An Introduction to Statistics

(statistics 427)

J. Calvin Berry
Mathematics Department
University of Louisiana at Lafayette

https://www.ucs.louisiana.edu/~jcb0773/

August 2024 revision

This text is currently in the middle of a major revision. Several sections are incomplete and you should watch out for typos and other editorial errors. Updates will be posted as soon as they are ready.

© 2002, 2024 J. Calvin Berry

20240807

Table of Contents

1 Introduction 1
- 1.1 Basic ideas 1
- 1.2 Sampling 6
- 1.3 Experimentation 15
- 1.4 Summary 20
- 1.5 Problems 21

2 Descriptive Statistics 23

Part 1: Tabular and graphical summary 23
- 2.1 Introduction 23
- 2.2 Types of variables 24
- 2.3 Describing qualitative data 25
- 2.4 Describing discrete quantitative data 36
- 2.5 Describing continuous quantitative data 46

Part 2: Numerical summary values 48
- 2.6 Numerical summary for quantitative data 48
- 2.7 Quantiles and percentiles in general 62
- 2.8 The mean and standard deviation 63
- 2.9 A measure of relative position 67
- 2.10 Summary 69
- 2.11 Problems 72

3 Probability Models 73
- 3.1 Introduction 73
- 3.2 Probability models for a variable with a finite number of values 74
- 3.3 Probability models for discrete quantitative variables 79
- 3.4 Probability models for counts 81
  - Part 1: Binomial distribution 81
  - Part 2: Hypergeometric distribution 86
  - Part 3: Poisson distribution 90
- 3.5 Probability models for continuous quantitative variables 94

3.6 The normal distribution 102
3.7 Sampling distributions 111
3.7a The sampling distribution of the sample mean 112
3.7b The sampling distribution of the sample proportion 114
3.7c Large sample size properties of sampling distributions 115
3.7d Sampling from a normal population 117
3.8 Summary 121
3.9 Problems 121

4 Overview of Statistical Inference 123
4.1 Introduction 123
4.2 Inference based on randomization distributions 128
4.3 Exercises 144

5 Inference for a Mean or Median 145
5.1 Introduction 145
5.2 Inference for a population mean 145
5.2a Estimating a normal population mean 159
5.2b Tests of hypotheses about normal a population mean 169
5.2c Directional confidence bounds 184
5.3 Inference for a population median 186
5.4 Summary 192
5.5 Exercises 193

6 Comparing Two Means 197
6.1 Introduction 197
6.2 Comparing the means of two normal populations 198
6.2a Inference when the two standard deviations are equal 200
6.2b Inference when the two standard deviations are not equal 220
6.3 Inference based on ranks 221
6.4 Summary 223
6.5 Exercises 234

7 Comparing Several Means 235
7.1 Introduction 235
7.2 Comparing the means of k normal populations 237
7.2a Assumptions, notation, and the overall F–test 237
7.2b Model comparison F–tests 245
7.2c Confidence intervals for linear combinations of means 260
7.3 Summary 267
7.4 Exercises 267

8 Inference for a Proportion 269
8.1 Introduction 269
8.2 The sampling distribution of the sample proportion 269
8.3 Estimation of p 274
8.4 Testing a hypothesis about p 311
8.5 Directional confidence bounds 311
8.6 Summary 318
8.7 Exercises 321

9 Inference for Two Proportions 323
9.1 Introduction 323
9.2 Inference for two proportions (independent samples) 324
9.3 Inference for two proportions (paired samples) 345
9.4 Summary 351
9.5 Exercises 355

10 Descriptive Statistics for Bivariate Data 357
10.1 Introduction 357
10.2 Association and Correlation 358
10.3 Regression 371
10.4 Summary 394
10.5 Exercises 394

11 Inference for Bivariate Data 395
11.1 Inference for Regression 395
11.2 Summary 412
11.3 Exercises 412

Index 413