Statistics

ANALYSIS OF MESSY DATA

VOLUME 1

DESIGNED EXPERIMENTS

SECOND EDITION

A bestseller for nearly 25 years, **Analysis of Messy Data, Volume 1: Designed Experiments** helps you analyze the kinds of data sets encountered in the real world. Written by two long-time researchers and professors, this second edition has been fully updated to reflect the many developments that have occurred since the original publication.

New to the Second Edition

- Several modern suggestions for multiple comparison procedures
- Additional examples of split-plot designs and repeated measures designs
- The use of SAS-GLM to analyze an effects model
- The use of SAS-MIXED and JMP to analyze data in random effects experiments, mixed model experiments, and repeated measures experiments

The book explores various techniques for multiple comparison procedures, random effects models, mixed models, split-plot experiments, and repeated measures designs. The authors implement the techniques using several statistical software packages and emphasize the distinction between design structure and the structure of treatments. They introduce each topic with examples, follow up with a theoretical discussion, and conclude with a case study. Bringing a classic work up to date, this edition will continue to show you how to effectively analyze real-world, nonstandard data sets.

C3340

ISBN: 978-1-58488-334-0

90000

9 781584 883340

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

1. The Simplest Case: One-Way Treatment Structure in a Completely Randomized Design Structure with Homogeneous Errors 1
1.1 Model Definitions and Assumptions 1
1.2 Parameter Estimation 2
1.3 Inferences on Linear Combinations—Tests and Confidence Intervals 4
1.4 Example—Tasks and Pulse Rate 5
1.5 Simultaneous Tests on Several Linear Combinations 7
1.6 Example—Tasks and Pulse Rate (Continued) 9
1.7 Testing the Equality of All Means 11
1.8 Example—Tasks and Pulse Rate (Continued) 12
1.9 General Method for Comparing Two Models—The Principle of Conditional Error 13
1.10 Example—Tasks and Pulse Rate (Continued) 15
1.11 Computer Analyses 15
1.12 Concluding Remarks 16
1.13 Exercises 17

2. One-Way Treatment Structure in a Completely Randomized Design Structure with Heterogeneous Errors 21
2.1 Model Definitions and Assumptions 22
2.2 Parameter Estimation 22
2.3 Tests for Homogeneity of Variances 23
2.3.1 Hartley's *F*-Max Test 23
2.3.2 Bartlett's Test 24
2.3.3 Levene's Test 24
2.3.4 Brown and Forsythe's Test 24
2.3.5 O'Brien's Test 25
2.3.6 Some Recommendations 25
2.4 Example—Drugs and Errors 25
2.5 Inferences on Linear Combinations 29
2.6 Example—Drugs and Errors (Continued) 31

2.7 General Satterthwaite Approximation for Degrees of Freedom 33
2.8 Comparing All Means 34
2.9 Concluding Remarks 38
2.10 Exercises 39

3. Simultaneous Inference Procedures and Multiple Comparisons 43
3.1 Error Rates 44
3.2 Recommendations 45
3.3 Least Significant Difference 46
3.4 Fisher's LSD Procedure 47
3.5 Bonferroni's Method 48
3.6 Scheffé's Procedure 48
3.7 Tukey–Kramer Method 49
3.8 Simulation Methods 50
3.9 Šidák Procedure 51
3.10 Example—Pairwise Comparisons 51
3.11 Dunnett's Procedure 53
3.12 Example—Comparing with a Control 54
3.13 Multivariate t 55
3.14 Example—Linearly Independent Comparisons 56
3.15 Sequential Rejective Methods 57
3.15.1 Bonferroni–Holm Method 58
3.15.2 Šidák–Holm Method 58
3.15.3 Benjamini and Hochberg Method to Control FDR 58
3.16 Example—Linearly Dependent Comparisons 59
3.17 Multiple Range Tests 61
3.17.1 Student–Newman–Keul's Method 61
3.17.2 Duncan's New Multiple Range Method 64
3.18 Waller–Duncan Procedure 65
3.19 Example—Multiple Range for Pairwise Comparisons 66
3.20 A Caution 68
3.21 Concluding Remarks 69
3.22 Exercises 69

4. Basics for Designing Experiments 71
4.1 Introducing Basic Ideas 71
4.2 Structures of a Designed Experiment 77
4.2.1 Types of Design Structures 78
4.2.2 Types of Treatment Structures 80
4.3 Examples of Different Designed Experiments 83
4.3.1 Example 4.1: Diets 83
4.3.2 Example 4.2: House Paint 88
4.3.3 Example 4.3: Steel Plates 90
4.3.4 Example 4.4: Levels of N and K 92
4.3.5 Example 4.5: Blocks and Replications 94
4.3.6 Example 4.6: Row and Column Blocks 97
4.4 Concluding Remarks 98
4.5 Exercises 98

5. Multilevel Designs: Split-Plots, Strip-Plots, Repeated Measures and Combinations 101
5.1 Identifying Sizes of Experimental Units—Four Basic Design Structures 101
5.2 Hierarchical Design: A Multilevel Design Structure 114
5.3 Split-Plot Design Structures: Two-Level Design Structures 115
5.3.1 Example 5.1: Cooking Beans—The Simplest Split-Plot or Two-Level Design Structure 116
5.3.2 Example 5.2: Grinding Wheat—The Usual Split-Plot or Two-Level Design Structure 120
5.3.3 Example 5.3: Baking Bread—Split-Plot with Incomplete Block Design Structure 122
5.3.4 Example 5.4: Meat in Display Case—A Complex Split-Plot or Four-Level Design 125
5.4 Strip-Plot Design Structures—A Nonhierarchical Multilevel Design 131
5.4.1 Example 5.5: Making Cheese 131
5.5 Repeated Measures Designs 135
5.5.1 Example 5.6: Horse Feet—Basic Repeated Measures Design 136
5.5.2 Example 5.7: Comfort Study—Repeated Measures Design 139
5.5.3 Example 5.8: Crossover or Change-Over Designs 141
5.6 Designs Involving Nested Factors 142
5.6.1 Example 5.9: Animal Genetics 143
5.6.2 Example 5.10: Soybeans in Maturity Groups 143
5.6.3 Example 5.11: Engines on Aircraft 145
5.6.4 Example 5.12: Simple Comfort Experiment 147
5.6.5 Example 5.13: Multilocation Study with Repeated Measures 148
5.7 Concluding Remarks 150
5.8 Exercises 151

6. Matrix Form of the Model 155
6.1 Basic Notation 155
6.1.1 Simple Linear Regression Model 156
6.1.2 One-Way Treatment Structure Model 156
6.1.3 Two-Way Treatment Structure Model 158
6.1.4 Example 6.1: Means Model for Two-Way Treatment Structure 159
6.2 Least Squares Estimation 161
6.2.1 Least Squares Equations 162
6.2.2 Sum-to-Zero Restrictions 165
6.2.3 Set-to-Zero Restrictions 166
6.2.4 Example 6.2: A One-Way Treatment Structure 167
6.3 Estimability and Connected Designs 169
6.3.1 Estimable Functions 170
6.3.2 Connectedness 171
6.4 Testing Hypotheses about Linear Model Parameters 171
6.5 Population Marginal Means 173
6.6 Concluding Remarks 179
6.7 Exercises 179

7. Balanced Two-Way Treatment Structures 181
7.1 Model Definition and Assumptions 181
7.1.1 Means Model 181
7.1.2 Effects Model 182
7.2 Parameter Estimation 182
7.3 Interactions and Their Importance 183
7.4 Main Effects 183
7.5 Computer Analyses 184
7.6 Concluding Remarks 184
7.7 Exercises 184

8. Case Study: Complete Analyses of Balanced Two-Way Experiments 187
8.1 Contrasts of Main Effect Means 187
8.2 Contrasts of Interaction Effects 188
8.3 Paint–Paving Example 189
8.4 Analyzing Quantitative Treatment Factors 193
8.5 Multiple Comparisons 195
8.6 Concluding Remarks 195
8.7 Exercises 195

9. Using the Means Model to Analyze Balanced Two-Way Treatment Structures with Unequal Subclass Numbers 199
9.1 Model Definitions and Assumptions 199
9.2 Parameter Estimation 199
9.3 Testing Whether All Means Are Equal 201
9.4 Interaction and Main Effect Hypotheses 202
9.5 Population Marginal Means 204
9.6 Simultaneous Inferences and Multiple Comparisons 206
9.7 Concluding Remarks 207
9.8 Exercises 207

10. Using the Effects Model to Analyze Balanced Two-Way Treatment Structures with Unequal Subclass Numbers 209
10.1 Model Definition 209
10.2 Parameter Estimates and Type I Analysis 209
10.3 Using Estimable Functions in SAS 213
10.4 Types I–IV Hypotheses 217
10.5 Using Types I–IV Estimable Functions in SAS-GLM 222
10.6 Population Marginal Means and Least Squares Means 226
10.7 Computer Analyses 226
10.8 Concluding Remarks 228
10.9 Exercises 229

11. Analyzing Large Balanced Two-Way Experiments Having Unequal Subclass Numbers 231
11.1 Feasibility Problems 231
11.2 Method of Unweighted Means 232
11.3 Simultaneous Inference and Multiple Comparisons 233
11.4 An Example of the Method of Unweighted Means 235

11.5 Computer Analyses 236
11.6 Concluding Remarks 237
11.7 Exercises 238

12. Case Study: Balanced Two-Way Treatment Structure with Unequal Subclass Numbers 239
12.1 Fat–Surfactant Example 239
12.2 Concluding Remarks 243
12.3 Exercises 243

13. Using the Means Model to Analyze Two-Way Treatment Structures with Missing Treatment Combinations 245
13.1 Parameter Estimation 245
13.2 Hypothesis Testing and Confidence Intervals 247
13.2.1 Example 13.1 247
13.3 Computer Analyses 249
13.4 Concluding Remarks 253
13.5 Exercises 253

14. Using the Effects Model to Analyze Two-Way Treatment Structures with Missing Treatment Combinations 255
14.1 Type I and II Hypotheses 255
14.2 Type III Hypotheses 257
14.3 Type IV Hypotheses 259
14.4 Population Marginal Means and Least Squares Means 261
14.5 Computer Analyses 262
14.6 Concluding Remarks 262
14.7 Exercises 263

15. Case Study: Two-Way Treatment Structure with Missing Treatment Combinations 265
15.1 Case Study 265
15.2 Concluding Remarks 268
15.3 Exercise 268

16. Analyzing Three-Way and Higher-Order Treatment Structures 271
16.1 General Strategy 271
16.2 Balanced and Unbalanced Experiments 273
16.3 Type I and II Analyses 273
16.4 Concluding Remarks 274
16.5 Exercises 274

17. Case Study: Three-Way Treatment Structure with Many Missing Treatment Combinations 277
17.1 Nutrition Scores Example 277
17.2 An SAS-GLM Analysis 277
17.3 A Complete Analysis 282
17.4 Concluding Remarks 284
17.5 Exercises 285

18. Random Effects Models and Variance Components 287
18.1 Introduction 287
18.1.1 Example 18.1: Random Effects Nested Treatment Structure 289
18.2 General Random Effects Model in Matrix Notation 290
18.2.1 Example 18.2: One-Way Random Effects Model 290
18.3 Computing Expected Mean Squares 292
18.3.1 Algebraic Method 293
18.3.2 Computing Using Hartley's Method of Synthesis 295
18.4 Concluding Remarks 307
18.5 Exercises 308

19. Methods for Estimating Variance Components 309
19.1 Method of Moments 309
19.1.1 Applications. Example 19.1: Unbalanced One-Way Model 312
19.1.2 Example 19.2: Wheat Varieties in a One-Way Random Effects Model 313
19.1.3 Example 19.3: Data for Two-Way Design in Table 18.2 315
19.2 Maximum Likelihood Estimators 318
19.2.1 Example 19.4: Maximum Likelihood Solution for Balanced One-Way Model 319
19.3 Restricted or Residual Maximum Likelihood Estimation 322
19.3.1 Example 19.5: REML Solution for Balanced One-Way Model 323
19.4 MIVQUE Method 325
19.4.1 Description of the Method 325
19.4.2 Application. Example 19.6: MIVQUE for the Unbalanced One-Way Design 327
19.5 Estimating Variance Components Using JMP 329
19.6 Concluding Remarks 334
19.7 Exercises 334

20. Methods for Making Inferences about Variance Components 337
20.1 Testing Hypotheses 337
20.1.1 Using the Analysis of Variance Table 338
20.1.2 Example 20.1: Two-Way Random Effects TS in a CR DS 338
20.1.3 Example 20.2: Complex Three-Way Random Effects TS 339
20.1.4 Likelihood Ratio Test 344
20.1.5 Example 20.3: Wheat Varieties—One-Way Random Effects Model 345
20.1.6 Example 20.4: Unbalanced Two-Way 347
20.2 Constructing Confidence Intervals 347
20.2.1 Residual Variance σ_ε^2 348
20.2.2 General Satterthwaite Approximation 348
20.2.3 Approximate Confidence Interval for a Function of the Variance Components 349
20.2.4 Wald-Type Confidence Intervals for Variance Components 353
20.2.5 Some Exact Confidence Intervals 353
20.2.6 Example 20.5: Balanced One-Way Random Effects Treatment Structure 356

20.2.7 Example 20.6 357
20.2.8 Example 20.6 Continued 360
20.3 Simulation Study 361
20.4 Concluding Remarks 362
20.5 Exercises 363

21. Case Study: Analysis of a Random Effects Model 365
21.1 Data Set 365
21.2 Estimation 367
21.3 Model Building 368
21.4 Reduced Model 370
21.5 Confidence Intervals 374
21.6 Computations Using JMP 379
21.7 Concluding Remarks 380
21.8 Exercises 382

22. Analysis of Mixed Models 385
22.1 Introduction to Mixed Models 385
22.2 Analysis of the Random Effects Part of the Mixed Model 387
22.2.1 Method of Moments 387
22.2.2 Method of Maximum Likelihood 390
22.2.3 Method of Residual Maximum Likelihood 392
22.2.4 MINQUE Method 394
22.3 Analysis of the Fixed Effects Part of the Model 394
22.3.1 Estimation 394
22.3.2 Construction of Confidence Intervals 395
22.3.3 Testing Hypotheses 396
22.4 Best Linear Unbiased Prediction 397
22.5 Mixed Model Equations 397
22.6 Concluding Remarks 399
22.7 Exercises 399

23. Case Studies of a Mixed Model 401
23.1 Two-Way Mixed Model 401
23.2 Unbalanced Two-Way Mixed Model 409
23.3 JMP Analysis of the Unbalanced Two-Way Data Set 415
23.4 Concluding Remarks 417
23.5 Exercises 418

24. Methods for Analyzing Split-Plot Type Designs 421
24.1 Introduction 421
24.1.1 Example 24.1: Bread Recipes and Baking Temperatures 422
24.1.2 Example 24.2: Wheat Varieties Grown in Different Fertility Regimes 426
24.2 Model Definition and Parameter Estimation 428
24.3 Standard Errors for Comparisons among Means 430
24.4 A General Method for Computing Standard Errors of Differences of Means 434
24.5 Comparison via General Contrasts 436

24.6 Additional Examples 440
24.6.1 Example 24.3: Moisture and Fertilizer 440
24.6.2 Example 24.4: Regression with Split-Plot Errors 444
24.6.3 Example 24.5: Mixed-Up Split-Plot Design 448
24.6.4 Example 24.6: Split-Split-Plot Design 451
24.7 Sample Size and Power Considerations 455
24.8 Computations Using JMP—Example 24.7 459
24.9 Concluding Remarks 466
24.10 Exercises 467

25. Methods for Analyzing Strip-Plot Type Designs 471
25.1 Description of the Strip-Plot Design and Model 471
25.2 Techniques for Making Inferences 475
25.3 Example: Nitrogen by Irrigation 477
25.4 Example: Strip-Plot with Split-Plot 1 480
25.5 Example: Strip-Plot with Split-Plot 2 482
25.6 Strip-Plot with Split-Plot 3 483
25.7 Split-Plot with Strip-Plot 4 485
25.8 Strip-Strip-Plot Design with Analysis via JMP7 486
25.9 Concluding Remarks 491
25.10 Exercises 493

26. Methods for Analyzing Repeated Measures Experiments 499
26.1 Model Specifications and Ideal Conditions 500
26.2 The Split-Plot in Time Analyses 502
26.2.1 Example 26.1: Effect of Drugs on Heart Rate 502
26.2.2 Example 26.2: A Complex Comfort Experiment 510
26.2.3 Example 26.3: Family Attitudes 516
26.3 Data Analyses Using the SAS-Mixed Procedure 519
26.3.1 Example 26.1 521
26.3.2 Example 26.2 524
26.3.3 Example 26.3 525
26.4 Concluding Remarks 529
26.5 Exercises 529

27. Analysis of Repeated Measures Experiments When the Ideal Conditions Are Not Satisfied 535
27.1 Introduction 535
27.2 MANOVA Methods 537
27.3 *p*-Value Adjustment Methods 547
27.4 Mixed Model Methods 553
27.4.1 Maximum Likelihood Method 557
27.4.2 Restricted Maximum Likelihood Method 557
27.5 Summary 568
27.6 Exercises 570

28. Case Studies: Complex Examples Having Repeated Measures 573
28.1 Complex Comfort Experiment 573
28.2 Family Attitudes Experiment 583

28.3 Multilocation Experiment 592
28.4 Exercises 597

29. Analysis of Crossover Designs 599
29.1 Definitions, Assumptions, and Models 599
29.2 Two Period/Two Treatment Designs 600
29.3 Crossover Designs with More than Two Periods 609
29.4 Crossover Designs with More than Two Treatments 616
29.5 Summary 625
29.6 Exercises 625

30. Analysis of Nested Designs 627
30.1 Definitions, Assumptions, and Models 627
30.1.1 Example 30.1: Companies and Insecticides 627
30.1.2 Example 30.2: Comfort Experiment Revisited 629
30.1.3 Example 30.3: Coffee Price Example Revisited 629
30.2 Parameter Estimation 629
30.2.1 Example 30.1: Continuation 630
30.2.2 Example 30.2: Continuation 631
30.2.3 Example 30.3: Continuation 632
30.3 Testing Hypotheses and Confidence Interval Construction 633
30.3.1 Example 30.1: Continuation 634
30.4 Analysis Using JMP 636
30.5 Concluding Remarks 640
30.6 Exercises 640

Appendix 643

References 651

Index 655