

Estadística (Química)
Práctica 8 - Regresión Lineal

1. Como parte de una investigación se trata de establecer si se puede utilizar la concentración de estrona en saliva para predecir la concentración del esteroide en plasma libre. Se obtuvieron los siguientes datos de 14 varones sanos, siendo $X :=$ concentración de estrona en saliva (en pg/ml) e $Y :=$ concentración de estrona en plasma libre (en pg/ml).

X	Y	X	Y
7.4	30.0	14.0	49.0
7.5	25.0	14.5	55.0
8.5	31.5	16.0	48.5
9.0	27.5	17.0	51.0
9.0	39.5	18.0	64.5
11.0	38.0	20.0	63.0
13.0	43.0	23.0	68.0

- (a) Represente estos datos en un diagrama de dispersión.
- (b) ¿Es razonable suponer que existe una relación lineal entre X e Y que permita predecir Y en función de X ?
- (c) Calcule el coeficiente de determinación R^2 . Interprete el significado de esta medida en este ejemplo.
2. En una experiencia para calibrar un instrumento con el objeto de medir la resistencia eléctrica de cierto material, se obtuvieron las siguientes mediciones:

X	60	70	80	100
Y	38	64	72	110
	44	70	76	118
	50		82	

donde X es la resistencia eléctrica (en ohms) determinada por un método suficientemente exacto como para ser considerado sin error e Y es la medición leída en el instrumento.

- (a) Encontrar un intervalo de confianza del 95% para el valor esperado de Y cuando $X = 90$ ohms.
- (b) Si se toma una nueva porción de material, se mide su resistencia eléctrica y ésta resulta 85 ohms, hallar un intervalo de predicción del 95% para la medición aún no observada.
- (c) Suponiendo que se hace una nueva observación independiente de las anteriores y ésta resulta igual a 73, encontrar una región de confianza del 90% para el verdadero valor de la resistencia eléctrica que corresponderá a ese material (este es el objeto fundamental de la experiencia, pues se desea calibrar el instrumento de manera que luego se mida sólo Y y con esto se pueda tener una idea de cuál es la resistencia eléctrica X).
3. Se analizaron 8 soluciones estándares de plata por espectrometría de absorción atómica de llama. Se obtuvieron los siguientes resultados:

Concentración (ng/ml)	0	10	20	30	40	50	60	70
Absorbancia	0.257	0.314	0.364	0.413	0.468	0.528	0.574	0.635

- (a) Realice el gráfico de calibración, determine la pendiente y la ordenada al origen de la recta de cuadrados mínimos.
- (b) Calcule el IC al 95% para la pendiente. Lo mismo para la ordenada al origen.

- (c) Se han hecho 4 análisis para una nueva muestra. Los valores de absorbancia observados son: 0.308, 0.314, 0.347 y 0.312. Estime la concentración de plata en esa muestra y calcule un IC al 95% para dicha concentración.

4. Un investigador de marketing estudió las ventas anuales de cierto producto que fue introducido en el mercado hace 15 años. Los datos de la siguiente tabla son

X = años transcurridos desde el lanzamiento al mercado del producto

Y = ventas en miles de unidades.

X	Y
1	217
2	204
3	215
4	251
5	312
6	293
7	269
8	420
9	337
10	399
11	419
12	320
13	488
14	460
15	371

- (a) Represente estos datos en un diagrama de dispersión. ¿Es razonable suponer que existe una relación lineal entre X e Y que permita predecir Y en función de X ?
- (b) Realice un ajuste por cuadrados mínimos.
- (c) Obtenga un gráfico de residuos versus valores ajustados. ¿Le parece razonable el supuesto de igualdad de varianzas para todas las observaciones?
- (d) Suponga que $Var(Y_i) = kx_i$, defina los pesos adecuados y realice una estimación de la recta por mínimos cuadrados pesados.
5. El archivo `aire.txt` contiene tres variables `ozono`, `temp` y `grupo`, correspondientes a el nivel de ozono y la temperatura medidos en 108 días elegidos al azar en una misma latitud, longitud y altitud. La variable `grupo` es una variable categórica que indica nivel de temperatura.
- (a) Suponga que sólo dispone de la variable `ozono`, medida en la forma indicada. ¿Puede asumirse que los datos de ozono corresponden a una variable aleatoria con distribución normal? Para responder a esta pregunta haga un `qqplot`, y un test de normalidad con nivel 0.10. Halle un intervalo de confianza de nivel 0.95 para la media poblacional del ozono. Justifique la elección del método seguido para hallarlo. Estime la varianza poblacional del ozono.
- (b) Sabiendo que las mediciones de ozono se realizaron en 4 tipos de días (clasificados en días de temperatura baja, media, media alta y alta) y que esta categorización se encuentra en la variable `grupo`, realice un ANOVA para ver si la categoría influye en el nivel medio de ozono presente en un día dado. Verifique los supuestos realizados, concluya a nivel 0.05. Escriba el modelo ajustado, y dé un estimador de la varianza de los datos. Compare este modelo con el modelo de errores independientes del ítem anterior.

- (c) Si ahora uno toma en cuenta, para modelar el comportamiento de la variable **ozono**, que se cuenta con la temperatura medida en el mismo momento que fue medida la variable **ozono**, (guardada en la variable **temp**), proponga y ajuste un modelo de regresión lineal para estas dos variables. Escriba el modelo, ajuste los parámetros, chequee la bondad del ajuste. Escriba el modelo ajustado, y dé un estimador de la varianza de los datos. Compare con el modelo anova anterior.