ESTADÍSTICA BIDIMENSIONAL: RECTA DE REGRESIÓN LINEAL Y RECTA TUKEY.

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

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

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

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

INTRODUCCIÓN: A través de unas actividades y utilizando como herramienta habitual una calculadora gráfica, en nuestro caso el modelo CFX 9970 G de CASIO, realizamos el análisis crítico y el estudio de los parámetros obtenidos en la observación simultánea de dos variables, con un enfoque más investigador e innovador, dando prioridad al razonamiento, permitiéndonos más tiempo para pensar y analizar lo que hacemos y los resultados que obtenemos.

ACTIVIDAD 1

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

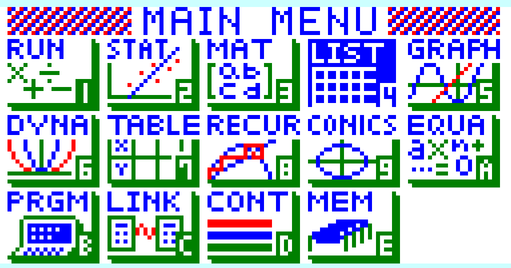
T	54	40	66	70	60	58	63	60
D	3	2	6	8	4	3	7	4

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

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

AC Presiona los cursores ▶ ◀ ▼ ▲ que consideres adecuados en el MENÚ INICIAL de presentación para seleccionar el modo **LIST.**

(o bien presionar directamente la tecla 4)

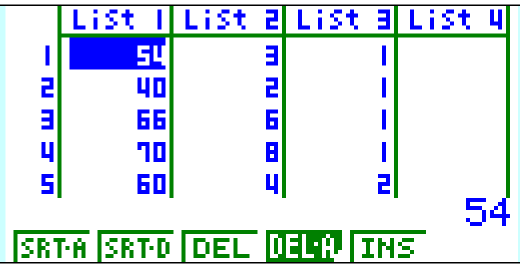


En el momento que presiones EXE entraremos en la alternativa de LISTAS.

El cursor aparecerá en la celda del primer elemento de la “List1”

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

Una pantalla podría ser ésta:



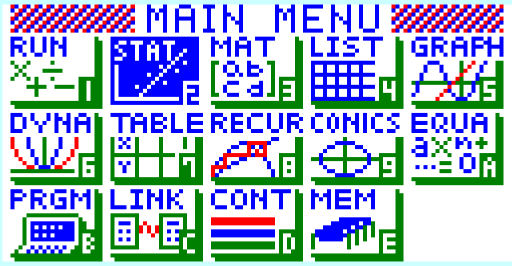
Por ejemplo, la fila 5 indica que 2 alumnos han obtenido 60 puntos en el Test sobre visión espacial y un 4 en Dibujo.

Si queremos obtener los valores de diferentes medidas de centralización y de dispersión tenemos que informar a la calculadora que los valores de los datos xi los tenemos señalados en la **List1**, los valores de yi los tenemos en la **List2** y que las frecuencias se encuentran en la List3. Para ello bastará con realizar los siguientes pasos:

1.- Entrar en el MODO ESTADÍSTICO

MENU ◀ ◀

o bien presionar directamente la tecla 2)



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

2.- Adecuar las listas a las variables:

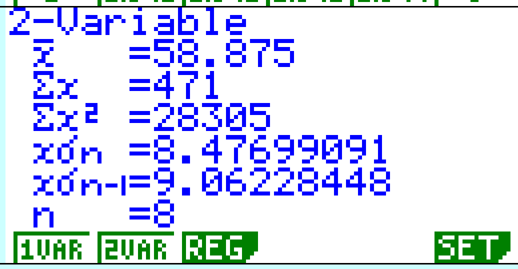
CALC F2 SET F6

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

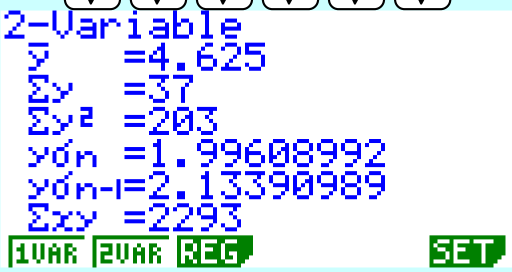
EXE

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

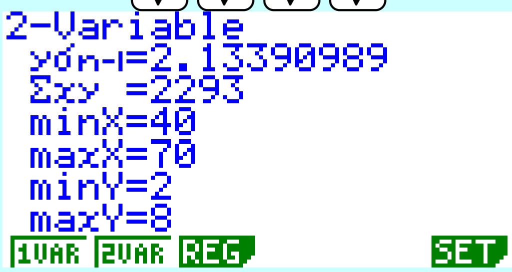
2VAR F2



▼ ▼ ▼ ▼ ▼ ▼



▼ ▼ ▼ ▼



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

TEST SOBRE VISIÓN ESPACIAL (T)

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

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

Σx^2 ➡ La suma de los cuadrados de los puntos obtenidos en el test sobre visión espacial por los 8 alumnos ha sido de 28305 puntos cuadrados.

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

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

(*) NOTA: ¿cómo se obtiene el valor 75% aparecido anteriormente?

Hay 6 elementos de 8 posibles incluidos en ese intervalo: $\frac{6}{8}\cdot\frac{100}{100} = \frac{75}{100} = 75\%$

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

CALIFICACIONES EN LA ASIGNATURA DE DIBUJO (D)

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

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

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

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

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

(*) NOTA: ¿cómo se obtiene el valor 62'5% aparecido anteriormente?

Hay 5 elementos de 8 posibles dentro de ese intervalo: $\frac{5}{8}\cdot\frac{100}{100} = \frac{62'5}{100} = 62'5\%$

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

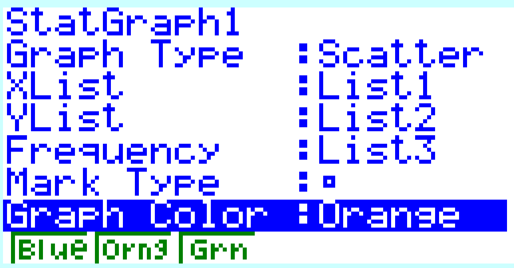
Otra de las ventajas de la calculadora gráfica es que nos permite representar con total sencillez el diagrama de dispersión o nube de puntos, para poder analizarlo rápidamente.

EXIT EXIT GRPH F1

Seleccionamos la forma de la gráfica:

SET F6

que podría ser la **GPH1** que aparece al margen:

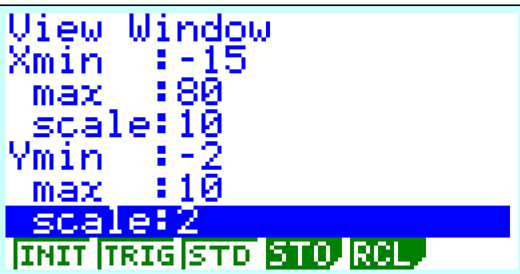


No hay que olvidar que hay que trabajar en MANUAL para introducirle el valor de los parámetros de visualización de pantalla que queramos:

EXE SHIFT SET UP MENU Man F2 EXE

SHIFT V-Window F3

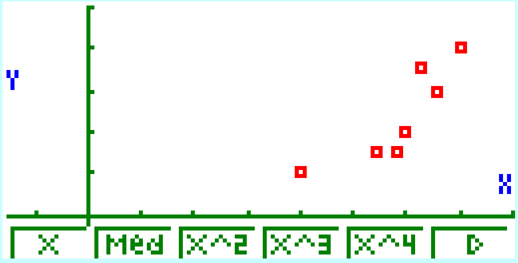
La escala adecuada para una correcta visualización de esta gráfica podría ser la señalada al margen:



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

GRPH GPH1
[EXE] [F1] [F1]

Podría ser la que se indica adjunta



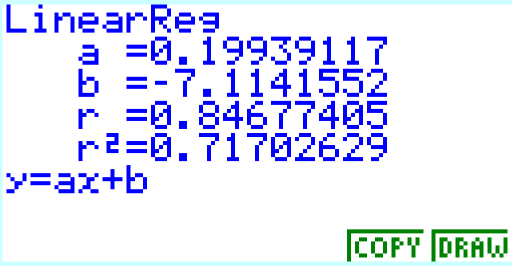
Si queremos dibujar la recta de regresión de y sobre x, conocer su ecuación y su coeficiente de correlación, bastará con realizar estos sencillos pasos:

X
[F1]

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

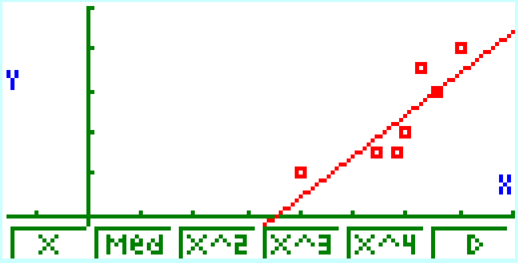
y = 0'19939117x – 7'1141552

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



DRAW
[F6]

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



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

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

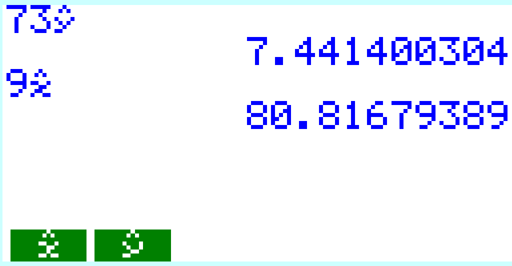
NOTA: Al estar "r" cercano a 1 los valores que alcancemos estarán próximos al valor estimado.

Estos cálculos han de realizarse desde la opción RUN del menú inicial.

STAT
[MENU] [1] [OPTN] [F5]

$\bar{y}$
[7] [3] [F2] [EXE]

$\bar{x}$
[9] [F1] [EXE]



ANÁLISIS CRITICO DE LOS RESULTADOS:

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

ACTIVIDAD 2

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

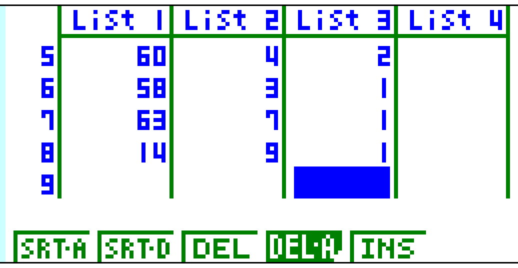
T	14	54	40	66	70	60	58	63	60
D	9	3	2	6	8	4	3	7	4

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

RESOLUCIÓN:

MENU 4

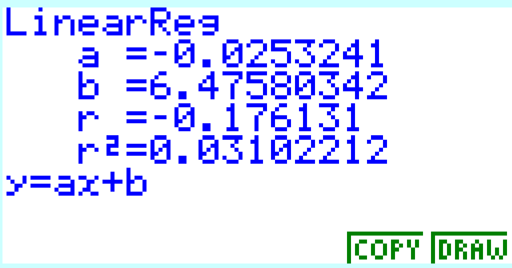
Llevamos el cursor al final de las columnas e introducimos el valor T, D (14,9) con frecuencia 1



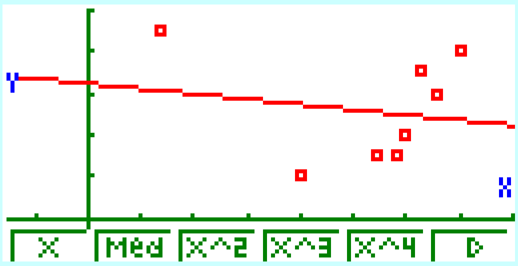
Representamos gráficamente y dibujamos la recta de regresión de y sobre x:

MENU 2 GRPH F1 GPH1 F1 X F1

obteniéndose la ecuación de dicha recta:



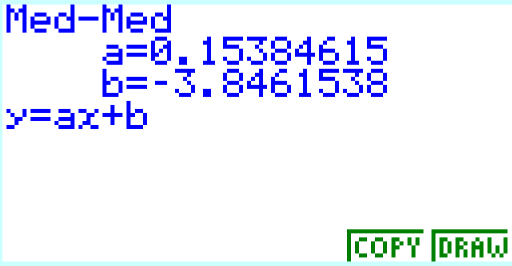
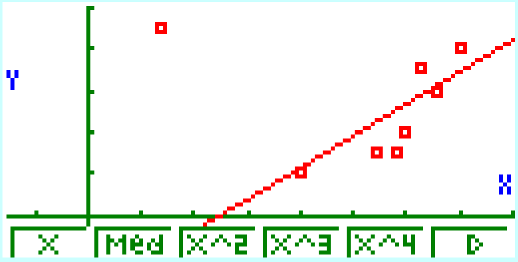
DRAW F6



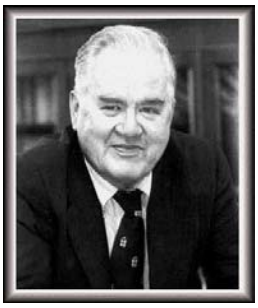
Como es una recta basada en las **medias** ($\bar{x}$, $\bar{y}$), podremos observar que se ajusta muy mal a la **nube de puntos** debido a ese punto extremo (14, 9), que afecta sensiblemente a dichas medias.

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

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

Med F2	
La opción **DRAW** dibujará la dicha recta con suma facilidad DRAW F6	

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



John Wilder Tukey
(1915 - 2000)

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