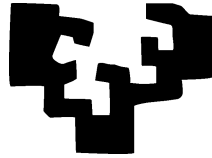


eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

Programa de la asignatura

# Estadística: Modelos Lineales (15765)

Curso 2011–2012

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 preciso 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 & Smith (1998), Seber (1977) (hay una segunda edición, Seber & Lee (1998), pero para algunas cosas utilizaremos la primera) y Stapleton (1995). Una obra enciclopédica es Trocóniz (1987), con multitud de detalles sobre casi cada aspecto del modelo de regresión y ANOVA. Pueden también utilizarse las notas del curso que no debieran disuadir de consultar la bibliografía. Son una ayuda, no un sustituto.

Como el curso está basado en R, será también de utilidad Faraway (2005), que muestra el modo de hacer ciertos cálculos; pero debe complementarse con un texto más detallado en cuanto a la teoría.

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. Se ofrece a los alumnos espacio de trabajo en el Laboratorio de Economía Cuantitativa (LEC), que proporciona lo necesario para realizar todo el trabajo del curso.

El programa está sobredimensionado; algunos temas no llegan a verse, dependiendo del número y disposición de días festivos en cada curso concreto y de la preparación de base de cada promoción.

**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, y una o dos tardes a la semana en el LEC. Los alumnos disponen además de tutoría electrónica, basada en MOODLE (<http://moodle3.ehu.es>).

**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 <http://www.et.bs.ehu.es/~etptupaf> y en MOODLE.

Bilbao, 27 de septiembre de 2011

## Temario

### I. REGRESIÓN LINEAL

#### 1. REPASO DE ALGEBRA LINEAL Y MATRICIAL

Repaso de conceptos que deben conocerse de cursos precedentes. Matrices, aplicaciones lineales, sistemas de ecuaciones lineales. Vectores y valores propios, diagonalización, tipos de matrices simétricas (definidas positivas, etc.) y propiedades de las mismas.

Clase práctica: El programa R.

BIBLIOGRAFÍA: Dalgaard (2002), Ugarte et al. (2008), Chambers & Hastie (1992). También pueden utilizarse Venables & Ripley (1999) (excelente), Krause & Olson (1997) (nivel más introductorio) y Fox (2002). Específicamente sobre R: Venables et al. (1997) y Spector (2008).

Sobre álgebra lineal y matricial pueden ser de utilidad Abadir & Magnus (2005), Seber (2007), Gentle (2007) o Searle (1982), además de los apéndices a libros como Seber (1977), Searle (1971) y Peña (2002) y el resumen Petersen & Pedersen (2008).

#### 2. LA ESTIMACIÓN MÍNIMO CUADRÁTICA COMO PROBLEMA GEOMÉTRICO.

Nociones previas: espacios vectoriales, repaso de álgebra matricial, proyección ortogonal. Propiedades de las proyecciones. Planteamiento geométrico 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.

BIBLIOGRAFÍA: Seber (1977) Sec. 3.1.

#### 3. 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.

#### 4. 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 & Lee (1998) Sec. 3.8.

#### 5. 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 & Lee (1998) Cap. 2 y Sec. 3.3.

#### 6. 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 & Lee (1998) Sec. 3.4-3.5 y Sec. 4.1 a 4.3, Myers (1990) Sec. 3.1 a 3.5.

#### 7. 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 & Lee (1998) Sec. 4.3 y 5.1.

#### 8. 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 & Lee (1998) Sec. 3.9 y 9.7

#### 9. 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 & Lee (1998) Sec. 12.5.2, Faraway (2005) 9.3 y Hoerl & Kennard (1970).

#### 10. 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 & Lee (1998) Sec. 10.2.

#### 11. 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 & Weisberg (1982), *passim*.

#### 12. 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 & Lee (1998) Cap. 12, parte.

#### 13. 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 & Lee (1998), Cap. 12, parte.

#### 14. 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 & Lee (1998), Cap. 9 y 10, parte.

## II. EXTENSIONES DEL MODELO

### 15. 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 & Lee (1998), Sec. 10.5.2, Faraway (2005), Cap. 7.

### 16. REGRESIÓN LOGÍSTICA

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

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

### 17. INTRODUCCIÓN A LA REGRESIÓN NO PARAMÉTRICA

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

BIBLIOGRAFÍA: Hastie & Tibshirani (1991), Cap. 3 y 4.

## III. ANÁLISIS DE VARIANZA

### 18. 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.

### 19. 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

2011

MARTES	JUEVES	VIERNES
<div style="border: 1px solid black; display: inline-block; padding: 2px;">Sep 27</div> <div style="float: right; text-align: right;">1</div> Repaso Algebra Lineal y Matricial.	29 <span style="float: right;">2</span> Repaso Algebra Lineal y Matricial. <i>Entrega Tarea 1</i>	30 <span style="float: right;">3</span> El modelo lineal: introducción. R en el LEC.
<div style="border: 1px solid black; display: inline-block; padding: 2px;">Oct 4</div> <div style="float: right; text-align: right;">4</div> Proyecciones.	6 <span style="float: right;">5</span> Proyecciones. Obtención $\hat{\beta}$ . <i>Entrega Tarea 2</i>	7 <span style="float: right;">6</span> Propiedades estimadores <b>Vence Tarea 1</b>
11th <span style="float: right;">7</span> Estimación $\sigma_\epsilon^2$ . Descomp. suma cuadrados. Uso <code>lsfit</code> en R.	13th <span style="float: right;">8</span> Interpretación geométrica. $R^2$ . <i>Entrega Tarea 3</i>	14th <span style="float: right;">9</span> Estimación condicional. <b>Vence Tarea 2</b>
18th <span style="float: right;">10</span> Estimación condicional. R bajo emacs.	20 <span style="float: right;">11</span> Regresión lineal con normalidad. <i>Entrega Tarea 4</i>	21 <span style="float: right;">12</span> Distribución de estadísticos. <b>Vence Tarea 3</b>
25 EUSKADIKO EGUNA	27 <span style="float: right;">13</span> Distribución de estadísticos	28 <span style="float: right;">14</span> Distribución de estadísticos <b>Vence Tarea 4</b>
<div style="border: 1px solid black; display: inline-block; padding: 2px;">Nov 1</div> <div style="float: right; text-align: right;">15</div> TODOS LOS SANTOS	3 <span style="float: right;">15</span> Contraste de hipótesis	4 <span style="float: right;">16</span> Contraste de hipótesis



MARTES	JUEVES	VIERNES
8 <b>17</b> Contraste de hipótesis e intervalos de confianza, <i>Entrega Tarea 5</i>	10th <b>18</b> Inferencia simultánea.	11th <b>19</b> Inferencia simultánea.
15th <b>20</b> Multicolinealidad. Diagnóstico.	17th <b>21</b> Regresión sesgada. <i>Entrega Tarea 6</i>	18th <b>22</b> Regresión sesgada <b>Vence Tarea 5</b>
22 <b>23</b> Residuos <i>Entrega Tarea 7</i>	24 <b>24</b> Diagnósticos. Influencia	25 <b>25</b> Especificación de modelos. Criterios ( $R^2$ , $\bar{R}^2$ , $C_p$ , AIC...) <b>Vence Tarea 6</b>
29 <b>26</b> Especificación de modelos. Estrategias. <i>Entrega Tarea 8</i>	Dic 1 <b>27</b> Incumplimiento de las hipótesis	2 <b>28</b> Transformaciones: Box-Cox, Box-Tidwell, regresión polinómica, etc. <b>Vence Tarea 7</b>
6 DÍA CONSTITUCIÓN	8 INMACULADA	9 <b>29</b> Regresión logística <i>Entrega Tarea 9</i>
13th <b>30</b> Regresión logística	15th <b>31</b> Modelos lineales generalizados (GLM).	16th <b>32</b> Regresión no lineal <b>Vence Tarea 8</b>

MARTES	JUEVES	VIERNES
20 Regresión no paramétrica: <i>kernels</i> y otros suavizadores. <i>Entrega Tarea 10</i>	22 Regresión no paramétrica <b>Vence Tarea 9</b>	23 Regresión semiparamétrica
<b>33</b>	<b>34</b>	<b>35</b>

**2012**

LUNES	MIÉRCOLES	VIERNES
Ene 10th Análisis de Varianza. Modelo un tratamiento. Introducción. <i>Entrega Tarea 11</i>	12th Análisis de Varianza. Modelo un tratamiento. Contrastes.	13th Análisis de Varianza. Modelo aditivo con dos tratamientos cruzados.
<b>34</b>	<b>35</b>	<b>36</b>
17th Análisis de Varianza. Modelos con dos y más tratamientos, aditivos y con interacción. <b>Vence Tarea 10</b>	19th Modelos incompletos. Cuadrado latino y grecolatino.	20 Análisis de Varianza. Modelos anidados. Otros diseños. <b>Vence Tarea 11.</b>
<b>37</b>	<b>38</b>	<b>39</b>
24 Análisis de Varianza. Otros diseños.	26 Dudas, preguntas.	27
<b>40</b>	<b>41</b>	<b>42</b>

# Bibliografía

- Abadir, K. & Magnus, J. (2005), *Matrix Algebra*, Cambridge Univ. Press.
- Chambers, J. M. & Hastie, T. J. (1992), *Statistical Models in S*, Wadsworth & Brooks/Cole, Pacific Grove, Ca.
- Cook, R. D. & Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman and Hall, New York.
- Dalgaard, P. (2002), *Introductory Statistics with R*, Statistics and Computing, Springer-Verlag. Signatura: 519.682 DAL.
- Draper, N. R. & Smith, H. (1998), *Applied Regression Analysis*, third edn, Wiley. Signatura: 519.233.5 DRA.
- Faraway, J. J. (2005), *Linear Models with R*, Chapman & Hall/CRC. Signatura: 519.233 FAR.
- Fox, J. (2002), *An R and S-Plus Companion to Applied Regression*, Sage Pub.
- Gentle, J. (2007), *Matrix Algebra: Theory, Computations, and Applications in Statistics*, Springer.
- Hastie, T. J. & Tibshirani, R. J. (1991), *Generalized Additive Models*, 2nd. edn, Chapman & Hall, London.
- Hoerl, A. E. & Kennard, R. W. (1970), ‘Ridge regression: Biased estimation for non-orthogonal problems’, *Technometrics* **12**, 55–67.
- Hosmer, D. W. & Lemeshow, S. (1989), *Applied Logistic Regression*, Wiley.
- Kleinbaum, D. G. (1994), *Logistic Regression. A Self-Learning Test*, Springer Verlag.
- Krause, A. & Olson, M. (1997), *The Basics of S-Plus*, Springer Verlag. Signatura: 519.682 KRA.
- Myers, R. H. (1990), *Classical and Modern Regression with Applications*, PWS-KENT Pub. Co., Boston.

- Peña, D. (2002), *Regresión y Diseño de Experimentos*, Alianza Editorial.
- Petersen, K. B. & Pedersen, M. S. (2008), *The Matrix Cookbook*, Librement disponible en Internet.  
**URL:** <http://matrixcookbook.com>
- Peña, D. (2002), *Análisis de Datos Multivariantes*, McGraw-Hill.
- Searle, S. R. (1971), *Linear Models*, Wiley, New York.
- Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, Wiley.
- Seber, G. (2007), *A Matrix Handbook for Statisticians*, Wiley.
- Seber, G. A. F. (1977), *Linear Regression Analysis*, Wiley, New York.
- Seber, G. A. F. & Lee, A. J. (1998), *Linear Regression Analysis*, Wiley.
- Spector, P. (2008), *Data Manipulation with R*, Springer.
- Stapleton, J. H. (1995), *Linear Statistical Models*, Wiley, New York.
- Theil, H. (1971), *Principles of Econometrics*, Wiley, New York.
- Trocóniz, A. F. (1987), *Modelos Lineales*, Serv. Editorial UPV/EHU, Bilbao.
- Ugarte, M., Militino, A. & Arnholt, A. (2008), *Probability and Statistics with R*, CRC Press.
- Venables, B., Smith, D., Gentleman, R. & Ihaka, R. (1997), *Notes on R: A Programming Environment for Data Analysis and Graphics*, Dept. of Statistics, University of Adelaide and University of Auckland. Librement disponible en Internet.  
**URL:** <http://cran.at.r-project.org/doc/R-intro.pdf>.
- Venables, W. & Ripley, B. (1999), *Modern Applied Statistics with S-Plus*, third edn, Springer-Verlag, New York.