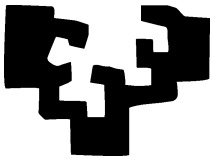
eman ta zabal zazu



Universidad del País Vasco
Euskal Herriko Unibertsitatea

Programa de la asignatura

Estadística: Modelos Lineales (15765)

Curso 2007–2008

Profesor: Fernando TUSELL

Dpto. Economía Aplicada III (Econometría y Estadística)
Facultad de CC.EE. y Empresariales

Descripción

Objetivos de la asignatura. Proporcionar una base teórica y práctica que faculte al alumno(a) para hacer un uso productivo de los modelos estadísticos lineales de regresión y análisis de varianza. Si el tiempo lo permite, se hace también una breve introducción a la regresión no paramétrica.

Prerrequisitos. Un curso introductorio de Estadística, al nivel, por ejemplo, de *Estadística para Economistas*. Es útil tener alguna competencia informática, pero no imprescindible: todo lo que es preciso aprender para realizar las tareas se enseña a lo largo del curso.

Orientación bibliográfica. Además de los libros citados como bibliografía de cada capítulo, hay muy buenos manuales de uso general como Draper and Smith (1998) y Stapleton (1995). Una obra enciclopédica es Trocóniz (1987), con multitud de detalles sobre casi cada aspecto del modelo de regresión y ANOVA. Puede también utilizarse Núñez and Tusell (2003), pero la disponibilidad de notas del curso no debiera disuadir de consultar la bibliografía. Son una ayuda, no un sustituto.

Salvo que se indique lo contrario, todos los libros y artículos están disponibles en Biblioteca. De algunos se incluye la signatura topográfica para facilitar la búsqueda.

Evaluación y desarrollo del curso. En un curso normal se realizan entre diez y once tareas semanales o decenales, que se corrigen en todo o en parte (dependiendo del número de alumnos matriculados) y se devuelven y comentan en clase. Hay además un examen final. La nota es un promedio de todo ello.

Las prácticas se realizan con R. Los alumnos disponiendo de ordenadores personales reciben, si lo desean, una copia de R y algún software adicional.

Tutoría presencial. Aunque se anima a los alumnos a plantear sus dudas en clase y se destina tiempo a ello, el profesor está disponible en su despacho (2-12, Edificio Despachos) en el horario que se publica para cada trimestre en el tablón de anuncios. Los alumnos disponen además de tutoría electrónica, basada en el sistema E-KASI.

Actualizaciones. La versión más moderna de este programa, de las notas empleadas en el curso, de los ficheros de datos, de algún software de libre uso (como R) y de los enunciados de las tareas está disponible en E-KASI.

Bilbao, 11 de julio de 2007

Temario

I. REGRESIÓN LINEAL

1. LA ESTIMACIÓN MÍNIMO CUADRÁTICA COMO PROBLEMA GEOMÉTRICO.

   Nociones previas: espacios vectoriales, repaso de álgebra matricial, proyección ortogonal. Propiedades de las proyecciones. Planteamiento geómetrico del problema de estimación mínimo cuadrática. Ecuaciones normales. Obtención de los estimadores en el caso matriz de diseño de rango completo. Inversas generalizadas. Obtención de estimadores mínimo cuadráticos en el caso de rango deficiente.

   Clase práctica: El programa R.

   BIBLIOGRAFÍA: Seber (1977) Sec. 3.1. Becker et~al. (1988), Chambers and Hastie (1992). También pueden utilizarse Venables and Ripley (1999) (excelente), A.Krause and M.Olson (1997) (nivel más introductorio) y Fox (2002). Específicamente sobre R: Venables et~al. (1997); puede encontrarse en `http://cran.r-project.org/`, junto a otra mucha documentación.

2. PROPIEDADES DE LOS ESTIMADORES MÍNIMO CUADRÁTICOS.

   Insesgadez. Consistencia. Eficiencia en la clase de los insesgados: teorema de Gauss-Markov. Estimación de la varianza de la perturbación. Descomposición de la suma de cuadrados. Coeficiente de correlación múltiple. Interpretación geométrica.

   Clase práctica: el editor `emacs`. Utilización interactiva y batch de R.

   BIBLIOGRAFÍA: Seber (1977) Sec. 3.2.

3. ESTIMACIÓN CONDICIONAL.

   Algunos lemas adicionales sobre proyecciones. La estimación condicionada como problema geométrico. Estimación mínimo cuadrática bajo restricciones lineales $A\vec{\beta} = \vec{c}$.

   Clase práctica: Utilización de R bajo `emacs`.

   BIBLIOGRAFÍA: Seber and Lee (1998) Sec. 3.8.

4. REGRESIÓN LINEAL CON PERTURBACIÓN ALEATORIA NORMAL (I).

Algunos teoremas previos sobre independencia de diversas formas lineales y cuadráticas. Independencia entre S^2 y el vector de estimadores de los parámetros.

BIBLIOGRAFÍA: Seber and Lee (1998) Cap. 2 y Sec. 3.3.

5. REGRESIÓN LINEAL CON PERTURBACIÓN ALEATORIA NORMAL (II).

   Distribución de diversos estadísticos en el muestreo. Contrastes de hipótesis: sobre parámetros aislados, sobre el vector de estimadores completo, sobre parte del vector de estimadores.

   BIBLIOGRAFÍA: Seber and Lee (1998) Sec. 3.4-3.5 y Sec. 4.1 a 4.3, Myers (1990) Sec. 3.1 a 3.5.

6. REGRESIÓN LINEAL CON PERTURBACIÓN ALEATORIA NORMAL (III).

   Contraste de hipótesis lineales generales. Contrastes cuando la matriz de diseño es de rango deficiente. Funciones estimables. Hipótesis contrastables. Contraste de hipótesis simultáneas: problemas que plantea. Métodos de Bonferroni, rango Studentizado, y de Scheffé*.

   BIBLIOGRAFÍA: Seber and Lee (1998) Sec. 4.3 y 5.1.

7. MULTICOLINEALIDAD.

   Interpretación gráfica. Correlación múltiple y parcial. Formas de detectar multicolinealidad en la matriz de diseño. Efecto sobre el vector de estimadores y sobre diferentes funciones lineales de los parámetros. Forma de tomar muestras adicionales óptimas para corregir la multicolinealidad en la matriz de diseño.

   BIBLIOGRAFÍA: Seber and Lee (1998) Sec. 3.9 y 9.7

8. REGRESIÓN SESGADA.

   Motivación. Regresión *ridge*: compromiso varianza-sesgo. Existencia de una solución que domina a la MCO en términos de ECM. Regresión en componentes principales. Motivación. Aplicaciones. Regresión en variables latentes*.

   BIBLIOGRAFÍA: Seber and Lee (1998) Sec. 12.5.2 y Hoerl and Kennard (1970).

9. ANÁLISIS DE RESIDUOS.

   Tipos de residuos: MCO, internamente studentizados, externamente studentizados (o "studentizados borrados"). Distribuciones de los mismos. Gráficos de residuos como ayuda en el diagnóstico. Contrastes de presencia de *outliers*. Residuos BLUS*. Ejemplos.

   BIBLIOGRAFÍA: Theil (1971) pág. 205-206 (residuos BLUS). Myers (1990), Cap. 5, Seber and Lee (1998) Sec. 10.2.

10. ANÁLISIS DE INFLUENCIA.

    Motivación. Nociones sobre robustez e influencia. Análisis de influencia en el modelo lineal: curvas de influencia empírica (EIC) y de influencia muestral (SIC). Distancia de Cook. Factores de incremento de varianza (VIF).

    BIBLIOGRAFÍA: Cook and Weisberg (1982), *passim.*

11. SELECCIÓN DE MODELOS DE REGRESIÓN (I).

    Modelos escasos y sobreparametrizados. Efectos sobre varianza y sesgo de los estimadores. Criterios de comparación: coeficiente de correlación múltiple, corregido o no. C_p de Mallows. Suma de cuadrados predictiva y validación cruzada*. Comparación.

    BIBLIOGRAFÍA: Myers (1990), Cap. 4, Seber and Lee (1998) Cap. 12, parte.

12. SELECCIÓN DE MODELOS DE REGRESIÓN (II).

    Métodos de selección de variables: Fuerza bruta (prueba de todos los subconjuntos), regresión escalonada hacia adelante (forward) y hacia atrás (backwards). Detalles sobre la implementación en R. Simultaneidad y niveles de significación de entrada y salida de variables. Ejemplos.

    BIBLIOGRAFÍA: Myers (1990), Cap. 4; Seber and Lee (1998), Cap. 12, parte.

13. EFECTOS DEL INCUMPLIMIENTO DE LAS HIPÓTESIS.

    Regresores estocásticos. Perturbaciones no homoscedásticas. Perturbaciones no normales. Perturbaciones autocorreladas. Una discusión muy somera, remitiendo a la asignatura de Econometría en el caso de regresores estocásticos y autocorrelación en las perturbaciones.

    BIBLIOGRAFÍA: Seber and Lee (1998), Cap. 9 y 10, parte.

II. EXTENSIONES DEL MODELO

14. TRANSFORMACIONES. REGRESIÓN CON VARIABLE CUALITATIVA

    Familia de transformaciones de Box-Tidwell*. Gráficos de residuos-variables y su empleo en la elección de una transformación. Transformaciones de la variable respuesta: Box-Cox*.

    BIBLIOGRAFÍA: Seber and Lee (1998), Sec. 10.5.2.

15. REGRESIÓN LOGÍSTICA

Introducción al modelo logístico. Motivación, estimación, contraste de hipótesis. Generalizaciones. Modelos lineales generalizados (GLM).

BIBLIOGRAFÍA: Hosmer and Lemeshow (1989), Cap. 4; Kleinbaum (1994).

16. INTRODUCCIÓN A LA REGRESIÓN NO PARAMÉTRICA

Motivación. Estimadores *kernel*. Estimación pro *backfitting*. Modelos aditivos.

BIBLIOGRAFÍA: Hastie and Tibshirani (1991), Cap. 3 y 4. Eubank (1988), Cap. 5 (parte).

III. ANÁLISIS DE VARIANZA

17. ANÁLISIS DE VARIANZA (I).

Planteamiento del problema. Supuestos. El modelo de análisis de varianza equilibrado de un tratamiento. El modelo de análisis de varianza equilibrado de dos tratamientos aditivos. Ejemplos.

BIBLIOGRAFÍA: Seber (1977), Sec. 9.1 y 9.2.

18. ANÁLISIS DE VARIANZA (II).

Investigación de interacciones. Modelos completos, cuadrados latinos y grecolatinos. Bloques aleatorizados. Modelos anidados. Relación del modelo de análisis de varianza y el modelo de regresión lineal.

Seber (1977), Sec. 9.2 y 10.1.

Calendario previsto

2007

MARTES	JUEVES	VIERNES
Sep 25 **1** Repaso Algebra Lineal y Matricial *Entrega Tarea 1*	27 **2** El modelo lineal: introducción.	28 **3** Proyecciones. Sesión introductoria R.
Oct 2 **4** Proyecciones. Obtención $\hat{\beta}$. *Entrega Tarea 2*	4 **5** Propiedades estimadores **Vence Tarea 1**	5 **6** Estimación σ^2_ϵ. Descomp. suma cuadrados. Uso `lsfit` en R.
9 **7** Interpretación geométrica. R^2. *Entrega Tarea 3*	11th **8** Estimación condicional. **Vence Tarea 2**	12th EL PILAR
16th **9** Estimación condicional. R bajo `emacs`.	18th **10** Regresión lineal con normalidad. *Entrega Tarea 4*	19th **11** Distribución de estadísticos. **Vence Tarea 3**
23 **12** Contraste de hipótesis	25 **13** Contraste de hipótesis e intervalos de confianza, *Entrega Tarea 5*	26 **14** Inferencia simultánea. **Vence Tarea 4**
30 **15** Inferencia simultánea.	Nov 1 TODOS LOS SANTOS	2 **16** Multicolinealidad. Diagnóstico.

MARTES	JUEVES	VIERNES
6 **17** Regresión sesgada. *Entrega Tarea 6* **Vence Tarea 5**	8 **18** Regresión sesgada	9 **19** Residuos *Entrega Tarea 7*
13th **20** Diagnósticos. Influencia **Vence Tarea 6**	15th **21** Especificación de modelos. Criterios (R^2, $\overline{R}^2$, C_p, AIC...)	16th **22** Especificación de modelos. Estrategias. *Entrega Tarea 8*
20 **23** Incumplimiento de las hipótesis **Vence Tarea 7**	22 **24** Transformaciones: Box-Cox, Box-Tidwell, regresión polinómica, etc.	23 **25** Regresión logística *Entrega Tarea 9*
27 **26** Regresión logística **Vence Tarea 8**	29 **27** Modelos lineales generalizados (GLM).	30 **28** Regresión no lineal
Dic 4 **29** Regresión no paramétrica: *kernels* y otros suavizadores. *Entrega Tarea 10*	6 DÍA CONSTITUCIÓN	7 FIESTA FACULTAD
11th **30** Regresión no paramétrica **Vence Tarea 9**	13th **31** Regresión semiparamétrica	14th **32** Regresión semiparamétrica. Modelos aditivos

MARTES	JUEVES	VIERNES
18th **33** Análisis de Varianza. Modelo un tratamiento. Introducción. *Entrega Tarea 11*	20 **34** Análisis de Varianza. Modelo un tratamiento. Contrastes.	21 **35** Análisis de Varianza. Modelo aditivo con dos tratamientos cruzados.

2008

MARTES	JUEVES	VIERNES
Ene 8 **36** Análisis de Varianza. Modelos con dos y más tratamientos, aditivos y con interacción. **Vence Tarea 10**	10th **37** Modelos incompletos. Cuadrado latino y grecolatino.	11th **38** Análisis de Varianza. Modelos anidados.
15th **39** Análisis de Varianza. Otros diseños. Nociones diseño experimental.	17th **40** Análisis de Varianza. Otros diseños.	18th **41** Dudas, preguntas. **Vence Tarea 11**.

Bibliografía

A.Krause and M.Olson (1997). *The Basics of* S *and* S-PLUS. Springer Verlag, Signatura: 519.682 KRA.

Becker, R., Chambers, J., and Wilks, A. (1988). *The New S Language. A Programming Environment for Data Analysis and Graphics*. Pacific Grove, California: Wadsworth & Brooks/Cole.

Chambers, J. and Hastie, T. (1992). *Statistical Models in S*. Pacific Grove, Ca.: Wadsworth & Brooks/Cole.

Cook, R. and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.

Draper, N. and Smith, H. (1998). *Applied Regression Analysis*. Wiley, third edition, Signatura: 519.233.5 DRA.

Eubank, R. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

Fox, J. (2002). *An R and S-Plus companion to applied regression*. Sage Pub.

Hastie, T. and Tibshirani, R. (1991). *Generalized Additive Models*. London: Chapman & Hall, second edition.

Hoerl, A. and Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 55–67.

Hosmer, D. and Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley.

Kleinbaum, D. (1994). *Logistic Regression. A Self-Learning Test*. Springer Verlag.

Myers, R. (1990). *Classical and Modern Regression with Applications*. Boston: PWS-KENT Pub. Co.

Núñez, V. and Tusell, F. (2003). Regresión Lineal y Análisis de Varianza. Notas de clase.

Peña, D. (2002). *Regresión y diseño de experimentos*. Alianza Editorial.

Seber, G. (1977). *Linear Regression Analysis*. New York: Wiley.

Seber, G. and Lee, A. (1998). *Linear Regression Analysis*. Wiley.

Stapleton, J. (1995). *Linear Statistical Models*. New York: Wiley.

Theil, H. (1971). *Principles of Econometrics*. New York: Wiley.

Trocóniz, A. F. (1987). *Modelos Lineales*. Bilbao: Serv. Editorial UPV/EHU.

Venables, B., Smith, D., Gentleman, R., and Ihaka, R. (1997). *Notes on R: A Programming Environment for Data Analysis and Graphics*. Dept. of Statistics, University of Adelaide and University of Auckland, Available at `http://cran.at.r-project.org/doc/R-intro.pdf`.

Venables, W. and Ripley, B. (1999). *Modern Applied Statistics with* S-PLUS. New York: Springer-Verlag, third edition.