

# Problemas de Regresión

1. Se plantaron 8 pinos de 0.3 metros de altura en medios controlados y se los sometió a distintas intensidades de irrigación para simular el efecto de las diferentes precipitaciones pluviales. Al acabar el año se midieron las alturas. En la tabla siguiente se muestran las alturas medias (en metros) ( $y_i$ ) al acabar el año y la cantidad de lluvia (en metros) simulada por cada valor  $x_i$ . Suponemos que  $Y$ , la altura del árbol al acabar el año, es una variable aleatoria con media  $\beta_0 + \beta_1 x$ , donde  $x$  es la precipitación, y con varianza constante  $\sigma^2$  por todo valor de  $x$ . Hallar las mejores estimaciones lineales sin sesgo de  $\beta_0$  y  $\beta_1$  y hallar una estimación sin sesgo de  $\sigma^2$ .

$y_i$	$x_i$
0.4826	0.2540
0.5588	0.3556
0.6350	0.4572
0.7874	0.5588
0.8382	0.6604
0.9906	0.7620
1.1176	0.8636
1.1430	0.9652

- Estimar los valores  $b_0$  y  $b_1$  para la regresión lineal de la altura del pino en función de la cantidad de lluvia.
  - Representa gráficamente los datos junto con la recta de regresión.
  - Hallar un intervalo de confianza al 95% de confianza para los parámetros  $\beta_0$  y  $\beta_1$ .
  - Calcular la estimación de la varianza común de los errores de la regresión  $\sigma^2$ .
  - Hallar el coeficiente de regresión y el coeficiente de regresión ajustado.
  - Estudiar si el modelo es homocedástico gráficamente y usando el test correspondiente.
  - Estudiar la normalidad de los residuos.
  - Estudiar la correlación de los residuos.
  - Hallar las observaciones “outliers”, los “leverages” y las observaciones influyentes.
2. Los siguientes datos relacionan la producción de biomasa de soja con la radiación solar interceptada acumulada durante un período de ocho semanas después de la emergencia. La producción de biomasa es el peso seco medio en gramos de muestras independientes de cuatro plantas.

$X$ (Radiación solar)	$Y$ Biomasa de la planta
29.7	16.6
68.4	49.1
120.7	121.7
217.2	219.6
313.5	375.5
419.1	570.8
535.9	648.2
641.5	755.6

- Estimar los valores  $b_0$  y  $b_1$  para la regresión lineal de la biomasa de la planta en función de la radiación solar.
- Representa gráficamente los datos junto con la recta de regresión.
- Hallar un intervalo de confianza al 95% de confianza para los parámetros  $\beta_0$  y  $\beta_1$ .
- Calcular la estimación de la varianza común de los errores de la regresión  $\sigma^2$ .
- Hallar el coeficiente de regresión y el coeficiente de regresión ajustado.
- Estudiar si el modelo es homocedástico gráficamente y usando el test correspondiente.

- g) Estudiar la normalidad de los residuos.
  - h) Estudiar la correlación de los residuos.
  - i) Hallar las observaciones “outliers”, los “leverages” y las observaciones influyentes.
3. Se probó un modelo de simulación para el flujo máximo de agua de las cuencas hidrográficas comparando el flujo máximo medido de 10 tormentas con predicciones del flujo máximo obtenido del modelo de simulación.  $Q_o$  y  $Q_p$  son los flujos máximos observados y pronosticados, respectivamente. Se registraron cuatro variables independientes:
- $X_1$ : area de la cuenca ( $m^2$ ),
  - $X_2$ : pendiente promedio de la cuenca (en porcentaje),
  - $X_3$ : índice de absorbencia superficial (0 = absorbencia completa, 100 = sin absorbencia), y
  - $X_4$ : intensidad de pico de lluvia calculada en intervalos de media hora.

$Q_o$	$Q_p$	$X_1$	$X_2$	$X_3$	$X_4$
28	32	.03	3.0	70	.6
112	142	.03	3.0	80	1.8
398	502	.13	6.5	65	2.0
772	790	1.00	15.0	60	.4
2294	3075	1.00	15.0	65	2.3
2484	3230	3.00	7.0	67	1.0
2586	3535	5.00	6.0	62	.9
3024	4265	7.00	6.5	56	1.1
4179	6529	7.00	6.5	56	1.4
710	935	7.00	6.5	56	.7

Consideramos  $Y = \ln\left(\frac{Q_o}{Q_p}\right)$  como variable dependiente, consideramos la regresión de  $Y$  como función de  $X_1$ ,  $X_2$ ,  $X_3$  y  $X_4$ . Se pide:

- a) Estimar los valores  $b_0, b_1, b_2, b_3, b_4$  para la regresión lineal de  $Y$  en función de  $X_i$ ,  $i = 1, 2, 3, 4$ .
- b) Hallar un intervalo de confianza al 95% de confianza para los parámetros  $\beta_i$ ,  $i = 0, 1, 2, 3, 4$ .
- c) Calcular la estimación de la varianza común de los errores de la regresión  $\sigma^2$ .
- d) Hallar el coeficiente de regresión y el coeficiente de regresión ajustado.
- e) Estudiar si el modelo es homocedástico gráficamente y usando el test correspondiente.
- f) Estudiar la normalidad de los residuos.
- g) Estudiar la correlación de los residuos.
- h) Contrastar la linealidad y la aditividad del modelo.
- i) Hallar las observaciones “outliers”, los “leverages” y las observaciones influyentes.