

Introduction to Linear Regression Analysis^{^ J:, ,1 / , ,, "Z}

Fourth Edition $\approx/\approx ;j. \approx;\!T; \quad \dots \dots \dots! \quad \dots \dots \dots;$

DOUGLAS C. MONTGOMERY $n \quad \prime, \dots$

Arizona State University
Department of Industrial Engineering
Tempe, AZ

ELIZABETH A. PECK

The Coca-Cola Company (retired)
Atlanta, GA

G. GEOFFREY VINING

Virginia Tech
Department of Statistics
Blacksburg, VA

HOCHSCHULE
LIECHTENSTEIN
Bibliothek



A JOHN WILEY & SONS, INC. PUBLICATION

Contents

<

H.

Preface	17
1. Introduction	1
1.1 Regression and Model Building,	1
1.2 Data Collection,	5
1.3 Uses of Regression,	9
1.4 Role of the Computer,	10
2. Simple Linear Regression	12
2.1 Simple Linear Regression Model,	12
2.2 Least-Squares Estimation of the Parameters,	13
2.2.1 Estimation of β_0 and β_1 ,	13
2.2.2 Properties of the Least-Squares Estimators and the Fitted Regression Model,	17
2.2.3 Estimation of tr^2 ,	20
2.2.4 Alternate Form of the Model,	21
2.3 Hypothesis Testing on the Slope and Intercept,	22
2.3.1 Use of t Tests,	22
2.3.2 Testing Significance of Regression,	23
2.3.3 Analysis of Variance,	25
2.4 Interval Estimation in Simple Linear Regression,	28
2.4.1 Confidence Intervals on β_0 , β_1 , and σ^2 ,	28
2.4.2 Interval Estimation of the Mean Response,	30
2.5 Prediction of New Observations,	33
2.6 Coefficient of Determination,	35
2.7 Using SAS for Simple Linear Regression,	36
2.8 Some Considerations in the Use of Regression,	37
2.9 Regression Through the Origin,	41
2.10 Estimation by Maximum Likelihood,	47

- 2.11 Case Where the Regressor x is Random, 49
 - 2.11.1 x and y Jointly Distributed, 49
 - 2.11.2 x and y Jointly Normally Distributed: Correlation Model, 49
- Problems, 54

3. Multiple Linear Regression **63**

- 3.1 Multiple Regression Models, 63
- 3.2 Estimation of the Model Parameters, 66
 - 3.2.1 Least-Squares Estimation of the Regression Coefficients, 66
 - 3.2.2 Geometrical Interpretation of Least Squares, 74
 - 3.2.3 Properties of the Least-Squares Estimators, 75
 - 3.2.4 Estimation of σ^2 , 76
 - 3.2.5 Inadequacy of Scatter Diagrams in Multiple Regression, 77
 - 3.2.6 Maximum-Likelihood Estimation, 79
- 3.3 Hypothesis Testing in Multiple Linear Regression, 80
 - 3.3.1 Test for Significance of Regression, 80
 - 3.3.2 Tests on Individual Regression Coefficients, 84
 - 3.3.3 Special Case of Orthogonal Columns in X , 89
 - 3.3.4 Testing the General Linear Hypothesis, 90
- 3.4 Confidence Intervals in Multiple Regression, 93
 - 3.4.1 Confidence Intervals on the Regression Coefficients, 93
 - 3.4.2 Confidence Interval Estimation of the Mean Response, 94
 - 3.4.3 Simultaneous Confidence Intervals on Regression Coefficients, 96
- 3.5 Prediction of New Observations, 99
- 3.6 Using SAS for Basic Multiple Linear Regression, 101
- 3.7 Hidden Extrapolation in Multiple Regression, 101
- 3.8 Standardized Regression Coefficients, 105
- 3.9 Multicollinearity, 109
- 3.10 Why Do Regression Coefficients Have the Wrong Sign?, 112
- Problems, 114

4. Model Adequacy Checking **122**

- 4.1 Introduction, 122
- 4.2 Residual Analysis, 123

CONTENTS

4.2.1	Definition of Residuals, 123	•••ir/oi! :. !
4.2.2	Methods for Scaling Residuals, 123	M'/1 ?.-
4.2.3	Residual Plots, 129	<• >>/
4.2.4	Partial Regression and Partial Residual Plots, 134	∴
4.2.5	Using MINITAB and SAS for Residual Analysis, 137	
4.2.6	Other Residual Plotting and Analysis Methods, 138	
4.3	PRESS Statistic, 141	
4.4	Detection and Treatment of Outliers, 142	
4.5	Lack of Fit of the Regression Model 145	∴rC .
4.5.1	Formal Test for Lack of Fit, 145	
4.5.2	Estimation of Pure Error from Near Neighbors, 149	- ,BV ∴)o /q^-nVj U
Problems,	153	»lr-,*<rV .irU^thuIU. «.JH irtt mi »tHmnr,"1 Cfc
5.	Transformations and Weighting to Correct Model Inadequacies	160
5.1	Introduction, 160	
5.2	Variance-Stabilizing Transformations, 161	•
5.3	Transformations to Linearize the Model, 164	
5.4	Analytical Methods for Selecting a Transformation, 171	^Kf->i1
5.4.1	Transformations on y: The Box-Cox Method, 171	
5.4.2	Transformations on the Regressor Variables, 174	
5.5	Generalized and Weighted Least Squares, 176	uhoitn\ !P
5.5.1	Generalized Least Squares, 177	'1 U-v
5.5.2	Weighted Least Squares, 179	> -.'P
5.5.3	Some Practical Issues, 180	• > F__t d
Problems,	183	"r- *'A«M
6.	Diagnostics for Leverage and Influence	189
6.1	Importance of Detecting Influential Observations, 189	
6.2	Leverage, 190	"• '_"
6.3	Measures of Influence: Cook's <i>D</i> , 193	
6.4	Measures of Influence: <i>DFFIT</i> S and <i>DFBETAS</i> , 195	...!H..ti
6.5	A Measure of Model Performance, 197	
6.6	Detecting Groups of Influential Observations,, 198	j,, mijutfci)*;V m
6.7	Treatment of Influential Observations, 199	
Problems,	199	U_m i *m
7.	Polynomial Regression Models	201
7.1	Introduction, 201	
7.2	Polynomial Models in One Variable, 201	
7.2.1	Basic Principles, 201	

7.2.2	Piecewise Polynomial Fitting (Splines), 207	
7.2.3	Polynomial and Trigonometric Terms, 213	
7.3	Nonparametric Regression, 214	
7.3.1	Kernel Regression, 214	
7.3.2	Locally Weighted Regression (Loess), 215	
7.3.3	Final Cautions, 219	
7.4	Polynomial Models in Two or More Variables, 220	
7.5	Orthogonal Polynomials, 226	
	Problems, 231	
8.	Indicator Variables	237
8.1	General Concept of Indicator Variables, 237	
8.2	Comments on the Use of Indicator Variables, 249	
8.2.1	Indicator Variables versus Regression on Allocated Codes, 249	
8.2.2	Indicator Variables as a Substitute for a Quantitative Regressor, 250	
8.3	Regression Approach to Analysis of Variance, 251	
	Problems, 256	
9.	Variable Selection and Model Building	261
9.1	Introduction, 261	
9.1.1	Model-Building Problem, 261	
9.1.2	Consequences of Model Misspecification, 262	
9.1.3	Criteria for Evaluating Subset Regression Models, 265	
9.2	Computational Techniques for Variable Selection, 270	
9.2.1	All Possible Regressions, 270	
9.2.2	Stepwise Regression Methods, 277	
9.3	Strategy for Variable Selection and Model Building, 283	
9.4	Case Study: Gorman and Toman Asphalt Data Using SAS, 286	
	Problems, 300	
10.	Validation of Regression Models	305
10.1	Introduction, 305	
10.2	Validation Techniques, 306	
10.2.1	Analysis of Model Coefficients and Predicted Values, 306	
10.2.2	Collecting Fresh Data—Confirmation Runs, 308	

13.1.1	Linear Regression Models, 397	
13.1.2	Nonlinear Regression Models, 398	
13.2	Origins of Nonlinear Models, 399	
13.3	Nonlinear Least Squares, 403	
13.4	Transformation to a Linear Model, 405	
13.5	Parameter Estimation in a Nonlinear System, 408	
13.5.1	Linearization, 408	
13.5.2	Other Parameter Estimation Methods, 414	
13.5.3	Starting Values, 415	
13.5.4	Computer Programs, 416	
13.6	Statistical Inference in Nonlinear Regression, 417	
13.7	Examples of Nonlinear Regression Models, 419	
13.8	Using SAS PROC NLIN, 420	
	Problems, 423	
14.	Generalized Linear Models	427
14.1	Introduction, 427	
14.2	Logistic Regression Models, 428	
14.2.1	Models with a Binary Response Variable, 428	
14.2.2	Estimating the Parameters in a Logistic Regression Model, 430	
14.2.3	Interpretation of the Parameters in a Logistic Regression Model, 433	
14.2.4	Statistical Inference on Model Parameters, 436	
14.2.5	Diagnostic Checking in Logistic Regression, 444	
14.2.6	Other Models for Binary Response Data, 446	
14.2.7	More Than Two Categorical Outcomes, 447	
14.3	Poisson Regression, 449	
14.4	The Generalized Linear Model, 454	
14.4.1	Link Functions and Linear Predictors, 455	
14.4.2	Parameter Estimation and Inference in the GLM, 456	
14.4.3	Prediction and Estimation with the GLM, 460	
14.4.4	Residual Analysis in the GLM, 461	
14.4.5	Overdispersion, 464	
	Problems, 465	
15.	Other Topics in the Use of Regression Analysis	475
15.1	Regression Models with Autocorrelated Errors, 475	
15.1.1	Source and Effects of Autocorrelation, 475	
15.1.2	Detecting the Presence of Autocorrelation, 476	
15.1.3	Parameter Estimation Methods, 479	

15.2	Effect of Measurement Errors in the Regressors, 486	
15.2.1	Simple Linear Regression, 486	
15.2.2	Berkson Model, 488	
15.3	Inverse Estimation—The Calibration Problem, 488	
15.4	Bootstrapping in Regression, 493	
15.4.1	Bootstrap Sampling in Regression, 494	
15.4.2	Bootstrap Confidence Intervals, 494	
15.5	Classification and Regression Trees (CART), 500	
15.6	Neural Networks, 502	
15.7	Designed Experiments for Regression, 505	
	Problems, 507	
APPENDIX A.	Statistical Tables	511
APPENDIX B.	Data Sets For Exercises	529
APPENDIX C.	Supplemental Technical Material	546
C.1	Background on Basic Test Statistics, 546	
C.2	Background from the Theory of Linear Models, 548	
C.3	Important Results on SS_R and $SS_{Re\&}$, 552	
C.4	Gauss-Markov Theorem, $\text{Var}(e) = \mathbf{cr}^2\mathbf{I}$, 558	
C.5	Computational Aspects of Multiple Regression, 560	
C.6	Result on the Inverse of a Matrix, 562	
C.7	Development of the PRESS Statistic, 562	
C.8	Development of S^2 , 564	
C.9	Outlier Test Based on t -Student, 565	
C.10	Independence of Residuals and Fitted Values, 568	
C.11	The Gauss-Markov Theorem, $\text{Var}(e) = \mathbf{V}$, 569	
C.12	Bias in MS_{Res} When the Model Is Underspecified, 571	
C.13	Computation of Influence Diagnostics, 572	
C.14	Generalized Linear Models, 573	
APPENDIX D.	Introduction to SAS	584
D.1	Basic Data Entry, 584	
D.2	Creating Permanent SAS Data Sets, 589	
D.3	Importing Data from an EXCEL File, 590	
D.4	Output Command, 591	
D.5	Log File, 591	
D.6	Adding Variables to an Existing SAS Data Set, 593	
	References	594
	Index	609