

Contents

1	Introduction	1
1.1	Concepts	1
1.1.1	Why Study Statistics?	2
1.2	Methods of Describing Data	3
1.2.1	Types of Data	3
1.3	Measurements	4
1.3.1	Measuring Devices	5
1.3.2	Standardization of Measuring Devices	7
1.3.3	Variability in Measurements	7
1.3.4	Bias in Measurement	8
1.3.5	Error in Measurement	8
1.4	Exercises	9
2	Frequency Distributions	11
2.1	Introduction	11
2.2	Frequency Distributions	12
2.2.1	Ungrouped Distribution	12
2.2.2	Grouped Distribution	13
2.2.3	Constructing a Frequency Distribution	14
2.2.4	Other Forms of Frequency Distribution	15
2.2.5	Cumulative Frequency Distributions	17
2.3	Graphical Representation of Data	18
2.3.1	The Dotplot	18
2.3.2	The Stem and Leaf Display	18
2.3.3	Histograms	21
2.3.4	Polygons	23
2.3.5	Cumulative Frequency Polygon - The Ogive	23
2.4	Presentation of Data: Charts and Diagrams	24
2.4.1	The Bar Chart	25
2.4.2	Multiple Bar Chart	26
2.4.3	Component Bar Chart	28
2.4.4	Pie Charts	28
2.5	Exercises	30

3	Numerical Description of Data	33
3.1	Introduction	33
3.1.1	Properties OF $\sum$	34
3.2	Measures of Center or Central Tendency	34
3.2.1	The Mean, $\bar{x}$	35
3.3	Weighted Means	36
3.3.1	Geometric Mean	36
3.3.2	Mean of Grouped Data	36
3.3.3	The Median	37
3.4	Percentiles	39
3.4.1	Quartiles	40
3.4.2	Checking for Outliers with Quartiles	40
3.5	The Boxplot	41
3.5.1	Mean of Grouped Data	42
3.5.2	The Median of Grouped Data	43
3.5.3	The Mode	44
3.5.4	Comparisons Between Mean, Median, and Mode	45
3.6	Relationship Between Mean, Median, and Mode	46
3.7	Measures of Variation or Dispersion	49
3.7.1	The Range	49
3.7.2	Variance and Standard Deviation	50
3.7.3	Case I: Variance of Ungrouped Data	50
3.7.4	Calculating the Variance of Grouped Data	52
3.7.5	Use of Coding to Simplify Calculations	53
3.8	Empirical Rule	54
3.9	Exercises	55
4	Probability and Probability Distributions	59
4.1	Introduction	59
4.2	Counting Methods	60
4.2.1	Permutation	60
4.2.2	Combinations	62
4.2.3	Probability of an Event	63
4.3	Marginal & Conditional Probabilities	65
4.4	Laws of Probability	66
4.4.1	The Addition Law of Probability	66
4.4.2	Multiplication Law	67
4.5	Relationship Between Probability and Odds	69
4.6	Specificity, Sensitivity of Tests	69
4.7	Receiver Operating Characteristics(ROC) Curves	74
4.7.1	Example	75
4.8	Probability Distributions	77
4.8.1	Mean and Variance of X	78
4.8.2	Cumulative Probability Distribution Function	79
4.9	The Binomial Distribution	80

4.10	The Poisson Distribution	85
4.10.1	Fitting of Poisson Distribution to a Sample of Data	86
4.10.2	Use of Recursion Formula	87
4.11	The Normal Distribution	88
4.11.1	Areas Under the Normal curve	89
4.12	Normal Approximations to the Binomial	94
4.13	The Hypergeometric Distribution	96
4.14	Sampling Distributions	98
4.14.1	Introduction	98
4.14.2	Simulation	103
4.14.3	Sampling Distribution of $\bar{x}$: A Summary	104
4.14.4	Sampling Distribution of Population Proportion	106
4.15	Exercises	106
5	Estimation and Hypotheses Testing	115
5.1	Confidence Intervals	115
5.1.1	Building a Confidence Interval	116
5.1.2	Sample Size Determination	118
5.1.3	Case of σ Not Known and n Large ($n \geq 30$)	119
5.1.4	Case of σ Not Known and n Small ($n < 30$)	119
5.1.5	Confidence Interval for a Population Proportion	121
5.1.6	Sample Size Determination for Estimating a Proportion	121
5.2	Confidence Interval for the Difference of Two Population Means	123
5.2.1	Distribution of Difference of Two Means	123
5.2.2	Case When $(n_1, n_2) < 30$ and σ_1, σ_2 Are Unknown	125
5.3	Confidence Interval for the Difference of Two Population Proportions	127
5.4	Hypothesis Testing	128
5.4.1	Concepts and Definitions	128
5.4.2	Test of Significance	130
5.4.3	Example Seed	131
5.4.4	The Level of Significance	133
5.4.5	Types of Alternative Hypotheses	134
5.5	Tests for Means and Proportion	135
5.5.1	p Values	136
5.5.2	Testing for a Binomial Proportion	141
5.6	Tests Concerning Two Population Means	143
5.6.1	Testing Differences Between Two Population Means	143
5.7	The Mann Whitney U -Test	150
5.8	Comparison of Two Binomial Proportions	152
5.9	Paired T -test	154
5.10	Chapter Summary	157
5.11	Exercises	160

6	Analysis of Variance (ANOVA)	169
6.1	Introduction	169
6.1.1	Analysis of Variance of Table 6.1	170
6.1.2	Assumptions of the One-Way ANOVA Model	171
6.1.3	An Example	172
6.1.4	Extending the Two-Sample t Test	175
6.2	Multiple Comparisons Procedures	175
6.2.1	Fisher's Least Significant Difference (LSD)	176
6.2.2	Experiment-Wise Error Rate (EER)	178
6.2.3	The Tukey Test	178
6.3	Contrasts	181
6.4	Partitioning the Treatments SS	182
6.5	Tests of Homogeneity of Variances	185
6.5.1	Bartlett's Test for Data in Table 6.3	186
6.6	Nonparametric Test	189
6.7	ANOVA with Unequal Replication	191
6.8	One Factor with Quantitative Levels	194
6.9	Orthogonal Polynomials for Unequal Spacing	198
6.10	Two-Factor Analysis of Variance	200
6.10.1	An Example: Bacteria Counts	200
6.11	Replication and Sample Size Determination	202
6.11.1	One- or Two-Sample t Test	203
6.11.2	One- and Two-Sample Z Test Example	205
6.11.3	Two-Sample t Test Example	207
6.11.4	Sample Size in One-Factor ANOVA	208
6.12	Exercises	211
7	Regression Analysis	217
7.1	Introduction	217
7.2	Model Assumptions	218
7.3	Estimating the Parameters of the Simple Model	219
7.3.1	The Ordinary Least Squares (OLS) Method	219
7.3.2	Interpretations of Parameter Estimates	225
7.4	Inferences on Parameter Estimates	225
7.4.1	Confidence Interval for β_0 and β_1	227
7.4.2	Confidence Interval Estimation for β_0	229
7.5	Residuals	230
7.6	Prediction of Y from X	231
7.6.1	Mean Prediction	231
7.6.2	Individual Prediction	232
7.6.3	Percentage Variation	233
7.7	Adequacy of the Regression Model	234
7.7.1	Examinations of Residuals	235
7.8	Correlation Coefficient	241
7.8.1	Properties of r	241

7.8.2	General Hypotheses Concerning ρ	242
7.9	Multiple Regression	243
7.9.1	Example	244
7.9.2	Partial F tests	247
7.10	Outliers and Influential Observations	249
7.10.1	Example	250
7.10.2	Multicollinearity	251
7.11	Rank Correlation	252
7.12	Concordance Correlation	253
7.13	Multiple and Partial Correlations	254
7.13.1	Partial Correlations	255
7.14	Comparisons of Regressions	257
7.14.1	Example	258
7.14.2	Alternative Approach	261
7.15	Fitting Parallel Lines	265
7.16	Nonlinear Regression	269
7.16.1	Example 7.16.1	277
7.16.2	Example 7.16.3: Exponential Response Model	284
7.16.3	Polynomial Model	286
7.17	Other Special Nonlinear Models	295
7.17.1	The Logistic Growth Model	296
7.17.2	The Gompertz Growth Model	297
7.18	Polynomial Regressions	298
7.19	Exercises	300
8	Categorical Data Analysis	307
8.1	Introduction	307
8.1.1	Tests of Goodness of Fit	307
8.2	The 2×2 Contingency Table	309
8.3	Fisher's Exact Test	312
8.4	Combining Several 2×2 Tables	315
8.5	The General $R \times C$ Contingency Table	319
8.6	Fitting the Poisson Distribution	326
8.7	Fitting the Binomial Distribution	327
8.7.1	Example	328
8.8	Exercises	330
9	Experimental Design	337
9.1	Introduction	337
9.2	Experimental Design	339
9.2.1	Planning, Design, and Layout	340
9.2.2	Management	340
9.2.3	Data Recording	341
9.2.4	Scrutiny and Editing of Data	341
9.3	Principles of Experimentation	341

9.3.1	Randomization	341
9.3.2	Replication	342
9.3.3	Blocking	343
9.4	Methods of Increasing the Accuracy of an Experiment	346
9.5	Random and Fixed Effect Models	347
9.6	Control of Error	347
9.7	Summary of Principles of Experimental Designs	348
9.8	Observational Studies	348
9.8.1	Prospective Studies	348
9.8.2	Relative Risk	349
9.8.3	Retrospective Studies	350
9.8.4	Odds Ratio	351
9.9	Exercises	354
10	The Completely Randomized Design	355
10.1	Introduction	355
10.1.1	Analysis of Variance of Table 10.2	357
10.2	Example 10.1	358
10.2.1	Students' t Test	360
10.3	Multiple Comparisons of Means	361
10.3.1	Duncan's Multiple Range Test	361
10.3.2	Tukey's Test	362
10.3.3	Scheffé's Test	362
10.4	The Unbalanced Case	367
10.4.1	Example 10.2	367
10.4.2	Analysis	367
10.5	Tests on Individual Treatment Means	369
10.5.1	Analysis	371
10.5.2	Computing Contrasts SS	373
10.6	Design with a Quantitative Treatment	376
10.6.1	Analysis of Variance for the Experiment	377
10.6.2	Use of Orthogonal Polynomials	378
10.7	Model Adequacy Checking	386
10.7.1	Example 10.62	388
10.8	Exercises	390
11	The Randomized Complete Block Design	395
11.1	Introduction	395
11.1.1	Why RCBD?	398
11.1.2	Model and Analysis for the RCBD	402
11.1.3	Analysis	403
11.1.4	Calculation of the Error SS from Residuals	406
11.2	Missing Values in a RCBD Design	407
11.2.1	Summary of Results of Analysis	411
11.3	Partitioning Treatment SS in a RCBD	412

11.3.1	Analysis	419
11.1	Paired Comparisons	422
11.4.1	Analysis	424
11.5	Test for Non Additivity	425
11.5.1	Alternative Implementation of Tukey's Additivity Test	429
11.6	Departures from Assumptions in Analysis of Variance	430
11.6.1	Square Root Transformation	430
11.6.2	The Logarithmic Transformation	431
11.6.3	Arc Sine Transformation	431
11.6.4	Box-Cox Transformation	438
11.7	Relative Efficiency of RCBD	440
11.8	Group Balanced Block Design	440
11.8.1	Randomization	441
11.8.2	An Example	442
11.9	Exercises	446
12	Multiple Blocking Designs	449
12.1	The Latin or Euler Square Design	449
12.2	The Model for the Latin Square Designs	451
12.2.1	Analysis	453
12.3	Missing Values in Latin Square Designs	459
12.4	Graeco-Latin Square Designs	461
12.4.1	The Completely Orthogonalized Square	465
12.4.2	Relative Efficiencies of Latin Square Design	467
12.5	Multiple Latin Squares	468
12.5.1	Example	471
12.6	Crossover Designs	474
12.6.1	Crossover Design Analysis With Carryover Effects	484
12.6.2	Analysis of Variance for the Experiment	487
12.7	Exercises	498
13	Analysis of Covariance	503
13.1	Introduction	503
13.1.1	Model and Assumptions	505
13.2	Further Analysis	508
13.2.1	Standard Errors	508
13.3	Test for Parallelism of Regression Lines	509
13.3.1	Testing for Parallelism with MINITAB	512
13.4	Covariance Analysis in a RCBD	516
13.4.1	Factorial Treatment Case	520
13.5	Estimation of Missing Observations	523
13.5.1	Another Missing Value Example	527
13.6	Exercises	528

14	Factorial Treatments Designs	531
14.1	Definitions	531
14.1.1	Factorial Design	531
14.2	Experiments at Two Levels: The 2^n Series	532
14.2.1	Factorial Effects in the 2^2 Factorial	533
14.2.2	The 2^3 factorial system	538
14.2.3	Yates Algorithm	542
14.2.4	Factorial in Complete Blocks	546
14.3	The 3^n Factorial Designs	547
14.3.1	The 3^2 Factorial	548
14.3.2	An Example of a 3^2 Design	549
14.3.3	The 3^3 factorial Design	552
14.4	Other Factorial Systems	554
14.4.1	Summary of Results	568
14.5	Single Replicate Experiments	571
14.6	Confounding in the Factorial System	575
14.6.1	Replications in 2^n Confounding	579
14.7	Partial Confounding	585
14.7.1	Confounding in the 3^n Series	592
14.8	Fractional Replication	593
14.8.1	Constructing a 2^{n-1} Fractional Factorial Design	593
14.8.2	Calculating the Effects:	595
14.8.3	The One-Quarter Fraction of the 2^n Design: 2^{n-2}	596
14.8.4	An Example:	597
14.9	2^{n-p} Resolution III and IV Designs	599
14.10	Logistic Regression for a Factorial Study	600
14.11	Exercises	604
15	The Split-Plot Design	609
15.1	Introduction	609
15.1.1	Summary of Results	615
15.1.2	Summary of the Results of the Experiment	621
15.1.3	Missing Data in Split-Plot Design	621
15.2	Latin Square Split-Plot Example	622
15.3	The Split-Split-Plot Design	624
15.4	The Strip-Plot Design	630
15.5	Exercises	635
16	Incomplete Block Design	639
16.1	Introduction	639
16.2	Balanced Incomplete Block Design	641
16.3	Statistical Model for a Balanced Incomplete Block (BIB) Design	642
16.4	Constructing a Balanced Incomplete Block (BIB) Design	647
16.5	Efficiency of Incomplete Block Designs	648

16.6	Lattice Design	649
16.6.1	Construction of Lattice Design	649
16.7	Relative Efficiency for Lattice Design	657
16.8	Exercises	658
17	Quantal Bioassay	661
17.1	Introduction	661
17.2	The Logistic Regression Approach	666
17.3	Using the Probit Model	671
17.4	Parallel-Line Bioassay	674
17.5	Use of Joint Model	676
17.5.1	Example 17.3	679
17.5.2	HIV Status Data Example	683
17.5.3	Interpretation of Parameters	688
17.5.4	ROC Curve for the Analysis	689
17.6	Exercises	691
18	Repeated Measures Design	697
18.1	Introduction	697
18.2	Single-Factor Experiments with Repeated Measures	698
18.2.1	Correlation Within Subjects	703
18.3	Two Factors with Repeated Measures on One Factor	703
18.3.1	Calculations	707
18.3.2	Multivariate Approach	707
18.4	Exercises	716
19	Survival Analysis	719
19.1	Introduction	719
19.2	Censoring	719
19.3	Describing Event Times	720
19.4	Estimating the Survival Function $S(t)$	720
19.4.1	The Kaplan Meier Method	720
19.4.2	Computing Survival Probabilities	723
19.4.3	The Life Table Method	724
19.4.4	Another Example	729
19.5	Testing Survival Times Between Two Groups	732
19.6	Hazard Function	735
19.6.1	Types of Hazard Functions	736
19.7	Proportional Hazards Model	744
19.8	Exercises	745
20	Combined Analysis of Experimental Data	749
20.1	Introduction	749
20.1.1	General Analysis of Series of Experiments	750
20.2	Analysis of Experiments Over Seasons	750

20.2.1	Analysis	751
20.2.2	Combined Seasonal Analysis.....	752
20.2.3	Partitioning the Interaction SS	754
20.2.4	Effect of Failure of Homogeneity Assumption	757
20.2.5	Example with Homogeneity Assumption Violated ..	757
20.2.6	Combined Seasonal Analysis.....	758
20.3	Experiment at Several Sites	763
20.3.1	RCB Design Example	763
20.4	Combined Analysis	764
20.5	Split-Plot Example	768
20.6	Experiments Conducted Over Several Years	771
20.6.1	Analysis	772
20.7	Exercises	776
Appendix: Statistical Tables		779
Bibliography		789
Credits		791
Index		793