# Introduction to Statistics

Online Edition

Primary author and editor:
David M. Lane[1]

Other authors:
David Scott[1], Mikki Hebl[1], Rudy Guerra[1], Dan Osherson[1], and Heidi Zimmer[2]

[1]Rice University; [2]University of Houston, Downtown Campus

Section authors specified on each section.

This work is in the public domain. Therefore, it can be copied and reproduced without limitation.

1. Introduction .....10

What Are Statistics .....11

Importance of Statistics .....13

Descriptive Statistics .....15

Inferential Statistics .....20

Variables .....26

Percentiles .....29

Levels of Measurement .....34

Distributions .....40

Summation Notation .....52

Linear Transformations .....55

Logarithms....58
Statistical Literacy....61
Exercises....62
2. Graphing Distributions....65
Graphing Qualitative Variables....66
Graphing Quantitative Variables....75
Stem and Leaf Displays....76
Histograms....82
Frequency Polygons....86
Box Plots....92
Bar Charts....101
Line Graphs....105
Dot Plots....109
Statistical Literacy....113
References....115
Exercises....116
3. Summarizing Distributions....123
What is Central Tendency?....124
Measures of Central Tendency....131
Median and Mean....134
Additional Measures of Central Tendency....136
Comparing Measures of Central Tendency....140
Measures of Variability....144

Shapes of Distributions ..... 152
Effects of Linear Transformations ..... 154
Variance Sum Law I ..... 156
Statistical Literacy ..... 158
Exercises ..... 159
4. Describing Bivariate Data ..... 164
Introduction to Bivariate Data ..... 165
Values of the Pearson Correlation ..... 170
Properties of Pearson's r ..... 175
Computing Pearson's r ..... 176
Variance Sum Law II ..... 178
Statistical Literacy ..... 180
Exercises ..... 181
5. Probability ..... 185
Remarks on the Concept of "Probability" ..... 186
Basic Concepts ..... 189
Permutations and Combinations ..... 198
Binomial Distribution ..... 203
Poisson Distribution ..... 207
Multinomial Distribution ..... 208
Hypergeometric Distribution ..... 210
Base Rates ..... 212
Statistical Literacy ..... 215

Exercises ....216
6. Research Design ....222
Scientific Method ....223
Measurement ....225
Basics of Data Collection ....231
Sampling Bias ....235
Experimental Designs ....238
Causation ....242
Statistical Literacy ....245
References ....246
Exercises ....247
7. Normal Distributions ....248
Introduction to Normal Distributions ....249
History of the Normal Distribution ....252
Areas Under Normal Distributions ....256
Standard Normal Distribution ....259
Normal Approximation to the Binomial ....263
Statistical Literacy ....266
Exercises ....267
8. Advanced Graphs ....272
Quantile-Quantile (q-q) Plots ....273
Contour Plots ....289
3D Plots ....292

Statistical Literacy .....297
Exercises .....298
9. Sampling Distributions .....299
Introduction to Sampling Distributions.....300
Sampling Distribution of the Mean .....307
Sampling Distribution of Difference Between Means.....311
Sampling Distribution of Pearson's r.....316
Figure 2. The sampling distribution of r for N = 12 and ρ = 0.90. .....318
Sampling Distribution of p.....319
Statistical Literacy .....322
Exercises .....323
10. Estimation .....328
Introduction to Estimation .....329
Degrees of Freedom .....330
Characteristics of Estimators.....333
Confidence Intervals.....336
Introduction to Confidence Intervals .....337
t Distribution.....339
Confidence Interval for the Mean .....343
Difference between Means .....349
Correlation .....356
Proportion.....358
Statistical Literacy .....360

Exercises .....362
11. Logic of Hypothesis Testing .....369
Introduction .....370
Significance Testing .....375
Type I and II Errors .....377
One- and Two-Tailed Tests .....379
Interpreting Significant Results .....383
Interpreting Non-Significant Results .....385
Steps in Hypothesis Testing .....388
Significance Testing and Confidence Intervals .....389
Misconceptions .....391
Statistical Literacy .....392
References .....393
Exercises .....394
12. Testing Means .....398
Testing a Single Mean .....399
Differences between Two Means (Independent Groups) .....406
All Pairwise Comparisons Among Means .....412
Specific Comparisons (Independent Groups) .....418
Difference Between Two Means (Correlated Pairs) .....428
Specific Comparisons (Correlated Observations) .....432
Pairwise Comparisons (Correlated Observations) .....436
Statistical Literacy .....438

References....439
Exercises....440
13. Power....447
Introduction to Power....448
Example Calculations....450
Factors Affecting Power....454
Statistical Literacy....458
Exercises....459
14. Regression....461
Introduction to Linear Regression....462
Partitioning the Sums of Squares....468
Standard Error of the Estimate....473
Inferential Statistics for b and r....476
Influential Observations....482
Regression Toward the Mean....487
Introduction to Multiple Regression....495
Statistical Literacy....507
References....508
Exercises....509
15. Analysis of Variance....515
Introduction....516
Analysis of Variance Designs....518
Between- and Within-Subjects Factors....519

One-Factor ANOVA (Between Subjects) ..... 521
Multi-Factor Between-Subjects Designs ..... 532
Unequal Sample Sizes ..... 544
Tests Supplementing ANOVA ..... 553
Within-Subjects ANOVA ..... 562
Statistical Literacy ..... 569
Exercises ..... 570
16. Transformations ..... 576
Log Transformations ..... 577
Tukey Ladder of Powers ..... 580
Box-Cox Transformations ..... 588
Statistical Literacy ..... 594
References ..... 595
Exercises ..... 596
17. Chi Square ..... 597
Chi Square Distribution ..... 598
One-Way Tables (Testing Goodness of Fit) ..... 601
Contingency Tables ..... 605
Statistical Literacy ..... 608
References ..... 609
Exercises ..... 610
18. Distribution-Free Tests ..... 616
Benefits ..... 617

Randomization Tests: Two Conditions .....618
Randomization Tests: Two or More Conditions .....620
Randomization Tests: Association (Pearson's r) .....622
Randomization Tests: Contingency Tables: (Fisher's Exact Test) .....624
Rank Randomization: Two Conditions (Mann-Whitney U, Wilcoxon Rank Sum) .....626
Rank Randomization: Two or More Conditions (Kruskal-Wallis) .....631
Rank Randomization for Association (Spearman's ρ) .....633
Statistical Literacy .....636
Exercises .....637
19. Effect Size .....639
Proportions .....640
Difference Between Two Means .....643
Proportion of Variance Explained .....647
References .....653
Statistical Literacy .....654
Exercises .....655
20. Case Studies .....657
21. Glossary .....659

# 1. Introduction

This chapter begins by discussing what statistics are and why the study of statistics is important. Subsequent sections cover a variety of topics all basic to the study of statistics. The only theme common to all of these sections is that they cover concepts and ideas important for other chapters in the book.

A. What are Statistics?
B. Importance of Statistics
C. Descriptive Statistics
D. Inferential Statistics
E. Variables
F. Percentiles
G. Measurement
H. Levels of Measurement
I. Distributions
J. Summation Notation
K. Linear Transformations
L. Logarithms
M. Exercises