

Estadística I

Tema 2: Análisis de datos bivariantes

Grado en Administración de Empresas 08/09

1. En la siguiente tabla se representa la distribución conjunta de frecuencias (relativas) de 2 variables: “calificación en Estadística I”, y “número de horas semanales dedicadas al estudio de la asignatura”.

# de h: \ nota:	Suspenso	Aprobado	Notable	Sobresaliente
2	0.20	0.15	0.08	0.03
3	0.12	0.07	0.02	0.02
4	0.04	0.10	0.02	0.00
5	0.00	0.05	0.05	0.05

- a) Obtén las dos distribuciones de frecuencias marginales (relativas).
- b) Obtén todas las distribuciones condicionadas.
2. Tenemos la siguiente distribución conjunta de frecuencias (relativas) correspondiente a observaciones bivariantes (X, Y) , donde X = número de hijos, e Y = renta mensual (en euros).

# de hijos \ renta:	0-1000	1000-2000	2000-3000	> 3000
0	0.15	0.05	0.03	0.02
1	0.10	0.20	0.10	0.05
2	0.05	0.10	0.05	0.03
≥ 3	0.02	0.03	0.02	0.00

- a) Obtén las distribuciones de frecuencias marginales.
- b) Obtén la distribución condicionada de $Y|X = 2$.
- c) Obtén la distribución condicionada de $X|1000 < Y < 2000$.
3. La siguiente tabla muestra la distribución conjunta de frecuencias relativas de la variable X , que representa el número de tarjetas de crédito que posee una persona, y la variable Y , que refleja el número de compras semanales pagadas con tarjeta de crédito.

		Núm. compras por semana				
		0	1	2	3	4
Núm. tarjetas	1	0.08	0.13	0.09	0.06	0.03
	2	0.03	0.08	0.08	0.09	0.07
	3	0.01	0.03	0.06	0.08	0.08

- a) Si se sabe que en el estudio han participado 300 personas, hallar la distribución conjunta de frecuencias absolutas.
- b) Hallar la distribución marginal de Y . ¿Cuál es el número medio y la desviación típica del número de compras semanales pagadas con tarjeta de crédito?
- c) Obtener la distribución del número de tarjetas de crédito que poseen las personas de dicho estudio. ¿Cuál es el número más frecuente de tarjetas de crédito que posee una de estas personas?
- d) Calcular la distribución del número de compras semanales pagadas con tarjetas de crédito que realizan las personas que poseen tres tarjetas. ¿Cuál es la media de esta distribución?

4. Se ha realizado una encuesta a 80 hombres casados en la que se les ha preguntado por el número de hermanos (X) y el número de hijos (Y) que han tenido. Los resultados son los que aparecen a continuación:

x_i	y_i	$fr(x_i, y_i)$	x_i	y_i	$fr(x_i, y_i)$	x_i	y_i	$fr(x_i, y_i)$
0	1	4	2	3	3	4	0	2
1	1	3	3	1	6	4	1	7
1	3	4	3	2	12	4	2	15
2	0	2	3	3	5	4	4	1
2	2	9	3	4	2	5	2	5

- a) Construye una tabla de doble entrada y representa la distribución mediante un diagrama de dispersión.
- b) Según el diagrama anterior, ¿qué valor crees que tendrá la covarianza entre X e Y ? ¿Y el coeficiente de correlación lineal? Justifica tu respuesta.
- c) Calcula el coeficiente de correlación lineal.
5. Los siguientes datos representan el número de años de experiencia (x) y los beneficios anuales (y , en millones de pesetas) de varias compañías:

años experiencia	5	15	24	16	19	3	6	12	27	13
beneficio	4	4	9	7	6	2	3	3	7	5

- a) Representa el diagrama de dispersión para estos datos.
- b) Sobre dicho gráfico, representa las siguientes 4 rectas:
- $y = 1.66 - 0.24x$
 - $y = -1.66 + 0.24x$
 - $y = 1.66 + 0.24x$
 - $y = -1.66 - 0.24x$
- c) ¿Cuál de las rectas anteriores parece ser la recta de regresión?
- d) Obtén el coeficiente de correlación, $r_{(x,y)}$
- e) La relación (lineal) es, ¿positiva o negativa?, ¿débil o fuerte?
- f) Obtén la otra recta de regresión, para los años en función de los beneficios e interprétala.
6. Basándonos en los siguientes datos, queremos analizar la relación entre la *distancia al destino* (x) y la *carga* (y):

distancia	22.4	36.8	14.4	27.2	16.0	35.2	8.0	19.2	9.6	25.6
carga	6.8	10.5	4.0	7.9	8.1	9.5	3.1	7.2	4.5	9.3

- a) Representa el diagrama de dispersión para estos datos.
- b) Estima por mínimos cuadrados la recta de regresión para la carga en función de la distancia y represéntala sobre el gráfico del apartado a).
7. Los siguientes datos muestran el número de pasajeros (por hora) que llegan al aeropuerto (x) y su tiempo de espera (y , en minutos):

# pasajeros	105	511	401	622	330	211	332	332
tiempo espera	44	214	193	299	143	112	155	131
# pasajeros	435	275	55	128	97	187	266	
tiempo espera	208	138	34	73	52	103	110	

- Obtén la recta de regresión para el tiempo de espera en función del número de pasajeros que llegan.
- Obtén el coeficiente de correlación, $r_{(x,y)}$.
- ¿Qué porcentaje de la variabilidad de los tiempos de espera queda **explicada** por su dependencia lineal del número de pasajeros que llegan? esto es, obtén el coeficiente de determinación R^2 .

8. Tenemos los siguientes datos, que representan el número de items defectuosos (x) y el tiempo de inspección (y , en minutos):

# items defectuosos	17	9	12	7	8	10	14	18	19	6
tiempo inspección	48	50	43	36	45	49	55	63	55	36

- Representa el diagrama de dispersión para estos datos.
- Obtén la recta de regresión para el tiempo de inspección en función del número de items defectuosos y represéntala sobre el gráfico del apartado a).
- Obtén los residuos, $r_i = y_i - \hat{y}_i$
- Representa el gráfico de los residuos r_i frente a los valores predichos \hat{y}_i , y utilízalo para decidir si la recta de regresión se ajusta correctamente a los datos.
- ¿Cuál es el valor del coeficiente de determinación R^2 ?

9. Se han recogido los siguientes datos en una cierta ciudad, que representan el tamaño de la unidad familiar (x) y el número de paquetes de detergente consumidos por semana (y):

tamaño familia	5	8	7	3	2	4	5	5	6	7
# paquetes	2	3	4	2	1	2	3	2	4	5

Determina el coeficiente de correlación $r_{(x,y)}$ entre x e y . La relación (lineal) es, ¿positiva o negativa?, ¿débil o fuerte?

10. Los siguientes datos muestran el número de inventarios por año y el porcentaje de ventas de 5 compañías:

Compañía	# inventarios	% ventas
A	3	10
B	4	8
C	5	12
D	6	15
E	7	13

Determina el coeficiente de correlación $r_{(x,y)}$ entre el número de inventarios y el porcentaje de ventas. La relación (lineal) es, ¿positiva o negativa?, ¿débil o fuerte?

11. En la siguiente tabla se recogen los precios de la gasolina (por litro y por año, en promedio, en céntimos de dólar) y el precio del crudo (por barril y por año, en promedio, en dólares), durante el periodo comprendido entre los años 1975-1989:

año	1975	1976	1977	1978	1979	1980	1981	1982
precio gas.	57	59	62	63	86	119	131	116
precio crudo	10.38	10.89	11.96	12.46	17.72	28.07	35.24	28.99
año	1983	1984	1985	1986	1987	1988	1989	
precio gas.	113	112	86	90	90	100	115	
precio crudo	28.63	26.75	14.55	17.90	14.67	17.67	22.23	

- Obtén, para estos datos, la recta de regresión para el precio de la gasolina en función del precio del crudo.
 - Representa el diagrama de dispersión para estos datos y la recta ajustada en el apartado anterior.
 - Si el precio del crudo cae a los 15\$, ¿cuál es el precio estimado de la gasolina?
 - ¿Tiene sentido hacerse la pregunta anterior para un precio del crudo de 0?
 - ¿Puedes emplear la recta de regresión obtenida en el apartado a) para predecir a futuro el precio del crudo a partir del precio de la gasolina?
12. El dueño de una tienda de *souvenirs* sospecha que sus ventas semanales (y) están relacionadas con las fluctuaciones (semanales) del índice *Dow Jones* (x). Para corroborar su sospecha recolecta los siguientes datos:

<i>Dow Jones</i>	58.3	62.9	46.3	48.2	58.2	65.8
ventas sem.	2215	2518	1781	1823	2117	2703
<i>Dow Jones</i>	36.7	32.3	52.7	39.3	58.7	39.3
ventas sem.	1423	1532	1879	1713	2122	2346

- Representa el diagrama de dispersión para estos datos y ajusta la recta de regresión para las ventas semanales en función de la fluctuación (semanal) del *Dow Jones*.
- ¿Qué puedes decir sobre la sospecha del dueño de la tienda?
- ¿Crees que la asociación entre x e y significa que x afecta a y ? Justifica tu respuesta.