

e-UMAB

MODELOS LINEALES

Francesc Carmona Pontaque



UNIVERSITAT DE BARCELONA



e-UMAB

MODELOS LINEALES

Francesc Carmona Pontaque

Electronic-University Mathematical Books

3

Publicacions i Edicions



UNIVERSITAT DE BARCELONA



Carmona Potanque, Francesc

Modelos lineales . – (Electronic-University Mathematical Books ; 3)

Bibliografia. Índex
ISBN 84-475-2893-6

I. Títol II. Col·lecció
1. Models lineals (Estadística)

Consejo editor:

T. Aluja
M. J. Bayarri
F. Carmona
C. M. Cuadras (coordinador)
F. R. Fernández
J. Fortiana
G. Gómez
W. González-Manteiga
M. J. Greenacre
J. M. Oller
J. Puerto
A. Satorra

e-UMAB
Electronic-University Mathematical Books

© PUBLICACIONS I EDICIONS DE LA UNIVERSITAT DE BARCELONA, 2005
Adolf Florensa, s/n; 08028 Barcelona; Tel. 934 035 442; Fax 934 035 446;
lcuenca@ub.edu; <http://www.publicacions.ub.es>

Copia impresa del libro electrónico con ISBN: 84-475-2894-4

Impresión: Gráficas Rey, S.L.

Depósito legal:

ISBN: 84-475-2893-6

Impreso en España / Printed in Spain

*A la meva esposa Carme
i els nostres fills Mireia i Guillem.*

*"Soñemos con un mundo unido
sin ninguna otra soberanía
que la del pueblo universal.
No hacer daño nunca, nunca, a nadie."*

José María de Llanos (Padre Llanos)

Presentación

La teoría y aplicaciones de los modelos lineales ocupan un papel fundamental en la Estadística. Tales modelos engloban la regresión simple, múltiple y polinómica, el análisis de la varianza, el diseño de experimentos, el estudio de curvas de crecimiento, los modelos log-lineales, y algunos contrastes sobre medias como caso particular. Basta consultar revistas especializadas como *Biometrics*, para comprobar que muchos problemas de estadística aplicada se pueden enfocar linealmente, siguiendo la omnipresente ecuación: Observación = Modelo + Error.

Algunos han creído que por el hecho de ser el modelo “lineal”, su tratamiento era más bien fácil. En realidad es todo lo contrario. Este tipo de modelo, que se adecua tan bien a la naturaleza, exige un estudio riguroso y posee múltiples facetas que por sí sólo constituye una especialidad en Estadística.

La obra de mi compañero y amigo Francesc Carmona, que hace más de veinticinco años fue un destacado alumno mío, nace precisamente de las clases que sobre el mismo tema impartí en la Facultad de Matemáticas de la Universidad de Barcelona, y que él continuó, ampliando y mejorando la materia. Diversos profesores editamos entonces unos apuntes, que luego ampliamos y publicamos dentro de la colección Publicaciones de Bioestadística y Biomatemática, editada por el Departamento de Estadística. Hacía falta convertir estos apuntes en un libro de verdad, una labor que ha sido llevada a cabo con entusiasmo por Francesc Carmona, consiguiendo una visión ampliada, moderna y mejorada del anterior material didáctico.

Me complace enormemente presentar el libro **Modelos Lineales**, editado en la colección e-UMAB de EUB, por estar muy bien escrito y documentado y ser muy completo. En efecto, además de contener los temas clásicos, incluye los modelos no paramétricos, el análisis de residuos, numerosos ejemplos ilustrativos, instrucciones en el lenguaje de programación R y adecuados hipervínculos. Esta obra es una contribución didáctica de alto nivel, que será de gran utilidad para investigadores, profesores y alumnos de Estadística.

Dr. Carles M. Cuadras
ccuadras@ub.edu

Prólogo

Las páginas que siguen constituyen una parte de las exposiciones teóricas y prácticas de asignaturas que se han impartido a lo largo de algunos años en varias licenciaturas y cursos de doctorado. En particular en la licenciatura de Matemáticas, la licenciatura de Biología y la diplomatura de Estadística de la Universidad de Barcelona. Se ha intentado un cierto equilibrio entre las explicaciones teóricas y los problemas prácticos. Sin embargo, nuestra intención siempre ha sido fundamentar sólidamente la utilización de los modelos lineales como base de las aplicaciones de la regresión, el análisis de la varianza y el diseño de experimentos. Por ello, en este libro la base matemática y estadística es considerable y creemos importante la correcta definición de los conceptos y la rigurosidad de las demostraciones. Una sólida base impedirá cometer ciertos errores, habituales cuando se aplican los procedimientos ciegamente.

Por otra parte, la aplicación práctica de los métodos de regresión y análisis de la varianza requiere la manipulación de muchos datos, a veces en gran cantidad, y el cálculo de algunas fórmulas matriciales o simples. Para ello es absolutamente imprescindible la utilización de algún programa de ordenador que nos facilite el trabajo. En una primera instancia es posible utilizar cualquier programa de hojas de cálculo que resulta sumamente didáctico. También se puede utilizar un paquete estadístico que seguramente estará preparado para ofrecer los resultados de cualquier modelo lineal estándar como ocurre con el paquete SPSS. En cambio, en este libro se ha optado por incluir algunos ejemplos con el programa R. Las razones son varias. En primer lugar, se trata de un programa que utiliza el lenguaje S, está orientado a objetos, tiene algunos módulos específicos para los modelos lineales y es programable. R utiliza un lenguaje de instrucciones y al principio puede resultar un poco duro en su aprendizaje, sin embargo superada la primera etapa de adaptación, su utilización abre todo un mundo de posibilidades, no sólo en los modelos lineales, sino en todo cálculo estadístico. Además, la razón más poderosa es que el proyecto R es GNU y, por tanto, de libre distribución. De modo que los estudiantes pueden instalar en su casa el programa R y practicar cuanto quieran sin coste económico alguno. Por otra parte, el paquete S-PLUS es una versión comercial con el mismo conjunto de instrucciones básicas.

El tratamiento de algunos temas tiene su origen en unos apuntes de C.M. Cuadras y Pedro Sánchez Algarra (1996) que amablemente han cedido para su actualización en este libro y a los que agradezco profundamente su colaboración. También es evidente que algunas demostraciones tienen su origen en el clásico libro de Seber [66].

Por último, este libro ha sido escrito mediante el procesador de textos científico $\text{\LaTeX}$ y presentado en formato electrónico. Gracias a ello se puede actualizar con relativa facilidad. Se agradecerá la comunicación de cualquier errata, error o sugerencia.

Barcelona, 6 de mayo de 2004.

Dr. Francesc Carmona
fcarmona@ub.edu

Índice general

1. Las condiciones	15
1.1. Introducción	15
1.2. Un ejemplo	15
1.3. El modelo	17
1.4. El método de los mínimos cuadrados	18
1.5. Las condiciones de Gauss-Markov	20
1.6. Otros tipos de modelos lineales	21
1.7. Algunas preguntas	21
1.8. Ejemplos con R	22
1.9. Ejercicios	24
2. Estimación	27
2.1. Introducción	27
2.2. El modelo lineal	27
2.3. Suposiciones básicas del modelo lineal	29
2.4. Estimación de los parámetros	30
2.5. Estimación de la varianza	35
2.6. Distribuciones de los estimadores	36
2.7. Matriz de diseño reducida	38
2.8. Matrices de diseño de rango no máximo	40
2.8.1. Reducción a un modelo de rango máximo	40
2.8.2. Imposición de restricciones	41
2.9. Ejercicios	41
3. Funciones paramétricas estimables	45
3.1. Introducción	45
3.2. Teorema de Gauss-Markov	46
3.3. Varianza de la estimación y multicolinealidad	50
3.4. Sistemas de funciones paramétricas estimables	51
3.5. Intervalos de confianza	53
3.6. Ejercicios	54

4. Complementos de estimación	59
4.1. Ampliar un modelo con más variables regresoras	59
4.1.1. Una variable extra	59
4.1.2. Una interpretación	61
4.1.3. Más variables	63
4.2. Mínimos cuadrados generalizados	64
4.3. Otros métodos de estimación	66
4.3.1. Estimación sesgada	66
4.3.2. Estimación robusta	67
4.3.3. Más posibilidades	68
4.4. Ejercicios	68
5. Contraste de hipótesis lineales	69
5.1. Hipótesis lineales contrastables	69
5.2. El modelo lineal de la hipótesis	70
5.3. Teorema fundamental del Análisis de la Varianza	73
5.3.1. Un contraste más general	78
5.3.2. Test de la razón de verosimilitud	80
5.4. Cuando el test es significativo	81
5.5. Contraste de hipótesis sobre funciones paramétricas estimables	81
5.6. Elección entre dos modelos lineales	82
5.6.1. Sobre los modelos	82
5.6.2. Contraste de modelos	83
5.7. Ejemplos con R	85
5.8. Ejercicios	86
6. Regresión lineal simple	91
6.1. Estimación de los coeficientes de regresión	91
6.2. Medidas de ajuste	94
6.3. Inferencia sobre los parámetros de regresión	96
6.3.1. Hipótesis sobre la pendiente	96
6.3.2. Hipótesis sobre el punto de intercepción	97
6.3.3. Intervalos de confianza para los parámetros	97
6.3.4. Intervalo para la respuesta media	98
6.3.5. Predicción de nuevas observaciones	98
6.3.6. Región de confianza y intervalos de confianza simultáneos	99
6.4. Regresión pasando por el origen	99
6.5. Correlación	100
6.6. Carácter lineal de la regresión simple	101
6.7. Comparación de rectas	104
6.7.1. Dos rectas	104
6.7.2. Varias rectas	108
6.7.3. Contraste para la igualdad de varianzas	111
6.8. Un ejemplo para la reflexión	112
6.9. Ejemplos con R	115
6.10. Ejercicios	117

7. Una recta resistente	121
7.1. Recta resistente de los tres grupos	121
7.1.1. Formación de los tres grupos	121
7.1.2. Pendiente e intercepción	122
7.1.3. Ajuste de los residuos e iteraciones	123
7.1.4. Mejora del método de ajuste	127
7.2. Métodos que dividen los datos en grupos	127
7.3. Métodos que ofrecen resistencia	129
7.4. Ejercicios	131
8. Regresión lineal múltiple	133
8.1. El modelo	133
8.2. Medidas de ajuste	134
8.3. Inferencia sobre los coeficientes de regresión	136
8.4. Coeficientes de regresión estandarizados	141
8.5. Multicolinealidad	144
8.6. Regresión polinómica	145
8.6.1. Polinomios ortogonales	147
8.6.2. Elección del grado	148
8.7. Comparación de curvas experimentales	150
8.7.1. Comparación global	150
8.7.2. Test de paralelismo	151
8.8. Ejemplos con R	152
8.9. Ejercicios	156
9. Diagnóstico del modelo	161
9.1. Residuos	161
9.1.1. Estandarización interna	161
9.1.2. Estandarización externa	163
9.1.3. Gráficos	164
9.2. Diagnóstico de la influencia	166
9.2.1. Nivel de un punto	166
9.2.2. Influencia en los coeficientes de regresión	167
9.2.3. Influencia en las predicciones	168
9.3. Selección de variables	169
9.3.1. Coeficiente de determinación ajustado	169
9.3.2. Criterio C_p de Mallows	169
9.3.3. Selección paso a paso	170
9.4. Ejemplos con R	170
9.5. Ejercicios	172
10. Regresión robusta	175
10.1. Minimizar una función objetivo	175
10.1.1. Funciones objetivo	176
10.2. Regresión robusta mínimo-cuadrada recortada	178

10.3. Ejemplos con S-PLUS	179
10.4. Ejercicios	183
11. Análisis de la Varianza	185
11.1. Introducción	185
11.2. Diseño de un factor	186
11.2.1. Comparación de medias	186
11.2.2. Un modelo equivalente	189
11.3. Diseño de dos factores sin interacción	192
11.4. Diseño de dos factores con interacción	198
11.5. Descomposición ortogonal de la variabilidad	203
11.5.1. Descomposición de la variabilidad en algunos diseños	205
11.5.2. Estimación de parámetros y cálculo del residuo	207
11.6. Diagnóstico del modelo	209
11.7. Diseños no balanceados y observaciones faltantes	212
11.8. Ejemplos con R	213
11.9. Ejercicios	219
12. Análisis de Componentes de la Varianza	223
12.1. Introducción	223
12.2. Contraste de hipótesis	224
12.2.1. Los test F	226
12.2.2. Estimación de los componentes de la varianza	227
12.3. Los modelos más sencillos	228
12.3.1. Diseño de un factor con efectos fijos	228
12.3.2. Diseño de un factor con efectos aleatorios	231
12.3.3. Diseño de dos factores sin interacción con efectos fijos	235
12.3.4. Diseño de dos factores sin interacción con efectos aleatorios	238
12.3.5. Diseño de dos factores aleatorios con interacción	240
12.3.6. Diseño de tres factores aleatorios y réplicas	241
12.3.7. Diseño anidado de dos factores aleatorios	242
12.3.8. Resumen	244
12.4. Correlación intraclásica	245
12.5. Ejemplos con R	246
12.6. Ejercicios	247
A. Matrices	251
A.1. Inversa generalizada	251
A.2. Derivación matricial	252
A.3. Matrices idempotentes	252
A.4. Matrices mal condicionadas	253
B. Proyecciones ortogonales	255
B.1. Descomposición ortogonal de vectores	255
B.2. Proyecciones en subespacios	257

C. Estadística multivariante	259
C.1. Esperanza, varianza y covarianza	259
C.2. Normal multivariante	260
Bibliografía	261
Índice alfabético	265