Modelos no paramétricos y de regresión/Estadística II | Semestre 2019-1

Tarea 3

Fecha de entrega: 12 de septiembre

1. Con los datos de desarrollo humano de las entidades del país (que pueden descargar aquí), considerar un modelo RLS para explicar la esperanza de vida con el logaritmo del ingreso:

   a) Calcular los EMV de β_0, β_1 y σ^2. Reportar los valores de las expresiones utilizadas (promedios, sumas de cuadrados o productos cruzados).

   b) Estimar las varianzas de los estimadores de β_0 y β_1 del inciso anterior.

   c) Interpretar los resultados en el contexto de los datos.

   d) Calcular los intervalos de confianza 90 % para β_0 y β_1. interpretar los intervalos calculados.

   e) Calcular un intervalo de confianza 90 % para σ^2. Reportar los cuantiles utilizados.

   - **Punto extra:** Encontrar el intervalo de confianza 90 % de menor longitud para σ^2. Describir con detalle el procedimiento a seguir para encontrar tal intervalo.

   f) Estimar la media de los años de esperanza de vida para un ingreso anual de $35,000$ USD PPC.

   g) Construir intervalos de confianza para la estimación anterior e interpretar los resultados. Reportar los errores estándar estimados y los cuantiles utilizados.

   h) Construir intervalos de predicción para una nueva observación dado un ingreso anual de $35,000$ USD PPC.

   i) Calcular SC_{TC}, SC_{reg} y SC_{error}.

   j) ¿Qué tan bueno es el ajuste del modelo? Calcular el coeficiente R^2 a partir de las sumas del inciso anterior.

2. Utilizar los resultados del ejercicio anterior para responder lo siguiente.

   a) Calcular intervalos de simultáneos de confianza 90 % para β_0 y β_1, con el método de Bonferroni.

   b) Calcular intervalos simultáneos de confianza 90 % para β_0 y β_1, con el método de Hotelling-Scheffé.

   c) Comparar la longitud de los intervalos simultáneos, ¿cuáles son mejores?

3. Utilizar los resultados del ejercicio 2 para responder lo siguiente. Suponer que se quiere hacer inferencia sobre la media de la esperanza de vida $E(Y \mid X = x)$, para niveles de ingreso $x = 10000$, 15000, 20000, 25000, 30000 USD PPC anuales.

   a) Estimar puntualmente $E(Y \mid X = x)$ y su error estándar, para los valores de x que se indican en el enunciado.

b) Construir intervalos simultáneos de confianza 90 % para las cinco medias del enunciado, utilizando el método de Bonferroni.

c) Construir intervalos simultáneos de confianza 90 % para las cinco medias del enunciado, utilizando el método de Hotelling-Scheffé.

d) Comparar las longitudes de los intervalos simultáneos, ¿cuáles son mejores? ¿la respuesta coincide con la del ejercicio anterior?

4. En el modelo RLS con errores normales $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, n$ mostrar que:

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_n)^2 = \hat{\beta}_1^2 S_{xx}$$

5. En el modelo RLS con errores normales, encontrar la prueba basada en el cociente de verosimilitudes generalizadas para contrastar las hipótesis

$$H_0 : \beta_0 = 0 \qquad \text{vs.} \qquad H_1 : \beta_0 \neq 0.$$

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	$P(> F)$
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 marcadas con X).

b) ¿Cuántas observaciones se utilizaron en el ajuste?

c) Hacer el contraste de $H_0 : \beta_1 = 0$. Considerar $\alpha = 0{,}1$.

d) Estimar a σ^2 puntualmente y por intervalo. Considerar 90 % de confianza.

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

f) ¿Qué porcentaje de la variabilidad es explicada por el modelo?

7. 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. Se quiere ajustar un modelo RLS para explicar el peso en términos de la estatura. A continuación de muestra un resumen de la información en la muestra.

$$\bar{x}_n = 164{,}9, \qquad \bar{y}_n = 59{,}3 \qquad S_{xx} = 2875{,}7, \qquad S_{yy} = 1423{,}5, \qquad S_{xy} = 1222{,}5$$

Responder lo siguiente:

a) (0.5 puntos) Estimar los parámetros del modelo β_0, β_1 y σ^2.

b) (0.25 puntos)¿Hay evidencia de que la estatura tiene algún efecto en el peso esperado de una persona? Usar $\alpha = 0{,}05$.

c) (0.5 puntos) Estimar a σ^2 puntualmente y por intervalo. Considerar 95 % de confianza.

d) (0.5 puntos) Llenar la siguiente tabla ANOVA con la información del modelo ajustado:

FV	GL	SC	CM	F	$P(>F)$
Regresión	X	X	X	X	X
Error	X	X	X		
Total	X	X			

e) (0.25 puntos) ¿Qué porcentaje de la variabilidad es explicada por el modelo?

8. En el modelo RLS, mostrar la igualdad

$$R^2 = r_{xy}^2,$$

donde R^2 es el coeficiente de determinación del modelo y r_{xy} es el coeficiente de correlación de las observaciones (x_i, y_i), $i = 1, \ldots, n$.