

Helge Toutenburg

Statistical Analysis of Designed Experiments

Second Edition

With Contributions by Thomas Nittner



Springer

Contents

Preface	v
1 Introduction	1
1.1 Data, Variables, and Random Processes	1
1.2 Basic Principles of Experimental Design	3
1.3 Scaling of Variables	5
1.4 Measuring and Scaling in Statistical Medicine	7
1.5 Experimental Design in Biotechnology	8
1.6 Relative Importance of Effects—The Pareto Principle . .	9
1.7 An Alternative Chart	10
1.8 A One-Way Factorial Experiment by Example	15
1.9 Exercises and Questions	19
2 Comparison of Two Samples	21
2.1 Introduction	21
2.2 Paired t -Test and Matched-Pair Design	22
2.3 Comparison of Means in Independent Groups	25
2.3.1 Two-Sample t -Test	25
2.3.2 Testing $H_0 : \sigma_A^2 = \sigma_B^2 = \sigma^2$	25
2.3.3 Comparison of Means in the Case of Un- equal Variances	26
2.3.4 Transformations of Data to Assure Homogeneity of Variances	27
2.3.5 Necessary Sample Size and Power of the Test . .	27

2.3.6	Comparison of Means without Prior Testing $H_0 : \sigma_A^2 = \sigma_B^2$; Cochran–Cox Test for Independent Groups	27
2.4	Wilcoxon’s Sign–Rank Test in the Matched–Pair Design	29
2.5	Rank Test for Homogeneity of Wilcoxon, Mann and Whitney	33
2.6	Comparison of Two Groups with Categorical Response	38
2.6.1	McNemar’s Test and Matched–Pair Design	38
2.6.2	Fisher’s Exact Test for Two Independent Groups	39
2.7	Exercises and Questions	41
3	The Linear Regression Model	45
3.1	Descriptive Linear Regression	45
3.2	The Principle of Ordinary Least Squares	47
3.3	Geometric Properties of Ordinary Least Squares Estimation	50
3.4	Best Linear Unbiased Estimation	51
3.4.1	Linear Estimators	52
3.4.2	Mean Square Error	53
3.4.3	Best Linear Unbiased Estimation	55
3.4.4	Estimation of σ^2	57
3.5	Multicollinearity	60
3.5.1	Extreme Multicollinearity and Estimability	60
3.5.2	Estimation within Extreme Multicollinearity	61
3.5.3	Weak Multicollinearity	63
3.6	Classical Regression under Normal Errors	67
3.7	Testing Linear Hypotheses	69
3.8	Analysis of Variance and Goodness of Fit	73
3.8.1	Bivariate Regression	73
3.8.2	Multiple Regression	79
3.9	The General Linear Regression Model	83
3.9.1	Introduction	83
3.9.2	Misspecification of the Covariance Matrix	85
3.10	Diagnostic Tools	86
3.10.1	Introduction	86
3.10.2	Prediction Matrix	86
3.10.3	Effect of a Single Observation on the Estimation of Parameters	91
3.10.4	Diagnostic Plots for Testing the Model Assumptions	96
3.10.5	Measures Based on the Confidence Ellipsoid	97
3.10.6	Partial Regression Plots	102
3.10.7	Regression Diagnostics by Animating Graphics	104

3.11	Exercises and Questions	110
4	Single-Factor Experiments with Fixed and Random Effects	111
4.1	Models I and II in the Analysis of Variance	111
4.2	One-Way Classification for the Multiple Comparison of Means	112
4.2.1	Representation as a Restrictive Model	115
4.2.2	Decomposition of the Error Sum of Squares	117
4.2.3	Estimation of σ^2 by MS_{Error}	120
4.3	Comparison of Single Means	123
4.3.1	Linear Contrasts	123
4.3.2	Contrasts of the Total Response Values in the Balanced Case	126
4.4	Multiple Comparisons	132
4.4.1	Introduction	132
4.4.2	Experimentwise Comparisons	132
4.4.3	Select Pairwise Comparisons	135
4.5	Regression Analysis of Variance	142
4.6	One-Factorial Models with Random Effects	145
4.7	Rank Analysis of Variance in the Completely Randomized Design	149
4.7.1	Kruskal-Wallis Test	149
4.7.2	Multiple Comparisons	152
4.8	Exercises and Questions	154
5	More Restrictive Designs	157
5.1	Randomized Block Design	157
5.2	Latin Squares	165
5.2.1	Analysis of Variance	167
5.3	Rank Variance Analysis in the Randomized Block Design	172
5.3.1	Friedman Test	172
5.3.2	Multiple Comparisons	175
5.4	Exercises and Questions	176
6	Multifactor Experiments	179
6.1	Elementary Definitions and Principles	179
6.2	Two-Factor Experiments (Fixed Effects)	183
6.3	Two-Factor Experiments in Effect Coding	188
6.4	Two-Factorial Experiment with Block Effects	196
6.5	Two-Factorial Model with Fixed Effects—Confidence Intervals and Elementary Tests	199
6.6	Two-Factorial Model with Random or Mixed Effects	203
6.6.1	Model with Random Effects	203

6.6.2	Mixed Model	207
6.7	Three-Factorial Designs	211
6.8	Split-Plot Design	215
6.9	2^k Factorial Design	219
6.9.1	The 2^2 Design	219
6.9.2	The 2^3 Design	222
6.10	Exercises and Questions	225
7	Models for Categorical Response Variables	231
7.1	Generalized Linear Models	231
7.1.1	Extension of the Regression Model	231
7.1.2	Structure of the Generalized Linear Model	233
7.1.3	Score Function and Information Matrix	236
7.1.4	Maximum Likelihood Estimation	237
7.1.5	Testing of Hypotheses and Goodness of Fit	240
7.1.6	Overdispersion	241
7.1.7	Quasi Loglikelihood	243
7.2	Contingency Tables	245
7.2.1	Overview	245
7.2.2	Ways of Comparing Proportions	246
7.2.3	Sampling in Two-Way Contingency Tables	249
7.2.4	Likelihood Function and Maximum Likelihood Estimates	250
7.2.5	Testing the Goodness of Fit	252
7.3	Generalized Linear Model for Binary Response	254
7.3.1	Logit Models and Logistic Regression	254
7.3.2	Testing the Model	257
7.3.3	Distribution Function as a Link Function	258
7.4	Logit Models for Categorical Data	258
7.5	Goodness of Fit—Likelihood Ratio Test	260
7.6	Loglinear Models for Categorical Variables	261
7.6.1	Two-Way Contingency Tables	261
7.6.2	Three-Way Contingency Tables	264
7.7	The Special Case of Binary Response	267
7.8	Coding of Categorical Explanatory Variables	270
7.8.1	Dummy and Effect Coding	270
7.8.2	Coding of Response Models	273
7.8.3	Coding of Models for the Hazard Rate	274
7.9	Extensions to Dependent Binary Variables	277
7.9.1	Overview	277
7.9.2	Modeling Approaches for Correlated Response	279
7.9.3	Quasi-Likelihood Approach for Correlated Binary Response	280
7.9.4	The Generalized Estimating Equation Method by Liang and Zeger	281

7.9.5	Properties of the Generalized Estimating Equation Estimate $\hat{\beta}_G$	283
7.9.6	Efficiency of the Generalized Estimating Equation and Independence Estimating Equation Methods	284
7.9.7	Choice of the Quasi-Correlation Matrix $R_i(\alpha)$	285
7.9.8	Bivariate Binary Correlated Response Variables	285
7.9.9	The Generalized Estimating Equation Method	286
7.9.10	The Independence Estimating Equation Method	288
7.9.11	An Example from the Field of Dentistry	288
7.9.12	Full Likelihood Approach for Marginal Models	293
7.10	Exercises and Questions	294
8	Repeated Measures Model	295
8.1	The Fundamental Model for One Population	295
8.2	The Repeated Measures Model for Two Populations	298
8.3	Univariate and Multivariate Analysis	301
8.3.1	The Univariate One-Sample Case	301
8.3.2	The Multivariate One-Sample Case	301
8.4	The Univariate Two-Sample Case	306
8.5	The Multivariate Two-Sample Case	307
8.6	Testing of $H_0 : \Sigma_x = \Sigma_y$	308
8.7	Univariate Analysis of Variance in the Repeated Measures Model	309
8.7.1	Testing of Hypotheses in the Case of Compound Symmetry	309
8.7.2	Testing of Hypotheses in the Case of Sphericity	311
8.7.3	The Problem of Nonsphericity	315
8.7.4	Application of Univariate Modified Approaches in the Case of Nonsphericity	316
8.7.5	Multiple Tests	317
8.7.6	Examples	318
8.8	Multivariate Rank Tests in the Repeated Measures Model	324
8.9	Categorical Regression for the Repeated Binary Response Data	329
8.9.1	Logit Models for the Repeated Binary Response for the Comparison of Therapies	329
8.9.2	First-Order Markov Chain Models	330
8.9.3	Multinomial Sampling and Loglinear Models for a Global Comparison of Therapies	332
8.10	Exercises and Questions	339
9	Cross-Over Design	341
9.1	Introduction	341
9.2	Linear Model and Notations	342

9.3	2×2 Cross-Over (Classical Approach)	343
9.3.1	Analysis Using t -Tests	344
9.3.2	Analysis of Variance	348
9.3.3	Residual Analysis and Plotting the Data	352
9.3.4	Alternative Parametrizations in 2×2 Cross-Over	356
9.3.5	Cross-Over Analysis Using Rank Tests	368
9.4	2×2 Cross-Over and Categorical (Binary) Response	368
9.4.1	Introduction	368
9.4.2	Loglinear and Logit Models	372
9.5	Exercises and Questions	384
10	Statistical Analysis of Incomplete Data	385
10.1	Introduction	385
10.2	Missing Data in the Response	390
10.2.1	Least Squares Analysis for Complete Data	390
10.2.2	Least Squares Analysis for Filled-Up Data	391
10.2.3	Analysis of Covariance—Bartlett's Method	392
10.3	Missing Values in the X -Matrix	393
10.3.1	Missing Values and Loss of Efficiency	394
10.3.2	Standard Methods for Incomplete X -Matrices	397
10.4	Adjusting for Missing Data in 2×2 Cross-Over Designs	400
10.4.1	Notation	400
10.4.2	Maximum Likelihood Estimator (Rao, 1956)	402
10.4.3	Test Procedures	403
10.5	Missing Categorical Data	407
10.5.1	Introduction	407
10.5.2	Maximum Likelihood Estimation in the Complete Data Case	408
10.5.3	Ad-Hoc Methods	409
10.5.4	Model-Based Methods	410
10.6	Exercises and Questions	412
A	Matrix Algebra	415
A.1	Introduction	415
A.2	Trace of a Matrix	418
A.3	Determinant of a Matrix	418
A.4	Inverse of a Matrix	420
A.5	Orthogonal Matrices	421
A.6	Rank of a Matrix	422
A.7	Range and Null Space	422
A.8	Eigenvalues and Eigenvectors	423
A.9	Decomposition of Matrices	425
A.10	Definite Matrices and Quadratic Forms	427
A.11	Idempotent Matrices	433

A.12	Generalized Inverse	434
A.13	Projections	442
A.14	Functions of Normally Distributed Variables	443
A.15	Differentiation of Scalar Functions of Matrices	446
A.16	Miscellaneous Results, Stochastic Convergence	449
B	Theoretical Proofs	453
B.1	The Linear Regression Model	453
B.2	Single-Factor Experiments with Fixed and Random Effects	475
C	Distributions and Tables	479
	References	487
	Index	497