

GEOMETRÍA DINAMICA Y ESTADÍSTICA CON LA CLASSPAD 300 Y LA FX-9860G SD

5. ESTADÍSTICA DESCRIPTIVA Y ANÁLISIS DE REGRESIÓN

CASIO®

MAURICIO CONTRERAS

ESTADÍSTICA DESCRIPTIVA Y REGRESIÓN CON LA CLASSPAD 300

Introducción


Para adquirir una visión crítica de las fuentes de información y desarrollar capacidades como la de búsqueda, selección, control y evaluación de la información, es necesario realizar experiencias que permitan conocer y aplicar técnicas matemáticas adecuadas y recursos tecnológicos apropiados.

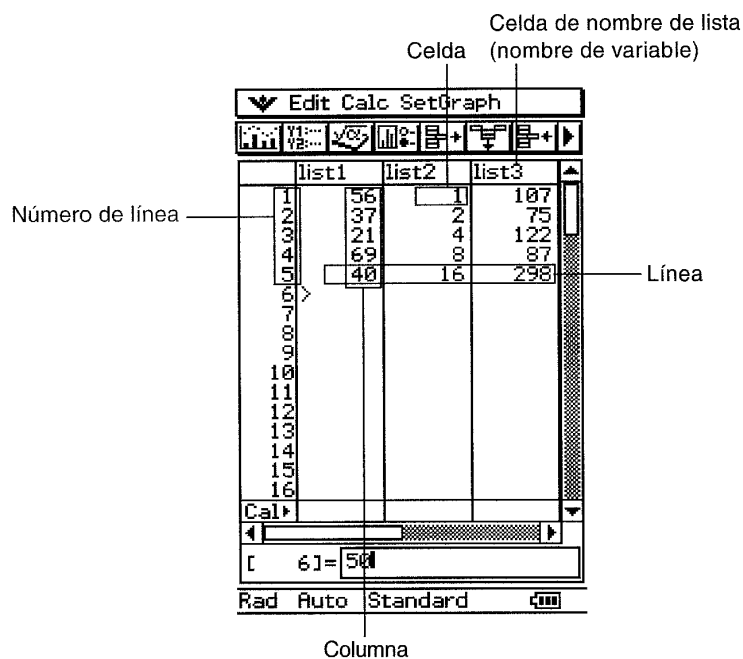
La ClassPad 300 permite generar números aleatorios y hacer simulaciones, seleccionar muestras, representar los datos, calcular medidas de centralización y dispersión, ajustar una curva a un conjunto de datos estadísticos, etc.

En esta sesión estudiaremos algunas de las posibilidades de la ClassPad 300 para el estudio de la Estadística en ESO y Bachillerato.

1. Estadística descriptiva y análisis de regresión

1. LA APLICACIÓN ESTADÍSTICA

- En el menú de aplicaciones toca el botón . Se inicia la aplicación Estadística y se muestra la ventana del editor de listas.

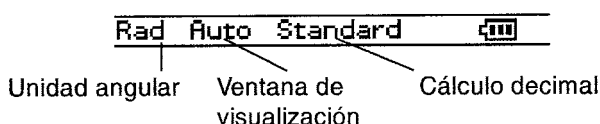


- En la siguiente tabla tienes los comandos disponibles en los menús y botones del editor de listas:

Función	Botón	Acción
Edit / Abrir lista	—	Abre una lista existente
Edit / Cerrar lista	—	Cierra la lista seleccionada
Edit / Ir al principio	—	Salta a la línea 1 de la lista actual
Edit / Ir al final	—	Salta a la línea siguiente de la última línea en la lista actual.

Función	Botón	Acción
Edit / Ordenar (Asc.)		Ordena los datos de una lista en orden ascendente
Edit / Ordenar (Desc.)		Ordena los datos de una lista en orden descendente
Edit / Borrar / Celda		Borra una celda
Edit / Borrar / Columna		Borra todos los datos de una lista
Edit / Borrar / Variable lista	—	Borra una lista de la memoria
Edit / Insertar celda		Inserta una celda en una lista
		Convierte una expresión matemática a un valor.
		Dibuja un gráfico estadístico.
		Muestra la ventana del editor de gráficos.
		Muestra la ventana de la aplicación Principal
▼ /Preferencias / Ventana vis.		Muestra el cuadro de diálogo de la ventana de visualización.
▼ /Preferencias / Adm. de variable		Muestra el administrador de variables.
ConfGraf / Opciones		Muestra el cuadro de diálogo de configuración de gráficos estadísticos.
		Muestra dos columnas en la ventana del editor de listas
		Muestra tres columnas en la ventana del editor de listas
		Muestra cuatro columnas en la ventana del editor de listas

- La barra de estado muestra la configuración del editor de listas, indicando las unidades de ángulos, el tipo de ventana de estadística y la configuración del cálculo decimal.



2. TRABAJANDO CON LISTAS

- En la ventana del editor de listas, toca la celda del nombre de lista en la parte superior de la lista list1. De esta forma se selecciona la celda de nombre de lista.
- Introduce el nombre de lista (hola), de ocho bytes como máximo, y pulsa [EXE]. Si introduces un nombre de lista existente, aparecerán los datos de dicha lista. Si introduces un nombre de lista sin indicar una carpeta, el nombre de variable se almacena en la carpeta actual.
- En la ventana del editor de listas, selecciona la celda "Cal" de la lista "hola" e introduce la lista {1,2,3} utilizando para ello el teclado virtual [mth]. Toca el botón [EXE].

	list1	list2	list3
1	1		
2	2		
3	3		
4	4		
5	5		
Cal	"C1,2..."		
Cal= {1,2,3}			

- Multiplica el valor de cada celda de la lista "hola" por 2 y escribe los resultados en la lista list2. Para ello:




1) Selecciona la celda "Cal" de la lista list2.

2) En el cuadro "Cal=" introduce la fórmula $hola \times 2$ y pulsa [EXE] para efectuar el cálculo. Aparecen en la list2 los valores de la lista hola multiplicados por 2.

	list1	list2	list3
1	1		
2	2		
3	3		
4	4		
5	5		
Cal	"C1,2..."		
Cal= {1,2,3}			

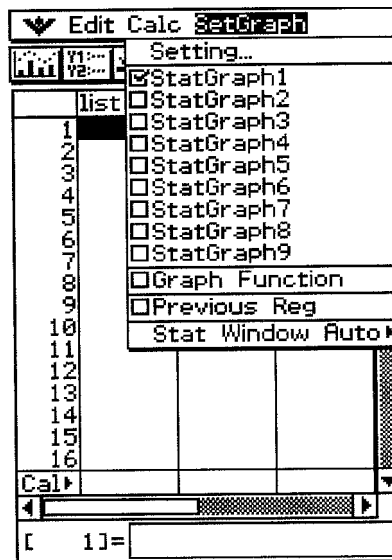
	list1	list2	list3
1	1	2	
2	2	4	
3	3	6	
4	4		
5	5		
Cal	"C1,2..."	"list1..."	
Cal= list1×2			

- Selecciona la celda de nombre de la lista list3 y cambia el nombre por "num3", utilizando el teclado virtual. A continuación, toca la celda "Cal" de dicha lista e introduce la lista {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} utilizando el teclado virtual [mth].
- Selecciona la celda 6 de la lista num3. Selecciona el comando Edit/Ir al principio.
- Selecciona la celda 4 de la lista num3. Selecciona el comando Edit/Ir al final.
- Selecciona la celda 5 de la lista num3. Selecciona el comando Edit / Borrar / Celda. Se borra dicha celda y las situadas por debajo se desplazan hacia arriba.
- Selecciona la celda 6 de la lista num3. Selecciona el comando Edit / Insertar celda. En la nueva celda introduce el valor 12 y pulsa [EXE].
- De la misma forma, selecciona la celda 9 de la lista num3 e inserta una nueva celda. Introduce en ella el valor 14 y pulsa [EXE].
- Selecciona el comando Edit / Ordenar (Asc.). En el siguiente cuadro de diálogo selecciona 1 lista y pulsa el botón [Acep.]. En la lista desplegable selecciona la lista main/num3 y toca el botón [Acep.]. La lista se ordena de menor a mayor.
- De la misma forma, selecciona el comando Edit / Ordenar (Desc.) y en el siguiente cuadro de diálogo selecciona 1 lista y toca el botón [Acep.]. En la lista desplegable, selecciona la lista main/num3 y toca el botón [Acep.] para ordenar la lista en orden decreciente.

- Con el cursor en una celda cualquiera de la lista num3, selecciona el comando Edit / Cerrar lista. La lista seleccionada desaparece de la pantalla y se reemplaza por una columna en blanco.
- Con el cursor en cualquier celda de la lista list1, selecciona el comando Edit / Abrir lista. En el cuadro List= situado en la parte inferior de la pantalla introduce el nombre de la lista, num3, usando para ello el menú virtual [abc]. Toca el botón [EXE] para que se visualice la lista.
- Selecciona la celda de nombre de la lista num3 y toca el cuadro "list=" en la parte inferior de la ventana del editor de listas. Pulsa la tecla [CLEAR] para que se borre la lista. De esta forma, se cierra la lista seleccionada.
- Selecciona la celda de nombre de la lista list3, pulsa [KEYBOARD] y, con ayuda del teclado virtual [abc], escribe el nombre de la lista num3 en la celda de nombre de lista. Toca el botón [EXE] y observa como se abre la lista num3.
- Selecciona la celda de nombre de la lista list5. En el cuadro "Calc=" introduce la fórmula $3 \cdot \text{num3} + 4$ y pulsa [EXE] para efectuar el cálculo. Aparecen los resultados en la lista list5.
- Selecciona el comando Edit / Ordenar (Asc.). En el siguiente cuadro de diálogo, selecciona 2 listas y pulsa el botón [Acep.]. En la lista desplegable que aparece a continuación del mensaje "Seleccionar lista de base", selecciona el nombre num3 y pulsa el botón [Acep.]. En la lista desplegable que aparece a continuación del mensaje "Seleccionar segunda lista", selecciona list5 y pulsa el botón [Acep.]. Observa como se ordenan las listas seleccionadas.
- Pulsa sucesivamente los botones  para cambiar el número de listas visibles en la pantalla del editor de listas. Observa que se pueden visualizar dos, tres o cuatro listas. Pulsa el botón  o el cuadro de desplazamiento horizontal situado en la parte inferior de la pantalla para visualizar más listas. Pulsa el botón  o el cuadro de desplazamiento vertical situado a la derecha de la pantalla para visualizar más valores de cada una de las listas.
- Selecciona la lista list5 y elige el comando Edit / Borrar / Columna. En el siguiente cuadro de diálogo toca el botón [Acep.]. De esta forma se borran los datos de la lista y ésta permanece vacía en la memoria.
- Selecciona la lista num3 y elige el comando Edit / Borrar / Variable Lista. En el siguiente cuadro de diálogo toca el botón [Acep.]. De esta forma se borra la lista de la memoria.
- Selecciona el comando Edit / Borrar todo. En el siguiente cuadro de diálogo, toca el botón [Acep.]. De esta forma se borran todos los datos del editor de listas y aparecen las listas 1 a 6 en blanco.

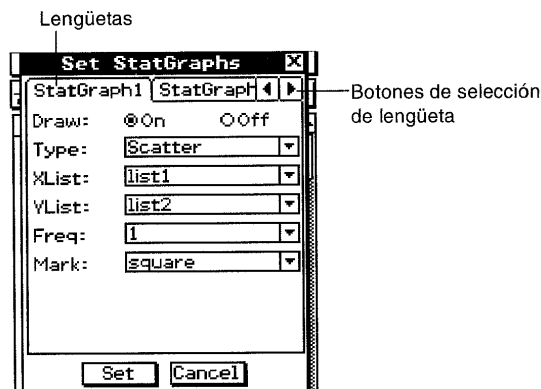
3. CONFIGURANDO GRÁFICOS ESTADÍSTICOS

- Toca el menú ConfGraf en la barra de menús del editor de listas. Aparece el menú de configuración de gráficos estadísticos.
- En la siguiente tabla tienes las opciones de dicho menú.



Función	Acción
Opciones...	Muestra un cuadro de diálogo para indicar el tipo de gráfico y la lista de datos para cada configuración de gráfico estadístico
Opciones... / <input checked="" type="checkbox"/> Gráfico estadís. 1 a <input checked="" type="checkbox"/> Gráfico estadís. 9	Selecciona una configuración de gráfico estadístico
<input checked="" type="checkbox"/> Función gráfica	Superpone un gráfico de función sobre un gráfico estadístico
<input type="checkbox"/> Función gráfica	Desactiva la superposición de un gráfico de función.
<input checked="" type="checkbox"/> Regresión previa	Representa gráficamente los resultados del último cálculo de regresión realizado.
<input type="checkbox"/> Regresión previa	Desactiva la representación gráfica de los últimos resultados de cálculo de regresión.
Vent. estad. autom. / On	Obtiene una configuración automática de las opciones de la ventana de visualización de estadísticas.
Vent. estad. autom. / Off	Configura las opciones de la ventana de visualización de estadísticas manualmente.

- En la ventana del editor de listas, selecciona el comando ConfGraf / Opciones... Aparece el cuadro de diálogo de configuración de gráficos estadísticos. En él hay una ficha para cada gráfico estadístico, denominadas Gráfico estadís. 1 a Gráfico estadís. 9.



- Toca la solapa de cada ficha para configurar cada gráfico estadístico. Toca los botones de selección On Off para activar o desactivar cada uno de los gráficos estadísticos.

- En la ListaX, selecciona el nombre de la lista (list1 a list6, o un nombre de lista) a usar como datos del eje X. La opción por defecto es list1.
- En la ListaY, selecciona el nombre de la lista (list1 a list6, o un nombre de lista) a usar como datos del eje Y. La opción por defecto es list2.
- En la lista Frec., selecciona 1 si todos las parejas (x, y) tienen frecuencia 1, o selecciona la lista (list1 a list6 o nombre de lista) que contiene las frecuencias de cada pareja (x, y).
- Configura las opciones de gráficos estadísticos, eligiendo el tipo de gráfico de la lista desplegable Tipo. Los tipos disponibles se muestran en la siguiente tabla:

Tipo	Gráfico
Disper.	Dispersión
Línea xy	Línea xy
GrafPNormal	Gráfico de probabilidad normal
Histogr.	Histograma
RegrLin	Gráfico de regresión lineal
RegrCuadr	Gráfico de regresión cuadrática
RegrCuart	Regresión de orden cuatro
RegrExp.	Regresión exponencial ($y = a \cdot e^{bx}$)
RegrPot	Regresión potencial
RegrLogis	Regresión logística

Tipo	Gráfico
CajaMed	Diagrama de caja Mediana
CajaMod	Diagrama de caja modificado
DistNor	Curva de distribución normal
L. Trazos	Línea a trazos
MedMed	Gráfico Med-Med
RegrCúbic	Regresión cúbica
RegrLog	Regresión logarítmica
Rexp.ab	Regresión exponencial ($y = a \cdot b^x$)
RegrSin	Regresión sinusoidal

- En la lista Marca, selecciona la forma que deben tener los puntos del gráfico. Pueden ser: cuadrado (□), cruz (×), punto grueso (■) y punto (•).
- Finalmente, toca el botón [Def.] para definir la configuración de los gráficos estadísticos.


4. GRÁFICOS ESTADÍSTICOS DE UNA VARIABLE

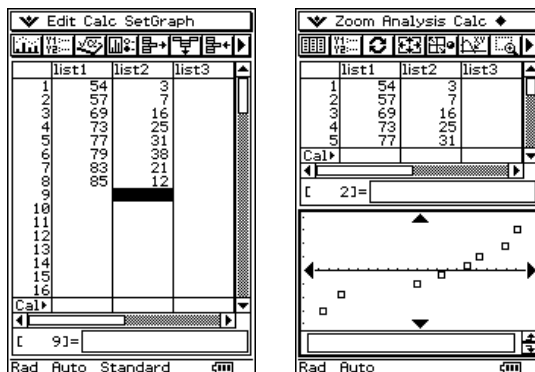
- Una compañía de seguros quiere realizar un estudio sobre la esperanza de vida de los españoles para ajustar sus cuotas de seguros. Para ello contrata los servicios de una empresa de investigación, que inicia el estudio con una muestra de 153 individuos, que resultaron tener una esperanza de vida X en años expresada en la siguiente tabla:

X	54	57	69	73	77	79	83	85	
F	3	7	16	25	31	38	21	12	$\sum F = 153$


Dibuja el gráfico de probabilidad normal, el histograma, el diagrama de cajas Med, el diagrama de cajas modificado, la curva de distribución normal y el gráfico de línea a trozos.

- **Gráfico de probabilidad normal**
- En el editor de listas, introduce los valores de X en la lista list1 y las frecuencias en la lista list2.

- A continuación, selecciona el comando *ConfGraf / Opciones...* En la siguiente ventana selecciona el tipo de gráfico *GrafPNormal*, *ListaX = list1*, *Marca = □ cuadrado*. Toca el botón [Def.]. Toca el botón  para dibujar el gráfico. Observa el resultado.



Histograma

- Selecciona el comando *ConfGraf / Opciones...* En la siguiente ventana selecciona el tipo de gráfico *Histogr.*, *ListaX = list1*, *Freq = list2*. Toca el botón [Def.]. Toca el botón  para dibujar el gráfico. Pulsa el botón [Acep.] para aceptar las dimensiones del intervalo. Observa el resultado.

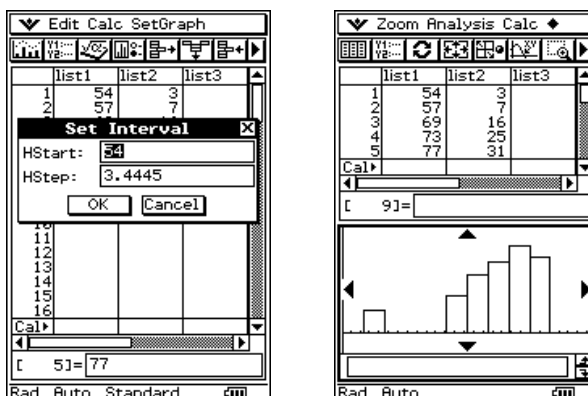
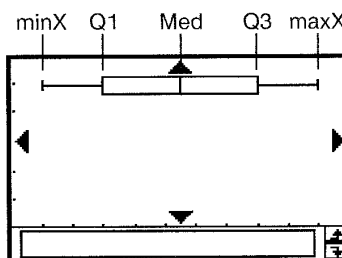

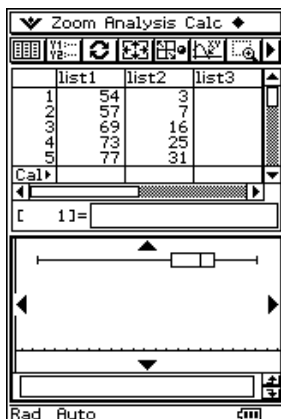


Diagrama de cajas Med

- Este gráfico se llama también "de cajas y bigotes" y representa los valores mínimo (minX) y máximo (maxX), así como los cuartiles Q_1 , Q_3 y la mediana *Med*, tal como se indica en la siguiente figura:

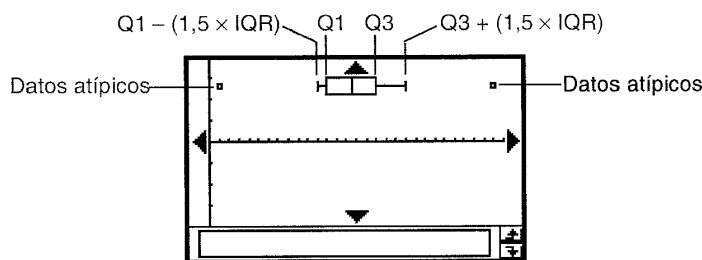



- Selecciona el comando *ConfGraf / Opciones...* En la siguiente ventana selecciona el tipo de gráfico *CajaMed.*, *ListaX = list1*, *Freq = list2*. Toca el botón [Def.]. Toca el botón  para dibujar el gráfico. Observa el resultado.

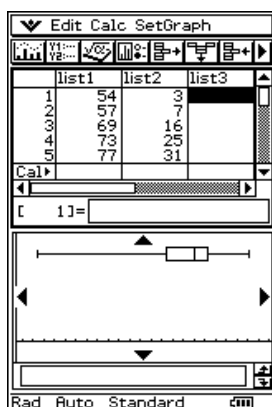


• **Diagrama de cajas modificado**

- Este gráfico utiliza la fórmula $1,5 \times RI$, siendo $RI = \text{rango intercuartílico} = Q_3 - Q_1$, para mostrar los valores alejados (outliers) que son aquellos valores que no siguen el mismo patrón que el resto de datos. Si un valor es menor que $Q_1 - 1,5 \times RI$, se dice que es un valor alejado por la izquierda. Si un valor es mayor que $Q_3 + 1,5 \times RI$, se dice que es un valor alejado por la derecha. Los valores alejados se representan por círculos en el diagrama.




- Selecciona el comando *ConfGraf / Opciones...* En la siguiente ventana selecciona el tipo de gráfico *CajaMod.*, *ListaX = list1*, *Freq = list2*. Toca el botón [Def.]. Toca el botón  para dibujar el gráfico. Observa el resultado.

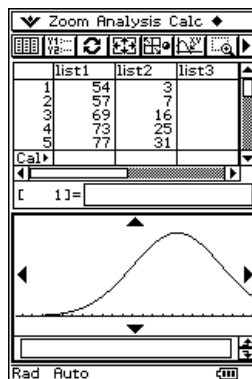


• **Curva de distribución normal**


- La curva de distribución normal se representa gráficamente utilizando la función de distribución

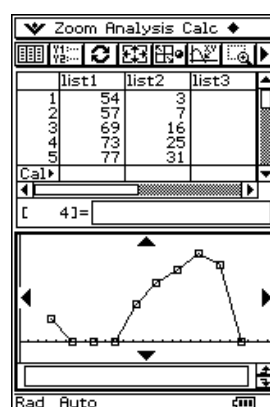
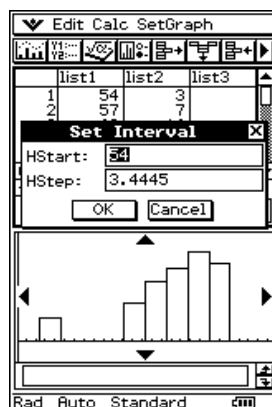
normal siguiente:
$$y = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

- Selecciona el comando ConfGraf / Opciones... En la siguiente ventana selecciona el tipo de gráfico DistNor., ListaX = list1, Freq = list2. Toca el botón [Def.]. Toca el botón  para dibujar el gráfico. Observa el resultado.



• **Gráfico de Línea a trazos**

- En este gráfico se conectan mediante segmentos los puntos centrales de cada barra del histograma.
- Selecciona el comando ConfGraf / Opciones... En la siguiente ventana selecciona el tipo de gráfico L.Trazos, ListaX = list1, Freq = list2. Toca el botón [Def.]. Toca el botón  para dibujar el gráfico. Toca el botón [Acep.] para aceptar las dimensiones del intervalo. Observa el resultado.








5. GRÁFICOS ESTADÍSTICOS DE DOS VARIABLES

- Diagrama de dispersión y gráfico de línea xy
- Dibuja el diagrama de dispersión de los siguientes datos y conecta los puntos para obtener un gráfico de línea xy. Sigue los siguientes pasos:

List1	0,5	1,2	2,4	4,0	5,2
List2	-2,1	0,3	1,5	2,0	2,4



- 1) En el menú de aplicaciones, toca el botón  para abrir el editor de listas.
- 2) Introduce en la listas list1 y list2 los datos anteriores.

- 3) Selecciona el comando *ConfGraf / Opciones...* o toca el botón .
- 4) En el cuadro de diálogo de configuración de gráficos, selecciona las siguientes opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec.=1, Marca: Cuadrado. Toca el botón [Def.].
- 5) Toca el botón  para trazar el gráfico de dispersión.
- 6) Toca la ventana de listas para activarla y selecciona el comando *ConfGraf / Opciones...* o toca el botón .
- 7) En la siguiente ventana configura el gráfico estadístico con las siguientes opciones: Dibujo: On, Tipo: Línea xy, ListaX=list1, ListaY=list2, Frec.=1, Marca=cuadrado. Toca el botón [Def.].
- 8) Toca el botón  para dibujar el gráfico Línea xy.

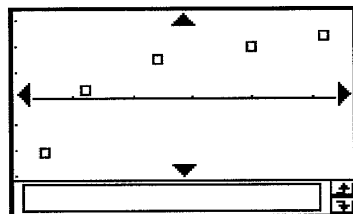


Diagrama de dispersión

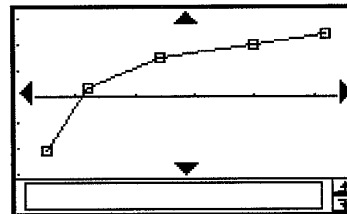




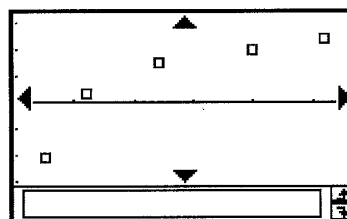
Gráfico de línea xy

• **Diagrama de dispersión y gráfico de regresión**

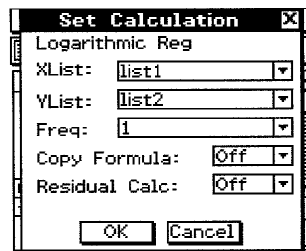
- Dibuja el diagrama de dispersión de los siguientes datos, realiza una regresión logarítmica de los datos para ver los parámetros de regresión y dibuja el gráfico de regresión. Sigue los siguientes pasos:

List1	0,5	1,2	2,4	4,0	5,2
List2	-2,1	0,3	1,5	2,0	2,4


- 1) Selecciona el comando *ConfGraf / Opciones...* o toca el botón .
- 2) En el cuadro de diálogo de configuración de gráficos, selecciona las siguientes opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec.=1, Marca: Cuadrado. Toca el botón [Def.].
- 3) Toca el botón  para trazar el gráfico de dispersión.

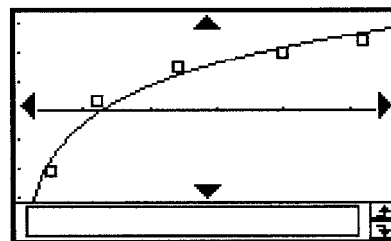
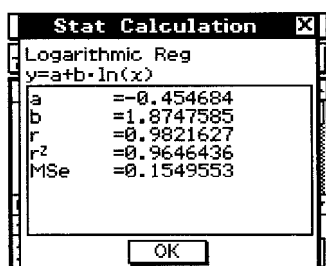


- 4) Selecciona el comando *Calc / Reg. Logarítmica*. En la ventana Definir cálculo introduce los valores de la siguiente ventana y toca el botón [Acep.].



5) En la ventana Calc. Estadístico tienes la función de regresión y el coeficiente de determinación r^2 que indica la fiabilidad de las aproximaciones hechas con la función de regresión. Toca el botón [Acep.].


6) Toca el botón  para que se dibuje la curva de regresión sobre el diagrama de dispersión.




• **Gráfico de regresión**

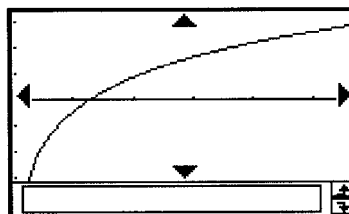
• Dibuja la curva de regresión de los siguientes datos, sin realizar los cálculos de regresión. Sigue los siguientes pasos:

List1	0,5	1,2	2,4	4,0	5,2
List2	-2,1	0,3	1,5	2,0	2,4

1) Selecciona el comando ConfGraf / Opciones... o toca el botón .

2) En el cuadro de diálogo de configuración de gráficos, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrLog., ListaX=list1, ListaY=list2, Frec.=1. Toca el botón [Def.].

3) Toca el botón  para trazar el gráfico de regresión.

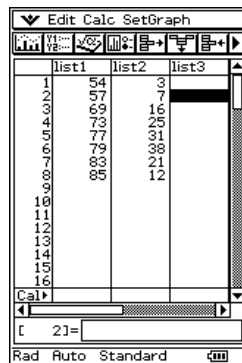



• **Gráfico de regresión lineal**

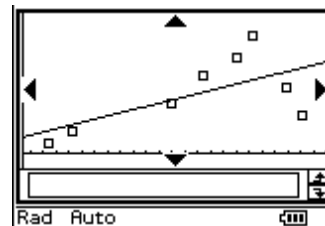
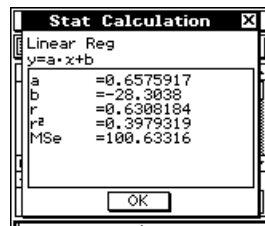
• Halla la ecuación de la recta de regresión lineal correspondiente a los siguientes datos. Sigue los siguientes pasos:

List1	54	57	69	73	77	79	83	85
List2	3	7	16	25	31	38	21	12

1) Introduce en las listas list1 y list2 del editor de listas los valores de la tabla.




- 2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones.. En la siguiente ventana elige las opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec=1, Marca=cuadrado. Toca el botón [Def.]. Observa el gráfico de dispersión.
- 3) Con la ventana de gráficos activada, selecciona el comando Calc. / Regresión lineal.
- 4) Toca el botón [Acep.] de la ventana Definir cálculo para aceptar las opciones que aparecen.
- 5) En la ventana Calc. Estadístico se muestra la ecuación de la recta de regresión lineal de mínimos cuadrados, el valor del coeficiente de correlación lineal r y el valor del coeficiente de determinación r^2 . Toca el botón [Acep.].
- 6) Toca el botón  para que se dibuje la recta de regresión.

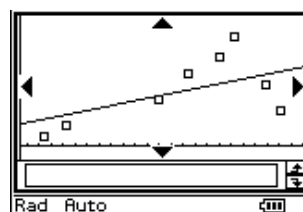


• **Gráfico Med-Med**

• Halla la ecuación de la recta mediana-mediana correspondiente a los siguientes datos. Sigue los siguientes pasos:


List1	54	57	69	73	77	79	83	85
List2	3	7	16	25	31	38	21	12

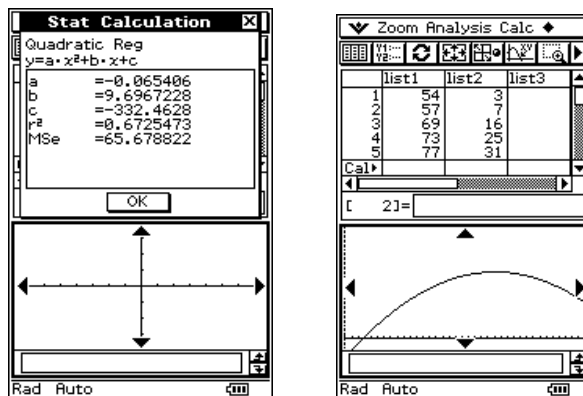
- 1) Con la ventana de gráficos activada, selecciona el comando Calc. / Línea MedMed.
- 2) Toca el botón [Acep.] de la ventana Definir cálculo para aceptar las opciones que aparecen.
- 3) En la ventana Calc. Estadístico se muestra la ecuación de la recta mediana-mediana. Toca el botón [Acep.].
- 4) Toca el botón  para que se dibuje la recta mediana-mediana.




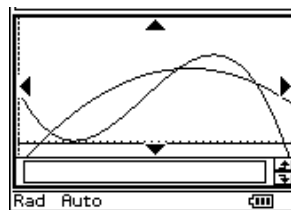
- **Gráficos de regresión cuadrática, cúbica y cuártica**
- Halla la ecuación de las curvas de regresión cuadrática, cúbica y cuártica correspondiente a los siguientes datos. Sigue los siguientes pasos:


List1	54	57	69	73	77	79	83	85
List2	3	7	16	25	31	38	21	12

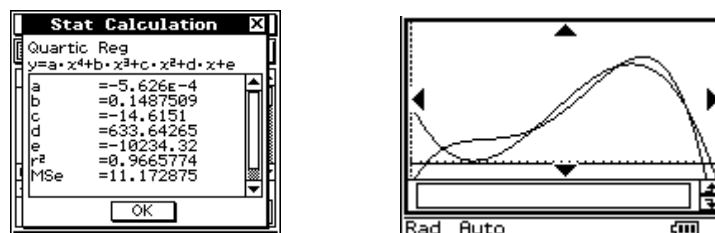
- 1) Con la ventana de gráficos activada, selecciona el comando Calc. / Regresión cuadrática.
- 2) Toca el botón [Acep.] de la ventana Definir cálculo para aceptar las opciones que aparecen.
- 3) En la ventana Calc. Estadístico se muestra la ecuación de la recta mediana-mediana. Toca el botón [Acep.].
- 4) Toca el botón  para que se dibuje la curva de regresión.



- 5) Toca la ventana de listas para activarla y selecciona el comando ConfGraf / Opciones... En la siguiente ventana introduce las opciones: Dibujo: On, Tipo: ReqrCubic, ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Def.].
- 6) Toca el botón  para trazar el gráfico de regresión. Observa la curva de regresión obtenida.



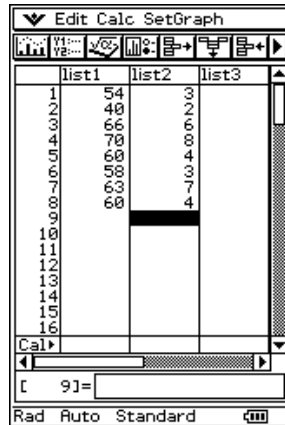
- 7) Con la ventana de gráficos activada, selecciona el comando Calc / Reqr. cuarto orden. En la siguiente ventana, toca el botón [Acep.].
- 8) Aparece la ecuación de la curva de regresión de cuarto grado. Toca el botón [Acep.]
- 9) Toca el botón  para que se dibuje la curva de regresión. Observa que se aproxima a los datos mejor que las curvas anteriores. Esto está indicado por el coeficiente de determinación r^2 que es superior a los casos anteriores.




- **Gráficos de regresión logarítmica, exponencial y potencial**
- La siguiente tabla recoge las puntuaciones obtenidas en un test sobre visión espacial (T) y sus correspondientes calificaciones en la asignatura de Dibujo (D). Halla la ecuación de las curvas de regresión logarítmica, exponencial y potencial correspondiente a dichos datos. Sigue los siguientes pasos:

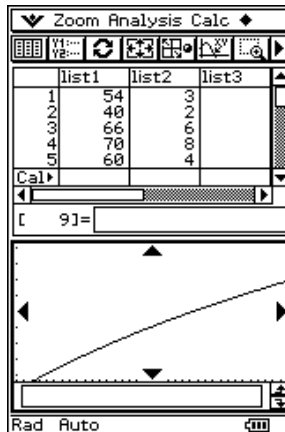
T	54	40	66	70	60	58	63	60
D	3	2	6	8	4	3	7	4

1) En el editor de listas, en las listas list1 y list2 introduce los valores de la tabla anterior.




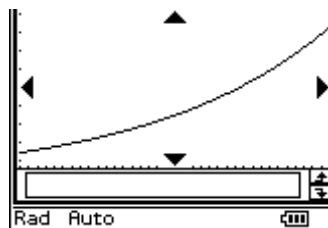
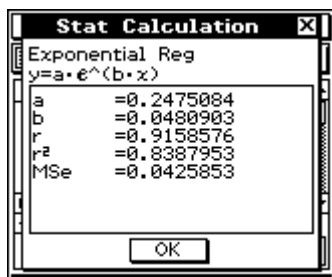
2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrLog, ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Def.].

3) Toca el botón  para dibujar el gráfico de regresión logarítmica.




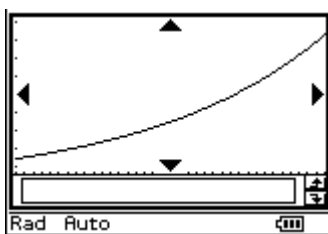
- 4) Con la ventana de gráficos activada, selecciona el comando \blacklozenge / Borrar todo.
- 5) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. exponencial. En la siguiente ventana selecciona las siguientes opciones ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].
- 6) En la ventana Calc. Estadístico se muestra la ecuación de la curva de regresión exponencial, que es de la forma $y = a \cdot e^{bx}$. Se indica también el valor del coeficiente de determinación r^2 que indica la bondad del ajuste de la nube de puntos por la curva de regresión.


- 7) Toca el botón [Acep.] y toca el botón  para dibujar la curva de regresión exponencial. Observa el resultado.

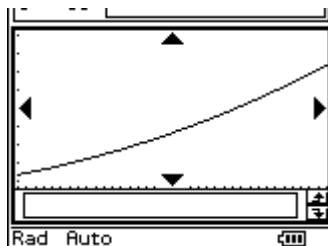
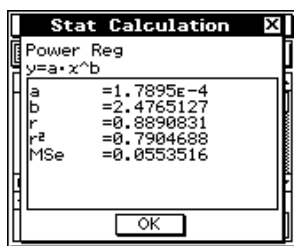


- 8) Con la ventana de gráficos activada, selecciona el comando \blacklozenge / Borrar todo.
 9) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana selecciona las siguientes opciones: Dibujo: On, Tipo=Rexp.ab, ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].

- 10) Toca el botón  para dibujar el gráfico de regresión exponencial, que responde a la función $y = a \cdot b^x$. Observa el resultado.



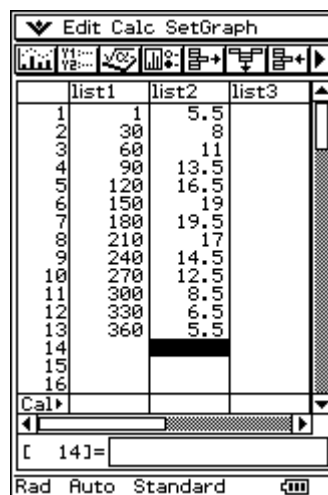
- 11) Con la ventana de gráficos activada, selecciona el comando \blacklozenge / Borrar todo.
 12) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. potencial. En la siguiente ventana selecciona las opciones ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].
 13) En la ventana Calc. Estadístico aparece la ecuación de la curva de regresión, que es de la forma $y = a \cdot x^b$ y el valor del coeficiente de determinación r^2 . Toca el botón [Acep.].
 14) Toca el botón  para dibujar la curva de regresión exponencial. Observa el resultado.





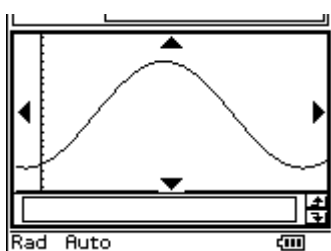
- **Gráfico de regresión sinusoidal**
- Calcula la curva de regresión correspondiente al número de horas de luz solar en Alaska durante un año, tal como se indica en la siguiente tabla. Sigue los siguientes pasos:

F	1	30	60	90	120	150	180	210	240	270	300	330	360
T	5,5	8	11	13,5	16,5	19	19,5	17	14,5	12,5	8,5	6,5	5,5

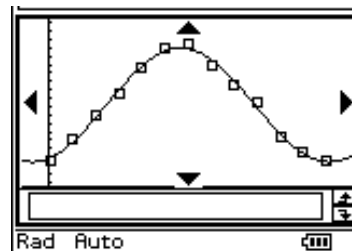
- 1) En el editor de listas, en las listas list1 y list2 introduce los valores de la tabla anterior.



- 2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrSin, ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Def.].
- 3) Toca el botón  para dibujar el gráfico de regresión sinusoidal.
- 4) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. Sinusoidal. En la siguiente ventana toca el botón [Acep.] y se mostrará una ventana con la ecuación de la curva de regresión, que es de la forma $y=a \cdot \sin(b \cdot x+c)+d$.
- 5) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana selecciona las opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec=1, Marca=cuadrado. Toca el botón [Def.].
- 6) Toca el botón  para visualizar simultáneamente el gráfico de regresión sinusoidal y la nube de puntos. Observa el resultado.



Stat Calculation	
Sinusoidal Reg	
$y=a \cdot \sin(b \cdot x+c)+d$	
a	=6.7757882
b	=0.016269
c	=-1.199583
d	=12.178001
MSe	=0.2133771
OK	

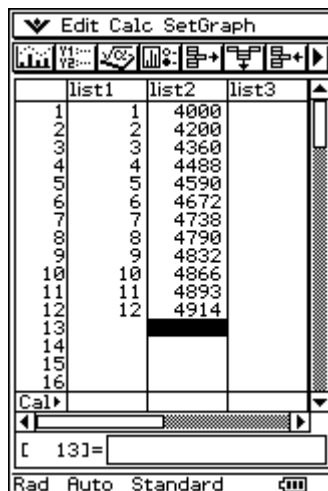


• **Gráfico de regresión logística**

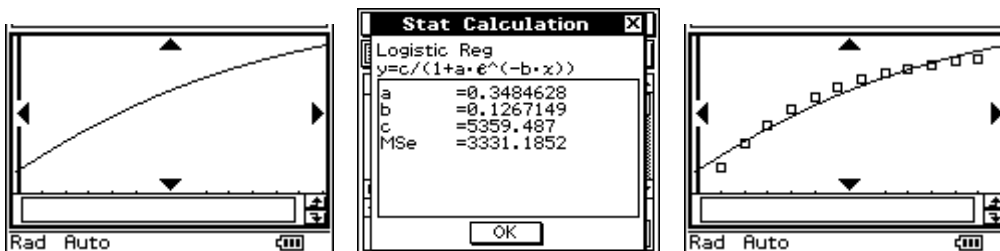
- El número de árboles de un bosque durante 12 años crece de la forma que se indica en la siguiente tabla. Calcula la curva de regresión logística. Sigue los siguientes pasos:

Año	1	2	3	4	5	6	7	8	9	10	11	12
Nº árboles	4000	4200	4360	4488	4590	4672	4738	4790	4832	4866	4893	4914

- 1) En el editor de listas, en las listas list1 y list2 introduce los valores de la tabla anterior.



- 2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrLogis, ListaX=list1, ListaY=list2. Toca el botón [Def.].
- 3) Toca el botón para dibujar el gráfico de regresión logística.
- 4) Con la ventana de gráficos activada, selecciona el comando Calc / Regresión logística. En la siguiente ventana toca el botón [Acep.] y se mostrará una ventana con la ecuación de la curva de regresión, que es de la forma $y = \frac{C}{1+a \cdot e^{-b \cdot x}}$.
- 5) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana selecciona las opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec=1, Marca=cuadrado. Toca el botón [Def.].
- 6) Toca el botón para visualizar simultáneamente el gráfico de regresión logística y la nube de puntos. Observa el resultado.







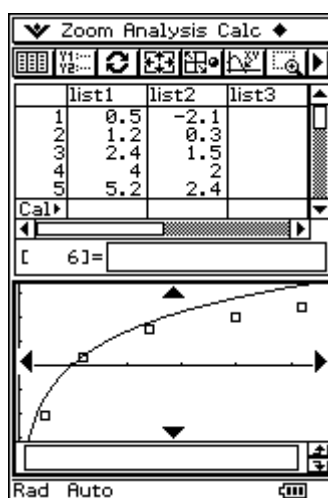
• **Gráfico de función y gráfico estadístico superpuestos**

- Dibuja el diagrama de dispersión de los siguientes datos y superpón dicho gráfico con el gráfico de la función $y=2 \cdot \ln(x)$. Sigue los siguientes pasos:

List1	0,5	1,2	2,4	4,0	5,2
List2	-2,1	0,3	1,5	2,0	2,4













- 1) En la ventana del editor de listas, introduce los datos de la tabla en las listas list1 y list2.
- 2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... y elige las siguientes opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec=1, Marca=cuadrado. Toca el botón [Def.].

- 3) Toca el botón  para dibujar la nube de puntos. Observa el resultado.
- 4) Toca la ventana de listas, para hacer que ésta sea la ventana activa, y luego toca el botón  para abrir la ventana del editor de gráficos.
- 5) Introduce la fórmula de la función $y=2 \cdot \ln(x)$ en la línea y1=.
- 6) Selecciona el comando  / Cerrar para cerrar la ventana del editor de gráficos.
- 7) Selecciona el comando ConfGraf / Función gráfica.
- 8) Toca el botón  para dibujar el gráfico de la función. De esta forma se visualiza simultáneamente el gráfico de la función y la nube de puntos.



6. BARRA DE HERRAMIENTAS DE GRÁFICOS ESTADÍSTICOS

También es posible utilizar los botones de la barra de herramientas de gráficos estadísticos. En la siguiente tabla tienes una explicación de las acciones de dichos botones:

Botón	Acción
	Muestra la ventana del editor de listas
	Muestra la ventana del editor de gráficos
	Vuelve a dibujar el gráfico visualizado
	Muestra el cuadro de diálogo de la ventana de visualización
	Alterna la opción Vent. estad. autom. entre automática y manual.
	Inicia una operación de trazo.
	Inicia la operación de zoom de cuadro.
	Amplía la imagen en pantalla
	Reduce la imagen en pantalla
	Muestra el cuadro de diálogo de configuración de gráficos estadísticos
	Muestra la ventana Principal.
	Muestra el administrador de variables.

7. CÁLCULOS ESTADÍSTICOS

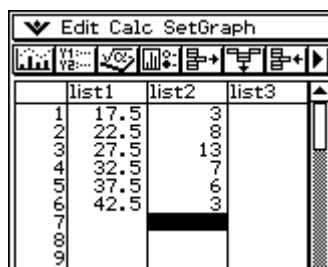
- **Cálculos estadísticos de una variable**

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

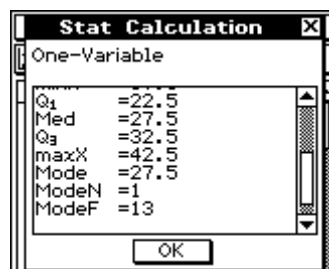
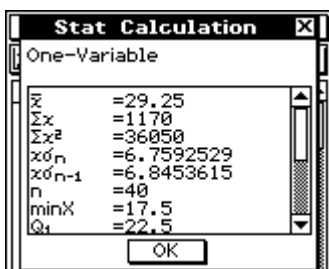
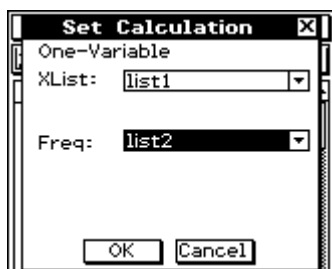
Puntuaciones	[15-20)	[20-25)	[25, 30)	[30, 35)	[35, 40)	[40, 45)
Nº de alumnos	3	8	13	7	6	3

Calcula los parámetros estadísticos: media, varianza, desviación típica, mínimo, máximo, primer cuartil, mediana, tercer cuartil, moda. Sigue los siguientes pasos:

- 1) En la ventana del editor de listas, introduce las marcas de clase en la lista list1 y las frecuencias en la list2.



- 2) En la barra de menús, selecciona el comando Calc / Una variable.
- 3) En el cuadro de diálogo que aparece, selecciona el nombre list1 como ListaX y selecciona el nombre list2 como Frec y toca el botón [Acep.].



- 4) Aparece una ventana con los valores de los parámetros estadísticos, que son los siguientes:

\bar{x}	Media	Q_1	Primer cuartil
Σx	Suma de datos	Med	Mediana
Σx^2	Suma de cuadrados	Q_3	Tercer cuartil
$x\sigma_n$	Desviación típica poblacional	MaxX	Máximo
$x\sigma_{n-1}$	Desviación típica muestral	Mode	Moda
N	Tamaño muestral	ModeN	Número de elementos iguales a la moda
MinX	mínimo	ModeF	Frecuencia de la moda

- **Cálculos estadísticos de dos variables**

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

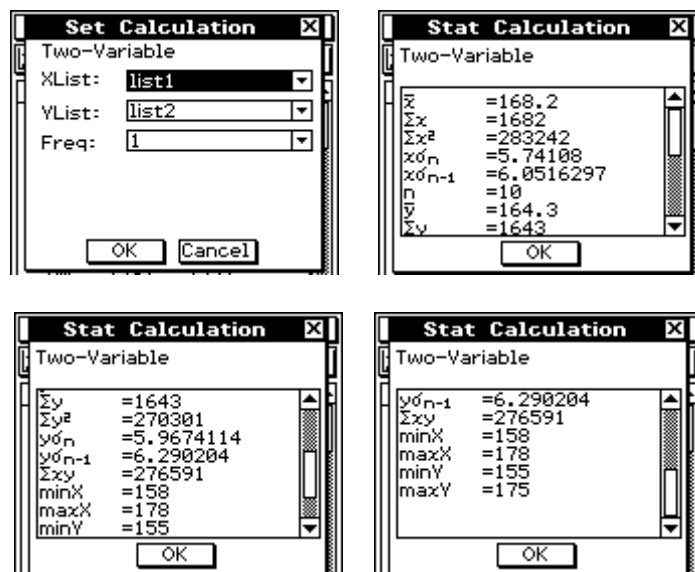
Hijas	158	162	164	165	168	169	172	172	174	178
Madres	163	155	160	161	164	158	175	169	166	172

Calcula los parámetros estadísticos de dichas variables: media, desviación típica, suma de productos, mínimo, máximo. Sigue los siguientes pasos:

- 1) En la ventana del editor de listas, introduce los datos de las chicas en la lista list1 y los datos de las madres en la list2.

	list1	list2	list3
1	158	163	
2	162	155	
3	164	160	
4	165	161	
5	168	164	
6	169	158	
7	172	175	
8	172	169	
9	174	166	
10	178	172	
11			
12			

- 2) En la barra de menús, selecciona el comando Calc / Dos variables.
- 3) En el cuadro de diálogo que aparece, selecciona el nombre list1 como ListaX, el nombre list2 como ListaY y Frec=1. Toca el botón [Acep.].



- 4) Aparece una ventana con los valores de los parámetros estadísticos, que son los siguientes:

\bar{x}	Media de la lista X	\bar{y}	Media de la lista Y
Σx	Suma de la lista X	Σy	Suma de la lista Y
Σx^2	Suma de cuadrados de la lista X	Σy^2	Suma de cuadrados de la lista Y
$x\sigma_n$	Desviación típica poblacional de X	$y\sigma_n$	Desviación típica poblacional de Y
$x\sigma_{n-1}$	Desviación típica muestral de X	$y\sigma_{n-1}$	Desviación típica muestral de Y
N	Tamaño muestral	Σxy	Suma de productos de las listas X e Y
MinX	Mínimo de la lista X	minY	Mínimo de la lista Y
MaxX	Máximo de la lista X	maxY	Máximo de la lista Y

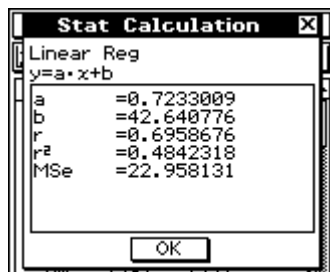
• **Cálculos estadísticos de regresión**

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

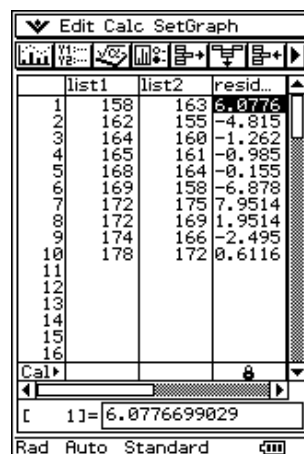
Hijas	158	162	164	165	168	169	172	172	174	178
Madres	163	155	160	161	164	158	175	169	166	172

Halla la recta de regresión de mínimos cuadrados, determina el coeficiente de correlación, analiza los residuos entre los datos reales y el modelo de mínimos cuadrados y estima la estatura de la madre cuando la estatura de la hija es de 167 cm. Investiga otros modelos de regresión. Sigue los siguientes pasos:


- 1) Con la ventana de listas activa, selecciona el comando Calc / Regresión lineal. En la siguiente ventana selecciona las opciones ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].
- 2) Aparece una ventana con la ecuación de la recta de regresión y el valor del coeficiente de correlación, así como el valor del coeficiente de determinación.

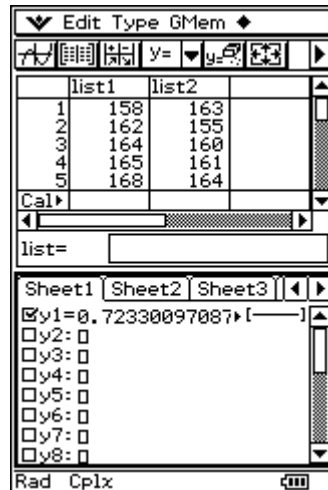




- 3) Como el coeficiente de determinación es pequeño, el modelo lineal es poco explicativo de los datos. Para analizar los residuos, selecciona el comando Calc. / Regresión lineal. En la siguiente ventana selecciona la opción Calc. Residual / On y toca el botón [Acep.].
- 4) Toca el botón [Acep.] de la siguiente ventana. Para ver los residuos, toca el cuadro de nombre de la list3. En la caja List= introduce el nombre "residual", utilizando el teclado virtual [abc]. Toca el botón [Ejec.].

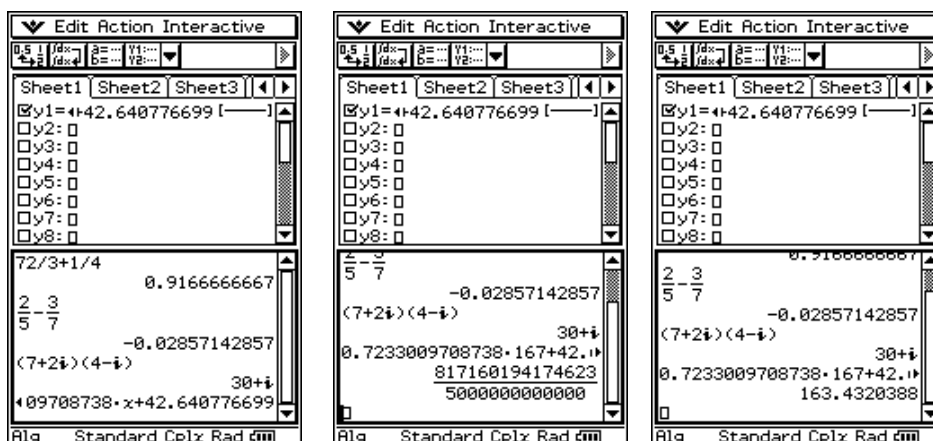


- 5) En la lista "resid..." se muestran las distancias entre los puntos reales y el modelo de regresión. Si la distancia es positiva, el punto está por encima del modelo de regresión; si es negativa, está por debajo del modelo.


- 6) Vamos a estimar el valor de list2 cuando list1=167. Para ello, selecciona el comando Calc / Regresión lineal. En el siguiente cuadro de diálogo selecciona de la lista desplegable Copiar fórmula la línea (y1 a y20) en la que quieres copiar la fórmula. Por ejemplo, la línea y1. Toca el botón [Acep.].
- 7) En la siguiente ventana toca el botón [Acep.]. Toca el botón  y comprueba que la ecuación de la recta de regresión se ha copiado en la línea Y1=.

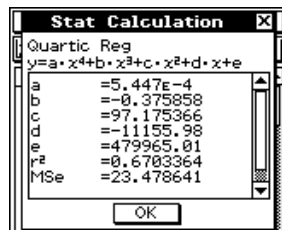



- 8) Para obtener la estatura de la madre correspondiente a una estatura de 167 cm para la hija, basta averiguar $Y1(167)$. Selecciona el cuadro de marcación de la función $Y1$ y desactiva todos los demás.
- 9) Toca el botón  para abrir la ventana Principal. Con el editor de gráficos y la ventana Principal en pantalla, selecciona la expresión a la derecha de $Y1=$ y arrástrala con el lápiz táctil hasta una entrada vacía de la ventana Principal. Observa que la ecuación de la recta de regresión aparece en la ventana principal. Selecciona la variable x y, en su lugar, con ayuda del teclado virtual, introduce la estatura de la hija, 167 cm, y toca el botón [Ejec.]. Selecciona la expresión anterior y toca el botón  para obtener el resultado en forma decimal.

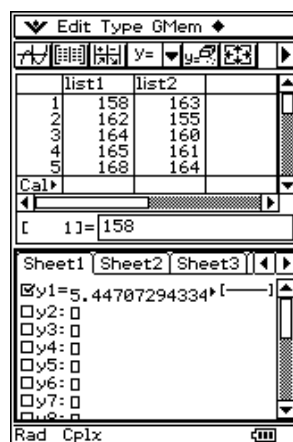




Por tanto, según el modelo lineal, la estatura de la madre, cuando la de la hija es de 167 cm, es de 163,4 cm. Pero esta estimación es poco fiable, debido al valor del coeficiente de determinación.

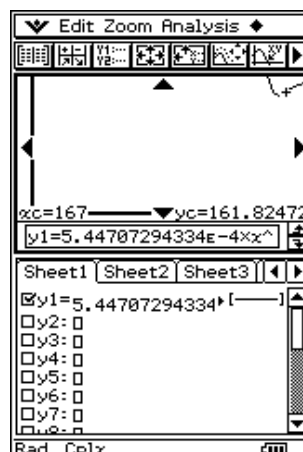
- 10) Cierra la ventana Principal y el editor de gráficos, con el comando  / Cerrar. Con la ventana de listas activa, selecciona el comando Calc. / Regr. cuarto orden. En el siguiente cuadro de diálogo selecciona la opción Copiar fórmula / y1. Toca el botón [Acep.].



- 11) Observa que el valor del coeficiente de determinación, r^2 , indica que el modelo de cuarto grado es mejor que el lineal. Toca el botón  y comprueba que la curva de regresión se ha copiado en la línea Y1=.



- 12) Toca el botón  para abrir la ventana de visualización. Introduce 180 como valores máximos para X y para Y. Toca el botón  para dibujar la función de regresión. Con la ventana de gráficos activa, selecciona el comando Análisis / Trazo. A continuación pulsa con el teclado 167. Aparece el cuadro de introducción de valores. Toca el botón [Acep.]. En pantalla se muestra el valor de la función de regresión, 161, 82472. Por tanto, si la estatura de la hija es de 167 cm, estimamos que la estatura de la madre será de, aproximadamente, 162 cm.



- Podemos ensayar otros modelos de regresión (med-med, cuadrática, cúbica, logarítmica, exponencial, potencial, sinusoidal y logística), como puedes ver en el menú Calc. cuando la ventana de listas está activa.

2. Actividades

- 1) Representa con el tipo de gráfico más adecuado los gastos de Sanidad en % del PIB de la Unión Europea. Utiliza para ello los datos de la siguiente tabla:

Países	1996	1998	2000
Alemania	8,9	9,3	7,9
Austria	6,5	7,4	7,8
Bélgica	7,6	8,1	8,1
Dinamarca	6,5	6,6	6,6
España	6,9	7,2	7,3
Finlandia	7,1	7,5	7,8
Francia	8,9	9,4	9,7
Grecia	4,2	4,4	5,5
Holanda	8,3	8,8	8,8
Irlanda	6,6	7,1	7,6
Luxemburgo	6,6	6,6	6,5
Portugal	6,0	6,9	6,9
Reino Unido	8,0	8,2	8,1
Suecia	8,0	9,3	7,9

- 2) En una fábrica se producen bombillas de tres tipos A, B y C. La duración de las mismas, en meses, viene expresada en la siguiente tabla. Calcula, para cada tipo de bombilla, los siguientes parámetros estadísticos: a) media, b) mediana, c) cuartiles, d) desviación típica. Dibuja el histograma correspondiente a cada tipo de bombilla.

Tipo A	Tipo B	Tipo C
15	25	20
25	35	30
35	45	40
45	55	50
55	65	60
65	75	70
75	85	80
85	95	90

- 3) Se piensa que los gastos en ocio de las familias están relacionados con su composición. Tomando una muestra de once familias se han obtenido los siguientes resultados:

Número miembros	2	3	3	4	4	2	2	6	3	2	3
Euros/mes	68,5	71,5	69,5	62,5	60	68	71	63	60	65	51

¿Qué gastos se han de esperar para una familia de 5 miembros?

- 4) La evolución del IPC (índice de precios al consumo) y de la tasa de inflación en los nueve primeros meses de 1997 fue:

IPC	0,7	1,1	1,7	2	1,9	1,9	2,9	2,9	3,8
Tasa de inflación	6	6	6,3	6,2	5,8	4,9	4,9	4,5	4.4

- Halla la media del IPC y de la tasa de inflación.
 - Halla las desviaciones típicas del IPC y de la tasa de inflación.
 - ¿Puedes predecir la tasa de inflación cuando el IPC fuese 3?
 - ¿Puedes predecir qué IPC habría para una tasa de inflación 5?
 - ¿Qué diferencia hay entre IPC y tasa de inflación?
- 5) En la siguiente tabla tienes los resultados de las elecciones al Congreso en Junio de 1977.
- Con estos resultados, ¿crees que hay correlación entre los votos emitidos y los escaños asignados a cada partido?
 - La ley d'Hont es la que permite ese proceso de distribución de escaños. Según la ley citada, y sin saber cómo se aplica, ¿sabrías prever cuántos escaños obtendría un partido que obtuviera en las elecciones del año 2008 el 42% de los votos emitidos?

Candidatura	Votos	Porcentaje válido	Diputados
UCD	6310391	34,44	166
PSOE	5371866	29,32	118
PCE	1709890	9,33	19
AP	1504771	8,21	16
PDPC	514647	2,81	11
PNV	296193	1,62	8
PSP-US	816582	4,46	6
UC-DCC	172791	0,94	6
EC-FED	143954	0,79	1
EE	61417	0,34	1
CAIC	37183	0,20	1
Independientes	29634	0,16	1

- 6) La siguiente tabla muestra las concentraciones de dióxido de azufre y humo, medidas en microgramos/m³ en una zona de tráfico denso de la ciudad de Valencia, durante cada mes de 2000

	Dióxido de azufre	Humo
Enero	26	44
Febrero	19	58
Marzo	22	49
Abril	20	31
Mayo	22	33
Junio	24	28
Julio	28	36
Agosto	19	38
Septiembre	24	50
Octubre	20	60
Noviembre	23	55
Diciembre	21	79

- Obtén los parámetros estadísticos correspondientes a las concentraciones de azufre y humo.
- Describe gráficamente los datos, mediante histogramas.
- Halla los percentiles de las dos distribuciones e interpreta los resultados.
- Dibuja el diagrama de dispersión y la recta de regresión que explique la concentración de humo en función de la concentración de dióxido de azufre.
- ¿Se puede estimar la concentración de humo cuando la de dióxido de azufre es de 30 microgramos/m³.

- 7) En la siguiente tabla se muestra la evolución de la tasa de paro durante el año 2000:

	Ciudad de Valencia	Comunidad Valenciana	España
Enero	14,4%	12,6%	12,9%
Febrero	15%	12,6%	12,8%
Marzo	14,7%	12,4%	12,6%
Abril	14,3%	12,2%	12,1%
Mayo	14,1%	12,1%	11,8%
Junio	13,9%	11,9%	11,5%
Julio	13,5%	11,4%	11,0%
Agosto	13,4%	11,7%	10,9%
Septiembre	13,3%	11,0%	11,1%
Octubre	13,3%	11,0%	11,1%
Noviembre	12,8%	10,5%	11,0%
Diciembre	12,0%	10,1%	10,9%

- a) Representa gráficamente las tres series temporales, utilizando un diagrama de líneas.
- b) Halla los parámetros estadísticos de cada serie y coméntalos.
- c) Dibuja el diagrama de dispersión y la recta de regresión cuando se estudian simultáneamente las tasas de paro en la Comunidad Valenciana y en la Ciudad de Valencia.
- d) Si la tasa de paro en la ciudad de Valencia es del 14,6%, ¿cuál es la tasa de paro en la Comunidad Valenciana?. ¿Es fiable la estimación?.
- e) ¿Se puede predecir la tasa de paro en la Comunidad Valenciana cuando la tasa de paro en España es del 11,3%?. ¿Por qué?.
- 8) El consumo de energía per cápita en miles de kw/h y la renta per cápita en miles de dólares en seis países de la Unión Europea son los siguientes:

	Consumo (Y)	Renta (X)
Alemania	5,7	11,1
Bélgica	5,0	8,5
Dinamarca	5,1	11,3
España	2,7	4,5
Francia	4,6	9,9
Italia	3,1	6,5

- a) Representa gráficamente los datos mediante un diagrama de puntos.
- b) Halla la recta de regresión del consumo (Y) sobre la renta (X).
- c) ¿Se puede predecir el gasto en consumo cuando la renta es de 12 miles de dólares?.

ESTADÍSTICA DESCRIPTIVA Y REGRESIÓN CON LA FX-9860G SD

Introducción

Vamos a estudiar a continuación las utilidades estadísticas de la nueva calculadora FX-9860G SD que incorpora una hoja de cálculo especialmente adecuada para el trabajo en Estadística.

1.- Estadística descriptiva

1. MEDIA Y DESVIACIÓN TÍPICA

- En un estudio genético, la comida regular se coloca en cada uno de 20 frascos y se anota el número de moscas de un genotipo particular que comen de cada frasco. Se cuenta también el número de moscas para otro conjunto de 20 frascos que contienen zumo de vino. Los datos recogidos son los siguientes:

Número de moscas (comida regular)									
15	20	31	16	22	22	23	33	38	28
25	20	21	23	29	26	40	20	19	31

Número de moscas (Zumo de vino)									
6	19	0	2	11	12	13	12	5	16
2	7	13	20	18	19	19	9	9	9

- Haz una comparación visual de las dispersiones respecto a sus centros de las dos distribuciones.
- Calcula la media y la desviación típica para cada conjunto de datos.

a) Abrimos el Editor de Hoja de calculo S-SHT. Introducimos la primera lista de datos en la columna A y la segunda lista en la columna B de la siguiente forma:

SHEE	A	B	C	D
1	15	6		
2	20	19		
3	31	0		
4	16	2		
5	22	11		
				15

FILE EDIT DEL INS CLR

SHEE	A	B	C	D
6	22	12		
7	23	13		
8	33	12		
9	38	5		
10	28	16		
				28

FILE EDIT DEL INS CLR

SHEE	A	B	C	D
11	25	2		
12	20	7		
13	21	13		
14	23	20		
15	29	18		
				29

FILE EDIT DEL INS CLR

SHEE	A	B	C	D
16	26	19		
17	40	19		
18	20	9		
19	19	9		
20	31	9		
				31

FILE EDIT DEL INS CLR

Para representar los datos gráficamente, accedemos al menú GRAPH. Asignamos el primer gráfico, GPH1 a la columna A y el segundo gráfico, GPH2 a la columna B y seleccionamos el diagrama de cajas [MedBox] como tipo de gráfico.

```

StatGraph1
Graph Type:MedBox
XCellRange:A1:A20
Frequency :1
Outliers :Off
    
```

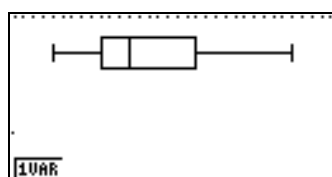
GPH1 GPH2 GPH3

```

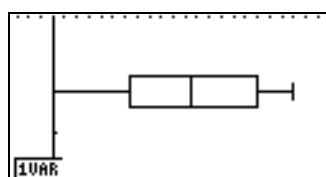
StatGraph2
Graph Type:MedBox
XCellRange:B1:B20
Frequency :1
Outliers :Off
    
```

GPH1 GPH2 GPH3

Los diagramas de caja para cada conjunto de datos son los siguientes:

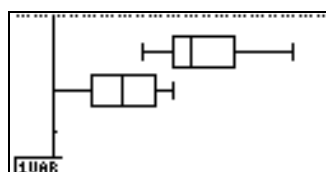
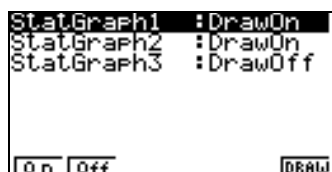


comida regular



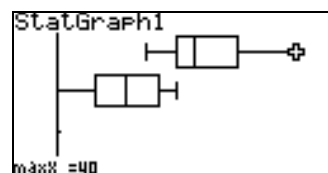
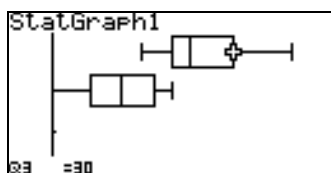
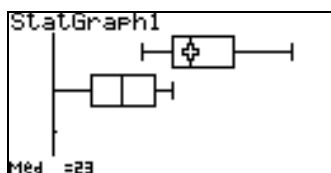
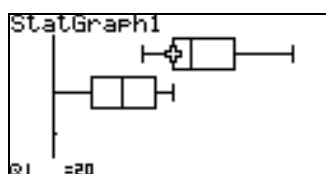
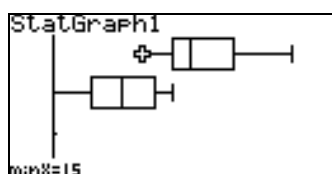
zumo de vino

Una comparación gráfica puede obtenerse al dibujar ambos gráficos en la misma pantalla. Elegimos SEL para seleccionar los dos gráficos, tal como sigue:

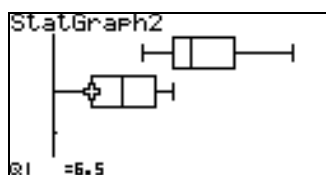
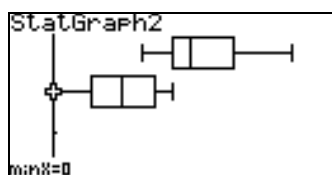


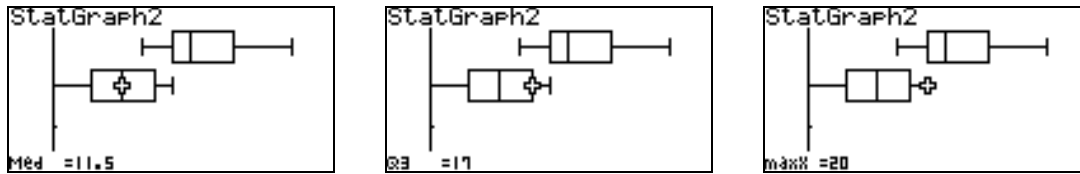
Observa que los centros de las medianas de ambas distribuciones son diferentes (están representadas por líneas verticales en el centro de las cajas). Los datos de las moscas asociadas al zumo de vino (gráfico de abajo) representa un conjunto más simétrico, ya que la línea que representa a la mediana está exactamente en el centro de la caja. Ambos conjuntos de datos parecen estar dispersos de la misma forma. Pulsando [SHIFT] [F1] (TRACE) podemos recorrer los diagramas de caja y ver las diferencias entre los parámetros.

Un diagrama de cajas es una representación gráfica de los datos que usa cinco medidas, la media, el primer y el tercer cuartil y el máximo y el mínimo de los datos. Los cuartiles dividen el conjunto de datos en cuatro partes iguales. El segundo cuartil es la mediana. Un diagrama de caja ayuda a visualizar el centro, la dispersión y la simetría del conjunto de datos. En este ejemplo, la construcción de los diagramas de caja permite comparar las dispersiones respecto al centro de las dos distribuciones. Observa que en el conjunto de moscas que comen la comida regular, el mínimo es 15 y el máximo es 40. La mediana es 23 moscas.

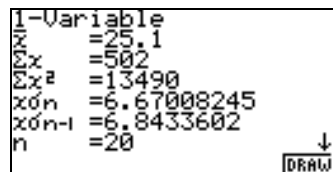


Por otra parte, para el conjunto de moscas que toman zumo de vino, el mínimo es 0 y el máximo es 20. La mediana es 11,5 moscas. Q1 y Q3 son los cuartiles.

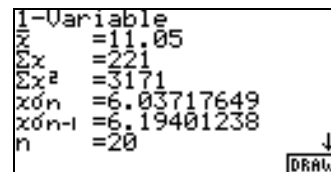




b) Pulsando [1Var] después de dibujar los diagramas de caja correspondientes a cada conjunto de datos, podemos ver los estadísticos univariantes asociados a los datos elegidos. Obtenemos los siguientes parámetros estadísticos correspondientes a cada conjunto de datos:



comida regular



zumo de vino

Observa que en el conjunto de moscas que comen comida regular, la media \bar{x} es 25.1 y la desviación típica $x\sigma_{n-1}$ es 6.8433602. En el conjunto de moscas que toman zumo de vino, la media \bar{x} es 11.05 y la desviación típica $x\sigma_{n-1}$ es 6.19401238.

Una observación que podemos hacer es que las desviaciones típicas de ambos conjuntos de datos difieren tan solo en 0.65. Una perspectiva gráfica puede ayudar también a hacer esta observación. En los diagramas de caja vemos que la dispersión de cada conjunto de datos es aproximadamente la misma.

Observamos también que la media del número de moscas que comen comida regular es mayor que la media del número de moscas que toman zumo de vino. Esta observación es también consistente con los diagramas de caja obtenidos.

- María inspecciona los precios para un cuarto de cierta marca de aceite de motor. Los datos, en dólares por cuarto, se resumen en la siguiente tabla:

Precio por cuarto	0.99	1.09	1.19	1.29	1.39	1.49
FRECUENCIA	2	3	7	10	14	4

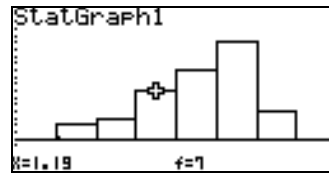
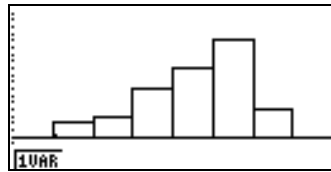
- Representa los datos gráficamente.
- ¿Cuál es la media y la desviación típica de los precios?

a) En este ejemplo, nuestros datos incluyen información del precio y frecuencia. Un histograma es un gráfico que puede resumir esta información. Los precios por cuarto se señalan en el eje horizontal y las frecuencias en el eje vertical.

En el editor de Hoja de cálculo, introducimos los precios en la columna A y las correspondientes frecuencias en la columna B. Para el gráfico de los datos, seleccionamos "histograma" [HIST]. En la siguiente pantalla indicamos el inicio del histograma y la anchura de cada intervalo de datos:



Si recorremos mediante [TRACE] el histograma de izquierda a derecha, para cada intervalo, la calculadora muestra el extremo de la izquierda como x y la frecuencia del intervalo como f , tal como se muestra en la siguiente pantalla:



- b) La media de los precios \bar{x} es aproximadamente \$1.2975 y la desviación típica de los precios σ_{n-1} es aproximadamente 0.13085027. La baja desviación típica nos dice que el precio no está demasiado disperso y se acerca a la media. Tenemos la siguiente pantalla:

```

1-Variable
x̄ = 1.2975
Σx = 51.9
Σx² = 68.008
x̄σn = 0.12920429
x̄σn-1 = 0.13085027
n = 40
    
```

- Para estudiar la composición de las familias de Winslow, Arizona, se seleccionaron al azar 40 matrimonios anotándose el número de niños por familia. Los datos obtenidos son los siguientes:

3	1	0	4	1	3	2	2	0	2	0	2	2	1
4	3	1	1	3	4	2	1	3	0	1	0	2	5
1	2	3	0	0	1	2	3	1	2	0	2		

- Construye un histograma para representar los datos.
 - Calcula la media de niños por familia.
 - Calcula la desviación típica del número de niños por familia.
- a) La frecuencia de cada uno de los datos es 1, porque cada dato representa el número de niños en una sola familia. Como los datos son enteros, ajustamos la configuración del histograma de forma que la anchura de cada barra sea 1.

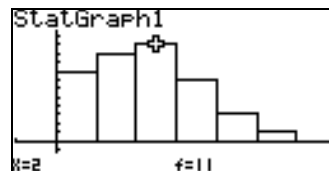
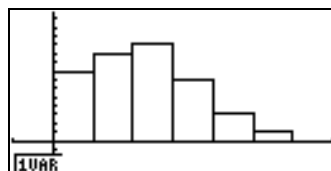
```

StatGraph1
Graph Type:Hist
XCellRange:A1:A40
Frequency:1
    
```

```

SH Histogram Settings
Start:0
Width:1
Draw:[EXE]
    
```

El histograma obtenido es el siguiente:



- El número medio de niños por familia es, aproximadamente, 1,75 o redondeando, 2 por familia.
- La desviación típica del número de niños por familia es alrededor de 1,32. Tenemos la siguiente pantalla de parámetros estadísticos:

```

1-Variable
x̄ = 1.75
Σx = 70
Σx² = 190
x̄σn = 1.2990381
x̄σn-1 = 1.31558702
n = 40
    
```


Nota: Las respuestas a los tres ejercicios anteriores se pueden obtener también introduciendo los datos en el menú STAT. Los gráficos y cálculos se obtienen usando comandos similares.

ACTIVIDADES:

1. Considera los siguientes datos:

LLUVIA MENSUAL EN SEATTLE, WASHINGTON											
JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
5.4	4.0	3.8	2.5	1.8	1.6	0.9	1.2	1.9	3.3	5.7	6.0

LLUVIA MENSUAL EN PHEONIX, ARIZONA											
JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
0.7	0.7	0.9	0.2	0.1	0.1	0.8	1.0	0.9	0.7	0.7	1.0

- a. Haz una comparación visual de las dispersiones respecto al centro de las dos distribuciones.
- b. Calcula la media y la desviación típica para cada conjunto de datos.

2. Los siguientes datos son los resultados de un examen en una clase de Estadística:

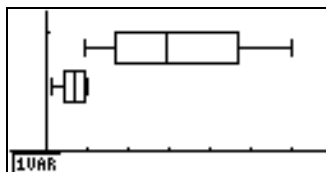
PUNTUACIONES EN EL TEST (en %)	99	74	85	93	80	70
NÚMERO DE ESTUDIANTES	2	3	7	10	14	4

¿Cuál es la media y la desviación típica de las puntuaciones del test?

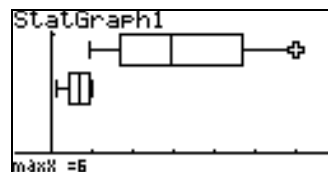
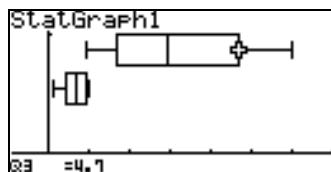
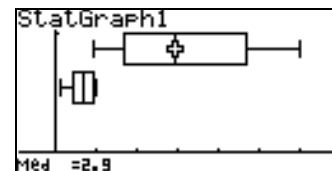
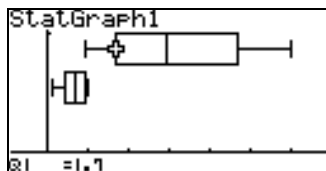
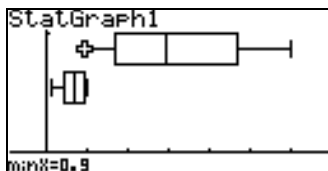
SOLUCIONES:

Actividad 1.

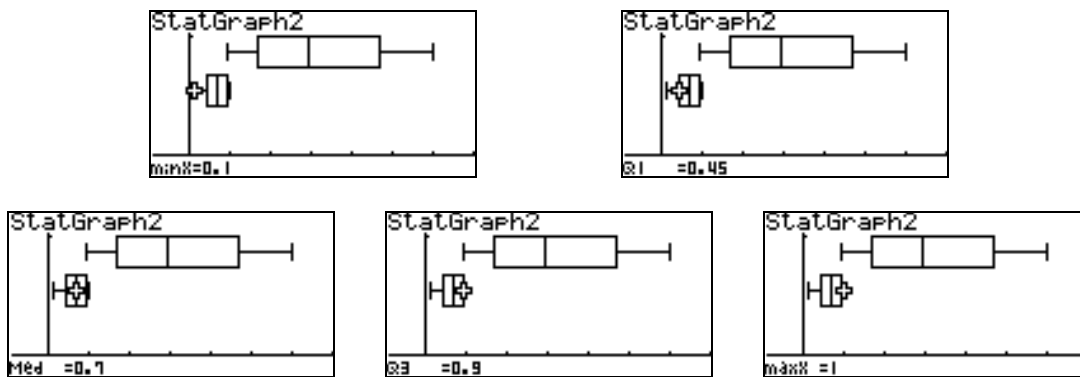
a) Los diagramas de caja son los de la siguiente figura. Elige SEL para que se muestren simultáneamente los dos gráficos en la pantalla:



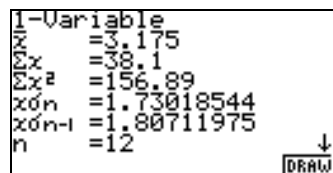
El diagrama de caja de arriba corresponde a la lluvia en Seattle. El mínimo y máximo valor son 0.9 y 6 respectivamente, con mediana 2.9.



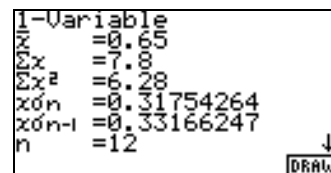
Por otra parte, el diagrama de caja de abajo corresponde a la lluvia en Arizona. Observa que el mínimo y el máximo valor son pequeños en comparación con los de Seattle, 0.1 y 1 respectivamente, con mediana 0.7.



b) Para los datos de lluvia en Seattle, la media \bar{x} es 3.175 y la desviación típica $x\sigma_{n-1}$ es 1.80711975. Por otro lado, para los datos de Arizona, la media \bar{x} es 0.65 y la desviación típica $x\sigma_{n-1}$ es 0.33. La lluvia en Seattle es más dispersa, tiene mayor desviación típica. Estos resultados se confirman por la forma de los diagramas de caja.



Seattle



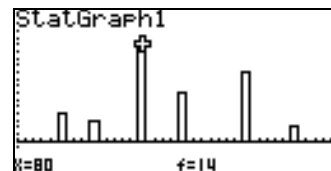
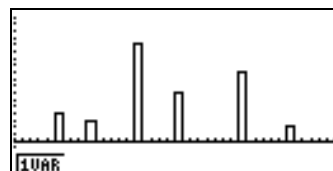
Arizona

Actividad 2.

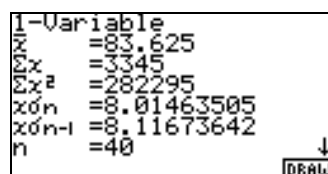
a) Tenemos la siguiente configuración para dibujar el histograma:



El histograma de las puntuaciones del test es el siguiente:



b) La media de las puntuaciones del test es aproximadamente 83.625, mientras que la desviación típica es aproximadamente 8.11673642.



2. USANDO LA HOJA DE CÁLCULO PARA HALLAR MEDIA Y DESVIACIÓN TÍPICA

- La media con la hoja de cálculo

¿Cómo se calcula la media? La media de un conjunto de datos se obtiene mediante la fórmula:

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

es decir, se suman todos los datos y el valor resultante se divide entre el número de datos.

- a) Cinco estudiantes miden el tiempo que utilizan en ir a la escuela por la mañana. Sus datos se muestran en la siguiente tabla, en la que se indican sus nombres (en la columna A) y sus tiempos (en la columna B). Para calcular la media, mueve el cursor hasta la celda C1.

SHEET	A	B	C	D
1	John	8		
2	Mary	17		
3	Joan	5		
4	Ann	26		
5	Carl	4		

Pulsa = para introducir la fórmula de la función. Al hacerlo se muestra una barra de menús como la siguiente:

SHEET	A	B	C	D
1	John	8		
2	Mary	17		
3	Joan	5		
4	Ann	26		
5	Carl	4		

=

GRAB	\$:	If	CEL	REL
------	----	---	----	-----	-----

Pulsa ahora (CEL) para abrir un menú con seis funciones de hoja de cálculo:

Min Max Mean Med Sum Prod

Pulsa (Mean). Teclea el rango de valores - B1 a B5 en nuestro caso - y pulsa I. Aparece el resultado.

SHEET	A	B	C	D
1	John	8	12	
2	Mary	17		
3	Joan	5		
4	Ann	26		
5	Carl	4		

=CellMean(B1:B5)

Observa que tecleando el nombre de la función y los parámetros "=CellMean(B1:B5)" obtienes un resultado idéntico, pero acabas con un mensaje de error. La función media debe ser seleccionada desde el menú.

- b) Incrementa uno de los valores con 1 unidad. ¿Qué le ocurre a la media? Cambiando diferentes valores en la columna B, ¿se obtienen diferentes efectos?
- c) Incrementa todos los valores 1 unidad. ¿Qué le ocurre a la media?
- d) La media de un conjunto de valores es 63. Incrementa un valor en 5. Disminuye otro valor en 5. Predice el valor de la media.

- e) Pregunta el número de calzado de los alumnos de tu clase. Teclea sus respuestas en una hoja de cálculo. Calcula el número medio de calzado de tu clase.
- f) Usando una balanza, pesa a los estudiantes de tu clase. Puedes introducir sus pesos en una hoja de cálculo. Calcula la media.
- g) Selecciona todos los estudiantes con los zapatos del mismo número de calzado. Pesa sus zapatos. Calcula su media. Comprueba que si el tamaño de los zapatos es exactamente igual a la media de los tamaños, su peso medio es exactamente igual a la media de los pesos de todos los zapatos. Si la diferencia de tamaños respecto a la media es la más grande, cabe esperar que también lo sea la diferencia en pesos.

• La desviación típica con la hoja de cálculo

Observa la siguiente figura. Los dos conjuntos de datos tienen la misma media.

SHEET	A	B	C	D
1	8	12		
2	17			
3	5			
4	26			
5	4			

=CellMean(A1:A5)

SHEET	A	B	C	D
1	14	12		
2	10			
3	11			
4	13			
5	12			

=CellMean(A1:A5)

Sin embargo, los valores de la columna A a la izquierda son más dispersos respecto de la media que los de la derecha. Expresamos este hecho mediante la desviación típica que mide la dispersión de los datos respecto de la media. Se calcula con la fórmula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

- a) Para calcular la desviación típica del conjunto de datos de la derecha con la hoja de cálculo, sigue los siguientes pasos:
 1. Calcula la media en la celda B1.
 2. Calcula la diferencia entre los valores y la media (en la celda C1). Después copia la fórmula de C1 a la columna C, tomando la referencia a B1 como referencia absoluta \$B\$1
 3. Calcula los cuadrados de los valores de la columna C en la columna D.
 4. Suma los cuadrados (pon el resultado en la celda E1).
 5. Divide el resultado por el número de elementos (en la celda F1).
 6. Calcula la raíz cuadrado del resultado anterior (en la celda G1).

SHEET	A	B	C	D	E	F	G
1	14	12	2	4	10	2	1.4142
2	10		-2	4			
3	11		-1	1			
4	13		1	1			
5	12		0	0			

La desviación típica de dicho conjunto de datos es 1.4142.

- b) Calcula la media y la desviación típica para los datos de la izquierda. Observa que, aunque la media es la misma, la desviación típica es diferente:

SHEE	A	B	C	D	E	F	G
1	8	12	-4	16	350	70	8.3666
2	17		5	25			
3	5		-7	49			
4	26		14	196			
5	4		-8	64			

La desviación típica, 8.3666, es ahora mucho mayor, porque los datos están más dispersos.

- c) Los cálculos anteriores muestran el algoritmo completo para obtener la desviación típica. Pero no necesitamos realizarlos con tanto detalle. La hoja de cálculo dispone de una función que permite el cálculo directo.

Por ejemplo, supongamos que hemos introducido los datos en la columna A.

SHEE	A	B	C	D
1	8			
2	17			
3	5			
4	26			
5	4			

8

FILE EDIT DEL INS CLR D

Pulsa [F6] (para que aparezca más opciones del menú desplegable). En dicho menú, pulsa CALC para mostrar su submenú. En el submenú selecciona 1VAR. La nueva ventana obtenida es la siguiente:

```

1-Variable
x̄ = 12
Σx = 60
Σx² = 1070
x̄σn = 8.36660026
x̄σn-1 = 9.35414346
n = 5
    
```

En ella podemos observar los valores de los distintos parámetros estadísticos.

- d) Halla la desviación típica del conjunto anterior en la celda G1. Selecciona al azar dos celdas en la columna A. Aumenta un valor en 7 unidades. Disminuye el otro valor en 7 unidades. La media queda igual. ¿Qué ocurre con la desviación típica?
- e) Se divide la clase en cuatro grupos de estudiantes. Cada grupo puede cambiar una par diferente de celdas:
1. Aumenta A1 en 4 unidades; disminuye A4 en 4 unidades.
 2. Disminuye A1 en 4 unidades; aumenta A4 en 4 unidades.
 3. Aumenta A2 en 6 unidades; disminuye A3 en 6 unidades.
 4. Disminuye A2 en 6 unidades; aumenta A3 en 6 unidades.

Observa que todos los pares de operaciones preservan la media.

Cada grupo referirá a los otros lo que ocurre con la desviación típica: ¿Aumenta? ¿Disminuye?

Cuando los valores cambiados están cerca de la media, la desviación típica disminuye. Cuando los valores están lejos de la media, la desviación típica aumenta.

2.- Análisis de regresión

1. REGRESIÓN LINEAL I

Considera los datos de la siguiente tabla que representan los valores de las acciones de la compañía *Vanguard Index Trust* desde 1987 a 1997.

- Llamamos x =años e y =valor de las acciones de Vanguard Index Trust. Dibuja un diagrama de dispersión para estos datos.
- Calcula la pendiente de la recta que pasa por los dos puntos que representan el valor de la acción en 1987 y en 1991. Haz lo mismo para los puntos que representan los datos en 1991 y en 1995.
- ¿Cuál de las pendientes calculadas en (b) es mayor en valor absoluto? ¿Qué significa esto?
- Halla la recta que mejor se ajusta a los datos. ¿Cuál es esta recta? Interpreta la pendiente de esta recta.
- Si tu fueras un dirigente de esta empresa, ¿cuál de las tres pendientes que has calculado utilizarías para convencer a alguien para invertir?
- ¿Cuál es la tendencia en los datos?
- Suponiendo que la tendencia continua, ¿cuál será el valor de la acción en 2006?

AÑO	Valor (dólares)
1987	54.26
1988	63.07
1989	82.81
1990	80.08
1991	103.27
1992	113.20
1993	123.11
1994	130.29
1995	170.32
1996	185.81
1997	201.40

Solución:

Abrimos el Editor de Hoja de cálculo e introducimos los datos. Introducimos los años en la primera columna y los valores de las acciones en la segunda columna, tal como se indica en la figura:

SHEE	A	B	C	D
1	1987	54.26		
2	1988	63.07		
3	1989	82.81		
4	1990	80.08		
5	1991	103.27		

FILE EDIT DEL INS CLR ▾

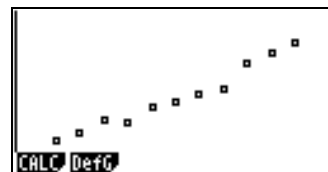
SHEE	A	B	C	D
6	1992	113.2		
7	1993	123.11		
8	1994	130.29		
9	1995	170.32		
10	1996	185.81		

FILE EDIT DEL INS CLR ▾

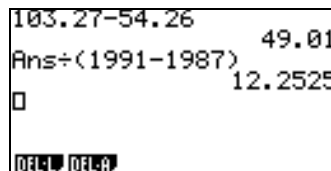
SHEE	A	B	C	D
10	1996	185.81		
11	1997	201.4		
12				
13				
14				

FILE EDIT DEL INS CLR ▾

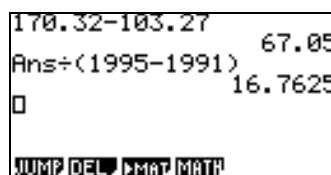
- Pulsamos GRPH y SET para configurar los rangos de celdas del diagrama de puntos:



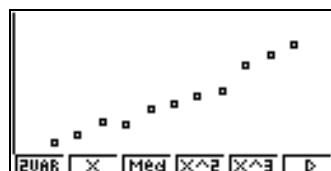
- b) Calculamos las pendientes de las rectas en el Editor RUN. La pendiente de la recta que une los puntos (1987,54.26) y (1991,103.27) se calculan así:



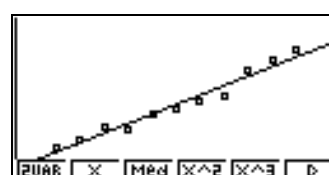
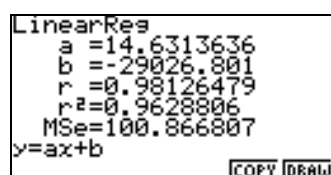
La pendiente de la recta es 12.2525. Por otro lado, la recta que une los puntos (1991, 103.27) y (1995,170.32) tiene pendiente 16.7625 y se calcula así:



- c) La pendiente de la recta que une los puntos (1991, 103.27) y (1995,170.32) es mayor en valor absoluto que la pendiente de la recta que une (1987,54.26) y (1991,103.27). Esto significa que entre los años 1991 y 1995, el valor de cada una de las acciones aumentó alrededor de 16.7625 dólares por término medio. Este valor es superior en 4.51 al comprendido entre los años 1987 y 1991, en los cuales el valor de cada acción aumentó alrededor de 12.2525 dólares por término medio).
- d) Para hacer la recta de mejor ajuste, introducimos el comando Linear regression mientras el gráfico de dispersión se muestra en pantalla:



La recta de mejor ajuste es $y = 14.6313636x - 29026.801$. La recta se dibuja en el diagrama de puntos, tal como vemos en la siguiente figura:

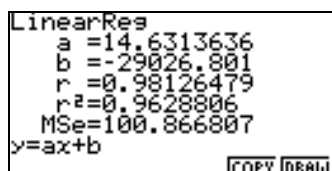


La pendiente de la recta de mejor ajuste puede interpretarse como el incremento medio del valor de la acción para un año; en nuestro caso este incremento es de 14.6313636 dollars, aproximadamente.

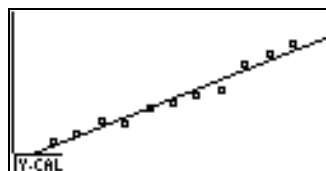
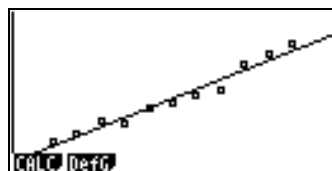
¿Cuál es la bondad del ajuste? La respuesta se puede obtener calculando el coeficiente de correlación r que es la medida de la intensidad de la relación lineal que existe entre dos variables. Cuanto más próximo a 1 sea el valor de $|r|$ más perfecta es la relación lineal entre las variables. En la pantalla obtenida anteriormente, hemos visto que r es 0.98126479, lo que indica que el ajuste lineal entre las dos variables es bueno.

- e) Puede ser más razonable usar la pendiente 14.6313636 de la recta de mejor ajuste, la cual indica el crecimiento en el valor de la acción para un incremento de x de un año y muestra la tendencia de los datos.
- f) La tendencia de los datos es que el valor de la acción aumenta cuando x aumenta.
- g) Suponiendo que la tendencia continua, podemos usar la recta de mejor ajuste para determinar el valor de la acción en 2006.

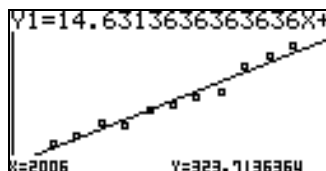
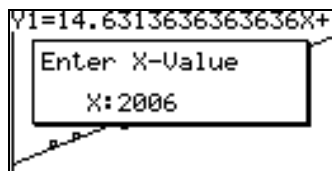
Primero, copiamos los cálculos que hemos obtenido respecto de la recta de mejor ajuste en el menú gráfico con el Editor de Hoja de cálculo.



A continuación dibujamos el gráfico (en la pantalla de la izquierda en la siguiente figura). Después, pulsamos [SHIFT] (G-Solv) y [Y-CAL], lo que nos permite aproximar el valor de la acción sustituyendo el año 1997 usando la recta de mejor ajuste (pantalla de la derecha en la figura).

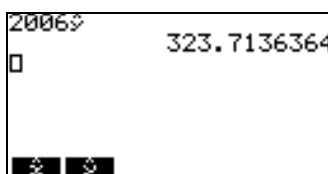


Podemos indicar el valor de x, por ejemplo, x=2006, y la calculadora muestra el correspondiente valor de y en el gráfico.



En el año 2006, el valor aproximado de la acción es de \$323.71.

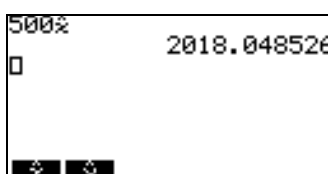
El valor de y correspondiente a x = 2006 usando la recta de mejor ajuste puede también obtenerse en el Editor RUN, usando el comando OPTN STAT como se muestra a continuación:



En el Editor RUN, también puede obtener el valor de x correspondiente a un valor dado de y. Por ejemplo, el apartado (g) que hemos visto puede sustituirse por el siguiente:

- h) Suponiendo que la tendencia continua, ¿cuando el valor de la acción valdrá aproximadamente \$500?

En el Editor RUN, obtenemos el siguiente cálculo:



Usando la recta de regresión, estimamos que el valor de la acción puede valer alrededor de \$500 en el año 2018.

Observaciones: Observa que los cálculos en el Editor RUN pertenecen a los datos guardados en el Editor de Hoja de cálculo (como los cálculos que hemos visto en las dos pantallas anteriores), pero no puede ser utilizado en el modo e-activity. Este puede ser utilizado solamente cuando usamos el Editor de Hoja de cálculo del menú principal.

ACTIVIDAD

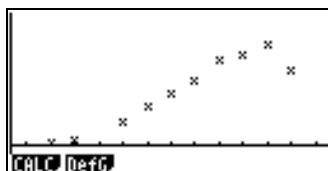
Los siguientes datos indican el número de personas muertas a causa del SIDA en Estados Unidos desde 1982 hasta 1992.

años	1982	1983	1985	1986	1987	1988	1989	1990	1991	1992
Muertos	843	1651	6681	11535	15451	19656	26151	28053	30579	22660

- a) Basándote en estos datos, compara el número de muertos interpolado en 1990 con el dato real 28053.
- b) Usa el modelo de regresión lineal para estimar el número de muertos a causa del SIDA en el año 2006.
- c) Si la tendencia continua, ¿es posible que los muertos a causa del SIDA lleguen a 200000? ¿En qué año?

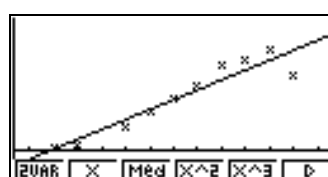
Solución:

- a) En el Editor de Hoja de cálculo, introducimos los datos y dibujamos el diagrama de dispersión.



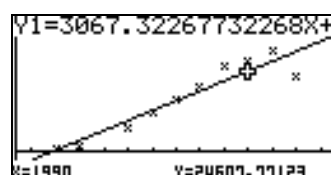
Entonces la recta de mejor ajuste es la siguiente, dibujada sobre el diagrama de dispersión:

```
LinearReg
a =3067.32267
b =-6.079E+06
r =0.94248765
r^2=0.88828297
MSe=1.4805E+07
y=ax+b
```

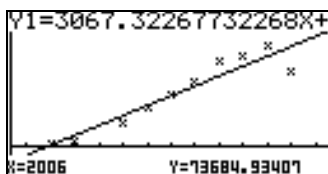


La recta de regresión es $y = 3067.32267x - 6079000$. Usando la recta de regresión, el número interpolado de muertos en 1990 es 24,607.77. El número real de muertos es 28053. Una diferencia de aproximadamente 3445. Observa que la recta que hemos obtenido tiene coeficiente de correlación, $r = 0.94$, lo que sugiere una correlación positiva alta; no es perfecta, esperamos que exista alguna diferencia entre los datos interpolados y los valores reales.

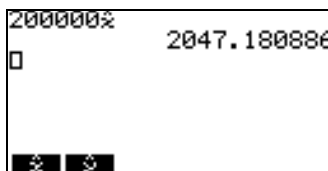
```
Y1=3067.32267732268X+
Enter X-Value
X:1990
```



- b) El número estimado de muertos en 2006 es alrededor de 73,685.



c) En el Editor RUN, obtenemos aproximadamente que en 2047, los muertos serán alrededor de 200,000.



2. REGRESIÓN LINEAL II

- Las pepitas de oro

Hay un río cerca del campus. La mina de oro cercana está agotada desde hace algunos años, pero algunos buscadores de oro ocasionales refinan el agua del río para buscar pepitas de oro. Los estudiantes han creado un Club de Buscadores de Oro como una especie de entretenimiento que puede también ayudarles a recoger fondos para la fiesta de Navidad. Van en fines de semana, refinan la arena del río y consiguen records en sus hazañas (ver la tabla de la hoja de cálculo).

SHEET	A	B	C	D
1	Name	Days	Gold	
2	Ann	14	28	
3	Bob	35	66	
4	David	22	38	
5	Fiona	29	70	
6	Frank	6	22	
7	Kathy	15	27	
8	Mick	17	28	
9	Sarah	20	47	
10	Tim	12	14	
11	Wena	29	68	

Como vemos en la tabla, no todos los miembros del club son igual de activos. Fiona - la presidenta del club - opina que hay una relación entre el número de visitas al río y la cantidad de oro recolectado por cada persona. Quiere demostrarlo a los otros. Ha estudiado un curso de Estadística y quiere aprovechar su conocimiento del método denominado Regresión Lineal.

a) Fiona introduce los datos en una hoja de calculo de su calculadora. Primero muestra a los otros que los datos pueden representarse en forma de nube de puntos. Para hacerlo, selecciona [GRAPH]. Del submenú selecciona [SET].



Los días consumidos en el río forman la variable independiente x (las celdas B2 a B11), la cuenta de pepitas de oro forma la variable dependiente y (las celdas C2 a C11). Regresando al submenú gráfico y seleccionado q - el gráfico aparece en pantalla.



La gráfica consiste en puntos aislados, uno por cada buscador de oro. Si la forma de la nube no es complicada, tiene una tendencia: para valores muy grandes de x, los valores de y tienden a ser muy grandes.

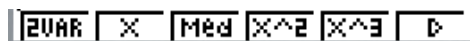
Esta tendencia puede expresarse por una recta de regresión, cuya ecuación general es $y = ax + b$ donde y representa la cuenta aproximada de pepitas encontradas en x días. Los parámetros a y b se calculan desde la tabla de valores usando las fórmulas:

$$a = \frac{n \left(\sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2}$$

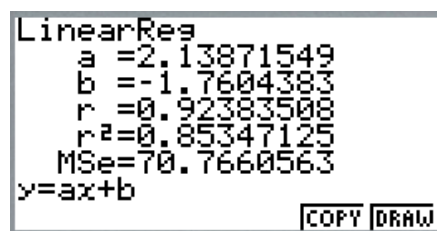
$$b = \frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n}$$

En general, hemos de considerar todos los n pares de elementos de los dos conjuntos $x_1, x_2, x_3, \dots, x_n$ (número de visitas al río por persona) y $y_1, y_2, y_3, \dots, y_n$ (número de pepitas de oro encontradas por cada individuo). Combinando los números y usando las fórmulas, obtenemos el resultado.

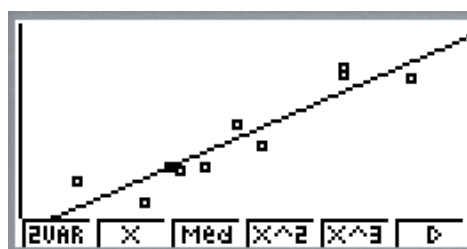
Fiona no se asusta de estas espantosas fórmulas, porque sabe que la calculadora puede obtener a y b directamente. Para hacerlo, pulsa [CALC]. El nuevo menú aparece en la parte inferior de la pantalla:



Después pulsa w (x), la calculadora muestra el siguiente resultado:



La función que permite aproximar el número de pepitas de oro después de x días está definida como $y = 2.13871549x - 1.7604383$. Para ver su gráfica, pulsa [DRAW].



ACTIVIDAD

¿Cuántas pepitas podemos esperar encontrar después de 30 días de búsqueda?

Solución:

Como hemos hecho una estimación, no necesitamos cálculos muy exactos. Dos cifras decimales son suficientes:

$$(2.14 \times 30) - 1.76 = 62.4$$

Esperamos encontrar alrededor de 62 pepitas.

• **El fertilizante**

Durante los últimos años, un granjero está utilizando un fertilizante en sus campos. Mantiene registros de la cantidad de fertilizante usado (en toneladas) y producción (en toneladas de cosecha). La tabla muestra estos registros.

SHEET	A	B	C	D
1	Year	Fert	Crop	
2	2000	2	71	
3	2001	1	36	
4	2002	3	79	
5	2003	4	104	
6	2004	2.5	73	

- a) Usando regresión lineal, muestra la dependencia entre las cantidades de fertilizante y las cosechas. Dibuja el diagrama de puntos.
- b) Haz el cálculo correspondiente. Registra los parámetros a y b.
- c) Dibuja la recta de regresión.
- d) El granjero ha aceptado una oferta de un comprador a granel de 80 toneladas de productos de la última cosecha. ¿Cuántas toneladas de fertilizante es recomendable que compre el granjero?

Solución:

El parámetro a vale 21.2, b vale 19.6. Hemos de resolver la ecuación $80 = 21.2x + 19.6$. El resultado es 2.84 (aproximadamente 3 toneladas).

• **Resistencia deportiva**

En un acontecimiento deportivo, un médico del equipo mide el tiempo empleado por los deportistas de diferente edad en una carrera. Todos los deportistas empiezan en el mismo momento. El médico del equipo anota el tiempo que tarda cada persona en pararse por estar exhausto. Los registros son los de la siguiente hoja de cálculo.

SHEET	A	B	C	D
1	Age	Time		
2	34	17.5		
3	21	20		
4	27	19.8		
5	18	22.9		
6	42	14.5		
7	33	16		
8	28	20		
9	50	12.2		
10	36	18.3		
11	44	13.2		

- a) Dibuja un diagrama de dispersión basado en estos datos. Haz el cálculo de los parámetros de la recta de regresión y dibújala.

- b) ¿Cuánto tiempo cabe esperar que resista una persona de 40 años?
- c) ¿Qué edad aproximada corresponde a un tiempo de resistencia de 19 minutos?
- d) ¿Por qué la recta de regresión es decreciente?

Solución:

a) El parámetro a es aproximadamente igual a -0.32; b es aproximadamente igual a 28.15:

$$y = -0.32x + 28.15$$

$$y = 15.35$$

- b) La persona de 40 años puede resistir alrededor de 15 minutos.
- c) Como a = -0.32 y b = 28.15, hemos de resolver la ecuación: $19 = -0.32x + 28.15$, cuya solución es $x = 28.125$. La persona capaz de resistir 19 minutos tiene, aproximadamente, 28 años de edad.
- d) La duración de la carrera decrece con la edad. Por ello, la recta de regresión es decreciente.

3. COEFICIENTE DE CORRELACIÓN I

Las pepitas de oro 2

Recuerda los datos del problema "Las pepitas de oro"

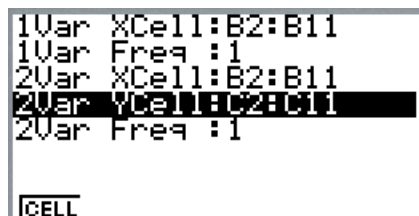
SHEE	A	B	C	D
1	Nombre	Days	Gold	
2	Ann	14	28	
3	Bob	35	66	
4	David	22	38	
5	Fiona	29	70	
6	Frank	6	22	
7	Kathy	15	27	
8	Mick	17	28	
9	Sarah	20	47	
10	Tim	12	14	
11	Wena	29	68	

¿Por qué Fiona confía en la existencia de una relación entre las dos variables? Porque conoce la forma de obtener el coeficiente de correlación con su calculadora.

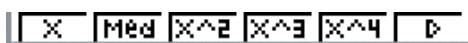
Para ver la relación entre dos conjuntos de números, los datos deben ser introducidos en una tabla de hoja de cálculo de la calculadora. Para ello, selecciona [CALC]. En el submenú CALC, primero selecciona [SET] para especificar el rango de valores



Selecciona el rango B2 a B11 para la x; y el rango C2 a C11 para la y. Sin especificar ambos rangos, la calculadora mostraría un mensaje de error o produciría resultados incorrectos.



Regresa a la hoja seleccionando [EXE]. Selecciona [CALC] [REG]. El siguiente submenú ofrece una variedad de métodos de regresión. Elegimos el más sencillo, la regresión lineal pulsando [X].



El cálculo es ejecutado.

```

LinearReg
a =2.13871549
b =-1.7604383
r =0.92383508
r²=0.85347125
MSe=70.7660563
y=ax+b
COPY
    
```

Observa el valor de r obtenido. Este valor se llama coeficiente de correlación y su fórmula es verdaderamente compleja:

$$r = \frac{n \left(\sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2 \right]}}$$

Por tanto, hemos de considerar todos los n pares de elementos de los dos conjuntos $x_1, x_2, x_3, \dots, x_n$ (el número de visitas al río por persona) y $y_1, y_2, y_3, \dots, y_n$ (el número de pepitas de oro encontradas por cada individuo). Sustituyendo los datos en la fórmula anterior obtenemos el resultado. Afortunadamente, la calculadora permite su obtención, siempre que especifiquemos los rangos de x y de y. El resultado es siempre un número real entre -1 y +1 (0.92383508 en nuestro caso). Los valores -1 y +1 (y valores cercanos a los extremos del intervalo) indican una relación entre los conjuntos de datos comparados. Los valores del centro del intervalo indican que no hay una relación (con valores 0 o cercanos a 0).

• **El fertilizante 2**

Recuerda los datos del problema "El fertilizante". Se muestran en la siguiente tabla.

SHEE	A	B	C	D
1	Year	Fert	Crop	
2	2000	2	71	
3	2001	1	36	
4	2002	3	79	
5	2003	4	104	
6	2004	2.5	73	

Calcula el coeficiente de correlación y utiliza el resultado para extraer una conclusión sobre la fiabilidad de la estimación para el presente año (evaluando los parámetros a y b de la recta de regresión). Observa que la especificación del rango es fundamental porque tenemos tres columnas de números. Nuestros rangos a seleccionar son la cantidad de fertilizante (B2 a B6) y la cosecha (C2 a C6).

• **Resistencia deportiva 2**

Recuerda los datos del problema "Resistencia deportiva"

SHEE	A	B	C	D
1	Age	Time		
2	34	17.5		
3	21	20		
4	27	19.8		
5	18	22.9		
6	42	14.5		
7	33	16		
8	28	20		
9	50	12.2		
10	36	18.3		
11	44	13.2		

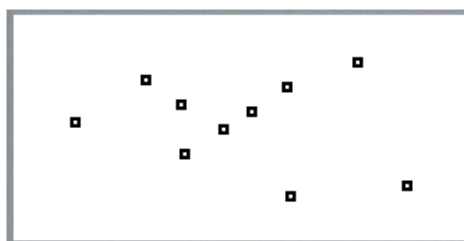
Calcula el coeficiente de correlación entre los dos conjuntos de datos. ¿Podemos utilizar la regresión lineal para hacer estimaciones en este caso?

• **Estatura y televisión**

Un grupo de estudiantes recoge datos sobre su estatura (en metros) y la cantidad de horas de ven la televisión diariamente. Están interesados en ver si hay una relación entre estos dos conjuntos de datos.

SHEE	A	B	C	D
1	Name	Height	Time	
2	Beata	1.42	2.5	
3	Ed	2.02	0.5	
4	Hugo	1.74	2.75	
5	Lea	1.62	1.5	
6	Nick	1.8	3.5	
7	Nora	1.81	0.25	
8	Raj	1.69	2.25	
9	Tanya	1.93	4.25	
10	Tibor	1.55	3.75	
11	Wanda	1.61	3	

El coeficiente de correlación $r = -0.2367005$ indica que no hay una relación importante. Lo mismo podemos observar en el diagrama de dispersión.



En este caso se dice que no hay correlación entre las variables. En general esto ocurre cuando el coeficiente de correlación r toma valores cercanos a 0.

4. COEFICIENTE DE CORRELACIÓN II

• **Las granjas en EEUU**

Los siguientes datos indican el número de granjas en los Estados Unidos durante los años 1910 a 1999:

Años	1910	1920	1930	1940	1950	1959	1969	1978	1987	1999
Número de granjas (en millones)	6.4	6.5	6.3	6.1	5.4	3.7	2.7	2.3	2.1	1.9

- Dibuja un diagrama de dispersión de los datos.
- Determina la función que mejor se ajusta a los datos.
- Usa la respuesta del apartado (b) para estimar el número de granjas en 1900 y en 1975.

Solución:

- Accedemos al Editor de Hoja de cálculo. Tomamos como coordenada x de cada punto el número de años transcurridos desde 1900 y como coordenada y el número de granjas. Los años los introducimos en la columna A (1910 se introduce como 10, 1920 como 20 y así sucesivamente) y el número de granjas en millones lo introducimos en la columna B:

SHEE	A	B	C	D
1	10	6.4		
2	20	6.5		
3	30	6.3		
4	40	6.1		
5	50	5.4		

10
FILE EDIT DEL INS CLR ▾

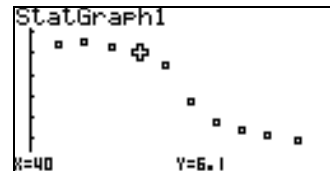
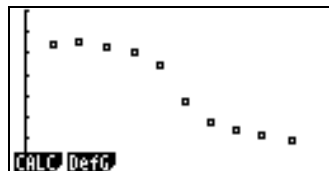
SHEE	A	B	C	D
6	59	3.7		
7	69	2.7		
8	78	2.3		
9	87	2.1		
10	99	1.9		

59
FILE EDIT DEL INS CLR ▾

Para dibujar el diagrama de dispersión, accedemos al menú GRPH. Asignamos el primer gráfico, GPH1 y especificamos las columnas que representan la x y la y, seleccionamos [Scatter] como tipo de gráfico.

```
StatGraph1
Graph Type: Scatter
XCellRange: A1:A10
YCellRange: B1:B10
Frequency : 1
Mark Type : □
GPH1 GPH2 GPH3
```

Observa que el gráfico de puntos puede recorrerse con las teclas de cursor, lo que es una buena forma de consultar los datos introducidos.

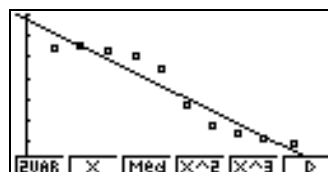


- b) Una vez dibujado el gráfico, el siguiente paso consiste en explorar la relación entre x e y buscando funciones que ajusten los datos aproximadamente. Pulsamos [CALC] y elegimos el tipo de función.

Por ejemplo, supongamos que seleccionamos el modelo lineal. Los coeficientes de regresión se calculan de la siguiente forma:

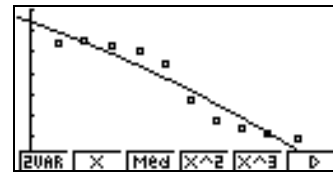
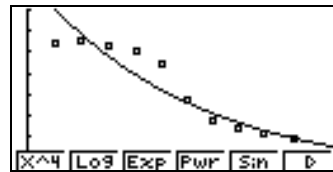
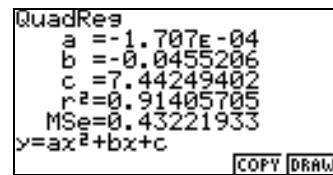
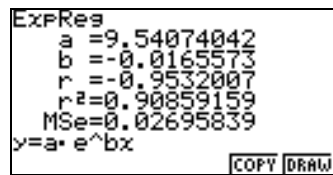
```
LinearRes
a = -0.0640032
b = 7.80897744
r = -0.9538903
r^2 = 0.90990678
MSe = 0.39645518
y = ax + b
COPY DRAW
```

El modelo lineal obtenido es $y = -0.0640032x + 7.80897744$. Observa que el coeficiente de correlación r es aproximadamente -0.9538903 . El coeficiente de correlación mide el grado de bondad del ajuste por una función lineal. En principio, observamos que el valor $|r| \approx 0.95$ indica una muy buena regresión; pero es posible que los datos se puedan ajustar por otro modelo funcional. De hecho, si dibujamos la recta sobre la nube de puntos, podemos observar que la recta no parece ajustarse del todo bien. Algunos puntos no están en la recta; de hecho, la nube de puntos no se parece mucho a una recta.



Así, un ajuste mejor se puede obtener mediante una curva. Experimentamos otros tipos de funciones (cuadráticas, exponenciales, cúbicas, cuárticas) y observamos la calidad del ajuste.

En las siguientes pantallas se muestran los modelos de regresión exponencial y cuadrático:

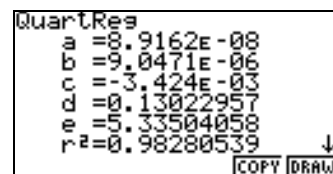
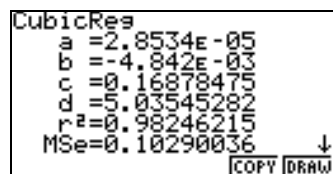


Como en el modelo lineal, observamos que algunos puntos no están cerca de las curvas exponencial y cuadrática. Los coeficientes de correlación correspondientes a los modelos exponencial y cuadrático son aproximadamente -0.9532007 y -0.9560633 (alrededor de -0.95 , la misma aproximación que la obtenida con el modelo lineal)

Otra exploración produce los siguientes resultados: con la curva cúbica se obtiene $r \approx -0.9911923$ y con la curva cuártica $r \approx -0.9955864101$. En ambos casos se obtiene un coeficiente de regresión: $|r| \approx 0.99$, que es muy cercano a 1. Sin embargo, el valor de $|r|$ para la curva cuártica es más cercano a 1 que el de la curva cúbica. ¡Por tanto, el mejor ajuste es aparentemente la curva cuártica!

La función cuártica de ajuste óptimo viene dada por la expresión:

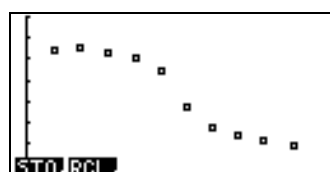
$$0.000000089162x^4 + 0.000009047x^3 - 0.003424x^2 + 0.13022957x + 5.33504058$$



Los valores negativos de r indican que la relación entre x e y es inversa. Así, cuando x crece, la y decrece. Esta relación puede visualizarse en el diagrama de puntos. Observa que el valor de r no aparece en estas pantallas. En realidad, no es necesario conocerlo. Basta saber el valor de r^2 , para conocer la calidad del ajuste. Como el valor de r^2 es cercano a 1, el ajuste es bueno.

Una buena forma de visualizar el ajuste por la curva es dibujarla sobre la nube de puntos. Esto es especialmente útil para comparar los gráficos de los diferentes modelos funcionales respecto de los datos.

Una vez obtenido y guardado en memoria el diagrama de dispersión con el nombre Pict1, pulsamos [OPTN]:



El diagrama se puede usar ahora como dibujo [SHIFT SET UP] de la siguiente forma:

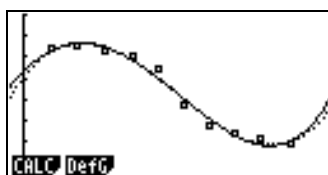


Los modelos funcionales obtenidos por regresión se pueden copiar a la lista de funciones gráficas y se pueden guardar como pantallas de dibujo.

Por ejemplo, copiamos las funciones cúbica y cuártica obtenidas por regresión al Editor de Gráficos. Asignamos los trazos "----" para la función cúbica y "_" para la función cuártica. Pulsamos [DefG] para dibujar ambos gráficos con el diagrama de puntos como dibujo:

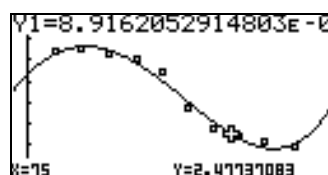
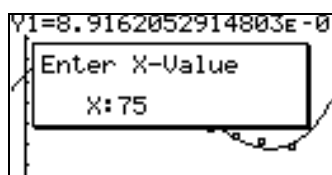
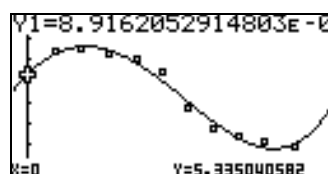
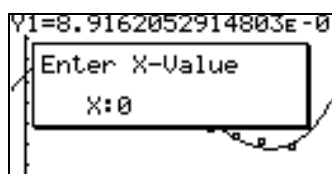


Podemos visualizar que ambas curvas de ajuste pasan muy cerca de los puntos y confirma esto que las dos podrían usarse como curvas de ajuste.



Es difícil averiguar, en algunos casos, qué modelo es mejor usar, fijándose solamente en los gráficos o en la forma de las nubes de puntos. Para decidir, por ejemplo, cuál de los dos modelos anteriores (cúbico o cuártico) es mejor, hemos de referirnos al valor del coeficiente de correlación para determinar la función de mejor ajuste.

- c) Usamos la función cuártica obtenida en (b) para estimar el número de granjas en 1900 y en 1975. Después de pulsar [DefG], pulsamos [DRAW], después [SHIFT G-Solv Y-Cal]. Especificamos el valor de x para obtener el correspondiente valor de y.



En el año 1900, habrán alrededor de 5.335 millones de granjas, y en 1975, alrededor de 2.477 millones de granjas.

• El Índice de Precios al Consumo

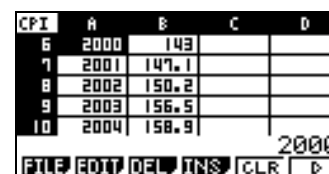
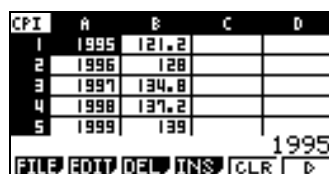
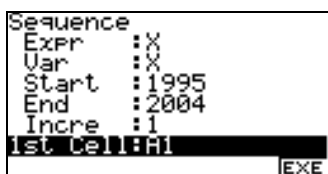
Los siguiente datos indican los niveles del Índice de Precios al Consumo (IPC) en Diciembre de distintos años:

Año	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
IPC	121.2	128.0	134.8	137.2	139.0	143.0	147.1	150.2	156.5	158.9

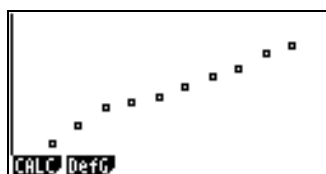
- Dibuja un diagrama de puntos de los datos.
- Determina el modelo (exponencial, logarítmico, potencial o lineal) que mejor describe la relación entre los años y el IPC.
- Usa el modelo obtenido en (b) para predecir el IPC para Diciembre de 2006.

Solución:

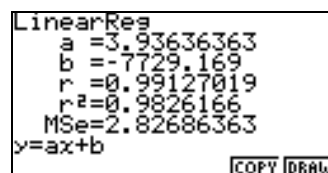
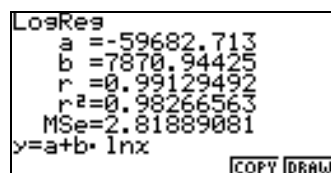
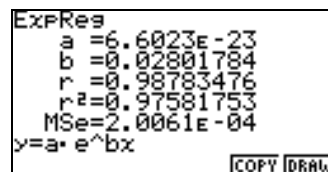
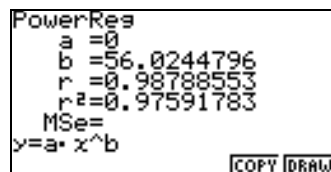
- Accedemos al Editor de Hoja de cálculo, e introducimos los años en la primera columna y los niveles del IPC en la segunda columna. Aplicamos la siguiente fórmula de sucesión para generar los años, pulsando [EDIT] y después [SEQ].



El diagrama de puntos es el siguiente:



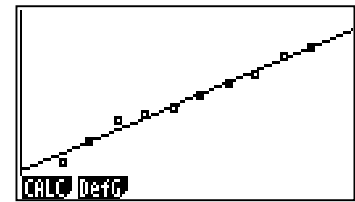
- Los coeficientes de regresión para varios tipos de regresión son los siguientes:



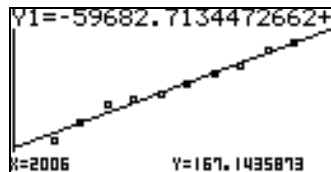
Resumimos los resultados obtenidos en la siguiente tabla:

Tipo de curva	COEFICIENTE DE CORRELACIÓN
Potencial	0.98788553
Exponencial	0.98783476
Logarítmica	0.99129492
Lineal	0.99127019

Los cuatro tipos de curvas dan una buena correlación. El hecho de que r sea positiva indica que la relación entre x e y es directa. Cuando x crece, y crece. El valor de $|r|$ para todas las curvas es cercano a 1; sin embargo, el mejor ajuste es aparentemente la curva logarítmica. El modelo logarítmico es: $-59682.713 + 7870.94425 \ln x$. El gráfico de la curva logarítmica superpuesto en el diagrama de dispersión es el siguiente:



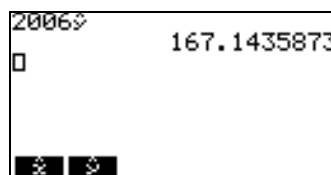
- c) Usamos el modelo logarítmico para determinar el nivel del IPC en 2006. Obtenemos aproximadamente 167.1435873.



Observaciones:

Los cálculos y gráficos de regresión obtenidos en el Editor de Hoja de cálculo se pueden obtener también en el Editor STAT usando los mismos comandos. Además, cuando trabajamos fuera de una hoja de e-activity, es posible usar el modelo funcional para hacer cálculos a través de los editores GRAPH, TABLE y RUN. Para hacer esto, la función debe copiarse primero en el Editor de Gráficos.

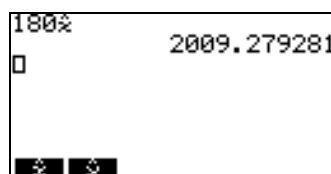
Por ejemplo, en el apartado (c) del problema anterior, calculamos el valor de y para $x=2006$ en el Editor RUN, de la siguiente forma:



Trabajando en el Editor GRAPH o RUN podemos no sólo hacer la y , sino también la x .

Por ejemplo, supongamos que queremos determinar aproximadamente en qué año el IPC será 180.

Usando el Editor RUN, la respuesta es el año 2009:



En el Editor GRAPH obtenemos una respuesta similar usando una ventana apropiada:

