# Table of Contents


1 Data Collection 1

1.1 Introduction to Statistics 1

1.1.1 Statistics Definition and Statistical Thinking 1

1.1.2 Understanding the Process of Statistics 4

1.1.3 Difference between Qualitative and Quantitative Variables 6

1.1.4 Difference between Discrete and Continuous Variables 7

1.1.5 Determining the Level of Measurement of a Variable 10

1.1.6 Exercise Set 11

1.2 Designed Experiments versus Observational Studies 12

1.2.1 Difference between an Observational Study and an Experiment 12

1.2.2 Types of Observational Studies 17

1.2.3 Exercise Set 19
1.3 Simple Random Sampling 20
1.3.1 Sampling 20
1.3.2 Simple Random Sample 21
1.3.3 Technology Step by Step 27
1.3.4 Exercise Set 29
1.4 Other Sampling Methods 32
1.4.1 Stratified Sample 33
1.4.2 Systematic Sample 35
1.4.3 Cluster Sample 37
1.4.4 Convenience Sampling 38
1.4.5 Exercise Set 39
1.5 Bias in Sampling 40
1.5.1 Describe the Sources of Bias in Sampling 40
1.5.2 Exercise Set 45
1.6 The Design of Experiments 47
1.6.1 Explaining the Characteristics of an Experiment 47
1.6.2 Describing the Steps in Designing an Experiment 49
1.6.3 Explain the Completely Randomized Design 50
1.6.4 Explaining the Matched Pairs Design 52
1.6.5 Describing the Randomized Block Design 53
1.6.6 Exercise Set 56
1.7 References 58

2 Organize and Summarize Data 59

2.1 Data Organization 59
2.1.1 Introduction 59
2.1.2 Data Processing 60
2.1.3 Level of Measurement 60
2.1.4 Comparison between Qualitative and Quantitative Data 61
2.1.5 Organizing Qualitative Data in Table Form 61
2.1.6 Construction of Bar Graphs 64
2.1.7 Pie Charts 73
2.1.8 Exercise Set 79
2.2 Quantitative Data Organizing 80
2.2.1 Introduction 80
2.2.2 Discrete Data Organized in Table Form 81

2.2.3 Construct Histograms for Discrete data 84
2.2.4 Continuous Data Organized in Table Form 88
2.2.5 Construct Histograms for Continuous Data 92
2.2.6 Stem and Leaf Plot 96
2.2.7 Dot Plots 100
2.2.8 Identification of the Shape 101
2.2.9 Exercise Set 104
2.3 Further Expansion of Displays of Quantitative Data 106
2.3.1 Frequency Polygons 106
2.3.2 Cumulative Frequency and Relative frequency 110
2.3.3 Construction of Ogives over Cumulative Frequency and Relative frequency 111
2.3.4 Time-series Graphs 115
2.3.5 Data Misleading 115
2.3.6 Exercise Set 116
2.4 Measures of Central Tendency 117
2.4.1 Mean 117
2.4.2 Median 120
2.4.3 Mode 121
2.4.4 Exercise Set 122
2.5 Measures of Dispersion 123
2.5.1 Variance and Standard Deviation 123
2.5.2 The Empirical Rule 126
2.5.3 Chebyshev's Inequality 127
2.5.4 Exercise Set 127
2.6 Position Measures and Outliers 129
2.6.1 Z- Score 129
2.6.2 Interpret Percentile 130
2.6.3 Outliers of a Data set 134
2.6.4 Exercise Set 136
2.7 References 137

3 Probability and Probability Distributions 138

3.1 Introduction 138
3.2 Probability Rules 139
3.2.1 Identifying Events and the Sample Space of a Probability Experiment 142
3.2.2 A Probability Model 143

| | | | |
|---|---|---|---|
| | 3.2.3 | Two Important Rules of Probability | 143 |
| | 3.2.4 | Compute and Interpret Probabilities using the Empirical Method | 144 |
| | 3.2.5 | Determining the Probability using the Classical Method | 146 |
| | 3.2.6 | Exercise Set | 151 |
| 3.3 | Disjoint Events | | 152 |
| | 3.3.1 | Addition Rule for Disjoint Events | 153 |
| | 3.3.2 | Use the General Addition Rule | 157 |
| | 3.3.3 | Determine the Probability of an Event using the Complement Rule | 160 |
| | 3.3.4 | Exercise Set | 163 |
| 3.4 | Independence and the Multiplication Rule | | 165 |
| | 3.4.1 | Using the Multiplication Rule for Independent Events | 166 |
| | 3.4.2 | Exercise Set | 169 |
| 3.5 | Conditional and Independent Probabilities | | 170 |
| | 3.5.1 | Conditional Probability | 170 |
| | 3.5.2 | Exercises Set | 173 |
| 3.6 | Counting Techniques | | 174 |
| | 3.6.1 | Use the Multiplication Rule to Solve Counting Problems | 174 |
| | 3.6.2 | Solve Counting Problems using Permutation and Combinations | 179 |
| | 3.6.3 | Calculating Probabilities involving Permutations and Combinations | 184 |
| | 3.6.4 | Technology used in the Computation | 185 |
| | 3.6.5 | Exercise Set | 186 |
| 3.7 | Bayes' Theorem | | 187 |
| | 3.7.1 | Bayes' Theorem for 2 Events | 188 |
| | 3.7.2 | Bayes' Theorem for 3 Events | 188 |
| | 3.7.3 | Bayes' Theorem for n Events | 188 |
| | 3.7.4 | Exercise Set | 190 |
| 3.8 | Discrete Random Variables | | 192 |
| | 3.8.1 | Identify Discrete Probability Distribution | 192 |
| | 3.8.2 | How to Graph a Discrete Probability Distribution | 193 |
| | 3.8.3 | Calculate the Standard Deviation of a Discrete Random Variable | 196 |
| | 3.8.4 | Exercise Set | 197 |

3.9 Binomial Probability Distribution 198
3.9.1 Criteria for Binomial Probability Experiment 198
3.9.2 How to Graph a Binomial Probability Distribution 200
3.9.3 Exercise Set 201
3.10 Poisson Probability Distribution 202
3.10.1 Function of Poisson Probability Distribution 202
3.10.2 Exercise Set 204
3.11 Properties of Normal Distribution 205
3.11.1 Applications of Normal Distribution 208
3.11.2 Exercise Set 209
3.12 Normal Approximation to Binomial Probability Distribution 211
3.12.1 Exercise Set 213
3.13 References 213

4 Estimating the Value of a Parameter and its Hypothesis Testing 214

4.1 Introduction 214
4.1.1 Random Sample 215
4.1.2 Definition of Statistic 215
4.1.3 Moments of the Sample Mean and Variance 215
4.1.4 Sampling Distribution 219
4.1.5 The Central Limit Theorem 220
4.1.6 Normal Approximations of Some Distributions 222
4.1.7 Sampling Distributions for the Normal 222
4.1.8 The T Result 224
4.1.9 The F Result for Variance Ratios 226
4.1.10 Sampling Distribution of Sample Proportion 226
4.1.11 Sampling Distribution of $\hat{p}$ 229
4.1.12 Exercise Set 233
4.2 Estimating the Value of a Parameter 234
4.2.1 Point Estimate 235
4.2.2 The Method of Moments 235
4.2.3 The Method of Maximum Likelihood 238
4.2.4 Unbiasedness 248
4.2.5 Mean Square Error 249
4.2.6 Asymptotic Distribution of MLEs 251
4.2.7 Exercise Set 253
4.3 Confidence Interval 255

4.3.1 Introduction 255
4.3.2 Sample Size 257
4.3.3 Confidence Intervals for the Normal Distribution 258
4.3.4 Confidence Intervals for Binomial and Poisson 261
4.3.5 Confidence Intervals for Two-sample Problems 270
4.3.6 Paired Data 273
4.3.7 Exercise Set 275
4.4 Hypothesis Testing 276
4.4.1 The Testing Procedure 276
4.4.2 Hypotheses 276
4.4.3 One-sided and Two-sided Tests 276
4.4.4 Test Statistics 277
4.4.5 Error 277
4.4.6 P-values 278
4.4.7 Testing the Value of a Population Mean 278
4.4.8 Testing the Value of a Population Variance 279
4.4.9 Testing the Value of a Population Proportion 280
4.4.10 Testing the Value of the Mean of a Poisson Distribution 282
4.4.11 Paired Data 283
4.4.12 Exercise Set 284
4.5 References 284

5 Inference on Two Samples and Categorical Data 285

5.1 Introduction 285
5.2 Inference about Two Population Proportions 286
5.2.1 Distinguish between Independent and Dependent Sampling 286
5.2.2 Test Hypothesis regarding Two Proportions from Independent Samples 287
5.2.3 Construct and Interpret Confidence Intervals for the Difference between Two Population Proportions 295
5.2.4 Determine the Sample Size Necessary for Estimating the Difference between Two Population Proportions 297
5.2.5 Exercise Set 299
5.3 Inference about Two Means: Dependent Samples 301

5.3.1 Test Hypothesis for a Population Mean from Matched-Pairs Data 301
5.3.2 Construct and Interpret the Confidence Intervals about the Population Mean Difference in Matched-Pairs Data 307
5.3.3 Exercise Set 308
5.4 Inference about Two Means: Independent Samples 309
5.4.1 Test Hypotheses Regarding the Difference of Two Independent Means 310
5.4.2 Construct and Interpret Confidence Intervals Regarding the Difference of Two Independent Means 316
5.4.3 Exercise Set 318
5.5 Inference about Two Population Standard Deviations 319
5.5.1 Find Critical Values of the F-distribution 319
5.5.2 Test Hypotheses regarding Two Population Standard Deviations 323
5.5.3 Exercise Set 327
5.6 Goodness-of-Fit Test 329
5.6.1 Perform a Goodness-of-Fit Test 329
5.6.2 Exercise Set 339
5.7 Tests for Independence and the Homogeneity of Proportions 341
5.7.1 Perform a Test for Independence 341
5.7.2 Perform a Test for Homogeneity of Proportions 350
5.7.3 Exercise Set 355
5.8 Inference about Two Population Proportions: Dependent Samples 358
5.8.1 Test Hypotheses Regarding Two Proportions from Dependent Samples 358
5.8.2 Exercise Set 362
5.9 References 364

6 Correlation and Regression 365

6.1 Scatter Diagrams and Correlation 365
6.1.1 Draw and Interpret Scatter Diagrams 365
6.1.2 Explaining the Properties of the Linear Correlation Coefficient 367
6.1.3 Compute and Interpret the Linear correlation Coefficient 368

| | | | |
|---|---|---|---|
| | 6.1.4 | Differentiate Between Correlation and Causation | 371 |
| | 6.1.5 | Exercise Set | 372 |
| 6.2 | Least Square Regression | | 374 |
| | 6.2.1 | Find the Least Squares Regression Line and use that Line to make Predictions | 374 |
| | 6.2.2 | Interpret the Slope and the y Intercept of the Least Squares Regression Line | 376 |
| | 6.2.3 | Exercise Set | 376 |
| 6.3 | Diagnostics on the Least Squares Regression | | 378 |
| | 6.3.1 | Calculate and Interpret the Coefficient of Determination | 378 |
| | 6.3.2 | Perform Residual Analysis on a Regression Model | 382 |
| | 6.3.3 | Identify Influential Observations | 386 |
| | 6.3.4 | Exercise Set | 387 |
| 6.4 | Inference on the Least Squares Regression Model | | 390 |
| | 6.4.1 | Introduction | 390 |
| | 6.4.2 | State the Requirements of the Least Squares Regression | 391 |
| | 6.4.3 | Compute the Standard Error of the Estimate | 393 |
| | 6.4.4 | Verify that the Residuals Are Normally Distributed | 395 |
| | 6.4.5 | Conduct Inference on the Slope of the Least Squares Regression Model | 395 |
| | 6.4.6 | Construct a Confidence Interval about the Slope of the Least Squares Regression Model | 398 |
| | 6.4.7 | Exercise Set | 399 |
| 6.5 | Confidence and Prediction Intervals and Multiple Regression | | 401 |
| | 6.5.1 | Construct a Confidence Interval for Mean Response | 401 |
| | 6.5.2 | Construct Prediction Intervals for an Individual Response | 402 |
| | 6.5.3 | Multiple Regression | 403 |
| | 6.5.4 | Perform an F-test for Lack of Fit | 407 |
| | 6.5.5 | Construct Confidence and Prediction Intervals | 410 |
| | 6.5.6 | Work with Multiple Regression Models with Interaction | 410 |
| | 6.5.7 | Exercise Set 6.5 | 414 |
| 6.6 | References | | 415 |

7 Appendix 416

7.1 Probabilities for the Standard Normal distribution 416
7.2 Percentage Points for the Standard Normal distribution 419
7.3 Percentage Points for the t distribution 420
7.4 Probabilities for the $x^2$ distribution 421
7.5 Probabilities for the $x^2$ distribution 423
7.6 Probabilities for the $x^2$ distribution 424
7.7 Percentage Points for the $x^2$ distribution 426
7.8 Percentage Points for the $x^2$ distribution 427
7.9 Percentage Points for the F distribution 428
7.10 10% Points for the F distribution 429
7.11 5% Points for the F distribution 430
7.12 2 1/2 % Points for the F distribution 431
7.13 1% Points for the F distribution 432

Index 434