Statistics in Theory and Practice

Robert Lupton

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY
Contents

Preface ix

1 Introduction 3

2 Preliminaries 5
  2.1 Basic Definitions 5
  2.2 The Distribution of \( g(x, y) \) Given the Distribution of \( x \) and \( y \) 9
  2.3 Characteristic Functions 10

3 Some Common Probability Distributions 13
  3.1 Gaussian Distribution 13
  3.2 Multivariate Gaussian Distributions 13
  3.3 Log-Normal Distribution 14
  3.4 Poisson Distribution 14
  3.5 Binomial Distribution 17
  3.6 Multinomial Distribution 20
  3.7 Cauchy Distribution 21
  3.8 Beta Distribution 22

4 Distributions Related to the Gaussian 24
  4.1 \( \chi^2 \) Distribution 24
  4.2 \( \chi^2 \) Distribution: Linear Constraints 25
  4.3 Student’s \( t \)-Distribution 27
  4.4 \( F \)-distribution 28

5 Sampling 30
  5.1 Estimating \( \mu \) and \( \sigma \) 30
  5.2 Efficiency of Estimators 33

6 Distributions of Sample Statistics 36
  6.1 Distribution of \( s^2 \) if \( x = N(0, \sigma^2) \) 37
  6.2 Student’s \( t \)-Distribution 39
  6.3 Sampling from a Finite Population 40
  6.4 Asymptotic Distribution of the Median 42
  6.5 Bootstraps and Jackknives 43

7 Bayes’ Theorem and Maximum Likelihood 50
  7.1 Bayes’ Theorem and Postulate 50
  7.2 Maximum Likelihood Estimators 50
8 Confidence Intervals
  8.1 Introduction 53
  8.2 The Choice of Confidence Intervals 55
  8.3 Classical and Bayesian Approaches to Confidence Intervals 56
9 Hypothesis Testing
  9.1 Do Two Samples Have the Same Mean? 61
  9.2 Comparing the Means of Paired Samples 64
  9.3 Do Two Samples Have the Same Variance? 64
  9.4 The Theory of Tests 65
  9.5 Likelihood Ratio Tests 69
10 The Theory of Maximum Likelihood Estimators
  10.1 Efficiency of ML Estimators and the Minimum Variance
       Bound 72
  10.2 Example: Statistical Parallaxes by MLE 75
  10.3 Maximum Likelihood, Chi-Squared, Least Squares, and
       All That 77
  10.4 Uncertainties in Parameters 78
11 Least Squares Fitting for Linear Models
  11.1 Relationship to MLE 81
  11.2 A Simple Example 82
  11.3 Linear Estimators: Theory 84
  11.4 Goodness of Fit 85
  11.5 Errors in the Parameters 87
  11.6 Fitting Models to Data with Non-Gaussian Errors 92
  11.7 Fitting Models with Errors in Both x and y 92
12 Hypothesis Testing in the Linear Model
  12.1 Significance of Parameters 98
  12.2 F-tests and Linear Regression 102
  12.3 The Distribution of Pearson's Correlation Coefficient 104
13 Rank Correlation Coefficients
  13.1 Spearman's Correlation Coefficient 107
  13.2 Kendall's Correlation Coefficient 108
  13.3 Distribution of \( r_s \) When \( \rho = 0 \) 109
  13.4 Distribution of Kendall's \( t \) When \( \tau = 0 \) 111
  13.5 Rank Correlation Coefficients When \( \rho \neq 0 \) 112
  13.6 Efficiency of Rank Tests 112
Contents

14 Tests of Fit 114
14.1 Binned Data: $\chi^2$ Tests 114
14.2 Signs Tests 116
14.3 Unbinned Data: Smirnov and Kolmogorov-Smirnov Tests 117
14.4 The Kolmogorov-Smirnov Test 119
14.5 Efficiency 121

15 Robust Tests for Means 123
15.1 Robustness of Tests 123
15.2 Distribution-Free Procedures for Equality of Means 123
15.3 Wilcoxon's $U$ Statistic 125

Epilogue 127

Some Numerical Exercises 128
References 130
Answers 132
Symbol Index 181
Problem Index 185
Index 186