

Proyecto MaTeX

Distribuciones Bidimensionales

Fco Javier González Ortiz

Directorio

- [Tabla de Contenido](#)
- [Inicio Artículo](#)



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



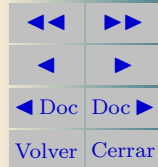
Tabla de Contenido

1. Introducción
2. Diagramas de dispersión
3. Covarianza
4. Coeficiente de correlación
 - Propiedades del coeficiente de correlación
5. Rectas de regresión
 - 5.1. Propiedades de las rectas de regresión
6. Ejercicios
 - Soluciones a los Ejercicios
 - Soluciones a los Tests



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES





1. Introducción

En el capítulo de **Estadística Descriptiva** el alumno estudió las técnicas para resumir información del conjunto de datos para una variable X .

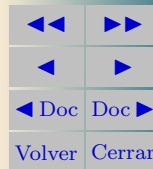
Ahora bien, los datos que tratamos de estudiar pueden incluir valores de varias variables relacionadas entre sí. Por ejemplo:

- en un individuo su altura, su peso y su edad,
- en un gas su presión, su volumen y su temperatura,
- en un vehículo su potencia, su velocidad y su consumo, etc.

Por ello en este capítulo estudiaremos las técnicas para resumir información de la distribución del conjunto de datos de los que se conocen dos variables X e Y , llamadas distribuciones **bidimensionales o bivariadas**.

MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



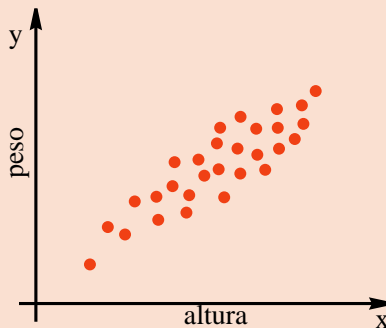
2. Diagramas de dispersión

Al igual que ocurre en el caso unidimensional, también es posible hacer gráficos de distribuciones de frecuencias bidimensionales.

Hay un gráfico bidimensional especialmente útil; éste es el diagrama de dispersión que es simplemente un dibujo cartesiano de la muestra observada.

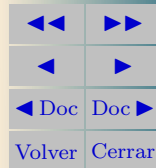
Como ilustración, la figura proporciona el diagrama de dispersión de los datos de la muestra de 33 alumnos donde se han medido su altura X y su peso Y .

A la vista del diagrama de dispersión de la figura parece constatarse que el peso aumenta con la altura.



MaTeX

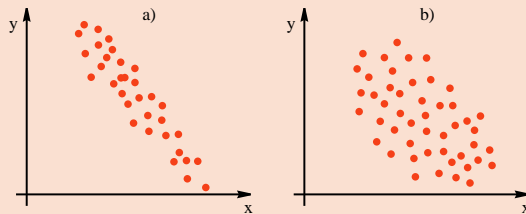
DISTRIBUCIONES
BIDIMENSIONALES



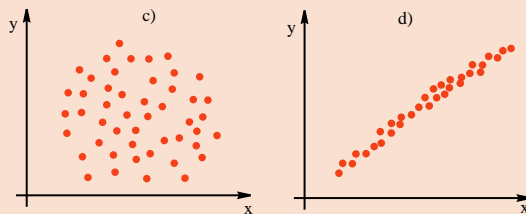
3. Covarianza

En las siguientes figuras se muestran cuatro diagramas de dispersión. En algunos se aprecia que los puntos están más alineados que en otros, es decir, en algunos de ellos hay mayor **grado de asociación lineal**.

En el gráfico a) se aprecia alto grado de asociación lineal, mientras que en b) hay mayor dispersión y muy poco grado de asociación lineal.

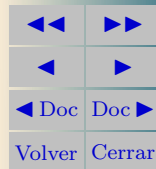


En el c) hay mucha dispersión y nulo grado de asociación lineal, mientras que en d) los puntos casi están alineados.



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



La medida de asociación lineal más simple entre dos variables es la **covarianza**. Viene definida por

$$Cov(X, Y) = S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} \quad (1)$$

y si se opera, se obtiene la expresión más simplificada:

$$S_{xy} = \frac{\sum_{i=1}^n x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y} \quad (2)$$

Veamos un ejemplo de su cálculo para dos variables x e y

x_i	y_i	$x_i y_i$
10	0,5	5,0
30	1,0	30,0
60	3,0	180,0
90	5,0	450,0
120	6,5	780,0
310	16	1445,0

$$\bar{x} = \frac{\sum x_i}{n} = \frac{310}{5} = 62$$

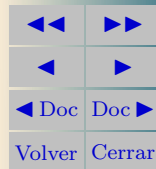
$$\bar{y} = \frac{\sum y_i}{n} = \frac{16}{5} = 3,2$$

$$S_{xy} = \frac{1445}{5} - \bar{x} \cdot \bar{y} = 90,6$$



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



4. Coeficiente de correlación

El inconveniente de la covarianza como medida de la asociación lineal entre dos variables es que depende de las unidades de X e Y , por ello se define el **coeficiente de correlación** entre dos variables r_{xy} , por

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} \quad (3)$$

Se define como el cociente entre la covarianza de X e Y y el producto de las desviaciones típicas.

• Propiedades del coeficiente de correlación

- Este número no tiene dimensiones y su valor está entre

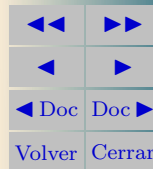
$$-1 \leq r_{xy} \leq 1$$

- Los valores extremos 1 y -1 se alcanzan solamente si todos los datos se sitúan exactamente sobre una recta.
- Si la relación lineal es muy pequeña, el valor de r_{xy} es próximo a cero.



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



Ejemplo 4.1. De un muelle cuelgan pesas, obteniéndose los siguientes alargamientos:

Pesos (g)	10	30	60	90	120
Alargamiento (cm)	0.5	1	3	5	6.5

Calcula e interpreta el coeficiente de correlación entre estas variables.

Solución: Sea x los pesos e y los alargamientos,

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	0,5			
30	1,0			
60	3,0			
90	5,0			
120	6,5			
310	16			

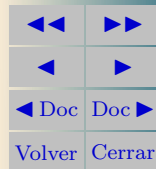
$$\bar{x} = \frac{310}{5} = 62 \quad \bar{y} = \frac{16}{5} = 3,2$$

□



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



Ejemplo 4.1. De un muelle cuelgan pesas, obteniéndose los siguientes alargamientos:

Pesos (g)	10	30	60	90	120
Alargamiento (cm)	0.5	1	3	5	6.5

Calcula e interpreta el coeficiente de correlación entre estas variables.

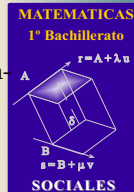
Solución: Sea x los pesos e y los alargamientos,

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	0,5	100		
30	1,0	900		
60	3,0	3600		
90	5,0	8100		
120	6,5	14400		
310	16	27100		

$$\bar{x} = \frac{310}{5} = 62 \quad \bar{y} = \frac{16}{5} = 3,2$$

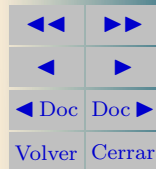
$$S_x^2 = \frac{27100}{5} - \bar{x}^2 \quad S_x = 39,70$$

□



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



Ejemplo 4.1. De un muelle cuelgan pesas, obteniéndose los siguientes alargamientos:

Pesos (g)	10	30	60	90	120
Alargamiento (cm)	0.5	1	3	5	6.5

Calcula e interpreta el coeficiente de correlación entre estas variables.

Solución: Sea x los pesos e y los alargamientos,

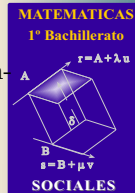
x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	0,5	100	0,25	
30	1,0	900	1,00	
60	3,0	3600	9,00	
90	5,0	8100	25,00	
120	6,5	14400	42,25	
310	16	27100	77,50	

$$\bar{x} = \frac{310}{5} = 62 \quad \bar{y} = \frac{16}{5} = 3,2$$

$$S_x^2 = \frac{27100}{5} - \bar{x}^2 \quad S_x = 39,70$$

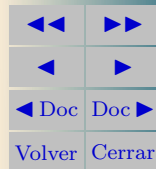
$$S_y^2 = \frac{77,50}{5} - \bar{y}^2 \quad S_y = 2,29$$

□



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



Ejemplo 4.1. De un muelle cuelgan pesas, obteniéndose los siguientes alargamientos:

Pesos (g)	10	30	60	90	120
Alargamiento (cm)	0.5	1	3	5	6.5

Calcula e interpreta el coeficiente de correlación entre estas variables.

Solución: Sea x los pesos y y los alargamientos,

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	0,5	100	0,25	5,0
30	1,0	900	1,00	30,0
60	3,0	3600	9,00	180,0
90	5,0	8100	25,00	450,0
120	6,5	14400	42,25	780,0
310	16	27100	77,50	1445,0

$$\bar{x} = \frac{310}{5} = 62 \quad \bar{y} = \frac{16}{5} = 3,2$$

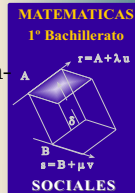
$$S_x^2 = \frac{27100}{5} - \bar{x}^2 \quad S_x = 39,70$$

$$S_y^2 = \frac{77,50}{5} - \bar{y}^2 \quad S_y = 2,29$$

$$S_{xy} = \frac{1445}{5} - \bar{x}\bar{y} \quad S_{xy} = 90,6$$

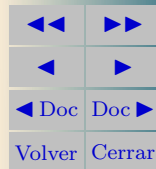
$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{90,6}{39,7 \cdot 2,29} = 0,995$$

El valor de r_{xy} es próximo a uno, luego el grado de asociación lineal entre el peso y el alargamiento es alto. \square



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



5. Rectas de regresión

Los modelos de **regresión** provienen de los trabajos de Galton en biología a finales del siglo XIX.

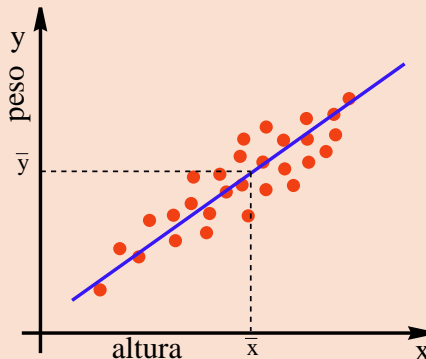
Galton estudió la dependencia de la estatura de los hijos (y) respecto a la de sus padres (x), encontrando lo que denominó “regresión” a la media.

Los padres bajos tienen hijos bajos, pero en promedio más altos que sus padres, y los padres altos tienen hijos altos, pero en promedio más bajos que sus padres.

La recta de regresión corresponde a la recta

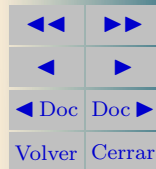
$$\hat{y}_i = a x_i + b$$

que mejor se aproxima a los puntos del diagrama de dispersión para la altura X y el peso Y .



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES





Denominamos **recta de regresión** a la recta

$$\hat{y} = ax + b$$

que “mejor” se aproxima a la nube de puntos del diagrama de dispersión.

Para hallar a y b se impone el criterio de que la suma de las distancias cuadráticas $\sum d_i^2$ sea mínima, donde para cada valor de x_i , d_i es la diferencia entre el valor observado y_i y el valor que predice la recta $\hat{y}_i = ax_i + b$.

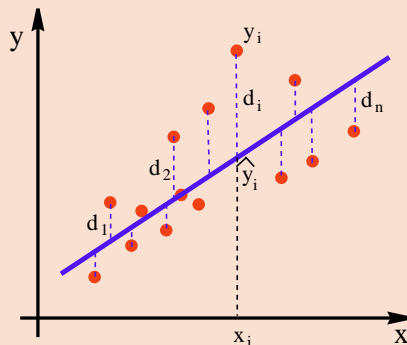
Como tenemos dos variables, según se tome x o y como variable independiente, tenemos dos rectas:

Recta de regresión de y sobre x :

$$r_{y/x} \equiv y - \bar{y} = \frac{S_{xy}}{S_x^2} (x - \bar{x}) \quad (4)$$

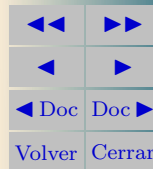
Recta de regresión de x sobre y :

$$r_{x/y} \equiv x - \bar{x} = \frac{S_{xy}}{S_y^2} (y - \bar{y}) \quad (5)$$



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES





5.1. Propiedades de las rectas de regresión

- a) Las dos rectas de regresión se cortan en el punto de las medias de las variables (\bar{x}, \bar{y}) .
- b) El producto de las pendientes de las rectas es el cuadrado del coeficiente de correlación.

$$\frac{S_{xy}}{S_x^2} \cdot \frac{S_{xy}}{S_y^2} = \left(\frac{S_{xy}}{S_x \cdot S_y} \right)^2 = r_{xy}^2$$

- c) Las rectas de regresión se usan para predecir el valor de una variable cuando se conoce la otra, y se debe cumplir que el coeficiente de correlación sea próximo a -1 o a 1.

Test. Responder a las siguientes cuestiones.

1. Si S_{xy} es positiva, entonces el coeficiente de correlación es:

(a) $r_{xy} > 0$ (b) $r_{xy} < 0$ (c) no se puede saber

2. La pendiente de la recta de regresión de y respecto de x es:

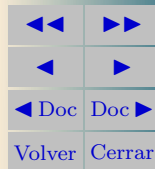
(a) $\frac{S_{xy}}{S_y^2}$ (b) $\frac{S_{xy}}{S_x^2}$ (c) Ninguna de ellas

3. La pendientes de las rectas de regresión pueden ser de signo contrario.

(a) verdadero (b) falso

MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



4. Si las pendientes de las rectas de regresión son iguales.

(a) $r_{xy} = 1$

(b) $S_x = S_y$

(c) Ninguna de ellas

Ejemplo 5.1. En el año 2000 los ingresos de determinadas empresas, en millones de € y de empleados en miles, fue

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

Determinar el coeficiente de correlación y la recta de regresión de los ingresos respecto de los empleados.

Solución: Sea x los empleados e y los ingresos,

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
16	5,7	256	32,49	91,2
29	3,8	841	14,44	110,2
17	1,9	289	3,61	32,3
6	1,0	36	1,00	6,0
9	1,0	81	1,00	9,0
77	13,4	1503	52,54	248,7

$$\bar{x} = 15,4$$

$$\bar{y} = 2,68$$

$$S_x^2 = \frac{1503}{5} - \bar{x}^2 \quad S_x = 7,96$$

$$S_y^2 = \frac{52,54}{5} - \bar{y}^2 \quad S_y = 1,82$$

$$S_{xy} = \frac{248,7}{5} - \bar{x}\bar{y} \quad S_{xy} = 8,47$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = 0,58 \quad y - 2,68 = \frac{8,47}{63,44}(x - 15,4)$$

□



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES





6. Ejercicios

Ejercicio 1. El índice de mortalidad y de una muestra de población que consumía diariamente x cigarrillos aparece en la tabla adjunta, donde se estudiaron siete muestras distintas de población que consumía distinto número de cigarrillos:

Nº de cigarrillos (x)	3	5	6	15	20	40	45
Índice de mortalidad (y)	0.2	0.3	0.3	0.5	0.7	1.4	1.5

Estudiar la correlación. ¿Qué índice de mortalidad se podría predecir para un consumidor de de 32 cigarrillos diarios?

Ejercicio 2. 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

- Calcular el coeficiente de correlación.
- Estimar la tensión de una persona de 60 años.

Ejercicio 3. La tabla siguiente representa una muestra de la que se conocen

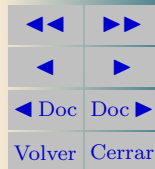
$$\sum_{i=1}^3 x_i \cdot y_i = 42 \quad S_{xy} = 2$$

x	3	4	5
y	a	3	b

Hallar a y b .

MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



Ejercicio 4. Se observaron las edades de 5 niños y sus pesos respectivos, obteniéndose la tabla:

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

- Hallar el coeficiente de correlación y las dos rectas de regresión
- ¿Qué peso corresponderá a un niño de 5 años?
- ¿Qué edad corresponderá a un niño de 22 kg?

Ejercicio 5. Los ingresos en determinadas empresas en miles de euros y de empleados en miles es el siguiente:

Ingresos (x)	34,2	22,8	11,4	6	6
Empleados (y)	16	29	17	6	9

- Estudiar la correlación entre las variables.
- Hallar la recta de regresión de los ingresos, en miles de euros, respecto de los empleados, en miles.

Ejercicio 6. Hallar la recta de regresión y respecto de x a partir de:

$$\sum_{i=1}^5 x_i = 80,4 \quad \sum_{i=1}^5 y_i = 77$$

$$\sum_{i=1}^5 x_i^2 = 1891,44 \quad \sum_{i=1}^5 y_i^2 = 1503 \quad \sum_{i=1}^5 x_i \cdot y_i = 1492,2$$



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



Soluciones a los Ejercicios

Ejercicio 1. Sea x el n° de cigarrillos e y el índice de mortalidad,

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
3	0,2	9	0,04	0,6
5	0,3	25	0,09	1,5
6	0,3	36	0,09	1,8
15	0,5	225	0,25	7,5
20	0,7	400	0,49	14,0
40	1,4	1600	1,96	56,0
45	1,5	2025	2,25	67,5
134	4,9	4320	5,17	148,9

$$\bar{x} = \frac{310}{7} = 19,143 \quad \bar{y} = \frac{16}{7} = 0,700$$

$$S_x^2 = \frac{27100}{7} - \bar{x}^2 \quad S_x = 15,83$$

$$S_y^2 = \frac{77,50}{7} - \bar{y}^2 \quad S_y = 0,50$$

$$S_{xy} = \frac{1445}{7} - \bar{x}\bar{y} \quad S_{xy} = 7,871$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = 0,997$$

Con la recta de regresión de y sobre x : $r_{y/x} \equiv y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$

$$y - 0,7 = \frac{7,871}{15,83^2}(x - 19,143) \quad \hat{y}_{x=32} = 1,10$$

Ejercicio 1



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



Ejercicio 2. Sea x la edad e y la presión arterial,

$$\left. \begin{array}{l} \bar{x} = \frac{460}{10} = 46 \\ S_x^2 = \frac{23420}{10} - \bar{x}^2 \\ S_x = 15,03 \\ S_y^2 = \frac{1974,77}{10} - \bar{y}^2 \\ S_y = 2,13 \\ S_{xy} = \frac{6708,3}{10} - \bar{x} \cdot \bar{y} \\ S_{xy} = 31,89 \end{array} \right\} r_{xy} = \frac{S_{xy}}{S_x S_y} = 0,995$$

Con la recta de regresión de y sobre x : $r_{y/x} \equiv y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$

$$y - 13,89 = \frac{31,89}{15,03^2}(x - 46)$$

Para una edad de $x = 60$ años el modelo lineal predice una presión arterial de:

$$\hat{y}_{x=60} = 15,86$$

Ejercicio 2

MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



Ejercicio 3. Como $\sum_{i=1}^3 x_i \cdot y_i = 42$

$$3 \cdot a + 3 \cdot 4 + 5 \cdot b = 42 \Rightarrow 3 \cdot a + 5 \cdot b = 30$$

Siendo $\bar{x} = 4$ $\bar{y} = \frac{3 + a + b}{3}$, sustituyendo en la expresión de la covarianza

$$S_{xy} = \frac{\sum_{i=1}^3 x_i \cdot y_i}{3} - \bar{x} \cdot \bar{y} = 2$$

$$\frac{42}{3} - 4 \cdot \frac{3 + a + b}{3} = 2 \Rightarrow a + b = 6$$

Resolviendo las dos ecuaciones con a y b :

$$\boxed{a = 0} \quad \boxed{b = 6}$$

Ejercicio 3



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES





Ejercicio 4. Sea x la edad en años e y el peso en kg,

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2,0	15	4,00	225	30,0
4,5	19	20,25	361	85,5
6,0	25	36,00	625	150,0
7,2	33	51,84	1089	237,6
8,0	34	64,00	1156	272,0
27,7	126	176,09	3456	775,1

$$\bar{x} = \frac{27,7}{5} = 5,54 \quad \bar{y} = \frac{126}{5} = 25,2$$

$$S_x^2 = \frac{176,09}{5} - \bar{x}^2 \quad S_x = 2,13$$

$$S_y^2 = \frac{3456}{5} - \bar{y}^2 \quad S_y = 7,49$$

$$S_{xy} = \frac{775,1}{5} - \bar{x}\bar{y} \quad S_{xy} = 15,412$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = 0,967$$

$$y - 25,2 = \frac{15,412}{2,13^2} (x - 5,54)$$

$$x - 5,54 = \frac{15,412}{7,49^2} (y - 25,2)$$

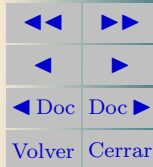
Para un niño de 5 años obtenemos $\hat{y}_{(x=5)} = 23,36$ kg.

Para un niño de 22 kg obtenemos $\hat{x}_{(y=22)} = 4,66$ años.

Ejercicio 4

MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



Ejercicio 5. Sean x los ingresos en miles de euros e y empleados en miles, Realizamos, como en ejercicios anteriores,, el cálculo de los estadísticos:

$$\bar{x} = \frac{80,4}{5} = 16,08 \quad \bar{y} = \frac{77}{5} = 15,4$$

$$S_x^2 = \frac{1891,44}{5} - \bar{x}^2 \quad S_x = 10,94$$

$$S_y^2 = \frac{1503}{5} - \bar{y}^2 \quad S_y = 7,96$$

$$S_{xy} = \frac{1492,2}{5} - \bar{x}\bar{y} \quad S_{xy} = 50,808$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = 0,583 \quad x - 16,08 = \frac{50,808}{7,96^2}(y - 15,4)$$

Siendo el coeficiente de correlación tan bajo no será conveniente utilizar la recta de regresión para realizar predicciones.

Ejercicio 5



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES



Ejercicio 6.

$$\bar{x} = \frac{80,4}{5} = 16,08 \quad \bar{y} = \frac{77}{5} = 15,4$$

$$S_x^2 = \frac{1891,44}{5} - \bar{x}^2 \quad S_x = 10,94$$

$$S_y^2 = \frac{1503}{5} - \bar{y}^2 \quad S_y = 7,96$$

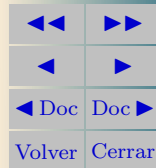
$$S_{xy} = \frac{1492,2}{5} - \bar{x}\bar{y} \quad S_{xy} = 50,808$$

$$y - 15,4 = \frac{50,808}{10,94^2}(x - 16,08)$$

Ejercicio 6



MaTeX

DISTRIBUCIONES
BIDIMENSIONALES

Soluciones a los Tests

Solución al Test: Como

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

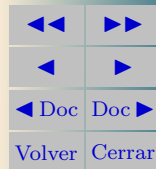
al ser las desviaciones S_x y S_y siempre positivas, el coeficiente de correlación y la covarianza tienen el mismo signo.

Final del Test



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



Solución al Test: Las pendientes de las rectas de regresión son respectivamente,

$$\frac{S_{xy}}{S_x^2} \quad \frac{S_{xy}}{S_y^2}$$

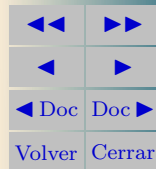
luego, ambas serán positivas o negativas, ya que ambas tienen el mismo signo que la covarianza.

Final del Test



MaT_EX

DISTRIBUCIONES
BIDIMENSIONALES



Solución al Test: Si las pendientes de las rectas de regresión son iguales, entonces

$$\frac{S_{xy}}{S_x^2} = \frac{S_{xy}}{S_y^2} \implies S_x^2 = S_y^2$$

y como las desviaciones típicas son positivas se tiene que son iguales

$$S_x = S_y$$

Final del Test



MaTEX

DISTRIBUCIONES
BIDIMENSIONALES



Índice alfabético

asociación lineal, 5

coeficiente de correlación, 7

covarianza, 5, 6

diagrama de dispersión, 4

regresión, 12

propiedades, 14

rectas de, 13



MaT_EX

DISTRIBUCIONES
BIDIMENSIONALES

