Introduction to Regression Analysis

M.A. Golberg & H.A. Cho
Department of Mathematical Sciences
University of Nevada
Las Vegas, USA
Introduction to Regression Analysis

M.A. Golberg & H.A. Cho

Published by

WIT Press
Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK
Tel: 44 (0) 238 029 3223; Fax: 44 (0) 238 029 2853
E-Mail: witpress@witpress.com
http://www.witpress.com

For USA, Canada and Mexico

WIT Press
25 Bridge Street, Billerica, MA 01821, USA
Tel: 978 667 5841; Fax: 978 667 7582
E-Mail: infousa@witpress.com
http://www.witpress.com

British Library Cataloguing-in-Publication Data

A Catalogue record for this book is available from the British Library

ISBN: 978-1-85312-624-6

Library of Congress Catalog Card Number: 98-86352

The texts of the papers in this volume were set individually by the authors or under their supervision. Only minor corrections to the text may have been carried out by the publisher.

No responsibility is assumed by the Publisher, the Editors and Authors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Revised and updated 2010
Reprinted 2012 by Lightning Source, UK.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Publisher.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>ix</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 A Brief History of Regression</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Genealogy of Regression</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 The Method of Least Squares</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Typical Applications of Regression Analysis</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Use of Regression Analysis</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Data Sets</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Computer Usage</td>
<td>3</td>
</tr>
<tr>
<td>2 Some Basic Results in Probability and Statistics</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Probability Spaces</td>
<td>5</td>
</tr>
<tr>
<td>2.3 Random Variables</td>
<td>7</td>
</tr>
<tr>
<td>2.4 The Probability Distribution of X</td>
<td>7</td>
</tr>
<tr>
<td>2.5 Some Random Variables and their Distributions</td>
<td>8</td>
</tr>
<tr>
<td>2.6 Joint Probability Distributions</td>
<td>12</td>
</tr>
<tr>
<td>2.7 Expectation</td>
<td>15</td>
</tr>
<tr>
<td>2.7.1 Moments</td>
<td>16</td>
</tr>
<tr>
<td>2.7.2 Moment Generating Function</td>
<td>18</td>
</tr>
<tr>
<td>2.8 The Normal and Related Random Variables</td>
<td>21</td>
</tr>
<tr>
<td>2.8.1 Normal Random Variables</td>
<td>21</td>
</tr>
<tr>
<td>2.8.2 Chi-Square Random Variables</td>
<td>22</td>
</tr>
<tr>
<td>2.8.3 t and F-Distributions</td>
<td>23</td>
</tr>
<tr>
<td>2.8.4 Lognormal Random Variables</td>
<td>25</td>
</tr>
<tr>
<td>2.9 Statistical Estimation</td>
<td>25</td>
</tr>
<tr>
<td>2.9.1 The Method of Moments</td>
<td>25</td>
</tr>
<tr>
<td>2.9.2 Maximum Likelihood Estimation</td>
<td>28</td>
</tr>
<tr>
<td>2.9.3 Least Squares Estimation</td>
<td>31</td>
</tr>
<tr>
<td>2.9.4 Bayesian Estimation</td>
<td>31</td>
</tr>
<tr>
<td>2.10 Properties of Estimators</td>
<td>33</td>
</tr>
<tr>
<td>2.10.1 Consistency and Unbiasedness</td>
<td>33</td>
</tr>
<tr>
<td>2.10.2 Sufficiency</td>
<td>35</td>
</tr>
<tr>
<td>2.11 Confidence Intervals</td>
<td>37</td>
</tr>
<tr>
<td>2.11.1 Exact Confidence Intervals</td>
<td>37</td>
</tr>
</tbody>
</table>
CONTENTS

2.11.2 Approximate Confidence Intervals ........................................... 38
2.12 Hypothesis Testing ........................................................................ 39
  2.12.1 Best Tests ........................................................................... 40
  2.12.2 Generalized Likelihood Ratio Tests ......................................... 43
2.13 Hypothesis Testing and Confidence Intervals .............................. 44
2.14 Exercises ..................................................................................... 45

3 Simple Linear Regression ..................................................................... 51
  3.1 Introduction .................................................................................. 51
  3.2 The Error Model .......................................................................... 52
    3.2.1 Algebraic Derivation of the Least Squares Estimators ............... 56
  3.3 Estimating $\sigma^2$ ........................................................................ 65
  3.4 Properties of $\left(\hat{\beta}_0, \hat{\beta}_1, s^2\right)$ .................................. 68
    3.4.1 Standard Errors of the Coefficients ........................................ 72
  3.5 The Gauss-Markov Theorem .......................................................... 73
  3.6 Confidence Intervals for $\left(\beta_0, \beta_1\right)$ ................................... 76
    3.6.1 Simultaneous Confidence Intervals .......................................... 78
  3.7 Hypothesis Tests for $\left(\beta_0, \beta_1\right)$ .......................................... 80
  3.8 The ANOVA Approach to Testing ................................................. 82
    3.8.1 Regression Through the Origin ............................................... 91
    3.8.2 Estimation and Testing for Regression through the Origin .......... 91
    3.8.3 Prediction ............................................................................ 95
  3.9 Assessing Model Validity .............................................................. 101
    3.9.1 The Lack of Fit Test (LOFT) .................................................. 104
    3.9.2 Residual Plots ....................................................................... 111
  3.10 Transformations .......................................................................... 113
    3.10.1 Transformations of $x$ ............................................................ 117
    3.10.2 Transformations in $x$ and $y$ ................................................ 118
    3.10.3 Box-Cox Transformations ....................................................... 119
  3.11 Exercises .................................................................................... 123

4 Random Vectors and Matrix Algebra .................................................. 129
  4.1 Introduction ................................................................................... 129
  4.2 Matrices and Vectors .................................................................... 129
    4.2.1 Some Special Matrices ............................................................ 130
  4.3 Fundamentals of Matrix Algebra .................................................... 131
    4.3.1 Matrix Addition ...................................................................... 131
    4.3.2 Properties of Matrix Addition ............................................... 131
    4.3.3 Matrix Multiplication ............................................................... 132
    4.3.4 Properties of Matrix Multiplication ......................................... 133
    4.3.5 Scalar Multiplication ................................................................. 134
    4.3.6 Powers of Matrices ................................................................. 134
    4.3.7 Matrix Trace ......................................................................... 134
  4.4 Matrices and Linear Transformations .......................................... 135
    4.4.1 Matrix Inversion ..................................................................... 135
    4.4.2 The Inverse of Partitioned Matrices ........................................ 139
    4.4.3 The Sherman-Morrison-Woodbury Formula .............................. 141
6 Residuals, Diagnostics and Transformations 249
6.1 Introduction ................................... 249
6.2 Residuals .................................... 250
6.2.1 Properties of $\hat{\epsilon}$ ............... 250
6.2.2 The Leverage $h_{ii}$ ..................... 251
6.3 Residual Plots ................................ 252
6.3.1 Normal Plots .............................. 252
6.3.2 Variable Plots ............................ 252
6.3.3 Partial Plots .............................. 253
6.4 PRESS Residuals ................................ 265
6.4.1 Deletion Statistics ....................... 266
6.4.2 Influence Diagnostics ................... 271
6.4.3 Influence on $\hat{\beta}$, $\hat{y}$ ............ 271
6.5 Transformations ................................ 276
6.5.1 Transformations in $\mathbf{x}$ ................ 276
6.5.2 The Box-Tidwell Method .................. 277
6.5.3 Transformations of $y$ ..................... 279
6.5.4 Linearizable Transformations .............. 280
6.5.5 Box-Cox Transformations .................. 280
6.5.6 Quick Estimates of $\lambda$ ............... 281
6.5.7 Variance Equalizing Transformations .... 288
6.5.8 Variance Stabilizing Transformations .... 299
6.6 Correlated Errors ................................ 301
6.6.1 The Durbin-Watson Statistic ............. 301
6.6.2 Correcting for Autocorrelation ............ 302
6.7 Generalized Least Squares .................... 305
6.8 Exercises .................................... 308

7 Further Applications of Regression Techniques 313
7.1 Introduction ................................... 313
7.2 Polynomial Models in One Variable ............. 313
7.2.1 Orthogonal Polynomials .................... 315
7.2.2 Piecewise Polynomial Models ............... 317
7.2.3 Multivariate Polynomial Models ......... 322
7.3 Radial Basis Functions ........................ 323
7.3.1 Types of Radial Basis Functions ............ 323
7.3.2 Fitting Methods for RBFs ................. 325
7.4 Dummy Variables ................................ 327
7.4.1 Further Comments on Dummy Variable ..... 334
7.5 Interactions .................................... 337
7.6 Logistic Regression Revisited .................. 341
7.6.1 Interpretation of Logistic Coefficients .... 345
7.6.2 Maximum Likelihood Estimation .......... 347
7.7 The Generalized Linear Model ................. 348
7.7.1 Linear Predictors and Link Functions .... 349
7.7.2 The Error Function ....................... 350
7.7.3 Parameter Estimation ...................... 351
7.8 Exercises .................................... 352
CONTENTS

8 Selection of a Regression Model 359
8.1 Introduction ................................... 359
8.2 Consequences of Model Mispecification ......................... 360
8.3 Criteria Functions ................................ 361
  8.3.1 Coefficient of Multiple Determination $R^2$ .............. 362
  8.3.2 Mallows’ $C_p$ ................................ 363
  8.3.3 The PRESS Statistic ................................ 366
  8.3.4 Standardized Residual Sum of Squares ..................... 367
  8.3.5 Other Criteria .................................... 367
8.4 Various Methods for Model Selection ........................... 367
  8.4.1 Evaluating All Possible Regressions ..................... 368
  8.4.2 Backward Elimination ................................ 369
  8.4.3 Forward Selection ................................ 372
  8.4.4 The Stepwise Regression Procedure ...................... 373
  8.4.5 Selection of Models - An Overview ...................... 375
8.5 Exercises .................................... 376

9 Multicollinearity: Diagnosis and Remedies 379
9.1 Introduction ................................... 379
9.2 Detecting Multicollinearity ................................ 380
9.3 Other Multicollinearity Diagnostics .......................... 384
  9.3.1 Consequences of Multicollinearity .................... 387
  9.3.2 Prediction ................................... 388
9.4 Combatting Multicollinearity ................................ 389
9.5 Biased Estimation ................................ 390
  9.5.1 Shrunken Estimators ............................ 390
  9.5.2 Ridge Regression ............................... 392
  9.5.3 Choosing the Ridge Parameter ....................... 397
  9.5.4 Generalized Ridge Regression ....................... 404
9.6 Other Alternatives to OLS ........................... 407
  9.6.1 Mixed Estimation ............................... 407
  9.6.2 Principal Components Estimation .................... 408
9.7 Exercises .................................... 409

Appendix 413

Bibliography 421

Index 429
To our wives
   Joyce
   Sookhyun

To our children
   Jonathan
   Stefany
   Katherine Soojin
Chapter 1

Introduction

Regression analysis is a collection of statistical techniques that serve as a basis for drawing inferences about relationships among interrelated variables. Since these techniques are applicable in almost every field of study, including the social, physical and biological sciences, business and engineering, regression analysis is now perhaps the most used of all data analysis methods. Hence, the goal of this text is to develop the basic theory of this important statistical method and to illustrate the theory with a variety of examples chosen from economics, demography, engineering and biology. To make the text relatively self contained we have included basic material from statistics, linear algebra and numerical analysis. In addition, in contrast to other books on this topic [27, 87], we have attempted to provide details of the theory rather than just presenting computational and interpretive aspects.

1.1 A Brief History of Regression

1.1.1 Genealogy of Regression

A well-known British anthropologist Sir Francis Galton (1822-1911) seems to be the first to introduce the word “regression” in his study on heredity. He found that on the average, heights of children do not tend toward the parents’ heights, but rather toward the average as compared to the parents. Galton termed this “regression to mediocrity in hereditary stature.” In the Journal of the Anthropological Institute, Vol. 15 (1885), pp. 246-263, it says that “... The experiments showed further that the mean filial regression towards mediocrity was directly proportional to the parental deviation from it.” Galton then described how to determine the relationship between children’s heights using parents’ heights. Today Galton’s analysis would be called a “correlation analysis,” a term for which he is also responsible. In most model-fitting situations today, there are no elements of “regression” in the original sense. Nevertheless, the word is so established that we continue to use it. For more related stories about the history of regression, we refer readers to the Statistical Encyclopedia or The History of Statistics by Stigler (1986) [110].
Chapter 9

Multicollinearity: Diagnosis and Remedies

9.1 Introduction

As we pointed out in Chapter 5 one of the more difficult problems that occurs in doing a regression analysis is that of dealing with multicollinearity. In Example 5.17 we illustrated some of the consequences of this phenomenon; i.e., the difficulty in interpreting the apparently contradictory facts of a good overall fit as indicated by a large value of $R^2$ and a significant observed $F$ ratio, simultaneously with insignificant values for the individual $t$ statistics, while in [76] it was shown how multicollinearity could seriously affect the accuracy of regression calculations. In addition, some of the problems with variable selection techniques as we observed in Chapter 8 are often associated with strong multicollinearity in the data.

In this Chapter we will examine this problem in greater detail. In particular, we will discuss methods for detecting multicollinearity, elaborate on its statistical consequences and examine some of the proposed remedies, particularly the method of ridge regression. This latter technique, which has spawned volumes of research in the past 25 years [28, 58, 59], is still controversial as are other forms of biased estimation [8, 116] and is by no means espoused by all statisticians as the work of Draper and Van Nostrand [28] and Draper and Smith [27] shows. However, because of its prominent role in current research and applications, no modern treatment of regression analysis would be complete without some discussion of its use. At present, it is fair to say that many of the computational problems associated with multicollinearity have been overcome through the development of more sophisticated computational techniques [5, 112] such as the QR decomposition and the advent of powerful computers. While the problem of building and interpreting models with this problem is far from being totally resolved [116] and perhaps never will be.
## Appendix

### Statistical Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table A.1</td>
<td>Standard Normal Distribution</td>
</tr>
<tr>
<td>Table A.2</td>
<td>Percentiles of the Student’s t-Distribution</td>
</tr>
<tr>
<td>Table A.3</td>
<td>Percentiles of the Chi-Square Distribution</td>
</tr>
<tr>
<td>Table A.4a</td>
<td>$F$-Distribution - Critical values of upper 10% points</td>
</tr>
<tr>
<td>Table A.4b</td>
<td>$F$-Distribution - Critical values of upper 5% points</td>
</tr>
<tr>
<td>Table A.4c</td>
<td>$F$-Distribution - Critical values of upper 1% points</td>
</tr>
<tr>
<td>Table A.5</td>
<td>Critical Values of the Durbin-Watson Statistic</td>
</tr>
</tbody>
</table>
Bibliography


Index

Adjusted R-square, 224
Alias, 361
Analysis of covariance model, 182
Analysis of variance, 82
general linear model, 224
  lack of fit analysis, 106
  m-way, 332
  table, 86
two-way, 332
Analysis of variance model, 182
Andrew’s test, 282, 286
Assessment function, 361
Atkinson’s modification, 278
Augmented coefficient matrix, 164
Autocorrelation
  Cochrane-Orcutt method, 302
  coefficient, 301
detecting, 301
  Durbin-Watson statistic, 302
  Hildreth-Lu procedure, 304
Auxillary regressions, 390
Average estimated variance, 367
Backward elimination, 369, 371, 409
Backward substitution, 162
Basis
  orthonormal, 148
Bayes rule, 31
Bayes’ theorem, 6
Best linear unbiased estimator, 55, 217
Best test, 41
Beta function, 24, 32
Beta weights, 195
Bias, 33, 47
Bias matrix, 361
Biased estimator, 389
binary 0-1 predictors, 341
binary 0-1 responses, 341
Binomial coefficient, 9
Bivariate normal distribution, 178
BLUE estimator, 217
Bonferroni inequality, 78
Box-Cox method, 299
Box-Cox transformation, 119
Box-Tidwell method, 118, 277
Canonical form, 196, 350
Cauchy-Schwarz inequality, 143, 144, 302, 310
Centered observations, 191
Centering matrix, 173
Central limit theorem, 22, 25, 41, 76
Characteristic polynomial, 150
Chi-square distribution, 22
Cholesky factorization, 166, 188, 307
Cochrane-Orcutt method, 302
Coefficient matrix, 161
  augmented, 164
Coefficient of determination, 83
  adjusted, 362
Complex conjugate, 150
Condition index, 387
Condition number, 169, 382
  ill-conditioned, 170
  optimally conditioned, 170
  well-conditioned, 170
Conditional Poisson distribution, 46
Confidence interval, 37
  approximate, 38
  exact, 37
  lower limit, 44
  one-sided interval, 44
Consistency, 33
Convergence
  in probability, 33
Cook’s distance, 272, 285
Correlation coefficient, 17
  sample, 83
Correlation form, 195, 214
Covariance, 16
Cramer’s rule, 75, 137, 193, 212
Criteria functions, 361
Cross-validation, 326
Cumulative distribution function, 8
  joint, 12
  properties, 10

Data
  asymmetric, 25
Degrees of freedom, 22, 23
  denominator, 24
  numerator, 24
Deletion statistics, 266
Density, 11
  conditional, 14
  joint discrete, 13
  posterior, 31
Dependent variable, 51
Design matrix, 167, 183, 352
  centered, 191
  centered and scaled, 194
  linearly dependent, 335
Design variable, 51
Determinant, 137
Deviance, 352
DFBETAS, 272
DFFITS, 273
Diagonalization, 149
Distribution
  Bernoulli, 9
  binomial, 9
  chi-square, 22
  double exponential, 55
  F, 24
  Laplace, 55
  logistic, 11
  lognormal, 25
  normal, 21
  of rare events, 9
  Poisson, 9
    conditional, 46
    standard normal, 19
  t, 23
  uniform, 10
Distribution function, 8
Dot product, 143
Double exponential distribution, 55
Dummy variable, 181, 328
  0-1 coding, 335
  analysis of variance model, 182
  levels, 335
  linearly dependent, 335
Durbin-Watson statistic, 301
test procedure, 302
Effective rank, 404
Efficiency, 34
Eigenvalue, 149
Eigenvector, 149
Error
  heteroscedastic, 52
  homoscedastic, 52
Error vector, 183
Errors in variable (EIV) model, 55
Estimable model, 336
Estimation
  Bayesian, 31
  bias, 47
  least squares, 52
  maximum likelihood, 25, 52
Estimator
  Bayes, 33
  best, 36
  best linear unbiased, 55
  bias, 33
  least squares, 31, 54, 74
  linear, 73
  maximum likelihood, 29
  method of moments, 26
  minimum variance unbiased, 34, 52
  minimum variance unbiased linear, 73
  properties
    consistent, 33
    efficient, 34
    sufficient, 35
    unbiased, 27, 33
Euclidean norm, 323
Events, 5
  complement event, 6
  independent events, 7
  mutually exclusive events, 5
Excluded category, 330
Expectation, 15
  linearity properties, 16
INDEX

Exponential family, 12, 46, 350, 358
exponential density, 12
Extra sum of squares principle, 238, 307
  regression sum of squares, 238
Extrapolation
  hidden, 243
F-distribution, 24
F-test, 234
  derivation of the F-test, 236
  full model, 234
  reduced model, 234
Factoring, 165
Factorization theorem, 35
Filter, 405
First order autoregressive, 301
Forward selection, 369, 372
Forward substitution, 162
Full rank model, 188
Fundamental theorem of algebra, 150
Fundamental theorem of calculus, 11
Gamma function, 22
Gauss-Markov theorem, 55, 73, 217, 295
Gauss-Newton method, 327
Gaussian elimination, 137, 162
  augmented coefficient matrix, 164
  backward substitution, 162
  factoring, 165
  forward substitution, 162
  partial pivoting, 163
  pivot element, 163
  round-off error, 163
General linear model, 179
Generalized cross validation, 403
Generalized least squares
  estimator, 306
  weighted mean, 308
Generalized likelihood ratio test, 43, 352
Generalized linear model, 348
  deviance, 352
  generalized likelihood ratio test, 352
  linear predictor, 349
  link function, 349
  weighted least squares, 352
Generalized ridge regression, 404
  estimators, 405
  filter, 405
Generation matrix, 173
Geometric mean, 122
Gram-Schmidt orthogonalization, 147
Gram-Schmidt process, 168, 196
Growth rate, 99
Hat matrix, 207, 250
Heteroscedasticity, 52
Hildreth-Lu procedure, 304
Homoscedasticity, 52
Hypothesis
  composite alternative, 40, 42
  simple, 40
  simple alternative, 40
Hypothesis testing, 39
  alternative hypothesis, 40
  best test, 41
  critical region, 40
  Neyman-Pearson lemma, 41
  null hypothesis, 40
  significance level, 41
  size of Type I error, 40
  Type I and II errors, 40
I Charts, 113
i.i.d. random variable, 15
Ill-conditioning, 167
Independence, 14
Independent variable, 51
Independent variable hull (IVH), 243
Inequality
  Cauchy-Schwarz, 143
  triangular, 144
Influence diagnostics, 271
  Cook's distance, 273
  DFBETAS, 272
  DFFITS, 273
Inner product, 143
  properties, 143
Interaction term, 204
Interactions, 337
  two-way, 337
Interpolation, 325
  noisy data, 323
  non-noisy data, 323
Interval estimation, 37
Iteration, 27, 352
James-Stein estimator, 391
Joint confidence region, 78, 80, 239
Kriging, 301, 323, 325
Lack of fit test, 104
   bias, 104
   lack of fit sum of squares, 105
   pure error sum of squares, 105
Lagrange multipliers, 74, 236, 247
Laplace distribution, 55
Laten root regression, 407
Law of total probability, 6
Least squares, 52
   equations, 187
   estimation, 31
   estimator, 54
   function, 244
   ordinary, 188
   orthogonal, 56
   weighted, 288
Least squares line, 54
   intercept, 54
   residual, 54
   slope, 54
Leverage, 251
   high, 251
   high leverage point, 251
Likelihood function, 29, 120
   log, 29
Likelihood ratio, 41
Likelihood ratio test
   generalized, 43, 352
   likelihood ratio, 41
Linear estimator, 217
Linear predictor, 349
Linear transformations, 135
   linear operator, 136
   linearity, 136
   mapping, 135
Link function, 349
   canonical, 350
   identity, 349
   log-log link, 349
   power family, 350
   probit, 349
Logistic coefficients, 345
Logistic distribution
   cdf of, 342
   Logistic regression, 279
Mallow’s Cp, 363
   properties, 365
Marginal distribution, 13
Matrix, 129
   block form, 139
   coefficient, 161
   columns, 129
   determinant, 137
   diagonal, 130
   diagonalization, 149
   dimension, 130
   elements, 130
      diagonal, 130
      off-diagonal, 130
   full rank, 138
   hat, 207
   idempotent, 134
   identity, 130
   inverse, 136
   non-symmetric, 130
   orthogonal, 148
   orthogonal projection, 135
   partitioned form, 139
   positive definite, 154
   positive semidefinite, 154
   power, 134
   projection, 134
   rank, 138
   rectangular, 130
   rows, 129
   square, 130
   submatrix, 139
   sum, 131
   symmetric, 130
   trace, 135
   transpose, 130
   triangular
      lower, 130
      upper, 130
Matrix addition, 131
   associativity, 131
Matrix multiplication, 132
   associative property, 133
coefficient matrix, 132
commute, 133
conformable, 133
distributive, 133
product, 132
Maximum likelihood estimate, 184
Maximum likelihood estimation, 25, 28, 52
Mean square, 24
Mean square error, 34
Mean value, 16
Median, 25
Method of moments, 26
Minimum absolute deviation line, 127
Minimum variance unbiased estimator, 34, 52, 217
Mixed estimation, 407
Model misspecification, 359
Model selection, 367
all possible regressions, 368
backward selection, 371
forward selection, 372
stepwise regression, 373
Moment generating function, 18
Moments
central, 16
n-th, 16
Multicollinearity, 189, 227, 376
Multiple linear model, 179
analysis of covariance, 182
analysis of variance, 182
design matrix, 183
dummy variable, 181
error random variable, 179
error vector, 183
independent variables, 179
least squares function, 244
linearity property, 181
normal equations, 211, 244
polynomial model of degree m, 181
quadratic model, 181
regression coefficients, 180
vector of observations, 183
vector of regression coefficients, 183
Multivariate normal distribution
degenerate, 160
nondegenerate, 156
Neural network modeling, 323
Neyman-Pearson lemma, 41
Nondecreasing function, 11
Nonlinear estimator, 389
Nonlinear regression, 327
Normal equations, 187
Normal plots, 252
Normal probability plots, 111
Numerical method
iteration, 27
Newton’s, 351
total least squares, 56
Odds ratio, 346
log odds, 346
odds, 346
Ordinary least squares estimate, 188
Orthogonal
basis, 146
hyperplane, 146
projection, 145
vector, 145
Orthogonal design, 196
Orthogonal least squares, 56
Orthogonal matrix, 148
properties, 148
Orthogonal polynomials, 316
Orthonormal, 148
Overfitting, 224
Parallelogram law, 176
Parameter space, 26
Partial F-tests, 235
Partial plots, 253
added variable plots, 253
regression plots, 253
residual plots, 253
Partial regression coefficient, 180
Penalized method, 365
Piecewise polynomials, 277
Pivot elements, 163
Pivotal quantity, 37
Pivoting, 163
Polynomial models, 313, 323
multivariate, 322
orthogonal, 315
piecewise (or spline), 313
Polynomial regression model, 181
Positive definite, 154
Positive semidefinite, 154
Posterior distribution, 31
Posterior mean, 32
Posterior probability, 6
Power family, 120, 350
Power transformation, 299
Powers of matrices, 134
Prediction
  confidence interval, 96
  error sum of squares, 265
  for a new observation, 95
  for the mean response, 95
  interval, 241
  prediction interval, 96
  standard error, 95
  variance, 240
PRESS residuals, 110
Principal components
  estimator, 407, 408
  regression, 407
Principal components form, 196
Prior distribution, 31
Prior probability, 6
Probability, 5
  conditional, 6
  of an event, 5
  posterior, 6
  prior, 6
Probability density function, 8
Probability mass function, 8
Probability space, 5
Probit function, 342
Probit link, 349
Proper value, 174
Pythagorean theorem, 144
QR decomposition, 170, 188, 196, 379
Quadratic form, 156
Qualitative variable, 327
Quantitative variable, 327
R-square
  adjusted, 224
Radial basis functions, 323
  Duchon splines, 324
  Gaussian, 324
  higher-order polyharmonic, 324
  multiquadrics, 324
  shape parameter, 324
  thin-plate spline (TPS), 324
  variance parameter, 324
Random noise, 227
Random sample, 15
Random variable, 7
  Bernoulli, 9
  binomial, 9
  canonical, 11
  continuous random variable, 7
  discrete, 7
  independent, 14
  logistic, 11
  lognormal, 25
  Poisson, 9
  range of, 7
  standard normal, 22
  uniform, 10
RBF approximations, 325
  interpolation, 325
  linear and nonlinear least squares, 325
  smoothing interpolation, 325
Regression coefficients, 180
  partial, 180
Regression function, 322
Regression model
  orthogonality
    canonical form, 196
    principal components form, 196
Regression surface, 183
Regression through the origin, 91
Regularization, 394
Relative risk, 346
Residual, 54
  PRESS, 265
  studentized
    externally, 251
    internally, 251
Residual plots, 297
  added variable plots, 262
  I Chart, 113, 256
  normal plots, 252
  partial plots, 253
  partial regression plots, 262
  variable plots, 252
Residual sum of squares, 186
  standardized, 367
Residuals
externally studentized, 110
outliers, 112
PRESS, 110
properties, 108
standardized, 110
studentized, 110, 112
Residuals plots, 112
histogram, 111
normal probability plots, 111
Response surface methodology, 323
response function, 322
response surface, 323
Response variable, 51
Ridge parameter, 326
Ridge regression, 326, 376, 379, 380
Bayes estimators, 402
constrained estimators, 402
estimate, 393
estimated ridge coefficients, 399
estimators, 392
generalized, 400
harmonic mean estimator, 400
nonstochastic estimators, 404
parameter, 393
PRESS, 402
ridge trace, 398
shrinkage parameter, 400
SRIDG, 401
Ridge trace, 398
Robust regression analysis, 73
Robustness, 38
Round-off error, 163
Sample correlations, 195
Sample proportion, 39
Sample space, 5
Sample variance, 23, 27
Scalar multiplication, 134
Seasonality, 333
Second derivative test, 30
Selecting the best subset, 359
Selection function, 361
Serial correlation, 112
Sherman-Morrison-Woodbury formula, 141
Shrinkage parameter, 400
Shunken estimator, 390
shrinking factors, 391
Significance level, 41
Simple linear regression, 51
error model, 52
Simultaneous confidence intervals, 78
Singular matrix, 185
Singular value decomposition, 167, 188, 380
singular vectors, 169
Smoothing parameter, 326
Spectral theorem, 149, 151, 210
Spline functions, 317
Spline models, 313, 323
Splines, 277
Standard deviation, 16
Standard error, 65, 72
Standard error of prediction, 95
Standard normal distribution, 19
Stepwise regression, 369, 373
Stepwise regression variable selection, 229
Studentized residual
externally, 251
internally, 251
Sufficiency, 35
jointly sufficient, 36
Sufficient statistic, 35
Sum of squares
regression, 84
residual, 84
total, 84
t-distribution, 23
Test
generalized likelihood ratio, 43
likelihood ratio, 41
one-sided hypotheses, 44
two-tailed, 42
uniformly most powerful, 42
Time series analysis, 301
Time series data, 333
Total least squares, 56
Total variance, 214
Trace, 135
Transformations, 113, 276
Atkinson’s modification, 278
Box-Cox method, 119, 280, 281
Box-Tidwell method, 118, 277, 281
in x, 276
intrinsically linear, 118
linearizable, 280
logarithmic, 25
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>of y, 279</td>
<td></td>
</tr>
<tr>
<td>power family, 120</td>
<td></td>
</tr>
<tr>
<td>variance equalizing, 288</td>
<td></td>
</tr>
<tr>
<td>variance stabilizing, 299</td>
<td></td>
</tr>
<tr>
<td>Tukey’s rule, 367</td>
<td></td>
</tr>
<tr>
<td>Two-tailed test, 43</td>
<td></td>
</tr>
<tr>
<td>Unbiased estimator, 23</td>
<td></td>
</tr>
<tr>
<td>Unbiasedness, 33</td>
<td></td>
</tr>
<tr>
<td>Uncorrelatedness, 16</td>
<td></td>
</tr>
<tr>
<td>Uniform measure, 10</td>
<td></td>
</tr>
<tr>
<td>Uniformly most powerful test, 42</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>dependent, 51</td>
<td></td>
</tr>
<tr>
<td>design, 51</td>
<td></td>
</tr>
<tr>
<td>dummy, 328</td>
<td></td>
</tr>
<tr>
<td>independent, 51</td>
<td></td>
</tr>
<tr>
<td>qualitative, 327</td>
<td></td>
</tr>
<tr>
<td>quantitative, 327</td>
<td></td>
</tr>
<tr>
<td>response, 51</td>
<td></td>
</tr>
<tr>
<td>Variable plots, 252</td>
<td></td>
</tr>
<tr>
<td>Variable selection problem, 359</td>
<td></td>
</tr>
<tr>
<td>Variance, 16, 17</td>
<td></td>
</tr>
<tr>
<td>Variance decomposition proportion, 387</td>
<td></td>
</tr>
<tr>
<td>Variance equalizing transformations</td>
<td></td>
</tr>
<tr>
<td>weighted least squares, 288</td>
<td></td>
</tr>
<tr>
<td>Variance inflation factor (VIF), 214</td>
<td></td>
</tr>
<tr>
<td>Variance multiplication factor, 72, 77</td>
<td></td>
</tr>
<tr>
<td>Variance stabilizing transformations, 299</td>
<td></td>
</tr>
<tr>
<td>Variance-covariance matrix, 156</td>
<td></td>
</tr>
<tr>
<td>Vector</td>
<td></td>
</tr>
<tr>
<td>angle, 145</td>
<td></td>
</tr>
<tr>
<td>basis, 146</td>
<td></td>
</tr>
<tr>
<td>canonical base, 146</td>
<td></td>
</tr>
<tr>
<td>column space spanned, 146</td>
<td></td>
</tr>
<tr>
<td>dimension, 146</td>
<td></td>
</tr>
<tr>
<td>dot product, 143</td>
<td></td>
</tr>
<tr>
<td>inner product, 143</td>
<td></td>
</tr>
<tr>
<td>length, 144</td>
<td></td>
</tr>
<tr>
<td>linear combination, 138</td>
<td></td>
</tr>
<tr>
<td>linearly dependent, 138</td>
<td></td>
</tr>
<tr>
<td>linearly independent, 138</td>
<td></td>
</tr>
<tr>
<td>orthogonal, 145</td>
<td></td>
</tr>
<tr>
<td>orthogonal base, 146</td>
<td></td>
</tr>
<tr>
<td>orthonormal, 148</td>
<td></td>
</tr>
<tr>
<td>span, 146</td>
<td></td>
</tr>
<tr>
<td>subspace, 146</td>
<td></td>
</tr>
</tbody>
</table>

- Vector differentiation, 186, 243
- Vector of observations, 183
  - centered, 191
- Vector of regression coefficients, 183
- Weighted least squares, 288
  - equations, 352
  - estimator, 291
  - iteratively reweighted, 294
  - sum of squares, 290
  - weights, 289
- Weighted mean, 308
- Zero intercept model, 94