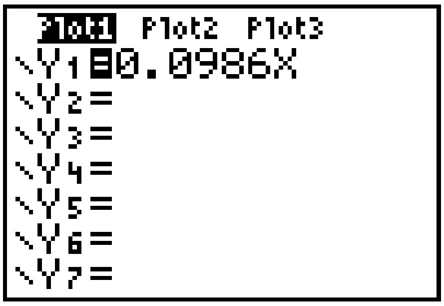
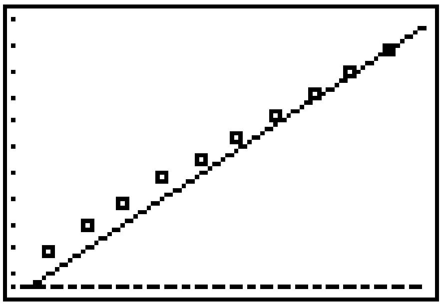
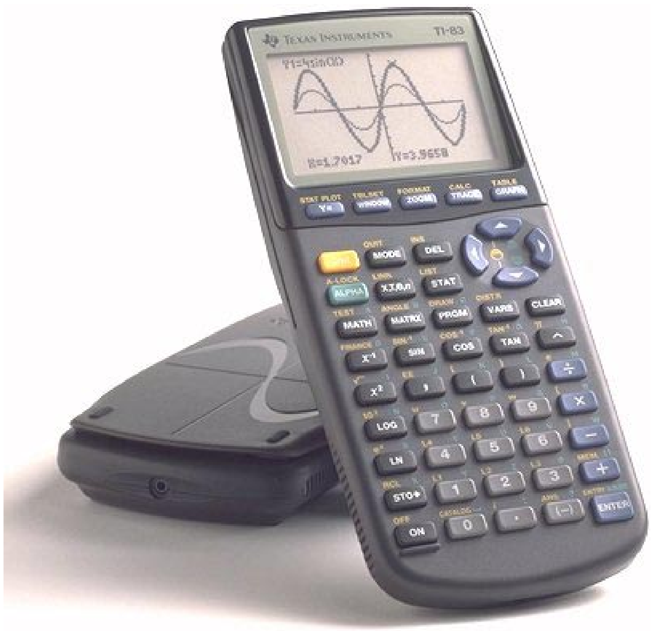
Estadística Descriptiva

TI 83





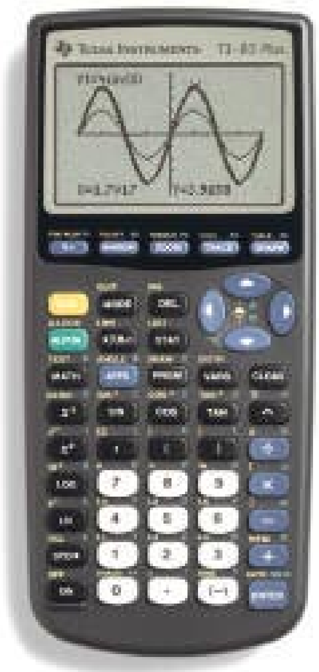


T^3 España

T^3 EUROPE

Onofre Monzó
José Antonio Mora
Pascual Pérez
Tomás Queralt
Julio Rodrigo
Salvador Caballero
Floreal Gracia
Fernando Juan
Alfred Mollá

ÍNDICE



Introducción

El proceso para el trabajo estadístico con una calculadora gráfica es muy similar al que se realizaría manualmente:

1. Selección de los datos y preparación de los cálculos. (Introducción de los datos en la calculadora)
2. Selección y realización de los cálculos pertinentes. (Obtención de los parámetros estadísticos)
3. Representación gráfica.
4. Conclusiones (y posible vuelta a los pasos 1, 2, 3 y 4).

Con la gran ventaja de la celeridad en los cálculos y en las representaciones gráficas, lo que permite que se pueda dedicar más tiempo al análisis y a la interpretación de ambos aspectos, y en analizar variaciones y análisis de resultados.

Tomaremos dos ejemplos de estadísticas aparecidas en prensa para ver la forma de proceder con una calculadora gráfica. En definitiva se trata de realizar un análisis estadístico de los datos publicados para realizar comparaciones y extraer conclusiones que permitan hacer afirmaciones a la vista de las tablas y los resultados estadísticos obtenidos.

Además, se pueden pasar datos de una calculadora a otra, lo que facilita seguir el trabajo en clase cuando por error algún alumno borra accidentalmente parte de los datos o todos. Se pasan los datos a la calculadora y se puede seguir el estudio estadístico sin tener que introducirlos de nuevo.

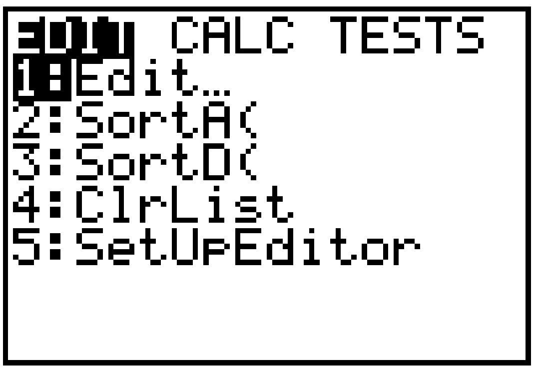
LLENO EN LOS ESTADIOS. (EL PAÍS, 29/8/95)

Equipo	**Abonos temporada 1995-96**	**Abonos temporada 1994-95**
Albacete	10.000	6.500
Athletic	31.000	31.000
At. Madrid	18.000	15.000
Barcelona	104.000	104.000
Betis	35.000	15.000
Celta	15.200	14.000
Compostela	9.000	8.200
Deportivo	29.500	26.000
Español	19.000	16.000
Mérida	8.000	2.700
Oviedo	11.000	10.000
Racing	10.000	9.000
Rayo	5.200	4.000
R. Madrid	65.000	65.000
R. Sociedad	20.000	13.500
Salamanca	13.000	3.500
Sevilla	35.000	34.200
Sporting	18.000	14.000
Tenerife	13.000	12.000

1. Selección de los datos y preparación de los cálculos.

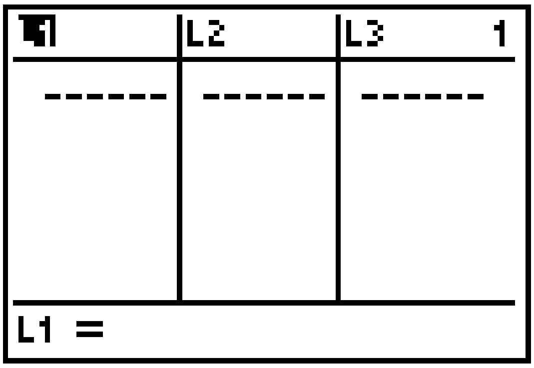
Pulsamos [STAT]

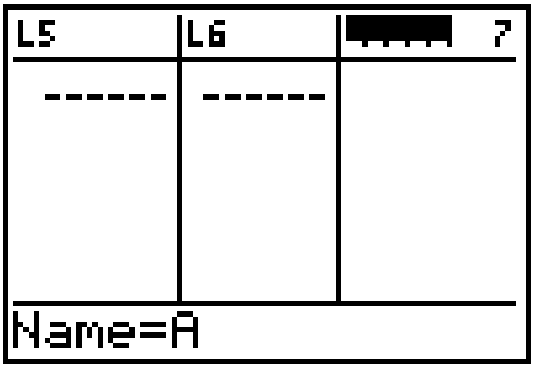
Tomamos la opción 1: Edit (para lo cual se pulsa la tecla 1 o [ENTER]). Se mostrará el editor de listas estadísticas.



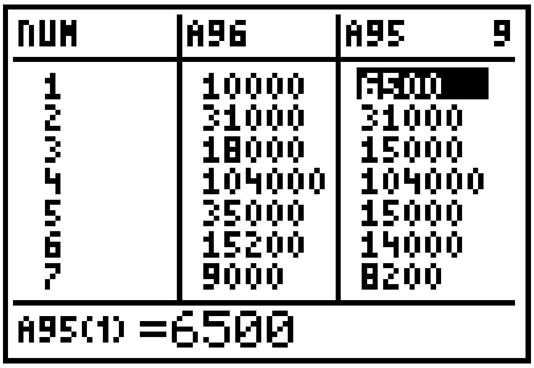
Pulsamos [▶] hasta que aparezca una lista sin nombre. Se muestra el indicador Name= en la línea de introducción y se activa el bloqueo alfabético. Pulsamos las letras o números para introducir el nombre.

El primer carácter no puede ser un número.



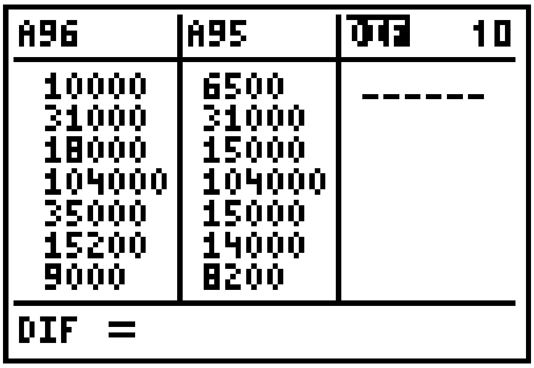


Para poner el nombre en las columnas: tecleamos NUM y pulsamos [ENTER] o [▼] para almacenar el nombre de la lista en la columna actual del editor de listas estadísticas. El nombre de la lista ahora será un elemento del menú LIST NAMES.

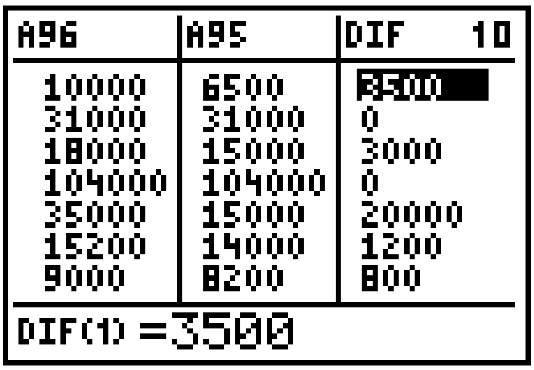


Ahora comenzamos a introducir los datos en las listas A96 y A95, en la lista NUM introducimos 1, 2, 3, 4,... que representan a los distintos equipos.

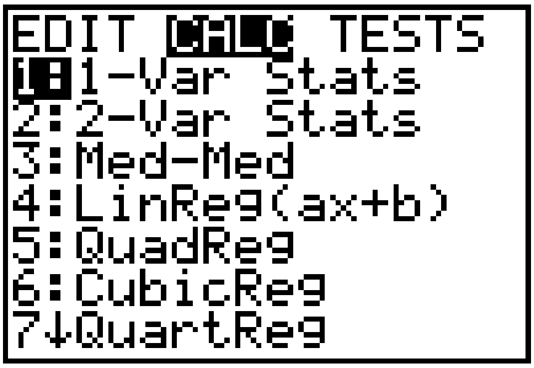
Podemos hacer una lista DIF con las diferencias de socios en las dos temporadas para lo cual nos situamos con el cursor en la etiqueta DIF y tecleamos LA96-LA95 (pulsando [ENTER] posteriormente):



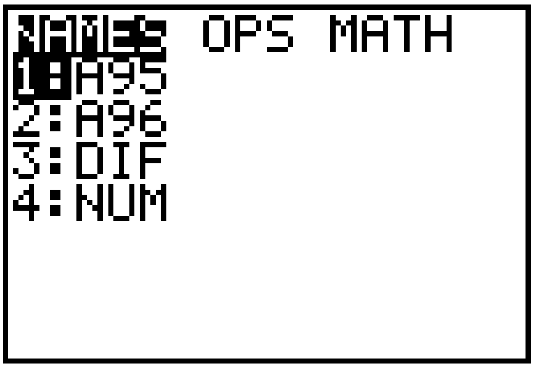
Con posterioridad volveremos a la edición de datos, pasemos ahora a ver algunas opciones de cálculos estadísticos. A partir de ahora cuando aparezca una tecla no diremos pulsar, quedará sobreentendido, así como tampoco diremos pulsar [ENTER], queda claro que esta es la forma de dar a entender a la calculadora que debe realizar cualquier opción.



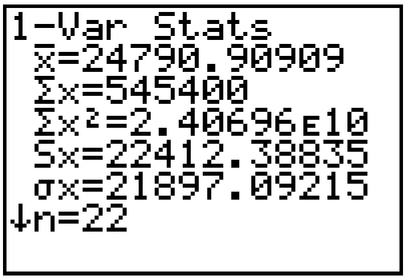
2. Selección y realización de los cálculos pertinentes.

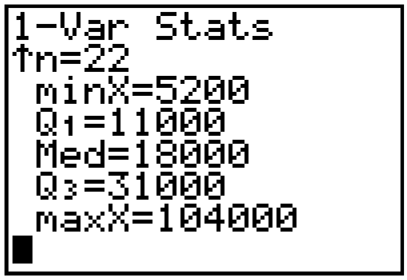


Pulsamos [STAT] opción CALC, opción 1: 1- Var Stats:

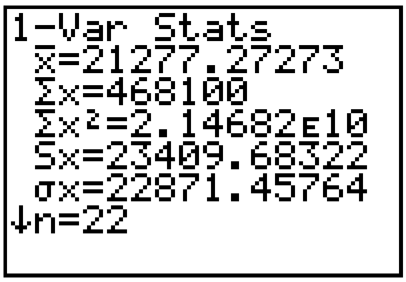


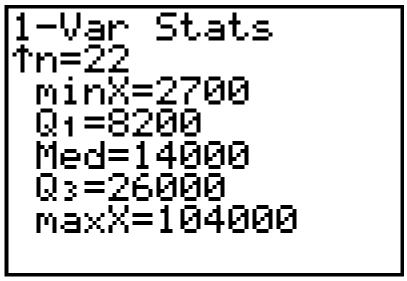
Con lo que nos aparecerá en la pantalla principal, a continuación introduciremos el nombre de la lista ([2nd] [STAT] 2:A96) donde están los datos y pulsaremos [ENTER] para ejecutarla, y tal como indica la flecha, podemos desplazarnos hacia abajo para continuar viendo más cálculos:





Hacemos ahora lo mismo pero seleccionando como datos los contenidos en A95 (temporada 94-95):





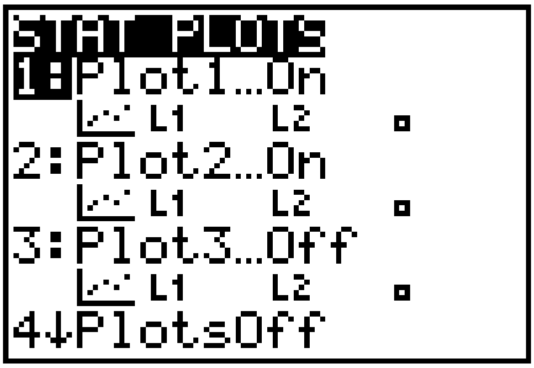
Una observación de resultados nos permite comparar fundamentalmente la media de abonados en las dos temporadas así como el rango intercuartílico. En la temporada 94-95 la media de abonados era de 21.277, con Q_1= 8.200 y Q_3= 26.000, siendo la mediana de 14.000. En la temporada 95-96 la media fue de 24.790, con Q_1= 11.000 y Q_3= 31.000, siendo la mediana de 18.000.

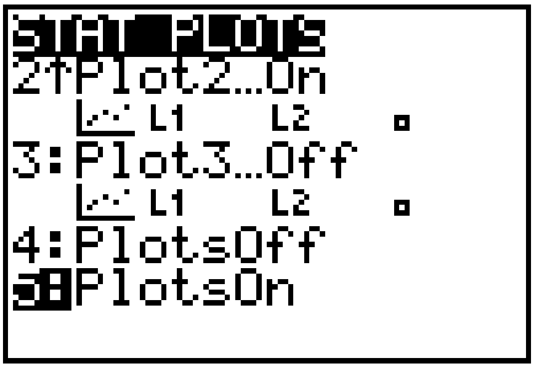
Estos resultados pueden observarse gráficamente. Veámoslo.

3. Representación gráfica.

Vamos a ver los menús de representación estadística STAT PLOTS (2nd Y=)

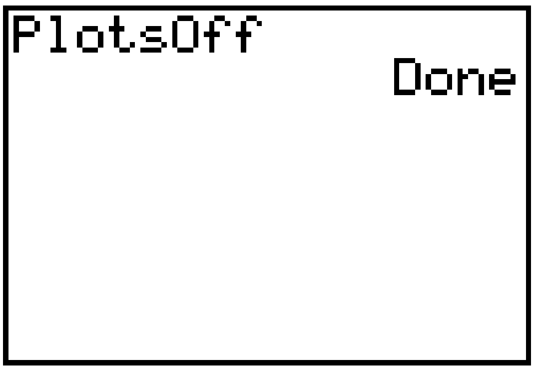
La pantalla STAT PLOTS nos muestra visibles 4 opciones y otra quinta a la que se accede con la teclas de desplazamiento.



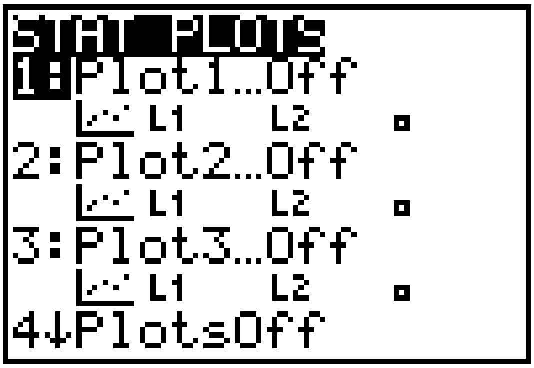


Las tres primeras indican el estado de las tres posibles representaciones estadísticas que puede realizar simultáneamente la calculadora. En nuestro caso todas las dos primeras están en ON (activadas) y la tercera en OFF (desactivada), las opciones 4 y 5 sirven para poner todas las representaciones en ON o todas en OFF. Para desactivarlas todas tomamos la opción 4: PlotsOff , pasando la orden a la pantalla principal, se ejecuta pulsando ENTER, y lo indica la calculadora por medio de la respuesta "Done".

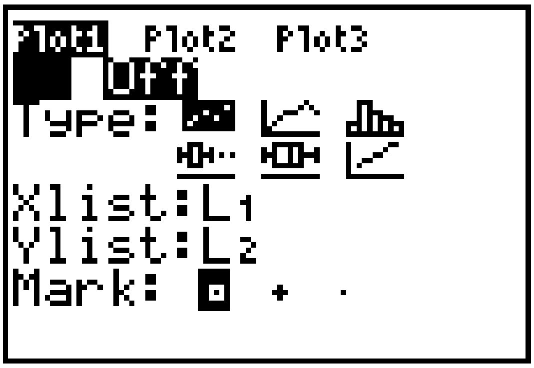
La TI-83 puede mostrar simultáneamente tres representaciones gráficas estadísticas, además de las funcionales.



Cuando volvemos a la pantalla de representaciones estadísticas ([2nd][Y=]) se aprecia que efectivamente las tres opciones están en OFF.



Tomamos la primera opción 1: Plot1... con lo que nos introducimos en la pantalla en la que habrá que decidir qué tipo de representación –gráfica, de las seis posibles– queremos y con qué listas (de las posibles, además de la lista constante 1).



Lo primero que tenemos que seleccionar es ON, para activarla. Después el tipo (Type:), como podemos ver está la representación de nube de puntos (diagrama de dispersión, scatter) en la que hay que decir qué lista es la X, qué lista es la Y y la marca para la representación (hay tres opciones).

La siguiente representación posible es como la anterior pero uniendo los puntos, las decisiones a tomar son las mismas que en la anterior opción, Xlist, Ylist y la marca (Mark).

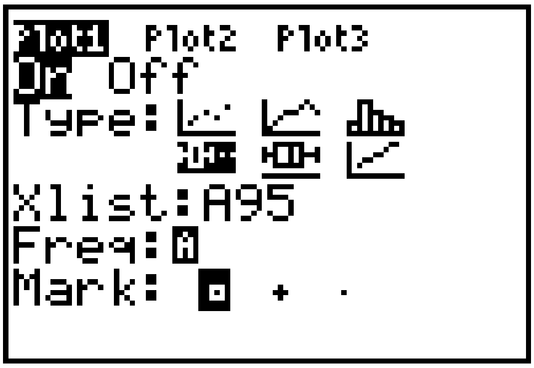
El tercer tipo es el Histograma. En el caso de tomar este tipo es necesario que determinemos Xlist y la frecuencia (Freq), que como vemos puede ser cualquiera de las listas o la lista constante de 1. El valor de la variable de ventana Xscl determina el ancho de cada barra, empezando en Xmin. ZoomStat ajusta Xmin, Xmax, además ajusta Xscl. Se debe cumplir la siguiente desigualdad:

$$(Xmax - Xmin)/Xscl \leqslant 47$$

El cuarto tipo es el del diagrama de caja modificado en que representa los datos igual que el diagrama de caja, excepto que los datos que están 1'5 x (*rango intercuartílico*) más allá de los cuartiles (el *rango intercuartílico* se define como la diferencia entre el tercer cuartil Q_3 y el primer cuartil Q_1). Estos puntos se representan de manera individual más allá de la línea (whisker).

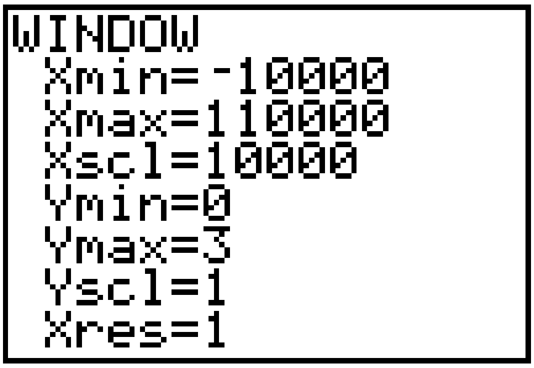
El quinto tipo es el del diagrama de caja en el que se representa el valor mínimos, el máximo y el rango intercuartílico, siendo necesario indicar como en el anterior Xlist y Freq.

El sexto tipo, gráfico de probabilidad normal, representa cada observación de **X** en **Data List** frente al cuartil correspondiente **z** de la distribución estándar normal. Si los puntos representados se aproximan a una línea recta, el gráfico indicará que los datos son normales.

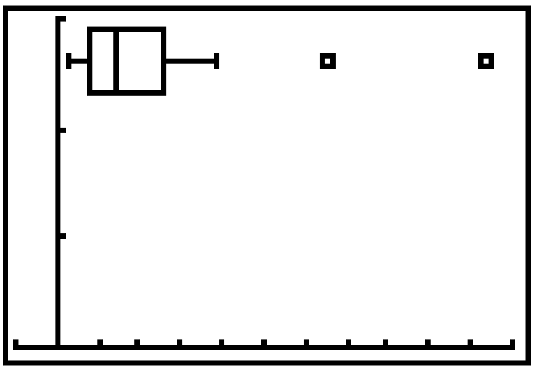


De acuerdo a lo que pretendíamos en nuestro estudio de abonados a los campos de fútbol vamos a tomar en principio la representación de caja modificado tomando como Xlist la lista A95 de abonados en el 94-95 con Freq 1 y Marca □.

Volvemos a la pantalla resumen de las opciones de representación estadística [2nd][Y=] y observamos que efectivamente está seleccionado el Plot 1 con la opción elegida de caja para la lista A95.

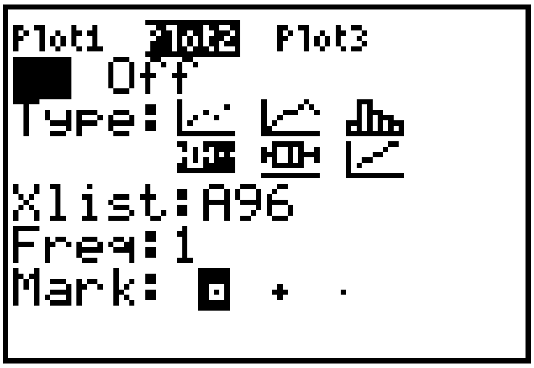


Para pasar a obtener la representación vamos a la ventana [WINDOW] y tecleamos el campo de representación que queremos, por ejemplo los valores que vemos aquí:



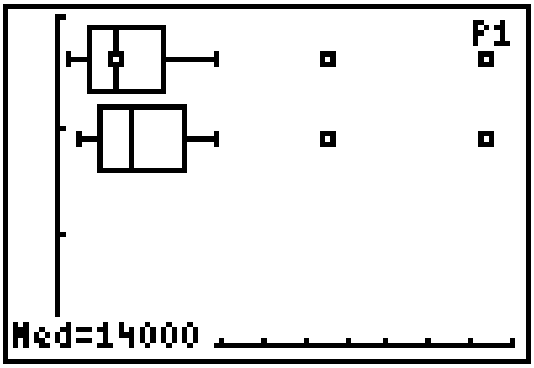
Y ahora al pulsar [GRAPH] obtenemos:

Pulsamos [TRACE] y al mover el cursor con las flechas de desplazamiento se sitúa (y nos lo indica en el extremo izquierdo) en minX, Q1, Med, Q3, 1.5* rango intercuartílico y los valores extremos, con lo que podemos visualizar el rango intercuartílico (en el que está el 50% de los datos), y comparar su gran concentración (caja muy estrecha) frente a los datos comprendidos entre Q3 y el maxX (25% de los datos), así como lo separado que está este valor máximo de los 50% centrales. (Este en realidad es un valor extremo ya que se encuentra a más de 1.5 veces el rango intercuartílico, con lo que para un estudio de evolución de los abonados debería eliminarse por la distorsión que produce).

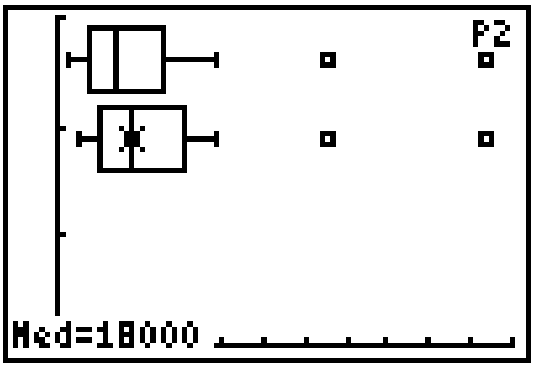


Para comparar los datos de abonados de los dos años, vamos a utilizar la misma representación.

La importancia de un estudio como el que estamos haciendo consiste en la comparación por lo que vamos a activar con el mismo tipo de representación de caja el Plot2 para la lista A96 (de abonados en el 95-96):



La activamos con On, en Xlist tomamos la L_2 y en Freq 1, como los datos se mueven en el mismo rango de valores no necesitaremos cambiar la ventana de representación por lo que pasamos directamente a [TRACE] y obtenemos :

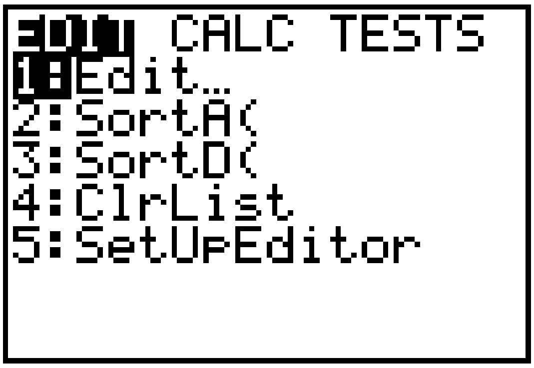


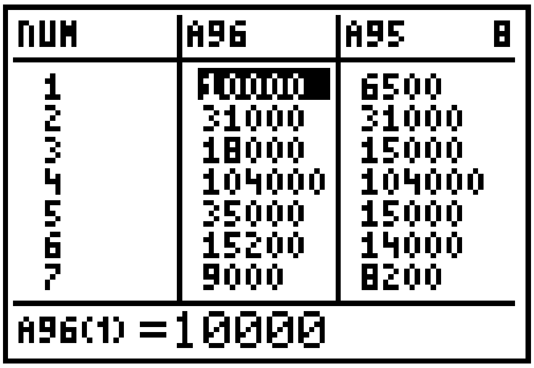
La indicación del extremo superior derecho de la pantalla, P1, nos indica tal como se aprecia en el cursor y en el dato inferior numérico, que está en disposición de recorrer al pulsar ◄ o ► la representación del Plot1, el de la temporada 94-95, cuando pulsemos ▲ o ▼ el cursor se traslada al segundo diagrama de caja, el de la temporada 95-96.

Ahora se puede observar el desplazamiento a la izquierda de la caja (que representa el rango intercuartílico), así como minX y maxX iguales o muy próximos en ambas distribuciones.

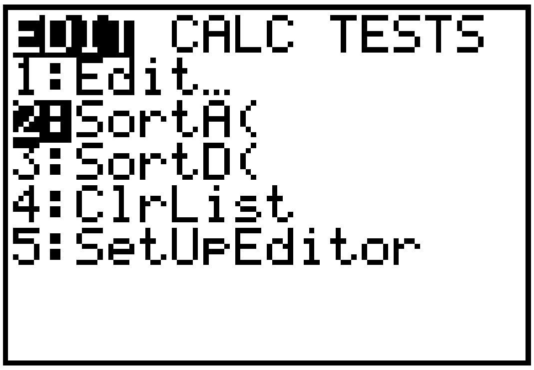
Tal como comentaba con anterioridad, si eliminamos (ojo, sólo a efectos de estudio estadístico) el dato del 104000 podremos observar mejor el aumento que se ha producido en los abonos. Para ello será necesario ir al paso 1. Edición y después al 3. Representación.

Veámoslo por pantallas sin indicar la teclas a pulsar:

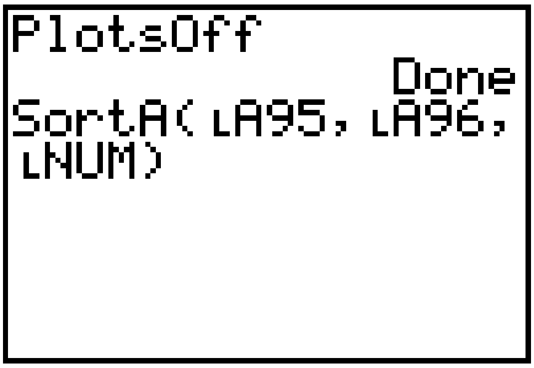




Y ahora podríamos bajar hasta el valor 104000 en ambas listas (y el valor 4 en NUM) y suprimirlo con DEL. Sin embargo vamos con anterioridad a ordenar las listas (ya que esto es algo muy cotidiano y útil en estadística), para ello pulsamos STAT y tomamos la opción 2:SortA(

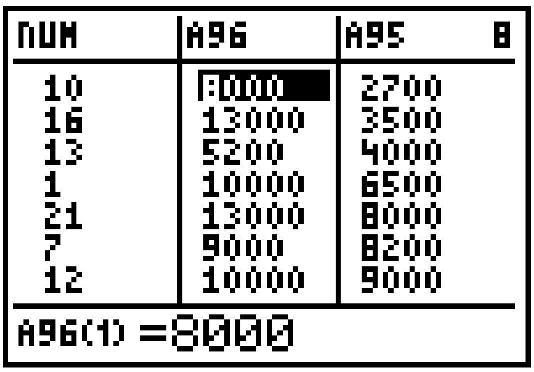


Esta orden pasa a la pantalla principal y a continuación seleccionamos (2nd STAT) A95, A96 y NUM para que ordene de menor a mayor los abonados arrastrando aparejados los datos de A96 y de NUM. Cuando hemos ejecutado la orden aparece Done en la calculadora. SortA (ʟA95) ordena sólo esta lista y no las demás, por lo que perderíamos la relación que existe.

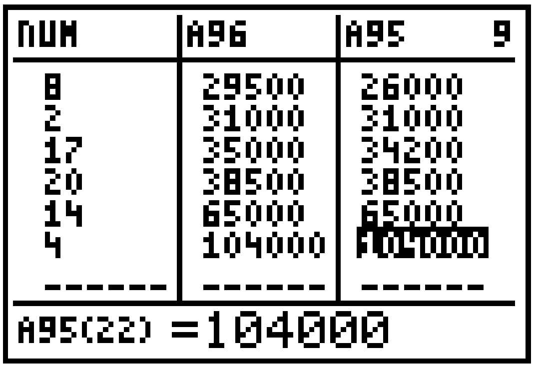


Veamos que ha ocurrido con los datos contenidos en las listas.

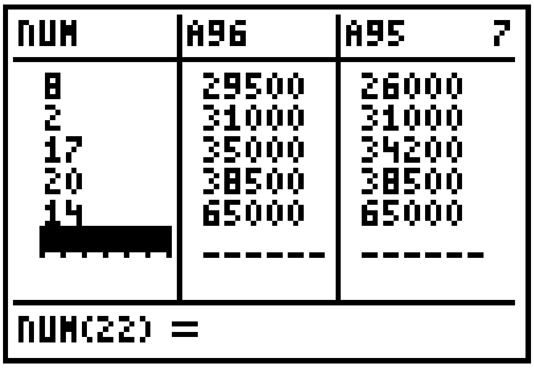
Efectivamente se ha ordenado A95, moviendo los datos asociados.



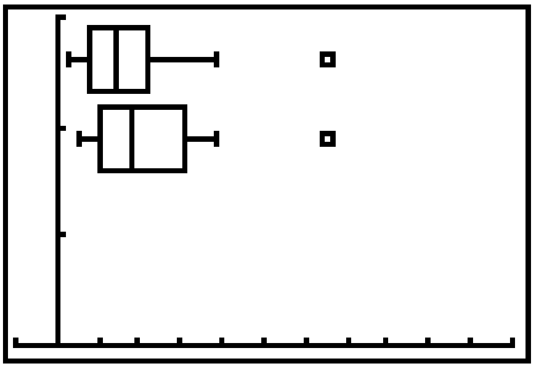
Vamos al final de la lista



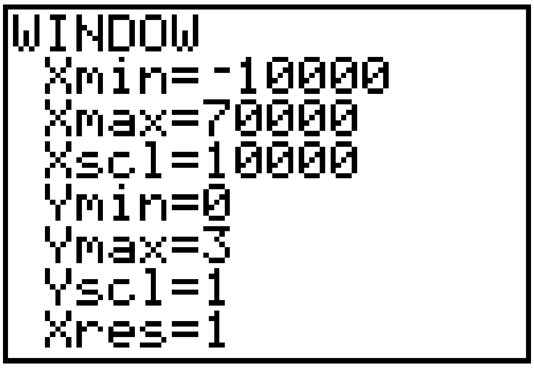
y con DEL eliminamos la última fila.

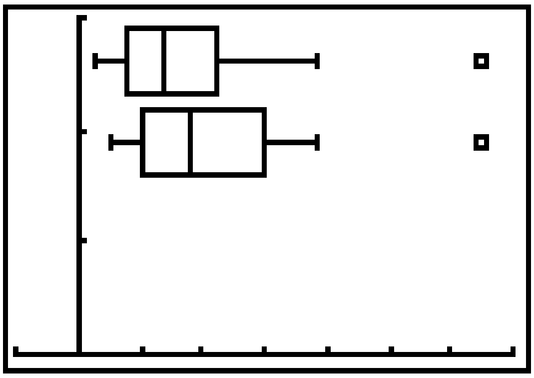


Ahora podemos ver el efecto de esta eliminación en los diagramas de caja. Pasando a representación mediante GRAPH



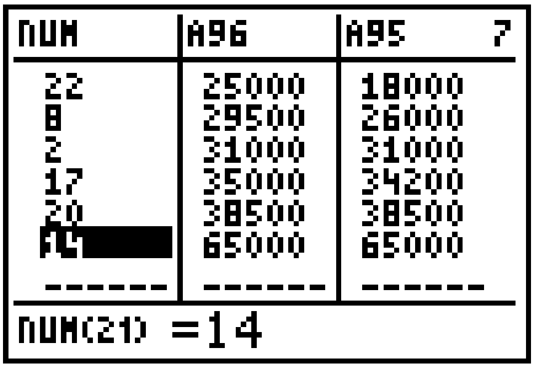
Si quisiéramos una representación mejor vamos a WINDOW y tomamos los siguientes valores para la representación, con lo que al ir después a la opción GRAPH, podemos ver con mayor claridad la distribución de abonados por cuartiles.

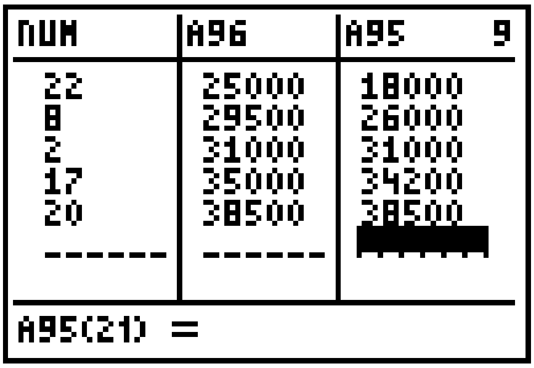




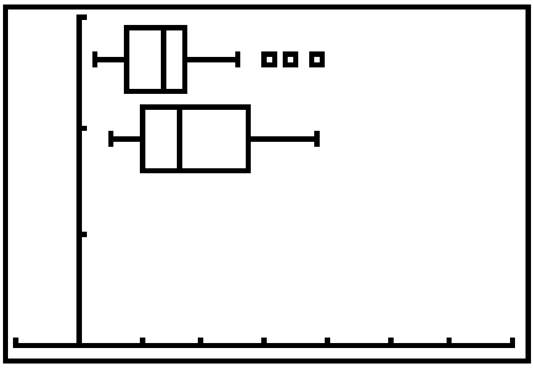
Se puede ahora quitar el otro valor que distorsiona los datos, el 65.000 del Real Madrid.

Vamos al Editor de Listas, y suprimimos la última fila, en NUM el 14 y en A96 y A95 el 65.000,

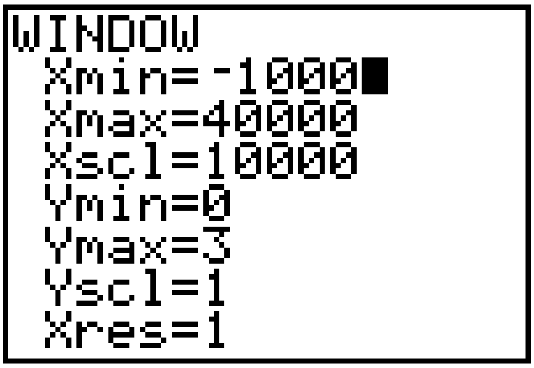




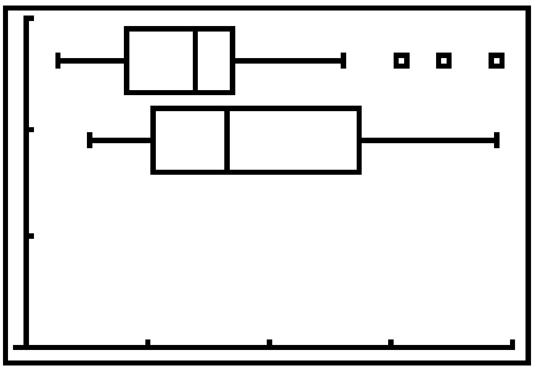
y como las opciones de representación son las mismas de antes vamos a la pantalla de representación gráfica con GRAPH y obtenemos:



y si queremos una mejor representación definimos el valor xmax=40.000 en WINDOW:

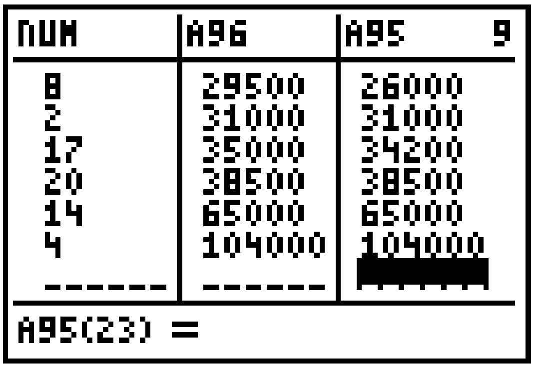


con lo que obtendríamos:

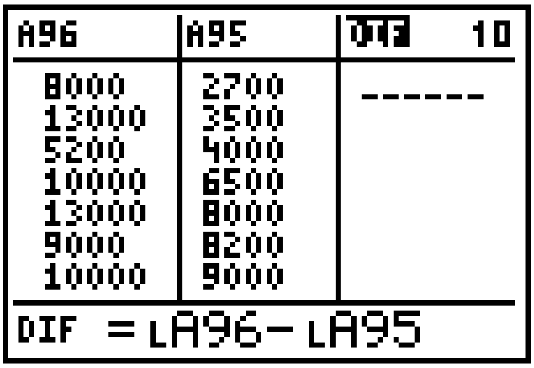


En este caso, descontados ya los dos equipos en los que casi es imposible que aumente el número de abonados, se aprecian mejor las diferencias entre ambas temporadas para el resto de los equipos.

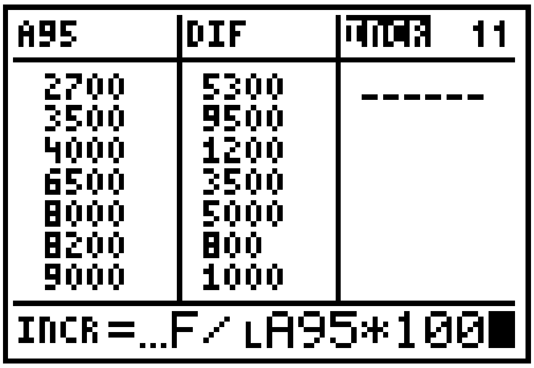
Vamos a continuar con otras posibilidades de análisis de los datos para lo cual primeramente volvemos a introducir los datos de los dos equipos suprimidos.



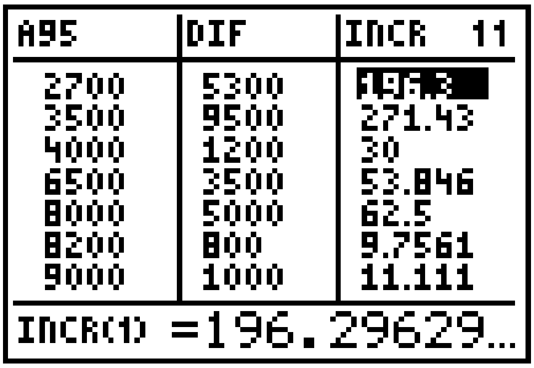
Y pasamos a la etiqueta de la lista DIF definiéndola de nuevo como A96-A95 (esta lista ya estaba creada, pero como con posterioridad ordenamos las dos listas habrá quedado desordenada).



Estos incrementos absolutos los pasamos ahora a porcentajes, para lo que ponemos nombre a una lista (INCR) y definimos INCR = DIF /A95*100

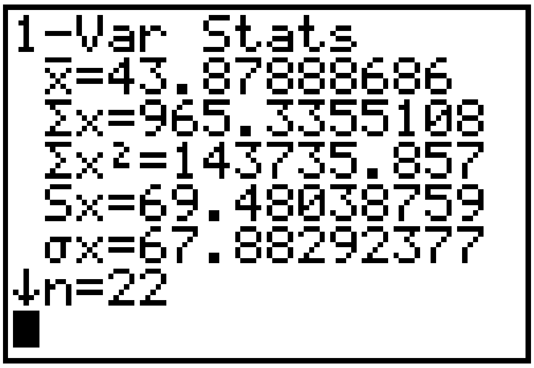


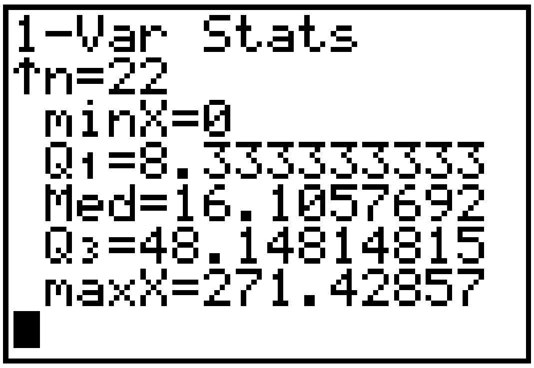
Y al ejecutarlo tenemos la columna de incrementos porcentuales.



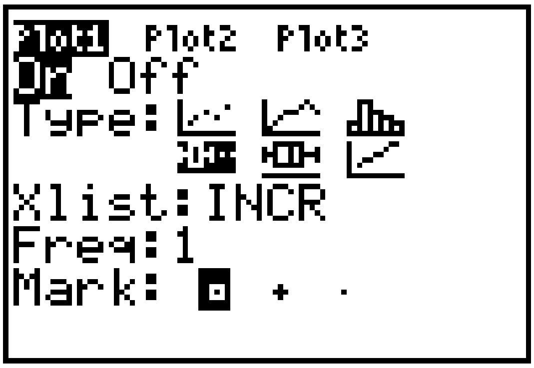
Ahora pasamos ya a las opciones de cálculo, ya que lo que pretendemos es información sobre los incrementos porcentuales, elegiremos la lista INCR.

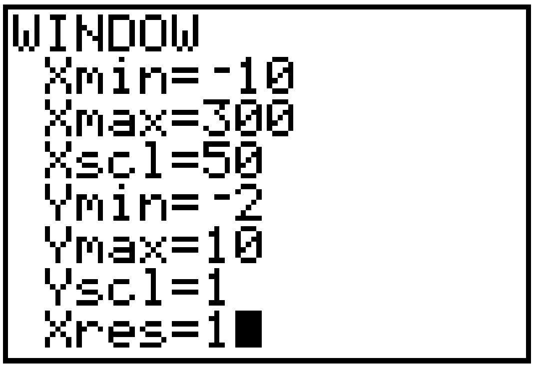
Con lo que obtenemos lo siguiente:



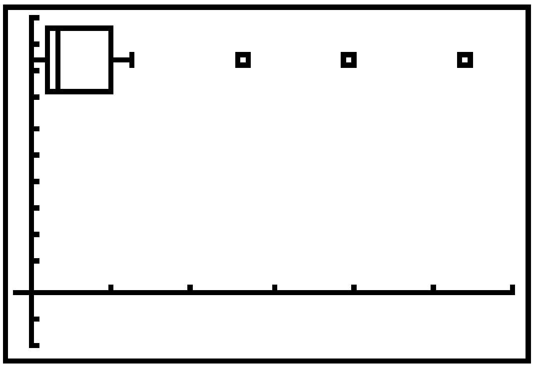


y si queremos representarlos gráficamente, elegimos el Plot1 y adecuamos el tamaño de la ventana con WINDOW:

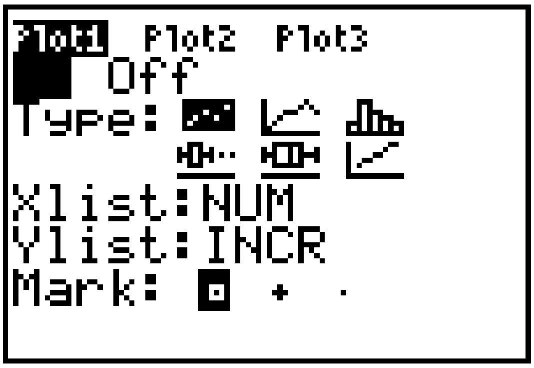


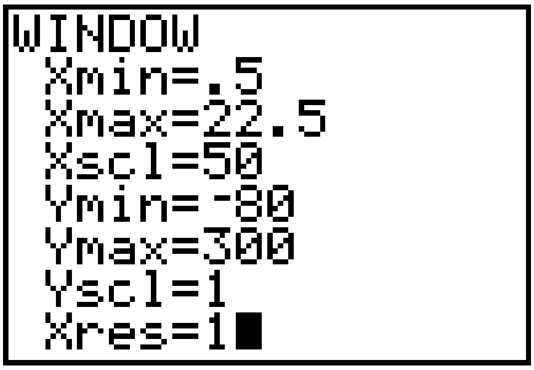


con lo que obtenemos:

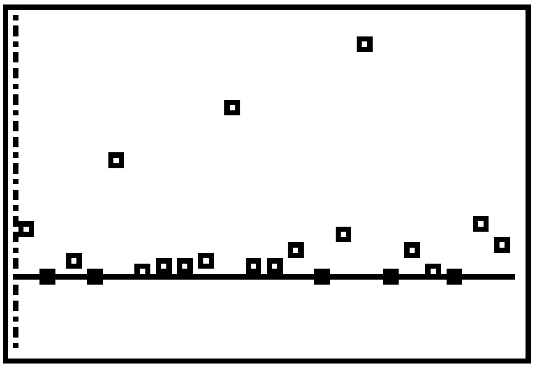


O también podemos tomar el Plot 1 con la definición:

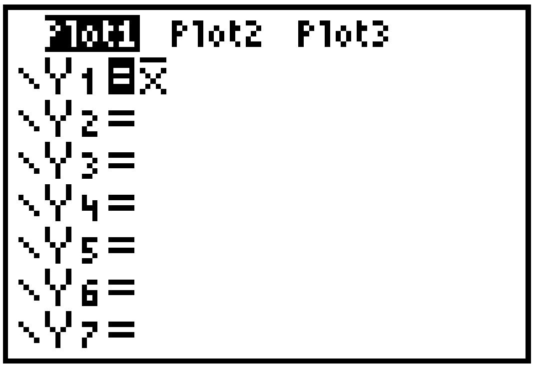




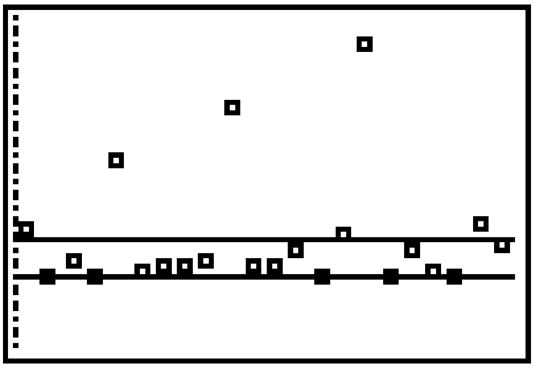
y obtenemos



Si queremos tener representada la media de incrementos introducimos en el editor de funciones [Y=], $y_1=\bar{x}$ ([VARS], 5: Statistics...., 2: $\bar{x}$)

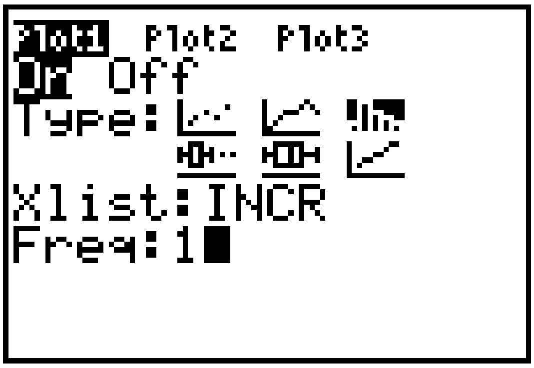


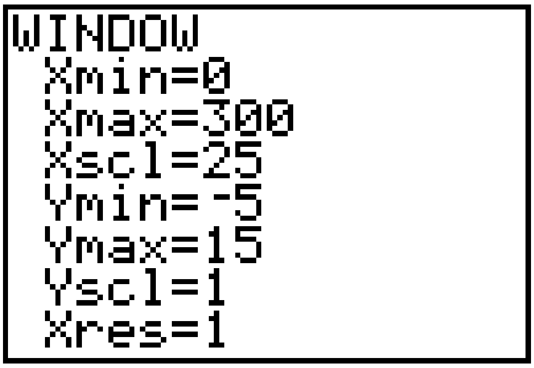
Con lo que acceder a la pantalla de gráficos tendremos:

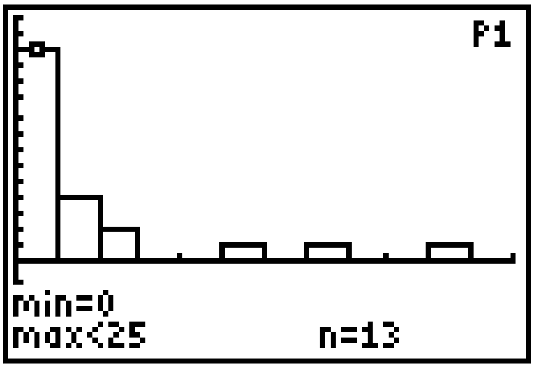


Quedando claramente muy por encima de la media los equipos 5, 10 y 16 que corresponden al Betis, Mérida y Salamanca respectivamente. (Se puede intentar explicar si han subido a 1ª división, etc.)

Cuando la opción de representación es HISTOGRAMA en la definición de la ventana de representación WINDOW, Xscl representa la amplitud de los rectángulos, por lo que si damos una amplitud por ejemplo de 25 al tomar la lista INCR con frecuencias 1 constante nos dará el diagrama agrupando los incrementos de abonados de 25 en 25:







Donde se indica que hay 13 equipos en el que su incremento porcentual de abonados en ambas temporadas está entre 0 y 25 (los intervalos los toma cerrados por la izquierda y abiertos por la derecha); 4 equipos han variado porcentualmente entre un 25 y un 50. Etc. Etc.

Podemos para concluir afirmar que hemos realizado un tratamiento de los datos en su sentido más amplio, los hemos organizado a nuestro gusto y hemos sacado consecuencias de los parámetros numéricos y de las representaciones gráficas. Sólo falta una fase imprescindible en la que la calculadora no puede ayudarnos, y es la reorganización y sistematización de este tratamiento para la realización de un informe. Sin embargo el amplio tratamiento de datos que ésta permite, nos da una visión mucho más amplia.

Pasemos a estudiar el segundo ejemplo.

HECTÁREAS QUEMADAS. (EL PAIS, 14 de enero de 1997)

La media de hectáreas quemadas por incendio el año pasado es la más baja desde 1970.

Año	Núm. Incend.	Has. Totales	Núm. Has. por incendio
1970	118	13.803	116,97
1971	65	5.571	85,71
1972	57	1.271	22,30
1973	150	4.659	31,06
1974	196	15.727	80,24
1975	249	8.009	32,16
1976	244	6.565	26,91
1977	199	8.712	43,78
1978	560	75.186	134,26
1979	524	90.946	173,56
1980	460	32.403	70,44
1981	727	31.370	43,15
1982	403	13.908	34,51
1983	495	15.579	31,47
1984	497	22.700	45,67
1985	523	39.753	76,01
1986	388	9.640	24,85
1987	377	6.019	15,97
1988	303	849	2,80
1989	405	1.651	4,08
1990	711	27.554	38,75
1991	871	44.426	51,01
1992	804	26.188	32,57
1993	613	25.966	42,36
1994	739	138.775	187,79
1995	478	2.231	4,67
1996	374	731	1,85

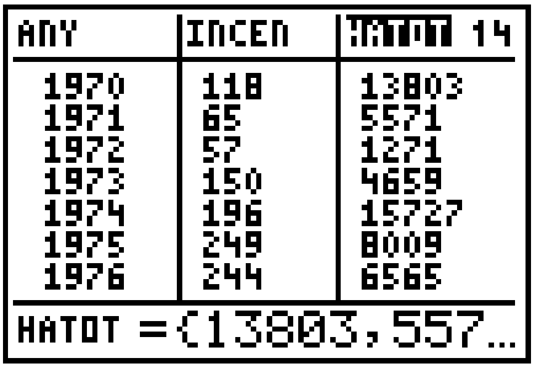
Una utilidad interesante de las TI-83 es la posibilidad de transmitir información de unas calculadoras a otras, cuando está información son datos numerosos es mucho más evidente su interés, pues permite:

1. Al profesorado llevar los datos preparados y pasárlos al alumnado.
2. Reponer los datos a cualquiera que accidentalmente manipule mal los datos u otra causa que haga necesario volver a introducir los datos.

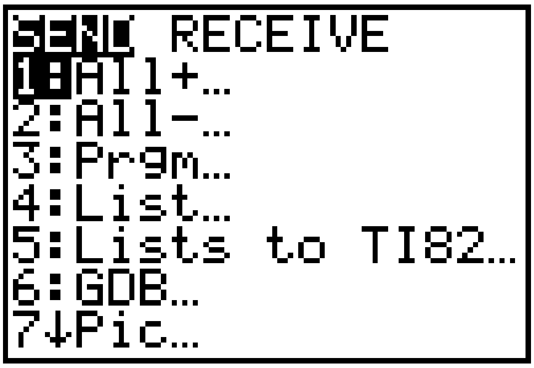
Vamos a analizar con detalle cómo proceder:

4. Conexión y transmisión de datos entre dos TI-83.

En la calculadora emisora tenemos los datos de la tabla y los queremos pasar a otra calculadora:

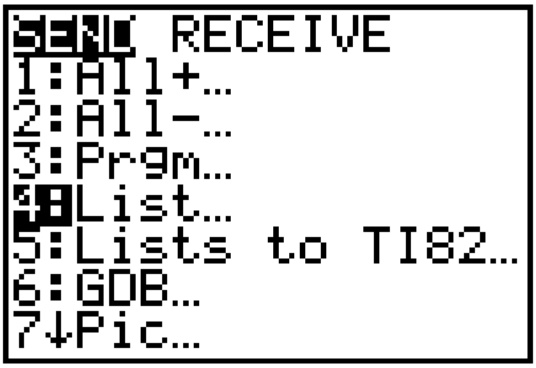


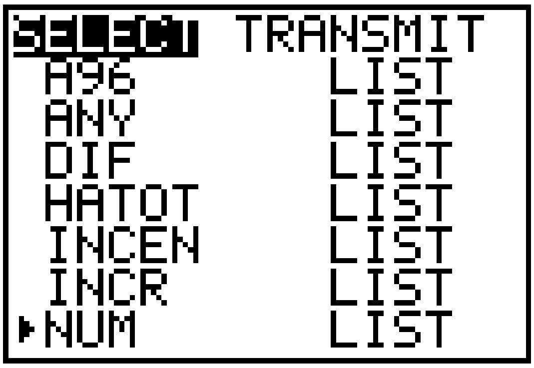
Conectamos las calculadoras con su cable de conexión y pulsamos en ambas 2nd LINK.



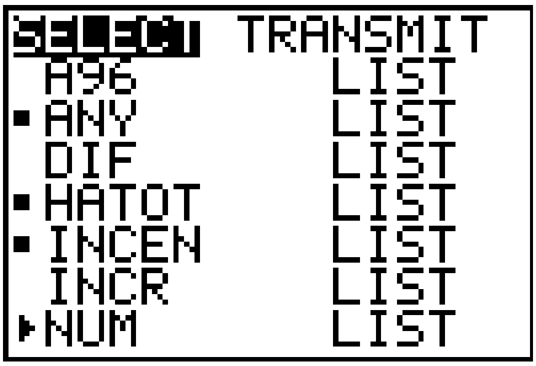
Vemos en la pantalla dos menús, el SEND que será el que utilice la calculadora emisora y el RECEIVE para la calculadora receptora.

La calculadora emisora toma la opción 4:List (que nos permitirá seleccionar, una a una, la información a enviar)

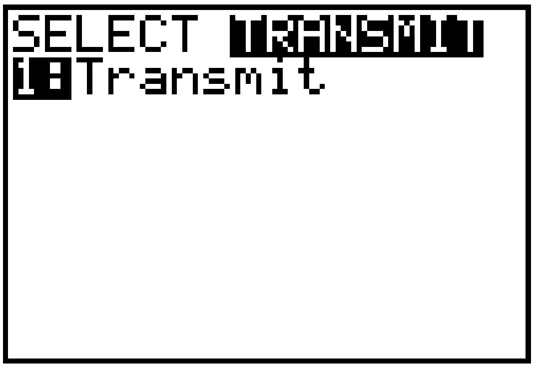




Con las flechas de desplazamiento del cursor seleccionamos la información a transmitir, en nuestro caso ANY, INCEN y HATOT (para indicar la selección es necesario pulsar ENTER cuando la flecha indique el elemento deseado)



En este momento con la calculadora receptora hay que seleccionar el menú RECEIVE, y aceptar la única opción con lo que la calculadora se queda esperando apareciendo Waiting … en pantalla.

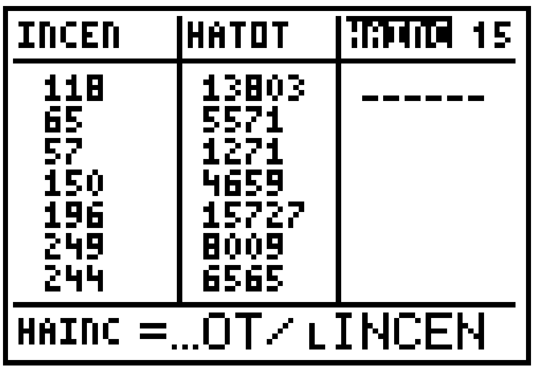


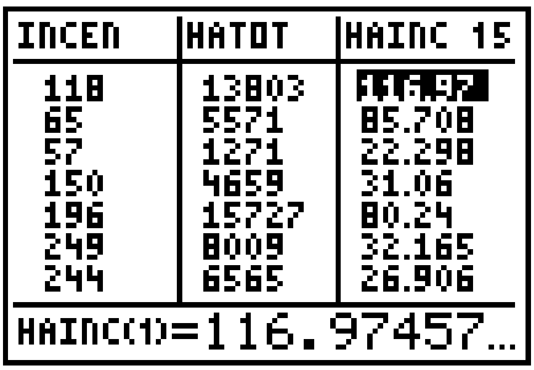
En ese momento la calculadora emisora seleccionará el submenú TRANSMIT

Con lo que si no hay problemas de conexión, hay que apretar bien los cables de conexión, tanto en la calculadora emisora como en la receptora aparecerán las listas enviadas y la confirmación Done.
Si la calculadora receptora ya tuviera listas con los mismos nombres lo indicará –la opción 2 overwrite, sobreescribir– permite copiar los datos encima.
En ese momento ya tenemos en ambas calculadoras las 3 listas de datos iguales como podemos comprobar seleccionando el editor de listas de la calculadora receptora.

Vamos pues a estudiar los datos introducidos de incendios en los últimos 27 años. En este ejemplo utilizaremos la calculadora sin una explicación exhaustiva de como acceder a los diferentes menús, submenús y opciones.

Introducimos en HAINC el número de hectáreas quemadas por incendio por medio de la fórmula HATOT/INCEN





Opciones de cálculo unidimensional para la lista INCEN:

```
1-Var Stats ʟINC
EN
```

```
1-Var Stats
 x̄=427.037037
 Σx=11530
 Σx²=6249064
 Sx=225.7744759
 σx=221.5540201
↓n=27
```

```
1-Var Stats
↑n=27
 minX=57
 Q1=244
 Med=405
 Q3=560
 maxX=871
```

Se puede ver una media de 427 incendios por año con una gran dispersión de 221. Y un rango intercuartílico comprendido entre 244 y 560 incendios como datos que abarcan un 50 % de los observados.

Cálculo unidimensional para la lista HATOT de número de hectáreas quemadas:

```
1-Var Stats ʟHAT
OT
```

```
1-Var Stats
 x̄=24821.92593
 Σx=670192
 Σx²=4.26548E10
 Sx=31634.52867
 σx=31043.17693
↓n=27
```

```
1-Var Stats
↑n=27
 minX=731
 Q1=5571
 Med=13908
 Q3=31370
 maxX=138775
```

Podemos apreciar que la media de hectáreas anualmente quemadas es de 21822 y que el rango intercuartílico se sitúa entre5571 y 27554, siendo la mediana de 13803 hectáreas. Vayamos ahora a HAINC:

```
1-Var Stats ʟHAI
NC
```

```
1-Var Stats
 x̄=53.8887472
 Σx=1454.999618
 Σx²=140767.794
 Sx=48.97389484
 σx=48.05841421
↓n=27
```

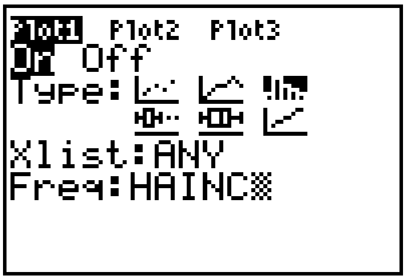
```
1-Var Stats
↑n=27
 minX=1.9545455
 Q1=24.84536082
 Med=38.7538678
 Q3=76.00956023
 maxX=187.78755
```

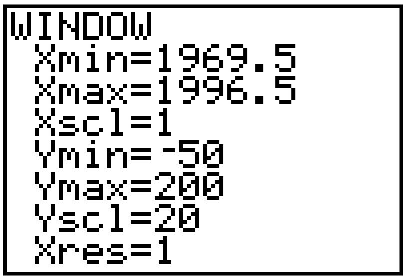
Las hectáreas quemadas por termino medio por incendio es de 53.89 estando el rango intercuartílico entre 24.85 y 76.01. Con una mediana de 38.75 hectáreas quemadas.

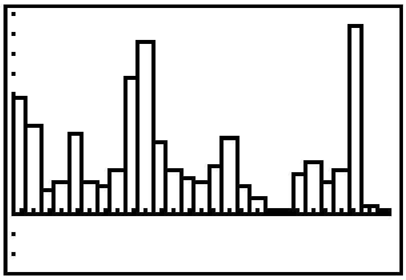
Se puede por tanto ver el dato tan fabuloso obtenido de los últimos 2 años 1995 y 1996 en el que el nº de hectáreas por incendio se sitúa muy por debajo del primer cuartil.

Pasemos a las opciones gráficas para la analizar la lista HAINC con mayor detalle:

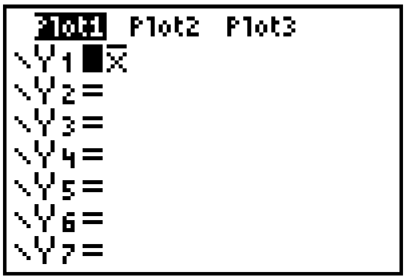
Determinemos la representación que deseamos

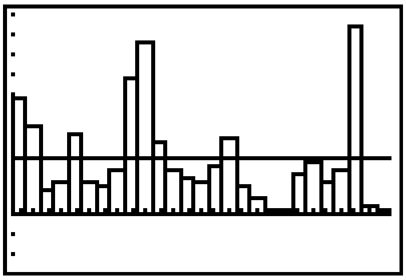
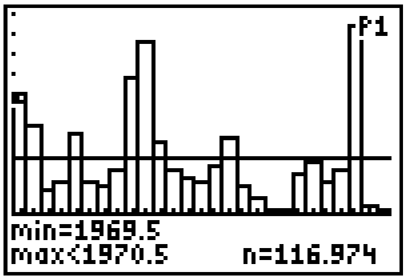






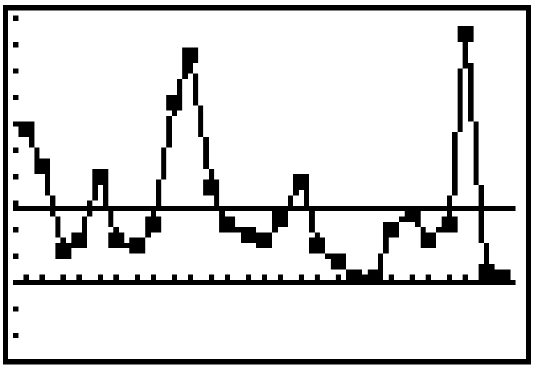
Y si queremos ver los resultados comparándolos con la media:



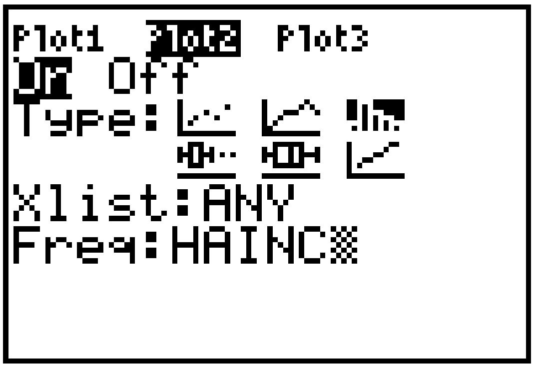



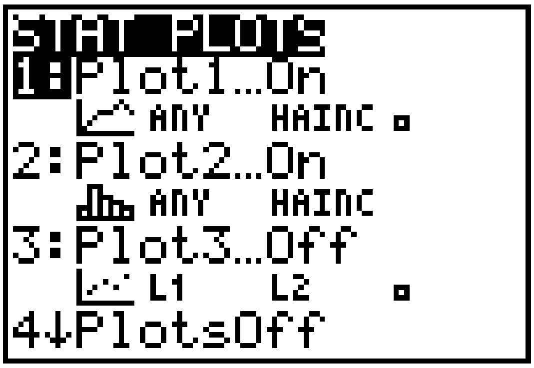
Podemos ver los valores superiores fundamentalmente el último, que fue el aciago 1994, así como los de los años 1978 y 1979. En el lado positivo los dos últimos años son apenas perceptibles, también resaltan como muy buenos los años 88 y 89.

Hagamos otra selección gráfica para ANY y HAINC.

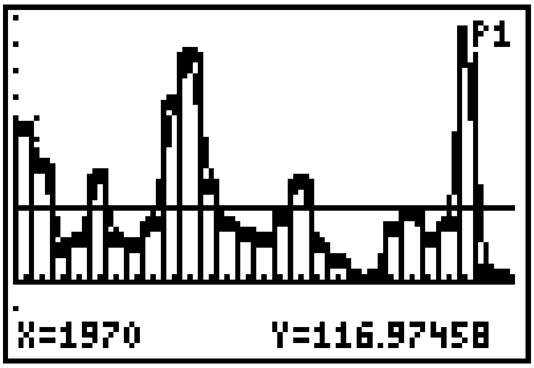


Y también podemos tener activas dos opciones de representación simultáneas.

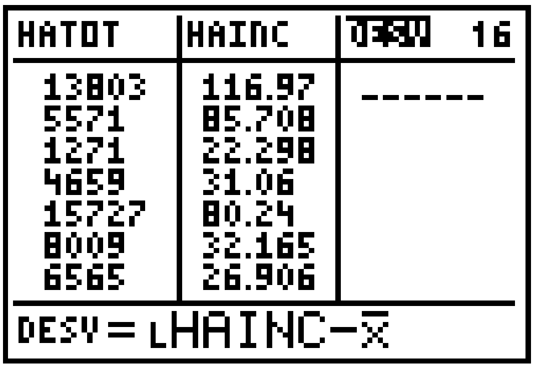


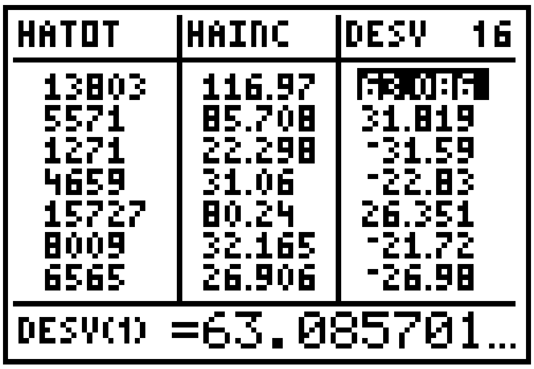


De forma que al representar se superponen:



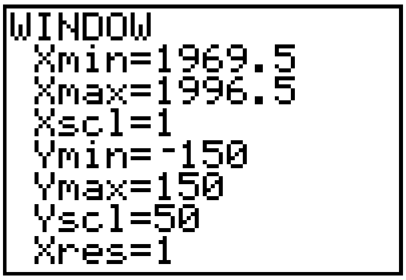
Podemos definir una nueva lista DESV =HAINC - $\bar{x}_{HAINC}$

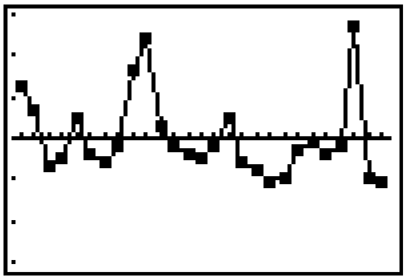




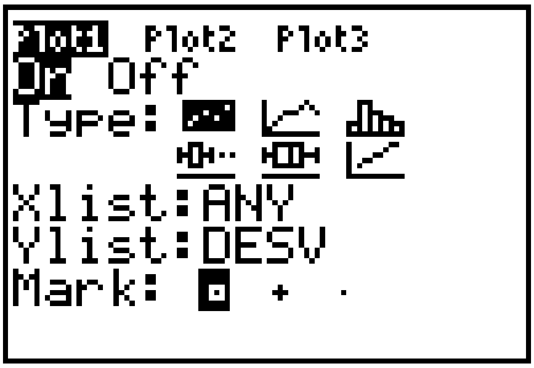
Ponemos los Plots en OFF

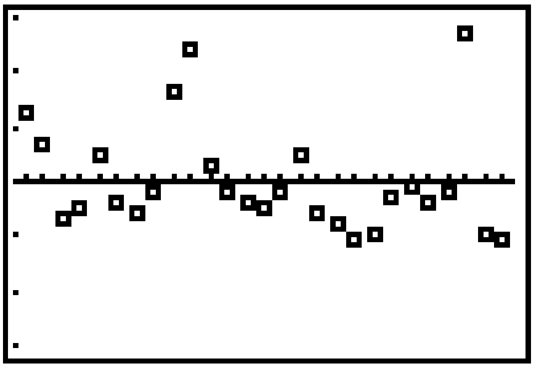
Y elegimos la representación:



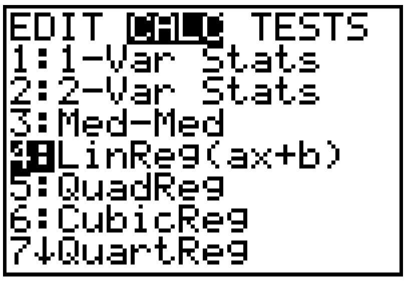


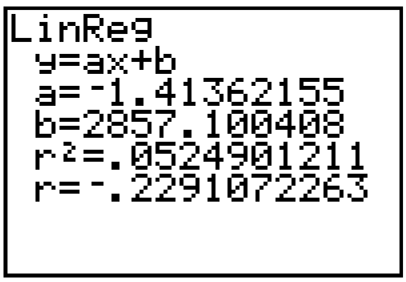
Si queremos verlo sin los segmentos de unión:





Y aunque nos salgamos un poco del tema de la sesión de estadística descriptiva (sólo en cuanto a lo que es posible hacer sin usar calculadora gráfica) puesto que los puntos que representan las diferencias respecto de la media parece que llevan (exceptuando el aciago 1994) un ritmo descendente podríamos ver la correlación entre las lista ANY (años) y por ejemplo HAINC (hectáreas por incendio) para ello seleccionamos la opción CALC de STAT y tomamos la 4:LinReg (ax+b)





Y, claro, da una correlación tan baja que no merece la pena seguir.

5. Regresión y correlación

Los datos estadísticos se representan frecuentemente en tablas, con líneas u otras formas geométricas y con fórmulas algebraicas. Aunque el análisis de datos se ha considerado tradicionalmente parte de la estadística, la búsqueda de un modelo matemático para pares de datos, lo que se conoce normalmente como ajuste de curvas puede ser estudiado en álgebra y geometría.

El ajuste de curvas constituye un área de las matemáticas que se utiliza con frecuencia en el mundo real. Un pediatra, por ejemplo, tiene que poder predecir cuánto debe medir y pesar un niño de acuerdo a su edad. Una compañía eléctrica tiene que estimar la demanda de corriente según las condiciones climáticas, en especial la temperatura. El ajuste de curvas proporciona al profesorado de álgebra y geometría tanto la oportunidad de presentar las matemáticas como algo con significado, como un contexto rico para elaborar conexiones matemáticas.

VELOCIDAD

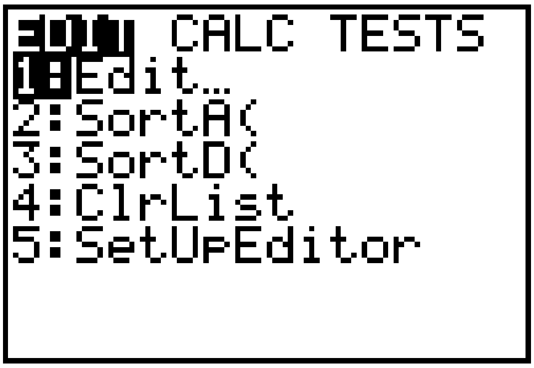
La tabla que tienes más abajo, representa la carrera de Carl Lewis en Tokio el 25 de julio de 1991.

a) Haz un diagrama de puntos con los tiempos de paso cada 10 m.
b) ¿Se observa tendencia lineal?
c) Intenta construir una recta que pase por el mayor número de puntos posible.
d) ¿Podríamos utilizar esta recta para predecir cuál será el tiempo de Carl Lewis en una carrera de 200 m.—400m.—? ¿Calcularlo?
e) Comprueba esta predicción con la marca real que encontrarás en la tabla de marcas olímpicas.
f) Comenta el resultado.

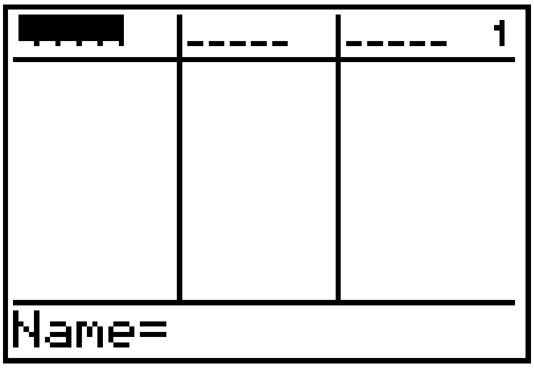
METROS	T. DE PASO
10	1,88
20	2,96
30	3,88
40	4,77
50	5,61
60	6,46
70	7,30
80	8,13
90	9,00
100	9,86

Introducción de datos

Antes de empezar la sesión para asegurarnos de tener las listas vacías , pulsando las teclas [ALPHA] y [+] llegaremos al menú MEMORY seleccionaremos la opción 4:ClrAllLists y pulsaremos [ENTER].

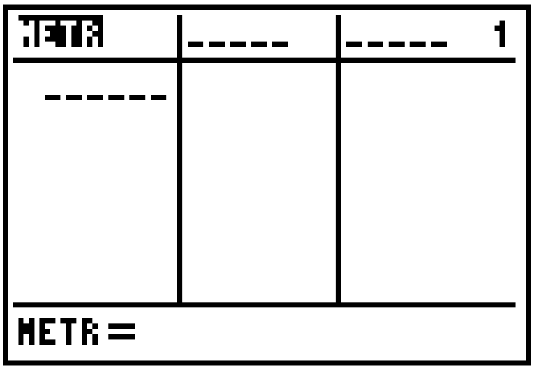


La introducción de datos para el cálculo de regresiones y correlaciones en la TI-83 se realiza a través del menú [STAT], submenú Edit, opción 1:Edit y después pulsaremos [ENTER].



Una vez situados en la edición de las listas:

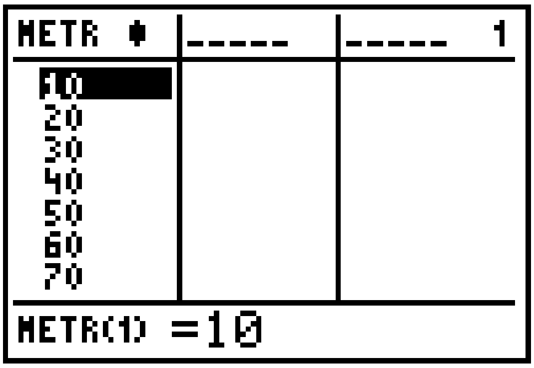
1.- Pondremos nombre a una de las listas (máximo 5 letras) y pulsaremos [ENTER].

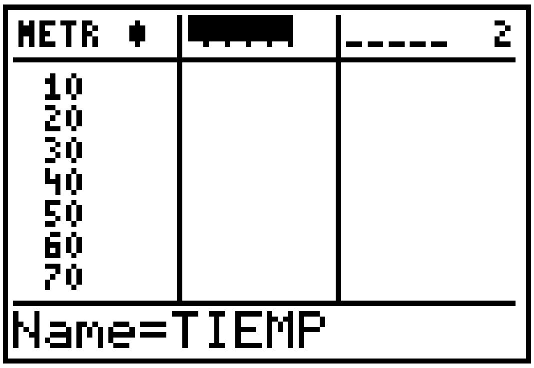


2.- Como la primera lista METR es una secuencia podemos utilizar el siguiente método para introducirla:

Una vez introducido el nombre de la lista pulsamos [ENTER] para pasar al contexto de edición de elementos y para seleccionar la opción seq(pulsaremos: [ALPHA] [+] [2nd] [STAT] [▶] hasta OPS y luego 5:seq([ENTER] ahora acabaremos de escribir la orden: "seq([ALPHA] [x^2] [,] [ALPHA] [x^2] [,] [1] [0] [,] [1] [0] [0] [,] [1] [0] [)] y pulsaremos [ENTER].

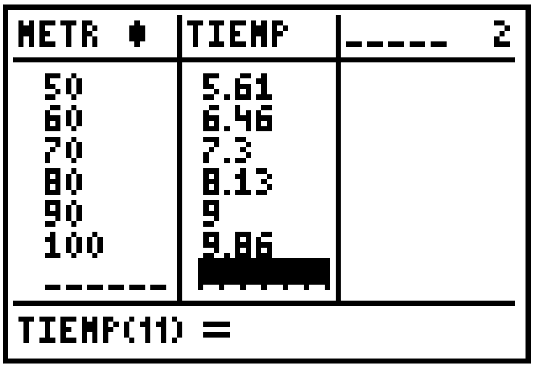
Con lo cual habremos anexado la fórmula a la lista





3.- Pondremos nombre a la otra lista :

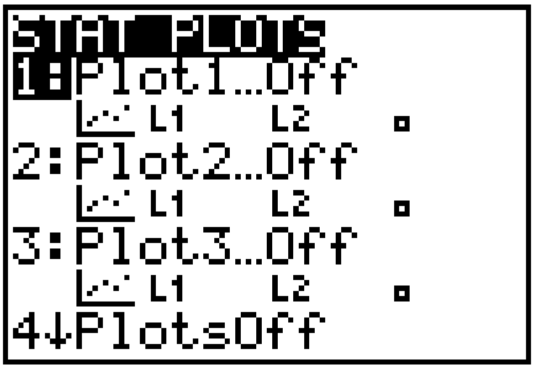
▲ ▶ ENTER TIEMP y ENTER.



4.- Pulsaremos ▼ para bajar al contexto de visualización elementos e introduciremos los valores de tabla pulsando detrás de cada uno ENTER.

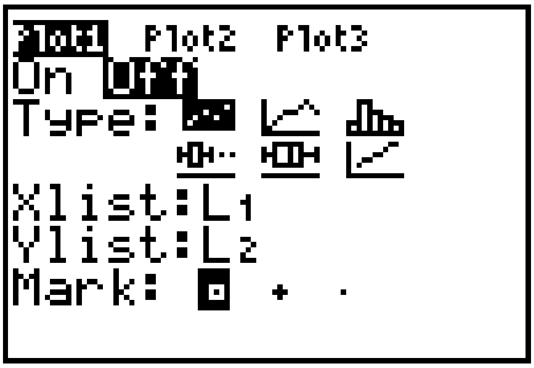
Dibujo del diagrama

5.- Una vez tengamos los datos en las dos tablas pasaremos a representar los datos, pero primero anularemos la selección de ecuaciones Y=.



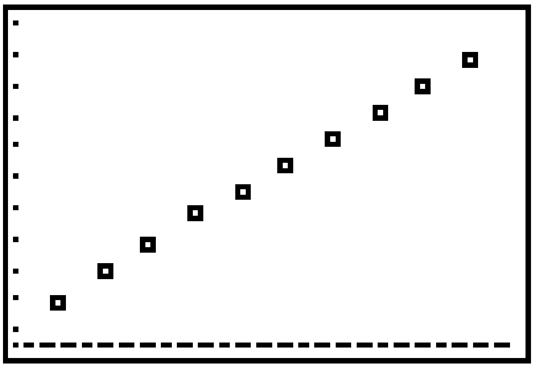
6.-Ahora iremos al menú STAT PLOT: 2nd Y=

pulsaremos ENTER y pasaremos a definir el gráfico estadístico que vamos a utilizar.



7.- Seleccionaremos las opciones On, LOG y en

Xlist: 2nd STAT 1:METR ENTER
Ylist: 2nd STAT 2:TIEMP ENTER, ahora ▼ y seleccionamos el tipo de marca
ð, + ó · y ENTER.



8.- Para garantizarnos que la visualización es la correcta:

ZOOM 9 ENTER :

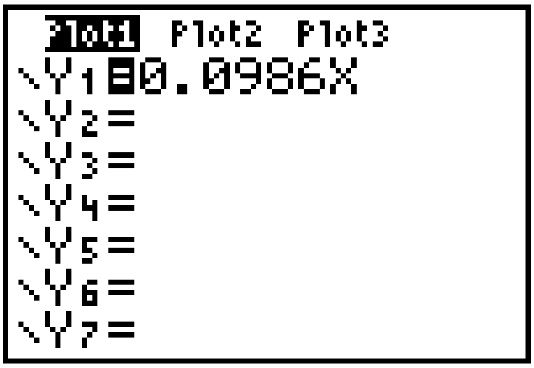
Análisis de los resultados:

Como se puede apreciar, si que se observa tendencia lineal y un modo de construir una recta, ya que sale con metros 0 y tiempo 0, es decir pasa por el (0'0), sería:

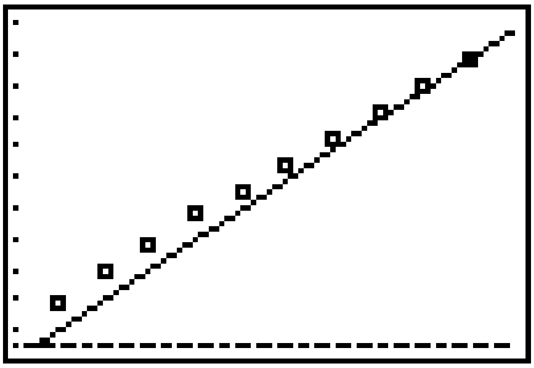
y = kx de forma que pase por el (100, 9.86),
es decir 9.86 = k100, entonces k = 9.86/100 = 0.0986,
con lo que y = 0.0986x

Para introducir estos valores, vamos al editor de ecuaciones [Y=] y en Y_1= ponemos la ecuación de nuestra recta:

∴ Y_1=0.0986[X,T,Θ,n]



Y ahora para visualizarla [GRAPH]



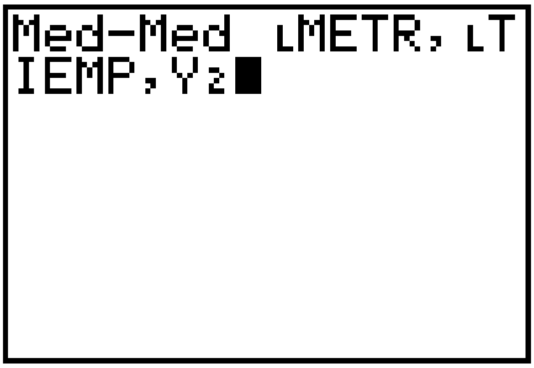
Para obtener cual es valor que correspondería a 200 metros salimos a la pantalla de operaciones [2nd][MODE] y seleccionamos [VARS][▸] 1: Function [ENTER] 1: Y_1 [ENTER]

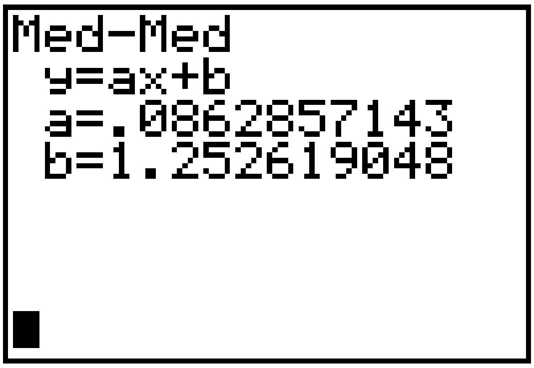
Con lo que obtendremos en la pantalla de operaciones Y_1 , ahora completaremos la expresión: Y_1(200) [ENTER] y obtenemos 19.72.

Pero también podemos utilizar alguna las regresiones del menú STAT CALC :

Med -Med (ax +b)

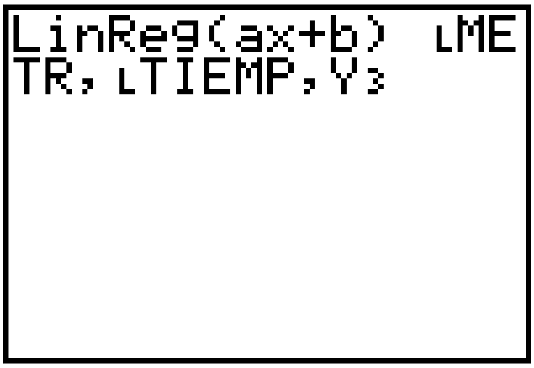
Med-Med ajusta la ecuación de la recta a los datos, utilizando la técnica de la línea mediana-mediana (recta de resistencia). Se dividen los datos en tres grupos de igual tamaño, se calcula la mediana de los tres, se traza una recta que pase por la primera y tercera medianas y se traza una paralela a esta que esté a un tercio de la distancia entre la segunda mediana y la primera recta.

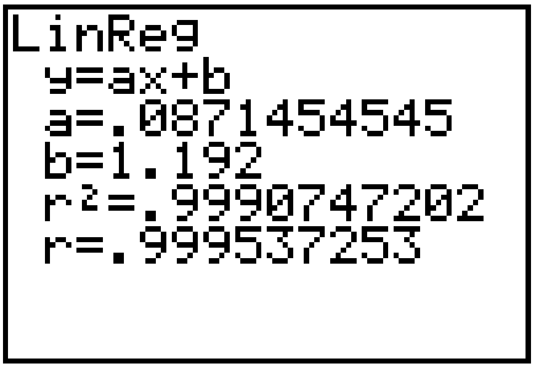




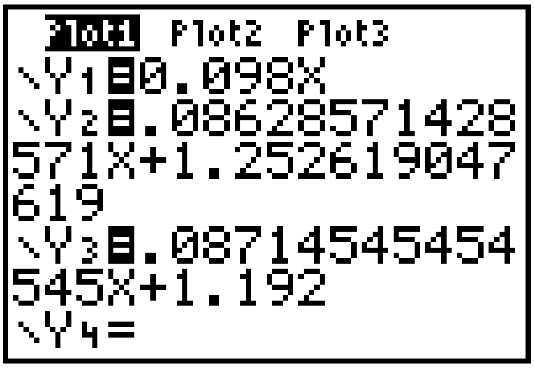
LinReg (ax+b)

LinReg ajusta la recta a los datos utilizando el método de los mínimos cuadrados. Si se establece el modo DiagnosticOn , se visualizaran r (coeficiente de correlación) y r^2 (coeficiente de determinación).

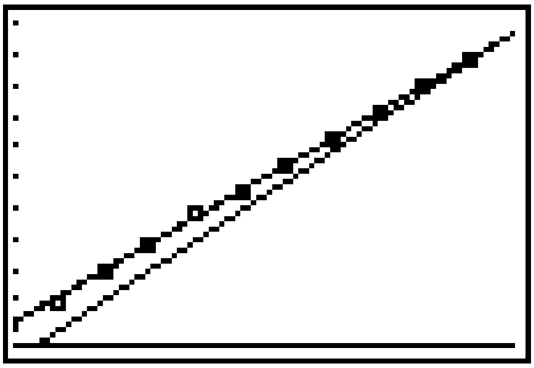




Comprobamos las ecuaciones en el editor de funciones [Y=]

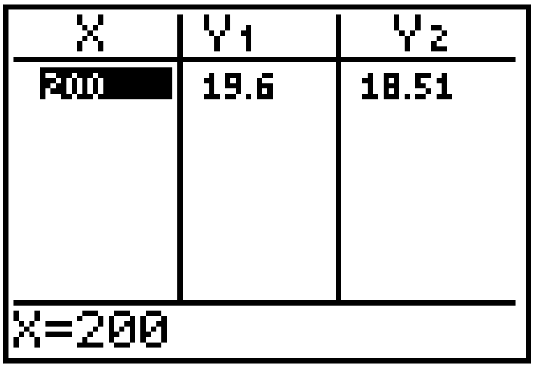


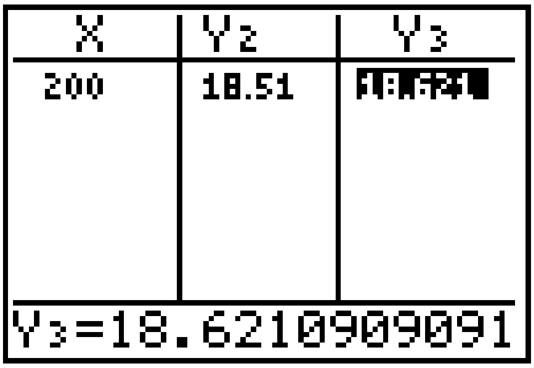
Y ahora las visualizamos con GRAPH



Y calculamos el valor mediante estas rectas para 200 metros colocando en TABLE SETUP Indpnt : Ask

Y poniendo 200 en la columna de las X en TABLE





Si comprobamos la marca olímpica comprobaremos que fue 20.01 lo que significa que para una carrera tan corta hemos tenido una desviación considerable debido a que al utilizar los datos de la carrera de los 100 metros hemos supuesto que la progresión seguida en esta carrera era la misma que la que se iba a dar en la de los 200, pero esto no es así ya que al ser una carrera más larga la velocidad máxima se alcanza antes de llegar a la meta y hay un trozo que se mantiene o incluso disminuye, por lo que deberíamos concluir que no debimos utilizar la recta construida con los datos de la carrera de 100 metros para extrapolar el resultado de la 200.

ALTURA Y NÚMERO DE CALZADO

1.- Anota tu número de zapatos y tu altura en centímetros en una hoja, Anota los resultados de toda la clase en la tabla siguiente:

	Altura	Nº calzado	Nº calzado previsto	Magnitud del error
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				

2.- Ahora representa los datos anteriores en diagrama de puntos (altura, número de calzado).

3.- ¿Parece que la mayoría de los puntos estén cerca de una recta?

4.- Traza una recta por en medio de la nube de puntos. Puedes utilizar la recta Med-med o la de regresión lineal que halla tu calculadora.

5.- Esta ecuación es un modelo algebraico de la relación entre la altura y el número de calzado.

Utiliza este modelo para calcular cuál debe ser tu número de calzado. Para ello coloca tu altura en la ecuación en lugar de x y redondea el resultado en ± 5 décimas.

¿Qué número de calzado te ha salido? ____

¿Cuánto se ha acercado este número al número de calzado que tienes en realidad?_____

6.- Utilizando esta ecuación, calcula el número de calzado de tus compañeros de clase. Redondea siempre en ± 5 décimas. Anota los resultados en la tercera columna de la tabla. En la cuarta columna, anota la diferencia entre los números que te hayan salido y el número de calzado que realmente tenga cada uno de tus compañeros.

7.- Según el *Libro Guinness de los Records*, la persona más alta que haya existido nunca fue Robert Wadlow, que media 2,72 metros.

¿Qué número de calzado da tu ecuación para él?______

La mujer viva más alta es Sandy Allen, que mide 2,31 metros.

¿Qué número de calzado da tu ecuación para ella ?______

8. ¿Qué número de calzado predice tu ecuación que tendría una persona de 1,27 metros ?
¿Crees que esta predicción es razonable? ¿Por qué sí o por qué no?

9. ¿Crees que sería razonable vender zapatos según la altura del cliente? ¿Por qué sí o por qué no?

CARRERAS DE 1500 m.

La tabla que te presentamos a continuación recoge las marcas olímpicas de la carrera de 1500 m. desde 1896 a 1992:

AÑO	MARCA (min.)
1896	4.553
1900	4.103
1904	4.09
1908	4.057
1912	3.947
1920	4.03
1924	3.893
1928	3.887
1932	3.853
1936	3.797
1948	3.83
1952	3.752
1956	3.687
1960	3.593
1964	3.635
1968	3.582
1972	3.605
1976	3.653
1980	3.64
1984	3.542
1988	3.599
1992	3.678

* Dibuja un diagrama de puntos que represente la información.

* ¿Hay tendencia lineal?

* Dibuja una recta que represente la nube de puntos.

* Si te fijas te darás cuenta que faltan las marcas de algunos años, podrías predecir cuál hubiera sido la marca.

* Después de cada salto hay un empeoramiento de la marca , Cómo lo explicas.

* Intenta predecir el tiempo de carrera para un atleta del año 3050. Comenta el resultado.

6. Algunas precauciones con la regresión y la correlación

Los análisis de regresión y de correlación son herramientas estadísticas bastante útiles cuando se utilizan apropiadamente. Sin embargo, su uso inapropiado sólo puede conducir a la obtención de resultados sin sentido. Por lo cual hay que hacer una serie de consideraciones a la hora de utilizar estas técnicas.

1. Antes de reunir los datos, deben revisarse cuidadosamente las suposiciones que fundamentan los análisis de regresión y de correlación. Aunque es raro encontrar que se cumplan las suposiciones a la perfección, se debe tener alguna idea acerca de la magnitud de la brecha que existe entre los datos que van a analizarse y las suposiciones del modelo propuesto, de modo que pueda decidir si debe elegir otro modelo; procédase con el análisis, pero tomando precauciones con la interpretación de los resultados; o bien, utilícese con confianza el modelo elegido.

2. En la regresión lineal simple y el análisis de correlaciones las dos variables de interés se miden sobre la misma entidad, llamada unidad de asociación. Si se tiene interés en la relación entre la estatura y el peso, por ejemplo, estas dos medidas se hacen en el mismo individuo. En general, carece de sentido hablar de correlación, por así decirlo, entre las estaturas de un grupo de individuos y el peso de otro grupo.

3. Sin importar qué tan grande es la indicación de una relación entre dos variables, no debe interpretarse esto como un caso de causa y efecto. Si, por ejemplo, se observa un coeficiente notable de correlación de la muestra entre las dos variables X y Y. puede significar una de varias cosas:

a) X causa Y.
b) Y causa X.
c) un tercer factor, sea directa o indirectamente, causa tanto a X como a Y
d) Ha ocurrido un evento improbable y se ha obtenido por casualidad un gran coeficiente de correlación de la muestra a partir de una población en la que, en efecto, X y Y no están correlacionadas.
e) La correlación es sencillamente disparatada, una situación que puede sura)

4. La ecuación de regresión de la muestra no debe utilizarse para predecir o estimar fuera del intervalo de valores de la variable independiente

representado en la muestra. Esta práctica, llamada extrapolación, tiene sus riesgos. La verdadera relación entre dos variables, aun cuando sea lineal sobre un intervalo de la variable independiente, a veces puede describirse mejor como una curva fuera de este intervalo. Si. por casualidad, se extrae la muestra sólo del intervalo donde la relación es lineal, se tiene únicamente una representación limitada de la población, por lo que proyectar los resultados de la muestra más allá del intervalo representado por ella puede conducir a conclusiones falsas.

7. Resumen de regresión y correlación

Hemos examinado dos importantes herramientas del análisis estadístico, la regresión y la correlación. Se sugiere el siguiente esquema para la aplicación de dichas técnicas.

1. *Identificación del modelo*. Se debe saber si el modelo de regresión o el de correlación es el apropiado para dar respuesta a sus preguntas.

2. *Revisión de las suposiciones*. La validez de las conclusiones depende de que también se ajusten los datos analizados al modelo elegido.

3. *Obtención de la ecuación de regresión*. Las calculadoras gráficas, evitan la realización de largos y tediosos cálculos para su obtención.

4. *Evaluación de la ecuación*. La utilidad de la ecuación de regresión para fines de estimación y predicción se determina por medio del análisis de varianza, el cual prueba el significado del cuadrado medio de la regresión. Se valora la intensidad de la relación entre dos variables bajo el modelo de correlación probando la hipótesis nula de que no existe correlación en la población. Si puede rechazarse esta hipótesis, puede concluirse, al nivel de significación elegido. que las dos variables están correlacionadas.

5. *Uso de la ecuación.* Una vez que se ha determinado que es probable que la ecuación de regresión describa adecuadamente la relación entre las dos variables, X y Y. puede utilizarse para cualquiera de fines:

a) predecir qué valor es probable que tenga Y, dado un valor particular de X, o bien,
b) estimar la media de la subpoblación de los valores de Y para un - valor particular de X.

8. Bibliografía

ALSINA CATALÀ, C. *¿Para qué aspectos concretos de la vida deben preparar las matemáticas?. UNO*, 1994, 1, 37-43.

BURRILL, G.(1992). *Data analisis and statistitcs acorss the curriculum* (Curriculum and evaluation standars for school mathematics addenda series. Grades 9-12). (NCTM. Reston, Virginia).

CABALLERO, S., GRILLES, M. Y MONZÓ, O. (1992). *Probabilidad, Estadística, Combinatoria* (3er curso de E.S.O.). Colección: Materiales para el desarrollo curricular. M26. (Generalitat Valenciana. Conselleria de Cultura Educació i Ciència)

CABALLERO, S., Y MONZÓ, O. (1993*). Probabilidad, Estadística, Combinatoria* (4º curso de E.S.O.). Colección: Materiales para el desarrollo curricular.M28 (Generalitat Valenciana. Conselleria de Cultura Educació i Ciència)

DANIEL, W. W. (1993). *Bioestadística.* (Editorial LIMUSA. México D.F.)

DEGROOT, M. H. (1988). *Probabilidad y Estadística*. (Addison-Wesley Iberoamericana, S.A.: Wilmington)

ENGEL, A. (1988). *Probabilidad y Estadística*. 2 Vol. (Mestral: Valencia)

GIL, D. Y DE GUZMAN, M. (1993). *Enseñanza de las ciencias y la matemática*. (Editorial Popular, S.A.. Madrid)

GRILLES, J.M. Y PLA, F. (1991). *Probabilidad y Estadística* (Materiales bachillerato experimental). (Generalitat Valenciana)

MONZÓ, O. Y QUERALT, T. (1995) *La inferencia estadística en los bachilleratos*. Actas 7ª JAEM, págs. 114-119. (SMPM Emma Castelnuovo. Madrid)

N.C.T.M. (1991*). Estándares curriculares y de evaluación para la educación matemática.* (SAEM "Thales". Sevilla).

SANTALÓ, L. A. (1977). *La educación matemática hoy.* (Teide. Barcelona).