TALLER DE ESTADÍSTICA

2. ANÁLISIS DE REGRESIÓN.

MAURICIO CONTRERAS

RELACIONES ESTADÍSTICAS ENTRE VARIABLES EN LA ESO

Introducción

Para contrastar si existe algún tipo de relación entre dos variables de una población, extraemos una muestra de la población y recogemos información sobre los valores de dichas variables. A partir de la información muestral podemos construir un diagrama de puntos o diagrama de dispersión. La forma del diagrama indicará si se puede ajustar o no una función a los datos obtenidos y qué tipo de función debemos utilizar en el ajuste. Las funciones de ajuste se utilizarán para hacer estimaciones y predicciones razonables, teniendo en cuenta los tamaños muestrales y el campo de variabilidad de los datos. En 3º y 4º de ESO se utilizarán los modelos funcionales más próximos a los alumnos (lineal y cuadrático). En Bachillerato pueden usarse otros modelos (cúbico, exponencial, logarítmico). La calculadora gráfica puede ayudarnos a representar gráficamente estas funciones y estudiar sus peculiaridades. El concepto de correlación es inseparable del concepto de regresión, ya que la correlación mide la bondad del ajuste. En las siguientes actividades analizarararemos la utilidad de la calculadora gráfica y de otros materiales manipulables para el tratamiento de estas cuestiones en el aula.

◆ Ajuste de un diagrama de puntos por una recta

- **<u>RUIDO Y VEHÍCULOS</u>**

Material: papel milimetrado, regla graduada, calculadora.

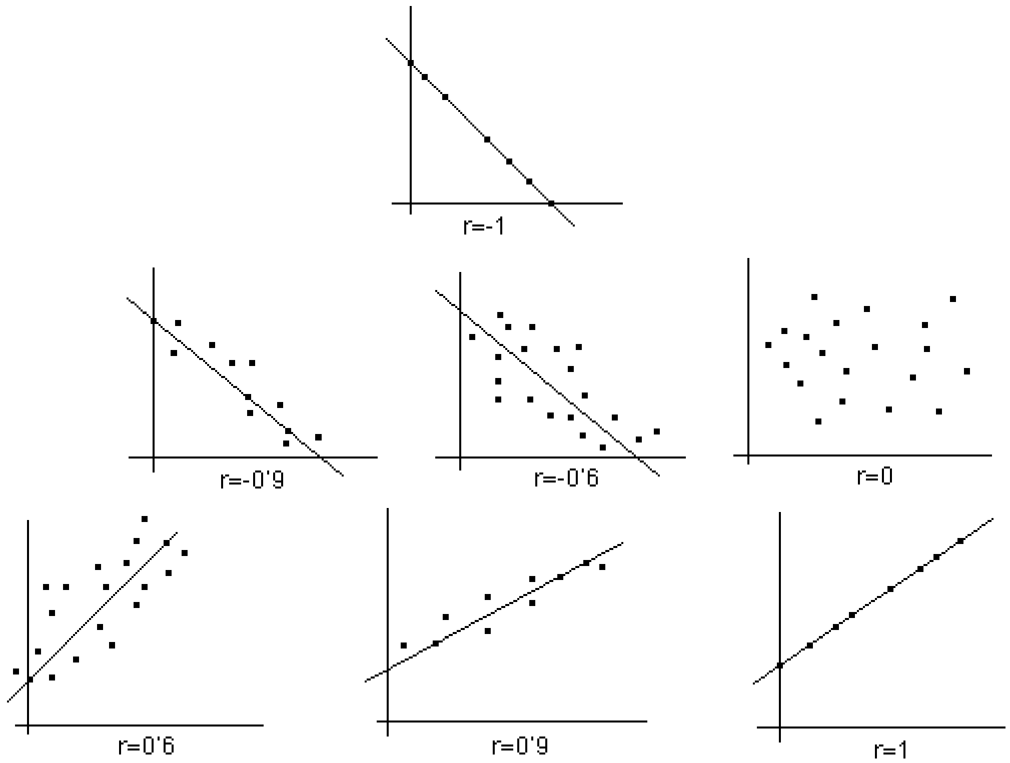
a) La siguiente tabla indica el número de turismos matriculados y el nivel de ruido por trimestre en la ciudad de Valencia durante el periodo 1998 – 2001. Representa gráficamente estos datos en una hoja de papel milimetrado, situando en el eje horizontal el número de turismos y en el eje vertical el número de decibelios.

		Turismos matriculados	**Nivel medio sonoro en dB**
1998	Primer trimestre	5622	70’92
	Segundo trimestre	6937	68’87
	Tercer trimestre	6747	69’41
	Cuarto trimestre	7247	71'20
1999	Primer trimestre	6850	70’75
	Segundo trimestre	7120	70’52
	Tercer trimestre	6530	71’31
	Cuarto trimestre	6900	70’97
2000	Primer trimestre	8023	71’45
	Segundo trimestre	8629	71’82
	Tercer trimestre	7804	70’99
	Cuarto trimestre	5883	71’98
2001	Primer trimestre	6502	70’67
	Segundo trimestre	7854	70’34
	Tercer trimestre	6956	71’93
	Cuarto trimestre	6929	72’19

- ✓ ¿Cómo debes elegir las escalas de los ejes?.
- ✓ ¿Tiene sentido unir los puntos de la gráfica?.
- ✓ A la vista de la gráfica, ¿crees que existe una relación entre el número de turismos y el nivel de ruido?.

✓ Si la relación anterior existe, ¿crees que se puede expresar mediante una fórmula?.

La gráfica obtenida se llama diagrama de dispersión o nube de puntos y muestra la relación de dependencia entre dos magnitudes. Si los puntos están próximos a una recta, se dice que hay correlación entre dichas magnitudes; si los puntos están muy dispersos y no parecen seguir ningún patrón, se dice que no hay correlación entre las dos magnitudes. Conforme los puntos estén más y más próximos a una recta la correlación es mayor. Si al aumentar una magnitud, aumenta la otra, la correlación es positiva; si al aumentar una magnitud, la otra disminuye, la correlación es negativa. Podemos cuantificar esta idea con un número r denominado coeficiente de correlación, medido en una escala de –1 a 1, de forma que si r está próximo a –1, la correlación es negativa, si r = 0 ó próximo a 0, no hay correlación y si r está próximo a 1, la correlación es positiva. He aquí unos ejemplos:



✓ ¿Qué valor del coeficiente de correlación asignarías al diagrama nº de turismos → nº de decibelios que has construido anteriormente?.

b) Intenta construir lo más aproximadamente que puedas una recta que se ajuste a la nube de puntos, de forma que la distancia de los puntos a dicha recta sea lo menor posible.

Esta recta se llama recta de ajuste o recta de regresión.

c) ¿Qué ocurrirá si continua aumentando el número de turismos matriculados?. ¿Qué nivel medio de ruido cabe esperar en el trimestre en que se matriculen 8000 turismos?. ¿Y cuando se matriculen 9000?. Si el ritmo de crecimiento se mantiene, ¿ha de pasar mucho tiempo para que se alcancen 10000 vehículos matriculados en un trimestre?.

d) ¿Qué grado de seguridad te merecen las estimaciones que has hecho anteriormente?.

e) ¿Te atreves a construir una fórmula que represente la recta de ajuste que has dibujado?. Compara los resultados con los de tus compañeros.

- **<u>ESTUDIO Y TV</u>**

a) Recoge información de tu clase sobre el número de horas diarias que dedican tus compañeros al estudio y a ver la televisión. Resume la información en una tabla de doble entrada:

HORAS ESTUDIO HORAS TV	$0 - 1/2$	$1/2 - 1$	1 – 2	2 – 3	3 – 4	4 – 5	5 – 6
$0 - 1/2$							
$1/2 - 1$							
1 – 2							
2 – 3							
3 – 4							
4 – 5							
5 – 6							

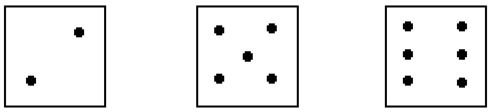
b) Representa gráficamente la nube de puntos y asígnale un coeficiente de correlación.

Te sugerimos que hagas lo siguiente:

Por los extremos de cada intervalo traza rectas paralelas a los ejes. De esta forma, la gráfica queda dividida en casillas.

Pues bien, dibuja en cada casilla tantos puntos como indique la frecuencia respectiva.

Así, si la frecuencia es 2, 5 ó 6, dibujarás en la casilla correspondiente:



c) Dibuja lo más aproximadamente que puedas una recta de ajuste.

d) Maribel dedica al estudio una hora y media diaria. ¿Podrías decir, de manera aproximada, cuánto tiempo dedica diariamente a ver la televisión?. ¿Es fiable la estimación realizada?.

e) Intenta encontrar una fórmula que represente la recta de ajuste.

- **<u>NOVIOS</u>**

Durante el último mes, en un juzgado, se celebraron diez matrimonios. Las edades de los novios y de las novias se recogen en la siguiente tabla.

Edad del novio	20	23	24	24	25	26	28	28	33	39
Edad de la novia	18	21	20	22	22	23	28	26	32	28

a) Calcula la edad media del novio y la novia y la desviación típica. ¿Qué variable está más dispersa?.

b) Dibuja la nube de puntos.

c) Con una regla traza la recta que mejor se ajuste a la nube de puntos.

d) Di si la correlación es fuerte o débil, positiva o negativa.

e) Si la edad del novio es de 30 años, ¿se puede predecir cuál será la edad de la novia?.

- **ABUELOS**

Las personas de un grupo conversan sobre sus edades y los abuelos que les queda a cada una. Los resultados numéricos de dicha conversación se recogen en la siguiente tabla:

Edad	6	10	15	21	26	30	35	39	46	50
Nº abuelos	4	4	3	4	2	3	2	1	1	0

Dibuja una nube de puntos edades de las personas → número de abuelos que les quedan. Expresa verbalmente la relación existente entre las dos variables.

- **CALIDAD DE VIDA**

La tabla siguiente muestra la clasificación de diez países de la ONU según su riqueza y la esperanza de vida (en años) de su población femenina:

País	Clasificación según riqueza	Esperanza de vida de las mujeres
Argentina	42	75
Bosnia	60	76
Etiopía	180	50
Guatemala	114	66
Haití	143	56
India	152	60
Irak	77	66
Japón	2	82
Reino Unido	19	79
Vietnam	165	69

¿Se puede predecir qué puesto ocupa en la clasificación según riqueza un país en el que la esperanza de vida femenina es de 80 años?.

- **TENSIÓN**

La distribución de edades y presión arterial de 10 personas es:

Edad (X)	30	28	35	42	51	42	63	32	70	67
Tensión (Y)	11'5	11'3	12'5	13'5	14'6	13	16'6	12	16'9	17

a) Representa la nube de puntos y justifica que sí tiene sentido ajustar a la nube una recta de regresión.

b) Calcula el coeficiente de correlación lineal.

c) ¿Cómo calificarías la correlación entre las variables?.

d) Determina la ecuación de la recta de regresión.

e) Estima la tensión esperada en una persona de 60 años.

- **PADRES E HIJOS**

La altura de 10 padres y de su primer hijo varón está reflejada en la siguiente tabla:

Talla del padre X	172	184	175	184	180	176	170	189	171	185
Talla del hijoY	174	180	174	186	190	178	170	195	178	182

Representa la nube de puntos. ¿Existe correlación entre ellos?. En caso afirmativo calcula la recta de regresión. ¿Qué altura cabe esperar en un hijo cuyo padre mide 182 cm?.

- **TEMPERATURAS**

La latitud en grados y la temperatura máximas en ºC en un mismo día del año son:

Ciudad	**Latitud (X)**	**Temperatura (Y)**
Acapulco	17	30
Barcelona	41	20
Calcuta	22	32
Dakar	15	30
Estambul	41	18
Jerusalén	32	25
Karachi	25	31
A Coruña	43	15

Representa la nube de puntos. ¿Existe correlación entre ellos?. En caso afirmativo, calcula la recta de regresión.

- **CUIDADO CON LAS ESTADÍSTICAS**

¿Qué opinas de algunas de estas afirmaciones?. Todas ellas se han hecho basándose en datos estadísticos, de forma que las magnitudes que se citan están correlacionadas.

1) Los niños con los pies grandes tienen mejor ortografía. ¿Significa esto que el tamaño del pie nos informa sobre la calidad de la ortografía de los niños?.

2) En el sur de Francia, los municipios con mayor tasa de divorcio tienen generalmente menor tasa de mortalidad. ¿Será bueno divorciarse para vivir más años?.

3) Los países que añaden flúor al agua tienen tasas de cáncer mayores que otros que no lo añaden. ¿Es el flúor, en concentraciones elevadas, perjudicial para la salud?.

4) Los accidentes de circulación se producen, generalmente, en vehículos que circulan con velocidad moderada. Pocos accidentes ocurren a 180 km/h. ¿Quiere esto decir que es más seguro circular a gran velocidad?.

5) En una determinada región del sur de Italia se observó con el paso del tiempo que hubo un fuerte crecimiento de la población al mismo tiempo que aumentó el número de cigüeñas. ¿Significa esto que a los niños los traen las cigüeñas?.

Ten en cuenta que la correlación entre dos magnitudes puede deberse al azar o a otras causas. El hecho de que dos variables estén correlacionadas no quiere decir que una sea causa de la otra necesariamente.

ANÁLISIS DE DATOS Y ANÁLISIS DE REGRESIÓN CON LA CALCULADORA GRÁFICA

Introducción

Este trabajo es un resumen del desarrollo en el aula de una unidad didáctica pensada para estudiantes de 1º curso de Bachillerato, en la modalidad de Matemáticas aplicadas a las ciencias sociales, aunque las actividades podrían tratarse también en el bachillerato científico – tecnológico.

La filosofía que se ha seguido en el diseño de la unidad es la de estudiar los temas transversales del currículum como hilo conductor para el aprendizaje de las matemáticas y como vehículo interdisciplinar. Así mismo se ha pretendido optimizar el uso de la calculadora gráfica en el aula, que se ha revelado como un recurso poderoso e imprescindible para el análisis y tratamiento de datos.

La necesidad de considerar una unidad didáctica de este tipo está relacionada con la ubicación geográfica del centro. Posiblemente no tiene sentido tratar la contaminación acústica en un entorno rural, pero si en una gran ciudad, ya que este tipo de contaminación –a la que normalmente no se le da importancia– va creciendo con el paso del tiempo, sobre todo en los grandes núcleos urbanos.

El análisis del fenómeno y sus posibles causas debe llevar a la reflexión sobre qué podemos hacer para mejorar la situación, y, sobre todo, tomar conciencia de que ser respetuoso con el medio ambiente implica necesariamente un consumo racional ligado a las auténticas necesidades de los individuos. Se trata, pues, de un tema interdisciplinar que ha resultado ser muy motivador para los estudiantes.

OBJETIVOS

- Conocer las actitudes frente al entorno y medio ambiente.
- Analizar, representar e interpretar datos estadísticos.
- Analizar procesos de tratamiento de datos.
- Utilizar técnicas estadísticas para hacer predicciones y extraer conclusiones de los datos.
- Fomentar hábitos de respeto al medio ambiente.
- Fomentar un uso racional de los instrumentos tecnológicos.

CONTENIDOS

CONCEPTOS

- La contaminación y la educación vial.
- El respeto al medio ambiente.
- Correlación y dependencia estadística.
- Análisis de regresión.
- Regresión lineal y no lineal.
- Predicciones estadísticas. Fiabilidad.
- Coeficiente de correlación y coeficiente de determinación.

PROCEDIMIENTOS

- Lectura e interpretación de la información contenida en las gráficas.
- Reflexión sobre el proceso de tratamiento de datos.
- Construcción de gráficas a partir de experiencias, tablas y fórmulas.
- Manejo de listas: operaciones con listas en la calculadora gráfica.
- Construcción e interpretación de diagramas estadísticos: barras, histogramas de frecuencias, nubes de puntos.
- Utilización de modelos de ajuste para hacer predicciones.
- Utilización apropiada de la calculadora gráfica para construir diagramas estadísticos, nubes de puntos y modelos de ajuste.
- Uso apropiado de la calculadora gráfica para representar e interpretar datos estadísticos.

ACTITUDES

- Reflexión y valoración crítica de la información.
- Respeto al entorno y al medio ambiente.
- Colaboración y participación en las tareas comunitarias.
- Toma de decisiones meditada y responsable.
- Reconocimiento de la potencia del uso razonable de la tecnología.

TIEMPO Y ORGANIZACIÓN

El tiempo previsto es, aproximadamente, de 12 sesiones (tres semanas).

Esta unidad didáctica forma parte de un bloque más amplio dedicado a la Estadística y el Tratamiento de datos.

En todas las sesiones los estudiantes trabajan por parejas.

EVALUACIÓN

El instrumento básico de evaluación es la observación por parte del profesor del trabajo de los estudiantes en cada grupo y del trabajo individual en cada actividad. La observación se centrará en los siguientes aspectos:

- Comprensión de la actividad.
- Proceso de resolución.
- Conclusiones obtenidas.
- Participación en la actividad: colaboración, discusión, respeto a los otros, grado de convivencia.
- Exposición de las conclusiones y decisiones adoptadas.
- Adecuación de la actividad.

1. Contaminación acústica

Los siguientes datos proceden del Servicio de Medio Ambiente del Ayuntamiento de Valencia. Las medidas del nivel sonoro se toman en tres puntos de la ciudad: Nuevo Centro, Plaza de España y Pista de Silla. La información indica niveles promedios durante un día para las distintas franjas horarias. Las mediciones se han hecho en el cuarto trimestre de 1997.

El nivel sonoro se mide a partir de un mínimo que se establece en el umbral de percepción del oído humano, este es de 0 dB (decibelios), entorno a 130 dB se sitúa el nivel del dolor. La O.C.D.E. establece como recomendable valores inferiores a los 65 dB.

Niveles sonoros de referencia: Sonido de fondo en el campo: entre 15 y 20 dB.
Sonido en una biblioteca: en torno a 35 dB.
Sonido de una conversación: en torno a 65 dB.
Sonido del tráfico: en torno a 70 dB.
Sonido de un avión despegando: en torno a 120 dB.

Nivel sonoro (ruido) en dB. Media por hora del día.			
HORA	**NOU CENTRE**	**PLAÇA D'ESPANYA**	**AV. AUSIAS MARCH**
0	69,1	67,8	66,5
1	67,5	67,1	65,7
2	66,5	65,9	65,4
3	66,1	65,4	65,0
4	68,1	66,1	65,8
5	71,2	68,1	67,4
6	73,2	70,8	69,8
7	73,7	72,6	70,8
8	73,8	73,3	71,2
9	73,6	73,7	71,3
10	73,5	73,9	71,5
11	73,5	74,1	71,5
12	73,5	74,3	71,4
13	73,3	74,0	70,6
14	73,4	73,7	70,0
15	73,9	73,5	70,7
16	73,9	73,9	71,5
17	74,1	74,2	71,8
18	73,8	74,2	71,8
19	73,7	74,1	71,2
20	73,2	73,6	70,8
21	72,4	72,3	69,4
22	71,2	70,7	68,0
23	70,6	69,2	67,1

a) *¿Cuál es el nivel de ruido en cada uno de los puntos de referencia a las 6 de la mañana?. ¿Y a las 14 horas?. ¿A qué horas del día hay un nivel de ruido superior a 72 dB?. ¿En qué lugares?.*

b) *¿Cómo evoluciona el nivel sonoro a lo largo del día?. ¿Dónde hay más contaminación acústica, en Nuevo Centro o en la plaza de España?.*

c) *¿Cuál es el mayor nivel de ruido durante el día?. ¿A qué hora?. ¿Dónde?. ¿Y el menor?. ¿A qué hora?. ¿Dónde?. ¿Se cumple la recomendación de la O.C.D.E.?.*

2. Un primer análisis de los datos

Para analizar los datos, introducimos en la TI-83 las siguientes listas:

HORA=seq{X, X, 0, 23},
NC (datos de la columna Nou Centre),
PE (columna Plaça d'Espanya),
AM (correspondiente a la columna Av. Ausias March).

Obtenemos en primer lugar los parámetros estadísticos correspondientes a las variables NC, PE y AM.

1-Var Stats LNC	**1-Var Stats LPE**	**1-Var Stats LAM**
$\bar{x} = 71.95$	$\bar{x} = 71.52083333$	$\bar{x} = 69.425$
$\sum x = 1726.8$	$\sum x = 1716.5$	$\sum x = 1666.2$
$\sum x^2 = 124395.26$	$\sum x^2 = 122991.15$	$\sum x^2 = 115804$
$S_x = 2.570738348$	$S_x = 3.132158468$	$S_x = 2.359670205$
$\sigma_x = 2.516611478$	$\sigma_x = 3.066210904$	$\sigma_x = 2.309987374$
n = 24	n = 24	n = 24
minX = 66.1	minX = 65.4	minX = 65
$Q_1 = 70.9$	$Q_1 = 68.65$	$Q_1 = 67.25$
Med = 73.35	Med = 73.4	Med = 70.65
$Q_3 = 73.7$	$Q_3 = 73.95$	$Q_3 = 71.35$
maxX = 74.1	maxX=74.3	maxX = 71.8

En los tres casos, la variable tiene muy poca dispersión: está muy concentrada entorno a la media y a la mediana. La media más alta corresponde a Nuevo centro ($\bar{x} = 71.95, \sigma_x = 2.517$), seguida de la Plaza de España ($\bar{x} = 71.52, \sigma_x = 3.07$). Los niveles de contaminación acústica son parecidos en estos dos puntos. En cambio, es algo más baja en Ausias March ($\bar{x} = 69.425, \sigma_x = 2.31$). También la mediana parece indicar que la contaminación acústica es ligeramente inferior en este punto de la ciudad, como podemos ver en la siguiente tabla:

VARIABLES	**MEDIANA**	**RANGO INTERCUARTÍLICO**	
		CUARTIL INFERIOR	**CUARTIL SUPERIOR**
NC	Med = 73.35	$Q_1 = 70.9$	$Q_3 = 73.7$
PE	Med = 73.4	$Q_1 = 68.65$	$Q_3 = 73.95$
AM	Med = 70.65	$Q_1 = 67.25$	$Q_3 = 71.35$

Para confirmar este estudio comparativo, utilizamos la TI-83 para construir los diagramas de caja correspondientes a las tres variables.

Plot 1	Plot 2	Plot 3
On Type………..caja con outliers Xlist………..NC Freq………..1 Mark……….□	On Type………..caja con outliers Xlist………..PE Freq………..1 Mark……….□	On Type………..caja con outliers Xlist………..PE Freq………..1 Mark……….□

Al pulsar ZOOM y seleccionar ZoomStat, obtenemos:

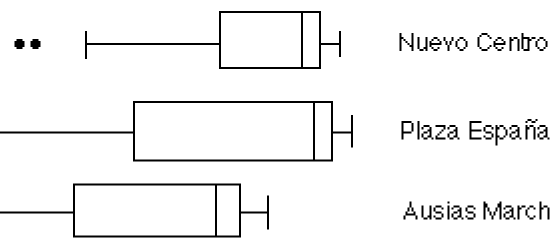


Figura 1

Estos diagramas de caja confirman que, efectivamente, el nivel de contaminación acústica en Ausias March es inferior a los otros dos puntos. Pero además muestran que las tres distribuciones son asimétricas a la izquierda (las tres presentan una cola hacia la izquierda), lo que se traduce en que la media es inferior a la mediana en los tres casos, como podemos observar en la siguiente tabla:

Nuevo Centro	Plaza España	Ausias March
$\bar{x} = 71.95$	$\bar{x} = 71.52083333$	$\bar{x} = 69.425$
Med = 73.35	Med = 73.4	Med = 70.65

Por otra parte vemos dos outliers en el diagrama correspondiente a la variable NC, lo que indica que la contaminación en Nuevo Centro es más alta que en los otros dos puntos: aunque presenta valores similares a los de PE, los datos de Nuevo Centro son menos dispersos. Mediante TRACE y las teclas de cursor, observamos que los valores outliers son x = 66.5 db (que se da a las 2 de la madrugada) y x = 66.1 db (valor mínimo, que se da a las 3 de la madrugada).

Aunque no está al alcance de los estudiantes de Bachillerato, podemos comparar las tres variables NC, PE y AM mediante un análisis de la varianza ANOVA de una vía, lo que es posible hacer en la TI – 83 con la instrucción ANOVA(LNC, LPE, LAM) pulsando la siguiente secuencia de teclas:

[STAT] TESTS F: ANOVA([2ND] LIST NC , [2ND] LIST PE , [2ND] LIST AM) ENTER

El resultado se muestra en la siguiente pantalla:

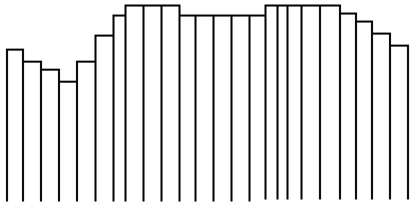
```
One – way ANOVA
F = 5.977486032
p = 0.0040354497
Factor
  df = 2
  SS = 87.6186111
  MS = 43.8093056
Error
  df = 69
  SS = 505.704583
  MS = 7.32905193
  S x p = 2.70722218
```

Como el p – valor es p = 0.004 < 0.05 rechazamos, con un nivel de significación del 5%, la hipótesis nula de que los tres grupos de datos sean iguales. Es decir, con un nivel de significación del 5% podemos asegurar que los tres grupos son diferentes. Por tanto, el nivel de ruido en Ausias March no es el mismo que en Nuevo Centro o Plaza de España.

Con objeto de estudiar el nivel de contaminación acústica en cada uno de los tres puntos por separado, según la hora del día, utilizamos la TI–83 para construir el histograma correspondiente. Previamente en WINDOW introducimos el valor Xscl = 1, con objeto de que la anchura de cada rectángulo sea igual a 1 hora. A continuación elegimos [2nd] [STAT PLOT] **4:** PlotsOff para desactivar todos los gráficos estadísticos e introducimos los valores de la siguiente tabla:

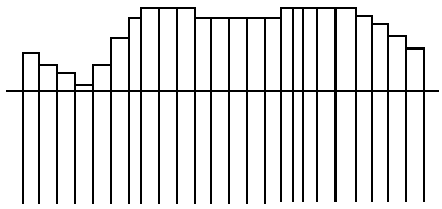
Plot 1	Plot 2	Plot 3
On Type………….histograma Xlist………….HORA Freq…………..NC	Off Type………….histograma Xlist…………..HORA Freq…………...PE	Off Type…………histograma Xlist………….HORA Freq…………..AM

Con objeto de ver mejor el diagrama, introducimos en WINDOW estos valores: Ymin = 50, Ymax = 90. Pulsando GRAPH, obtenemos:

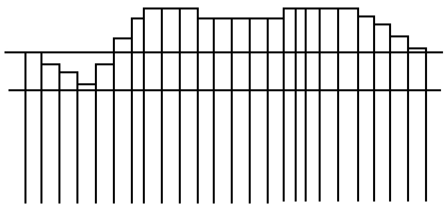


que es el gráfico correspondiente a Nuevo Centro. Utilizando TRACE y los cursores observamos que el valor mínimo es x = 66**.**1 a las 3 horas de la madrugada. Comprobamos también que la anchura de cada rectángulo es igual a 1 hora.

¿Se cumple en este punto de la ciudad la recomendación de la OCDE?. Para averiguarlo introducimos en el menú Y= la función $Y_1 = 65$. Al pulsar la tecla GRAPH, obtenemos:



Pero lo peor no es esto. Si volvemos a introducir en el menú Y= la función $Y_2 = \overline{x}$ (para lo que hay que pulsar VARS y seleccionar 5**:** Statistics y 2**:** $\overline{x}$), obtenemos esto otro:



lo que indica no solo que no se cumple la recomendación de la OCDE en ningún momento del día, sino que, además, durante casi todo el día el nivel de ruido en Nuevo Centro es superior a su propia media. De hecho, sólo hay cinco horas del día por debajo de la media (con TRACE y las teclas de cursor descubriremos que son desde las 0 hasta las 5 de la madrugada).

Si desactivamos el Plot 1 y activamos (On) el Plot 2 (con el Plot 3 desactivado) observaremos una situación parecida en la Plaza de España, aunque ahora nos encontramos con 7 horas por debajo de la media (desde las 23 horas hasta las 5 de la madrugada). Si desactivamos el Plot 1 y el Plot 2 y activamos (On) el Plot 3, volvemos a obtener algo similar en Ausias March, pero con 8 horas por debajo de la media (desde las 21 horas hasta las 5 de la madrugada).

PERO EN TODOS LOS CASOS EL NIVEL DE RUIDO ES SUPERIOR O IGUAL A 65 DB, ES DECIR, NO SE CUMPLE NUNCA LA RECOMENDACIÓN DE LA OCDE.

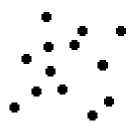
3. Análisis de regresión

Los diagramas de caja de la figura 1 parecen indicar una relación entre los niveles de contaminación acústica en Nuevo Centro y Plaza de España. Con objeto de averiguar si esto es realmente así, utilizaremos la TI–83 para construir el diagrama de dispersión de las variables NC y PE. Desactivamos todos los gráficos mediante [2nd] [STAT PLOT] **4:** PlotsOff [ENTER]. A continuación definimos de nuevo el Plot 1, de acuerdo con los siguientes valores:

Plot1
On Type……………..dispersión Xlist……………...NC Ylist……………...PE Mark……………..□

En el menú Y = borramos las dos funciones $Y_1 = 65$ e $Y_2 = \bar{x}$ mediante la tecla CLEAR. A continuación pulsamos ZOOM 9**:** ZoomStat y obtenemos la siguiente figura:

El diagrama de dispersión parece indicar una correlación lineal positiva fuerte entre las dos variables, si bien hay un cierto "apelotonamiento" de puntos en la parte superior. Activando TRACE y mediante ZOOM 1**:** Zbox señalaremos con ayuda de las teclas de cursor, un rectángulo alrededor de esa zona de puntos con objeto de ampliarla:



Utilizando TRACE y las teclas de cursor observaremos que dichos puntos corresponden a parejas de valores altos, como (73.4, 73.7), (73.9, 73.5), (73.9, 73.9) etc, siendo la primera componente el valor de la contaminación en Nuevo Centro y la segunda la contaminación en la Plaza de España. El hecho de que este apelotonamiento de los datos se produzca para valores altos parece estar de acuerdo con los diagramas de caja de las dos variables, en los que se observa que la mediana está muy cerca del tercer cuartil (asimetría a la izquierda). Esto indica que durante la mayor parte de las horas el nivel de ruido se mantiene en cotas altas. Y esto ocurre en las dos zonas consideradas.

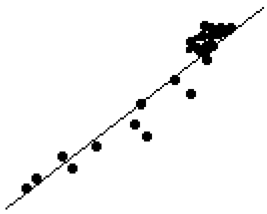
Mediante la instrucción LinReg(ax+b) LNC , LPE obtendremos los coeficientes a y b de la recta de regresión mínimo cuadrática, el coeficiente de determinación, r^2, y el coeficiente de correlación, r. Este comando se obtiene pulsando la siguiente secuencia de teclas: [STAT] CALC 4: LinReg(ax+b) [2nd] [LIST] NC , [2nd] [LIST] PE. Así obtenemos estos resultados:

LinReg
y = a x + b
a = 1.164177632
b = –12.24174726
r^2 = 0.912991638
r = 0.9555059592

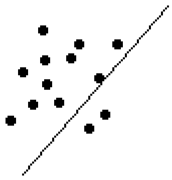
Por tanto, la correlación lineal es positiva fuerte, es decir los datos se podrían ajustar por una recta de ecuación Y = 1.16 X –12.24.

Se podría utilizar esta recta para saber qué nivel de ruido habrá en la Plaza de España cuando se conoce el nivel de ruido en Nuevo Centro. Así, si el nivel de ruido en Nuevo Centro fuese de 75 db, entonces en la Plaza de España sería de Y= 1.16 × 75 –12.24 = 74.76 ≅ 74.7 db.

Podemos comprobar gráficamente la bondad del ajuste de la nube de puntos por la recta obtenida. Para ello en el menú Y = definimos la función Y_1 = como la recta de regresión pulsando [VARS] 5: Statistics EQ 1: RegEQ. A continuación volvemos a la pantalla gráfica mediante ZOOM 9: ZoomStat. Con ello obtenemos:



Pero si hacemos un zoom en la zona de mayor densidad de la nube (utilizando [ZOOM] 1:Zbox obtenemos esta imagen:

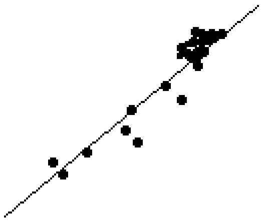


Así que parece que la recta deja más puntos a un lado de la nube que a otro. Recordemos sin embargo que la variable NC presenta dos outliers (para las 2 y las 3 de la madrugada). Si eliminamos esas dos observaciones extremas y repetimos el análisis de regresión, ¿obtendremos una recta que pase más cerca de todos los puntos?.

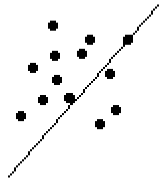
Para eliminar los dos outliers vamos primero a ordenar descendentemente las listas HORA, NC y PE. De esta forma, las dos últimas observaciones de la lista NC serán los dos outliers que queremos eliminar. Tecleamos: [STAT] 3: SortD([2nd] [LIST] NC , [2nd] [LIST] PE , [2nd] [LIST] HORA) para introducir el comando SortD(LNC, LPE, LHORA) con el cual se ordenará de mayor a menor la lista NC, arrastrando a las listas PE y HORA.

A continuación vamos al editor de listas y comprobamos que, efectivamente, la variable NC está ordenada de mayor a menor. Eliminamos las dos últimas observaciones de las listas NC y PE situando el cursor en cada dato y pulsando DEL. Por último, repetimos el análisis de regresión y la gráfica de correlación, junto con la recta de ajuste, obteniendo estos resultados:

LinReg
y=ax+b
a=1.269994917
b=–19.9739044
r^2=0.8827734315
r=0.939560233



nube de puntos sin outliers



zoom

Como vemos la situación es peor que la anterior, ya que hemos obtenido un coeficiente de determinación r^2 más pequeño. Tal vez se ajuste mejor a la nube de puntos la recta mediana – mediana. Para obtener su ecuación introducimos de nuevo los outliers que habíamos eliminado en el editor de listas y pulsamos [STAT] CALC 3: Med –Med [2^{nd}] [LIST] NC [2^{nd}] [LIST] PE . De esta forma introducimos el comando Med–Med LNC, LPE en la pantalla principal. Al pulsar ENTER obtenemos:

Med – Med
y = a x + b
a = 1.221153846
b= –16.13798077

Si mantenemos en el menú Y= la función Y_1 = RegEQ (recta de regresión mínimo-cuadrática) y en Y_2 introducimos la función ax+b (a y b se encuentran mediante [VARS] 5: Statistics EQ 2: a y 3: b), al pulsar de nuevo [ZOOM] 9: ZoomStat obtenemos la recta mediana-mediana y la de regresión junto con el diagrama de dispersión:

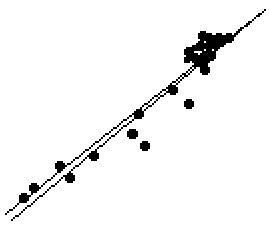
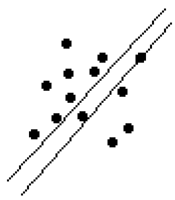


diagrama de dispersión



zoom

Podemos apreciar que la recta mediana – mediana está más centrada en la nube que la recta de mínimos cuadrados. Sin embargo, hay algo en la forma de la nube que hace que no quedemos satisfechos del todo. ¿Qué es?. Su falta de homocedasticidad, es decir el hecho de que la nube no sea uniforme, que la densidad de la nube no sea la misma en todas sus partes. Significa esto que posiblemente la correlación entre las variables NC y PE no es lineal, lo que nos conduciría a ensayar otros modelos (¿parabólico?, ¿cúbico?, ¿exponencial?, ¿logarítmico?). La búsqueda del modelo se convierte en una auténtica investigación para los estudiantes, donde afortunadamente la exploración se ve beneficiada por la enorme potencia de la calculadora gráfica.

Tras muchas pruebas y después de un largo periodo de búsqueda, probamos el modelo de regresión exponencial. Mediante el comando ExpReg LNC , LPE (que se obtiene pulsando [STAT] CALC 0:ExpReg [2^{nd}] LIST NC , [2^{nd}] LIST PE), llegamos al siguiente resultado:

ExpReg
y = a * b ^ x
a = 21.5868463
b = 1.016775173
r^2 = 0.9178692207
r = 0.9580549153

Vemos que el modelo exponencial explica casi el 92% de la variabilidad de los datos ($r^2 \approx 0.92$) ligeramente superior al modelo lineal ($r^2 \approx 0.91$). Sin embargo, al construir el diagrama de dispersión, junto con la curva de regresión (introduciendo la función $Y_1 = a * b \wedge X$ en el menú Y =, utilizando la tecla VARS), no se aprecian prácticamente diferencias con el modelo lineal. Por tanto, podemos aceptar el modelo lineal como explicativo de los datos, a pesar de la falta de homocedasticidad.

4. Una discusión importante

Eva.- ¿Por qué hay una correlación tan fuerte entre los niveles de contaminación acústica en Nuevo Centro y Plaza de España y en cambio no la hay entre Nuevo Centro y Ausias March?.

Raúl.- Como entre los dos primeros puntos hay una avenida, el flujo de vehículos debe ser parecido en los dos puntos. En cambio, Ausias March está bastante alejado de los dos puntos citados.

María.- La explicación reside en que los valores de las variables son cercanos a 70 dB, característica del tráfico de vehículos.

Eva no se ha dado cuenta de que ha introducido dos listas equivocadas en sus cálculos: la correlación entre el nivel de ruido en Nuevo Centro y Ausias March es más alta que entre Nuevo Centro y Plaza de España.

Raúl revela una confusión conceptual importante: está pensando que la correlación mide el grado de igualdad entre dos variables. Cuando otros compañeros discuten sus argumentos, él se aferra a la idea de que debe haber mayor correlación entre niveles de ruido parecidos.

María tiene razón: no hay más que leer las tablas para darse cuenta.

5. Buscando las causas

Llegados a este punto, concluimos que en todo este asunto de la contaminación acústica debe tener mucha influencia el creciente número de vehículos que cada vez crean más problemas de tráfico. Claro que, también el incremento de la población puede ser un factor que haga aumentar el nivel de ruido… Así que decidimos estudiar las relaciones con estas dos nuevas variables.

- **RUIDO Y VEHÍCULOS**

a) *La siguiente tabla indica el número de turismos matriculados y el nivel de ruido por trimestre en la ciudad de Valencia durante el periodo 1995 – 1998. Representa gráficamente estos datos, situando en el eje horizontal el número de turismos y en el eje vertical el número de decibelios. (Fuente: Revista DADES, Ayuntamiento de Valencia)*

  - *¿Cómo debes elegir las escalas de los ejes?.*
  - *¿Tiene sentido unir los puntos de la gráfica?.*
  - *A la vista de la gráfica, ¿crees que existe una relación entre el número de turismos y el nivel de ruido?.*
  - *Si la relación anterior existe, ¿crees que se puede expresar mediante una fórmula?.*
  - *¿Qué valor del coeficiente de correlación asignarías al diagrama nº de turismos → nº de decibelios que has construido anteriormente?.*

		Turismos matriculados	Nivel medio sonoro en dB
1995	Primer trimestre	4225	69,92
	Segundo trimestre	5687	69,97
	Tercer trimestre	3873	67,91
	Cuarto trimestre	5188	70,00
1996	Primer trimestre	4396	70,25
	Segundo trimestre	5099	70,52
	Tercer trimestre	4951	70,11
	Cuarto trimestre	5447	70,67
1997	Primer trimestre	4994	70,45
	Segundo trimestre	5955	70,82
	Tercer trimestre	5638	70,89
	Cuarto trimestre	5776	70,98
1998	Primer trimestre	5622	70,67
	Segundo trimestre	6937	71,34
	Tercer trimestre	6747	71,23
	Cuarto trimestre	7247	71,79

b) *Intenta construir lo más aproximadamente que puedas una recta que se ajuste a la nube de puntos, de forma que la distancia de los puntos a dicha recta sea lo menor posible.*

c) *¿Qué ocurrirá si continua aumentando el número de turismos matriculados?. ¿Qué nivel medio de ruido cabe esperar en el trimestre en que se matriculen 8000 turismos?. ¿Y cuando se matriculen 9000?. Si el ritmo de crecimiento se mantiene, ¿ha de pasar mucho tiempo para que se alcancen 10000 vehículos matriculados en un trimestre?.*

d) *¿Qué grado de seguridad te merecen las estimaciones que has hecho anteriormente?.*

e) *¿Te atreves a construir una fórmula que represente la recta de ajuste que has dibujado?. Compara los resultados con los de tus compañeros.*

La actividad propuesta no tendría sentido desarrollarla si solo se dispone de papel milimetrado y calculadora científica, ya que el tiempo necesario para desarrollarla sería tal vez excesivo, además de que la representación gráfica del diagrama de dispersión es, en este caso, bastante compleja. Así que utilizamos directamente la calculadora gráfica.

En el editor de listas definimos dos nuevas listas de nombres TM (Turismos matriculados) y NMS (Nivel medio sonoro) e introducimos en ellas los valores de la tabla anterior. A continuación desactivamos todos los gráficos con [2nd] STAT PLOT 4**:** PlotsOff y definimos el plot 1 de la siguiente forma:

Plot 1
On Type…………………..dispersión Xlist…………………..TM Ylist…………………..NMS Mark………………….□

Comprobamos que no hay ninguna función en el menú Y = (si es necesario borramos las que estén, utilizando CLEAR) y pulsamos ZOOM 9**:** ZoomStat, obteniendo la nube de puntos:

Parece que entre las variables TM y NMS hay, efectivamente, una correlación lineal positiva fuerte. Para confirmarlo, recurrimos al análisis de regresión, mediante el comando LinReg(ax+b) LTM , LNMS (que se obtiene pulsando [STAT] CALC 4**:** LinReg(ax+b) [2nd] LIST TM , [2nd] LIST NMS). Los resultados son los siguientes:

LinReg
y = a x + b
a = 7.6706638E – 4
b = 66.2615862
r^2 = 0.6988615903
r = 0.8359794198

Esto indica que un modelo lineal explicaría aproximadamente el 70% de la variabilidad de los datos ($r^2 \approx 0.70$). Y la recta que ajusta los datos (ajuste mínimo – cuadrático) tiene por ecuación: y = 0.00077 x + 66.

Así, si el número de vehículos es de TM = 8000, entonces el nivel medio sonoro es:
$y = 0.00077 \times 8000 + 66 = 72.16$ db.

Si el número de vehículos matriculados es de TM = 9000, entonces el nivel medio sonoro es:
$y = 0.00077 \times 9000 + 66 = 72.93$ db.

No sabemos cuando la tasa de matriculación de vehículos en Valencia ciudad alcanzará los 10000 por trimestre, pero suponiendo que TM = 10000, entonces el nivel de ruido será:
$y = 0.00077 \times 10000 + 66 = 73.7$ db

¡Resulta que en el año 1997 ya se alcanzaron los 74.3 db en algunos puntos de la ciudad (por ejemplo en la Plaza de España)!. Un vistazo a la siguiente tabla (en la que hemos hallado la media anual de NMS en el periodo 95 – 98 indica claramente que, si se mantiene la tendencia, no tardará mucho el momento en que se alcancen los 10000 vehículos matriculados por trimestre:

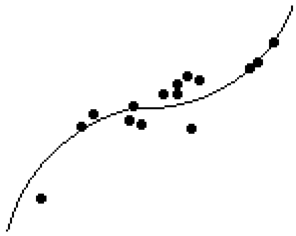
AÑO	NMS (media)
1995	69.45
1996	70.39
1997	70.78
1998	71.26

Claro que… puede que el ajuste lineal no sea el mejor, porque la forma de la nube de puntos recuerda bastante a una cúbica, de forma que a lo mejor lo apropiado es hacer una regresión cúbica.

Y efectivamente, así ocurre, porque al introducir en la TI – 83 el comando CubicReg LTM , LNMS (por el procedimiento habitual), obtenemos:

CubicReg
$y = ax^3 + bx^2 + cx + d$
a = 3.435214E −10
b = −5.8925E −6
c = 0.0337447284
d = 6.1088409
R^2 = 0.8367535319

La regresión cúbica explica el 84% de la variabilidad de los datos. De forma que obtendríamos mayor fiabilidad en nuestras predicciones si ajustásemos la nube de puntos por una función polinómica de tercer grado.



6. ¿Contribuimos nosotros a la contaminación?

Con objeto de implicar más a los estudiantes y para que reflexionen más sobre las propias actitudes en torno al medio ambiente, investigamos la posible relación entre el número de motocicletas matriculadas en los últimos años y el nivel medio sonoro. Es muy frecuente que los estudiantes acudan al centro utilizando este tipo de vehículos. ¿Influyen mucho las motos en el nivel de ruido de la ciudad?

- **RUIDO Y MOTOCICLETAS**

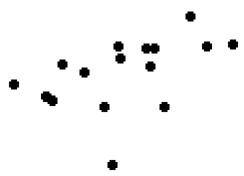
La siguiente tabla muestra el número de motocicletas matriculadas cada trimestre en la ciudad de Valencia, durante el período 1995 – 1998. Haz un estudio parecido al de las actividades anteriores para averiguar el tipo de correlación que existe entre el número de motos matriculadas y el nivel de ruido. (Fuente: revista DADES, Ayuntamiento de Valencia).

		Motocicletas matriculadas	**Nivel medio sonoro en dB**
1995	Primer trimestre	210	69,92
	Segundo trimestre	271	69,97
	Tercer trimestre	202	67,91
	Cuarto trimestre	154	70,00
1996	Primer trimestre	121	70,25
	Segundo trimestre	265	70,52
	Tercer trimestre	152	70,11
	Cuarto trimestre	169	70,67
1997	Primer trimestre	197	70,45
	Segundo trimestre	268	70,82
	Tercer trimestre	261	70,89
	Cuarto trimestre	227	70,98
1998	Primer trimestre	226	70,67
	Segundo trimestre	355	71,34
	Tercer trimestre	328	71,23
	Cuarto trimestre	310	71,79

- *Dibuja el diagrama de dispersión y asígnale un coeficiente de correlación.*
- *Dibuja lo más aproximadamente que puedas la recta de ajuste.*
- *¿Qué nivel de ruido cabe esperar cuando el número de motos matriculadas sea de 500?.*
- *¿Hasta qué punto es fiable la estimación anterior?.*
- *Compara la recta de ajuste con la que obtuviste en la actividad anterior.*
- *¿Qué crees que influye más en el nivel medio de ruido, el número de turismos matriculados o el de motocicletas?.*

En el editor de listas introducimos una nueva variable MM (= motos matriculadas). Después desactivamos todos los gráficos estadísticos mediante [2nd] STAT PLOT 4**:** PlotsOff. A continuación obtenemos el diagrama de dispersión, definido por los siguientes valores:

Plot 1
On Type………………………dispersión Xlist………………………MM Ylist………………………NMS Mark………………………□



El diagrama presenta gran dispersión y muestra que la correlación es bastante débil. Introduciendo el comando LinReg(ax+b) LMM , LNMS (mediante el procedimiento habitual), obtenemos estos resultados:

LinReg
y = a x + b
a = 0**.**0067997118
b = 68**.**89076694
r^2 = 0**.**2744348122
r = 0**.**5238652625

Para regocijo de los “moteros”, el coeficiente de determinación es realmente bajo, de manera que el modelo lineal sólo explicaría un 27% de la variabilidad de los datos. La influencia de las motocicletas en el nivel medio de ruido existe, pero es realmente escasa. Podríamos hablar de correlación positiva débil. De todas formas, también podemos hacer una lectura al revés: puesto que la influencia se puede cuantificar en un 27%, si eliminamos las motos ganaríamos un 27% de salud auditiva. Claro que esto no parece convencer demasiado a nuestros estudiantes…

Evidentemente, con este coeficiente de determinación no hay ninguna fiabilidad, de manera que no tendría mucho sentido hacer predicciones. No obstante, si nos atrevemos a hacerlas, para un ritmo de matriculación trimestral de 500 motos (MM = 500), el nivel de ruido sería:
$y = 0.0068 \times 500 + 68.9 = 72.3$ db. Pero esta estimación es realmente muy poco fiable.

Por último, decidimos investigar la relación entre el crecimiento de la población y el nivel de ruido. Parece evidente que a mayor población, mayor incremento en el nivel sonoro. Así que nos pusimos manos a la obra.

- **RUIDO Y POBLACIÓN**

El nivel sonoro de una ciudad ¿depende del número de habitantes?. En la siguiente tabla tienes datos de la evolución de la población mayor de 15 años en la ciudad de Valencia durante el período 1995 – 1998. Se señala también el nivel medio de ruido en ese periodo. Investiga el tipo de correlación existente entre dichas magnitudes. (Fuente: revista DADES, Aytto. de Valencia).

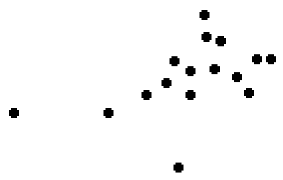
		Población de 16 y más años (en miles)	**Nivel medio sonoro en dB**
1995	Primer trimestre	603,0	69,92
	Segundo trimestre	606,4	69,97
	Tercer trimestre	609,7	67,91
	Cuarto trimestre	613,1	70,00
1996	Primer trimestre	608,0	70,25
	Segundo trimestre	609,1	70,52
	Tercer trimestre	610,2	70,11
	Cuarto trimestre	611,3	70,67
1997	Primer trimestre	612,4	70,45
	Segundo trimestre	613,3	70,82
	Tercer trimestre	614,2	70,89
	Cuarto trimestre	609,4	70,98
1998	Primer trimestre	610,2	70,67
	Segundo trimestre	611,0	71,34
	Tercer trimestre	611,7	71,23
	Cuarto trimestre	—	71,79

- *¿Crees que la tendencia de la población es creciente o decreciente?.*
- *Dibuja el diagrama de dispersión y asígnale un coeficiente de correlación.*
- *Con los datos de que dispones, ¿crees que es útil dibujar una recta de ajuste?. ¿Por qué?.*
- *Dibuja una recta de ajuste. Explica las dificultades que encuentres.*
- *¿Qué nivel de ruido cabe esperar cuando el número de habitantes mayores de 15 años sea de 615000?. ¿Consideras fiable esta estimación?.*
- *¿Qué crees que influye más en el nivel medio de ruido, el número de vehículos matriculados o el de habitantes?.*
- *¿Qué medidas crees más adecuadas para mejorar la situación?.*

En el editor de listas introducimos la nueva variable POB con los valores de la población de la ciudad de Valencia con edad mayor o igual que 16 años. ¿Qué hacer con el dato faltante del cuarto trimestre de 1998?. Como es desconocido, hemos optado por sustituirlo por la media de los tres primeros trimestres de ese año, que aproximadamente es 611. (No habría sido "justo" sustituirlo por la media de todos los años, porque es evidente que ha habido un crecimiento de la población). De todas formas, no es fácil adoptar una decisión en estos casos y se suscita siempre mucha discusión entre los estudiantes. Hay quien prefiere poner la media de los cuartos trimestres…

A continuación construimos el diagrama de dispersión, con el procedimiento habitual:

Plot 1
On Type………………………dispersión Xlist………………………POB Ylist………………………NMS Mark………………………□



El diagrama es más disperso que el de las motos. De forma que parece que no hay correlación lineal entre las dos variables. Hacemos un análisis de regresión, utilizando el comando LinReg(ax+b) LPOB , LNMS (con el procedimiento de siempre) y obtenemos unos resultados que confirman nuestras sospechas:

LinReg
y = a x + b
a = 0.0995975338
b = 9.690604984
r^2 = 0.1032109371
r = 0.3212645905

No hay correlación lineal entre las variables. Puesto que $r^2 = 0.10$, el modelo lineal sólo explica el 10% de la variabilidad de los datos, lo que es realmente poco.

En contra de lo que pudiera parecer por la intuición, el incremento de la población no tiene excesiva influencia en el nivel medio sonoro. LO QUE REALMENTE TIENE INFLUENCIA ES EL INCREMENTO EN EL NÚMERO DE VEHÍCULOS MATRICULADOS. ESTO QUIERE DECIR QUE PARA DISMINUIR LOS NIVELES DE CONTAMINACIÓN ACÚSTICA Y SITUARLOS, POR EJEMPLO, EN LOS NIVELES RECOMENDADOS POR LA O.C.D.E., DEBERÍAN ESTABLECERSE MEDIDAS PARA:

A) LIMITAR EL NÚMERO DE MATRÍCULAS

B) HACER VEHÍCULOS MENOS RUIDOSOS.

C) LIMITAR EL TRÁFICO EN LAS CIUDADES.

Como la primera medida es impopular y la segunda no está al alcance de la técnica por el momento, la única posibilidad es limitar el tráfico en las ciudades. LO QUE NO ES FÁCIL.

7. Conclusiones

La calculadora gráfica se ha mostrado en estas actividades como un instrumento imprescindible para el análisis de datos. Hubiera sido muy difícil dibujar los diagramas de dispersión con lápiz y papel o hacer los cálculos para obtener los parámetros estadísticos o las rectas de regresión con una calculadora científica. El tiempo dedicado al desarrollo de estas actividades habría sido también más extenso si no se dispone de una calculadora gráfica.

La calculadora gráfica permite ahorrar tiempo y dedicarlo más a la reflexión sobre los datos y a los procesos seguidos, así como a las decisiones que se deben adoptar.

Por otra parte, la calculadora gráfica no es un simple instrumento de cálculo, sino que, por sí misma genera conceptos y permite formular conjeturas. Como con ella se pueden procesar gran cantidad de datos en poco tiempo, es posible hacer simulaciones y comprobar conjeturas en un tiempo relativamente corto. Lo que supone una ayuda inestimable al estudiante, porque le permite dedicar más tiempo a la reflexión. De esta forma los contenidos se ven modificados, porque el foco de atención deja de ser el cálculo para ser el proceso y la comprensión de los conceptos.

Además la calculadora gráfica permite acercar al estudiante a conceptos, procedimientos, técnicas y contenidos que hasta ahora no estaban al alcance de la educación secundaria, especialmente por su dificultad de cálculo. Así ocurre, por ejemplo, con la llamada “Estadística Matemática” o Inferencia Estadística y el Análisis de Regresión.

Todo este proceso también podría desarrollarse si se utilizase un ordenador en el aula con software adecuado (y hay donde elegir: EXCEL, STATGRAPHICS, SPSS para Windows, etc). Sin embargo, todos sabemos las limitaciones que supone trabajar con ordenador. (Hay que trasladar a los estudiantes al aula de Informática, que no siempre está disponible por problemas de horario, además la capacidad del aula de Informática no siempre coincide con el número de estudiantes de un grupo, el ordenador no es tan fácilmente transportable como la calculadora,…). Parece, pues, que el uso de la calculadora gráfica tiene grandes ventajas que sería lamentable no aprovechar.

Teniendo en cuenta la evolución del mercado y, sobre todo, las necesidades crecientes de su uso en la educación matemática, no es ninguna exageración afirmar que su uso estará muy generalizado dentro de 4 ó 5 años. Tan generalizado como lo está hoy el uso de las calculadoras científicas.

ANÁLISIS DE REGRESIÓN CON LA CALCULADORA GRÁFICA CLASSPAD 300 CASIO

Introducción

Para ajustar una nube de puntos a una función usando un modelo lineal o de otro tipo es necesario utilizar recursos tecnológicos apropiados. La ClassPad 300 permite ensayar diversos modelos de regresión (polinómicos, exponenciales, logarítmicos, trigonométricos, etc), además de obtener los parámetros estadísticos necesarios para la obtención de la función de regresión y hacer estimaciones y predicciones.

A continuación estudiaremos algunas de las posibilidades de la ClassPad 300 para el análisis de datos y el análisis de regresión en ESO y Bachillerato.

1. Gráficos de regresión

a) <u>GRÁFICOS ESTADÍSTICOS DE DOS VARIABLES</u>

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

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

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

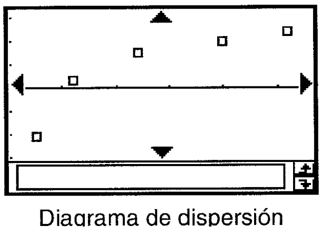


Diagrama de dispersión

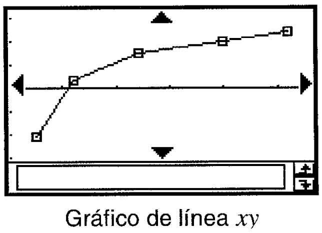


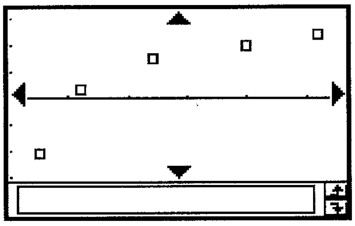
Gráfico de línea *xy*

- **Diagrama de dispersión y gráfico de regresión**

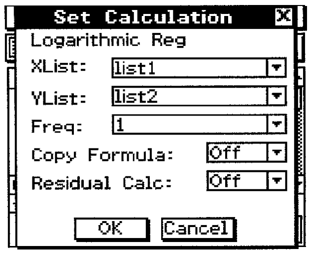
- Dibuja el diagrama de dispersión de los siguientes datos, realiza una regresión logarítmica de los datos para ver los parámetros de regresión y dibuja el gráfico de regresión. Sigue los siguientes pasos:

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

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

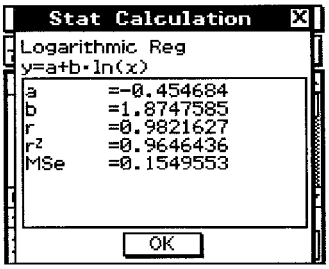


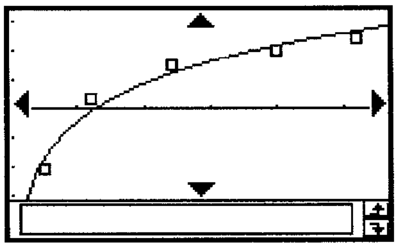
4) Selecciona el comando Calc / Reg. Logarítmica. En la ventana Definir cálculo introduce los valores de la siguiente ventana y toca el botón [Acep.].



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

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



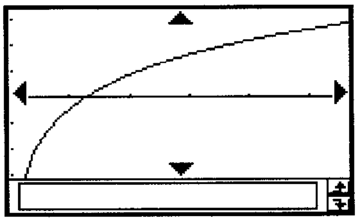


- **Gráfico de regresión**

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

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

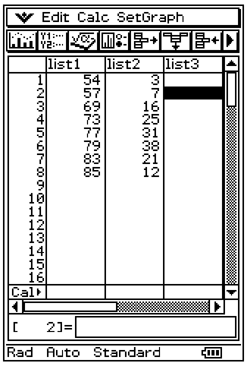
1) Selecciona el comando ConfGraf / Opciones... o toca el botón .
2) En el cuadro de diálogo de configuración de gráficos, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrLog., ListaX=list1, ListaY=list2, Frec.=1. Toca el botón [Def.].
3) Toca el botón para trazar el gráfico de regresión.



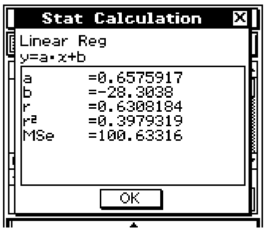
- **Gráfico de regresión lineal**
- Halla la ecuación de la recta de regresión lineal correspondiente a los siguientes datos. Sigue los siguientes pasos:

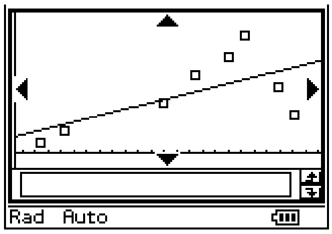
List1	54	57	69	73	77	79	83	85
List2	3	7	16	25	31	38	21	12

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



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

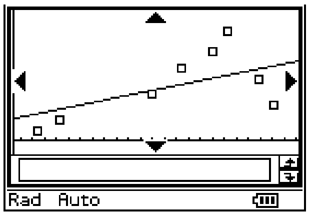




- **Gráfico Med-Med**
- Halla la ecuación de la recta mediana-mediana correspondiente a los siguientes datos. Sigue los siguientes pasos:

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

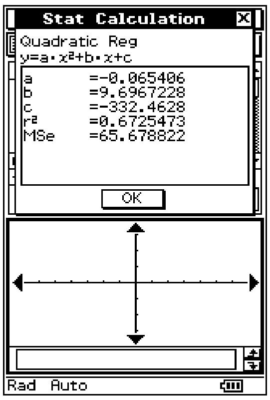
1) Con la ventana de gráficos activada, selecciona el comando Calc. / Línea MedMed.
2) Toca el botón [Acep.] de la ventana Definir cálculo para aceptar las opciones que aparecen.
3) En la ventana Calc. Estadístico se muestra la ecuación de la recta mediana-mediana. Toca el botón [Acep.].
4) Toca el botón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

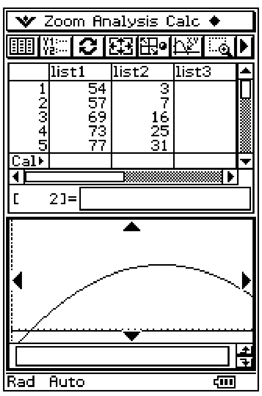


- **Gráficos de regresión cuadrática, cúbica y cuártica**
- Halla la ecuación de las curvas de regresión cuadrática, cúbica y cuártica correspondiente a los siguientes datos. Sigue los siguientes pasos:

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

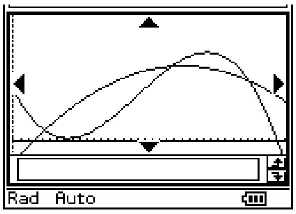
1) Con la ventana de gráficos activada, selecciona el comando Calc. / Regresión cuadrática.
2) Toca el botón [Acep.] de la ventana Definir cálculo para aceptar las opciones que aparecen.
3) En la ventana Calc. Estadístico se muestra la ecuación de la recta mediana-mediana. Toca el botón [Acep.].
4) Toca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





5) Toca la ventana de listas para activarla y selecciona el comando ConfGraf / Opciones... En la siguiente ventana introduce las opciones: Dibujo: On, Tipo: RegrCubic, ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Def.].

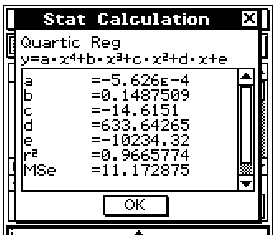
6) Toca el botón  para trazar el gráfico de regresión. Observa la curva de regresión obtenida.

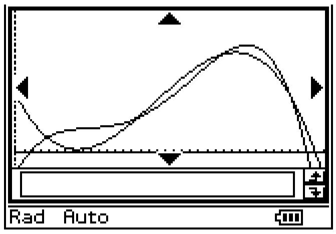


7) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. cuarto orden. En la siguiente ventana, toca el botón [Acep.].

8) Aparece la ecuación de la curva de regresión de cuarto grado. Toca el botón [Acep.]

9) Toca el botón  para que se dibuje la curva de regresión. Observa que se aproxima a los datos mejor que las curvas anteriores. Esto está indicado por el coeficiente de determinación r^2 que es superior a los casos anteriores.

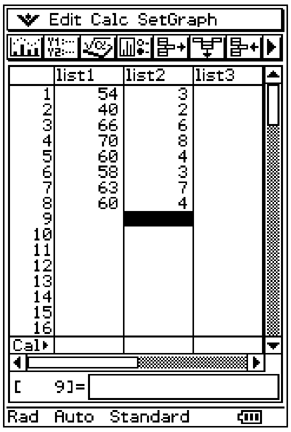




- **Gráficos de regresión logarítmica, exponencial y potencial**
- La siguiente tabla recoge las puntuaciones obtenidas en un test sobre visión espacial (T) y sus correspondientes calificaciones en la asignatura de Dibujo (D). Halla la ecuación de las curvas de regresión logarítmica, exponencial y potencial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

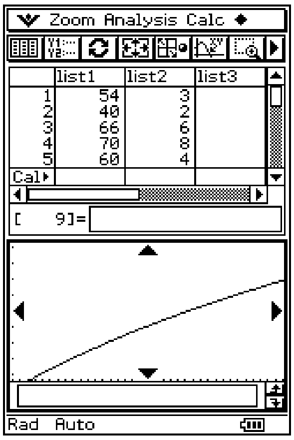
T	54	40	66	70	60	58	63	60
D	3	2	6	8	4	3	7	4

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



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

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

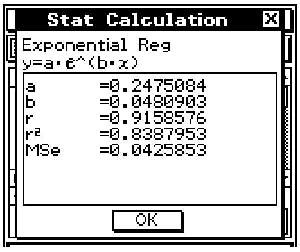


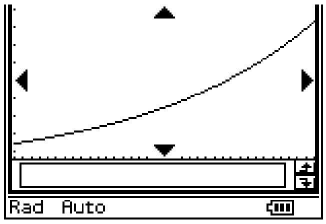
4) Con la ventana de gráficos activada, selecciona el comando ♦ / Borrar todo.

5) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. exponencial. En la siguiente ventana selecciona las siguientes opciones ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].

6) En la ventana Calc. Estadístico se muestra la ecuación de la curva de regresión exponencial, que es de la forma $y = a \cdot e^{bx}$. Se indica también el valor del coeficiente de determinación r^2 que indica la bondad del ajuste de la nube de puntos por la curva de regresión.

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

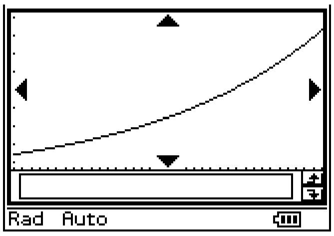




8) Con la ventana de gráficos activada, selecciona el comando ♦ / Borrar todo.

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

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

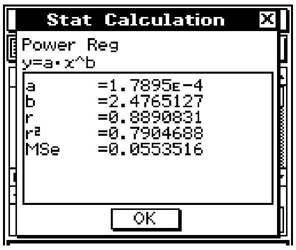


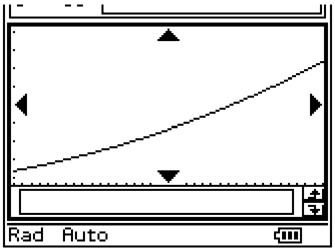
11) Con la ventana de gráficos activada, selecciona el comando ♦ / Borrar todo.

12) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. potencial. En la siguiente ventana selecciona las opciones ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].

13) En la ventana Calc. Estadístico aparece la ecuación de la curva de regresión, que es de la forma $y = a \cdot x^b$ y el valor del coeficiente de determinación r^2. Toca el botón [Acep.].

14) Toca el botón  para dibujar la curva de regresión exponencial. Observa el resultado.

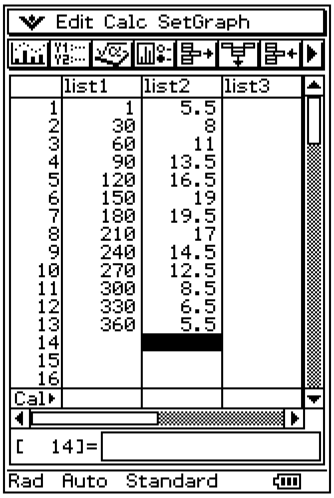




- **Gráfico de regresión sinusoidal**
- Calcula la curva de regresión correspondiente al número de horas de luz solar en Alaska durante un año, tal como se indica en la siguiente tabla. Sigue los siguientes pasos:

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

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



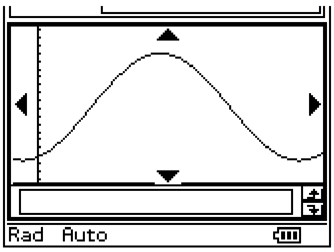
2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrSin, ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Def.].

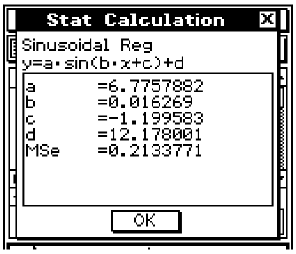
3) Toca el botón  para dibujar el gráfico de regresión sinusoidal.

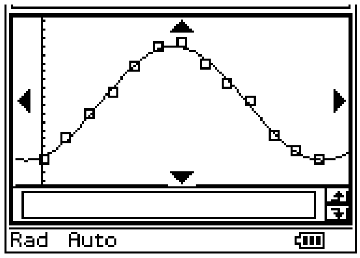
4) Con la ventana de gráficos activada, selecciona el comando Calc / Regr. Sinusoidal. En la siguiente ventana toca el botón [Acep.] y se mostrará una ventana con la ecuación de la curva de regresión, que es de la forma y=a·sin(b·x+c)+d.

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

6) Toca el botón para visualizar simultáneamente el gráfico de regresión sinusoidal y la nube de puntos. Observa el resultado.



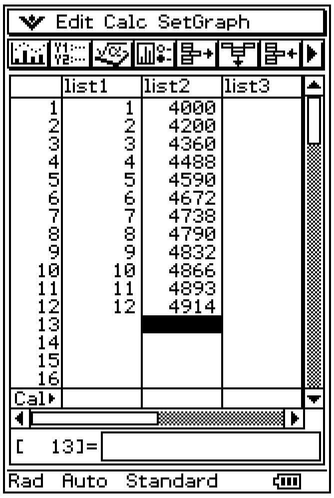




- **Gráfico de regresión logística**
- El número de árboles de un bosque durante 12 años crece de la forma que se indica en la siguiente tabla. Calcula la curva de regresión logística. Sigue los siguientes pasos:

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

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



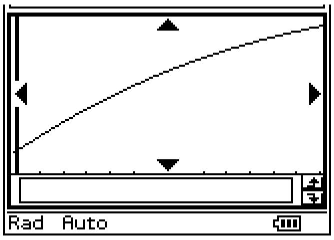
2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... En la siguiente ventana, selecciona las siguientes opciones: Dibujo: On, Tipo: RegrLogis, ListaX=list1, ListaY=list2. Toca el botón [Def.].

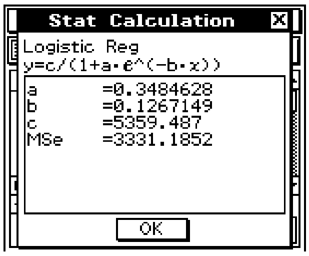
3) Toca el botón para dibujar el gráfico de regresión logística.

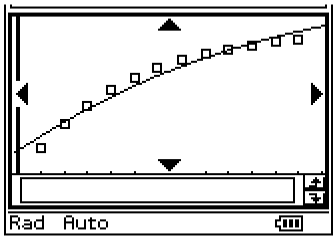
4) Con la ventana de gráficos activada, selecciona el comando Calc / Regresión logística. En la siguiente ventana toca el botón [Acep.] y se mostrará una ventana con la ecuación de la curva de regresión, que es de la forma $y = \dfrac{C}{1 + a \cdot e^{-b \cdot x}}$.

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

6) Toca el botón para visualizar simultáneamente el gráfico de regresión logística y la nube de puntos. Observa el resultado.



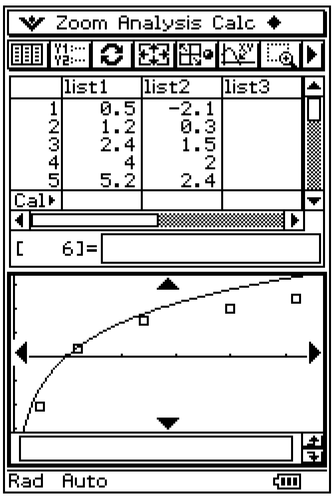




- **Gráfico de función y gráfico estadístico superpuestos**
- Dibuja el diagrama de dispersión de los siguientes datos y superpón dicho gráfico con el gráfico de la función y=2·ln(x). Sigue los siguientes pasos:

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

1) En la ventana del editor de listas, introduce los datos de la tabla en las listas list1 y list2.
2) Con la ventana de listas activada, selecciona el comando ConfGraf / Opciones... y elige las siguientes opciones: Dibujo: On, Tipo: Disper., ListaX=list1, ListaY=list2, Frec=1, Marca=cuadrado. Toca el botón [Def.].
3) Toca el botón para dibujar la nube de puntos. Observa el resultado.
4) Toca la ventana de listas, para hacer que ésta sea la ventana activa, y luego toca el botón para abrir la ventana del editor de gráficos.
5) Introduce la fórmula de la función y=2·ln(x) en la línea y1=.
6) Selecciona el comando / Cerrar para cerrar la ventana del editor de gráficos.
7) Selecciona el comando ConfGraf / Función gráfica.
8) Toca el botón para dibujar el gráfico de la función. De esta forma se visualiza simultáneamente el gráfico de la función y la nube de puntos.



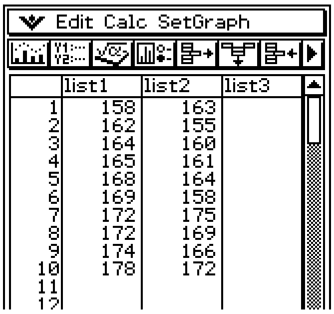
b) CÁLCULOS ESTADÍSTICOS

- **Cálculos estadísticos de dos variables**
- Las estaturas de 10 chicas y de sus respectivas madres son las siguientes, datos en cm:

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

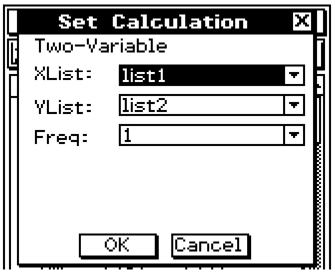
Calcula los parámetros estadísticos de dichas variables: media, desviación típica, suma de productos, mínimo, máximo. Sigue los siguientes pasos:

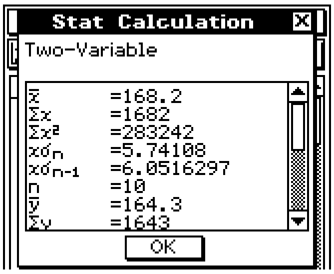
1) En la ventana del editor de listas, introduce los datos de las chicas en la lista list1 y los datos de las madres en la list2.

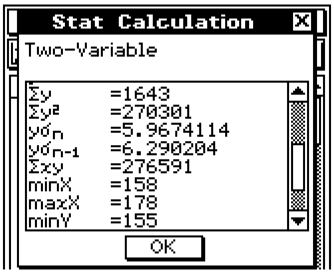


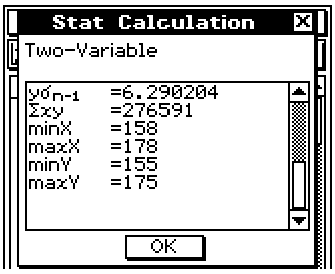
2) En la barra de menús, selecciona el comando Calc / Dos variables.

3) En el cuadro de diálogo que aparece, selecciona el nombre list1 como ListaX, el nombre list2 como ListaY y Frec=1. Toca el botón [Acep.].









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

$\bar{x}$	Media de la lista X	$\bar{y}$	Media de la lista Y
Σx	Suma de la lista X	Σy	Suma de la lista Y
Σx^2	Suma de cuadrados de la lista X	Σy^2	Suma de cuadrados de la lista Y

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

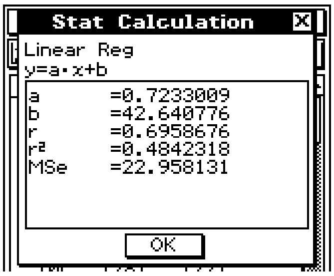
- **Cálculos estadísticos de regresión**
- Las estaturas de 10 chicas y de sus respectivas madres son las siguientes, datos en cm:

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

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

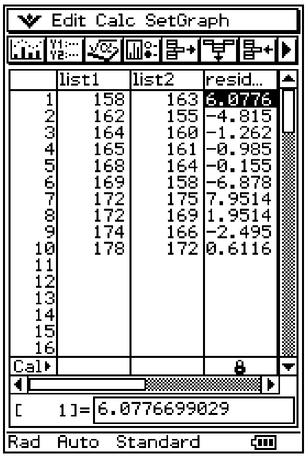
1) Con la ventana de listas activa, selecciona el comando Calc / Regresión lineal. En la siguiente ventana selecciona las opciones ListaX=list1, ListaY=list2, Frec=1. Toca el botón [Acep.].

2) Aparece una ventana con la ecuación de la recta de regresión y el valor del coeficiente de correlación, así como el valor del coeficiente de determinación.



3) Como el coeficiente de determinación es pequeño, el modelo lineal es poco explicativo de los datos. Para analizar los residuos, selecciona el comando Calc. / Regresión lineal. En la siguiente ventana selecciona la opción Calc. Residual / On y toca el botón [Acep.].

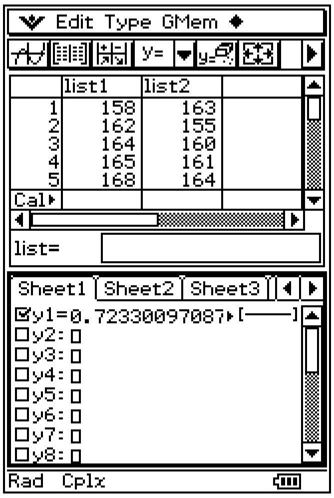
4) Toca el botón [Acep.] de la siguiente ventana. Para ver los residuos, toca el cuadro de nombre de la list3. En la caja List= introduce el nombre "residual", utilizando el teclado virtual [abc]. Toca el botón [Ejec.].



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

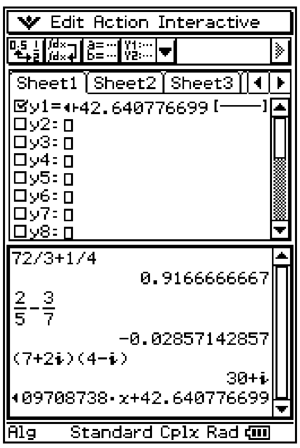
6) Vamos a estimar el valor de list2 cuando list1=167. Para ello, selecciona el comando Calc / Regresión lineal. En el siguiente cuadro de diálogo selecciona de la lista desplegable Copiar fórmula la línea (y1 a y20) en la que quieres copiar la fórmula. Por ejemplo, la línea y1. Toca el botón [Acep.].

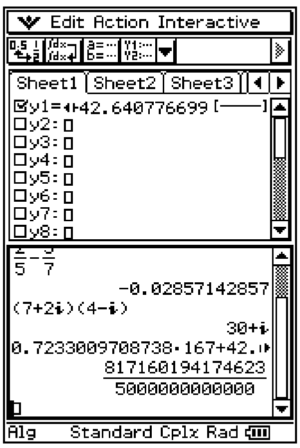
7) En la siguiente ventana toca el botón [Acep.]. Toca el botón Y1: Y2: y comprueba que la ecuación de la recta de regresión se ha copiado en la línea Y1=.

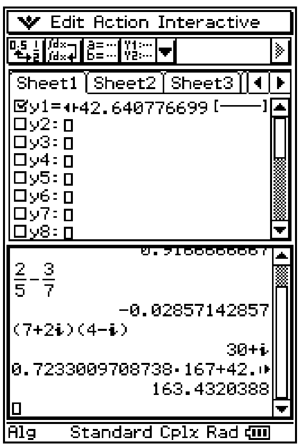


8) Para obtener la estatura de la madre correspondiente a una estatura de 167 cm para la hija, basta averiguar Y1(167). Selecciona el cuadro de marcación de la función Y1 y desactiva todos los demás.

9) Toca el botón Main para abrir la ventana Principal. Con el editor de gráficos y la ventana Principal en pantalla, selecciona la expresión a la derecha de Y1= y arrástrala con el lápiz táctil hasta una entrada vacía de la ventana Principal. Observa que la ecuación de la recta de regresión aparece en la ventana principal. Selecciona la variable x y, en su lugar, con ayuda del teclado virtual, introduce la estatura de la hija, 167 cm, y toca el botón [Ejec.]. Selecciona la expresión anterior y toca el botón para obtener el resultado en forma decimal.

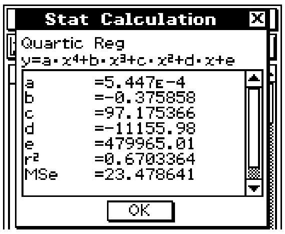




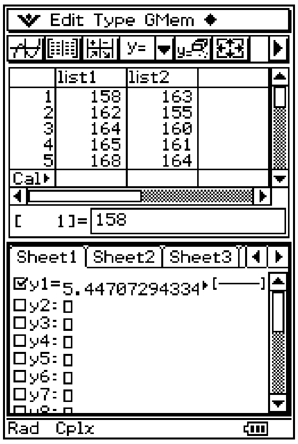


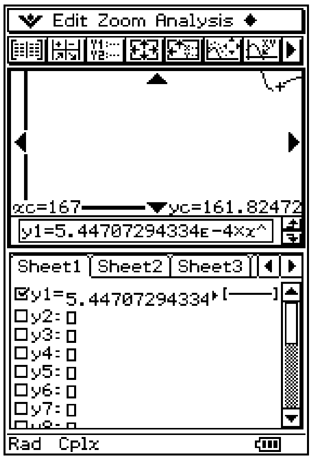
Por tanto, según el modelo lineal, la estatura de la madre, cuando la de la hija es de 167 cm, es de 163,4 cm. Pero esta estimación es poco fiable, debido al valor del coeficiente de determinación.

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



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





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

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

ACTIVIDADES SOBRE DESCRIPCIÓN, ANÁLISIS DE DATOS Y REGRESIÓN CON LA CALCULADORA GRÁFICA

En las siguientes páginas te mostramos un conjunto de tablas con informaciones y datos procedentes de diversas fuentes. En cada caso se pide que utilices la calculadora gráfica para efectuar un análisis estadístico de los datos, obteniendo los parámetros más significativos, confeccionando los diagramas más representativos, comparando las poblaciones, analizando las series temporales (si las hay) y ajustando, si es el caso, una curva de regresión.

1. El clima

1. Precipitació total en l/m2 i nombre de dies de precipitació. 1994-1998

	1994		1995		1996		1997		1998
	l/m2	dies	l/m2	dies	l/m2	dies	l/m2	dies	l/m2
Gener	1,1	4	19,0	1	38,3	17	90,7	21	109,1
Febrer	4,4	5	16,9	4	29,0	9	4,1	1	26,5
Març	0,7	4	4,5	6	14,3	12	6,6	2	5,5
Abril	55,1	7	8,3	5	17,8	9	41,8	11	11,1
Maig	13,4	9	1,9	4	29,9	10	24,8	10	68,2
Juny	3,1	4	9,9	11	3,3	3	15,7	6	4,9
Juliol	2,1	3	3,6	2	1,7	5	11,1	7	0,2
Agost	1,6	7	23,0	8	11,1	6	24,2	5	13,6
Setembre	163,4	8	51,8	8	90,3	11	84,0	14	16,9
Octubre	83,4	13	29,9	9	4,1	3	10,4	10	7,8
Novembre	22,4	6	5,1	10	69,4	9	14,5	11	20,3
Desembre	3,8	5	105,8	16	77,6	17	29,3	13	123,3
Total	354,5	75	279,7	84	386,8	111	357,2	111	407,4

Font: Institut Nacional de Meteorologia. Centre Meteorològic Zonal de València

2.Temperatures absolutes en °C. 1994-1998

	1994		1995		1996		1997		1998	
	Màx	Mín	Màx	Mín	Màx	Mín	Màx	Mín	Màx	Mín
Gener	24,6	1,8	24,0	3,0	21,8	6,0	18,8	2,4	22,5	2,2
Febrer	23,5	3,6	26,6	4,8	23,4	2,8	26,4	5,2	24,4	6,5
Març	27,6	7,4	28,0	5,8	26,8	5,0	28,4	7,2	29,2	6,0
Abril	28,0	5,0	24,6	6,2	25,2	7,6	32,5	8,4	27,6	7,6
Maig	33,8	10,8	34,4	11,8	28,8	11,2	29,6	10,0	28,0	10,4
Juny	33,6	12,6	30,7	14,0	34,0	14,2	36,2	15,6	38,2	15,5
Juliol	35,5	19,6	36,0	16,6	37,8	16,6	32,2	17,0	36,5	19,6
Agost	42,5	22,3	33,2	18,8	37,6	17,0	34,2	17,2	34,4	16,6
Setembre	35,2	11,8	34,3	11,6	32,2	13,6	33,5	17,8	37,2	16,0
Octubre	25,7	12,8	28,0	13,2	30,2	10,6	34,5	9,2	31,3	10,6
Novembre	25,2	8,5	29,9	7,0	25,6	5,4	26,2	6,8	29,0	4,2
Desembre	23,2	1,0	23,6	5,8	24,2	1,6	23,0	2,6	24,1	2,6
Total	42,5	1,0	36,0	3,0	37,8	1,6	36,2	2,4	38,2	2,2

Font: Institut Nacional de Meteorologia. Centre Meteorològic Zonal de València

2. Actividades económicas

1. Passatgers segons línies, CVT. 1996-1998

	1996	1997	1998
València-Torrent-El Vedat	2.680.132	2.523.123	2.462.600
València-Paterna-Plantío	792.610	778.033	772.152
Paterna-B.La Coma	-	-	-
Esta.Ademús-N.Facultats.	127.233	409.204	550.177
València-Manises	1.985.664	1.957.387	1.811.485
València-Accés Ademús	738.163	655.652	525.933
València-Rocafort-Sta.Bàrbara	5.675	5.070	2.195
València-Serra	106.432	102.658	100.490
València-Portaceli	45.000	45.000	73.196
Esta.Ademús-B.La Coma-P.Tecnològic	-	36.908	163.010
Esta.Ademús-Rocafort-Sta Bàrbara	-	-	6.445
València-Aeroport	80.720	-	-
València-L' Eliana-Benaguasil	118.255	121.526	127.614
València-Llíria-Gestalgar	619.250	585.266	563.299
València-Paiporta-Picanya	34.471	31.080	12.336
FGV	127.233	-	-
Total	7.333.605	7.250.907	7.170.932

Font: Consorci Valencià de Transports

2. Desocupació registrada. 1990-1999

	València	**L'Horta**	**València / l'Horta**	**Comunitat Valenciana**	**L'Horta /C.Valenciana**
31/03/90	51.238	92.242	55,5	261.561	35,3
31/03/91	51.334	93.762	54,7	261.583	35,8
31/03/92	48.665	77.986	62,4	254.757	30,6
31/03/93	54.481	98.538	55,3	287.976	34,2
31/03/94	61.072	110.922	55,1	309.593	35,8
31/03/95	57.587	103.337	55,7	280.581	36,8
31/03/96	51.960	94.730	54,9	256.146	37,0
31/30/97	49.731	90.941	54.7	235.597	38.6
31/03/98	43.708	78.728	55,5	205.893	38,2
31/03/99	36.479	64.204	56,8	168.404	38,1

Font: Institut Nacional d'Ocupació. Oficina d'Estadística Ajuntament de València

3. Contaminación

Diòxid de sofre i fum, per mesos, segons zones. 1998

	Zona Trànsit dens		Zona Trànsit mitjà		Zona Trànsit baix	
	Diòxid de sofre	Fum	Diòxid de sofre	Fum	Diòxid de sofre	Fum
Gener	26	44	23	32	31	19
Febrer	19	58	23	40	15	24
Març	22	49	20	38	26	22
Abril	20	31	19	19	23	13
Maig	22	33	22	20	18	11
Juny	24	28	25	17	27	8
Juliol	28	36	30	19	33	9
Agost	19	38	27	21	14	8
Setembre	24	50	25	24	20	10
Octubre	20	60	20	35	23	19
Novembre	23	55	24	36	17	17
Desembre	21	79	19	51	13	34
Mitjana Anual	22	47	23	29	22	16

Nota: Concentracions en micrograms / m3
Font: Laboratori Municipal. Ajuntament de València

4. Educación

Batxillerat LOGSE. Centres segons tipus per districte. Curs 1998/99.

	Centres	Grups	Alumnes
Concertat	2	4	118
Privat	1	1	28
Públic	6	40	1.160
Total	9	45	1.306

Font: Conselleria de Cultura, Educació i Ciència. Oficina d'Estadística. Ajuntament de València.

Batxillerat LOGSE. Alumnes i règim del centre. Curs 1998/99.

	Grups 1r	Grups 2n	Total Grups	Alumnes 1r	Alumnes 2n	Total Alumnes
Públic	25	15	40	742	418	1.160
Concertat	3	1	4	93	25	118
Privat	1	-	1	28	-	28
Total	29	16	45	863	443	1.306

Font: Conselleria de Cultura, Educació i Ciència. Oficina d'Estadística. Ajuntament de València.