Estadística Bidimensional

José Mª Chacón Íñigo
I.E.S. Llanes, Sevilla

En esta nueva entrega de la sección "TIC" se muestran opciones y se abordan programas para trabajar en el aula la enseñanza y el estudio de Estadística bidimensional en el bachillerato. Se van a proponer sencillas actividades que se pueden realizar con software libre, con software comercial y con calculadoras gráficas (sin describir los programas ni comandos e instrucciones para utilizarlos).

PROPUESTAS PARA TRABAJAR CON SOFTWARE LIBRE

Gnumeric

Es una hoja de cálculo muy buena y muy fácil de usar. (Existe versión para Linux y para Windows; en Andalucía viene con las distribuciones de Guadalinex en los Centros Tics).

Dispone de una potente herramienta **"Análisis estadístico"**. Veamos su aplicación en la siguiente actividad.

En la tabla se muestran datos de automóviles de ocasión con respecto al kilometraje y precios de los modelos de cierta marca:

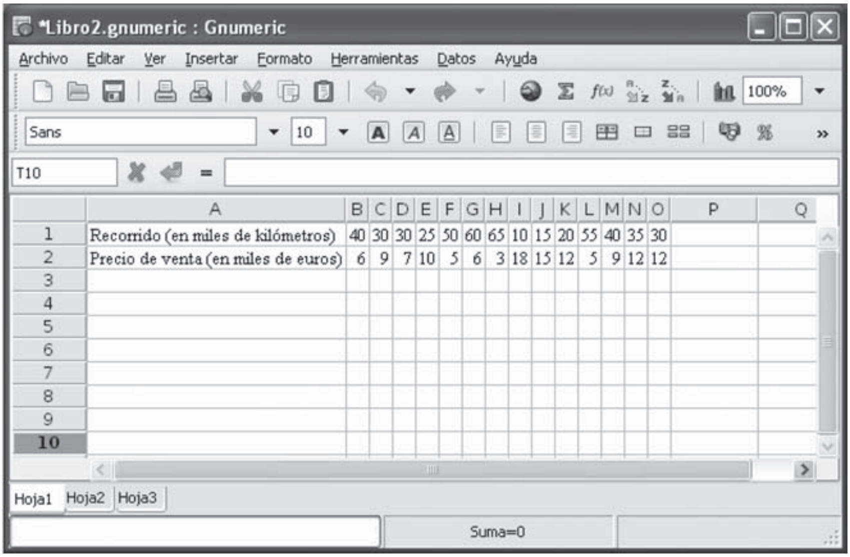
Recorrido (en miles de km.)	40	30	30	25	50	60	65	10	15	20	55	40	35	30
Precio de venta (en miles de euros)	6	9	7	10	5	6	3	18	15	12	5	9	12	12

a) Construye el diagrama de dispersión.

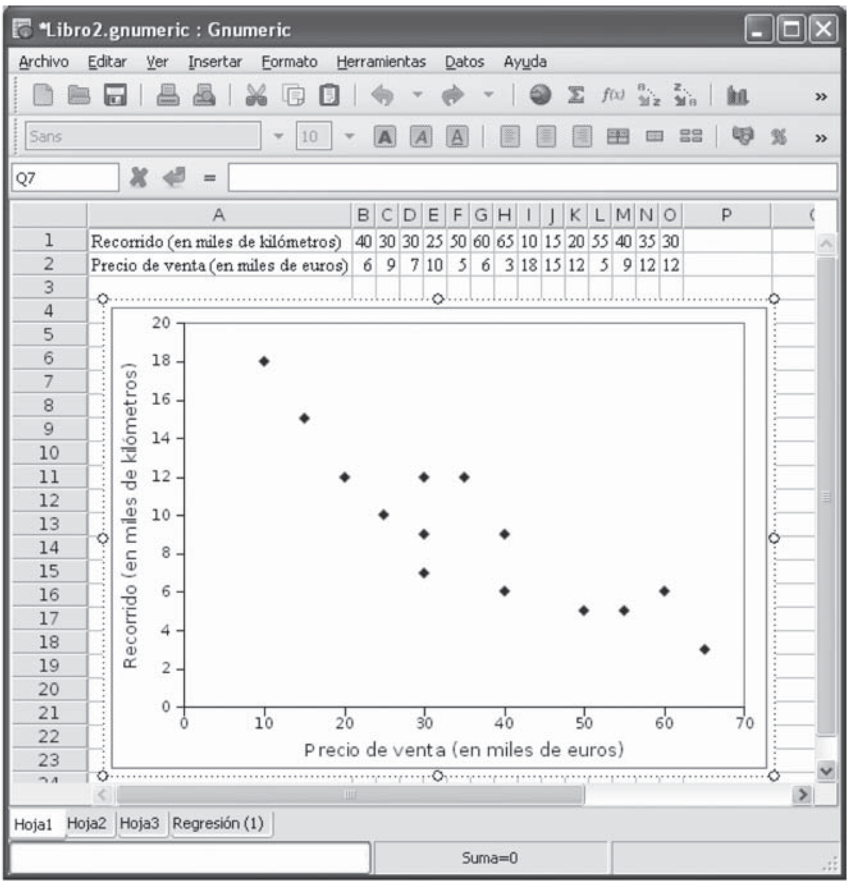
b) Estudia la correlación entre ambas variables.

c) Halla la recta de regresión de Y sobre X.

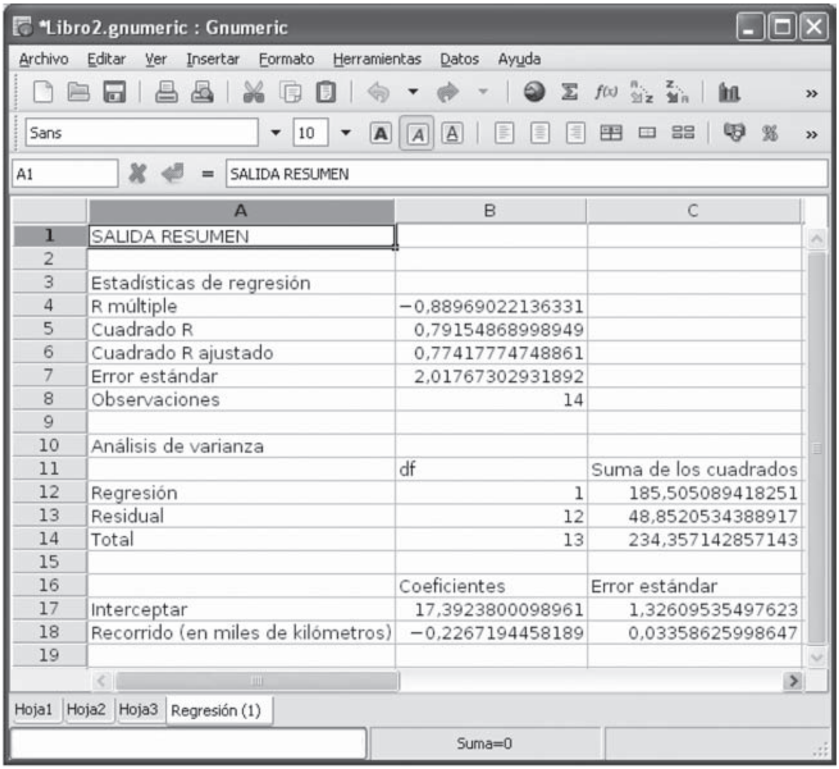
Introducimos los datos:



y construimos el diagrama de dispersión:



Activamos la herramienta **Análisis estadístico → Regresión:**



	A	B	C
1	SALIDA RESUMEN		
2			
3	Estadísticas de regresión		
4	R múltiple	−0,88969022136331	
5	Cuadrado R	0,79154868998949	
6	Cuadrado R ajustado	0,77417774748861	
7	Error estándar	2,01767302931892	
8	Observaciones	14	
9			
10	Análisis de varianza		
11		df	Suma de los cuadrados
12	Regresión	1	185,505089418251
13	Residual	12	48,8520534388917
14	Total	13	234,357142857143
15			
16		Coeficientes	Error estándar
17	Interceptar	17,3923800098961	1,32609535497623
18	Recorrido (en miles de kilómetros)	−0,2267194458189	0,03358625998647
19			

De esta hoja sólo comentaremos los resultados que habitualmente utilizamos en clase:

B4 Coeficiente de correlación lineal

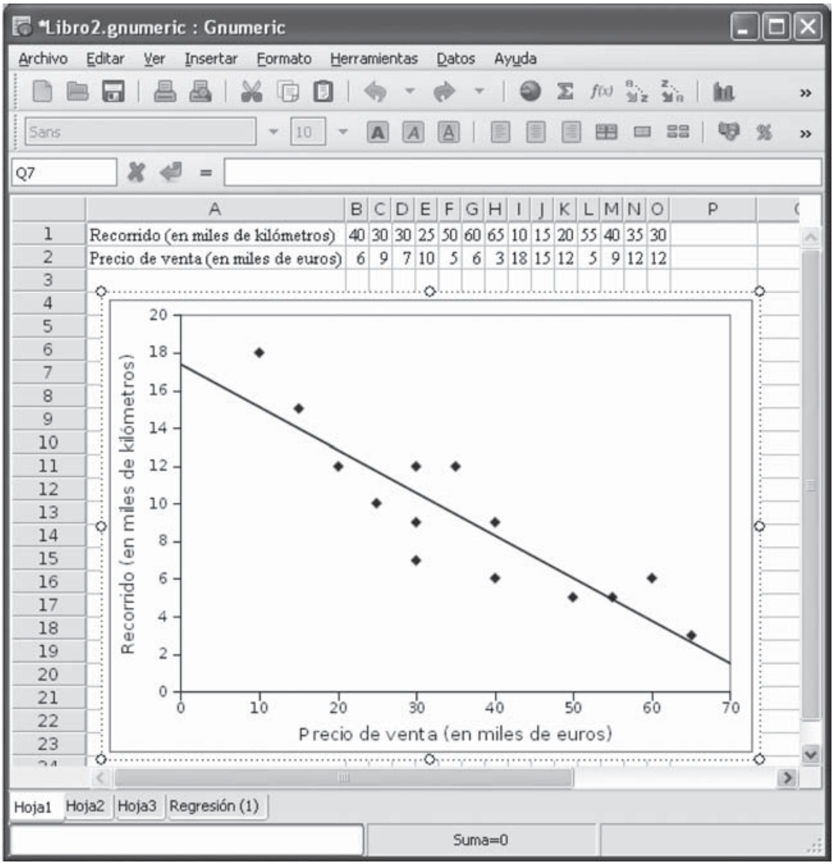
B5 Coeficiente de determinación

B8 Tamaño de la muestra analizada

B17 Ordenada en el origen de la recta de regresión de Y (precio) sobre X (kilómetros)

B18 Pendiente de la recta de regresión ($y = -0{,}227x + 17{,}392$)

Obsérvese que el coeficiente de correlación lineal es cercano a –1, lo que indica una relación fuerte e inversa, como podemos observar insertando la recta de regresión lineal:



WxMaxima

Conocidísimo y estupendo programa libre de cálculo simbólico; sus capacidades estadísticas son menos conocidas pero en la siguiente actividad vamos a observar como con pocas órdenes se pueden obtener muchos datos relativos a estadística bidimensional y gráficos.

Una compañía de seguros sospecha que el número de accidentes está en función de la edad del conductor. Para ello elige 10 personas de cada grupo de edad y contabiliza los accidentes totales del último año. Los datos fueron:

Edad	*20*	*25*	*30*	*35*	*40*	*45*
Nº de Accidentes	*10*	*11*	*9*	*7*	*4*	*5*

Realicemos un estudio bidimensional, observando los comandos y sus significados.

(%i1) A:matrix([20,10], [25,11], [30,9], [35,7], [40,4], [45,5])/* Definimos una matriz con dos columnas (bidimenisonal)*/;

$$(\%o1) \quad \begin{bmatrix} 20 & 10 \\ 25 & 11 \\ 30 & 9 \\ 35 & 7 \\ 40 & 4 \\ 45 & 5 \end{bmatrix}$$

(%i2) load(“descriptive”)/*Cargar el paquete DESCRIPTIVE que viene en la distribución de WxMaxima*/;

(%i3) range(A)/* Vector de rangos */;

(%o3) [25, 7]

(%i4) numer; true/* Pedimos los cálculos en forma decimal */;

(%o4) true

(%i5) mean(A)/* Vector de medias */;

(%o5) [32.5, 7.666666666666666]

(%i6) var(A)/* Vector de varianzas */;

$$(\%o6) \quad [72.91666666666666, 6.555555555555555]$$

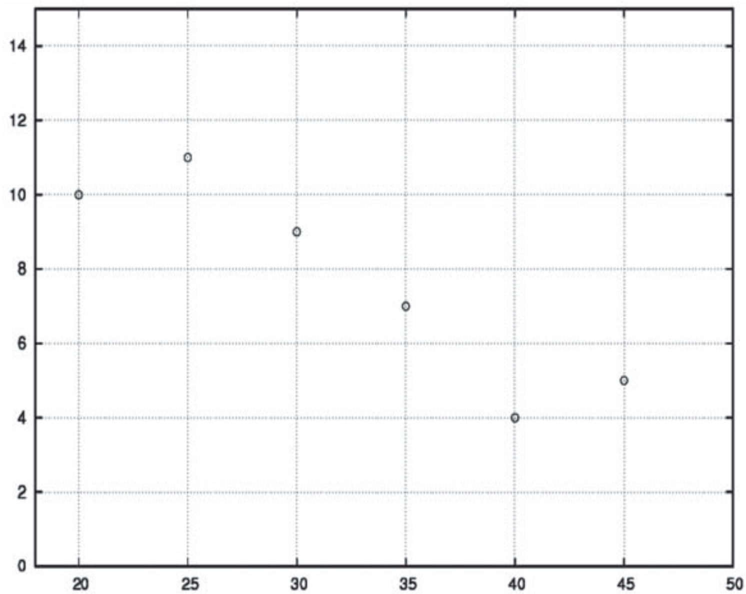
(%i7)cov(A)/* Matriz de covarianzas */;

$$(\%o7) \quad \begin{bmatrix} 72.91666666666652 & -20.0 \\ -20.0 & 6.555555555555557 \end{bmatrix}$$

(%i8) cor(A)/* Matriz de correlaciones */;

$$(\%o8) \quad \begin{bmatrix} 1.0 & -0.91476984760759 \\ -0.91476984760759 & 1.0 \end{bmatrix}$$

(%i9) scatterplot(A,point_type = circle)$

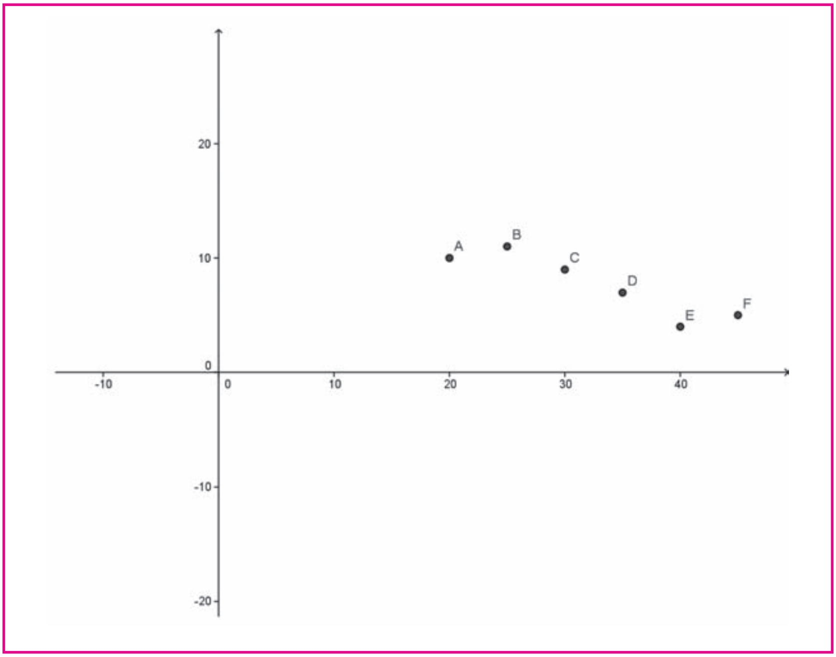


Geogebra

Sorprenderá el uso de este programa para trabajar con estadística bidimensional, ya que es bastante conocido como programa de geometría dinámica con herramientas de algebra; sin embargo desde la versión de desarrollo 3.1 (la próxima estable será la 3.2) incorpora nuevos comandos para hacer ajustes en una nube de puntos. (También se pueden hallar diferentes parámetros estadísticos unidimensionales).

Veamos a resolver el mismo problema que en el apartado anterior:

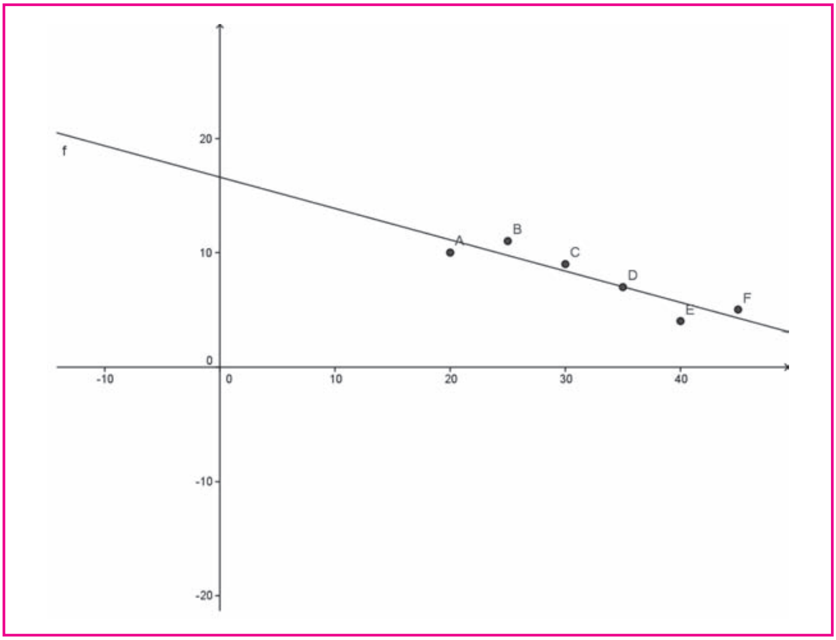
Definimos los puntos y los representamos.



Llamamos M al conjunto de esos puntos: M = {A,B,C,D,E,F}

Para hallar el coeficiente de correlación: r = CPearson[M]

Para hallar la recta de regresión lineal: f = AjusteLineal[M]



Objetos Libres
- A = (20, 10)
- B = (25, 11)
- C = (30, 9)
- D = (35, 7)
- E = (40, 4)
- F = (45, 5)

Objetos Dependientes
- M = {(20, 10), (25, 11), (30, 9), (35, 7), (40, 4), (45, 5)}
- f: 144x + 525y = 8705
- r = -0.91

A partir de ahí podemos añadir o mover puntos para ver el efecto en el coeficiente de correlación y en la recta.

PROPUESTA PARA TRABAJAR CON SOFTWARE LIBRE Y COMERCIAL

Comparación entre las hojas de cálculo de las suites ofimáticas Microsoft Office (Excel) y OpenOffice (Calc.)

Esta actividad de relación entre dos variables estadísticas muestra las similitudes a la hora de trabajar con las dos hojas de cálculo mencionadas.

Las edades de cinco alumnos de un centro escolar, y sus pesos respectivos se muestran a continuación:

Edad (Años)	3	4,5	6	7,2	8
Peso (Kg.)	17	19	25	33	34

a) Hallar las medias y desviaciones marginales.

b) Calcular el coeficiente de correlación lineal.

c) Hallar la recta de regresión de Y sobre X.

d) ¿Qué peso se espera que tenga un niño de 5 años?

e) Realizar un gráfico que contenga la nube de puntos y la recta de regresión.

Introducimos los datos en las celdas A, la edad, y en la B, el peso.

	A	B	C	D
1	**Edad**	**Peso**		
2	3	17		
3	4,5	19		
4	6	25		
5	7,2	33		
6	8	34		
7				
8			Media Edad	5,74
9			Media Peso	25,6
10			Desv. Típica Edad	1,808424729
11			Desv. Típica Peso	6,97423831
12				
13			Coeficiente de correlación	0,973328992
14			Pendiente recta regresión	3,753669276
15			Peso esperado para cinco años	22,82228474

Las funciones utilizadas en la hoja de cálculo (idénticas en Microsoft Excel y OpenOffice Calc) han sido:

=PROMEDIO(A2:A6)

=PROMEDIO(B2:B6)

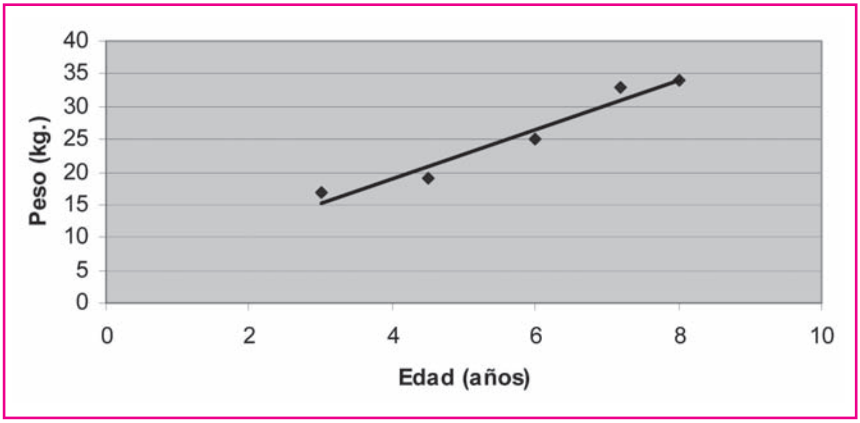
=DESVESTP(A2:A6)

=DESVESTP(B2:B6)

=COEF.DE.CORREL(A2:A6;B2:B6)

=PENDIENTE(B2;B6;A2:A6). La recta de regresión es y–25’6 = 3’75 (x–5’74).

=TENDENCIA(B2:B6,A2:A6,5)



PROPUESTA PARA TRABAJAR CON SOFTWARE COMERCIAL

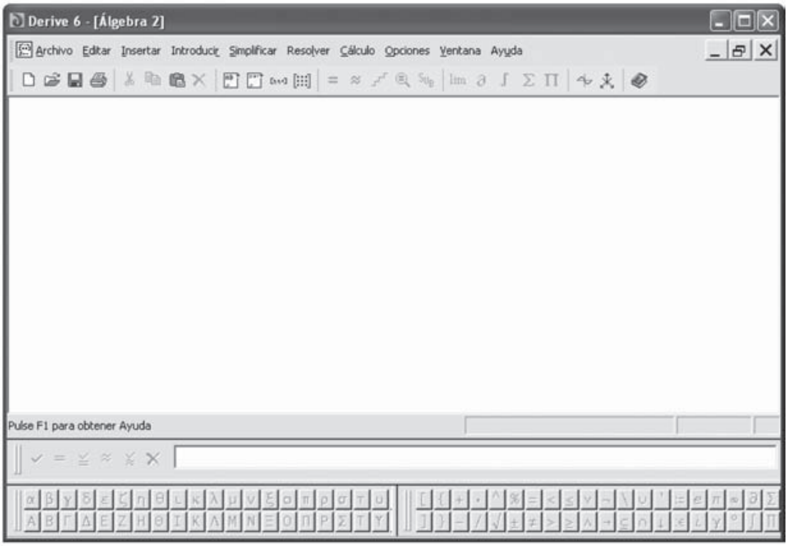
Derive

El clásico programa de cálculo simbólico comercial y no libre (aunque tiene versiones de evaluación temporal) también permite trabajar la Estadística bidimensional.

La siguiente tabla presenta datos correspondientes a la tasa de ocupación en función de la edad en España en el segundo cuatrimestre de 2008, según la Encuesta de Población Activa (EPA) del Instituto Nacional de Estadística. Realizar un diagrama de dispersión y ajustar por un gráfico de regresión lineal, cuadrático y cúbico.

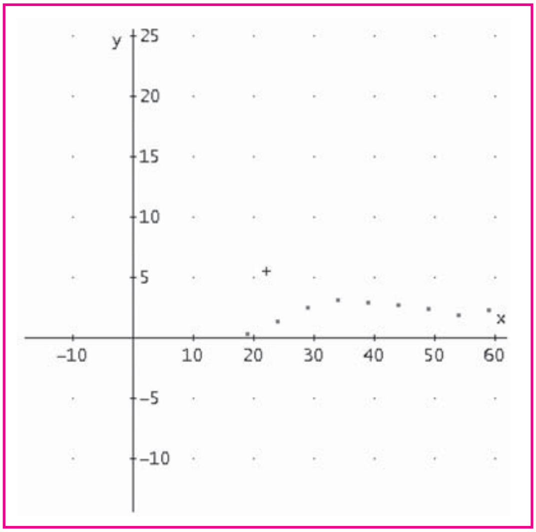
Edad (en años)		Millones de personas
19		0,3
24		1,4
29		2,5
34		3,1
39		2,9
44		2,7
49		2,4
54		1,9
59		2,3

Introducimos los datos en forma de matriz o tabla en la siguiente pantalla:



$$\#1: \quad \begin{bmatrix} 19 & 0.3 \\ 24 & 1.4 \\ 29 & 2.5 \\ 34 & 3.1 \\ 39 & 2.9 \\ 44 & 2.7 \\ 49 & 2.4 \\ 54 & 1.9 \\ 59 & 2.3 \end{bmatrix}$$

y los representamos:



Con la función FIT vamos a pedirle un ajuste lineal:

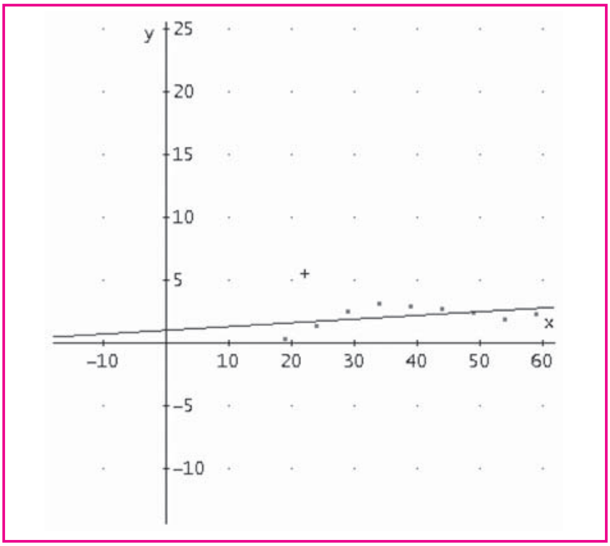
FIT([x,ax+b],#1)

(donde #1 es la etiqueta de la matriz de datos, aunque le podemos poner cualquier nombre)

y simplificando:

$$\#2: \quad \frac{89\cdot x}{3000} + \frac{3029}{3000}$$

Ahora la representamos sobre el gráfico de dispersión:

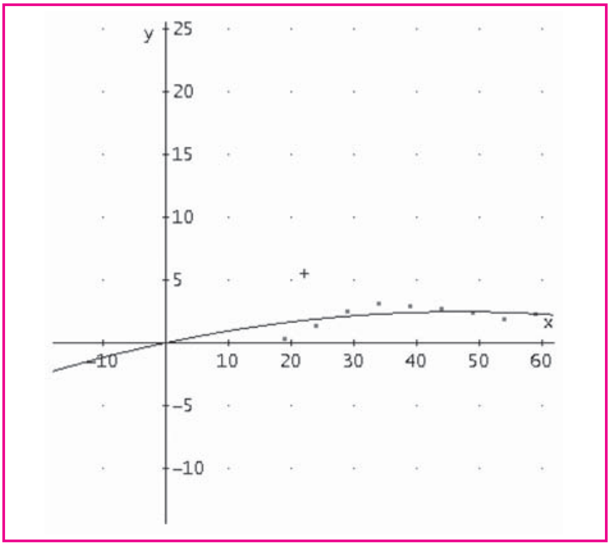


Obsérvese que el ajuste no es muy acertado (de hecho, con este mismo programa se puede hallar el coeficiente de correlación paso a paso y comprobar que es r = 0’48).

Intentemos un ajuste cuadrático con FIT ([x,aˆ2+bx+c],#1)

$$\#5: \quad \frac{973339079\cdot x}{9228612000} + \frac{10368361\cdot x^2}{9228612000}$$

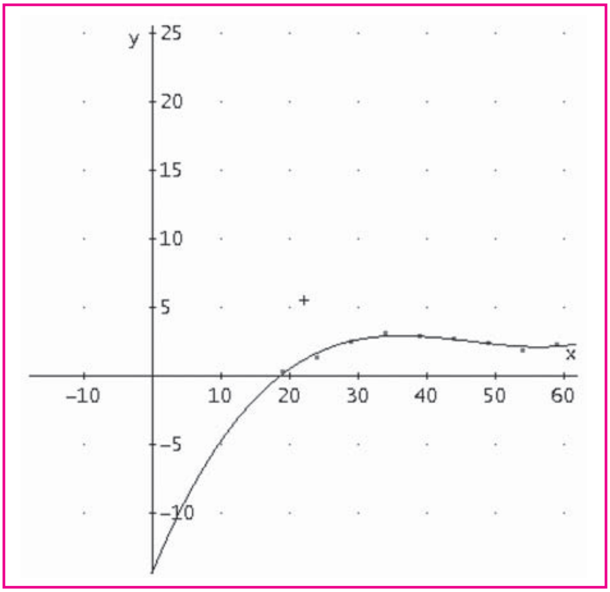
y representamos nuevamente sobre la nube de puntos:



Todavía podemos ajustar mejor:

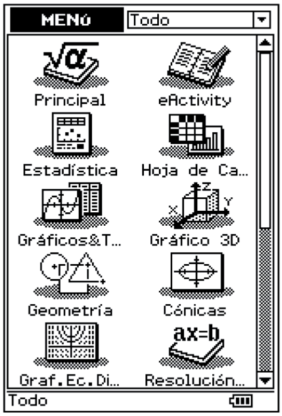
FIT([x,axˆ3+bxˆ2+cx+d],#1)

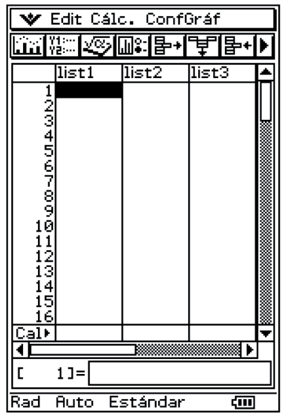
#6: $$\frac{49\cdot x3}{247500} - \frac{10583\cdot x2}{385000} + \frac{4199473\cdot x}{3465000} - \frac{8266099}{577500}$$



CALCULADORA GRÁFICA

Otro recurso con el que es posible incorporar nuevas herramientas al aula de matemáticas es la calculadora gráfica, ya que dispone también de buenas herramientas para el estudio estadístico bidimensional. Utilizamos la calculadora ClassPad 330-A de Casio.

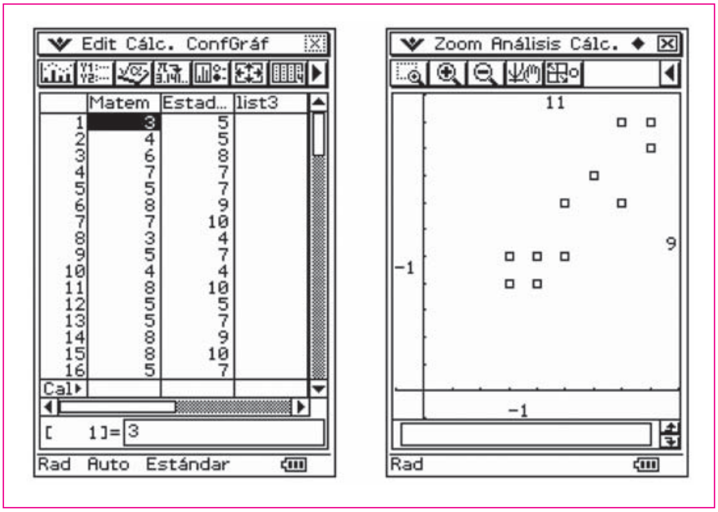




Las calificaciones obtenidas por un grupo de alumnos en las asignaturas de Estadística y Matemáticas han sido:

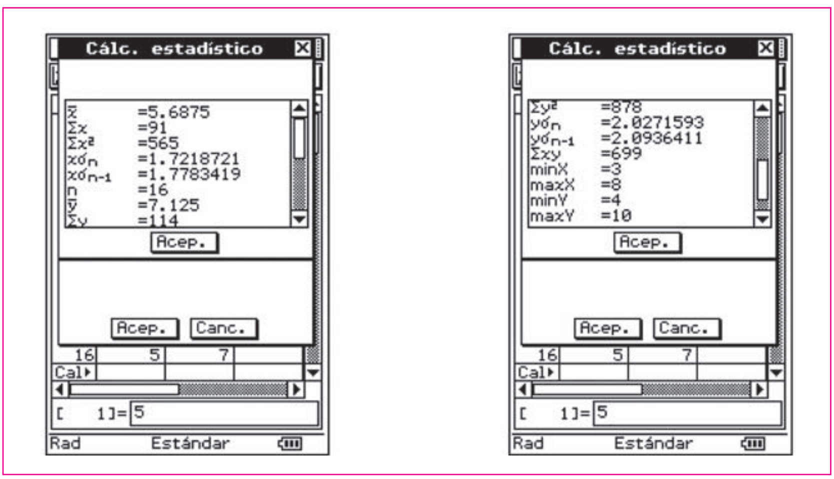
Matemáticas	3	4	6	7	5	8	7	3	5	4	8	5	5	8	8	5
Estadística	5	5	8	7	7	9	10	4	7	4	10	5	7	9	10	7

1. *Construye el diagrama de dispersión.*
2. *Estudia la correlación entre ambas variables.*
3. *Halla la recta de regresión de Y sobre X.*
4. *¿Qué nota se espera que obtenga un alumno en Estadística cuya nota en Matemáticas ha sido 2?*

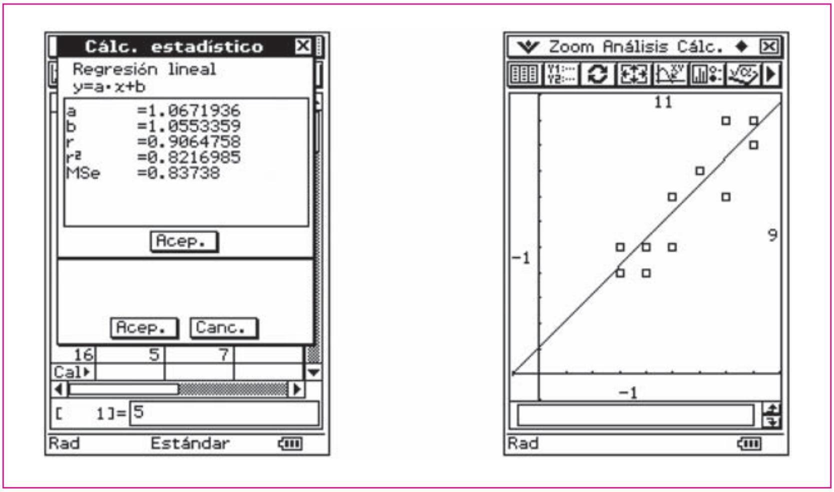


El primer paso para determinar el sentido y el grado de la correlación entre dos variables es dibujar el diagrama de dispersión. En esta actividad la nube es alargada lo que indica que hay correlación lineal, en principio; además es directa y parece que fuerte.

Datos marginales:

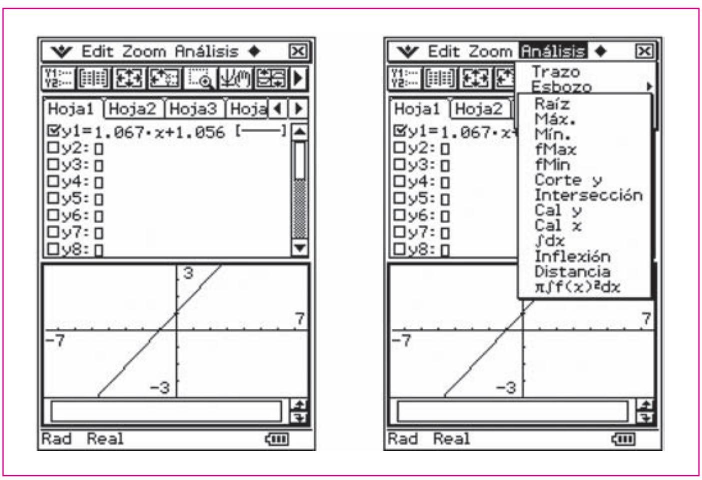


Para sacar el máximo provecho a la correlación vamos a hallar la recta de regresión que es la que mejor se ajusta a la nube de puntos.

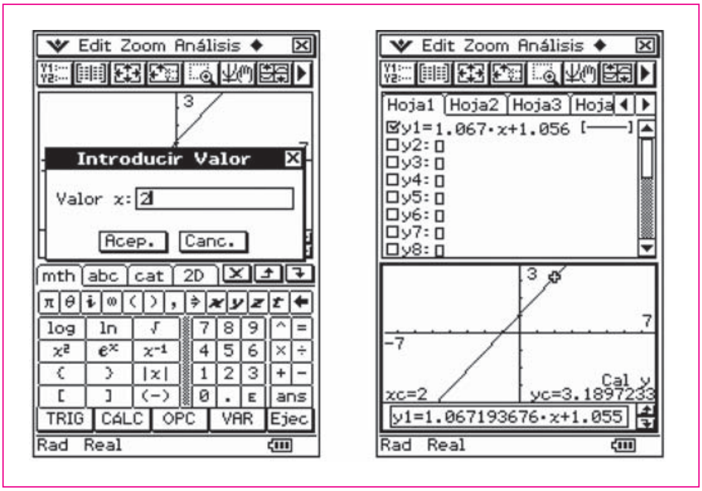


Obsérvese que aparece por primera vez el coeficiente de correlación lineal (al ser cercano a +1 la correlación es fuerte y directa como habíamos aventurado a con la observación del gráfico) y la expresión analítica de la recta.

Para hacer la estimación, en el menú **Gráficos** hallamos el valor esperado de *y* conocido *x*.



Por tanto se espera que obtenga un 3 aproximadamente.



REFERENCIAS

Gnumeric: http://projects.gnome.org/gnumeric/

xMaxima y WxMaxima: http://maxima.sourceforge.net/

Geogebra: http://geogebra.org

OpenOffice Calc: http://es.openoffice.org/

Microsoft Office Excel: http://office.microsoft.com/

Derive 6: versión de prueba que se puede buscar en la página http://education.ti.com/

Calculadoras ClassPad: http://www.aulacasio.com