



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TAREA 3

EJERCICIOS

Los ejercicios que siguen tienen por objeto proporcionar práctica adicional en el manejo de R (ó S-PLUS) y afianzar algunos conceptos sobre los rudimentos del modelo de regresión lineal. Valen las observaciones hechas en la Tarea 1 sobre el modo de presentar los resultados.

1. Esta pregunta ilustra algunos cosas aprendidas sobre proyecciones, problemas de rango incompleto y estimabilidad.
 - a) Crea una matriz X de orden 5×3 con los siguientes datos:

$$\begin{pmatrix} 5 & 6 & 11 \\ 2 & 8 & 10 \\ 3 & 6 & 9 \\ 2 & 4 & 6 \\ 1 & -2 & -1 \end{pmatrix}$$

Observa que la tercera columna es combinación lineal de las dos primeras.

- b) ¿Existe $(X'X)^{-1}$? ¿Por qué?
- c) ¿Cuáles son los valores propios de $(X'X)$?
- d) Sea $X_{(12)}$ la matriz formada por las dos primeras columnas de X y $X_{(13)}$ la formada por las columnas primera y tercera. ¿Generan las columnas de $X_{(12)}$ y de $X_{(13)}$ el mismo subespacio de R^5 ? ¿Generan las columnas de $X_{(12)}$ el mismo subespacio de R^5 que las columnas de X ?
- e) Computa:

$$M = (X'_{(12)}X_{(12)})^{-1}$$

$$N = (X'_{(13)}X_{(13)})^{-1}$$

$$P = X_{(13)}(X'_{(13)}X_{(13)})^{-1}X'_{(13)}$$

$$Q = X_{(12)}(X'_{(12)}X_{(12)})^{-1}X'_{(12)}$$

¿Qué proyecciones representan estas matrices? Compara P y Q . ¿Qué observas? ¿Es lógico?

- f) ¿Son M y N iguales? Toma M y forma una matriz M_0 de dimensión 3×3 orlándola con una fila y columna de ceros:

$$M_0 = \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix}$$

Realiza el producto:

$$(X'X)M_0(X'X)$$

y compara el resultado con $X'X$. ¿Qué observas? ¿Qué es M_0 respecto a $(X'X)$? Guarda M_0 .

- g) Toma un vector arbitrario \vec{y} de R^5 y computa:

$$\hat{\beta}_{(12)} = (X'_{(12)}X_{(12)})^{-1}X'_{(12)}\vec{y} \quad (1)$$

$$\hat{\beta} = (M_0)X'\vec{y} \quad (2)$$

- h) Calcula ahora:

$$\vec{u} = X_{(12)}\hat{\beta}_{(12)} \quad (3)$$

$$\vec{v} = X\hat{\beta} \quad (4)$$

Compara \vec{u} y \vec{v} . ¿Qué observas?

- i) ¿Es $\hat{\beta}_{(12)}$ único siempre que ajustamos mínimo-cuadráticamente \vec{y} sobre las dos primeras columnas de X ?
- j) ¿Es $\hat{\beta}$ único al ajustar mínimo-cuadráticamente \vec{y} sobre las tres columnas de X ?
- k) Imagina un modelo lineal $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$ con X dada en el apartado 1. Indica cuáles de las siguientes combinaciones lineales de parámetros serían estimables: $(\beta_0 + \beta_1)$, $(\beta_0 + \beta_2)$, $(\beta_1 + \beta_2)$, $(\beta_0 + \beta_1 + \beta_2)$, β_1 .
- l) Proporciona una restricción sobre los valores de los parámetros que convierta $\vec{\beta}$ en estimable. Comprueba, añadiendo dicha restricción a las ecuaciones normales, que el sistema resultante tiene solución única.
2. Se te proponen a continuación diversas situaciones para que expliques qué modelo lineal ajustarías, pronunciándote en particular sobre: i) La pertinencia o no de incluir una ordenada en el origen, β_0 , y ii) La plausibilidad de una especificación lineal.
- a) Las observaciones Y_i son datos de consumo de carburante de un mismo vehículo, medido en N ensayos diferentes. En cada ensayo se han recorrido X_i kilómetros.
- b) Las observaciones Y_i son datos de consumo de energía eléctrica total en una población, N días consecutivos. En cada día se ha registrado la temperatura media X_i .
- c) Las observaciones Y_i son datos de consumo total de la familia i -ésima en una población, durante un cierto periodo. Las observaciones X_i recogen la renta de la familia i -ésima en el periodo anterior.
- d) Las observaciones Y_i son datos de consumo calórico de un animal de tiro. Las X_i recogen el número de kilómetros recorridos y el número de kilogramos arrastrados (en condiciones cuidadosamente controladas: mismo trayecto y mismo carro).

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

1. Cuando se te pregunta si una matriz es singular (o si existe su inversa) has de dar un argumento que justifique tu respuesta. Decir que una matriz es singular “porque el determinante es cero” es tautológico. No obstante, puedes querer asegurarte de algo calculando un determinante. En R tienes una función `det` que lo hace directamente. En S-PLUS 6.0, disponible en `anbot0`, puedes cargar una biblioteca con algunas funciones matriciales que no forman parte de S-PLUS básico mediante

```
library(Matrix)
```

tras de lo cual tendrás a tu disposición una función `det`. (Nota que al utilizar esta librería has de definir las matrices con `Matrix` en lugar de `matrix`.)

2. En R puedes calcular una inversa generalizada mediante la función `ginv` de la biblioteca MASS (has de hacer un `library(MASS)` previo) (en S-PLUS tienes la función `ginverse`). No estaría de más que al menos la primera vez la calcularas por el procedimiento descrito en clase.
3. Los vectores y valores propios pueden ser calculados tanto en R como en S-PLUS mediante la función `eigen`. Observa que una matriz simétrica puede no parecerlo en coma flotante, y la función te puede dar valores propios tales como $1.2345234 + 2.39485-16i$ que a todos los efectos prácticos son reales.
4. Sobre R tienes, además de [9] o la versión un poco más antigua en Biblioteca [8], las notas [6] o su traducción española [7], ambas en <http://etdx01.bs.ehu.es>.

Referencias

- [1] R.A. Becker, J.M. Chambers, and A.R. Wilks. *The New S Language. A Programming Environment for Data Analysis and Graphics*. Wadsworth & Brooks/Cole, Pacific Grove, California, 1988.
- [2] J.M. Chambers and T.J. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, Ca., 1992.
- [3] J.H. Grafe. *Matemáticas Universitarias*. MacGraw-Hill, Madrid, 1985.
- [4] G.A.F. Seber. *Linear Regression Analysis*. Wiley, New York, 1977.
- [5] A. Fdez. Trocóniz. *Modelos Lineales*. Serv. Editorial UPV/EHU, Bilbao, 1987.
- [6] B. Venables, D. Smith, R. Gentleman, and R. Ihaka. *Notes on R: A Programming Environment for Data Analysis and Graphics*. Dept. of Statistics, University of Adelaide and University of Auckland, 1997. Available at <http://cran.at.r-project.org/doc/R-intro.pdf>.
- [7] B. Venables, D. Smith, R. Gentleman, R. Ihaka, and M. Mächler. *Notas sobre R: Un entorno de programación para análisis de datos y gráficos*, 2000. Traducción española de A. González y S. González.
- [8] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-PLUS*. Springer-Verlag, New York, 1994. Signatura: 681.03.068 VEN.
- [9] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-PLUS*. Springer-Verlag, New York, third edition, 1999.