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

Tarea 5

Fecha de entrega: 9 de octubre

Sean $\mathbf{y}$ un vector de dimensión n, $\mathbf{X}$ una matriz de dimensión $(p+1)\times n$ de rango completo (por columnas) y $\boldsymbol{\beta}$ un vector de dimensión $p+1$.

Notación:

- $\mathbf{I}_n$: matriz identidad de dimensión n.
- $\mathbf{1}_n$: vector de unos de dimensión n.
- $\mathbf{J}_n$: matriz de unos de dimensión $n\times n$.

1. Mostrar que
$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y}' - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

2. Si $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ y $\hat{y} = \mathbf{X}\hat{\boldsymbol{\beta}}$, mostrar que
$$(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{y}' - \hat{\mathbf{y}}'\hat{\mathbf{y}}.$$

3. Si $\mathbf{C}$ es una matriz de dimensión $(p+1)\times n$ tal que $\mathbf{CX}=\mathbf{0}$, mostrar que
$$\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{C}\right)'\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{C}\right) = (\mathbf{X}'\mathbf{X})^{-1} + \mathbf{CC}'.$$

4. Mostrar que las matrices $\mathbf{H}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ e $\mathbf{I}-\mathbf{H}$ son simétricas e idempotentes, de rangos $p+1$ y $n-p-1$, respectivamente.

5. Mostar que $\mathbf{X}'(\mathbf{I}-\mathbf{H})=\mathbf{0}$ y $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}-\mathbf{H})=\mathbf{0}$.

6. Justificar a partir de $\mathbf{X}'(\mathbf{I}-\mathbf{H})=\mathbf{0}$ que la suma de los residuos es 0, es decir,
$$\sum_{i=1}^{n} e_i = 0$$
donde e_i es la i-ésima componente del vector $\mathbf{e}=\mathbf{y}-\hat{\mathbf{y}}$.
Hint: Expresar la suma de los residuos como función de $\mathbf{a}'\mathbf{e}$, para algún vector $\mathbf{a}$, ¿qué relación tienen tal vector $\mathbf{a}$ y $\mathbf{X}$.

7. Mostrar que $\hat{\sigma}^2_{MCO}$ es insesgado y calcular su varianza. Calcular el ECM (error cuadrático medio de $\hat{\sigma}^2_{MCO}$. Según el ECM, ¿qué estimador es mejor?

8. La suma de cuadrados de regresión de define como
$$SC_{reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_n)^2$$
donde $\hat{y}_i$ es la i-ésima componente del vector $\hat{\mathbf{y}}$ y $\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$. Mostrar que
$$SC_{reg} = \mathbf{y}'(\mathbf{H}-\frac{1}{n}\mathbf{J})\mathbf{y}.$$

9. Mostrar que $\mathbf{H}-\frac{1}{n}\mathbf{J}$ es simétrica e idempotente de rango p. Además mostrar que
$$(\mathbf{I}-\mathbf{H})(\mathbf{H}-\frac{1}{n}\mathbf{J}).$$

⋆ Opcional: bajo los supuestos del modelo RLS y con los resultados sobre distribuciones de formas cuadráticas vistos en clase, justificar las siguientes afirmaciones.

- $SC_{TC}/\sigma^2 = \mathbf{Y}'(\mathbf{I}-\frac{1}{n}\mathbf{J})\mathbf{Y}/\sigma^2 \sim \chi^{*2}_{n-1,\lambda}$, calcular el parámetro de no centralidad.
- $SC_{reg}/\sigma^2 = \mathbf{Y}'(\mathbf{H}-\frac{1}{n}\mathbf{J})\mathbf{Y}/\sigma^2 \sim \chi^{*2}_{n-1,\lambda}$, calcular el parámetro de no centralidad.
- $SC_{reg} \perp SC_{error}$.