Ejercicios para entregar de Métodos de Regresión. Curso 2011/12.

Grado en Estadística y Empresa.

NORMAS: La resolución de los problemas puede realizarse en grupos de dos.

Notación matricial del modelo de regresión lineal múltiple:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}, \quad \text{donde } \mathbf{U} \sim NM_n(\mathbf{0}, \sigma^2\mathbf{I}),$$

siendo $\mathbf{Y}$ el vector de respuestas de dimensión $n \times 1$, $\mathbf{X}$ la matriz de regresores de dimensión $n \times (k+1)$, $\boldsymbol{\beta}$ el vector $(k+1) \times 1$ de parámetros a estimar.
El etimador por Mínimos Cuadrados Ordinarios (MCO) de $\boldsymbol{\beta}$ es $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, el vector de respuestas estimadas (hiperplano de regresión) es $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ y el vector de residuos es $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$.

Problema 1 *Con las hipótesis habituales del modelo de regresión lineal múltiple, demostrar que:*

1. *Los residuos de la estimación por MCO pueden expresarse como* $\mathbf{e} = \mathbf{M}\mathbf{U}$*, donde* $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ *es una matriz idempotente.*
2. *La suma de cuadrados de los residuos de la estimación MCO puede escribirse como:*
$$\mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\,\mathbf{X}'\,\mathbf{Y}.$$
3. *El vector de respuestas estimadas está incorrelado con el vector de residuos, es decir,* $\hat{\mathbf{Y}}'\mathbf{e} = 0$.

Problema 2 *Demostrar que la matriz de covarianzas entre el vector de estimadores de los coeficientes y la perturbación aleatoria del modelo de regresión lineal múltiple es* $\mathrm{Var}(\hat{\boldsymbol{\beta}}, \mathbf{U}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Problema 3 *Demostrar que la matriz de covarianzas entre los estimadores de los coeficientes y los residuos MCO del modelo de regresión lineal múltiple es una matriz de ceros.*

Problema 4 *Cuando las matrices de datos están centradas, se obtiene el modelo en desviaciones:*

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{U}, \quad \textit{donde } \mathbf{U} \sim NM_n(\mathbf{0}, \sigma^2\mathbf{I}),$$

siendo $\tilde{\mathbf{Y}} = \mathbf{Y} - \overline{y}\mathbf{1}$ *el vector de respuestas centrado de dimensión* $n \times 1$, $\overline{y} = \frac{1}{n}\mathbf{1}'\mathbf{Y}$, $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}\overline{\mathbf{x}}'$ *la matriz de regresores centrados de dimensión* $n \times k$, $\overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}'\mathbf{1}$ *es el vector de medias de los regresores de dimensión* $k \times 1$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ *el vector* $k \times 1$ *de parámetros a estimar. (Observad que* $\tilde{\mathbf{X}}$ *no contiene la columna de unos y que, por tanto, el parámetro* β_0 *no está).*
Demostrar que:

1. *Las ecuaciones normales dan lugar a la ecuación* $\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})\boldsymbol{\beta}$.
2. *El estimador MCO de* $\boldsymbol{\beta}$ *es* $\hat{\boldsymbol{\beta}} = \mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{XY}}$*, donde* $\mathbf{S}_{\mathbf{XX}}$ *es la matriz de covarianzas entre los regresores y* $\mathbf{S}_{\mathbf{XY}}$ *es el vector de covarianzas cruzadas entre los regresores y el vector de respuestas.*
3. $\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}$ *coincide con la suma de cuadrados totales (VT) del Teorema Fundamental de la Regresión.*
4. *El coeficiente de determinación se puede calcular como:*
$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}}{\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}}$$
(Indicación: Expresar R^2 *en función de los residuos centrados* $\tilde{\mathbf{e}} = \tilde{\mathbf{Y}} - \hat{\tilde{\mathbf{Y}}}$*).*

Problema 5 *De un modelo de regresión lineal múltiple sabemos que la matriz de regresores y el vector de residuos son:*

$$\mathbf{X} = \begin{pmatrix} 1 & 4 & A \\ 1 & -2 & 1 \\ 1 & 1 & 4 \\ 1 & -2 & 8 \\ 1 & B & 2 \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} C \\ 4 \\ 5 \\ -7 \\ -1 \end{pmatrix}.$$

Completar los valores que faltan (A, B, C) *en las matrices anteriores.*

Problema 6 *Se plantea un modelo de regresión lineal múltiple con 4 regresores, incluyendo la constante. Se estima el modelo a partir de una muestra de tamaño 100 y se obtiene un coeficiente de determinación corregido de* $\tilde{R}^2 = 0.7$*. Determinar el porcentaje de variabilidad explicado por el modelo de regresión.*

Problema 7 *En un modelo de regresión lineal se dispone de tres valores de la variable respuesta, 2, 4 y 8. Después de estimar el modelo, se obtiene que* $\sum_{i=1}^{3} \hat{y}_i^2 = 80$.

1. *Determinar cuánto vale la suma de los cuadrados de los residuos (VNE).*
2. *Calcular e interpretar el coeficiente de determinación.*

Problema 8 *Como resultado de contrastar conjuntamente determinadas restricciones en el siguiente modelo*

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, 2, \ldots, 20,$$

se ha obtenido que $VNE = 5758.092$ *y el estadístico* $F = 2.395187$. *Si además se sabe que*

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = 1723.96338,$$

determinar cuántas restricciones se han contrastado.

Problema 9 *Un grupo empresarial ha estimado un modelo de regresión lineal múltiple para explicar el número de uniddes vendidas en función del número de establecimientos que posee, de los gastos en publicidad (en miles de euros) y del precio unitario de cada unidad vendida (en euros). Se toma una muestra de 34 empresas de dicho grupo y se estima el modelo, obteniendo el siguiente resultado:*

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 2 \\ 1000 \\ 700 \\ 10 \end{pmatrix}, \quad \mathbf{e}'\mathbf{e} = 100, \quad (\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.1 & 0.40 & -0.1 & -0.10 \\ & 0.01 & -0.3 & 0.40 \\ & & 0.1 & 0.20 \\ & & & 0.02 \end{pmatrix}$$

Para un nivel de significación del 5% contrastar que el efecto sobre el número de unisdades vendidas es el mismo si se incrementa 1000 euros en publicidad que si se dispone de un establecimineto más.

Problema 10 *Dada la siguiente información:*

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 35.000000 & 193.057941 & 772.251242 \\ 193.057941 & 1067.570520 & 4270.38961 \\ 772.251242 & 4270.38961 & 17081.9886 \end{pmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} 2437.90284 \\ 13469.82610 \\ 53880.66190 \end{pmatrix}, \quad \mathbf{Y}'\mathbf{Y} = 170002.814,$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 11.4104 & 48.6772 & -12.6848 \\ 48.6772 & 548760.1560 & -137188.7850 \\ -12.6848 & -137188.7850 & 34296.9061 \end{pmatrix},$$

justificar si puede afirmarse que el modelo presenta multicolinealidad aproximada. Tomar $\alpha = 0.05$. *(Indicación: Las consecuencias de la multicolinealidad aproximada son: (1) imprecisión de los estimadores, debido a la inflación de la varianza, y (2) falta de significación individual de los estimadores, pero modelo globalmente significativo.)*