

CUESTIONES Y PROBLEMAS DE DISTRIBUCIONES DE FRECUENCIAS BIDIMENSIONALES PROPUESTOS EN EXÁMENES

1º) Qué ocurre cuando $r = 1$: a) Los valores teóricos no coinciden con los observados; b) La dependencia funcional existente viene dada por una recta decreciente; c) La varianza residual es 0; d) Ninguna de las anteriores [Feb. 2000]

2º) Cuando la covarianza entre dos variables es 0: a) Las variables son estadísticamente independientes; b) El grado de asociación lineal entre las variables es perfecto; e) El coeficiente de variación tendrá valores muy grandes; d) Ninguna de las anteriores. [Feb. 2000]

3º) El coeficiente de determinación lineal indica: a) La varianza total menos la varianza explicada por la regresión; b) La proporción de varianza explicada por la regresión lineal; c) El ajuste de y mediante $y = f(x)$; d) Ninguna de las anteriores. [Feb. 2000]

4º) Dada la regresión lineal simple Y/X , el coeficiente de regresión (b) indica: a) La variación que se produce en Y ante una variación de X en una unidad; b) Los resultados del ajuste; c) El grado de asociación lineal entre X e Y ; d) Ninguna de las anteriores. [Feb. 2000]

5º) Cuando $R^2 = 1$, señálese la respuesta correcta: a) La varianza residual es igual a la varianza de la variable dependiente; b) La suma de la varianza residual y la varianza dependiente también es la unidad; c) La varianza residual es nula; d) Ninguna de las anteriores. [Sep. 2000]

6º) Cuando existe una relación exacta entre las variables, señálese la opción correcta: a) la varianza explicada por la regresión coincide con la varianza de la variable dependiente; b) El coeficiente de correlación es igual a cero; c) La varianza residual es igual a la varianza de la variable dependiente; d) Ninguna de las anteriores. [Sep. 2000]

7º) Si a las variables estadísticas X e Y las sometemos a un cambio de origen y de escala, señálese la respuesta correcta: a) su covarianza queda afectada por el cambio de escala; b) su covarianza queda afectada por ambos cambios; c) su covarianza queda afectada por el cambio de origen; d) su covarianza no queda afectada por ninguno de los cambios. [Sep. 2000]

8º) El coeficiente de correlación lineal entre X e Y es 0,60. ¿Qué porcentaje de variación total queda sin explicar por la regresión? [Feb. 2001]

a) con estos datos no se puede obtener; b) 40 %; c) 64 %; d) ninguna de las anteriores

Aclaración: El coeficiente de determinación que es $0,60^2 = 0,36$ indica que el porcentaje de variación total explicado por la regresión es el 36%. Luego el porcentaje que queda sin explicar por la regresión es el 64%.

9º) A menudo las nubes de puntos de una distribución bidimensional indican un ajuste no lineal. En el caso de que el modelo de partida sea de tipo exponencial, ¿qué cambios habría que hacer para linealizar el modelo? [Feb. 2001]

a) $y' = \log y$; $a' = \log a$; $x' = \log x$; b) $y' = \log y$; $b' = \log b$;
c) $y' = \log y$; $a' = \log a$; $b' = \log b$; d) ninguna de las anteriores

Aclaración: El modelo exponencial ajustado sería del tipo $y = a \cdot b^x$; tomando logaritmos: $\log y = \log a + x \log b$. Haciendo $y' = \log y$; $a' = \log a$; $b' = \log b$, se convierte en un modelo lineal.

10º) El coeficiente de regresión (b) obtenido en la regresión Y/X indica: a) En cuánto varía Y cuando X lo hace en una unidad. b) El valor que alcanza Y cuando X es cero. c) En cuanto varía la variable independiente o endógena ante una variación de la variable dependiente o exógena. d) Ninguna de las anteriores. [Feb. 2002]

11º) De la siguiente distribución correspondiente a la edad de mujeres españolas que realizan algún tipo de estudios. Obténgase: a) La recta de regresión utilizando las formulas mínimo-cuadráticas. b) Obtenga la varianza residual. (Razone los resultados).

Edad	15-24	25-34	35-44	45-54	55-64	
Mujeres (millones)	2,3	2,4	2,1	1,8	1,5	[Feb. 2002]

Solución:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
20	2,3	400	5,29	46
30	2,4	900	5,76	72
40	2,1	1600	4,41	84
50	1,8	2500	3,24	90
60	1,5	3600	2,25	90
200	10,1	9000	20,95	382

$$a_{10} = 40; a_{01} = 2,02; a_{11} = 76,4; a_{20} = 1800; a_{02} = 4,19; m_{20} = 200; m_{02} = 0,1096; m_{11} = -4,4$$

Recta de regresión: $y - 2,02 = \frac{-4,4}{200}(x-40) \leftrightarrow y = -0,022x + 2,9$

Varianza residual: $S_{ry}^2 = 0,0128$

12º) Si la varianza residual es 0

- a) La varianza de la variable dependiente es menor que la varianza explicada por la regresión
- b) La varianza de la variable dependiente es mayor que la varianza explicada por la regresión
- c) La varianza de la variable dependiente es igual a la varianza explicada por la regresión
- d) Ninguna de las anteriores

13)

Una compañía quiere realizar un estudio sobre la influencia del gasto en investigación y desarrollo sobre las ventas netas. Para ello dispone de los siguientes datos sobre los últimos años:

Años	Gasto en I+D (millones)	Ventas Netas (millones)
1998	3,0	130
1999	3,3	155
2000	3,8	175
2001	4,2	210

- a) Obtenga un modelo lineal que permita predecir las ventas a partir de los gastos en I+D. Comente los resultados.
- b) Prediga las ventas del 2002, sabiendo que el gasto en I+D será de 4 millones y medio
- c) Juzgue la bondad del modelo estimado.

[Feb. 2002]

Solución: a) De la tabla:

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
3	130	390	9	16900
3,3	155	511,5	10,89	24025
3,8	175	665	14,44	30625
4,2	210	882	17,64	44100
14,3	670	2448,5	51,97	115650

Obtenemos:

$$\begin{array}{lll} a_{10} = 3,58 & m_{11} = 13,31 & R = 0,988 \\ a_{01} = 167,5 & m_{20} = 0,21 & \\ a_{11} = 612,125 & m_{02} = 856,25 & \\ a_{20} = 12,99 & b = 62,83 & \\ a_{02} = 28912,5 & & \end{array}$$

Y de aquí la recta de regresión Y/X : $y - 167,5 = 62,83(x - 3,58) \leftrightarrow y = 62,83x - 57,12$.

Luego la predicción sería: **Ventas = 62,83·Gastos - 57,12**

b) Si el gasto es de 4,5 millones, las ventas serían de $62,83 \cdot 4,5 - 57,12 = 225,62$ millones.

c) Puesto que el coeficiente de correlación $R = 0,9884$ es muy próximo a 1, el modelo es adecuado para efectuar la predicción.

14º) La técnica de regresión empleada en el proceso de ajuste, permite

a) Describir el grado de evolución conjunta entre sus correspondientes series de datos; b) Evaluar el grado de relación causal entre las variables independientes y dependiente del ajuste e) Imponer alguna restricción a los parámetros del modelo con el fin de evaluar la forma de la distribución d) Ninguna de las anteriores. [Sep 2002]

15º) El coeficiente de regresión b

a) Solo es invariante ante un cambio de origen b) Solo es invariante ante un cambio de escala
c) Es invariante ante un cambio de origen y de escala d) Ninguna de las anteriores. [Sep 2002]

16º) Un determinado parque de atracciones está realizando un estudio para conocer como influyen las inversiones en nuevas atracciones sobre el número total de visitas que recibe anualmente. Los resultados obtenidos para los últimos 4 años son:

INVERSIÓN (millones)	2,1	2,7	3,3	5,1
Nº DE VISITAS (miles)	35	38	43	61

Obtener un modelo lineal que permita predecir las visitas anuales a partir de la inversión realizada, así como juzgar la bondad del modelo estimado. Razone sus respuestas. [Sep 2002]

Solución.- Puesto que hemos de predecir las visitas anuales a partir de la inversión realizada, llamaremos variable X a la inversión y variable Y al nº de visitas. Ajustaremos pues una recta de regresión de Y/X:

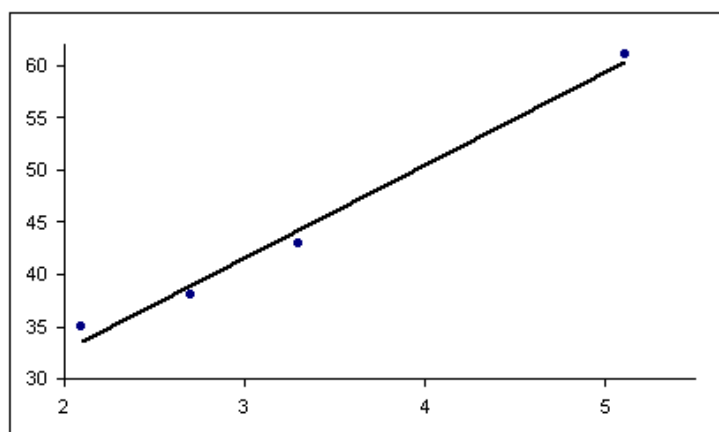
INVERSIÓN N° DE VISITAS					
x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2	
2,1	35	73,5	4,41	1225	
2,7	38	102,6	7,29	1444	
3,3	43	141,9	10,89	1849	
5,1	61	311,1	26,01	3721	
TOTALES	13,2	177	629,1	48,6	8239

$$\left. \begin{array}{l} a_{10} = 3,30 \\ a_{01} = 44,25 \\ a_{11} = 157,28 \\ a_{20} = 12,15 \\ a_{02} = 2059,75 \end{array} \right\} \begin{array}{l} m_{20} = 1,26 \\ m_{02} = 101,69 \\ m_{11} = 11,25 \end{array}$$

$$\text{Recta de regresión de Y/X: } y - 44,25 = 8,93(x - 3,3) \Leftrightarrow y = 8,93x + 14,79$$

Coeficiente de correlación: $R = 0,994$

La recta de regresión hallada es adecuada para hacer estimaciones por que el coeficiente de correlación es prácticamente 1.



17º) Los coeficientes b y b' correspondientes a la regresión de Y sobre X y de X sobre Y respectivamente:

- a) Tienen siempre el mismo signo b) Su signo depende del término independiente a .
c) No tienen siempre el mismo signo d) Ninguna de las anteriores [Feb 2003]

18º) La siguiente distribución corresponde a la edad de los hombres españoles fumadores y el consumo medio de cigarrillos diario. Obténgase a) la recta de regresión. b) Obtenga la varianza residual. (razone los resultados).

Edad hombres fumadores	Cigarrillos diarios
15 a 24	8
25 a 34	13
35 a 44	21
45 a 54	16

[Feb 2003]

Solución.- Consideramos que el nº de cigarrillos es función de la edad (y no al revés). Por eso pondremos x_i (variable independiente) a la edad e y_i (variable dependiente) al nº de cigarrillos.

Intervalos	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
[15, 25[20	8	400	64	160
[25, 35[30	13	900	169	390
[35, 45[40	21	1600	441	840
[45, 55[50	16	2500	256	800
	140	58	5400	930	2190

De la tabla se obtiene:

$$\begin{aligned} a_{10} &= 35 & m_{11} &= 40 \\ a_{01} &= 14,5 & m_{20} &= 125 \\ a_{20} &= 1350 & m_{02} &= 22,25 \\ a_{02} &= 232,5 \\ a_{11} &= 547,5 \end{aligned}$$

Entonces la recta de regresión es: $y - 14,5 = \frac{40}{125}(x - 35) \Leftrightarrow y = 0,32x + 3,3$.

La varianza residual: $S^2_{ry} = m_{02} - m_{11} \cdot b = 22,25 - 40 \cdot 0,32 = 9,45$

19º)

La siguiente distribución corresponde a la edad de los hombres españoles fumadores y el consumo medio de cigarrillos diario. Obténgase a) la recta de regresión. b) Obtenga la varianza residual. (razone los resultados).

Edad hombres fumadores	Cigarrillos diarios
15 a 24	8
25 a 34	13
35 a 44	21
45 a 54	16

[Feb 2003]

Solución.-

Consideramos que el nº de cigarrillos x_i (variable independiente) es función de la edad y_i (variable dependiente)

Intervalos	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
[15, 25[20	8	400	64	160
[25, 35[30	13	900	169	390
[35, 45[40	21	1600	441	840
[45, 55[50	16	2500	256	800
	140	58	5400	930	2190

De la tabla se obtiene:

$$\begin{array}{ll} a_{10} = 35 & m_{11} = 40 \\ a_{01} = 14,5 & m_{20} = 125 \\ a_{20} = 1350 & m_{02} = 22,25 \\ a_{02} = 232,5 & \\ a_{11} = 547,5 & \end{array}$$

Entonces la recta de regresión es: $y - 14,5 = \frac{40}{125}(x - 35) \leftrightarrow y = 0,32x + 3,3$.

La varianza residual: $S^2_{ry} = m_{02} - m_{11} \cdot b = 22,25 - 40 \cdot 0,32 = 9,45$

20º) Siempre que la covarianza entre dos variables X, Y es nula

- a) las variables son estadísticamente independientes.
- b) No se puede realizar afirmación sobre su independencia.
- c) las varianzas son nulas.
- d) Ninguna de las respuestas es correcta. [Sep 2003]

21º) En una regresión lineal simple el coeficiente b:

- a) Es la derivada de y con respecto a x.
- b) Nos determina en cuanto varia la variable x al variar en una unidad la variable y.
- c) Se calcula como $\frac{m_{02}}{m_{11}}$
- d) Ninguna de las respuestas es correcta. [Sep 2003]

22º) Cuando el valor de $r = 1$ (r = coeficiente de correlación lineal)

- a) la varianza residual también es igual a 1
- b) La dependencia funcional existente entre las variables viene dada por una recta creciente
- c) la varianza explicada por la regresión es igual a 1
- d) Ninguna de las respuestas es correcta [Sep 2003]

23º) Los coeficientes b y b' correspondientes a la regresión de Y sobre X y de X sobre Y respectivamente:

- a) El signo depende del termino independiente a.
- b) Tienen siempre el signo contrario
- c) Ninguna de las respuestas es correcta.
- d) Tienen siempre el mismo signo. [Feb 2004]

24º) El coeficiente de determinación se expresa como el cociente entre:

- a) Varianza explicada y la varianza de la variable x.
- b) Varianza explicada y la varianza de la variable y.
- c) Varianza de los residuos y la varianza de la variable x
- d) Varianza de la variable x y la varianza de los residuos [Feb 2004]

25º) Las varianzas residuales para dos problemas de regresión son 400 y 700. Las varianzas de la variable dependiente, para los mismos son 5000 y 12000. A partir de estos datos se opta por la segunda ecuación.

- a) La opción tomada es incorrecta dado que el coeficiente de determinación es menor en la segunda
- b) La opción tomada es correcta dado que la varianza residual es mayor en la segunda.
- c) Ninguna de las respuestas es correcta.
- d) La opción tomada es correcta dado que el coeficiente de determinación es mayor en la segunda. [Feb 2004]

26º) En la siguiente tabla se presentan las preferencias de vivienda por edad de una determinada muestra, entre los habitantes de un determinado municipio:

	Chalet Individual	Chalet Adosado	Piso
20-30	3	5	12
30-40	5	4	9
40-50	8	6	11
50-60	9	5	12
60-70	16	7	13

a) Halle la tabla de correlación de frecuencias relativas. b) calcule las distribuciones de frecuencias marginales y la media de la variable edad. c) Obtenga la distribución de Y (Preferencia en la vivienda) condicionado a que X (edad) sea 55 años. [Feb 2004]

Solución.-

a) Dividimos cada frecuencia absoluta por el total de individuos (125):

	Chalet Individual	Chalet Adosado	Piso	f_i
20-30	0,024	0,04	0,096	0,16
30-40	0,04	0,032	0,072	0,144
40-50	0,064	0,048	0,088	0,2
50-60	0,072	0,04	0,096	0,208
60-70	0,128	0,056	0,104	0,288
f_j	0'328	0'216	0'456	1

b)

Distribución de frecuencias marginales de la X (edad)				
	x_i	n_i	f_i	$x_i n_i$
20-30	25	20	0'16	500
30-40	35	18	0'144	630
40-50	45	25	0'2	1125
50-60	55	26	0'208	1430
60-70	65	36	0'288	2340
Total		125		6025

Distribución de frecuencias marginales de la Y (preferencia en la vivienda)		
	n_j	f_j
Chalet individual	41	0,328
Chalet adosado	27	0,216
Piso	57	0,456

$$\text{Media de } X = \frac{6025}{125} = 48'2$$

c) Puesto que 55 es la marca de clase del intervalo 50-60, lo tomaremos como representante de dicha clase y, en ese caso, la distribución de Y/X=55 sería:

	n_{4j}	$f_{j/4} = \frac{n_{4j}}{n_4}$
Chalet individual	9	0,346
Chalet adosado	5	0,192
Piso	12	0,462
Total	$n_4 = 26$	

27º) Hallar: a) La recta de regresión. b) los residuos. c) el coeficiente de determinación y d) la varianza explicada por la regresión., a partir de los siguientes datos:

Gastos en publicidad (Miles de €)	Ingresos de la empresa (Miles de €)
10	500
20	620
30	780
40	950

Razone los resultados obtenidos en todos los apartados.

[Feb 2004]

Solución.-

De la tabla:

Gastos x_i	Ingresos y_i	x_i^2	y_i^2	$x_i y_i$
10	500	100	250000	5000
20	620	400	384400	12400
30	780	900	608400	23400
40	950	1600	902500	38000
100	2850	3000	2145300	78800

obtenemos:

$$\begin{aligned}
 a_{10} &= 25 \\
 a_{01} &= 712,5 \\
 a_{11} &= 19700 & m_{11} &= 1887,5 \\
 a_{20} &= 750 & m_{20} &= 125 \\
 a_{02} &= 536325 & m_{02} &= 28668,75
 \end{aligned}$$

de donde:

- a) la recta Y/X : $y - 712,5 = 15,1(x - 25) \leftrightarrow y = 335 + 15,1x$;
b)

Residuos:

$$e_i = y_i - a - bx_i$$

14

-17

-8

11

c) el coeficiente de determinación $R^2 = \frac{m_{11}^2}{m_{20} m_{02}} \cong 0,994$;

d) la varianza explicada por la regresión $= \frac{m_{11}^2}{m_{20}} = 28501,25$

28º) Una determinada población del norte de España quiere analizar si el volumen de lluvia influye en el numero de visitantes que reciben en cada verano. Para ello disponen de la siguiente información, correspondiente a los últimos 4 años:

X	120	131	135	144
Y	25	22	21	17

siendo X el volumen de lluvia por metro cuadrado e Y el número de visitantes (expresado en miles). Calcular la recta de regresión y juzgue la bondad del ajuste. Justifique los resultados obtenidos.

[Sep 2004]

Solución.-

Efectuamos los cálculos:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
120	25	14400	625	3000
131	22	17161	484	2882
135	21	18225	441	2835
144	17	20736	289	2448
530	85	70522	1839	11165

de donde:

$$a_{10} = 132,5$$

$$a_{01} = 21,25$$

$$m_{20} = 74,25$$

$$a_{20} = 17630,5$$

$$m_{02} = 8,1875$$

$$b = \frac{m_{11}}{m_{20}} \cong -0,33$$

$$a_{02} = 459,75$$

$$m_{11} = -24,375$$

$$a_{11} = 2791,25$$

y la recta de regresión Y/X:

$$y - 21,25 = -0,33(x - 132,5) \quad \Leftrightarrow \quad y = -0,33x + 64,75$$

El coeficiente de correlación $R = \frac{m_{11}}{\sqrt{m_{20} \cdot m_{02}}} \cong -0,9886$, por lo tanto la dependencia es fuerte y la recta

de regresión puede usarse para hacer predicciones.

29º) A partir de la información de la siguiente tabla de correlación, indique la respuesta correcta:

	1	2	3	
2	1	4	1	6
3	2	4	2	8
4	1	2	1	4
	4	10	4	18

a) X e Y son independientes y $S_{xy} = 0$; b) X e Y son dependientes y $S_{xy} = 0$

c) X e Y son dependientes y $S_{xy} \neq 0$ d) Ninguna de las respuestas es correcta.

(Explicación: X e Y son dependientes porque, por ejemplo, $\frac{n_{11}}{N} = \frac{1}{18} \neq \frac{n_{1\cdot}}{N} \cdot \frac{n_{\cdot 1}}{N} = \frac{4}{18} \cdot \frac{6}{18}$. Además calculando S_{xy} se

$$\text{obtiene } S_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = \frac{104}{18} - \frac{52}{18} \cdot 2 = 0)$$

30º)

Calcular la covarianza de la siguiente distribución:

x	2	4	6	9
y	1	2	5	7

a) $COV(X, Y) = 0$

b) $COV(X, Y) = 25,75$

c) $COV(X, Y) = 6,06$

d) Ninguna de las respuestas es correcta

[Sep 2005]