

Tema 8: Regresión lineal y análisis de la varianza

1. Una compañía de energía eléctrica pretende desarrollar un modelo lineal para el consumo de energía en función de la temperatura media diaria durante los meses de invierno. En 9 días elegidos al azar se obtuvo la siguiente información:

Temperatura ($^{\circ}$ C)	0	2	4	8	13	-4	-6	-8	-11
Consumo (Kw/día)	70	79	67	66	63	97	82	90	107

- Obtener la recta de regresión estimada e interpretar sus coeficientes.
 - Descomponer la variabilidad de los datos de consumo. Obtener la varianza residual, el coeficiente de determinación y dar su interpretación.
 - Obtener la predicción del consumo de energía para un día con temperatura media 10 grados y estimar el error cuadrático medio de esta predicción.
 - Si se suponen las hipótesis adecuadas de normalidad, ¿se puede considerar, al nivel de significación $\alpha = 0.05$, que el consumo no depende linealmente de la temperatura?
2. Una compañía de seguros desea establecer una relación lineal, y el grado de dicha relación, para determinar el montante anual del seguro de vida del cabeza de familia en función del ingreso mensual familiar. Observada una muestra aleatoria de 10 familias elegidas de forma independiente, se obtuvo la siguiente información:

Ingreso (cientos de euros)	10	10	15	20	20	25	25	30	30	30
Seguro (decenas de euros)	50	35	55	55	70	65	65	60	75	90

- ¿Cabe pensar en la citada relación lineal? En caso afirmativo, estimar la recta de regresión, interpretar los coeficientes y obtener la predicción del montante del seguro de vida para un ingreso mensual de 2800 euros.
 - Suponiendo las hipótesis adecuadas de normalidad, realizar el contraste de regresión al nivel de significación $\alpha = 0.05$ e interpretar el resultado.
3. Los datos de la siguiente tabla representan las calificaciones medias (X) de 7 recién graduados y sus respectivos salarios iniciales (Y) en miles de euros:

X	2.75	2.85	2.95	3.05	3.2	3.4	3.6
Y	16.8	18.8	17.2	17.2	21.2	21.5	22.4

- Estimar la recta de regresión de Y sobre X e interpretar sus coeficientes.
- Determinar la varianza residual.
- Calcular e interpretar el coeficiente de determinación.
- Predecir el salario inicial de un estudiante con una calificación media de 3.25.
- Suponiendo las hipótesis adecuadas de normalidad, contrastar la hipótesis de que no existe relación lineal entre ambas variables al nivel de significación $\alpha = 0.01$.

4. En cierto estudio sobre la relación entre el diámetro de los guisantes (X) y el diámetro medio de sus descendientes (Y), Galton obtuvo los siguientes resultados:

D. Padres	21	20	19	18	17	16	15
D. Descendientes	17.26	17.07	16.37	16.4	16.13	16.17	15.98

- a) Determinar el modelo de regresión lineal estimado de Y sobre X e interpretar el valor estimado de la pendiente. Dar la predicción del diámetro de los guisantes cuyos progenitores tienen un diámetro de 18.5. Dar una medida de la bondad del ajuste de los datos a la recta estimada.
- b) Suponiendo las hipótesis adecuadas de normalidad, ¿puede deducirse, al nivel 0.05, que no hay relación lineal entre las variables consideradas? Relacionar este resultado con las conclusiones anteriores.
5. Una compañía farmacéutica investiga los efectos de 5 compuestos. El experimento consiste en inyectar los compuestos a 13 ratas de características similares y anotar los tiempos de reacción. Los animales se clasifican en 5 grupos de 4, 2, 2, 3 y 2 ratas, respectivamente, y a cada grupo se le administra un compuesto diferente, obteniéndose los resultados de la siguiente tabla:

Grupo	Tiempo de reacción (minutos)			
1	8.3	7.6	8.4	8.3
2	7.4	7.1		
3	8.1	6.4		
4	7.9	8.5	10.0	
5	7.1	8		

Suponiendo que se verifican las hipótesis de normalidad, aleatoriedad, independencia e igualdad de varianzas, contrastar la hipótesis de que los tiempos medios de reacción coinciden en los cinco grupos y, por tanto, la eficacia de los cinco compuestos es la misma.

6. Se quiere estudiar la eficacia de tres fertilizantes, A , B y C , en la producción de cierto fruto. Para ello se aplica el A en 8 parcelas, el B en 6, y el C en 12 parcelas. Las parcelas son de características similares en cuanto a fertilidad, por lo que se considera que las diferencias en la producción, si las hay, serán debidas al tipo de fertilizante. Las toneladas producidas en cada parcela en una determinada temporada son:

A	6	7	5	6	5	8	4	7				
B	10	9	9	10	10	6						
C	3	4	8	3	7	6	3	6	4	7	6	3

Suponiendo que las tres muestras proceden de poblaciones normales con varianzas iguales, contrastar la hipótesis de que los abonos son igualmente eficaces.

7. Los siguientes datos corresponden a observaciones del consumo medio (en Kw/h) realizado por 5 tipos de calefactores para mantener una habitación a una temperatura determinada durante todo un día:

Tipo	Consumo (en kw/h)					
1	14.5	14.1	14.6	14.2		
2	13.2	13.4	13.0			
3	13.7	13.6	14.1	13.8	14.0	
4	12.7	13.1	12.8	12.9	13.3	13.2
5	14.6	15.2	14.4	14.8	14.3	

Contrastar la hipótesis de igualdad de los consumos medios de los diferentes tipos de calefactores. ¿Bajo qué hipótesis se puede realizar este contraste?

8. En un tratamiento contra la hipertensión se seleccionaron 35 enfermos de características similares. Los enfermos se distribuyeron en cuatro grupos de 10 (P, A, B y AB). El grupo P tomó “placebo” (fármaco inocuo), el grupo A tomó un fármaco “A”, el grupo B un fármaco “B” y el grupo AB una asociación entre “A” y “B”. Para valorar la eficacia de los tratamientos, se registró el descenso de la presión diastólica desde el inicio del tratamiento hasta después de una semana de tratamiento. Los resultados, después de registrarse algunos abandonos, fueron:

P	10	0	15	-20	0	15	-5			
A	20	25	33	25	30	18	27	0	35	20
B	15	10	25	30	15	35	25	22	11	25
AB	10	5	-5	15	20	20	0	10		

A la vista de estos datos, ¿Puede afirmarse que el descenso de la presión diastólica coincide en los cuatro grupos? ¿Bajo qué hipótesis?