

Variables estadísticas bidimensionales: problemas propuestos

BENITO J. GONZÁLEZ RODRÍGUEZ (bjglez@ull.es)

DOMINGO HERNÁNDEZ ABREU (dhabreu@ull.es)

MATEO M. JIMÉNEZ PAIZ (mjimenez@ull.es)

M. ISABEL MARRERO RODRÍGUEZ (imarrero@ull.es)

ALEJANDRO SANABRIA GARCÍA (asgarcia@ull.es)

Departamento de Análisis Matemático
Universidad de La Laguna

Índice

7. Problemas propuestos

1

ULL

Universidad
de La Laguna



7. Problemas propuestos

1. Un experimento tiene por objeto averiguar la influencia de la presencia del compuesto X en la duración Y de una determinada reacción química. Para ello, se controla la variable X dándole valores predeterminados x_i (en décimas de mol) y se mide en cada caso el tiempo y_i de duración de la reacción (en minutos). Los resultados obtenidos aparecen en la siguiente tabla:

x_i	1	2	3	4	5	6	7
y_i	5.5	6	8.5	11	14	14.5	17.5

Se pide:

- Hallar la ecuación de la recta de regresión de Y sobre X .
- Dibujar la nube de puntos y las rectas de regresión.
- Estimar el tiempo de reacción para $x = 2.5$ y para $x = 8$.
- Hallar el coeficiente de correlación y comentar el resultado.

Solución. a) $r_{y/x} \equiv y - 11 = 2.09(x - 4)$; c) $y(2.5) = 7.9$, $y(8) = 19.4$; (d) $\rho = 0.99$.

2. Las notas obtenidas por 10 alumnos en Matemáticas y Música son:

Matemáticas	6	4	8	5	3.5	7	5	10	5	4
Música	6.5	4.5	7	5	4	8	7	10	6	5

Se pide:

- Calcular el coeficiente de correlación y las rectas de regresión.
- ¿Cuál sería la nota esperada en Música de un alumno que haya obtenido 8.3 en Matemáticas?

Solución. a) $\rho = 0.92$, $r_{y/x} : y - 6.30 = 0.82(x - 5.75)$, $r_{x/y} : x - 5.75 = 1.04(y - 6.30)$; b) $y = 8.4$.

3. Se han tomado cinco muestras de glucógeno, de una cantidad fija cada una. Se les ha aplicado una cantidad X de glucogenasa (en milimoles por litro) anotando en cada caso la velocidad de reacción Y medida en micromoles por minuto, obteniéndose así la siguiente tabla:

X	1	2	3	0.2	0.5
Y	18	35	60	8	10

Se pide:

- a) ¿Se deduce de estos datos que la velocidad de reacción aumenta con la concentración de glucogenasa?
- b) Si a una de las muestras le hubiésemos aplicado una concentración de glucogenasa de 5 milimoles por litro, ¿cuál hubiera sido la velocidad de reacción? ¿Con qué grado de predicción?
- c) Dibujar la nube de puntos y las rectas de regresión.

Solución. a) Sí: $\rho = 0.99 > 0$; b) $y = 94$.

4. Demostar que si la correlación es $\rho = \pm 1$, las dos rectas de regresión coinciden.
5. Las rectas de regresión de dos variables estadísticas X, Y son: $2x + y = 7$, $2x + 3y = 13$. Hallar los coeficientes de regresión y correlación. ¿Qué condición han de verificar dos rectas cualesquiera para ser de regresión?

Solución. $\beta_{y/x} = -0.67$, $\beta_{x/y} = -0.50$, $\rho = -0.58$.

6. Se ha medido el contenido de oxígeno Y en miligramos por litro del lago Worthen, en Austria, a una profundidad de X metros, obteniéndose los siguientes datos:

X	15	20	30	40	50	60	70
Y	6.5	5.6	5.4	6.0	4.6	1.4	0.1

Se pide:

- a) Ajustar una recta a los datos obtenidos.
- b) Estudiar la correlación entre ambas variables.
- c) Para una profundidad comprendida entre 75 y 80 metros, ¿qué contenido en oxígeno se podría predecir?
- d) Dibujar la nube de puntos y las rectas de regresión.

Solución. a) $r_{y/x} : y - 4.23 = -0.11(x - 40.71)$; b) $\rho = -0.90$; c) $0 \leq y \leq 0.5$.

7. Se han estudiado las calificaciones de 100 alumnos en las asignaturas de Matemáticas y Física, obteniéndose los siguientes datos: $\bar{x} = 110$, $\bar{y} = 2.5$, $\sigma_x = 10$, $\sigma_y = 0.5$. Además, se sabe que el coeficiente de correlación entre ambas variables es $\rho = 0.85$. Se quiere averiguar:

- a) ¿Qué nota se puede predecir en la asignatura de Física para un alumno que ha obtenido 125 puntos en Matemáticas?

- b) ¿Se puede decir que aquellos alumnos que obtienen mayor calificación en Matemáticas son los mismos que obtienen mayor calificación en Física?
- c) ¿Cuál es la ecuación de la recta de regresión de X sobre Y ?

Solución. a) $y = 3.1$; b) sí, $\rho > 0$; c) $r_{x/y} : x - 110 = 17(y - 2.5)$.

8. En una exploración biológica sobre un tejido se han observado dos caracteres cuantitativos (X, Y) , obteniéndose los siguientes resultados: $(0, 2)$, $(1, 6)$, $(3, 14)$, $(-1, -2)$, $(2, 10)$.
- a) Estudiar la correlación entre ambos caracteres.
- b) ¿Cómo se completarían los pares $(-3, \cdot)$, $(-2, \cdot)$, $(\cdot, 4)$? Utilizar para ello la recta de regresión ajustada a los datos observados.
- c) Dibujar la nube de puntos y las rectas de regresión.

Solución. a) $\rho = 1$; b) $(-3, 10)$, $(-2, -6)$, $(0.5, 4)$.

9. En un hospital se ha aplicado un medicamento A a 100 enfermos, y en otro hospital se ha aplicado otro medicamento B a otros 100 enfermos. El número de curados cada día durante los diez primeros días es el siguiente:

Medicamento A	8	7	6	5	4	3	3	2	1	1
Medicamento B	4	4	6	7	2	5	1	2	2	2

Se pide:

- a) Rectas de regresión de Y sobre X y de X sobre Y .
- b) Dibujar la nube de puntos y las rectas de regresión.
- c) Hallar el coeficiente de correlación e interpretarlo.

Solución. a) $r_{y/x} : y - 3.5 = 0.5(x - 4)$, $r_{x/y} : x - 4 = 0.7(y - 3.5)$; c) $\rho = 0.6$.

10. La estrona o foliculina es el estrógeno predominante en mujeres postmenopáusicas. Se realiza un estudio para establecer una ecuación mediante la cual se pueda utilizar la concentración de estrona en saliva para predecir la concentración del esteroide en plasma libre (ambas concentraciones medidas en picogramos por mililitro). Se extrajeron los siguientes datos de 14 varones sanos:

$X =$ concentración de estrona en saliva	$Y =$ concentración de estrona en plasma libre
7.4	30.0
7.5	25.0
8.5	31.5
9.0	27.5
9.0	39.5
11.0	38.0
13.0	43.0
14.0	49.0
14.5	55.0
16.0	48.5
17.0	51.0
18.0	64.5
20.0	63.0
23.0	68.0

- a) Dibujar la nube de puntos y las rectas de regresión.
- b) Determinar el coeficiente de correlación.
- c) Predecir el nivel de estrona en plasma libre de un varón cuyo nivel de estrona en saliva es de 17.5 picogramos por mililitro.

Solución. b) $\rho = 0.95$; c) $y = 56.4$.

11. Para estudiar el efecto de las aguas residuales de las alcantarillas que afluyen en un lago se toman medidas de la concentración de nitrato en el agua. La variable se viene monitorizando mediante un antiguo método manual. Se idea un nuevo método automático. Si se pone de manifiesto una alta correlación positiva entre las medidas tomadas empleando los dos métodos, entonces se hará uso habitual del método automático. Los datos obtenidos son los siguientes (las unidades son microgramos de nitrato por litro de agua):

X (manual)	Y (automático)
25	30
40	80
120	150
75	80
150	200
300	350
270	240
400	320
450	470
575	583

- Hallar las ecuaciones de las rectas que mejor aproximan la nube de puntos.
- Dibujar la nube de puntos y las rectas de regresión.
- Estimar la medición para $X = 250$.
- Determinar el coeficiente de correlación. ¿Es idóneo el modelo lineal?

Solución. a) $r_{y/x} : y - 250.3 = 0.93(x - 240.5)$, $r_{x/y} : x - 240.5 = 1.03(y - 250.3)$; c) $Y = 259$; (d) $\rho = 0.98$.

12. La pérdida de actividad de un preparado hormonal en el curso del tiempo viene dada por la siguiente tabla:

tiempo (meses)	1	2	2	3	4	4	5
actividad restante (%)	90	75	73	42	27	30	21

- Determinar la recta de regresión que da el porcentaje de actividad en función del tiempo.
- Dibujar la nube de puntos y las rectas de regresión.
- Calcular el coeficiente de correlación e interpretarlo.
- Estimar cuándo será nula la actividad. ¿Es fiable este resultado?

Solución. a) $r_{y/x} : y - 51.1 = -19.1(x - 3)$; c) $\rho = -0.98$; (d) $x = 6$.

13. La tabla adjunta muestra la natalidad por cada 1000 individuos en España, durante los años 1969-1973:

año	1969	1970	1971	1972	1973
natalidad (‰)	56	59	64	65	73

- Hallar la recta de regresión que da la natalidad con respecto al año.
- Dibujar la nube de puntos y las rectas de regresión.
- Hallar el coeficiente de regresión e interpretarlo.
- ¿Cuántos nacimientos se esperan para 1978?

Solución. a) $r_{y/x} : y - 63.4 = 4(x - 1971)$; c) $\rho = 0.97$; (d) $y = 91$.

14. El número de bacterias por unidad de volumen presentes en un cultivo después de x horas viene representado por la siguiente tabla:

número de horas	0	1	2	3	4	5	2	3
número de bacterias	12	19	23	34	56	62	25	33

- ¿Cuántas horas habrá que esperar hasta tener 100 bacterias?
- ¿Cuál será el número de bacterias al cabo de 6 horas?
- Calcular el coeficiente de correlación e interpretarlo.

Solución. a) $x = 8$; b) $y = 70$; c) $\rho = 0.96$.

15. Se realiza una batería de tests de inteligencia a estudiantes de Historia y en comparación con sus notas se obtienen las siguientes puntuaciones:

CI	96	100	102	110	122	130
notas	3	4	6	8	7	10

- ¿Qué nota se esperaría de un alumno con 115 puntos en el test de inteligencia?
- Estudiar la correlación e interpretarla.
- Dibujar la nube de puntos y las rectas de regresión.

Solución. a) $y = 7$; b) $\rho = 0.89$.

16. Un fabricante de tanques de plástico por el método de moldeo por soplado tuvo dificultades con tanques defectuosos cuyas paredes eran muy delgadas. Se sospecha que la causa de las paredes defectuosas era la variación en la presión del aire de soplado. La siguiente tabla muestra la presión de aire de soplado utilizado (en kilogramos por centímetro cuadrado) y el porcentaje de defectos:

presión (kg/cm ²)	8.6	8.9	9.2	8.2	8.4	9.4	9.4
defectos (%)	88.4	91.4	92.1	86.4	89.4	94.5	93

- a) ¿Podemos decir que la presión incide notablemente en el porcentaje de defectos aparecidos? ¿Qué valor mide esta relación? Calcularlo.
- b) ¿Qué porcentaje aproximado de errores cabe esperar para una presión de 9.0 kilogramos por centímetro cuadrado?
- c) Dibujar la nube de puntos y las rectas de regresión.

Solución. a) $\rho = 0.95$; b) $y = 91.5$.

17. Un estudio para analizar la relación entre las concentraciones de resina y las correspondientes lecturas de colorímetro arrojó los siguientes resultados:

$X =$ lectura del colorímetro	8	8	50	81	102	140	181
$Y =$ concentración de resina	0.17	0.2	0.71	1.09	1.38	1.95	2.5

- a) ¿Qué valor podríamos predecir para $X = 25$? Dibujar la nube de puntos y el modelo de predicción utilizado.
- b) ¿Es fiable el resultado obtenido en el apartado anterior? ¿Qué parámetro mide esta fiabilidad? Calcularlo y comentar su valor.

Solución. a) $Y = 0.39$; b) $\rho = 1$.

18. Se han medido las concentraciones de una cierta sustancia a distintas temperaturas, arrojando los siguientes resultados:

T (°C)	10	15	25	30	35
$c(T)$	2	1.6	0.77	0.5	0.2

- a) Determinar la concentración que podría predecirse para una temperatura de 20 grados.
- b) Determinar y comentar el grado de fiabilidad del método de predicción utilizado en el apartado anterior.

Solución. a) $c(20) = 1.23$; b) $\rho = -1$.

19. En la tabla siguiente se recogen las medidas de la presión sanguínea sistólica y la edad de 10 mujeres:

edad	56	42	70	35	64	47	53	49	38	42
presión	155	152	165	113	165	146	130	150	118	125

A partir de estos datos, dar una predicción de la presión sistólica de una mujer de 50 años. ¿Es fiable este resultado? Explicar la respuesta.

Solución. $y = 142$; $\rho = 0.82$.

20. Para analizar la relación entre la duración (en minutos) del primer resplandor emitido por los machos de la mariposa de la luz y la temperatura (en grados centígrados), se han recogido los siguientes datos:

temperatura (°C)	17.8	20.5	21.1	21.7	22	22.3	23.9	26.7
tiempo (min)	48	44	35	30	40	21	30	18

- a) ¿Cuánto tiempo durará el resplandor cuando la temperatura sea de 21 °C?
 b) ¿Es fiable el resultado anterior? ¿Qué valor mide esta fiabilidad? Determinarlo y comentar el resultado.

Solución. a) $y = 37$; b) $\rho = -0.84$.

21. Se han medido las alturas (en pulgadas) de 12 padres y sus hijos mayores, obteniéndose los siguientes datos:

$X =$ altura padre	65	63	67	64	68	62	70	66	68	67	69	71
$Y =$ altura hijo	68	66	68	65	69	66	68	65	71	67	68	70

- a) ¿Qué altura se puede esperar para un hijo cuyo padre mide 72 pulgadas? ¿Es fiable este resultado? Explicar la respuesta.
 b) ¿Se deduce de estos datos que los hijos con padres más altos tienen también mayor altura?
 c) Dibujar la nube de puntos y la recta de regresión de Y sobre X .

Solución. a) $y = 70$, $\rho = 0.7$; b) sí, $\rho > 0$.

22. Las rectas de regresión de dos variables aleatorias X, Y son: $17y - x + 67.5 = 0$, $y - 0.042x + 107.5 = 0$.

Se pide:

- a) Hallar la media de las variables.
 b) Hallar los coeficientes de regresión.
 c) ¿Es aceptable el modelo lineal? Justificar la respuesta.

Solución. a) $\bar{x} = -6153.85$, $\bar{y} = -365.96$; b) $\beta_{y/x} = 0.042$, $\beta_{x/y} = 17$; c) $\rho = 0.84$.

23. Se ha estudiado la concentración de un cierto producto en el organismo de los bebés de menos de 6 semanas de edad, obteniéndose la siguiente tabla:

semanas	1	2	3	5
cantidad de producto	350	300	270	220

- a) ¿Qué cantidad de producto se espera para un bebé de 3.5 semanas?
- b) ¿Es aceptable el valor obtenido en el apartado anterior? Justificar la respuesta.
- c) Al examinar un bebé resulta que tiene una cantidad de compuesto igual a 240. Estimar su edad.

Solución. a) $y = 261.4$; b) $\rho = -0.99$; c) $x = 4$.