DISTRIBUCIONES BIDIMENSIONALES

Lo que haces ahora de alguna manera está en correlación con el futuro.
Prepárate para ser útil a los demás: estudia, busca la verdad, haz el bien, fomenta la libertad.

1. Distribuciones bidimensionales

Las distribuciones bidimensionales estudian a la vez dos características (dos variables aleatorias) de una población. Ambas variables deben ser cuantitativas (denotadas genéricamente por X e Y, cuyos valores se dan en forma de pares, (x_i, y_i), asimilables a puntos del plano).
Así, por ejemplo:
a) En una población se pueden estudiar a la vez las variables: X = estatura en cm; Y = número de zapato. Algunos de esos pares podrían ser: (180, 43); (165, 40); (195, 45); … Normalmente los individuos altos tienen mayor talla de pie.
b) En el conjunto de los alumnos de un instituto se pueden estudiar las variables: X = tiempo diario dedicado a las "redes sociales"; Y = notas obtenidas en Matemáticas. Algunos de pares podrían ser: (1, 7); (3, 2); (2, 6), (2, 8), … Aunque no sea determinante, si un estudiante dedica mucho tiempo a las redes sociales dedicará menos al estudio y, consecuentemente, sus notas en Matemáticas y en otras asignaturas serán peores.

c) Para los distintos países se pueden estudiar las variables: X = porcentaje del PIB invertido en investigación; Y = porcentaje de población aficionada al ajedrez. Algunos pares de valores pueden ser: (2, 6); (3, 8); (2,3, 5); (1,3, 7), …

2. Correlación lineal

Al estudiar distribuciones bidimensionales, el objetivo es determinar si existe relación estadística entre las dos variables consideradas; es decir, ver si los cambios en una de las variables influyen en los cambios de la otra. Cuando sucede esto, se dice que ambas variables están correlacionadas o que hay correlación entre ellas. En este caso, la variable Y podría estimarse (deducirse) a partir de la X.

Si las variables aumentan o disminuyen conjuntamente, la correlación es directa. Si, por el contrario, al aumentar una de ellas disminuye la otra, la correlación será inversa.
Si la correlación es *fuerte*, a partir de una variable puede estimarse la otra con una fiabilidad (probabilidad) alta. Si la correlación es débil, la estimación de una variable a partir de la otra es poco fiable.
→ Aquí se estudiará solo la correlación lineal, que utiliza la ecuación de una recta para hacer la estimación.

Ejemplos:
a) La correlación entre la "estatura de las personas" y el "número de zapato" es directa y fuerte: a más altura de una persona suele corresponder un mayor número de zapato.

b) Las variables "tiempo dedicado a las redes sociales" y "nota obtenida en Matemáticas" están inversamente correlacionadas: a mayor tiempo dedicado a redes sociales suele corresponder una menor nota en Matemáticas. Para determinar si la correlación es fuerte hay que hacer un estudio detallado, pues es posible que se presenten excepciones significativas.

c) Las variables "porcentaje del PIB invertido en investigación" y "porcentaje de población aficionada al ajedrez" es posible que estén poco correlacionadas; hay países en los que la tradición ajedrecista está muy arraigada por motivos culturales.

Relación funcional y relación estadística

- Dos variables X e Y está relacionadas funcionalmente cuando conocida X se puede saber con exactitud el valor de Y. La relación funcional se cumple siempre: globalmente y para cada valor particular.

Por ejemplo, el tiempo que tarda en impactar contra el suelo un objeto que se deja caer puede saberle exactamente, pues depende de la altura: hay una fórmula física que relación ambas variables. Si un objeto se deja caer desde 4 metros siempre tarda más tiempo en llegar al suelo que si se deja caer desde 3 metros.

- Dos variables X e Y está relacionadas estadísticamente (correlacionadas) cuando conocida X se puede estimar aproximadamente el valor de Y. La relación estadística se cumple en general; para cada valor particular la respuesta puede ser múltiple.

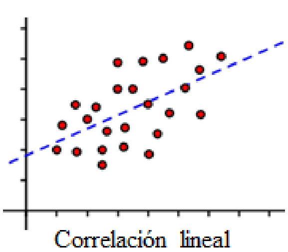
Por ejemplo, la altura de los niños "depende" de su edad: normalmente los niños de 4 años son más altos que los de 3 años; pero no en todos los casos. Globalmente puede admitirse que la relación "más años → más alto" es cierta; pero no siempre se cumple: se tiene una *certeza probable.*

Diagramas de dispersión

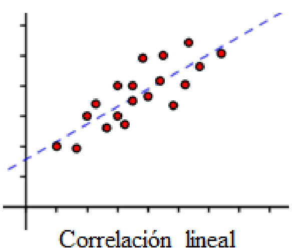
El primer paso para determinar el sentido y el grado de la correlación entre dos variables consiste en representar gráficamente, en el plano cartesiano, los pares de valores conocidos. Estos gráficos, que reciben el nombre de diagramas de dispersión, permiten visualizar la posición de los datos en el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

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

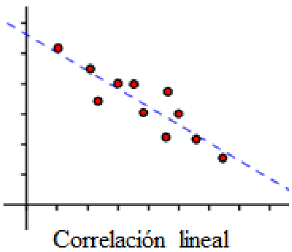
- Una nube de puntos alargada indica correlación lineal: los puntos se distribuyen en torno a una línea recta. La estrechez de la nube expresa que la correlación es fuerte.
- Si la recta que se ajusta a la nube tiene pendiente positiva, la correlación será directa: al crecer la variable X, lo hace también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



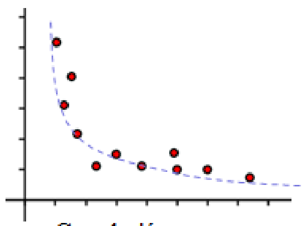
Correlación lineal directa, débil



Correlación lineal directa, fuerte



Correlación lineal inversa, fuerte



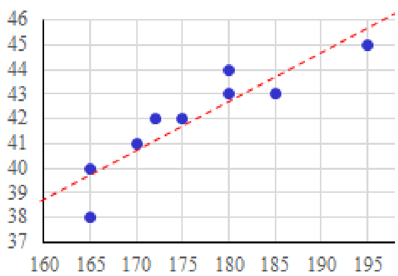
Correlación no lineal

Aquí se estudiará solamente la correlación de tipo lineal, cuando los puntos de la nube se distribuyen de alguna manera en torno a una línea recta.

Ejemplo:

En la tabla siguiente se da la estatura (en cm) y el número de zapato de nueve personas:

Estatura	180	165	195	170	172	165	185	180	175
N.º zapato	43	40	45	41	42	38	43	44	42



El diagrama de dispersión es el adjunto; se ha dibujado con Excel. (Se obtiene automáticamente, más adelante se indicará el proceso. Observa que los ejes no se cortan en el punto (0, 0); así es más fácil de entender).

La correlación entre ambas variables es directa y parece *fuerte.*

La línea recta se ha ajustado de manera aproximada.

3. Coeficiente de correlación lineal

Hasta ahora se ha calificado la correlación como fuerte o débil a partir de la nube de puntos; a partir de su apariencia visual, de un modo intuitivo.

La confirmación cuantitativa (objetiva) de estas conjeturas se obtiene a partir del cálculo de un coeficiente, llamado de correlación, que mide la dependencia estadística entre las variables consideradas.
Este coeficiente se halla a partir de los parámetros (media y desviación típica) de cada una de las variables consideradas, por separado y conjuntamente.

Parámetros de una distribución bidimensional
Cuando estos parámetros se consideran para cada una de las variables se llaman marginales.

- Medias marginales para cada una de las variables X e Y. Valen:
  $\bar{x} = \frac{\sum x_i}{n}$; $\bar{y} = \frac{\sum y_i}{n}$ $\to$ n es el número de observaciones

El punto $(\bar{x}, \bar{y})$ se llama centro medio de la distribución.

- Varianzas y desviaciones típicas
  $$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2; \quad \sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum y_i^2}{n} - \bar{y}^2.$$

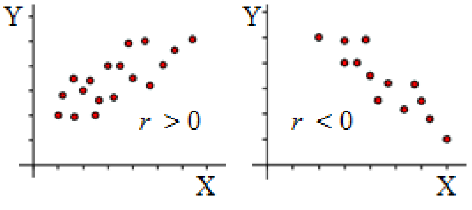
Las desviaciones típicas marginales, σ_x y σ_y, son la raíz cuadrada de las varianzas.

- La covarianza: La covarianza es un parámetro estadístico conjunto, pues, en su cálculo intervienen las dos variables a la vez. Se define como la media aritmética de los productos de las diferencias de los valores de cada variable respecto de su media marginal. Por tanto, vale:
  $$\sigma_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \to \sigma_{xy} = \frac{\sum x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y}.$$

Si $\sigma_{xy} > 0$, la correlación es directa; si $\sigma_{xy} < 0$, la correlación es inversa.

El coeficiente de correlación lineal
Da una medida de la fuerza de la correlación entre las dos variables estudiadas.

Su valor es $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$. Es la razón entre la covarianza de las variables X e Y y el producto de sus desviaciones típicas marginales.
El coeficiente de correlación cumple:
1) El valor de r no cambia al hacerlo la escala de medición.
2) El signo de r es el mismo que el de la covarianza: si $r > 0$, la correlación es directa; si $r < 0$, la correlación es inversa.
3) El valor de r está entre –1 y +1: $-1 \le r \le 1$
4) Si $|r|$ toma valores cercanos a 1, la correlación es fuerte.



5) El cuadrado de r, r^2, indica la proporción de la variación en la variable Y que puede ser explicada por los cambios de la variable X. A r^2 se le llama coeficiente de determinación.

Ejemplo:
Si $r = 0{,}8$, el coeficiente de determinación vale $r^2 = 0{,}8^2 = 0{,}64$. Esto significa que el 64 % de la variación de Y puede ser explicada a partir de la variación de X.

Cálculo del coeficiente de correlación lineal

→ Método manual:

Los cálculos que siguen solo tienen sentido didáctico: son largos y falibles.

Para el caso de las variables estatura y número de zapato pueden organizarse los datos como sigue:

	Estatura (X)	N. zapato (Y)	x_i^2	y_i^2	$x_i \cdot y_i$
	180	43	32400	1849	7740
	165	40	27225	1600	6600
	195	45	38025	2025	8775
	170	41	28900	1681	6970
	172	42	29584	1764	7224
	165	38	27225	1444	6270
	185	43	34225	1849	7955
	180	44	32400	1936	7920
	175	42	30625	1764	7350
Sumas	1587	378	280609	15912	66804

Aplicando las fórmulas (redondeado a centésimas):

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1587}{9} = 176,33...\,;$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{378}{9} = 42\,;$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \Rightarrow$$

$$\sigma_x = \sqrt{\frac{280609}{9} - (176,33...)^2} \approx 9,24$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{15912}{9} - 42^2} = 2\,;\ \sigma_{xy} = \frac{\sum x_i y_i}{n} - \bar{x}\cdot\bar{y} = \frac{66804}{9} - 176,33\cdot 42 \approx 16,81\,.$$

Por tanto, el coeficiente de correlación vale: $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \dfrac{16,81}{9,24\cdot 2} \approx 0,91$.

La correlación entre las variables “estatura” y “número de zapato” es directa y bastante fuerte. Como $r^2 = 0,91^2 = 0,8281$, puede deducirse que el 82,81 % de la variación en el número de zapato de las personas depende de la estatura.

→ Calculadora:

Las medias y desviaciones típicas marginales se pueden calcular una a una, con la calculadora en el MODO SD; o a la vez, poniendo la en el MODO REG Lin (LR *Linear regression*). (Tu calculadora puede ser diferente: lee sus instrucciones). Con la que tengo a mano (Casio *fx–82MS*) hay que proceder como sigue:

(1) Poner la calculadora en el MODO REG Lin.

–Pulsar $\boxed{\text{MODE}}$ $\boxed{3}$ $\boxed{\text{Lin}}$ $\boxed{1}$ (en la pantalla aparecerá REG).

–Borrar los datos de memoria: Pulsar $\boxed{\text{SHIFT}}$ $\boxed{\text{AC}}$.

(2) Introducir los pares de datos (x_i, y_i), tecleando: x_1 $\boxed{,}$ y_1 $\boxed{\text{M+}}$ x_2 $\boxed{,}$ y_2 $\boxed{\text{M+}}$ … x_n $\boxed{,}$ y_n $\boxed{\text{M+}}$

(3) Los parámetros se obtienen pulsando:

$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{1}$ → $\bar{x}$; $\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{2}$ → σ_x (desv. típica de X)

$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{\blacktriangleright}$ $\boxed{1}$ → $\bar{y}$; $\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{\blacktriangleright}$ $\boxed{2}$ → σ_y.

$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{\blacktriangleright}$ $\boxed{\blacktriangleright}$ $\boxed{3}$ → r (coeficiente de correlación lineal).

- Pulsando $\boxed{\text{SHIFT}}$ $\boxed{1}$ se obtienen otros resultados: ($\sum x_i^2$, con 1; $\sum x_i$, con 2; n, con 3).
- Y con $\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{\blacktriangleright}$ $\boxed{1}$ → ($\sum y_i^2$, con 1; $\sum y_i$, con 2; $\sum x_i y_i$, con 3).

Para el ejemplo anterior hay que teclear:

180 $\boxed{,}$ 43 $\boxed{\text{M+}}$ 165 $\boxed{,}$ 40 $\boxed{\text{M+}}$ … 175 $\boxed{,}$ 42 $\boxed{\text{M+}}$

Los resultados que se obtienen deben ser los indicados arriba (con diferencia debidas al redondeo).

→ Usando **GeoGebra** se obtiene el coeficiente de correlación de manera muy sencilla. Tecleando “CoeficienteCorrelación{(180,43), (165,40), (195,45), (170,41), (172,42), (165,38), (185,43), (180,44), (175,42)}” aparece r = 0,9021.

4. Recta de regresión lineal

Cuando la nube de puntos correspondiente a una distribución bidimensional tiene forma alargada, se le puede asociar una "línea de tendencia" recta, llamada recta de regresión (de Y sobre X). La ecuación de esa recta permite hacer estimaciones de la variable Y a partir de la X.

- La recta de regresión es la que mejor se ajusta a la nube de puntos. Se conoce con el nombre de recta de regresión mínimo cuadrática, pues es la que minimiza la suma de los cuadrados de los errores.

Es una recta ideal que asignaría a cada valor x_i de la variable X el promedio de los y_i correspondientes a x_i: (ver problema 11). En consecuencia, debe pasar por el punto $(\bar{x}, \bar{y})$, centro de gravedad de la distribución bidimensional: (ver problemas 13 y 14).
Las estimaciones son fiables cuando el valor de X está cercano a la media $\bar{x}$ y cuando $|r|$ está próximo a 1. (La probabilidad de acierto en las estimaciones se puede medir; aquí no se hará).
Su ecuación es:

$$y-\bar{y}=\frac{\sigma_{xy}}{\sigma_x^2}(x-\bar{x}).$$

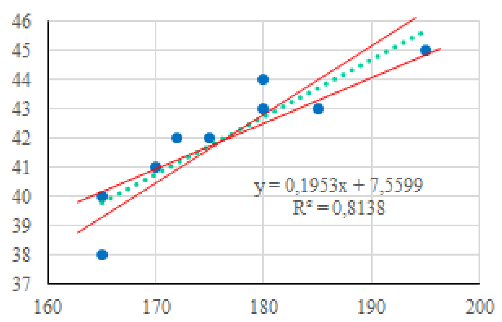
Recuerda que la ecuación general de una recta es $y=mx+n$, En las calculadoras (ordenadores) está recta suele expresarse en la forma y = A + Bx, siendo $B=\frac{\sigma_{xy}}{\sigma_x^2}$ y $A=\bar{y}-\frac{\sigma_{xy}}{\sigma_x^2}\cdot\bar{x}$; valores que también se obtienen de manera automática: teclas $\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{\blacktriangleright}$ $\boxed{\blacktriangleright}$ $\boxed{1 \to A}$; $\boxed{2 \to B}$.

- La recta de regresión de X sobre Y (que no es la misma que la de Y sobre X) permite estimar los valores de Y a partir de los de la variable X. Su ecuación es: $x-\bar{x}=\frac{\sigma_{xy}}{\sigma_y^2}(y-\bar{y})$. (Problema 10).

Ejemplo:
La ecuación de la recta de regresión correspondiente a correlación "estatura/n.º de zapato" estudiada antes es:

$$y-42=\frac{16,81}{9,24^2}(x-176,33) \Rightarrow y=0,1969x+7,263.$$

→ Con la calculadora se obtiene:
A = 7,559895833; B = 0,1953125; r = 0,902109775.



En el gráfico adjunto, cuyos resultados se han obtenido automáticamente con Excel, se observan pequeñas diferencias en los coeficientes de la recta (son debidas al redondeo).
La recta de regresión es la de trazo discontinuo verde; las trazadas en rojo son aproximaciones.

→ La ecuación de la recta permite inferir (deducir estadísticamente) la talla de zapato que usará un individuo de una determinada altura.
Así, por ejemplo, si el individuo mide 190 cm puede esperarse que use el número
$y=0,1969\cdot 190+7,263=44,674$ → talla 45, aprox.
Si mide 162 cm, usará el número $y=0,1969\cdot 162+7,263=39,1608$ → talla 39, aprox.

La fiabilidad de estas inferencias (deducciones) depende del coeficiente de determinación, que en Excel se denota por R^2 y que, en este caso, vale 0,8138; lo que significa que el 81,38 % de la variación en el número de zapato de las personas depende de su estatura. El resto, hasta el 100 % dependerá de otros factores (estatura de la madre, alimentación, …).

5. Interpretación conjunta: coeficiente de correlación, recta de regresión

Como se ha indicado desde el principio del tema, el análisis de regresión puede hacerse:

- Visualmente.

Representando los puntos de la distribución se obtiene el diagrama de dispersión, la nube de puntos: su forma permite deducir si hay correlación lineal y aproximar su intensidad.

- Algebraicamente.

Calculando el coeficiente de correlación lineal *r* (y el de determinación r^2); si el valor absoluto de *r* se acerca a 1, la correlación es fuerte. En este caso, la recta de regresión puede utilizarse para realizar estimaciones del fenómeno estudiado.

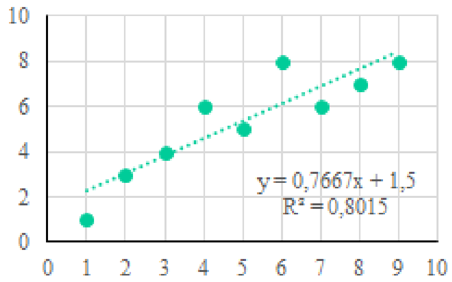
→ A continuación se comentan tres supuestos. (Los cálculos pueden hacerse *a mano* o utilizando herramientas informáticas. Lo importante no es hacer los cálculos, sino saber interpretar los resultados. No obstante, te recomiendo que los hagas; así puedes practicar).

Ejemplos:

a) Datos

X	Y
1	1
2	3
3	4
4	6
5	5
6	8
7	6
8	7
9	8

Nube de puntos



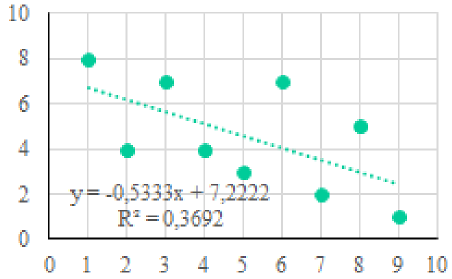
Interpretación

La nube de puntos es estrecha y "creciente", lo que indica una correlación positiva fuerte.
El coeficiente de correlación vale $r=\sqrt{0,8015}\approx 0,895$. La correlación entre ambas variables es muy fuerte.
La recta de regresión asigna valores fiables a Y a partir de la X.
Por ejemplo, a x = 7,5 le asigna, $y = 0,7667\cdot 7,5 + 1,5 \approx 7,25$

b) Datos

X	Y
1	8
2	4
3	7
4	4
5	3
6	7
7	2
8	5
9	1

Nube de puntos



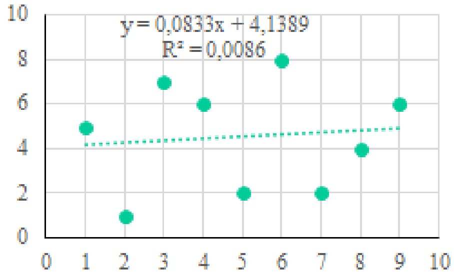
Interpretación

La nube de puntos no es estrecha, aunque sigue una tendencia decreciente. Esto sugiere una correlación negativa y no muy fuerte. Las variaciones de X *solo* explican el 36,92 % de las de Y, pues $r^2 = 0,3692 \to r \approx -0,608$. La recta de regresión asigna valores *poco* fiables a Y a partir de X. (A veces, un *poco* puede ser importante: dependerá de la naturaleza del problema).

c) Datos

X	Y
1	5
2	1
3	7
4	6
5	2
6	8
7	2
8	4
9	6

Nube de puntos



Interpretación

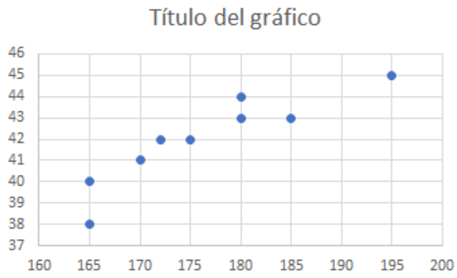
La nube de puntos no sigue ninguna tendencia claramente definida. Se debe admitir que no hay correlación lineal entre las variables X e Y.
El análisis puede darse por terminado: no se puede inferir nada de una variable a partir de la otra.

6. Recursos informáticos: Excel y GeoGebra

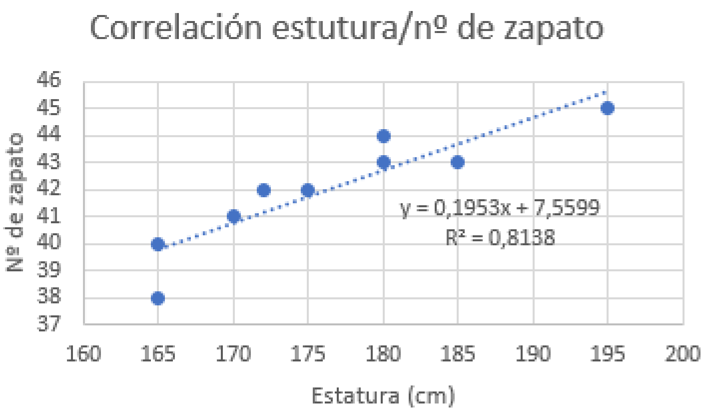
Para estudiar la posible correlación entre estatura y número de zapato, cuyos datos son:

Estatura	180	165	195	170	172	165	185	180	175
N.º zapato	43	40	45	41	42	38	43	44	42

En Excel se puede proceder como sigue:
1) Abrir Excel y teclear datos (pueden copiarse en bloque).
2) Abarcar la tabla de datos e ir a Insertar → Gráficos (elegir Dispersión).
De manera automática aparece el gráfico de la derecha.
Para presentarlo adecuadamente:
3) Cambiar título (pulsar sobre él y escribir lo que sea adecuado: Correlación estatura/nº de zapato).



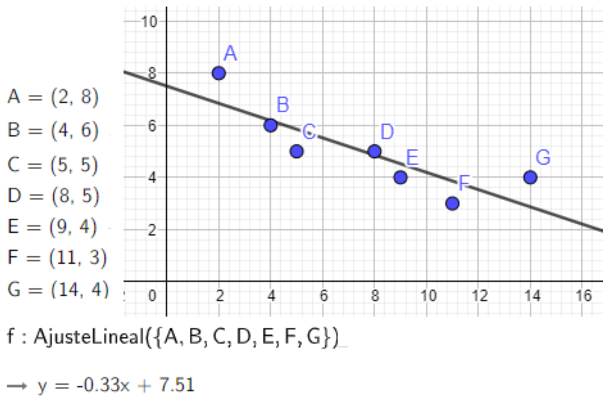
4) Para indicar leyenda en los ejes. Pulsar sobre el gráfico (Aparece un cuadro de opciones. Pinchar en + Elementos del gráfico → Títulos del eje (aquí se ha escrito: Estatura (cm) y Nº de zapato.
5) Para añadir la recta de regresión conviene dar formato al eje vertical. Para ello: (1) Pinchar sobre los datos del eje vertical; (2) botón derecho → Dar formato al eje → Opciones de eje, Límites: Mínimos 37; Máximos 46. Aparecerá una línea de trazos.
6) Pinchando sobre esa la línea, con el botón derecho, elegir Formato de línea de tendencia y en opciones, marcar: Lineal; Presentar ecuación en el gráfico; presentar el valor R cuadrado en el gráfico.
Se obtiene el gráfico adjunto.

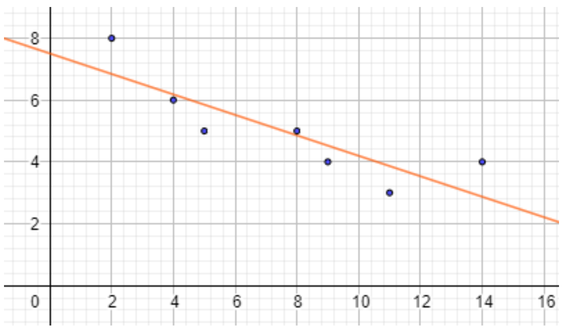


Con **GeoGebra**, para estudiar el conjunto de datos:

Variable X	2	4	5	8	9	11	14
Variable Y	8	6	5	5	4	3	4

1) Se teclean las coordenadas de los puntos.
2) Con el botón derecho del ratón se abarcan todos los puntos; en la barra de herramientas se elige Ajuste lineal (tercera opción por la izquierda): de manera automática se obtiene la recta de regresión con su ecuación. (Gráfico de la izquierda)

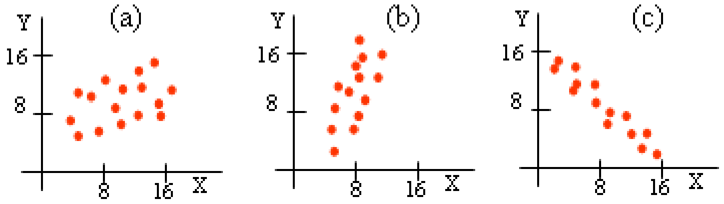




3) El tamaño y aspecto de los puntos y de la línea de tendencia se deciden en "propiedades". Así se obtiene el gráfico de la derecha.
Nota: Tecleando CoeficienteCorrelación({A,B,C,D,E,F,G}) se obtiene –0.85.

PROBLEMAS PROPUESTOS

1. a) Asocia las rectas de regresión: $y = -x + 16$, $y = 2x - 12$ e $y = 0{,}5x + 5$ a las nubes de puntos siguientes:



b) Asigna los coeficientes de correlación lineal $r = 0{,}4$, $r = -0{,}85$ y $r = 0{,}7$, a las mismas nubes de puntos.

2. Se han tomado ocho medidas de la temperatura de una batería y de su voltaje, y se obtuvieron los siguientes datos:

X: temperatura	10,0	10,0	23,1	23,5	34,0	34,5	45,0	45,6
Y: voltaje	430	425	450	460	470	480	495	510

a) Sin efectuar cálculos, razona cuál de las siguientes ecuaciones es la de la recta de regresión de *Y* sobre *X* para los datos anteriores:

$y = 350 - 2{,}1x$; $y = 460 - 2{,}1x$; $y = 406 + 2{,}1x$.

b) Para 25 grados, ¿qué voltaje sería razonable suponer?

3. El número de horas de estudio de una asignatura y la calificación obtenida en el examen correspondiente, fue, para 7 personas, la siguiente:

Horas	5	8	10	12	15	17	18
Calificación	3	6	5	6	9	7	9

a) Dibuja la nube de puntos y traza, aproximadamente, la recta de regresión asociada.
b) Indica el carácter y estima la fuerza de la correlación.

4. Calcula, paso a paso (sin utilizar la calculadora en modo estadístico), el coeficiente de correlación y la recta de regresión asociada a los datos del problema anterior.

5. a) Calcula la recta de regresión de Y sobre X en la distribución siguiente realizando todos los cálculos intermedios.

X	10	7	5	3	0
Y	2	4	6	8	10

b) ¿Cuál es el valor que correspondería según dicha recta a X = 7?

6. El departamento de control de calidad de una empresa de instalación de componentes electrónicos desea determinar la relación entre las semanas de experiencia de sus trabajadores y el número de componentes rechazados a esos trabajadores la semana anterior.

Trabajador examinado	A	B	C	D	E	F	G	H	I	J
Semanas de experiencia (*X*)	7	8	10	1	4	5	15	18	4	8
Número de rechazos (*Y*)	22	35	15	42	26	30	16	20	31	23

a) Representa el diagrama de dispersión asociado a esos datos. ¿Sugiere la gráfica alguna asociación lineal?
b) ¿Cómo calificarías la correlación?

7. Para los datos del problema anterior, halla con ayuda de la calculadora:
a) Las medias y desviaciones típicas marginales.
b) La covarianza.
c) El coeficiente de correlación lineal.
d) La recta de regresión de Y sobre X.
e) El número de rechazos que hay que esperar para una persona con 20 semanas de experiencia.
f) Detalla la solución con Excel.

8. Se midieron los valores de concentración de una sustancia A en suero fetal y los valores de su concentración en suero materno. Se obtuvieron los siguientes datos en una muestra de seis embarazadas a término:

Madre (*X*)	8	4	12	2	7	9
Feto (*Y*)	6	4	8	1	4	5

a) Calcula el coeficiente de correlación lineal.
b) Halla la expresión de la recta que permita estimar los valores fetales a partir de los maternos.

9. En seis alumnas de bachillerato se observaron dos variables: *X* = puntuación obtenida en un determinado test e *Y* = nota en un examen de Matemáticas. Los resultados se indican en la siguiente tabla:

Test: *X*	110	90	140	120	120	100
Examen: *Y*	6	5	9	7	8	6

a) Halla la recta de regresión.
b) Sabiendo que una alumna obtuvo 130 puntos en el test, pero no realizó el examen de Matemáticas, predice, si es posible, la nota que hubiese obtenido.

10. La altura, en cm, de 8 padres y del mayor de sus hijos varones, son:

Padre (X)	170	173	178	167	171	169	184	175
Hijo (Y)	172	177	175	170	178	169	180	187

a) Calcula la recta de regresión que permita estimar la altura de los hijos dependiendo de la del padre; y la del padre conociendo la del hijo.
b) ¿Qué altura cabría esperar para un hijo si su padre mide 174? ¿Y para un padre, si su hijo mide 190 cm?

11. Los años de siete árboles y el diámetro de su tronco, en cm, se dan en la siguiente tabla:

Años	2	4	5	8	10	14	20
Diámetro	10	15	17	20	23	25	27

a) Calcula, utilizando la recta de regresión, el diámetro que se puede predecir para árboles de 10 y 20 años.
b) Compara el resultado anterior con los valores observados en la tabla. Razona el porqué de las diferencias.

12. El número de bacterias por unidad de volumen, presentes en un cultivo después de un cierto número de horas, viene expresado en la siguiente tabla:

X: N.º de horas	0	1	2	3	4	5
Y: N.º de bacterias	12	19	23	34	56	62

Calcula:
a) Las medias y desviaciones típicas de las variables, número de horas y número de bacterias.
b) La covarianza de la variable bidimensional.
c) El coeficiente de correlación, dando una interpretación del resultado.
d) La recta de regresión de Y sobre X.

13. Un conjunto de datos bidimensionales (x, y) tiene un coeficiente de correlación $r = 0,8$. Las medias marginales valen: $\bar{x} = 2$; $\bar{y} = 4$. Indica si alguna de las siguientes ecuaciones puede corresponder a la recta de regresión de Y sobre X:

$y = -2x + 8$; $y = 0,8x + 2$; $y = 1,5x + 1$.

14. Halla el centro medio de una distribución sabiendo que sus rectas de regresión valen:
De Y sobre X: $y = 2x + 2$.
De X sobre Y: $x = 0,45y - 0,2$.

15. Una compañía de seguros de automóvil sospecha que el número de accidentes está en función de la edad del conductor. Para ello elige 100 personas de cada grupo de edad y contabiliza los accidentes totales del último año. Los datos fueron:

Edad	20	25	30	35	40	45
N.º accidentes	10	11	9	7	4	5

a) Representa gráficamente la nube de puntos asociada a estos datos. ¿Qué correlación se observa?
b Halla, sin utilizar la calculadora en el modo REG, el coeficiente de correlación lineal entre las variables medidas. Comenta su valor.

16. Se está experimentado la resistencia a la rotura de una determinad fibra textil. Para ello se ha medido el diámetro de la fibra y el peso que soporta hasta la rotura, obteniéndose los siguientes datos:

Diámetro en mm (X)	1	1,2	1,4	1,6	1,8	2
Peso a la rotura en kg (Y)	12,5	18	25	32	41	52

a) Representa el diagrama de dispersión asociado a esos datos. ¿Sugiere la gráfica alguna asociación lineal?
b) ¿Cómo calificarías la correlación?

17. Con los datos del problema anterior, halla:
a) La recta de regresión de Y sobre X y determina la resistencia a la rotura de una fibra de 2,5 mm de diámetro.
b) La recta de regresión de X sobre Y y determina el diámetro mínimo de una fibra para que soporte más de 60 kg.

18. La siguiente tabla indica las horas de asistencia a un curso de informática y las notas obtenidas por seis alumnos:

Horas de asistencia (X)	2	1	7	5	4	3
Notas (Y)	2	1	10	6	5	3

a) Representa la nube de puntos.
b) Halla el coeficiente de correlación entre X e Y, e interprétalo.
c) Halla la recta de regresión; y represéntala.
d) Si una persona asistiera seis horas al curso, ¿qué nota cabe esperar para ella?

19. El número de faltas de asistencia a clase en los dos últimos meses, en la asignatura de Matemáticas, de ocho estudiantes de 1º de bachillerato, y su calificación final en dicha asignatura han sido:

Faltas (X)	1	2	4	4	6	7	8	12
Calificaciones (Y)	8	9	6	4	3	5	3	3

a) Dibuja la nube de puntos asociada. ¿Qué tipo de correlación se observa entre las variables estudiadas?
b) Calcula, indicando todos los pasos intermedios, el coeficiente de correlación y la recta de regresión de *Y* sobre *X*.