
BIOSTATISTICAL ANALYSIS

FOURTH EDITION

JERROLD H. ZAR

Department of Biological Sciences
Northern Illinois University



PRENTICE HALL

Upper Saddle River, New Jersey 07458

CONTENTS

PREFACE *x*

1 INTRODUCTION *1*

- 1.1 Types of Biological Data 2
- 1.2 Accuracy and Significant Figures 5

- 1.3 Frequency Distributions 6
- 1.4 Cumulative Frequency Distributions 13

2 POPULATIONS AND SAMPLES *16*

- 2.1 Populations 16
- 2.2 Samples from Populations 17

- 2.3 Random Sampling 17
- 2.4 Parameters and Statistics 18

3 MEASURES OF CENTRAL TENDENCY *20*

- 3.1 The Arithmetic Mean 20
- 3.2 The Median 23
- 3.3 Other Quantiles 26
- 3.4 The Mode 27

- 3.5 Other Measures of Central Tendency 28
- 3.6 The Effect of Coding Data 29
- Exercises 31

4 MEASURES OF DISPERSION AND VARIABILITY *32*

- 4.1 The Range 32
- 4.2 Dispersion Measured with Quantiles 34
- 4.3 The Mean Deviation 34
- 4.4 The Variance 35
- 4.5 The Standard Deviation 39

- 4.6 The Coefficient of Variation 40
- 4.7 Indices of Diversity 40
- 4.8 The Effect of Coding Data 44
- Exercises 47

5 PROBABILITIES *48*

- 5.1 Counting Possible Outcomes 49
- 5.2 Permutations 50
- 5.3 Combinations 54
- 5.4 Sets 56

- 5.5 Probability of an Event 58
- 5.6 Adding Probabilities 59
- 5.7 Multiplying Probabilities 61
- Exercises 63

6 THE NORMAL DISTRIBUTION *65*

- 6.1 Symmetry and Kurtosis 67
- 6.2 Proportions of a Normal Distribution 72
- 6.3 The Distribution of Means 76

- 6.4 Introduction to Statistical Hypothesis Testing 79
- 6.5 Assessing Departures from Normality 86
- Exercises 89

7 ONE-SAMPLE HYPOTHESES 91

- 7.1 Two-Tailed Hypotheses Concerning the Mean 91
- 7.2 One-Tailed Hypotheses Concerning the Mean 96
- 7.3 Confidence Limits for the Population Mean 98
- 7.4 Reporting Variability about the Mean 100
- 7.5 Sample Size and Estimation of the Population Mean 105
- 7.6 Power and Sample Size in Tests Concerning the Mean 105
- 7.7 Sampling Finite Populations 108
- 7.8 Confidence Limits for the Population Median 110

8 TWO-SAMPLE HYPOTHESES 122

- 8.1 Testing for Difference between Two Means 122
- 8.2 Confidence Limits for Population Means 129
- 8.3 Sample Size and Estimation of the Difference between Two Population Means 131
- 8.4 Power and Sample Size in Tests for Difference between Two Means 132
- 8.5 Testing for Difference between Two Variances 136
- 8.6 Confidence Interval for the Population Variance Ratio 139
- 8.7 Sample Size and Power in Tests for Difference between Two Variances 140

9 PAIRED-SAMPLE HYPOTHESES 161

- 9.1 Testing Mean Difference 161
- 9.2 Confidence Limits for the Population Mean Difference 164
- 9.3 Power and Sample Size in Paired-Sample Testing of Means 164
- 9.4 Testing for the Difference between Variances from Two Correlated Populations 164

10 MULTISAMPLE HYPOTHESES: THE ANALYSIS OF VARIANCE 177

- 10.1 Single-Factor Analysis of Variance 178
- 10.2 Confidence Limits for Population Means 189
- 10.3 Power and Sample Size in Analysis of Variance 189
- 10.4 Nonparametric Analysis of Variance 195
- 10.5 Testing for Difference among Several Medians 200

11 MULTIPLE COMPARISONS 208

- 11.1 The Tukey Test 210
- 11.2 The Newman-Keuls Test 214
- 11.3 Confidence Intervals Following Multiple Comparisons 215
- 11.4 Comparison of a Control Mean to Each Other Group Mean 217

- 7.9 Hypotheses Concerning the Median 110
- 7.10 Confidence Limits for the Population Variance 110
- 7.11 Hypotheses Concerning the Variance 112
- 7.12 Power and Sample Size in Tests Concerning the Variance 113
- 7.13 Hypotheses Concerning the Coefficient of Variation 114
- 7.14 Hypotheses Concerning Symmetry and Kurtosis 115
- Exercises 120

- 8.8 Testing for Difference between Two Coefficients of Variation 141
- 8.9 Nonparametric Statistical Methods 145
- 8.10 Two-Sample Rank Testing 146
- 8.11 Testing for Difference between Two Medians 155
- 8.12 The Effect of Coding 155
- 8.13 Two-Sample Testing of Nominal-Scale Data 156
- 8.14 Testing for Difference between Two Diversity Indices 156
- Exercises 159

- 9.5 Paired-Sample Testing by Ranks 165
- 9.6 Confidence Limits for the Population Median Difference 169
- 9.7 Paired-Sample Testing of Nominal-Scale Data 169
- Exercises 175

- 10.6 Homogeneity of Variances 202
- 10.7 Homogeneity of Coefficients of Variation 204
- 10.8 The Effect of Coding 206
- 10.9 Multisample Testing for Nominal-Scale Data 206
- Exercises 206

- 11.5 Scheffé's Multiple Contrasts 219
- 11.6 Nonparametric Multiple Comparisons 223
- 11.7 Nonparametric Multiple Contrasts 226
- 11.8 Multiple Comparisons among Medians 226
- 11.9 Multiple Comparisons among Variances 228
- Exercises 230

- 12 TWO-FACTOR ANALYSIS OF VARIANCE 231**
- 12.1 Two-Factor Analysis of Variance with Equal Replication 232
 - 12.2 Two-Factor Analysis of Variance with Unequal Replication 245
 - 12.3 Two-Factor Analysis of Variance without Replication 248
 - 12.4 The Randomized Block Experimental Design 250
 - 12.5 Repeated-Measures Experimental Designs 255
 - 12.6 Multiple Comparisons and Confidence Intervals in Two-Factor Analysis of Variance 260
 - 12.7 Power and Sample Size in Two-Factor Analysis of Variance 261
 - 12.8 Nonparametric Randomized Block or Repeated-Measures Analysis of Variance 263
 - 12.9 Multiple Comparisons for Nonparametric Randomized Block or Repeated-Measures Analysis of Variance 267
 - 12.10 Dichotomous Nominal-Scale Data in Randomized Blocks or from Repeated Measures 268
 - 12.11 Multiple Comparisons with Dichotomous Randomized Block or Repeated-Measures Data 270
 - 12.12 Introduction to Analysis of Covariance 270
Exercises 271
- 13 DATA TRANSFORMATIONS 273**
- 13.1 The Logarithmic Transformation 275
 - 13.2 The Square Root Transformation 275
 - 13.3 The Arcsine Transformation 278
 - 13.4 Other Transformations 280
Exercises 280
- 14 MULTIWAY FACTORIAL ANALYSIS OF VARIANCE 282**
- 14.1 Three-Factor Analysis of Variance 283
 - 14.2 The Latin Square Experimental Design 286
 - 14.3 Higher-Order Factorial Analysis of Variance 287
 - 14.4 Blocked and Repeated-Measures Experimental Designs 288
 - 14.5 Factorial Analysis of Variance with Unequal Replication 298
 - 14.6 Multiple Comparisons and Confidence Intervals in Multiway Analysis of Variance 299
 - 14.7 Power and Sample Size in Multiway Analysis of Variance 300
Exercises 300
- 15 NESTED (HIERARCHICAL) ANALYSIS OF VARIANCE 303**
- 15.1 Nesting within One Main Factor 305
 - 15.2 Nesting in Factorial Experiments 308
 - 15.3 Multiple Comparisons and Confidence Intervals 310
 - 15.4 Power and Sample Size in Nested Analysis of Variance 311
Exercise 311
- 16 MULTIVARIATE ANALYSIS OF VARIANCE 312**
- 16.1 The Multivariate Normal Distribution 313
 - 16.2 Multivariate Analysis of Variance Hypothesis Testing 316
 - 16.3 Further Analysis 322
 - 16.4 Other Experimental Designs 322
Exercises 323
- 17 SIMPLE LINEAR REGRESSION 324**
- 17.1 Regression vs. Correlation 324
 - 17.2 The Simple Linear Regression Equation 326
 - 17.3 Testing the Significance of a Regression 333
 - 17.4 Confidence Intervals in Regression 337
 - 17.5 Inverse Prediction 342
 - 17.6 Interpretations of Regression Functions 344
 - 17.7 Regression with Replication and Testing for Linearity 345
 - 17.8 Power and Sample Size in Regression 350
 - 17.9 Regression through the Origin 351
 - 17.10 Data Transformations in Regression 353
 - 17.11 The Effect of Coding 357
Exercises 358

- 18 COMPARING SIMPLE LINEAR REGRESSION EQUATIONS 360**
- 18.1 Comparing Two Slopes 360
 - 18.2 Comparing Two Elevations 364
 - 18.3 Comparing Points on Two Regression Lines 368
 - 18.4 Comparing more than Two Slopes 369
 - 18.5 Comparing more than Two Elevations 372
 - 18.6 Multiple Comparisons among Slopes 372
 - 18.7 Multiple Comparisons among Elevations 373
 - 18.8 Multiple Comparisons of Points among Regression Lines 374
 - 18.9 An Overall Test for Coincidental Regressions 375
 - Exercises 375
- 19 SIMPLE LINEAR CORRELATION 377**
- 19.1 The Correlation Coefficient 377
 - 19.2 Hypotheses about the Correlation Coefficient 381
 - 19.3 Confidence Intervals for the Population Correlation Coefficient 383
 - 19.4 Power and Sample Size in Correlation 385
 - 19.5 Comparing Two Correlation Coefficients 386
 - 19.6 Power and Sample Size in Comparing Two Correlation Coefficients 388
 - 19.7 Comparing more than Two Correlation Coefficients 390
 - 19.8 Multiple Comparisons among Correlation Coefficients 392
 - 19.9 Rank Correlation 395
 - 19.10 Weighted Rank Correlation 398
 - 19.11 Correlation for Dichotomous Nominal-Scale Data 401
 - 19.12 Intraclass Correlation 404
 - 19.13 Concordance Correlation 407
 - 19.14 The Effect of Coding 410
 - Exercises 410
- 20 MULTIPLE REGRESSION AND CORRELATION 413**
- 20.1 Intermediate Computational Steps 414
 - 20.2 The Multiple Regression Equation 419
 - 20.3 Analysis of Variance of Multiple Regression or Correlation 422
 - 20.4 Hypotheses Concerning Partial Regression Coefficients 424
 - 20.5 Standardized Partial Regression Coefficients 426
 - 20.6 Partial Correlation 426
 - 20.7 Round-off Error and Coding Data 428
 - 20.8 Selection of Independent Variables 429
 - 20.9 Predicting Y Values 433
 - 20.10 Testing Difference between Two Partial Regression Coefficients 436
 - 20.11 "Dummy" Variables 436
 - 20.12 Interaction of Independent Variables 437
 - 20.13 Comparing Multiple Regression Equations 437
 - 20.14 Multiple Regression through the Origin 440
 - 20.15 Nonlinear Regression 440
 - 20.16 Descriptive vs. Predictive Models 442
 - 20.17 Concordance: Rank Correlation among Several Variables 443
 - Exercises 450
 - 21.3 Quadratic Regression 457
 - Exercises 459
- 21 POLYNOMIAL REGRESSION 452**
- 21.1 Polynomial Curve Fitting 452
 - 21.2 Round-off Error and Coding Data 457
- 22 TESTING FOR GOODNESS OF FIT 461**
- 22.1 Chi-Square Goodness of Fit 462
 - 22.2 Chi-Square Goodness of Fit for More than Two Categories 464
 - 22.3 Subdividing Chi-Square Analyses 466
 - 22.4 Chi-Square Correction for Continuity 468
 - 22.5 Bias in Chi-Square Calculations 470
 - 22.6 Heterogeneity Chi-Square 471
 - 22.7 The Log-Likelihood Ratio 473
 - 22.8 Kolmogorov-Smirnov Goodness of Fit for Discrete Data 475
 - 22.9 Kolmogorov-Smirnov Goodness of Fit for Continuous Data 478
 - 22.10 Sample Size Required for Kolmogorov-Smirnov Goodness of Fit for Continuous Data 481
 - Exercises 483

23 CONTINGENCY TABLES 486

- 23.1 Chi-Square Analysis of Contingency Tables 488
- 23.2 Graphing Contingency Table Data 490
- 23.3 The 2×2 Contingency Table 491
- 23.4 Heterogeneity Testing of 2×2 Tables 500
- 23.5 Subdividing Contingency Tables 502
- 23.6 Bias in Chi-Square Contingency Table Analyses 504

24 MORE ON DICHOTOMOUS VARIABLES 516

- 24.1 Binomial Probabilities 517
- 24.2 The Hypergeometric Distribution 523
- 24.3 Sampling a Binomial Population 524
- 24.4 Confidence Limits for Population Proportions 527
- 24.5 Goodness of Fit for the Binomial Distribution 530
- 24.6 The Binomial Test 533
- 24.7 The Sign Test 538
- 24.8 Power of the Binomial and Sign Tests 539

25 TESTING FOR RANDOMNESS 571

- 25.1 Poisson Probabilities 571
- 25.2 Confidence Limits for the Poisson Parameter 574
- 25.3 Goodness of Fit of the Poisson Distribution 575
- 25.4 The Binomial Test Revisited 578
- 25.5 Comparing Two Poisson Counts 582
- 25.6 Serial Randomness of Nominal-Scale Categories 583

26 CIRCULAR DISTRIBUTIONS: DESCRIPTIVE STATISTICS 592

- 26.1 Data on a Circular Scale 592
- 26.2 Graphical Presentation of Circular Data 595
- 26.3 Sines and Cosines of Circular Data 597
- 26.4 The Mean Angle 599
- 26.5 Angular Dispersion 602
- 26.6 The Median and Modal Angles 605
- 26.7 Confidence Limits for the Population Mean and Median Angles 605

27 CIRCULAR DISTRIBUTIONS: HYPOTHESIS TESTING 616

- 27.1 Testing Significance of the Mean Angle: Unimodal Distributions 616
- 27.2 Testing Significance of the Median Angle: Omnibus Test 621
- 27.3 Testing Significance of the Median Angle: Binomial Test 624
- 27.4 Testing Symmetry around the Median Angle 624
- 27.5 Two-Sample and Multisample Testing of Mean Angles 625

- 23.7 The Log-Likelihood Ratio for Contingency Tables 505
- 23.8 Three-Dimensional Contingency Tables 506
- 23.9 Log-Linear Models for Multidimensional Contingency Tables 512
- Exercises 514

- 24.9 Confidence Interval for the Population Median 542
- 24.10 The Fisher Exact Test 543
- 24.11 Comparing Two Proportions 555
- 24.12 Power and Sample Size in Comparing Two Proportions 558
- 24.13 Comparing more than Two Proportions 562
- 24.14 Multiple Comparisons for Proportions 563
- 24.15 Trends among Proportions 565
- Exercises 568

- 25.7 Serial Randomness of Measurements: Parametric Testing 586
- 25.8 Serial Randomness of Measurements: Nonparametric Testing 587
- Exercises 590

- 26.8 Diametrically Bimodal Distributions 607
- 26.9 Second-Order Analysis: The Mean of Mean Angles 608
- 26.10 Confidence Limits for the Second-Order Mean Angle 611
- Exercises 614

- 27.6 Nonparametric Two-Sample and Multisample Testing of Angles 630
- 27.7 Two-Sample and Multisample Testing of Median Angles 635
- 27.8 Two-Sample and Multisample Testing of Angular Distances 635
- 27.9 Two-Sample and Multisample Testing of Angular Dispersion 637
- 27.10 Parametric One-Sample Second-Order Analysis of Angles 638

27 CIRCULAR DISTRIBUTIONS: HYPOTHESIS TESTING (continued)

- | | |
|--|---|
| 27.11 Nonparametric One-Sample Second-Order Analysis of Angles 639 | 27.16 Parametric Angular Correlation and Regression 649 |
| 27.12 Parametric Two-Sample Second-Order Analysis of Angles 641 | 27.17 Nonparametric Angular Correlation 653 |
| 27.13 Nonparametric Two-Sample Second-Order Analysis of Angles 643 | 27.18 Goodness of Fit Testing for Circular Distributions 654 |
| 27.14 Parametric Paired-Sample Testing with Angles 645 | 27.19 Serial Randomness of Nominal-Scale Categories on a Circle 658 |
| 27.15 Nonparametric Paired-Sample Testing with Angles 647 | Exercises 660 |

APPENDIX A ANALYSIS OF VARIANCE HYPOTHESIS TESTING App1

- | | |
|---|--|
| A.1 Determination of Appropriate F 's and Degrees of Freedom App1 | A.4 Nested Analysis of Variance App7 |
| A.2 Two-Factor Analysis of Variance App5 | A.5 Split-Plot and Mixed Within-Subjects Analysis of Variance App6 |
| A.3 Three-Factor Analysis of Variance App6 | |

APPENDIX B STATISTICAL TABLES AND GRAPHS App11

- | | |
|---|--|
| Table B.1 Critical Values of Chi-Square Distribution App12 | Table B.16 Critical Values of Q' for Nonparametric Multiple Comparison Testing with a Control App108 |
| Table B.2 Proportions of the Normal Curve (One-Tailed) App17 | Table B.17 Critical Values of the Correlation Coefficient, r App109 |
| Table B.3 Critical Values of the t Distribution App19 | Table B.18 Fisher's z Transformation for Correlation Coefficients, r App111 |
| Table B.4 Critical Values of the F Distribution App21 | Table B.19 Correlation Coefficients, r , Corresponding to Fisher's z Transformation App113 |
| Table B.5 Critical Values of the q Distribution App58 | Table B.20 Critical Values of the Spearman Rank Correlation Coefficient, r_s App116 |
| Table B.6 Critical Values of q' for the One-Tailed Dunnett's Test App74 | Table B.21 Critical Values of the Top-Down Correlation Coefficient, r_T App118 |
| Table B.7 Critical Values of q' for the Two-Tailed Dunnett's Test App76 | Table B.22 Critical Values of the Symmetry Measure, g_1 App119 |
| Table B.8 Critical Values of d_{\max} for the Kolmogorov-Smirnov Goodness of Fit Test for Discrete or Grouped Data App77 | Table B.23 Critical Values of the Kurtosis Measure, g_2 App121 |
| Table B.9 Critical Values of D for the Kolmogorov-Smirnov Goodness of Fit Test for Continuous Distributions App83 | Table B.24 The Arcsine Transformation, p' App124 |
| Table B.10 Critical Values of D_δ for the δ -Corrected Kolmogorov-Smirnov Goodness of Fit Test for Continuous Distributions App87 | Table B.25 Proportions, p , Corresponding to Arcsine Transformations, p' App127 |
| Table B.11 Critical Values of the Mann-Whitney U Distribution App89 | Table B.26a Binomial Coefficients, ${}_nC_X$ App129 |
| Table B.12 Critical Values of the Wilcoxon T Distribution App101 | Table B.26b Proportions of the Binomial Distribution for $p = q = 0.5$ App132 |
| Table B.13 Critical Values of the Kruskal-Wallis H Distribution App104 | Table B.27 Critical Values of C for the Sign Test or for the Binomial Test with $p = 0.5$ App133 |
| Table B.14 Critical Values of the Friedman χ_r^2 Distribution App106 | Table B.28 Critical Values for Fisher's Exact Test App143 |
| Table B.15 Critical Values of Q for Nonparametric Multiple Comparison Testing App107 | Table B.29 Critical Values for Runs Test App171 |
| | Table B.30 Critical Values of C for the Mean Square Successive Difference Test App180 |
| | Table B.31 Critical Values for the Runs Up and Down Test App182 |

APPENDIX B STATISTICAL TABLES AND GRAPHS (continued)

Table B.32 Angular Deviation, s , As a Function of Vector Length, r App184

Table B.33 Circular Standard Deviation, s_0 , As a Function of Vector Length, r App186

Table B.34 Critical Values of Rayleigh's z App188

Table B.35 Critical Values of u for the V Test of Circular Uniformity App190

Table B.36 Critical Values of m for the Hodges-Ajne Test App191

Table B.37 Correction Factor, K , for the Watson and Williams Test App193

Table B.38 Critical Values of Watson's U^2 App195

Table B.39 Critical Values of R' for the Moore Test of Circular Uniformity App198

Table B.40 Common Logarithms of Factorials App199

Table B.41 Ten Thousand Random Digits App201

Figure B.1 Power and Sample Size in Analysis of Variance App205

ANSWERS TO EXERCISES *Ans1*

LITERATURE CITED *L1*

INDEX *I1*