Regresión múltiple y otras técnicas multivariadas | Semestre 2019-2

Tarea 04

Fecha de entrega: 6 de marzo

1. Mostrar que, para el modelo RLS, se cumple $SCR = S_{xx}\hat{\beta}_1^2$.

2. *a*) Mostrar que, para el modelo RLS, se cumple

$$F = \frac{R^2(n-2)}{1-R^2},$$

donde F es el estadístico del ANOVA y R^2 es el coeficiente de determinación del modelo.

   b) Suponer que para un conjunto de 20 pares de observaciones (x, y), $r = -0.74$. Si se ajusta un modelo RLS a estos datos, ¿qué se puede decir sobre la significancia del modelo?

3. Los siguientes datos representan la relación entre el número de errores de alineación y el número de remaches faltantes para 10 diferentes aeronaves.

Núm. remaches (x)	13	15	10	22	30	7	25	16	20	15
Núm. errores (y)	7	7	5	12	15	2	13	9	11	8

   a) Ajustar un modelo RLS e interpretar las estimaciones de los coeficientes del modelo.

   b) Graficar el diagrama de dispersión de los datos y sobreponer la recta de regresión ajustada.

   c) Reportar el R^2 del modelo ajustado. ¿La interpretación del R^2 es consistente con la gráfica del inciso anterior?

   d) Contrastar las hipótesis

$$H_0 : \beta_0 = 1 \qquad \text{vs.} \qquad H_1 : \beta_0 \neq 1.$$

Utilizar un tamaño de prueba $\alpha = 0.1$. Interpretar el resultado en el contexto del problema.

   e) Estimar puntualmente el número esperado de errores de alineación de un avión con 24 remaches faltantes y construir un IC 90 %.

4. Utilizar el conjunto de datos que se envía adjunto para responder lo siguiente.

   a) Graficar un diagrama de dispersión del número de cigarros por persona (variable `cigarrette`) contra la tasa de muertes por cáncer de pulmón (por cada 100,000 habitantes, variable `lung`). ¿La gráfica indica la posibilidad de una asociación lineal entre las variables?

   b) Ajustar un modelo RLS para explicar la distribución de `lung` como función de `cigarrettes`. Reportar las estimaciones de los coeficientes e interpretarlos en el contexto de los datos.

   c) Construir la tabla ANOVA para la hipótesis de significancia del efecto del número de cigarros por persona en la tasa de muertes por cáncer de pulmón.

d) Concluir sobre la significancia del modelo si se utiliza un tamaño de prueba $\alpha = 0.05$. Reportar el p-value de la prueba.

5. En un estudio se midió la estatura (X, en cm) y el peso (Y, en kg) de 50 mujeres de entre 20 y 24 años, y se ajustó un modelo RLS para explicar el la distribución del peso en como función de la estatura. A continuación de muestra un resumen de la información obtenida.

$$\bar{x}_n = 164.9, \qquad \bar{y}_n = 59.3 \qquad S_{xx} = 2875.7, \qquad S_{yy} = 1423.5, \qquad S_{xy} = 1222.5$$

Responder lo siguiente:

a) Reportar las estimaciones de los coeficientes del modelo e interpretarlas en el contexto de los datos.

b) ¿Hay evidencia de que la estatura tiene algún efecto en el peso esperado de una persona? Considerar un tamaño $\alpha = 0.05$.

c) Construir la tabla ANOVA para este modelo.

d) Calcular el R^2 e interpretar el resultado.

6. Se ajustó un modelo de regresión lineal simple a un conjunto de datos y se obtuvo la siguiente tabla ANOVA.

FV	GL	SC	CM	F	p-value
Regresión	1	X	20.11	X	X
Error	X	92.62	X		
Total	20	112.7			

Además se calculó $S_{xx} = 770.0$. Responda lo siguiente:

a) Completar la información de la tabla anterior (sólo las celdas señaladas).

b) Contrastar la significancia del modelo. Considerar un tamaño $\alpha = 0.1$.

c) Estimar $|\,\beta_1\,|$ y calcular estimar el error estándar del estimador.

d) Calcular el R^2 e interpretar el resultado.