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

Tarea 2

Fecha de entrega: 4 de septiembre

1. Utilizar el resultado
$$\frac{(n-2)\hat{\sigma}^2_{MCO}}{\sigma^2} \sim \chi^2_{n-2}$$
para encontrar $ECM\left(\hat{\sigma}^2_{MCO}\right)$ y $ECM\left(\hat{\sigma}^2_{MV}\right)$. En términos del Error Cuadrático Medio (ECM), ¿qué estimador es mejor?

   En el caso de $\hat{\beta}_0$ y $\hat{\beta}_1$ ya encontramos los MELI. Para σ^2 vamos a comparar los dos estimadores que hemos encontrado y decidir cuál de ellos es mejor, utilizamos el ECM en lugar de la varianza porque $\hat{\sigma}^2_{MV}$ es sesgado.

2. Utilizar el resultado
$$\frac{(n-2)\hat{\sigma}^2_{MCO}}{\sigma^2} \sim \chi^2_{n-2}$$
para encontrar un intervalo de confianza $100(1-\alpha)\,\%$ para σ^2.

   El resultado proporciona una cantidad pivotal para σ^2 a partir de la que podemos *despejar* un intervalo de confianza para este parámetro.

3. Considerar el modelo RLS con errores normales $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, \ldots, n$, y los estimadores de MV. ¿Qué pasa con $\hat{\beta}_0$ y $\hat{\beta}_1$ cuando se aplican las siguientes transformaciones a los datos?

   a) $Y^* = Y + c$, $c \in \mathbb{R}$.
   b) $Y^* = cY$, $c \in \mathbb{R} \setminus \{0\}$.
   c) $Y^* = cY + d$, $c \in \mathbb{R} \setminus \{0\}$ y $d \in \mathbb{R}$.
   d) $X^* = X + c$, $c \in \mathbb{R}$
   e) $X^* = cX$, $c \in \mathbb{R} \setminus \{0\}$.
   f) $X^* = cX + d$, $c \in \mathbb{R} \setminus \{0\}$ y $d \in \mathbb{R}$.

   En este ejercicio se deben explorar los cambios qué ocurren en las estimaciones de los parámetros cuando se aplican transformaciones de localización y de escala a los datos. Por ejemplo, las transformaciones del tipo $Y^* = cY$ se dan cuando se cambia la escala de la variable dependiente de kilómetros a metros y las transformaciones del tipo $X^* = cX + d$ se dan cuando la variable explicativa se transforma de grados centígrados a Farenheit.

4. 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.

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

   a)

   $$Cov\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\sigma^2 \frac{\bar{x}_n}{S_{xx}}$$

   Hint: Expresar a $\hat{\beta}_0$ y $\hat{\beta}_1$ como combinación lineal de las Y_i y utilizar las propiedades de la covarianza.

   b)

   $$V\left(\hat{\beta}_0 + \hat{\beta}_1 x_*\right) = \sigma^2 \left(\frac{1}{n} + \frac{\left(x_* - \bar{x}_n\right)^2}{S_{xx}}\right)$$

   Hint: Utilizar el resultado del inciso anterior.

   c)

   $$\sum_{i=1}^{n} \left(y_i - \bar{y}_n\right)^2 = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}_n\right)^2 + \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$$

   Hint: Sumar y restar $\hat{y}_i$ a cada sumando, agrupar términos, desarrollar el binomio al cuadrado, sumar y simplificar términos.

   d)

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

   Hint: Sustituir la expresión de $\hat{y}_i$ en términos de $\hat{\beta}_0$ y $\hat{\beta}_1$.