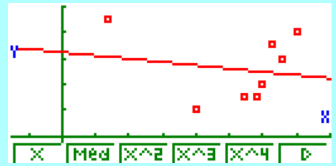
Representamos gráficamente y dibujamos la recta de regresión de y sobre x:

MENU 2 GRPH F1 GPH1 F1 X F1

obteniéndose la ecuación de dicha recta:



DRAW F6

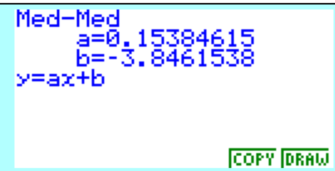


Como es una recta basada en las **medias** ($\bar{x}$, $\bar{y}$), podremos observar que se ajusta muy mal a la **nube de puntos** debido a ese punto extremo (14, 9), que afecta sensiblemente a dichas medias.

¿Qué representación utilizaremos en estos casos, con presencia de "outliers" (puntos fuera de línea), que se salen de un margen "razonable"?.

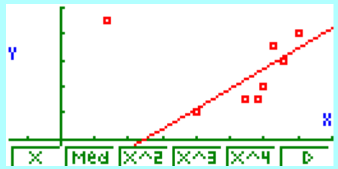
John Wilder Tukey desarrolló una recta basada en la **mediana** y que en la actualidad conocemos con el nombre de **recta de Tukey**, **recta Med**, **recta Med-Med** o **recta resistente** que, como podremos observar, se ajusta mucho mejor a la mayor parte de los puntos de la nube:

Med F2



La opción **DRAW** dibujará la dicha recta con suma facilidad

DRAW F6



En realidad ésta es una recta muy representativa en caso de presencia de "outliers", siempre y cuando tengamos una calculadora gráfica, ya que con LÁPIZ Y PAPEL su cálculo resulta ciertamente una labor ardua y costosa.

ACTIVIDAD 1

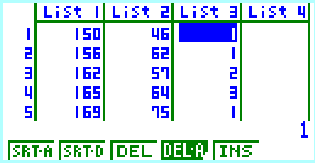
Al tallar y pesar a varias personas hemos obtenido los siguientes resultados:

Estatura en cm (x)	150	156	162	165	169	171	177	180	184	186
Peso en kg (y)	46	62	57	64	75	68	76	82	82	80
Frecuencia	1	1	2	3	1	2	1	1	1	1

a) Averigua el valor de los parámetros estadísticos que consideres más representativos:

En **List1** cargamos la variable **X**, en **List2** la variable **Y** y en **List 3** la frecuencia con la que aparece cada par (T, D).

Una pantalla podría ser ésta



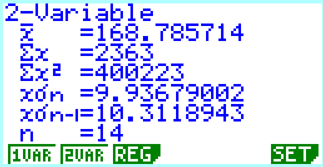
CALC SET

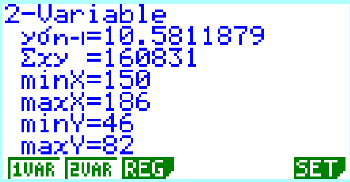
MENU 2 F2 F6

Adecuamos las listas a las variables

2VAR

EXE F2





b) Contesta a las siguientes cuestiones cortas, indicando el símbolo matemático que representa a cada una de ellas:

b1) ¿Cuántas personas conformaban el grupo consultado?.

n = 14 personas

b2) ¿Cuál es el peso medio de los individuos entrevistados?.

$\bar{y}$ = 67.5 kg

b3) ¿Cuál es el peso total de los individuos entrevistados?.

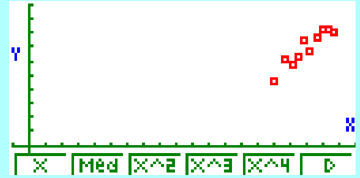
$\bar{y}$ = 945 kg

c) Calcula el valor de la covarianza:

$$Sxy = \frac{\sum x_i \cdot y_i \cdot f_i}{N} - \bar{x} \cdot \bar{y}$$

$$Sxy = \frac{160831}{14} - 168.785714 \cdot 67.5 = \mathbf{94.89287643}$$

d) Dibuja la nube de puntos correspondiente y COMENTA, de forma intuitiva, a la vista de la misma, la relación existente entre ambas variables.

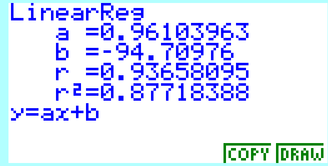


✹ Aparentemente, la relación es lineal, directa, bastante fuerte y positiva.

e) Cuantifica, de forma matemática y precisa, la relación existente entre la talla y el peso de la distribución. ¿Cómo se llama el parámetro que utilizas para ello?. Interpreta el resultado.

X

F1



RESOLUCIÓN:

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{94.89287643}{9.93679002 \cdot 10.1962878} = \mathbf{0.9365811514}$$

— Este parámetro se le conoce con el nombre de coeficiente de correlación lineal de Pearson.

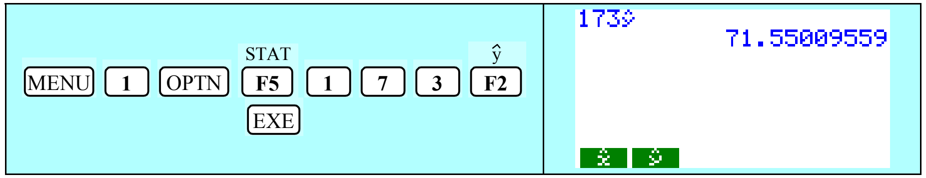
r = 0.94 ➩ Por lo que la relación es lineal, directa, bastante fuerte y positiva.

f) Si escogemos al azar un amigo de los encuestados y mide 173 cm, ¿cuál estimarás que será su peso?. Calcula la ecuación de la recta de regresión correspondiente.

RESOLUCIÓN:

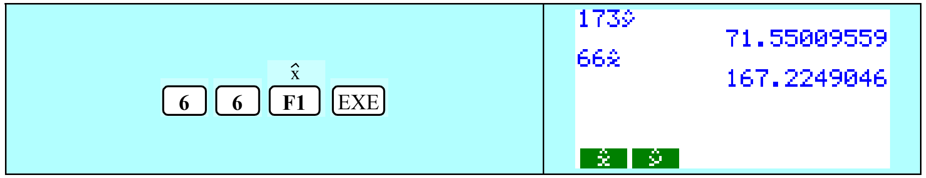
De la pantalla anterior obtenemos los valores teóricos de los parámetros **"a"** y **"b"** a la expresión y = ax + b

y = 0.96 x – 94.71



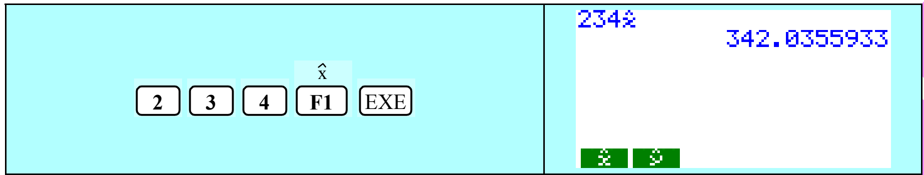
<u>SOLUCIÓN:</u> Se espera que dicho amigo tenga un peso aproximado de 71.37 kg [la recta de regresión explicaría el 93.66% de los casos (r = – 0.9366)].

g) Si otro amigo pesa 66 kg, ¿cuál es la altura que se espera tenga?.



<u>SOLUCIÓN:</u> Se espera que dicho amigo tenga una estatura de, aproximadamente, 167.24 cm [la recta de regresión explicaría el 93.66% de los casos (r = – 0.9366)].

h) Si otra persona amiga pesase 234 kg ¿cuánto se estima que tendría de estatura?.



SOLUCIÓN: Se esperaría que dicho amigo midiese 342.03 cm, PERO debemos de tener cierta desconfianza pues las observaciones de que disponemos (en las cuales se basan nuestros cálculos) son muy lejanas a esas alturas y esos pesos.

PROBLEMAS PROPUESTOS
Estadística bidimensional

1.- Las estaturas de 10 chicas y de sus respectivas madres son las siguientes, expresadas en centímetros:

Hijas (x)	158	162	164	165	168	169	172	172	174	178
Madres (y)	163	155	160	161	164	158	175	169	166	172

a) Dibuja la nube de puntos correspondiente y comenta de forma intuitiva, a la vista de la misma, la relación que existe entre ambas variables:

b) Cuantifica, de forma matemática y precisa, la relación existente entre la talla y el peso de la distribución.

c) Calcula la ecuación de la recta de regresión de "y" sobre "x".

d) Si una mujer que mide 1.63 tiene una hija ¿cuánto se espera que mida ésta?.

e) Si una chica mide 1.70 m ¿cuánto se espera que mida la madre?.

f) Dibuja la recta de regresión.

2.- Colgamos diversos pesos de un muelle y observamos que cuanto mayor sea el peso más se estira el muelle. La siguiente tabla nos da los pesos colgados (en gramos) y los correspondientes alargamientos del muelle (en centímetros):

Masa (x)	0	30	60	90	120	150	180	210	240	270
cm (y)	0	9	17	26	35	43	52	61	70	79

a) Dibuja la nube de puntos correspondiente y comenta de forma intuitiva, a la vista de la misma, la relación que existe entre ambas variables:

b) Cuantifica, de forma matemática y precisa, la relación existente entre la talla y el peso de la distribución.

c) Calcula la ecuación de la recta de regresión de "y" sobre "x".

d) Calcula la ecuación de la recta de regresión de "x" sobre "y".

e) Si nos dan un objeto que, al colgarlo, alarga el muelle 65 cm ¿cuál será su masa?.

f) Si el peso del objeto es de 876 gr ¿cuánto se espera que se alargue el muelle?.

3.- El consumo de energía "*per cápita*" en miles de kw/h y la renta "*per cápita*" en miles de dólares de seis países de la UE son las siguientes:

	Alem	Bélg	Dinam	España	Francia	Grecia
Consumo (y)	5.7	5.0	5.1	2.7	4.6	2.65
Renta (x)	11.1	8.5	11.3	4.5	9.9	4.4

a) Averigua con la calculadora científica el valor de los siguientes parámetros estadísticos:

$\bar{x}$	Sx_n	Sx_{n-1}	Σx	Σx^2	n
$\bar{y}$	Sy_n	Sy_{n-1}	Σy	Σy^2	

b) Dibuja el diagrama de dispersión correspondiente y comenta de forma intuitiva, a la vista del mismo, la relación que existe entre el consumo y la renta de los países indicados:

c) Indica el valor de la correlación existente entre entre el consumo y la renta de los países indicados. Interpreta el resultado.

d) ¿Qué predicción podemos hacer sobre el consumo "per cápita" de Italia si sabemos que su renta es 6500 dólares?.

4.- En una cofradía de pescadores las capturas registradas de cierta variedad de pescado ("x" en Kg) y el precio de subasta en lonja ("y" PTAS/Kg) son los siguientes:

x	2.000	2.400	2.500	3.000	2.900	2.800	3.160
y	300	280	274	220	240	250	200

a) Contesta las siguientes cuestiones cortas, indicando el símbolo matemático que representa cada una de ellas:

a1) ¿Cuántos observaciones se han realizado?.

a2) ¿Cuál es la media de la cantidad de capturas realizadas?.

a3) ¿Cuál es la cantidad de dinero ingresado por la lonja en el supuesto de que se venda todo?

a4) ¿Cuál es el precio medio de las capturas realizadas?.

a5) ¿Cuántas toneladas de pescado han entrado en la lonja?

b) Calcula el valor de la covarianza:

c) Dibuja el diagrama de dispersión correspondiente y COMENTA, a la vista del mismo, la dependencia entre el número de hematíes y la tasa de hemoglobina

d) Analiza el tipo de dependencia entre las dos variables. Justifica lo que haces.

e) Si un día determinado se capturan 3450 kg de dicha especie, ¿cuál es el precio del kg que se espera?. Deduce la ecuación de la recta de regresión correspondiente.

f) Si un día llegamos a la lonja y comprobamos que el precio se encuentra a 150 PTAS/kg. ¿Cuánto estimas que será la captura de pescado de ese día?. Deduce la ecuación de la recta de regresión correspondiente.

g) ¿Y si se capturan 5100 kg de dicha especie, ¿cuál es el precio del kg que se espera?.