TALLER 1 - MODELOS ACTUARIALES PARA SEGUROS GENERALES

1. Ravishanker & Dey (2002)[p.133] Cuando un objeto de masa uno es sujeto a una fuerza desconocida β por una longitud de tiempo t, su posición cambia a $\frac{t^2\beta}{2}$. Las posiciones observadas son $Y_1, Y_2, \ldots, Y_n$ medidas en los tiempos $t_1, t_2, \ldots, t_n$. Asumiendo un modelo lineal $Y_i = \frac{t_i^2\beta}{2} + \varepsilon_i$, $i = 1, 2, \ldots, n$ y que los errores ε_i son tal que $E(\varepsilon_i) = 0$ y $Var(\varepsilon_i) = \sigma^2$, obtenga el estimador vía mínimos cuadrados del parámetro β y la varianza de este estimador.

2. Ravishanker & Dey (2002)[p.133] Sea $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_p + \boldsymbol{\varepsilon}$, tal que $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ y $Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}_n$. Si $n = 10$, $p = 3$, $\boldsymbol{y}^t\boldsymbol{y} = 58$ y dadas las ecuaciones normales

$$
\begin{aligned}
4\hat{\beta}_1 + 2\hat{\beta}_2 - 2\hat{\beta}_3 &= 4 \\
2\hat{\beta}_1 + 2\hat{\beta}_2 + \hat{\beta}_3 &= 7 \\
-2\hat{\beta}_1 + \hat{\beta}_2 + 6\hat{\beta}_3 &= 9
\end{aligned}
$$

encuentre el estimador vía mínimos cuadrados del parámetro $\boldsymbol{\beta}$ y σ^2. ¿Cuáles son los estimativas para $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_1 - \hat{\beta}_2$ y $\hat{\beta}_1 + \hat{\beta}_3$?

3. Ravishanker & Dey (2002)[p.134] Sean $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$ variables aleatorias que denotan el rendimiento de un proceso de producción de seis días consecutivos. La máquina A fué usada los días 1, 3 y 5, mientras que la máquina B fué usada los días 2, 4, y 6. Considere los modelos

$$
\begin{aligned}
(i) \quad & Y_i = \beta_0 + (-1)^i\beta_1 + \epsilon_i, \\
(ii) \quad & Y_i = \beta_0 + (-1)^i\beta_1 + i\beta_2 + \epsilon_i,
\end{aligned}
$$

En cada caso asuma $E(\epsilon_i) = 0$ y $Var(\epsilon_i) = \sigma^2$. Obtenga el estimador mínimos cuadrados para β_1 bajo cada modelo y muestre que el cociente entre $Var(\hat{\beta}_1)$ bajo el modelo (i) y $Var(\hat{\beta}_1)$ bajo el modelo (ii) es 32/35.

4. Ravishanker & Dey (2002)[p.135] Considere el modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, en que

$$
\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix},
$$

y $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

   a) ¿$\beta_1 + \beta_3$ es estimable?. Si la respuesta es afirmativa, encuentre su varianza.

   b) ¿β_2 es estimable?. Si la respuesta es afirmativa, encuentre su varianza.

   c) Encuentre $\boldsymbol{P}_X$ y $\boldsymbol{I} - \boldsymbol{P}_X$.

5. Paula (2012)[p.107] Sean dos muestras independientes de tamaños n_1 y n_2 tomadas de dos poblaciones normales, respectivamente, con medias μ_1 y μ_2, y varianza igual. Si $\mu_1 = \alpha + \beta$ y $\mu_2 = \alpha - \beta$, muestre que la estadística F para probar $H_0 : \beta = 0$ contra $H_a : \beta \neq 0$ se puede escribir de forma simplificada como
$$F = \frac{(n-2)^{\frac{n_1 n_2}{n}}(\bar{y}_1 - \bar{y}_2)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{n_1 n_2}{n}(\bar{y}_1 - \bar{y}_2)^2}$$
con $\bar{y}$, $\bar{y}_1$ y $\bar{y}_2$ las medias muestrales respectivas.

6. Paula (2012)[p.175] En el archivo restaurante.dat (moodle)[a] se describen las facturaciones anuales (primera variable), como los gastos en publicidad (segunda variable), en miles de dólares, de una muestra aleatoria de 30 restaurantes (Montgomery, Peck e Vining, 2001, pgs. 197-200). El objetivo principal es intentar relacionar la facturación con los gastos en publicidad.

   a) Hacer un análisis descriptivo de los datos, en particular un diagrama de dispersión entre las variables.
   b) Ajuste un modelo de regresión lineal entre la facturación y el gasto, y verifique usando las técnicas de diagnóstico vistas en clase si existen problemas de supuestos.

7. Frees (2010)[p.56] El archivo WiscNursingHome.csv (moodle) [b] contiene información sobre ancianatos en el estado de Wisconsin (EEUU). Entre las variables que tiene está el identificador del ancianato (hospID), el año del reporte (CRYEAR), la exposición de los pacientes (suma de tiempo que han estado en el ancianto) expresada en años (TPY), el número de camas (NUMBED) y el número de pies cuadrados del ancianato (SQRFOOT). La variable dependiente es TPY. Para los datos reportados en el año 2000,

   a) Hacer un análisis descriptivo de los datos, en particular un diagrama de dispersión entre las variables.
   b) Ajuste un modelo de regresión lineal entre TPY y las variables NUMBED y SQRFOOT. ¿Las dos variables deberían considerarse en el modelo?, justifique.
   c) Verifique usando las técnicas de diagnóstico vistas en clase si existen problemas de supuestos para el modelo seleccionado.

8. Paula (2012)[p.104]Considere la distribucion dada por la función de densidad
$$f(y;\theta,\phi) = a(y,\phi)\exp\{\phi[\theta(y+1) - \theta\ln(\theta)]\},$$
con $\theta > 0$, $-\infty < y < \infty$, $\phi^{-1} > 0$ el parámetro de dispersión y $a(.,.)$ una función normalizadora. Muestre que esta distribución pertenece a la familia exponencial.

[a] Fuente: http://www.ime.usp.br/~giapaula/textoregressao.htm
[b] Fuente: https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression %20Modeling/BookWebDec2010/data.html

9. Paula (2012)[p.104] Considere la distribucion dada por la función de densidad

$$f(y;\theta,\phi) = \frac{\phi a(y,\phi)}{\pi(1+y^2)^{\frac{1}{2}}} \exp\left\{\phi[\theta y + (1-\theta^2)^{\frac{1}{2}}]\right\},$$

con $\theta > 0$, $-\infty < y < \infty$, $\phi^{-1} > 0$ el parámetro de dispersión y $a(.,.)$ una función normalizadora. Muestre que esta distribución pertenece a la familia exponencial.

10. Dobson (1990)[p.35] Considere la distribución de valor extremo (Gumbel)

$$f(y;\theta,\phi) = \frac{1}{\phi} \exp\left\{\left(\frac{y-\theta}{\phi}\right) - \exp\left(\frac{y-\theta}{\phi}\right)\right\},$$

con $\phi > 0$. Muestre si esta distibución pertenece o no a la familia exponencial.

Referencias

Dobson, A. (1990), *An introduction to generalized linear models*, Chapman and Hall.

Frees, E. (2010), *Regression modeling with actuarial and financial applications*, Cambridge University Press, Cambridge.

Paula, G. (2012), 'Modelos de regressão com apoio computacional'.

Ravishanker, N. & Dey, D. (2002), *A first course in linear model theory*, Chapman & Hall/CRC, Boca Ratón.