# CONTENTS

1 INTRODUCTION ................................. 1
   1.1 A BRIEF HISTORY OF THE NORMAL CURVE ... 2
   1.2 EMPIRICAL STUDIES REGARDING NORMALITY ... 5
   1.3 INFERENTIAL METHODS ........................ 6

2 GETTING STARTED .............................. 11
   2.1 PROBABILITY CURVES .......................... 11
   2.2 THE MEAN .................................. 12
   2.3 THE MEDIAN .................................. 16
   2.4 A WEIGHTED MEAN ............................. 18
   2.5 VARIANCE .................................... 19
   2.6 MEASURING ERROR ............................. 20
   2.7 FITTING A STRAIGHT LINE TO DATA .......... 22
      2.7.1 Two Views of the Computations ......... 25
   2.8 A SUMMARY OF KEY POINTS ................... 27

3 THE NORMAL CURVE AND OUTLIER DETECTION 29
   3.1 THE NORMAL CURVE .............................. 29
   3.2 DETECTING OUTLIERS ........................... 32
   3.3 THE BOXPLOT .................................. 34
   3.4 THE CENTRAL LIMIT THEOREM ................. 35
      3.4.1 Normality and the Median ............... 40
   3.5 THREE POINTS WORTH STRESSING ............. 44
   3.6 A SUMMARY OF KEY POINTS .................... 45

4 ACCURACY AND INFERENCE ...................... 47
   4.1 SOME OPTIMAL PROPERTIES OF THE MEAN ...... 47
   4.2 THE MEDIANVERSUS THE MEAN ................. 49
   4.3 REGRESSION .................................. 52
   4.4 CONFIDENCE INTERVALS ........................ 55
   4.5 CONFIDENCE INTERVALS FOR THE POPULATION MEAN ........................................ 57
   4.6 CONFIDENCE INTERVAL FOR THE SLOPE ........ 58
   4.7 A SUMMARY OF KEY POINTS .................... 62
5 HYPOTHESIS TESTING AND SMALL SAMPLE SIZES 63
  5.1 HYPOTHESIS TESTING .............................. 63
  5.2 THE ONE-SAMPLE $T$ TEST .......................... 68
  5.3 SOME PRACTICAL PROBLEMS WITH STUDENT’S $T$ 72
  5.4 THE TWO-SAMPLE CASE ............................ 77
  5.5 THE GOOD NEWS ABOUT STUDENT’S $T$ ............ 79
  5.6 THE BAD NEWS ABOUT STUDENT’S $T$ ............. 79
  5.7 WHAT DOES REJECTING WITH STUDENT’S $T$ TELL US? 81
  5.8 COMPARING MULTIPLE GROUPS ..................... 83
  5.9 COMPARING MEDIANS ............................. 83
  5.10 A SUMMARY OF KEY POINTS ...................... 84
  5.11 BIBLIOGRAPHIC NOTES .......................... 85

6 THE BOOTSTRAP 87
  6.1 TWO BOOTSTRAP METHODS FOR MEANS ............... 88
    6.1.1 The Percentile Method ...................... 88
    6.1.2 The Bootstrap $t$ Method .................... 92
  6.2 TESTING HYPOTHESES ............................. 96
    6.2.1 Why Does the Bootstrap $t$ Perform Well
          Compared to Student’s $T$? ................. 96
  6.3 COMPARING TWO INDEPENDENT GROUPS ................ 98
    6.3.1 Hypothesis Testing ......................... 98
  6.4 COMPARING MEDIANS ............................. 99
  6.5 REGRESSION ..................................... 99
    6.5.1 A Modified Percentile Bootstrap Method .... 100
    6.5.2 The Wild Bootstrap ......................... 102
  6.6 CORRELATION AND TESTS OF INDEPENDENCE ........ 103
  6.7 A SUMMARY OF KEY POINTS ...................... 107
  6.8 BIBLIOGRAPHIC NOTES .......................... 108

7 A FUNDAMENTAL PROBLEM 109
  7.1 POWER ........................................... 112
  7.2 ANOTHER LOOK AT ACCURACY ....................... 114
  7.3 THE GRAPHICAL INTERPRETATION OF VARIANCE .... 115
  7.4 OUTLIER DETECTION ............................. 115
  7.5 MEASURING EFFECT SIZE .......................... 117
  7.6 HOW EXTREME CAN THE MEAN BE? .................. 120
  7.7 REGRESSION ..................................... 120
  7.8 PEARSON’S CORRELATION ......................... 123
  7.9 MORE ABOUT OUTLIER DETECTION .................. 125
  7.10 A SUMMARY OF KEY POINTS ...................... 125
  7.11 BIBLIOGRAPHIC NOTES .......................... 126
11 ROBUST REGRESSION 193
11.1 THEIL-SEN ESTIMATOR .................. 194
11.2 REGRESSION VIA ROBUST CORRELATION AND VARIANCES ............ 197
11.3 $L_1$ REGRESSION ..................... 198
11.4 LEAST TRIMMED SQUARES .......... 200
11.5 LEAST TRIMMED ABSOLUTE VALUE .......... 203
11.6 LEAST MEDIAN OF SQUARES .......... 203
11.7 REGRESSION OUTLIERS AND LEVERAGE POINTS ............. 203
11.8 M-ESTIMATORS ....................... 206
11.9 THE DEEPEST REGRESSION LINE .......... 208
11.10 RELATIVE MERITS AND EXTENSIONS TO MULTIPLE PREDICTORS .......... 209
11.11 CORRELATION BASED ON ROBUST REGRESSION .... 210
11.12 ROBUST SMOOTHERS ................. 210
11.13 COMPARING REGRESSION LINES: MODERN IMPROVEMENTS ON ANCOVA ........... 211
11.14 CHOOSING A REGRESSION METHOD .......... 213
11.15 A SUMMARY OF KEY POINTS .......... 215
11.16 BIBLIOGRAPHIC NOTES .......... 215

12 ALTERNATIVE STRATEGIES AND SOFTWARE 217
12.1 RANKED-BASED METHODS FOR COMPARING TWO GROUPS .......... 217
12.2 EMPIRICAL LIKELIHOOD ................. 223
12.3 PERMUTATION TESTS ................. 224
12.4 PLOTS FOR COMPARING TWO GROUPS .......... 226
12.5 COMPARING MORE THAN TWO GROUPS .......... 229
12.6 RANK-BASED METHODS FOR COMPARING MULTIPLE GROUPS .......... 232
12.7 REGRESSION BASED ON RANKED RESIDUALS .......... 233
12.8 SOFTWARE ......................... 233
12.9 A SUMMARY OF KEY POINTS .......... 235
12.10 BIBLIOGRAPHIC NOTES .......... 236

APPENDIX A 237
REFERENCES 239
Fundamentals of Modern Statistical Methods
Substantially Improving Power and Accuracy
Wilcox, R.R.
2010, XVI, 249 p., Hardcover
ISBN: 978-1-4419-5524-1