Contents

Preface vii

Chapter 1 Descriptive Statistics 1

1.1. Introduction 1
1.2. Measures of Location 3
1.2.1. Mean 3
1.2.1.1. Some properties of mean 5
1.2.2. Median 6
1.2.3. Mode 7
1.3. Measures of Dispersion 8
1.3.1. Range 9
1.3.2. Variance and standard deviation 9
1.3.2.1. Some properties of variance and standard deviation 10
1.3.2.2. Two alternative formulas for the variance 11
1.3.3. Covariance $\mathrm{Cov}(X, Y)$, or $S_{X,Y}$ 11
1.4. Grouped Data 12
1.4.1. Computational formulas for grouped data 13
1.5. Graphics 15
1.5.1. Histogram 16

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

1.5.2. Relative frequency polygon 16
1.5.3. Cumulative frequency polygon 17
1.5.4. Percentiles and quartiles and interquartile range . 18
1.6. An Example 20
1.7. Proofs of the Results in this Chapter 22
1.8. Exercises and Problems 23

Chapter 2 Probability 25

2.1. Basic Concepts 25
2.1.1. Three components of probability 25
2.1.2. Definition of probability 26
2.1.3. Values of a probability 29
2.1.4. Sure event and impossible event 29
2.1.5. Complement of an event (or negation of an event) 29
2.2. Composite Event "A and B" 30
2.2.1. Conditional probability 32
2.2.2. Multiplication theorem 34
2.2.3. Bayes' theorem 35
2.2.4. Independence of events 37
2.2.5. A theorem of pairwise independence 38
2.2.6. Multiplication theorem when events are independent 39
2.3. Composite Event "A or B" 40
2.3.1. Addition theorem 40
2.3.2. Composite events of complements 41
2.3.3. De Morgan's laws 43
2.3.4. Addition theorem when events are mutually exclusive 44
2.4. Remarks on the Addition and Multiplication Theorems . . 44
2.4.1. Summary of the two theorems 44
2.4.2. The sssociative and distributive laws 44
2.5. Factorials, Permutations, and Combinations 46
2.5.1. Permutations 46
2.5.2. Combinations 48
2.6. Proofs of the Results in this Chapter 50
2.7. Exercises and Problems 51

Chapter 3 Random Variables 55

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

3.1. Definition of a Random Variable 55
3.1.1. Illustrative examples 56
3.2. Joint Probability Distribution 57
3.2.1. Marginal probability 58
3.2.2. Conditional probability 59
3.2.3. Independence of random variables 59
3.3. Expectation of a Random Variable 62
3.3.1. Some properties of the expectation of a random variable 65
3.3.2. Expectation of a linear function of random variables 66
3.3.3. Expectation of a sample mean 67
3.4. Variance of a Random Variable 67
3.4.1. Two properties of the variance 69
3.4.2. Variance of a linear function of independent random variables 70
3.4.3. Variance of sample mean 70
3.4.4. Expectation of sample variance 71
3.5. Covariance of a Bivariate Distribution 71
3.5.1. Some properties of the covariance 72
3.5.2. The variance of a linear function of random variables 72
3.6. Continuous Random Variables 74
3.6.1. Joint probability distribution of two random variables 74
3.7. Proofs of the Results in this Chapter 76
3.8. Exercises and Problems 78

Chapter 4 Probability Distributions 81

4.1. Introduction 81
4.2. Uniform Distribution (Discrete) 82
4.3. The Binomial Distribution, $B(n; p)$ 83
4.4. The Hypergeometric Distribution 91
4.4.1. Relation with the binomial distribution 93
4.4.2. Applications of the hypergeometric distribution . . 94
4.5. The Poisson Distribution 96
4.6. The Multinomial Distribution 99
4.6.1. The covariance between X_i and X_j 100

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

4.7. Continuous Uniform Distribution 103
4.8. The Exponential Distribution 105
4.9. The Normal Distribution 107
4.9.1. Density function and distribution function 109
4.9.2. The standard normal distribution $N(0,1)$ 111
4.10. Proofs of the Results in this Chapter 114
4.11. Exercises and Problems 121

Chapter 5 Statistical Inference—Interval Estimation 127

5.1. Introduction 127
5.1.1. The central limit theorem 129
5.1.2. Normal approximation for the binomial distribution 133
5.2. Interval Estimation of Population Means 136
5.2.1. Interval estimation of a population mean when the standard deviation is known 136
5.2.1.1. Length of the confidence interval 139
5.2.2. Interval estimation of a population mean when the standard deviation is unknown 141
5.2.3. Confidence interval for the difference between two population means when population variances are known 143
5.2.4. Confidence interval of the difference between two population means when the population variances are unknown 144
5.3. Confidence Intervals for Population Proportions 146
5.3.1. Confidence intervals for a single population proportion, p 146
5.3.2. Confidence interval for the difference between two population proportions 147
5.4. Exercises and Problems 147

Chapter 6 Hypothesis Testing — Fundamental Concepts 151

6.1. Introduction 151
6.2. Basic Elements in Hypothesis Testing 153
6.2.1. Statistical hypothesis 153
6.2.1.1. Simple hypothesis versus composite hypothesis 154

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

6.2.2. Statistical test and two hypotheses 154
6.2.3. Two kinds of errors 156
6.2.4. Level of significance and power of test 157
6.2.5. Statistic (or test statistic) and critical ratio 158
6.2.6. Critical region and region of acceptance 159
6.2.7. Computation of power, $1 - \beta$ 159
6.3. Illustrative Examples of Hypothesis Testing 160
6.4. Power of Test 168
6.4.1. Determination of sample size n 169
6.5. Power of Test Concerning a Single Probability 171
6.6. Exercises and Problems 172

Chapter 7 Testing Hypotheses Concerning Population Means and Population Proportions 175

7.1. Introduction 175
7.2. The "Student's" t Distribution 176
7.2.1. One-sample case 176
7.2.2. Two-sample case 178
7.3. Procedure for Hypothesis Testing 180
7.4. Test of Hypotheses Concerning Population Means 181
7.4.1. Test of hypotheses concerning a single population mean 182
7.4.2. Test of hypotheses concerning two population means 183
7.4.3. Two populations — before and after treatment . . 184
7.4.4. Some remarks regarding hypothesis testing 188
7.5. Test of Hypotheses Concerning Probabilities 191
7.5.1. Hypothesis testing concerning a single probability 192
7.5.2. Hypothesis testing concerning two probabilities . . 193
7.6. Exercises and Problems 196

Chapter 8 The Chi-Square Test 203

8.1. Introduction 203
8.1.1. The chi-square distribution 204
8.2. Some Theorems Concerning the Chi-Square Distribution . 205
8.3. Chi-Square Tests 207
8.3.1. One-way classification 207

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

8.3.1.1. The chi-square test versus the normal Z test I . . . 211
8.3.2. Two-way classification . . . 212
8.3.2.1. Conditional distribution of Y given X . . 216
8.3.2.2. The 2×2 contingency table . . . 220
8.3.2.3. The chi-square test versus the normal Z test II . . . 221
8.3.2.4. Case-control studies . . . 223
8.3.2.5. Relative risk and odds ratio . . . 226
8.3.3. Goodness of fit . . . 227
8.4. More Remarks about the Chi-Square Test . . . 231
8.5. Proof of the Results in this Chapter . . . 233
8.6. Exercises and Problems . . . 235

Chapter 9 Linear Regression 243

9.1. Introduction . . . 243
9.2. Estimation of a and b . . . 245
9.3. Underlying Assumptions . . . 252
9.4. Relevant Theorems . . . 254
9.5. Statistical Inference in Linear Regression . . . 256
9.5.1. Hypothesis testing regarding b . . . 256
9.5.2. Confidence interval for the expectation $E(Y|X_0) = a + bX_0$ . . . 258
9.5.3. Confidence belt for the regression line $E(Y|X) = a + bX$ . . . 259
9.6. Test of the Linear Regression Assumption . . . 265
9.6.1. The F distribution . . . 266
9.6.2. The F test of linear regression . . . 267
9.6.3. Test of hypothesis $b = 0$ . . . 268
9.7. Remarks . . . 272
9.7.1. A linear model . . . 272
9.7.2. Choice of values of X . . . 273
9.7.3. Interval of prediction for an individual Y . . . 273
9.7.4. Interpolation versus extrapolation . . . 274
9.8. Some Other Forms of Regression . . . 274
9.8.1. Curve-linear regression . . . 274
9.8.2. Logistic regression . . . 275
9.9. Proofs of the Results in this Chapter . . . 276

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

9.10. Exercises and Problems 278

Chapter 10 Correlation 281

10.1. The Correlation Coefficient, ρ_{XY} (or ρ) 281
10.1.1. The sample correlation coefficient, r_{XY} 283
10.2. Relationship Between Correlation and Regression 284
10.3. The Bivariate Normal Distribution 288
10.4. Statistical Inference About ρ_{XY} 291
10.4.1. Test of zero correlation 291
10.4.2. Confidence interval for ρ_{XY} 293
10.4.3. Spurious correlation? 296
10.5. Other Types of Correlation 300
10.5.1. Rank correlation coefficient (Spearman's correlation coefficient), r_s 300
10.5.2. Point biserial correlation coefficient, r_{pb} 302
10.5.3. The Φ coefficient 304
10.6. Proofs of the Results in this Chapter 306
10.7. Exercises and Problems 314

Chapter 11 Multiple Regression and Correlation 319

11.1. The Regression Equation 319
11.2. Estimation of a, b_1, b_2 320
11.3. Theorems and Inferences 323
11.4. Multiple Correlation Coefficient, $\rho_{y\cdot 12}$ 329
11.4.1. Test of hypothesis concerning true correlation coefficient $\rho_{y\cdot 12}$ 330
11.5. Partial Correlation 332
11.6. A General Multiple Regression 335
11.6.1. The regression equation and estimation of the constants 335
11.6.2. Some theorems for statistical inference 338
11.6.3. Multiple correlation coefficient, $\rho_{y\cdot 12\cdots s}$ 340
11.7. Multiple Logistic Regression 341
11.8. Proofs of the Results in this Chapter 342
11.9. Exercises and Problems 345

Chapter 12 One-Way Analysis of Variance 349

12.1. Introduction 349

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

12.2. One-Way Analysis of Variance 350
12.2.1. The layout and linear model 350
12.2.2. Least-squares estimates, sums of squares (ssq), and mean squares (msq) 352
12.2.3. Null hypothesis and test statistic 355
12.3. Examples 358
12.4. Multiple Comparisons 362
12.4.1. The least significant difference (LSD) 362
12.4.2. The studentized range 364
12.4.3. Contrasts 365
12.4.3.1. Bonferroni's inequality 366
12.4.4. The t test 368
12.5. Random-Effects Model 373
12.5.1. Partition of sums of squares 375
12.5.2. The null hypothesis 376
12.6. A Comparison of Two Models 378
12.7. Proofs of the Results in this Chapter 380
12.8. Exercises and Problems 385

Chapter 13 Two-Way Analysis of Variance — Fixed-Effects Models 389

13.1. Introduction 389
13.2. The Linear Model 391
13.3. One Observation Per Cell 394
13.3.1. Least squares estimates and sums of squares (ssq) 395
13.3.2. Null hypotheses and test statistics 396
13.3.3. Multiple comparisons 403
13.4. n Observations Per Cell 404
13.4.1. Linear model 406
13.4.2. Least squares estimates and sums of squares (ssq) 407
13.4.3. Null hypotheses and test statistics 408
13.4.4. Multiple comparisons 414
13.5. Proofs of Results in this Chapter 420
13.6. Exercises and Problems 425

Chapter 14 Two-Way Analysis of Variance Random-Effects Models and Mixed Model 429

14.1. Introduction 429

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

14.2. One Observation Per Cell — Random-Effects Model . . . 430
14.2.1. Sums of squares and mean squares 431
14.2.2. Null hypotheses and test statistics 432
14.3. n Observations Per Cell — Random-Effects Model 434
14.3.1. Sums of squares and mean squares 435
14.3.2. Null hypotheses and test statistics 437
14.4. A Comparison of Two Models — Fixed-Effects Model versus Random-Effects Model 440
14.4.1. Expectations and variances 440
14.4.2. Statistical inference and test statistics 442
14.5. Mixed Model 446
14.5.1. Effects and interactions 447
14.5.2. Estimation of effects and sums of squares 448
14.5.3. Expectations of mean squares 449
14.5.4. Hypotheses and test statistics 451
14.6. Proofs of the Results in this Chapter 456
14.7. Exercises and Problems 464

Chapter 15 Design of Experiment 467

15.1. Introduction 467
15.1.1. Randomization 469
15.1.2. Replication 469
15.2. Completely Randomized Design 470
15.2.1. Table of random numbers 471
15.2.2. Method of analysis 472
15.3. Randomized Blocks — Single Blocking 473
15.3.1. Some remarks 473
15.3.2. An illustration 474
15.3.3. Linear model and F ratios 475
15.3.4. Efficiency of blocking 476
15.3.4.1. Measurement of efficiency 477
15.3.4.2. Efficiency expressed in F ratio 479
15.4. Latin Squares — Double Blocking 479
15.4.1. Examples 481
15.4.2. Symmetry of the three factors 483
15.4.3. Replication of Latin squares 484
15.4.4. Observations in a Latin square 486
15.4.5. Linear model and hypotheses 487

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

15.4.6. Sums of squares and test statistics 489
15.4.7. Analysis of data — an example 492
15.4.8. Efficiency of Latin square 496
15.4.8.1. Preliminaries 496
15.4.8.2. Efficiency relative to randomized blocks 498
15.4.8.3. Efficiency relative to complete randomization 499
15.5. Graeco-Latin Square — Triple Blocking 503
15.6. Proofs of the Results in this Chapter 507
15.7. Exercises and Problems 514

Chapter 16 Analysis of Covariance 519

16.1. Introduction 519
16.2. Linear Model, Hypotheses, and Regression Equations . . . 522
16.2.1. Null hypotheses 522
16.2.2. Regression equations and regression lines 524
16.3. Partition of Sum of Squares and F Statistics 530
16.3.1. The F statistics 532
16.3.1.1. Test of hypothesis $H_1 : E(Y_i|X_i) = a_{iw} + b_w X_i$ 532
16.3.1.2. Test of hypothesis $H_2 : E(Y_i|X_i) = a_t + b_t X_i$ 533
16.3.1.3. Test of hypothesis $H_3 : E(\bar{Y}_i|\bar{X}_i) = a_b + b_b \bar{X}_i$ 533
16.3.1.4. Test of hypothesis $H_4 : b_w = b_b$ 535
16.3.1.5. Test of hypothesis $H_5 : b_w = 0$ 536
16.4. Computation Formulas of the Sums of Squares 537
16.5. An Example — Oral Contraceptives and Metabolic Rate in Women 540
16.6. Proofs of the Results in this Chapter 544
16.7. Exercises and Problems 549

Chapter 17 Non-Parametric Statistics 555

17.1. Introduction 555
17.2. Sign Test 556
17.3. Wilcoxon Signed Rank Test 558
17.3.1. Critical values of R^+ 558

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html

17.3.2. Normal approximation 560
17.4. Wilcoxon Two-Sample Rank-Sum Test 560
17.4.1. Normal approximation 563
17.4.2. Test for equality of two variances 564
17.5. Mann–Whitney Test 565
17.6. Kruskal–Wallis Test — One-Way Analysis of Variance . . 566
17.7. Friedman Test — Two-Way Analysis of Variance 569
17.8. Proofs of the Results in this Chapter 574
17.9. Exercises and Problems 580

Appendix: Tables 587

Bibliography 599

Hints and Answers to Selected Exercises
by T. Porco 601

Index 617

STATISTICAL METHODS OF ANALYSIS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/5224.html