Applied Regression Modeling
A Business Approach

Iain Pardoe
University of Oregon
Charles H. Lundquist College of Business
Eugene, Oregon

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION
## CONTENTS

Preface xiii  
Acknowledgments xv  
Introduction xvii  

### 1 Foundations 1
  
1.1 Identifying and summarizing data 1  
1.2 Population distributions 4  
1.3 Selecting individuals at random—probability 9  
1.4 Random sampling 10  
  1.4.1 Central limit theorem—normal version 11  
  1.4.2 Student's t-distribution 12  
  1.4.3 Central limit theorem—t version 14  
1.5 Interval estimation 14  
1.6 Hypothesis testing 17  
  1.6.1 The rejection region method 17  
  1.6.2 The p-value method 19  
  1.6.3 Hypothesis test errors 23  
1.7 Random errors and prediction 23  

Preface xiii  
Acknowledgments xv  
Introduction xvii  

I.1 Statistics in business xvii  
I.2 Learning statistics xix  

1 Foundations 1
# CONTENTS

1.8 Chapter summary 26
Problems 27

2 **Simple linear regression** 31

2.1 Probability model for \( X \) and \( Y \) 31
2.2 Least squares criterion 36
2.3 Model evaluation 40
  2.3.1 Regression standard error 41
  2.3.2 Coefficient of determination—\( R^2 \) 43
  2.3.3 Slope parameter 47
2.4 Model assumptions 54
  2.4.1 Checking the model assumptions 54
2.5 Model interpretation 59
2.6 Estimation and prediction 60
  2.6.1 Confidence interval for the population mean, \( E(Y) \) 61
  2.6.2 Prediction interval for an individual \( Y \)-value 62
2.7 Chapter summary 65
  2.7.1 Review example 66
Problems 70

3 **Multiple linear regression** 73

3.1 Probability model for \( (X_1, X_2, \ldots) \) and \( Y \) 73
3.2 Least squares criterion 77
3.3 Model evaluation 81
  3.3.1 Regression standard error 81
  3.3.2 Coefficient of determination—\( R^2 \) 82
  3.3.3 Regression parameters—global usefulness test 89
  3.3.4 Regression parameters—nested model test 93
  3.3.5 Regression parameters—individual tests 97
3.4 Model assumptions 105
  3.4.1 Checking the model assumptions 106
3.5 Model interpretation 109
3.6 Estimation and prediction 111
  3.6.1 Confidence interval for the population mean, \( E(Y) \) 111
  3.6.2 Prediction interval for an individual \( Y \)-value 112
3.7 Chapter summary 114
Problems 116

4 **Regression model building I** 121

4.1 Transformations 122
  4.1.1 Natural logarithm transformation for predictors 122
4.1.2 Polynomial transformation for predictors 128
4.1.3 Reciprocal transformation for predictors 130
4.1.4 Natural logarithm transformation for the response 134
4.1.5 Transformations for the response and predictors 137
4.2 Interactions 140
4.3 Qualitative predictors 146
4.3.1 Qualitative predictors with two levels 147
4.3.2 Qualitative predictors with three or more levels 153
4.4 Chapter summary 158
Problems 160

5 Regression model building II 165
5.1 Influential points 165
5.1.1 Outliers 165
5.1.2 Leverage 168
5.1.3 Cook's distance 171
5.2 Regression pitfalls 173
5.2.1 Autocorrelation 173
5.2.2 Multicollinearity 175
5.2.3 Excluding important predictor variables 177
5.2.4 Overfitting 180
5.2.5 Extrapolation 181
5.2.6 Missing Data 183
5.3 Model building guidelines 186
5.4 Model interpretation using graphics 188
5.5 Chapter summary 194
Problems 196

6 Case studies 201
6.1 Home prices 201
6.1.1 Data description 201
6.1.2 Exploratory data analysis 203
6.1.3 Regression model building 204
6.1.4 Results and conclusions 205
6.1.5 Further questions 210
6.2 Vehicle fuel efficiency 211
6.2.1 Data description 211
6.2.2 Exploratory data analysis 211
6.2.3 Regression model building 213
6.2.4 Results and conclusions 214
6.2.5 Further questions 219
## 7 Extensions

### 7.1 Generalized linear models
- **7.1.1** Logistic regression
- **7.1.2** Poisson regression

### 7.2 Discrete choice models

### 7.3 Multilevel models

### 7.4 Bayesian modeling
- **7.4.1** Frequentist inference
- **7.4.2** Bayesian inference

## Appendix A: Computer software help

### A.1 SPSS
- **A.1.1** Getting started and summarizing univariate data
- **A.1.2** Simple linear regression
- **A.1.3** Multiple linear regression

### A.2 Minitab
- **A.2.1** Getting started and summarizing univariate data
- **A.2.2** Simple linear regression
- **A.2.3** Multiple linear regression

### A.3 SAS
- **A.3.1** Getting started and summarizing univariate data
- **A.3.2** Simple linear regression
- **A.3.3** Multiple linear regression

### A.4 R and S-PLUS
- **A.4.1** Getting started and summarizing univariate data
- **A.4.2** Simple linear regression
- **A.4.3** Multiple linear regression

### A.5 Excel
- **A.5.1** Getting started and summarizing univariate data
- **A.5.2** Simple linear regression
- **A.5.3** Multiple linear regression

*Problems*

## Appendix B: Critical values for t-distributions

## Appendix C: Notation and formulas

### C.1 Univariate data
### C.2 Simple linear regression
### C.3 Multiple linear regression

## Appendix D: Mathematics refresher
D.1 The natural logarithm and exponential functions
D.2 Rounding and accuracy

Appendix E: Brief answers to selected problems

References
Glossary
Index