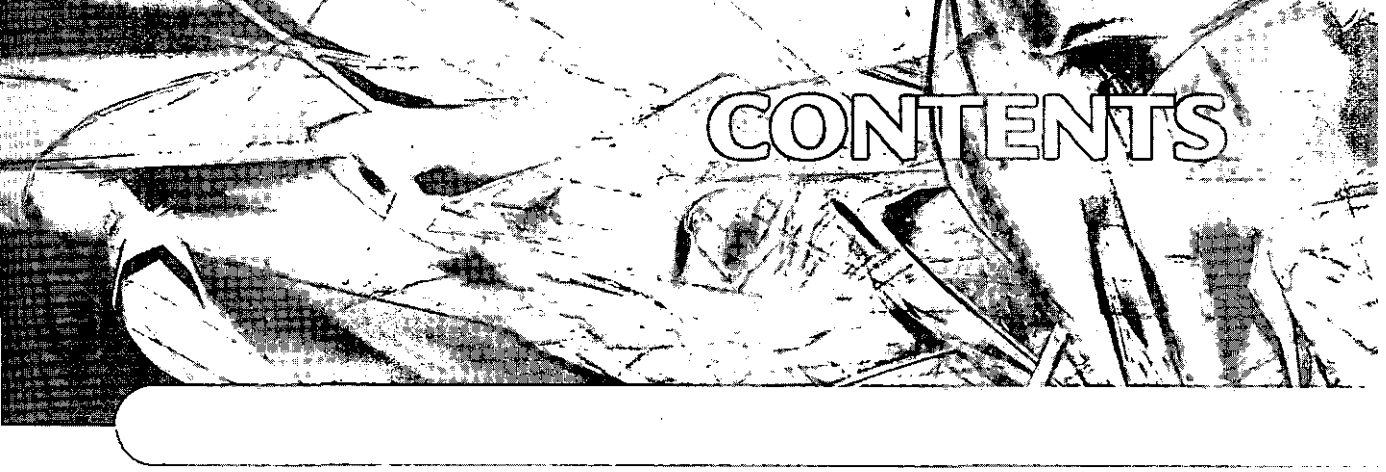
RESEARCH METHODS FOR INFORMATION SYSTEMS

Ronald S. King

MERCURY LEARNING AND INFORMATION
Dulles, Virginia
Boston, Massachusetts
New Delhi



Preface

UNIT 1 DESCRIPTIVE STATISTICS

Chapter I Introduction to Statistics

] .1 Introduction 4
1.2 What Is Statistics? 5
1.3 The Role ol Probability in Statistical Inference 6
1.4 "A'pes of Data 6
1.5 Mow to Perform a Statistical Study 8
1.6 Exercises 9

Chapter 2 Measures of Central Tendency 11

2.1 Sample Study 12
2.2 Introduction 12
2.3 Measures of Central Tendency L2
2.3.1 Mode 12
2.3.2 Median 13
2.3.3 Mean 13
2.4 Measures ol the Middle and Distributional Shape 14
2.5 Exercises 35

Chapter 3 Measures of Dispersion 17

3.1 Introduction 18
3.2 Measures of Variation IS
3.2.1 Range 18
3.2.2 Variance and standard deviation 18

3.3 Chebyshev's Inequality and the Empirical Rule 20
3.4' Comparing Variability 21
3.5 Measures of Distributional Shape: Skewness 22
3.6 Measures of Distributional Shape: Kurtosis 22
3.7 Exercises 25

Chapter 4 Frequency Distributions 29

4.1 Introduction 30
4.2 Sample Study 30
4.3 Presenting Qualitative Data 30
4.4 Exercises 35

Chapter 5 Grouped Frequency Distrihutions 37

5.1 Introduction 38
5.2 Summarizing Quantitative Data 38
5.3 Exercises 42

Chapter (i Data Mining 43

6.1- Introduction 44
6.2 Single Variatc Exploratory Data Analysis 44
6.2.1 Stem-and-lcaf plots 44
6.2.2 Quartiles, deciles, and percentiles 46
6.2.3 Box plots 47
6.2.4 Time plots 48
6.3 Bivariatc Exploratory Data Analysis 48
6.3.1 Pivot tables and pivot charts 48
6.3.2 Scatter diagrams 56
6.4 Exercises 59

UNIT 2 ELEMENTARY PROBABILITY 65

Chapter 7 Random Experiments, Counting Techniques, and Probability 67

7.1 Introduction 68
7.2 Random Experiments 68
7.3 Sample Spaces and Events 68
7.4 What Probability Means 69
7.5 Equally Likely Outcomes 70
7.6 Putting Events Together: Union, Intersection, and Complement 71
7.7 Venn Diagrams 72
7.8 The Axioms of Probability 72
7.8.1 Laws, or theorems, derived from the axioms 73
7.9 Counting Techniques: Permutations and Combinations 76

7.10 Counting Techniques and Probability 78
7.11 Conditional Probabilities 79
7.12 Independent Events 82
7.13 Exercises 84

Chapter 8 Probability Toolkit

8.1 Introduction 88
8.2 Random Variables 88
8.3 Probability Distributions 89
8.4 Expected Value 92
8.5 Variance 93
8.6 Independent Random Variables 94
8.7 Exercises 96

Chapter 9 Discrete Probability Distributions 99

9.1 Introduction 100
9.2 Problem Solving through Modeling 100
9.3 Important Discrete Distributions 102
9.3.1 The uniform discrete distribution 102
9.3.2 The binomial distribution 103
9.3.3 The hypergeometric distribution 109
9.3.4 The poisson distribution 11.2
9.3.5 Summary of discrete distributions 115
9.4 Exercises 116
Reference 117

Chapter 10 Continuous Probability Distributions 119

.10.1 Introduction .120
.10.2 Introduction to Continuous Probability Distributions 120
10.3 Expected Value and Variance of Continuous Distributions 124
10.4 Particular Continuous Distributions 125
10.4.1 The continuous uniform distribution 125
10.4.2 The exponential distribution 126
10.5 Exercises 129

Chapter 11 The Normal Distribution 131

11.1 Introduction 132
11.2 The Normal Distribution 132
.1 1.3 Exercises 142

Chapter 12 Distributional Approximations 145

12.1 Introduction 146
12.2 Review of Discrete and Continuous Distributions 146
12.2.1 Summary of discrete distributions 146
12.2.2 Summary of continuous distributions 1.47

12.3 Discrete Approximations of Discrete Distributions 147
12.4 Continuous Approximations of Discrete Distributions 149
12.4.1 Normal approximation of a Poisson distribution 149
.12.4.2 Normal approximation of a binomial distribution 15.1
12.5 Exercises J51

UNITS INTRODUCTION TO ESTIMATION 153

Chapter 13 Sampling Distributions 155

13.1 Introduction 156
13.2 An Example of a Sampling Distribution 156
13.3 The Sampling Distribution of X 158
13.4 The Central Limit Theorem 159
13.5 The Distribution of the Sample Median 162
13.6 Sampling Distributions of Measures of Dispersion 162
13.6.1 The expected value of the sample variance 162
13.6.2 The sample range 164
13.6.3 The distribution of the sample proportion 164
13.7 Exercises 165

Chapter 14 Point Estimation'and Interval Estimation 169

14.1 Introduction 1.70
14.2 Point Estimation 170
14.3 Interval Estimation 174
14.4 Exercises 179

Chapter 15 Introduction to Hypothesis Testing 181

15.1 Introduction 182
15.2 The Probability of a Type II Error 188
15.3 Exercises 192

UNIT 4 HYPOTHESIS TESTING 195

Chapter 16 Single Large Sample Tests 197

16.1 Introduction 198
16.2 Sample Study 198
16.3 Lower-Tail Tests for the Population Mean 199
16.4 Two-Tail Tests for the Population Mean 202
16.5 Exercises 204

Chapter 17 Single Small Sample Tests 207

17.1 Introduction 208
17.2 Small Sample Tests for the Population Mean 208
17.3 Hypothesis Tests with the Population Variance 209
17.4 Exercises 213

Chapter IS Independent Sample Tests 215

18'! Introduction 216
18.2 Tests with Two Population Variances 216
18.3 Independent Sample Test for the Difference of Two Means
18.4 Exercises 229

Chapter 19 Matched-Pair Tests 233

19.1 Introduction 234
19.2 Matched-pair Tests lor the Difference of Two Means 234
.19.3 t-Tests lor the Difference of Population Proportions 238
19.4 Exercises 238

Chapter 20 Hypothesis Testing versus Confidence Intervals 241

20.1 Introduction 242
20.2 Choosing a Confidence Level 242
20.3 Hypothesis Testing versus Confidence Intervals 242
20.4 Two-sided Confidence Intervals 243
20.5 One-sided Confidence Intervals 244
20.6 Confidence Intervals for Proportions 245
20.7 Exercises 247

UNIT 5 APPLICATIONS OF CHI-SQUARE STATISTICS 249

Chapter 21 Chi-Square Tests of Multinomial Data 251

21.1 Introduction 252
21.2 Chi-Square Tests of Multinomial Data 252
21.3 Exercises 257

Chapter 22 Chi-Square Tests of Independence 259

22.1 Introduction 260
22.2 Chi-Square Tests of Independence 260
22.3 Guidelines for Using the Chi-Square Test of Independence 265
22.4 Exercises 265

Chapter 23 Chi-Square Tests of Goodness-of-Fit and Missing Data 269

23.1 Introduction 270
23.2 Chi-Square Tests of Goodness-of-Eit 270
23.3 Chi-Square Analysis of Missing Data 273
23.4 Exercises 274

UNIT 6 REGRESSION AND CORRELATION ANALYSIS 277

Regression and Curve Fitting 278
Test of the CEHs Statistical Quality 278
Selection of the Best CER 278

Chapter 24 Correlation Analysis 279

24.1 Introduction 280
24.2 Scatter Diagrams 280
24.3 The Pearson Correlation Coefficient 283
24.4 Estimating the Population Correlation Coefficient 287
24.5 Partial Correlation Coefficients 288
24.6 Recommendations When Using Correlation Coefficients 290
24.7 Exercises 290
Reference 295

Chapter 25 Introduction to Simple Linear Regression 297

25.1 Introduction 298
25.2 Sample Study 298
25.3 The Regression Line and Regression Equation 299
25.4 Simple Linear Regression 300
25.5 Exercises 306
Simple Linear Regression Project I 309
Simple Linear Regression Project II 310
References 311

Chapter 26 Simple Linear Regression: Hypothesis Testing 313

26.1 Introduction 314
26.2 The Standard Error of Estimate s_i 314
26.3 Sampling Distribution and Hypothesis Tests for |3 315
26.3.1 The slope of the regression equation 315
26.4 Sampling Distribution and Hypothesis Tests for P 319
26.4.1 The Y-interccpt of the regression equation 319
26.5 Hypothesis Test for the Conditional Mean of the Regression Equation 319
26.6 The; Coefficient of Determination 321
26.7 Observations about Linear Regression 325
26.S Exercises 325

Chapter 27 Simple Linear Regression: Case Study 331

27.1 Introduction 332
27.2 The Statement for the Case Study 332
27.3 Case Study Analysis 332
27.4 Exercises 337

Chapter 28 Introduction to Multiple Linear Regression 339

28'1 Introduction 340
28.2 Sample Study 340
28.3 Multiple Linear Regression Model 340
28.4 The Relative Importance of Predictors -.346
28.5 The Significance of R^2 348
28.6 Inferences about the Regression Coefficients 349
28.7 Exercises 352
Reference 355

Chapter 29 Multiple Linear Regression: Case Study 357

29.1 Introduction 358
29.2 Case Study: Prediction of Rental Car's Basic Price 358
29.3 Forward, Stepwise, and Backward Selection 363
29.3.1 Forward selection 363
29.3.2 Stepwise selection 364
29.3.3 Backward elimination 364
29.3.4 Setwise selection 364
29.4 Exercises 365

Chapter 30 Multiple Linear Regression: Handling Violations of Restrictions 373

30.1 Introduction 374
30.2 Visual Tests for Verifying the Regression Assumptions 374
30.2.1 Test for linearity 374
30.2.2 Test for independent errors 374
30.2.3 Test for normally distributed errors 375
30.2.4 Test for homoseedastieitv 375
30.3 The Problem of Multicollinearity 375
30.4 Ridge Regression 376
30.5 Categorical Predictors 376
30.6 Cumlinear Regression 378
30.7 Transformations 378
30.8 Outliers 379
30.9 Exercises 380

UNIT 7 EXPERIMENTAL DESIGNS 383

Chapter 31 One-Way Analysis of Variance 385

31.1 Introduction 386
31.2 Sample Study: *(Winning Database Configurations, Continued)* 386
3.1.3 Introduction to A NOVA 386
3.1.4 Tests of Homogeneity of Variance 394

31.5 Multiple Comparisons 396
31.6 Exercises 398
Reference * 402

Chapter 32 Two-Way Analysis of Variance 403

32.1 Introduction 404
32.2 Two-Way ANOVA with One Entry Per Cell 41 J.
32.3 Randomized-Block Designs 412
32.4 Latin Square Design 414
32.5 Exercises 415
32.6 Two-Way ANOVA Project I 421
Reference 422

Chapter 33 Analysis of Covariance 423

33.1 Introduction 424
33.2 Exercises 431

Chapter 34 Experimental Designs 435

34.3 Introduction 436
34.2 Classification of Designs 436
34.3 Experimental Design Definitions 438
34.4 Avoiding Pitfalls 441
34.5 Experimental Goals 441
34.5.1 Experimental design in practice 442
34.6 Exercises 443
Experimental Design Project 445

UNIT 8 NONPARAMETRIC TESTS AND COMMONLY USED DISTRIBUTIONS 447

Chapter 35 Random Number Generation 451

35.1 Introduction 452
35.2 Random Number Generation 452
35.3 Desired Properties of a Good Generator 452
35.4 Linear Congruential Generators 453
35.5 Multiplicative Linear Congruential Generators 454
35.6 Extended Fibonacci Generators 454
35.7 Combined Generators 454
35.8 Seed Selection 455
35.9 Myths about Random Number Generation 455
35.10 Exercise 456
Random Number Project I 456
Random Number Project TI 456
References 457

Chapter 36 Random Van ate Generation 459

36:1 Introduction 460
36.2 The Inverse Transformation 460
36.3 The Rejection Method 461
36.4 The Composition Method -462
36.5 The Convolution Method 464
36.6 Exercises 464

Chapter 37 Testing for Randomness 467

37.1 Introduction 468
37.2 The Frequency Test 468
37.3 The Gap Test 470
37.4 The Poker Test 472
37.5 The Runs Test 473
37.6 Runs Above and Below a Central Value 475
37.7 Runs Up and Down 475
37.8 The Kohnogorov Goodness-of-Fit Test 476
37.9 The Kolmogorov-Smirnov Two-Sample Test 478
37.10 Exercises 480

Chapter 38 Nonparametric Substitutes for Some Familiar Parametric Tests 481

38.1 Introduction 482
38.2 The Mann-Whitney Test 482
38.3 The Wilcoxou Matched-Pairs Sigued-Rank Test 484
38.4 The Kruskal-Wallis Test 485
38.5 The Spearman Rank Correlation Coefficient 487
38.6 Exercises 490

Chapter 39 Commonly Used Distributions 493

39.1 Introduction 494
39.2 The Bernoulli Distribution 494
39.3 The Binomial Distribution 494
39.4 The Chi-Square Distribution 495
39.5 The Exponential Distribution 496
39.6 The F Distribution 497
39.7 The Gamma Distribution 498
39.8 The Geometric Distribution 499
39.9 The Normal Distribution 500
39.10 The Poisson Distribution 502
39.11 The Student *I* Distribution 502
39.12 The Continuous Uniform Distribution 503
39.13 The Discrete Uniform Distribution 504
39.14 Exercises 505
Reference 506

UNIT 9 RESEARCH METHODS 507

Chapter 40 A Guide to Research „ 509

40.1 Introduction 510
40.2 Conceptual Framework 511
40.3 Reliability, Validity, Utility, and Usage 512
40.4 The Scientific Method 513
40.4.1 Research 514
40.4.2 Problem 514
40.4.3 Project experimentation 514
40.4.4 Project conclusion 515
40.5 Topic Research 515
40.6 Project Research 516
40.7 Scientific Writing 517
40.8 Matters of Ethical Concern in Research 517
40.9 Exercises 518
40.10 Project 519
40.11 Problem Solving 520

Chapter 41 Survey and Field Research 521

4!. **I** Introduction 522
41.2 Types ol Surveys 522
41.2.1 Questionnaires 523
41.2.2 Interviews 523
41.2.3 Writing your own survey questions 524
41.3 Survey Research Sample 525
41.4 Sampling 525
41.4.1 Simple random sampling 526
41.4.2 Stratified sampling 527
41.4.3 Cluster sampling 528
41.4.4 Alternative sampling methods 529
41.4.4.1 Systematic Sampling 529
41.4.4.2 Double Sampling 529
41.5 Sampling Errors 530
41.6 Field Studies 530
41.7 Field Research Example 532
41.S Suivey Research Exercises 533
41.9 Sampling Exercises 533
41:10 Hypothetical Research Project 534
41.1.1 Field Study Exercises 535
41.12 Suivey Research Project 535
41.13 Sampling References 536

Chapter 42 A Methodology for Model Construction **537**

42^1 Introduction 538
42.2 Sample Study 538
42.3 Lessons Learned 539
42.3.1 Step 1: Validate your data -539
42.3.1.1 Statement of Problem 539
42.3.1.2 Purpose 540
42.3.2 Step 2: Select the variables and model 540
42.3.2.1 Operational Definitions 540
42.3.2.2 Questions Answered 541
42.3.2.3 Limitations 541
42.3.2.4 Judgment Analysis 542
42.3.3 Step 3: Perform preliminar)'analyses 543
42.3.3.1 Predictor Variables 543
42.3.3.2 Criterion Variables 544
42.3.3.3 Questions Asked 545
42.3.3.4 Method Used for Organizing Data 545
42.3.4 Step 4: Determine design and methodologies of the study 549
42.3.4.1 Subjects judged 549
42.3.4.2 Judges 549
42.3.4.3 Strategy Used for Obtaining Data 549
42.3.5 Step 5: Check the model 552
42.3.6 Step 6: Extract the equation 553
42.3.7 Conclusion 556
42.4 General Modeling Considerations 557
42.4.1 Planning the model building process 557
42.5 Development of the mathematical model 558
42.6 Verification and Maintenance of the Mathematical Model 558
42.7 Exercises 558
42.8 Clustering Project 559
42.9 JPC Project 562
References 563

Chapter 43 A Guide to Statistical Software **565**

43.1 Introduction 566
43.2 Design Constructs 566
43:2.1 Sets of programs 567
43.2.2 Sets of subroutines 567
43.2.3 Large, multiple-use programs 567
43.2.4 Application compilers 567

43.3 Problem Areas 568
43.4 Desirable Package Features 570
43.5 Evaluation Checklist * 570
43.6 Exercises 573
43.7 Statistical Computing Exercises 573

Chapter 44 Product Development 575

44.1 Introduction 576
44.2 Sample Study 576
44.3 Types of Research 577
44.4 Adequacy Testing Theory in a Field of Study 578
44.5 Product Development Methodology 579
44.6 Exercises 582
44.7 Conjoint Analysis Project 583

Chapter 45 The Axiomatic Research Method 585

45.1 Introduction 586
45.2 Sample Study 586
45.3 Axiomatic Development as a Research Method 586
45.4 Definition Jbr the Relational Data Model 587
45.5 Strong Relations 589
45.6 Strong Inter-related Relations 591
45.6.1 BCNF algorithm 592
45.6.2 Test for functional dependency preserving 592
45.6.3 To find a key 592
45.7 Summary 592
45.8 The Axiomatic Method as a Tool for Research 593
45.9 Exercises 593
45.10 Semantic Data Models Project 594
45.11 Extended Relational Models 594
Reference 595

UNIT 10 SIMULATION AND RESEARCH ISSUES 597

Chapter 46 Monte Carlo Simulation Overview 599

46.1 Introduction 600
46.2 Application: Determination of the Number of Production Units 600
46.3 Exercises 604
46.3.1 Games and Simulation Project 604
46.3.2 Risk Analysis Project 604

Chapter 47 How to Conduct a Simulation 607

47.1 Introduction 608
47.2 Manufacturing Example 608

47.3 Discrete Event Simulation 615
47.4 Discrete Event Simulation of a Simplified Token Ring 616
47.5 Summary * 619
47.6 Projects 620

Chapter 48 A Research Study Vignette 623

48.1 Introduction 624
48.2 The Vignette Setting 624
48.3 Preliminary Research Study Statement 624
48.4 Background Review 625
45.5 Formulating a Project that can be Resolved 625
48.6 Attribute Screening 626
48.7 Stuclv Design 62S
48.8 Reporting Results 631
48.9 Promoting Research Results 632
48.10 Vignette Closing 632
48.1.1 Chapter Summary 632
48.12 Exercises 633

Appendix A Statistical Tables 635

Table A-l The Normal Distribution 636
Table A-2 Binomial Probabilities 638
Table A-4a Critical Values of the t Distribution 644
Table A-4b Critical Values of the F Distribution . 647
Table A-5 Critical Values of tlie Studentized Range Statistic and Dunnetts Test 649
Table A-6 Critical Values of Dunn's Test 651
Table A-7 Critical Values of the Chi-Square Distribution 653
Table A-S Critical Values of the Binomial Test 654
Table A-9 Critical Values of the Mann-Whitney U Test 655
Table A-10 Critical Values of the Wilcoxon Ranked Sums Test 657
Table A-l 1 Critical Values of the Wilcoxon Signed Ranks Test 659
Table A-12 Critical Values of the Correlation Coefficient 660
Table A-l3 Transforming r to Z 661
Table A-14 Statistical Power of the Z Test 663
Table A-15 Statistical Power of the t Test for One Sample or Two Related Samples 664
Table A-16 Statistical Power of the t Test for Two Independent Samples 665
Table A-17 Statistical Power of the Analysis of Variance 667
Table A-IS Statistical Power of the Correlation Coefficient 670
Table A-I9 Required Sample Size 674
Table A-20 The Poisson Distribution 676

Table A-21 Critical Values of the Spearman Correlation Coefficient 680
Table A-22 Critical Values for Total Number of Runs (U) » 681
Table A-23 Critical Values for the Hartley Tost of Homogeniety of Variance 682
Table A-24 The Cochran Test for Homogoniety of Variances 683
Table A-25 Table of Percentage Points of Kolmogorov Statistics 685
Table A-26 Quautilcs ol the SmirnovTest Statistic for Two Samples of Equal Size n 686
Table A-27 Quantiles of the SmirnovTest Statistic for Two Samples of Different Size 687

Appendix B Data Files **691**

Table B-l American Cities Database 691
Table B-2 American Cities Database- Version 2 694
Table B-3 Auto Eile 696
Table B-4 Cost of Living 697
Table B-5 East Food **TOO**
Table B-6 Health File 702
Table B-7 Interest Rate Voltality 706
Table B-8 Stock Prices 708

Appendix C Articles **713**

Comparative Study ol Graduate Information Systems and MBA Students Cognitive Styles 7.14
Ethnographic Study ol MSIS Student Project Team Dynamics 723

Appendix D Solutions to Seleeted Exercises (On *Companion DVD)* **735**

Index **737**