

CONTENTS

1 Concepts and Examples of Research 1

- 1.1 Concepts 1
- 1.2 Examples 2
- 1.3 Concluding Remarks 5
- References 6

2 Classification of Variables and the Choice of Analysis 7

- 2.1 Classification of Variables 7
- 2.2 Overlapping of Classification Schemes 11
- 2.3 Choice of Analysis 12
- References 15

3 Basic Statistics: A Review 16

- 3.1 Preview 16
- 3.2 Descriptive Statistics 17
- 3.3 Random Variables and Distributions 19
- 3.4 Sampling Distributions of t , χ^2 , and F 24
- 3.5 Statistical Inference: Estimation 27
- 3.6 Statistical Inference: Hypothesis Testing 30
- 3.7 Error Rates, Power, and Sample Size 35
- Problems 37
- References 40

4 Introduction to Regression Analysis

41

- 4.1 Preview 41
- 4.2 Association versus Causality 42
- 4.3 Statistical versus Deterministic Models 45
- 4.4 Concluding Remarks 45
 - References 46

5 Straight-line Regression Analysis

47

- 5.1 Preview 47
- 5.2 Regression with a Single Independent Variable 47
- 5.3 Mathematical Properties of a Straight line 50
- 5.4 Statistical Assumptions for a Straight-line Model 51
- 5.5 Determining the Best-fitting Straight Line 55
- 5.6 Measure of the Quality of the Straight-line Fit and Estimate of σ^2 60
- 5.7 Inferences about the Slope and Intercept 61
- 5.8 Interpretations of Tests for Slope and Intercept 64
- 5.9 The Mean Value of Y at a Specified Value of X 66
- 5.10 Prediction of a New Value of Y at X_0 68
- 5.11 Assessing the Appropriateness of the Straight-line Model 70
- 5.12 Example: BRFSS Analysis 71
 - Problems 74
 - References 107

6 The Correlation Coefficient and Straight-line Regression Analysis

108

- 6.1 Definition of r 108
- 6.2 r as a Measure of Association 109
- 6.3 The Bivariate Normal Distribution 112
- 6.4 r^2 and the Strength of the Straight-line Relationship 113
- 6.5 What r^2 Does Not Measure 115
- 6.6 Tests of Hypotheses and Confidence Intervals for the Correlation Coefficient 117
- 6.7 Testing for the Equality of Two Correlations 120
- 6.8 Example: BRFSS Analysis 122
- 6.9 How Large Should r^2 Be in Practice? 123
 - Problems 125
 - References 127

7 The Analysis-of-Variance Table 129

- 7.1 Preview 129
- 7.2 The ANOVA Table for Straight-line Regression 129
- Problems 133

8 Multiple Regression Analysis: General Considerations 136

- 8.1 Preview 136
- 8.2 Multiple Regression Models 137
- 8.3 Graphical Look at the Problem 138
- 8.4 Assumptions of Multiple Regression 141
- 8.5 Determining the Best Estimate of the Multiple Regression Equation 143
- 8.6 The ANOVA Table for Multiple Regression 145
- 8.7 Example: BRFSS Analysis 146
- 8.8 Numerical Examples 148
- Problems 151
- References 164

9 Statistical Inference in Multiple Regression 165

- 9.1 Preview 165
- 9.2 Test for Significant Overall Regression 166
- 9.3 Partial F Test 167
- 9.4 Multiple Partial F Test 172
- 9.5 Strategies for Using Partial F Tests 175
- 9.6 Additional Inference Methods for Multiple Regression 180
- 9.7 Example: BRFSS Analysis 186
- Problems 188
- References 198

10 Correlations: Multiple, Partial, and Multiple Partial 199

- 10.1 Preview 199
- 10.2 Correlation Matrix 200
- 10.3 Multiple Correlation Coefficient 201
- 10.4 Relationship of $R_{Y|X_1, X_2, \dots, X_k}$ to the Multivariate Normal Distribution 203

10.5	Partial Correlation Coefficient	204
10.6	Alternative Representation of the Regression Model	212
10.7	Multiple Partial Correlation	212
10.8	Concluding Remarks	214
	Problems	215
	References	225

11 Confounding and Interaction in Regression **226**

11.1	Preview	226
11.2	Overview	226
11.3	Interaction in Regression	228
11.4	Confounding in Regression	236
11.5	Summary and Conclusions	242
	Problems	242
	References	256

12 Dummy Variables in Regression **257**

12.1	Preview	257
12.2	Definitions	257
12.3	Rule for Defining Dummy Variables	258
12.4	Comparing Two Straight-line Regression Equations: An Example	259
12.5	Questions for Comparing Two Straight Lines	261
12.6	Methods of Comparing Two Straight Lines	262
12.7	Method I: Using Separate Regression Fits to Compare Two Straight Lines	263
12.8	Method II: Using a Single Regression Equation to Compare Two Straight Lines	268
12.9	Comparison of Methods I and II	271
12.10	Testing Strategies and Interpretation: Comparing Two Straight Lines	272
12.11	Other Dummy Variable Models	273
12.12	Comparing Four Regression Equations	275
12.13	Comparing Several Regression Equations Involving Two Nominal Variables	277
	Problems	283
	References	307

13 Analysis of Covariance and Other Methods for Adjusting Continuous Data **308**

13.1	Preview	308
13.2	Adjustment Problem	309

13.3	Analysis of Covariance	310	
13.4	Assumption of Parallelism: A Potential Drawback	313	
13.5	Analysis of Covariance: Several Groups and Several Covariates	314	
13.6	Analysis of Covariance: Several Nominal Independent Variables	316	
13.7	Comments and Cautions	318	
13.8	Summary	321	
	Problems	321	
	References	338	

14 Regression Diagnostics

339

14.1	Preview	339	
14.2	Simple Approaches to Diagnosing Problems in Data	340	
14.3	Residual Analysis: Detecting Outliers and Violations of Model Assumptions	347	
14.4	Strategies for Addressing Violations of Regression Assumptions	355	
14.5	Collinearity	358	
14.6	Diagnostics Example	372	
	Problems	382	
	References	399	

15 Polynomial Regression

401

15.1	Preview	401	
15.2	Polynomial Models	402	
15.3	Least-squares Procedure for Fitting a Parabola	402	
15.4	ANOVA Table for Second-order Polynomial Regression	404	
15.5	Inferences Associated with Second-order Polynomial Regression	405	
15.6	Example Requiring a Second-order Model	406	
15.7	Fitting and Testing Higher-order Models	410	
15.8	Lack-of-fit Tests	410	
15.9	Orthogonal Polynomials	412	
15.10	Strategies for Choosing a Polynomial Model	422	
	Problems	423	

16 Selecting the Best Regression Equation

438

16.1	Preview	438	
16.2	Steps in Selecting the Best Regression Equation: Prediction Goal	439	
16.3	Step 1: Specifying the Maximum Model: Prediction Goal	439	
16.4	Step 2: Specifying a Criterion for Selecting a Model: Prediction Goal	442	
16.5	Step 3: Specifying a Strategy for Selecting Variables: Prediction Goal	444	
16.6	Step 4: Conducting the Analysis: Prediction Goal	454	

16.7	Step 5: Evaluating Reliability with Split Samples: Prediction Goal	454
16.8	Example Analysis of Actual Data	457
16.9	Selecting the Most Valid Model	463
	Problems	466
	References	480

17 One-way Analysis of Variance

481

17.1	Preview	481
17.2	One-way ANOVA: The Problem, Assumptions, and Data Configuration	484
17.3	Methodology for One-way Fixed-effects ANOVA	488
17.4	Regression Model for Fixed-effects One-way ANOVA	494
17.5	Fixed-effects Model for One-way ANOVA	497
17.6	Random-effects Model for One-way ANOVA	500
17.7	Multiple-comparison Procedures for Fixed-effects One-way ANOVA	503
17.8	Choosing a Multiple-comparison Technique	515
17.9	Orthogonal Contrasts and Partitioning an ANOVA Sum of Squares	516
	Problems	522
	References	543

18 Randomized Blocks: Special Case of Two-way ANOVA

545

18.1	Preview	545
18.2	Equivalent Analysis of a Matched-pairs Experiment	549
18.3	Principle of Blocking	553
18.4	Analysis of a Randomized-blocks Study	555
18.5	ANOVA Table for a Randomized-blocks Study	557
18.6	Regression Models for a Randomized-blocks Study	561
18.7	Fixed-effects ANOVA Model for a Randomized-blocks Study	565
	Problems	566
	References	578

19 Two-way ANOVA with Equal Cell Numbers

579

19.1	Preview	579
19.2	Using a Table of Cell Means	581
19.3	General Methodology	586
19.4	F Tests for Two-way ANOVA	592
19.5	Regression Model for Fixed-effects Two-way ANOVA	594

19.6	Interactions in Two-way ANOVA	599	
19.7	Random- and Mixed-effects Two-way ANOVA Models	607	
	Problems	610	
	References	629	

20 Two-way ANOVA with Unequal Cell Numbers **630**

20.1	Preview	630	
20.2	Presentation of Data for Two-way ANOVA: Unequal Cell Numbers		630
20.3	Problem with Unequal Cell Numbers: Nonorthogonality	632	
20.4	Regression Approach for Unequal Cell Sample Sizes	637	
20.5	Higher-way ANOVA	641	
	Problems	642	
	References	659	

21 The Method of Maximum Likelihood **661**

21.1	Preview	661	
21.2	The Principle of Maximum Likelihood	661	
21.3	Statistical Inference Using Maximum Likelihood	665	
21.4	Summary	677	
	Problems	678	
	References	680	

22 Logistic Regression Analysis **681**

22.1	Preview	681	
22.2	The Logistic Model	681	
22.3	Estimating the Odds Ratio Using Logistic Regression	683	
22.4	A Numerical Example of Logistic Regression	689	
22.5	Theoretical Considerations	698	
22.6	An Example of Conditional ML Estimation Involving Pair-matched Data with Unmatched Covariates	704	
22.7	Summary	707	
	Problems	708	
	References	712	

23 Polytomous and Ordinal Logistic Regression **714**

23.1	Preview	714	
23.2	Why Not Use Binary Regression?	715	

23.3	An Example of Polytomous Logistic Regression: One Predictor, Three Outcome Categories	715	
23.4	An Example: Extending the Polytomous Logistic Model to Several Predictors		721
23.5	Ordinal Logistic Regression: Overview	726	
23.6	A “Simple” Example: Three Ordinal Categories and One Dichotomous Exposure Variable	727	
23.7	Ordinal Logistic Regression Example Using Real Data with Four Ordinal Categories and Three Predictor Variables	731	
23.8	Summary	737	
	Problems	738	
	References	742	

24 Poisson Regression Analysis **743**

24.1	Preview	743	
24.2	The Poisson Distribution	743	
24.3	An Example of Poisson Regression	745	
24.4	Poisson Regression	748	
24.5	Measures of Goodness of Fit	753	
24.6	Continuation of Skin Cancer Data Example	756	
24.7	A Second Illustration of Poisson Regression Analysis	762	
24.8	Summary	765	
	Problems	766	
	References	780	

25 Analysis of Correlated Data Part 1: The General Linear Mixed Model **781**

25.1	Preview	781	
25.2	Examples	784	
25.3	General Linear Mixed Model Approach	792	
25.4	Example: Study of Effects of an Air Pollution Episode on FEV1 Levels		806
25.5	Summary—Analysis of Correlated Data: Part 1	818	
	Problems	819	
	References	824	

26 Analysis of Correlated Data Part 2: Random Effects and Other Issues **825**

26.1	Preview	825	
26.2	Random Effects Revisited	825	

26.3	Results for Models with Random Effects Applied to Air Pollution Study Data	829
26.4	Second Example—Analysis of Posture Measurement Data	839
26.5	Recommendations about Choice of Correlation Structure	859
26.6	Analysis of Data for Discrete Outcomes	861
	Problems	862
	References	882

27 Sample Size Planning for Linear and Logistic Regression and Analysis of Variance

883

27.1	Preview	883
27.2	Review: Sample Size Calculations for Comparisons of Means and Proportions	884
27.3	Sample Size Planning for Linear Regression	886
27.4	Sample Size Planning for Logistic Regression	889
27.5	Power and Sample Size Determination for Linear Models: A General Approach	893
27.6	Sample Size Determination for Matched Case-control Studies with a Dichotomous Outcome	908
27.7	Practical Considerations and Cautions	910
	Problems	911
	References	913

Appendix A—Tables

915

A.1	Standard Normal Cumulative Probabilities	916
A.2	Percentiles of the t Distribution	919
A.3	Percentiles of the Chi-square Distribution	920
A.4	Percentiles of the F Distribution	921
A.5	Values of $\frac{1}{2} \ln \frac{1+r}{1-r}$	928
A.6	Upper α Point of Studentized Range	930
A.7	Orthogonal Polynomial Coefficients	932
A.8A	Bonferroni Corrected Jackknife Residual Critical Values	933
A.8B	Bonferroni Corrected Studentized Residual Critical Values	933
A.9	Critical Values for Leverages	934
A.10	Critical Values for the Maximum of n Values of Cook's $(n - k - 1) d_i$	936

Appendix B—Matrices and Their Relationship to Regression Analysis

937

Appendix C—SAS Computer Appendix	949
Appendix D—Answers to Selected Problems	991
Index	1037