Tema 11.
ESTADÍSTICA BIDIMENSIONAL

- Introducción
- Cálculos estadísticos.
- Representación gráfica de datos estadísticos de dos variables.
- Actividades propuestas

INTRODUCCIÓN

La aplicación Estadística permite realizar diversos cálculos estadísticos y representar gráficamente datos estadísticos de variables bidimensionales.

Además, la calculadora lleva incorporadas funciones de manipulación de listas que permiten generar fácilmente todas las tablas de frecuencias de una distribución estadística bidimensional, así como estudiar la correlación entre variables. También es posible ver los distintos modelos de regresión para ajustar los tipos de correlación.

Con la aplicación Estadística podemos realizar las siguientes operaciones:

- Introducir y ordenar datos de tipo lista.
- Dibujar gráficos estadísticos de dos variables:
    - De dispersión (Disper.)
    - De línea xy (Línea xy)
    - De regresión lineal (RegrLin)
    - Med- Med (MedMed)
    - Regresión cuadrática (RegrCuadr)
    - Regresión cúbica (RegrCúbic)
    - Regresión de orden cuatro (RegrCuart)
    - Regresión logarítmica (RegrLog)
    - Regresión exponencial (RegrExp)
    - Regresión exponencial (RExp. ab)
    - Regresión sinusoidal (RegrSin)
    - Regresión logística (RegrLogis)
    - Regresión potencial (RegrPot)
- Cálculos estadísticos

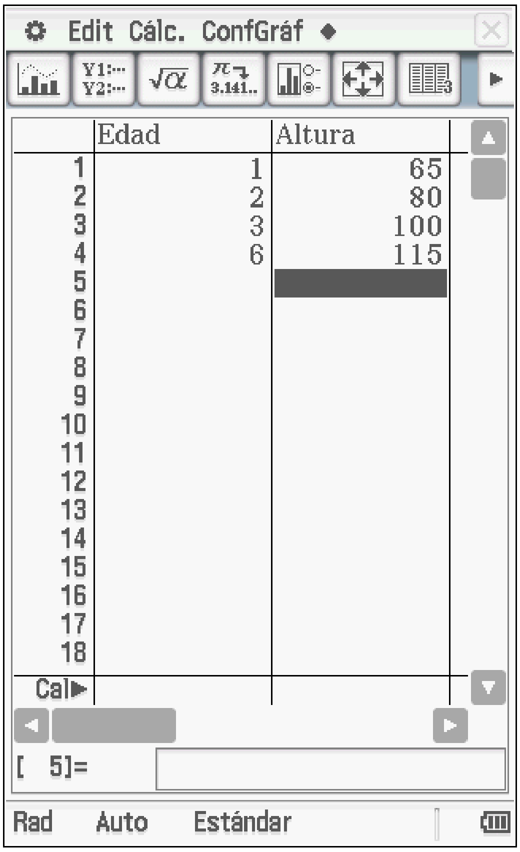
CÁLCULOS ESTADÍSTICOS

Recordemos que para arrancar la aplicación Estadística hay que pulsar **`Estadística`** en el menú de aplicaciones; con esto se muestra la ventana del editor de listas.

Como ejercicio o ejemplo consideremos la distribución que mide la altura en centímetros de diferentes niños

Años	1	2	3	6
Altura	65	80	100	115

Vamos a crear una tabla de datos (mostramos las pantallas que se obtienen, ya que el proceso es el descrito en el tema de Estadística unidimensional)

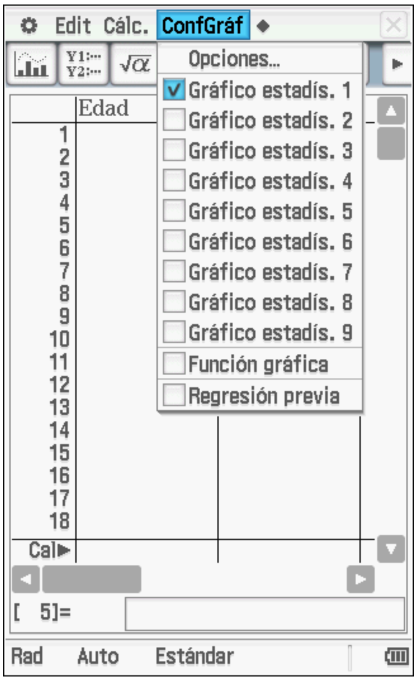


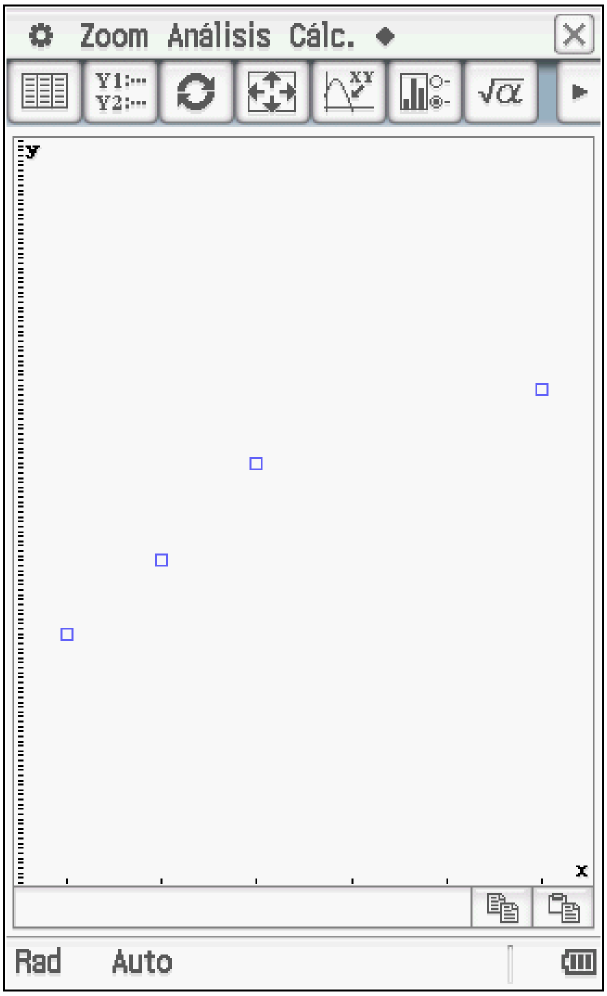
Recordemos que al estudiar distribuciones bidimensionales el objetivo perseguido es determinar si existe relación estadística entre las dos variables consideradas; es decir, ver si los cambios en una de las variables influyen en los cambios de la otra. Cuando sucede ésto, diremos que ambas variables están correlacionadas o que hay correlación entre ellas.

Si las variables crecen conjuntamente, la correlación es directa. Si, por el contrario, al aumentar una de ellas disminuye la otra, la correlación será inversa. La correlación puede calificarse como fuerte cuando el grado de dependencia es alto; y como débil en caso contrario.

El primer paso para determinar el sentido y el grado de la correlación entre dos variables consiste en representar gráficamente, en el plano cartesiano, los pares de valores. Estos gráficos, que reciben el nombre de diagramas de dispersión, permiten visualizar la posición de los datos en el plano. La forma de la nube de puntos asociada a cada diagrama permite establecer conjeturas sobre la correlación existente entre las variables estudiadas.

Para dibujar nuestra nube de puntos:





La forma de una nube de puntos sugiere qué tipo de correlación se da.

- Una nube estrecha y decreciente indica correlación lineal inversa y fuerte.
- Si la nube no adopta una forma definida no hay correlación o es muy débil.
- Una nube ancha y con tendencia a crecer sugiere una correlación lineal directa y débil.
- Si una nube presenta una forma clara pero no rectilínea, la correlación no es lineal, quizá sea exponencial o parabólica.

En general, dependiendo de la forma de la nube de puntos, puede asegurarse:

- Una nube de puntos alargada indica correlación lineal: los puntos se dsitribuyen en torno a una línea recta. La estrechez de la nube expresa que la correlación es fuerte.
- Si la recta que se ajusta a la nube tiene pendiente positiva, la correlación será directa: al crecer la variable X lo hace también la variable Y.
- Una recta con pendiente negativa indica que la correlación es inversa, al crecer X disminuye Y.

En nuestro ejemplo, la nube es alargada lo que indica que hay correlación lineal, en principio; además es directa y parece que fuerte.

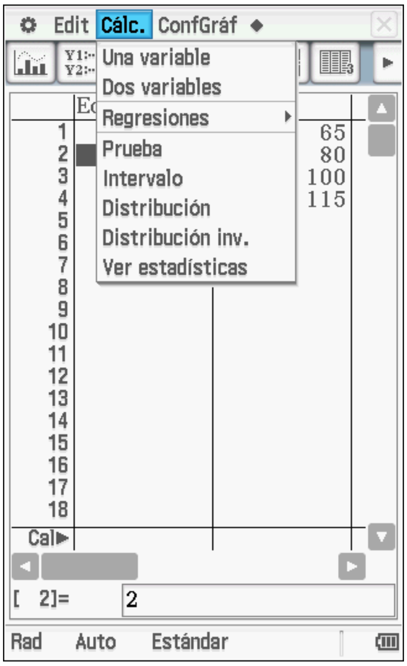
Por otra parte, los datos marginales (datos correspondientes a cada una de las variables) permiten el cálculo de los parámetros marginales de cada una de las variables.

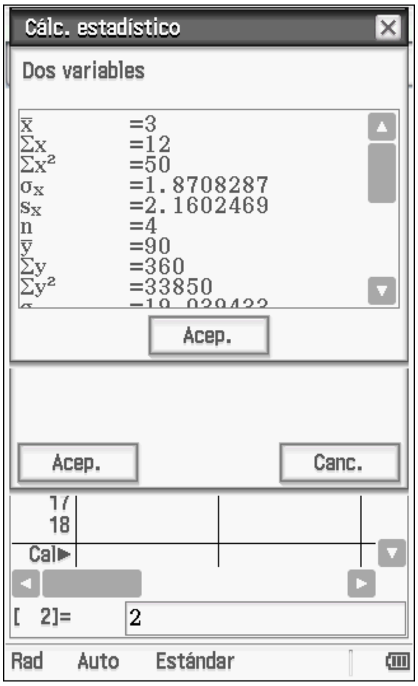
Para ver los resultados de cálculo de dos variables:

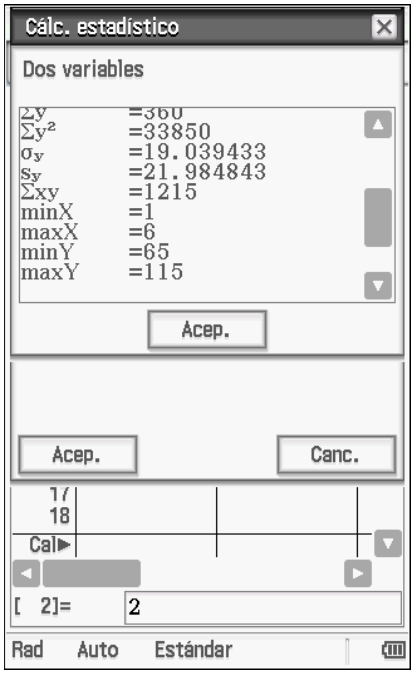
(1) En la barra de menús, toque **[Cálc]** y luego **[Dos Variables]**.

(2) En el cuadro de diálogo que aparece, especifique el nombre **[ListaX]** y el nombre **[ListaY]**.

(3) Pulse **[Acep.]**







Obsérvese que no aparece ningún parámetro estadístico conjunto como la covarianza, que recordemos es la media aritmética de los productos de las diferencias de los valores de cada variable respecto de su media marginal; esto no tiene importancia ya que se puede calcular con los datos anteriores, y sobre todo porque la covarianza no da una medida objetiva (comparable) de la correlación entre variables.

El criterio que se utiliza para medir la fuerza de la correlación entre dos variables es el coeficiente de correlación lineal, r que es la razón entre la covarianza de las variables X e Y y el producto de sus desviaciones típicas marginales. Recordemos alguna de sus propiedades fundamentales:

- El valor de r no cambia al hacerlo la escala de medición.
- Si r>o, la correlación es directa.
- Si r<0, la correlación es inversa.
- El valor de r está entre -1 y +1.
- Si r toma valores cercanos a -1 la correlación es fuerte (e inversa).
- Si r toma valores cercanos a +1 la correlación es fuerte (y directa).
- Si |r|=1, la correlación es perfecta. Hay dependencia lineal entre las variables X e Y.
- Si r toma valores cercanos a 0, la correlación es débil.

Este coeficiente lo calcularemos con la calculadora más adelante.

GRÁFICOS ESTADÍSTICOS DE DOS VARIABLES

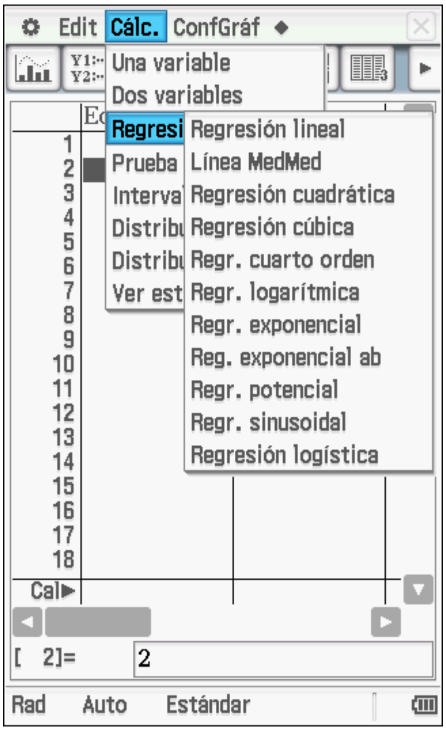
Para sacar el máximo provecho a la correlación vamos a hallar qué recta es la que mejor se ajusta a la nube de puntos; esta recta nos permitirá calcular qué valor de Y es el que cabe esperar para un valor conocido de X. O sea, podemos hacer estimaciones de una variable a partir de la otra.

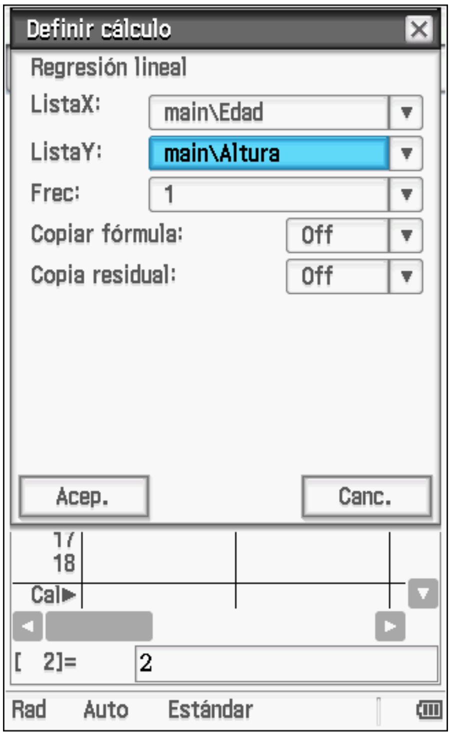
La recta de regresión es la que mejor se ajusta a la nube de puntos. Es una recta que asignaría a cada valor x de X el promedio de los y correspondientes a x. Pero además de la regresión lineal hay otros modelos de regresión (como hemos enumerado al principio).

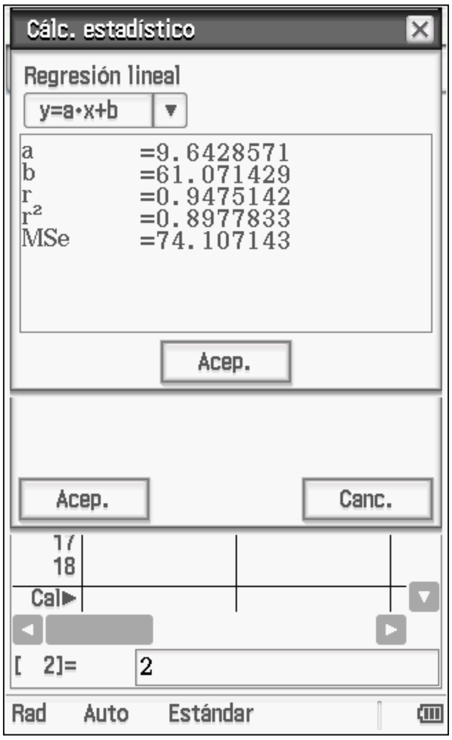
La regresión lineal utiliza el método de mínimos cuadrados para determinar la ecuación que se ajusta a sus puntos de datos, y devuelve valores para la pendiente y la intersección con el eje OY. La representación gráfica de esta relación es un gráfico de regresión lineal.

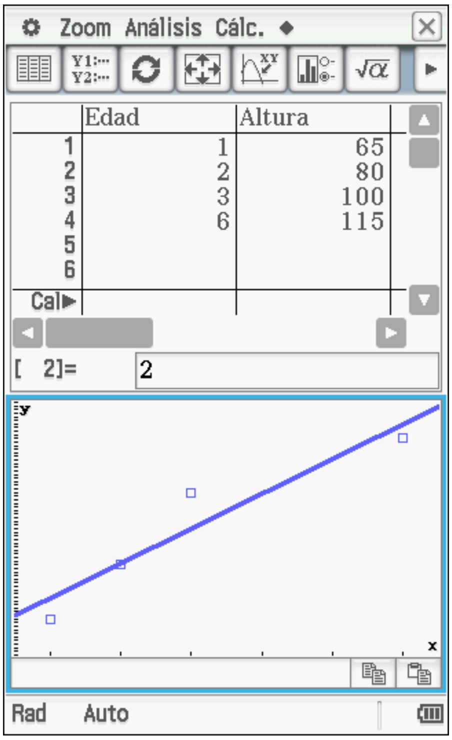
Para obtenerlo:

Desde la ventana de listas o del gráfico de regresión pulse **`[Cálc.]`**, **`[Regresiones]`**, **`[Regresión lineal]`**









Obsérvese que aparece por primera vez el coeficiente de correlación lineal; al ser cercano a +1 la correlación es fuerte y directa.

El modelo de regresión lineal tiene el siguiente significado:

a: Coeficiente de la regresión (pendiente).

b: Corte con el eje Y (término constante de la regresión).

r: Coeficiente de correlación.

r^2 : Coeficiente de determinación.

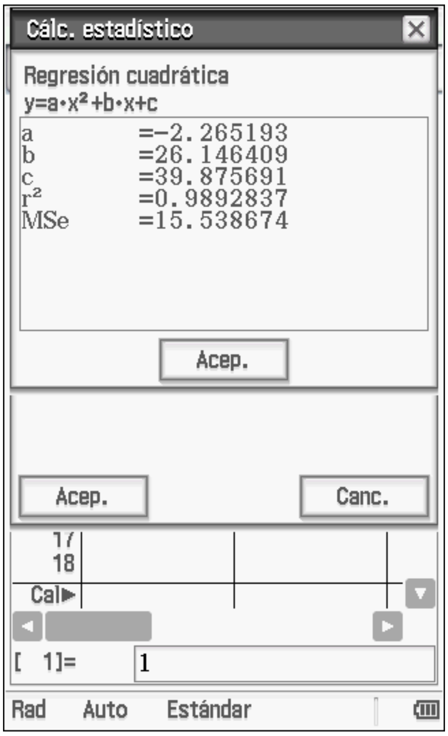
Mse: Error cuadrático medio.

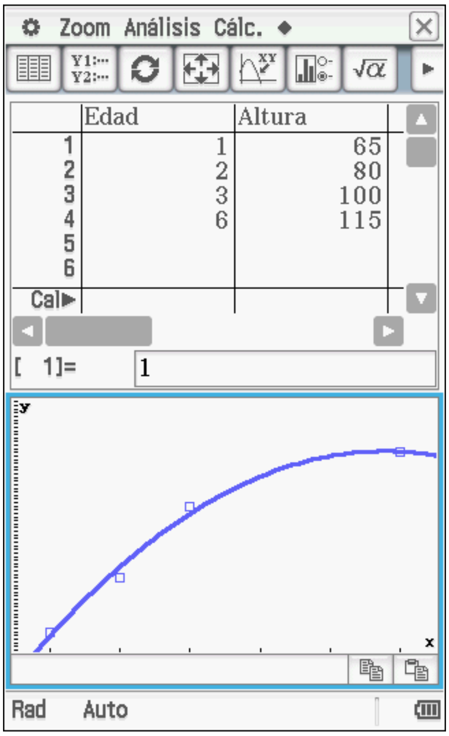
$$MSe = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - (a \cdot x_i + b))^2$$

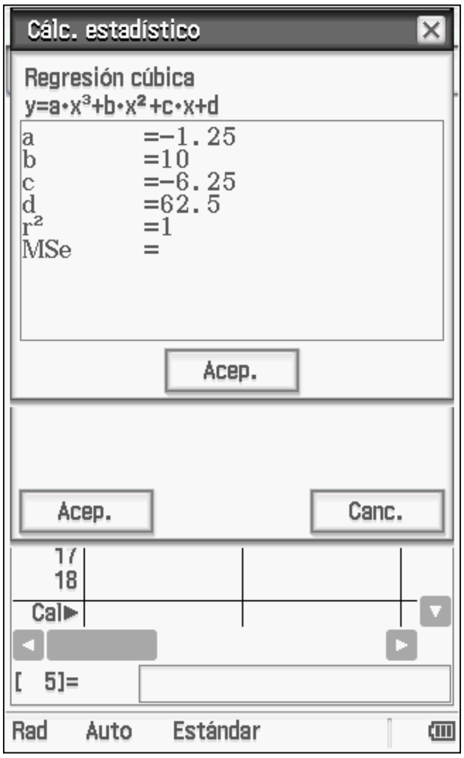
Regresión cuadrática y cúbica.

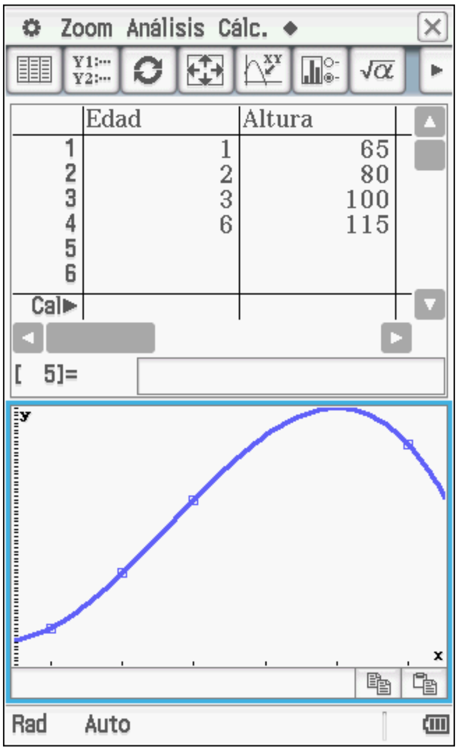
Se puede dibujar un gráfico de regresión cuadrática, cúbica basado en los puntos trazados. Estos gráficos utilizan el método de mínimos cuadrados para dibujar una curva que pase por tantos puntos de datos como sea posible. Estos gráficos pueden expresarse como expresiones de regresión cuadrática, cúbica y de orden cuatro. Para obtenerlos:

Desde la ventana de listas o del gráfico de regresión pulse **`[Cálc.]`**, **`[Regresiones]`**, **`[Regresión cuadrática (cúbica)]`**.









Se pueden ensayar todavía algunas regresiones más, pero el procedimiento es el mismo en todos los casos.

Vamos a dejar el ejemplo que nos ha servido de guía y vamos a ver un tipo de ***regresión especial.***

Cuando sospeche que los datos contienen valores extremos, deberá utilizar el gráfico Med-Med (que se basa en las medianas), en lugar del gráfico de regresión lineal. El gráfico Med-Med es similar al gráfico de regresión lineal, pero también minimiza los efectos de los valores extremos.

Consideremos el siguiente ejemplo

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

Observe este dato. Introduzca los valores en sendas tablas y dibuje la nube de puntos correspondiente; estudie si la recta de regresión es idónea y si no lo es siga los pasos:

Desde la ventana de gráficos o listas:

Pulse **`[Cálc][Línea MedMed]`**.

ACTIVIDADES.

1. Introducir los datos de dos variables mostrados a continuación. Después trace un diagrama de dispersión de los datos y luego conecte los puntos para producir un gráfico de línea xy.

List1	0’5	1’2	2’4	4’0	5’2
List2	–2’1	0’3	1’5	2’0	2’4

2. Introducir los datos de dos variables mostrados a continuación y trace un diagrama de dispersión de los datos. A continuación, realizar una regresión logarítmica de los datos para ver los parámetros de regresión, y luego dibujar el gráfico de regresión.

List1	0’5	1’2	2’4	4’0	5’2
List2	–2’1	0’3	1’5	2’0	2’4

3. En el ejercicio anterior dibuje el gráfico de regresión sin realizar el cálculo de regresión.
4. En una competición de patinaje artístico por parejas se otorgan dos notas: una a los ejercicios obligatorios (X) y otra a los ejercicios libres (Y).

   Las seis parejas que se disputan la final han obtenido los siguientes resultados:

X	5	5	6	7	7	7
Y	5	7	7	7	7	8

a) ¿Qué tipo de correlación hay entre las variables?
b) Dibuje la nube de puntos.
c) Calcule el coeficiente de regresión.
d) Ensaye los diferentes tipos de regresión que hemos visto en el tema.