Unidad 15 – Distribuciones bidimensionales. Correlación y regresión

PÁGINA 343

cuestiones iniciales

1. En un hospital se quiere estimar el peso de las niñas recién nacidas. Para ello, se seleccionan, de forma aleatoria, cien de estas, obteniéndose los siguientes resultados:

Intervalo (kg)	[1; 1,5)	[1,5; 2)	[2; 2,5)	[2,5; 3)	[3; 3,5)	[3,5; 4)	[4; 4,5)	[4,5; 5)
Nº de niñas	1	2	5	20	40	26	5	1

Calcula la media, la moda, la mediana, la varianza, la desviación típica y el número de niñas que están en los intervalos $(\bar{x}-\sigma, \bar{x}+\sigma)$; $(\bar{x}-2\sigma, \bar{x}+2\sigma)$; $(\bar{x}-3\sigma, \bar{x}+3\sigma)$. Comenta la simetría de esta distribución.

2. Las ventas de libros de aventuras en una librería tiene como valores $\bar{x} = 10$, $\sigma = 3$. ¿Cómo se clasificaría un día que se vendan 15 libros de aventuras?

3. Diez alumnos han realizado el último mes dos ejercicios de Matemáticas. Las notas son las de la tabla siguiente:

Primer ejercicio	4	7	6	9	4	7	9	4	8	10
Segundo ejercicio	5	8	5	10	3	6	8	4	8	10

Dibuja la nube de puntos. Ajusta a ojo una recta a la nube de puntos, y estima el valor que tendrá la posible correlación.

SOLUCIONES

1. Las soluciones quedan:

La media, $\bar{x} = \dfrac{324,5}{100} = 3,245$; La moda, $Mo = 3,25$; La mediana, $Me = 3,25$

La varianza, $\sigma^2 = \dfrac{1085,25}{100} - (3,245)^2 = 0,3225$; La desviación típica, $\sigma = \sqrt{0,3225} = 0,5679$;

Los intervalos pedidos son:

$(\bar{x}-\sigma, \bar{x}+\sigma) = (2,677; 3,813)$ En el intervalo $(\bar{x}-\sigma, \bar{x}+\sigma)$ hay 86 niñas.

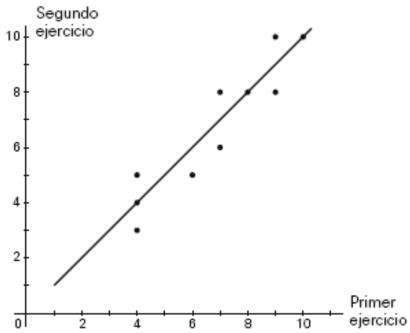
$(\bar{x}-2\sigma, \bar{x}+2\sigma) = (2,109; 4,381)$ En el intervalo $(\bar{x}-2\sigma, \bar{x}+2\sigma)$ hay 96 niñas.

$(\bar{x}-3\sigma, \bar{x}+3\sigma) = (1,5413; 4,949)$ En el intervalo $(\bar{x}-3\sigma, \bar{x}+3\sigma)$ hay 99 niñas.

La distribución es casi simétrica, ligeramente desplazada hacia la derecha.

2. Es un día raro ya que 15 se sitúa en el intervalo $(\overline{x}-2\sigma, \overline{x}+2\sigma)$. La puntuación típica $z=\dfrac{15-10}{3}=1,6667$ se aleja bastante de la media estándar que es cero.

3. La solución queda:

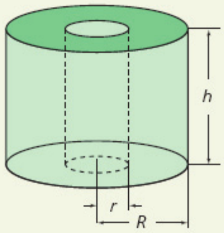


La nube de puntos aparece en la gráfica de la izquierda.
La recta ajustada *a ojo* puede ser bisectriz del cuadrante, $y=x$.
La correlación será positiva y fuerte, próxima a 1.

PÁGINA 357

A C T I V I D A D E S

Usando la simetría o los casos límites, intentar resolver los siguientes problemas:



1. **El bizcocho.** La figura representa la forma de un bizcocho especial. En la etiqueta figura que el volumen es:

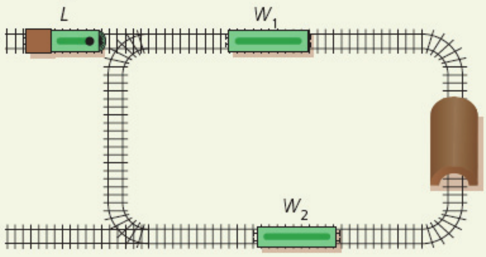
$$V = \pi \cdot h \cdot (R^2 + r^2)$$

¿Es cierto?

2. **Números felices.** El número 44 es un número feliz, pues:

$$44 \Rightarrow 4^2 + 4^2 = 32 \Rightarrow 3^2 + 2^2 = 13 \Rightarrow 1^2 + 3^2 = 10 \Rightarrow 1^2 + 0^2 = 1$$

Investiga sobre los números felices.



3. **Cambio de vagones.** La figura muestra unas vías de tren, en las cuales hay dos vagones W_1 y W_2, una locomotora L, situada en una vía muerta, y un túnel, por el cual sólo puede pasar la locomotora. La locomotora puede enganchar los vagones por delante y por detrás, e incluso puede arrastrar los dos vagones a la vez. El problema consiste en cambiar la posición de los dos vagones, dejando la locomotora en una de las dos vías muertas. Puedes simular este problema utilizando tres monedas diferentes.

4. **Doble frontón.** Un pelotari se encuentra en P y golpea la pelota. Esta debe llegar al pelotari que se encuentra en Q después de pegar en ambos frontones. Construye la trayectoria que debe seguir la pelota.

SOLUCIONES

1. Veamos los dos casos límite:

   1.^{er} $r = 0 \Rightarrow V = \pi \cdot h \cdot R^2 =$ volumen del cilindro.

   2.^{do} $r = R \Rightarrow V = \pi \cdot h \cdot (R^2 + R^2) = 2 \cdot \pi \cdot R^2 \cdot h$, pero si $r = R$ el volumen es cero.

   Luego la fórmula es falsa.

2. La solución queda:

   Números felices de 2 cifras:

$10 \Rightarrow 1^2 + 0^2 = 1$

$13 \Rightarrow 1^2 + 3^2 = 10 \Rightarrow 1^2 + 0^2 = 1$

$23 \Rightarrow 2^2 + 3^2 = 13 \Rightarrow 1^2 + 3^2 = 10 \Rightarrow 1^2 + 0^2 = 1$

$31 \Rightarrow 3^2 + 1^2 = 10 \Rightarrow 1^2 + 0^2 = 1$

$32 \Rightarrow 3^2 + 2^2 = 13 \Rightarrow 1^2 + 3^2 = 10 \Rightarrow 1^2 + 0^2 = 1$

$44 \Rightarrow 4^2 + 4^2 = 32 \Rightarrow 3^2 + 2^2 = 13 \Rightarrow 1^2 + 3^2 = 10 \Rightarrow 1^2 + 0^2 = 1$

Números felices de 3 cifras:

$100 \Rightarrow 1^2+0^2+0^2=1$ $\qquad$ $130 \Rightarrow 1^2+3^2+0^2=10 \Rightarrow 1^2+0^2=1$

$103 \Rightarrow 1^2+0^2+3^2=10 \Rightarrow 1^2+0^2=1$

Igualmente: 310; 301; 230; 203; 320; 302; 440; 404.

Números felices de 4 cifras:

Igual que los que hemos estado formando hasta ahora más otros como 1 339.

Así sucesivamente podemos seguir con los demás.

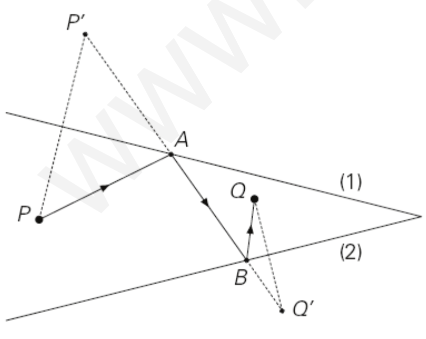
3. Después de varios intentos vemos que la situación final, para lograr el objetivo buscado, que debe quedar en la vía muerta superior es: W_1 W_2 L.

   Llamamos A al lugar en que inicialmente está el vagón W_1 y B al lugar donde está inicialmente el vagón W_2.

   Los pasos a seguir son:

   1.º L coge W_1 y lo lleva a la vía muerta de abajo.
   2.º L da la vuelta al circuito pasando por el túnel y empuja a W_2 hasta el punto A.
   3.º L coge W_1 y lo lleva junto a W_2.
   4.º L da la vuelta al circuito y empuja a ambos vagones a la vía muerta de arriba, quedando la situación que buscábamos, W_1 W_2 L.
   5.º L remolca a W_2 hasta el punto A.
   6.º L da la vuelta al circuito y engancha a W_1 llevándolo a la posición B.
   7.º L vuelve a la vía muerta de arriba y los vagones han cambiado de posición.

4. La solución queda:



Este problema es una doble simetría.

Construimos P', simétrico de P respecto a la banda (1), y Q' simétrico de Q respecto a la banda (2).

Unimos P' y Q' y llamamos A y B a los puntos en que la recta $P'Q'$ corta a las bandas.

La trayectoria pedida es $PABQ$.

PÁGINA 360

ACTIVIDADES FINALES

EJERCICIOS Y PROBLEMAS

1. Una encuesta sobre el gasto que 200 países harán durante el próximo quinquenio para proteger la capa de ozono, han dado los resultados de la tabla siguiente:

Gastos (millones de dólares)	[150-155)	[155-160)	[160-165)	[165-170)	[170-175)	[175-180)	[180-185)	[185-190)	[190-195)
x_i	152,5	157,5	162,5	167,5	172,5	177,5	182,5	187,5	192,5
f_i	7	14	24	37	42	35	23	13	5

Calcula la media, la desviación típica, así como el número de países que se encuentran en los intervalos $(\bar{x} - \sigma, \bar{x} + \sigma)$; $(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$ y $(\bar{x} - 3\sigma, \bar{x} + 3\sigma)$.

2. Se desea comparar la duración de dos marcas de pilas A y B. Para ello, elegimos dos muestras, compuestas por 10 pilas de cada una de las marcas. La duración en horas de cada una de ellas fue:

Marca A	25	28	26	34	30	28	24	27	22	23
Marca B	24	31	26	29	32	31	27	29	24	32

Calcula la media y la desviación típica de las duraciones de cada marca. ¿Qué marca será aconsejable elegir?

3. En las siguientes variables estadísticas bidimensionales, referidas a los alumnos de una clase, estima si hay o no correlación y, en caso de existir, señala si esta es positiva o negativa, fuerte o débil:

a) Estatura y calificación en Lengua

b) Peso y calificación en Educación Física

c) Número de horas diarias de estudio y número de asignaturas aprobadas en la última evaluación

d) Estatura y grado de concentración en el estudio

e) Peso y estatura

4. Se ha pasado una encuesta a los 20 vecinos de una urbanización de las afueras de una gran ciudad obteniéndose los siguientes resultados en los que el primer número se refiere al número de de viajes realizados por los padres y el segundo al número de viajes realizado por sus hijos:

(4, 1) (3, 4) (2, 5) (1, 6) (3, 2) (2, 6) (2, 6) (4, 2) (4, 1) (4, 2)
(1, 7) (1, 6) (4, 1) (1, 7) (2, 4) (2, 6) (3, 3) (4, 2) (1, 6) (2, 5)

a) Construye la tabla de doble entrada correspondiente.

b) Representa gráficamente los datos de esta tabla y, a la vista de la gráfica, estudia si existe correlación entre las variables y el tipo de la misma.

1. La solución queda:

La media, $\overline{x}=\dfrac{34\,425}{200}=172{,}125$; La moda, $Mo=172{,}25$; La mediana, $Me=172{,}25$

La varianza, $\sigma^2=\dfrac{59\,942\,750{,}1}{200}-(172{,}125)^2=86{,}735$; La desviación típica, $\sigma=\sqrt{86{,}735}=9{,}313$;

El número de países en:

$(\overline{x}-\sigma,\overline{x}+\sigma)=(162{,}8;181{,}4)$ es 138.
$(\overline{x}-2\sigma,\overline{x}+2\sigma)=(153{,}5;190{,}7)$ es 188.
$(\overline{x}-3\sigma,\overline{x}+3\sigma)=(144{,}2;200{,}1)$ es 200.

2. La solución es:

Las medias aritméticas son: $\overline{x}_A=\dfrac{267}{10}=26{,}7$ y $\overline{x}_B=\dfrac{285}{10}=28{,}5$.

Las desviaciones típicas son: $\sigma_A=\sqrt{\dfrac{7\,243}{10}-(26{,}7)^2}=3{,}38$ y $\sigma_B=\sqrt{\dfrac{8\,209}{10}-(28{,}5)^2}=2{,}94$.

a) Será aconsejable optar por la marca *B*, ya que tiene mayor media y, a su vez, menor desviación típica.

3. En cada caso queda:

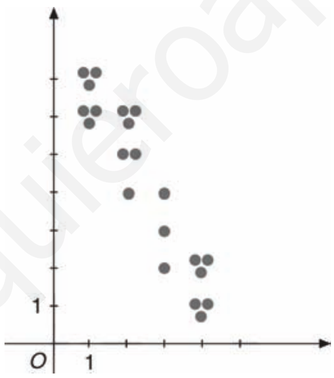
a) No existe correlación.
b) Existe correlación negativa y fuerte.
c) Existe correlación positiva y fuerte.
d) No existe correlación.
e) Existe correlación positiva y fuerte.

4. La solución queda:

a) La tabla de doble entrada es:

y viajes hijos / x viajes padres	1	2	3	4
1	—	—	—	3
2	—	—	1	3
3	—	—	1	—
4	—	1	1	—
5	—	2	—	—
6	3	3	—	—
7	2	—	—	—

b) El diagrama de dispersión es:



5. En una muestra de 100 familias se han estudiado las variables estadísticas X, número de miembros en edad laboral, e Y, número de ellos que se encuentran en activo. Los resultados obtenidos pueden verse en la tabla adjunta.

X \ Y	1	2	3
1	9	0	0
2	14	7	0
3	16	9	5
4	20	12	8

a) Construye la tabla bidimensional simple correspondiente y obtén las distribuciones marginales de X e Y.

b) Calcula la media y la derivación típica de las distribuciones marginales.

6. Se ha solicitado a un grupo de 50 individuos información sobre el número de horas que dedica diariamente a dormir y a ver la televisión. La clasificación de las respuestas ha permitido elaborar la siguiente tabla:

N.º horas dormidas X	6	7	8	9	10
N.º horas televisión Y	4	3	3	2	1
Frecuencias absolutas	3	16	20	10	1

a) Realiza el diagrama de dispersión correspondiente.
b) Calcula la media y la derivación típica de cada una de las variables.
c) Halla el porcentaje de individuos que ven la televisión por encima de la media.
d) Calcula el coeficiente de correlación lineal.

7. Se han hecho dos pruebas de Historia a un grupo de 10 alumnos/as de 3º de ESO. Los resultados obtenidos son:

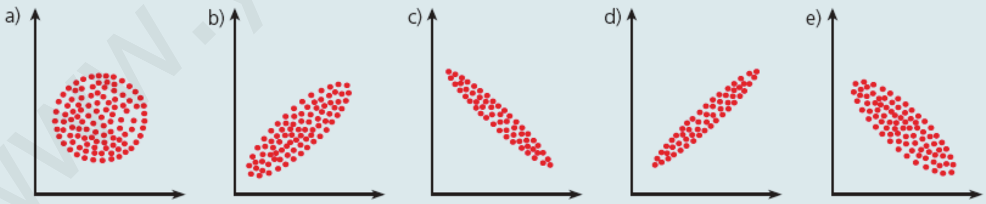
Alumno	1	2	3	4	5	6	7	8	9	10
A	14	12	15	12	13	12	17	7	9	14
B	14	13	17	15	16	12	12	10	14	20

Calcula la covarianza y el coeficiente de correlación. ¿Existe dependencia entre ambas pruebas?

8. En cinco estudios estadísticos se han obtenido los siguientes coeficientes de correlación lineal:

$r = -0{,}98$ $\quad r = 0{,}93$ $\quad r = 0{,}05$ $\quad r = 0{,}71$ $\quad r = -0{,}62$

Identifica, justificando la respuesta, la nube de puntos correspondiente a cada uno de ellos:



9. La estadística de ingresos de determinadas empresas, en miles de euros, y de empleados, en miles, es la siguiente:

Ingresos	5,7	3,8	1,9	1	1
Empleados	16	29	17	6	9

Estudia la correlación existente entre ambas variables y determina la recta de regresión de ingresos en función del número de empleados.

5. La solución queda:

a)

x \ y	1	2	3	
1	9	0	0	9
2	14	7	0	21
3	16	9	5	30
4	20	12	8	40
	59	28	13	

b)

x_i	f_i
1	9
2	21
3	30
4	40

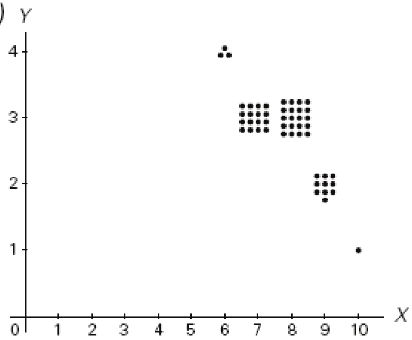
y_i	f_i
1	59
2	28
3	13

c) $\bar{x} = 3,01$ $\sigma_x = 0,98$

$\bar{y} = 1,54$ $\sigma_y = 0,71$

6. La solución queda:

a)



b) Para ambas variables queda:

$\bar{x} = \dfrac{390}{50} = 7,8$ horas dormidas y $\sigma_x = 0,89$

$\bar{y} = \dfrac{141}{50} = 2,82$ horas dormidas y $\sigma_y = 0,71$

c) El porcentaje de individuos por encima de la media es: $\dfrac{20+16+3}{50} = 0,78$, es decir, el 78%.

d) Para el cálculo de $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$, calculamos la covarianza: $\sigma_{xy} = \dfrac{1078}{50} - 7,8 \cdot 2,82 = -0,436$.

Así $r = \dfrac{-0,436}{0,89 \cdot 0,71} = -0,69$. La correlación no es muy fuerte y es negativa.

7. La solución queda:

La covarianza es $\sigma_{AB} = \dfrac{1819}{10} - 12,5 \cdot 14,3 = 3,15$.

El coeficiente de correlación es: $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \dfrac{3,15}{2,73 \cdot 2,72} = 0,424$.

La dependencia estadística es moderada.

8. La correspondencia de cada gráfico con su coeficiente de correlación es:

a) $r = 0,05$
b) $r = 0,71$
c) $r = -0,98$
d) $r = 0,93$
e) $r = -0,62$

9. La solución queda:

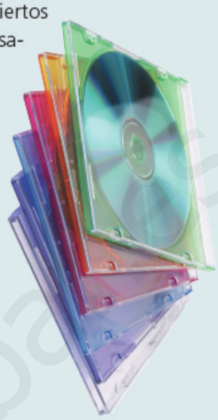
Los parámetros estadísticos son: $\overline{x} = 2,68$; $\overline{y} = 15,4$; $\sigma_x = 1,98$; $\sigma_y = 7,96$; $\sigma_{xy} = 8,47$.

a) La correlación es $r = \dfrac{8,47}{1,98 \cdot 7,96} = 0,54$.

b) La recta de regresión es: $y - 15,4 = \dfrac{8,47}{3,92}(x - 2,68)$.

ACTIVIDADES FINALES

10. Una compañía discográfica ha recopilado la siguiente información sobre el número de conciertos dados durante el verano por 15 grupos musicales y las ventas de discos de estos grupos (expresadas en miles de CDs), obteniendo los datos siguientes:



CD/conciertos	10-30	30-40	40-80
1-5	3	0	0
5-10	1	4	1
10-20	0	1	5

a) Calcula el número medio de CD vendidos por estos grupos.

b) ¿Cómo es el grado de dependencia lineal del número de conciertos dado por el grupo con respecto al número de discos que ha vendido?

c) Obtén la recta de regresión que explica la dependencia anterior.

d) Si un grupo musical ha vendido 18 000 CD, ¿qué número de conciertos es previsible que ofrezca?

11. Se observaron las edades de 5 niños y sus pesos respectivos, obteniéndose los siguientes resultados:

Edad, en años (X)	2	4,5	6	7,2	8
Peso, en kg (Y)	15	19	25	33	34

a) Halla el coeficiente de correlación y las rectas de regresión de Y sobre X y de X sobre Y.

b) ¿Qué peso corresponderá a un niño/a de 5 años? ¿qué edad corresponderá a un peso de 36 kg?

12. La siguiente tabla muestra los valores observados de dos variables X e Y en cinco individuos:

X	1	–1	a	2	3
Y	–2	–3	2	1	0

a) Halla a para que el coeficiente de correlación sea nulo.

b) Suponiendo que $a = 4$, determina la recta de regresión de Y sobre X, y estima el valor de Y cuando X tome el valor –2.

13. De dos variables X e Y, se sabe que la desviación típica de X es $\sqrt{3}$, la media y la desviación típica de Y valen 1 y 2, respectivamente, y la ecuación de la recta de regresión de Y sobre X es $2x + 3y = 6$. Hallar:

a) La media de X

b) La covarianza de X e Y

c) El coeficiente de correlación

d) La recta de regresión de X e Y

14. La estatura media de una muestra de padres es de 1,68 m con una desviación típica de 5 cm. En una muestra de sus hijos, la estatura media es de 1,70 m con una desviación típica de 7,5. El coeficiente de correlación entre las estaturas de padres e hijos es 0,7. Si un padre mide 1,80 m, ¿qué estatura se estima que tendrá su hijo?

10. Llamamos x al número de CDs vendidos e y al número de conciertos. Los datos en una tabla simple son:

x	3	8	8	8	15	15	
y	20	20	35	60	35	60	
f_i	3	1	4	1	1	5	$\ldots \to 15$

Los parámetros estadísticos son: $\bar{x}=9,8$; $\bar{y}=41$; $\sigma_x=4,62$; $\sigma_y=16,55$; $\sigma_{xy}=62,53$.

a) El número medio de CDs vendidos es $\bar{x}=9,8$.

b) El coeficiente de correlación es $r=\dfrac{62,53}{4,62\cdot 16,55}=0,82$. La dependencia lineal es moderada.

c) La recta de regresión es: $y-41=\dfrac{62,53}{21,34}(x-9,8)$.

d) Si $x=18 \Rightarrow y=65,03$ conciertos.

11. La solución queda:

Los parámetros estadísticos son: $\bar{x}=5,54$; $\bar{y}=25,2$; $\sigma_x=2,13$; $\sigma_y=7,49$; $\sigma_{xy}=15,42$.

a) El coeficiente de correlación es $r=\dfrac{15,42}{2,13\cdot 7,49}=0,97$.

b) La rectas de regresión son:

- De Y en X: $y-25,2=\dfrac{15,42}{4,54}(x-5,54)$
- De X en Y: $y-5,54=\dfrac{15,42}{51,12}(x-25,2)$

Un chico de $x=5$ pesa, aproximadamente, $y=27,03$ kg.

La edad que corresponderá a un peso de 36 kg es aproximadamente $x=8,72$ años.

12. La solución queda:

El coeficiente de correlación lineal es nulo si la covarianza es nula.
Por tanto:

$$\frac{3+2a}{5}-(-0,4)\cdot\frac{5+a}{5}=0 \quad \Rightarrow \quad \text{La solución es: } a=-2,083.$$

Los parámetros de las variables son: $\overline{x}=1,8$; $\overline{y}=-0,4$; $\sigma_x=1,72$; $\sigma_y=1,85$; $\sigma_{xy}=2,92$.

La recta de regresión de Y sobre X es: $y+0,4=\frac{2,92}{(1,72)^2}(x-1,8) \Rightarrow y=0,99x-2,18$

Si $x=-2$, el valor estimado de y es: $y=0,99(-2)-2,18=-4,16$

13. La solución queda:

a) La recta de regresión $2x+3y=6$ pasa por el punto $(\overline{x},\overline{y})$, por tanto:

$$2\overline{x}+3\cdot 1=6 \Rightarrow 2\overline{x}=3 \Rightarrow \overline{x}=\frac{3}{2}$$

b) El coeficiente de regresión $\frac{\sigma_{xy}}{\sigma_x^2}$ vale para la recta $2x+3y=6$, por tanto:

$$\frac{\sigma_{xy}}{\sigma_x^2}=-\frac{2}{3};\ \text{al ser } \sigma_x=\sqrt{3},\ \text{obtenemos } \sigma_{xy}=-2$$

c) El coeficiente de correlación es: $r=\frac{\sigma_{xy}}{\sigma_x\cdot\sigma_y}=\frac{-2}{\sqrt{3}\cdot 2}=-0,58$

d) La recta de regresión de X sobre Y es: $x-1,5=\frac{-2}{2^2}(y-1) \Rightarrow x=-0,5y+2$

14. Al ser el coeficiente de correlación $r=0,7$; obtenemos:

$$r=\frac{\sigma_{xy}}{\sigma_x\cdot\sigma_y} \Rightarrow 0,7=\frac{\sigma_{xy}}{5\cdot 7,5} \Rightarrow \sigma_{xy}=26,25$$

La recta de regresión de Y (estatura de los hijos) sobre X (estatura de los padres) es:

$$y-170=\frac{26,25}{5^2}(x-168) \Rightarrow y=1,05x-6,4$$

Si un padre mide 180 cm, se estima que su hijo tendrá: $y=1,05\cdot 180-6,4=182,6$ cm.

NOTA: Todos los datos se han convertido a centímetros.