Modelos no paramétricos y de regresión/Estadística II Semestre 2018-1

Tarea 7

Fecha de entrega: 30 de octubre

1. Mostrar que en el modelo RML:

   a) La correlación entre los residuos $e_1, \ldots, e_n$ y las respuestas $y_1, \ldots, y_n$, es $(1 - R^2)^{1/2}$.

   b) Si se ajusta un modelo RLS de los residuos contra las y, el coeficiente de y vale $1 - R^2$.

   c) Mostrar que la correlación entre los residuos y los valores ajustados $\hat{y}_1, \ldots, \hat{y}_n$, es cero.

2. En el siguiente cuadro se muestran los resultados de un análisis de varianza para un modelo RLM.

F.V.	**G.L.**	**S.C.**	**C.M.**	**F**
Reg.	3	X	1600.81	X
Error	36	146.9	X	
TC	X	X	X	

   Responder lo siguiente.

   a) Completar la información de la tabla anterior. Únicamente las celdas marcadas con X.

   b) ¿Con cuántas variables explicativas y cuántas observaciones se ajustó el modelo?

   c) Si se toma $\alpha = 0.01$, ¿el modelo ajustado es significativo?

   d) Estimar puntual y por intervalo (de confianza 99 %) la varianza del modelo.

   e) Calcula los coeficientes R^2 y R^2-ajustado del modelo.

3. El conjunto de datos `Publicidad` contiene: cuatro variables `Ventas`, `TV`, `Radio` e `Impresos` que corresponden a las ventas semanales y gasto en publicidad de un determinado producto en 200 mercados diferentes. El objetivo es modelar las ventas semanales a partir del gasto en publicidad, ambas variables expresadas en miles de USD.

   a) Ajustar un modelo RLM para explicar las ventas semanales a partir del gasto en publicidad. Reportar las estimaciones de $\boldsymbol{\beta}$, σ^2 y $V(\hat{\boldsymbol{\beta}})$.

   b) Interpretar $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$ en el contexto de los datos.

   c) Construir la tabla ANOVA y probar la significancia del modelo. Interpretar los resultados. Utilizar $\alpha = 0.01$.

   d) Construir intervalos de confianza 99 % para las componentes de $\boldsymbol{\beta}$, individuales y simultáneos (Bonferroni y Hotelling-Scheffé). Comparar las longitudes de los intervalos.

   e) Probar la significancia del modelo utilizando los intervalos de confianza simultáneos del inciso anterior y comparar los resultados con el inciso 3. Utilizar $\alpha = 0.01$.

   f) Calcular el R^2 con y sin ajuste e interpretar.

   g) Contrastar si el efecto de `TV` es el doble que el efecto de `Radio`. Interpretar los resultados en el contexto de los datos.

4. Considerar el siguiente conjunto de datos

Y	6	8	1	0	5	3	2	-4	10	-3	5
X_1	1	4	9	11	3	8	5	10	2	7	6
X_2	8	2	-8	-10	6	-6	0	-12	4	-2	-4

   a) Estimar $\boldsymbol{\beta}$ en el modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

b) Escribir la tabla del análisis de varianza y determinar la significacia del modelo. Usar $\alpha = 0.05$.

c) Calcular el coeficiente R^2 e interpretar el resultado en términos de la varianza explicada por el modelo.

d) Estimar la matriz de covarianzas de $\boldsymbol{\beta}$.

e) Estimar, puntualmente y por intervalo de confianza 95 %, el valor esperado de Y dado $X_1 = 3$ y $X_2 = 5$.

f) Calcular la longitud del intervalo de predicción 95 % para Y dado $X_1 = 3$ y $X_2 = 5$.

5. Considerar el conjunto de datos en el archivo `cuad`.

a) Hacer un gráfico de dispersión de X contra Y. ¿Es razonable suponer que la relación entre las variables es lineal?

b) Ajustar el modelo

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon.$$

Reportar la estimación de $\boldsymbol{\beta}$.

c) Construir la tabla del análisis de varianza y determinar la significancia del modelo. Usar $\alpha = 0.05$.

d) Graficar el modelo ajustado en sobre el gráfico de dispersión de los datos. ¿El modelo captura adecuadamente la relación entre X y Y?

e) Estimar, puntualmente y por intervalo de confianza 95 %, el valor de X, digamos x^*, para el cual $E(Y \mid X)$ se minimiza. Para construir el intervalo de confianza se debe utilizar el método delta.

f) Calcular un intervalo de confianza 95 % para $E(Y \mid X = x^*)$.

g) Calcular un intervalo de predicción de confianza 95 % para Y dado $X = x^*$.