**Universidad de la República** **Modelos Lineales**
**Facultad de Ingeniería - IMERL** **Primer semestre 2016**

# Práctico 3

1. Realice el análisis completo de los residuos del modelo de regresión parábolico de los datos de tráfico de los apuntes de Carmona.

2. Los datos que se presentan en la tabla `datos 5.3.csv` representan la cantidad de dinero aportado por los alumnos de 14 clases, de una universidad del sur. La variable regresora $x$ indica la cantidad de años desde la graduación. (Fuente: Datos analizados por The statistical Consulting Center, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1979.)

   *a*) Ajuste un modelo de regresión lineal simple. Grafique los residuos contra la variable regresora $x$.

   *b*) Obtenga los residuos "studentizados" y grafíquelos contra la variable regresora $x$.

   *c*) ¿Los gráficos anteriores sugieren la necesidad de realizar alguna modificación en el modelo?. Explique.

   *d*) Obtenga y analice los valores $h_{ii}$, elementos de la diagonal principal de la hat matrix.

   *e*) Obtenga y analice los residuos R-studentizados y compararlos con los residuos studentizados.

3. Probar que:

   *a*) $\frac{1}{n} \leq h_{ii} \leq 1$ y que $-0{,}5 \leq h_{ij} \leq 0{,}5$ para todo $j \neq i$.

   *b*) que en la regresión lineal simple $h_{ii} = \dfrac{1}{n} + \dfrac{(x_i - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$ y $h_{ij} = \dfrac{1}{n} + \dfrac{(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$.

   *c*) que en la regresión lineal múltiple $h_{ii} = \dfrac{1}{n}\left(1 + \underbrace{(\mathbf{x}_i - \overline{\mathbf{x}})'(\tilde{X}'\tilde{X})^{-1}(\mathbf{x_i} - \overline{\mathbf{x}})}_{D_i^2}\right)$ donde $\tilde{X}$ es la matriz centrada, $\mathbf{x}_i$ es la i-esima final de $X$, y por lo tanto $(\mathbf{x_i} - \overline{\mathbf{x}})$ es la i-esima fila de $\tilde{X}$, y $D_i^2$ es la *distancia de Mahalanobis* de $\mathbf{x_i}$ a la media $\overline{\mathbf{x}}$.

4. Probar que $\mathrm{Dffits}_i = |t_i|\sqrt{\dfrac{h_{ii}}{1 - h_{ii}}}$ donde $t_i$ son los residuos studentizados externamente.

5. Considere los datos presentados en la tabla `datos 3.8.txt`, extraídos de *Procedures and Analysis for Staffing Standards Development: Data/ Regression Analysis Handbook.*

   Para el modelo completo, con 5 variables explicativas, obtenga (a) los DFBETAS, (b) las distancias de Cook, (c) los Dffits y (d) los $h_{ii}$.

   A partir de las medidas obtenidas en los items anteriores, determine si existe la necesidad de chequear los datos de alguno de los hospitales.

6. Cuando la gasolina se bombea en el tanque de un automóvil, algunos vapores son ventilados a la atmósfera. Se realizó un experimento para determinar si la variable $y$ (cantidad de vapor), puede ser predicha usando las siguientes cuatro variables referidas a las condiciones iniciales del tanque y de la gasolina bombeada.

   - $x_1$ = Temperatura del tanque ($^o$F)
   - $x_2$ = Temperatura de la gasolina ($^o$F)

- $x_3$ = Presión del vapor en el tanque (psi)
- $x_4$ = Presión del vapor de la gasolina (psi)

Los datos estáan disponibles en el archivo Tabla 7.3.txt (Weisberg 1985, p.138).

*a*) Calcular $\widehat{\beta}$ y $\widehat{\sigma}^2$ y una estimación de $\text{Var}(\widehat{\beta})$.

*b*) Calcular el coeficiente de determinación y el coeficiente de determinación ajustado.

*c*) Realizar el test de significación del modelo.

*d*) Hallar los intervalos de confianza al 95 % para $\beta_0, \beta_1, \beta_2$ y $\beta_3$.

*e*) Realizar los siguientes tests de hipótesis:

1) $(H_0): \beta_j = 0$ para $j = 1, 2, 3, 4$.
2) $(H_0): \beta_1 = \beta_3 = 0$.
3) $(H_0): \beta_1 = \beta_2 = \frac{\beta_3}{12} = \frac{\beta_4}{12}$.
4) $(H_0): \beta_1 = \beta_2$.
5) $(H_0): \beta_3 = \beta_4$.
6) $(H_0): \beta_1 = \beta_2 \, \text{y} \, \beta_3 = \beta_4$.

7. En un intento por obtener el rendimiento máximo en una reacción química, se eligieron los valores de las siguientes variables para realizar un experimento.

- $x_1$ = Temperatura ($^o$F).
- $x_2$ = Concentración de un reactivo ( %)
- $x_3$ = Tiempo de reacción (horas)

Se observaron las siguientes variables de respuesta:

- $y_1$ = Porcentaje del material inicial que no presenta cambios.
- $y_2$ = Porcentaje que se convierte al producto deseado.

Los datos están disponibles en el archivo Tabla 7.4.txt (Box and Youle 1955, Andrews and Herzberg 1985, p.188.)

*a*) Usando la variable $y_1$:

1) Calcular $\widehat{\beta}$ y $\widehat{\sigma}^2$.
2) Calcular una estimación de $\text{Var}(\widehat{\beta})$.
3) Calcular el coeficiente de determinación y el coeficiente de determinación ajustado.
4) Realizar el test de significación del modelo.
5) Hallar los intervalos de confianza al 95 % para $\beta_0, \beta_1, \beta_2$ y $\beta_3$.
6) Para intentar encontrar el máximo rendimiento para $y_1$ se plantea el siguiente modelo:

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2 + \beta_7 x_1 x_2 + \beta_8 x_1 x_3 + \beta_9 x_2 x_3 + \epsilon$$

(i) Calcular $\widehat{\beta}$ y $\widehat{\sigma}^2$.

(ii) Calcular el coeficiente de determinación y el coeficiente de determinación ajustado.

*b*) Usando la variable $y_2$:

1) Encontrar un intervalo de confianza para $E(y_0) = \mathbf{x_0}'\beta$ donde $\mathbf{x_0}' = (1, 165, 32, 5)$.
2) Encontrar un intervalo de predicción para $y_0 = \mathbf{x_0}'\beta + \epsilon$ donde $\mathbf{x_0}' = (1, 165, 32, 5)$.
3) Testear $(H_0)\, 2\beta_1 = 2\beta_2 = \beta_3$.