# Table of Contents

**Preface**  xiii  
The Software  xiii  
*JMP Start Statistics*, Fifth Edition  xiv  
SAS  xv  
This Book  xv  

## 1 Preliminaries  1  
What You Need to Know  1  
  …about your computer  1  
  …about statistics  1  
Learning About JMP  1  
  …on your own with JMP Help  1  
  …hands-on examples  2  
  …using Tutorials  2  
  …reading about JMP  2  
Chapter Organization  3  
Typographical Conventions  5  

## 2 JMP Right In  7  
Hello!  7  
First Session  9  
  Tip of the Day  9  
  The JMP Starter (Macintosh)  9  
  The JMP Home Window (Windows)  10  
Open a JMP Data Table  11  
Launch an Analysis Platform  14  
Interact with the Report Surface  15  
Special Tools  18  
Customize JMP  19  
Modeling Type  20  
  Analyze and Graph  21  
Navigating Platforms and Building Context  22  
Contexts for a Histogram  22  

---

## Table of Contents

### 3 Data Tables, Reports, and Scripts 27

- **Overview** 27
- **The Ins and Outs of a JMP Data Table** 29
  - Selecting and Deselecting Rows and Columns 30
  - Mousing Around a Spreadsheet: Cursor Forms 30
- **Creating a New JMP Table** 32
  - Define Rows and Columns 32
  - Enter Data 35
  - The New Column Command 36
  - Plot the Data 37
- **Importing Data** 39
  - Importing Text Files 40
  - Importing Other File Types 44
  - Copy, Paste, and Drag Data 44
- **Moving Data Out of JMP** 45
- **Working with Graphs and Reports** 47
  - Copy and Paste 47
  - Drag Report Elements 47
  - Context Menu Commands 48
- **Juggling Data Tables** 49
  - Data Management 49
    - Give New Shape to a Table: Stack Columns 51
- **Creating Summary Statistics** 53
  - Create Summary Statistics with the Summary Command 54
  - Create Summary Statistics with Tabulate 56
- **Working with Scripts** 58

### 4 Formula Editor Adventures 61

- **Overview** 61
- **The Formula Editor Window** 63
  - The Formula Editor and the JMP Scripting Language 64
    - A Quick Example 64
- **Formula Editor: Pieces and Parts** 68
  - Terminology 68
  - The Formula Editor Control Panel 69
  - The Keypad Functions 71
  - The Formula Display Area 72
Function Definitions 73
Row Function Examples 74
Conditional Expressions and Comparison Operators 77
Summarize Down Columns or Across Rows 82
Random Number Functions 87
Local Variables and Table Variables 91
Some Nice Examples Involving Dates 93
Tips on Building Formulas 94
Examining Expression Values 94
Cutting, Dragging, and Pasting Formulas 94
Selecting Expressions 95
Tips on Editing a Formula 95
Exercises 96

5 What Are Statistics? 99
Overview 99
Ponderings 101
The Business of Statistics 101
The Yin and Yang of Statistics 101
The Faces of Statistics 102
Don’t Panic 103
Preparations 105
Three Levels of Uncertainty 105
Probability and Randomness 106
Assumptions 106
Data Mining? 107
Statistical Terms 108

6 Simulations 113
Overview 113
Rolling Dice 115
Rolling Several Dice 118
Flipping Coins, Sampling Candy, or Drawing Marbles 118
Probability of Making a Triangle 119
Confidence Intervals 124
Other JMP Simulations 125
Exercises 127

7 Univariate Distributions: One Variable, One Sample 129
Overview 129
Looking at Distributions 131
### Table of Contents

- Probability Distributions 133
  - True Distribution Function or Real-World Sample Distribution 134
  - The Normal Distribution 136

- Describing Distributions of Values 137
  - Generating Random Data 137
  - Histograms 138
  - Stem-and-Leaf Plots 139
  - Outlier and Quantile Box Plots 141
  - Mean and Standard Deviation 143
  - Median and Other Quantiles 144
  - Mean versus Median 144
  - Other Summary Statistics: Skewness and Kurtosis 145
  - Extremes, Tail Detail 145

- Statistical Inference on the Mean 146
  - Standard Error of the Mean 146
  - Confidence Intervals for the Mean 146
  - Testing Hypotheses: Terminology 148
  - The Normal z-Test for the Mean 150
  - Case Study: The Earth’s Ecliptic 151
  - Student’s t-Test 153
    - Comparing the Normal and Student’s t Distributions 154
    - Testing the Mean 155
    - The p-Value Animation 156
    - Power of the t-Test 158

- Practical Significance vs. Statistical Significance 160

- Examining for Normality 162
  - Normal Quantile Plots 162
  - Statistical Tests for Normality 165

- Special Topic: Practical Difference 168

- Special Topic: Simulating the Central Limit Theorem 170

- Seeing Kernel Density Estimates 172

- Exercises 173

8 The Difference Between Two Means 177

- Overview 177

- Two Independent Groups 179
  - When the Difference Isn’t Significant 179
  - Check the Data 179
  - Launch the Fit Y by X Platform 181
  - Examine the Plot 181
  - Display and Compare the Means 182
  - Inside the Student’s t-Test 183
  - Equal or Unequal Variances? 184
9 Comparing Many Means: One-Way Analysis of Variance 217
   Overview 217
   What Is a One-Way Layout? 219
   Comparing and Testing Means 221
      Means Diamonds: A Graphical Description of Group Means 222
      Statistical Tests to Compare Means 223
      Means Comparisons for Balanced Data 226
      Means Comparisons for Unbalanced Data 227
   Adjusting for Multiple Comparisons 231
   Are the Variances Equal Across the Groups? 234
   Testing Means with Unequal Variances 237
   Nonparametric Methods 237
      Review of Rank-Based Nonparametric Methods 238
      The Three Rank Tests in JMP 239
   Exercises 241

10 Fitting Curves through Points: Regression 245
   Overview 245
   Regression 247
      Least Squares 247
### Table of Contents

- Seeing Least Squares 248
- Fitting a Line and Testing the Slope 250
- Testing the Slope by Comparing Models 251
- The Distribution of the Parameter Estimates 254
- Confidence Intervals on the Estimates 255
- Examine Residuals 257
- Exclusion of Rows 257
- Time to Clean Up 259

#### Polynomial Models 259
- Look at the Residuals 260
- Higher-Order Polynomials 260
- Distribution of Residuals 261

#### Transformed Fits 262
- Spline Fit 263

- Are Graphics Important? 264
- Why It’s Called Regression 267
- What Happens When X and Y Are Switched? 269

#### Curiosities 271
- Sometimes It’s the Picture That Fools You 271
- High-Order Polynomial Pitfall 272
- The Pappus Mystery on the Obliquity of the Ecliptic 273

- Exercises 274

### 11 Categorical Distributions 277

#### Overview 277

- Categorical Situations 279

- Categorical Responses and Count Data: Two Outlooks 279

- A Simulated Categorical Response 282
  - Simulating Some Categorical Response Data 283
  - Variability in the Estimates 284
  - Larger Sample Sizes 286
  - Monte Carlo Simulations for the Estimators 287
  - Distribution of the Estimates 288

- The X² Pearson Chi-Square Test Statistic 289

- The G² Likelihood-Ratio Chi-Square Test Statistic 290
  - Likelihood Ratio Tests 291
  - The G² Likelihood Ratio Chi-Square Test 292

- Univariate Categorical Chi-Square Tests 292
  - Comparing Univariate Distributions 293
  - Charting to Compare Results 295

- Exercises 297
## 12 Categorical Models 299

**Overview** 299

- Fitting Categorical Responses to Categorical Factors: Contingency Tables 301
  - Testing with $G^2$ and $X^2$ 301
  - Looking at Survey Data 302
  - Car Brand by Marital Status 306
  - Car Brand by Size of Vehicle 307

- Two-Way Tables: Entering Count Data 308
  - Expected Values Under Independence 309
  - Entering Two-Way Data into JMP 309
  - Testing for Independence 310

- If You Have a Perfect Fit 312

- Special Topic: Correspondence Analysis—Looking at Data with Many Levels 313

- Continuous Factors with Categorical Responses: Logistic Regression 315
  - Fitting a Logistic Model 316
  - Degrees of Fit 320
  - A Discriminant Alternative 321
  - Inverse Prediction 322
  - Polytomous Responses: More Than Two Levels 324
  - Ordinal Responses: Cumulative Ordinal Logistic Regression 325

- Surprise: Simpson's Paradox: Aggregate Data versus Grouped Data 328

- Generalized Linear Models 331

- Exercises 336

## 13 Multiple Regression 339

**Overview** 339

- Parts of a Regression Model 341
- Regression Definitions 341

- A Multiple Regression Example 342
  - Residuals and Predicted Values 344
  - The Analysis of Variance Table 347
  - The Whole Model F-Test 348
  - Whole-Model Leverage Plot 348
  - Details on Effect Tests 349
  - Effect Leverage Plots 350

- Collinearity 351
  - Exact Collinearity, Singularity, Linear Dependency 355
  - The Longley Data: An Example of Collinearity 357
  - The Case of the Hidden Leverage Point 359

- Mining Data with Stepwise Regression 362

- Exercises 367

---

14  Fitting Linear Models  371
   Overview  371
   The General Linear Model  373
      Kinds of Effects in Linear Models  374
      Coding Scheme to Fit a One-Way ANOVA as a Linear Model  375
      Regressor Construction  378
      Interpretation of Parameters  379
      Predictions Are the Means  379
      Parameters and Means  379
      Analysis of Covariance: Continuous and Categorical Terms in the Same Model  380
      The Prediction Equation  383
      The Whole-Model Test and Leverage Plot  384
      Effect Tests and Leverage Plots  385
      Least Squares Means  387
      Lack of Fit  389
      Separate Slopes: When the Covariate Interacts with a Categorical Effect  391
   Two-Way Analysis of Variance and Interactions  394
   Optional Topic: Random Effects and Nested Effects  401
      Nesting  401
      Repeated Measures  403
      Method 1: Random Effects-Mixed Model  404
      Method 2: Reduction to the Experimental Unit  408
      Method 3: Correlated Measurements-Multivariate Model  410
      Varieties of Analysis  413
      Closing Thoughts  413
   Exercises  413

15  Design of Experiments  417
   Overview  417
   Introduction  419
      Key Concepts  419
      JMP DOE  420
   A Simple Design  421
      The Experiment  421
      Enter the Response and Factors  422
      Define the Model  424
      Is the Design Balanced?  427
      Perform Experiment and Enter Data  427
      Analyze the Model  428
      Flour Paste Conclusions  433
      Details of the Design - Confounding Structure  433
   Using the Custom Designer  434
      How the Custom Designer Works  434
Choices in the Custom Designer 435
An Interaction Model: The Reactor Data 436
  Analyzing the Reactor Data 438
  Where Do We Go From Here? 443
Some Routine Screening Examples 445
  Main Effects Only (a Review) 445
  All Two-Factor Interactions Involving A Single Factor 446
  Alias Optimal Designs 448
Response Surface Designs 449
  The Odor Experiment 449
  Response Surface Designs in JMP 449
  Analyzing the Odor Response Surface Design 451
  Plotting Surface Effects 453
  Specifying Response Surface Effects Manually 454
  The Custom Designer vs. the Response Surface Design Platform 455
Split Plot Designs 456
  The Box Corrosion Split-Plot Experiment 457
  Designing the Experiment 457
  Analysis of Split Plot Designs 459
Design Strategies 462
Design of Experiments Glossary 464
Exercises 468

16 Bivariate and Multivariate Relationships 471
Overview 471
Bivariate Distributions 473
Density Estimation 473
  Bivariate Density Estimation 474
  Mixtures, Modes, and Clusters 476
  The Elliptical Contours of the Normal Distribution 477
Correlations and the Bivariate Normal 479
  Simulating Bivariate Correlations 479
  Correlations Across Many Variables 481
  Bivariate Outliers 483
Outliers in Three and More Dimensions 485
Identify Variation with Principal Components Analysis 487
  Principal Components for Six Variables 490
  How Many Principal Components? 491
Discriminant Analysis 492
  Canonical Plot 494
  Discriminant Scores 494
  Stepwise Discriminant Variable Selection 497
Cluster Analysis 497
  Hierarchical clustering work: How Does it Work? 498
  A Real-World Example 501
Some Final Thoughts 503
Exercises 503

17 Exploratory Modeling 505
Overview 505
Recursive Partitioning (Decision Trees) 507
  Growing Trees 509
  Exploratory Modeling with Partition 515
  Saving Columns and Formulas 517
Neural Nets 518
  A Simple Example 519
  Modeling with Neural Networks 522
  Saving Columns 522
  Profiles in Neural 524
Exercises 528

18 Control Charts and Capability 533
Overview 533
What Does a Control Chart Look Like 535
Types of Control Charts 536
  Variables Charts 537
  Attributes Charts 537
  Specialty Charts 537
Control Chart Basics 538
Control Charts for Variables Data 539
  Variables Charts using Control Chart Builder 539
    The Control Chart Builder Work Space 540
    Control Chart Builder Examples 541
Control Charts for Attributes Data 543
Specialty Charts 546
  Presummarize Charts 546
  Levey-Jennings Charts 547
  Uniformly Weighted Moving Average (UWMA) Charts 547
  Exponentially Weighted Moving Average (EWMA) Chart 549
Capability Analysis 550
  What is Process Capability? 550
  Capability for One Process Measurement 552
  Capability for Many Process Measurements 554
  Capability for Time-Ordered Data 556
## A Few Words About Measurement Systems 559

### Exercises 559

### 19 Mechanics of Statistics 561

#### Overview 561

<table>
<thead>
<tr>
<th>Springs for Continuous Responses</th>
<th>563</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitting a Mean</td>
<td>563</td>
</tr>
<tr>
<td>Testing a Hypothesis</td>
<td>564</td>
</tr>
<tr>
<td>One-Way Layout</td>
<td>565</td>
</tr>
<tr>
<td>Effect of Sample Size Significance</td>
<td>565</td>
</tr>
<tr>
<td>Effect of Error Variance on Significance</td>
<td>566</td>
</tr>
<tr>
<td>Experimental Design’s Effect on Significance</td>
<td>567</td>
</tr>
<tr>
<td>Simple Regression</td>
<td>567</td>
</tr>
<tr>
<td>Leverage</td>
<td>569</td>
</tr>
<tr>
<td>Multiple Regression</td>
<td>570</td>
</tr>
<tr>
<td>Summary: Significance and Power</td>
<td>570</td>
</tr>
</tbody>
</table>

#### Mechanics of Fit for Categorical Responses 570

<table>
<thead>
<tr>
<th>How Do Pressure Cylinders Behave?</th>
<th>570</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimating Probabilities</td>
<td>572</td>
</tr>
<tr>
<td>One-Way Layout for Categorical Data</td>
<td>573</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>575</td>
</tr>
</tbody>
</table>

### Appendix A

#### Analyze and Graph Menu Commands 577

<table>
<thead>
<tr>
<th>Analyze Menu</th>
<th>577</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Graph Menu</td>
<td>582</td>
</tr>
</tbody>
</table>

### Answers to Selected Exercises 585

| Chapter 4, "Formula Editor Adventures" | 585 |
| Chapter 7, "Univariate Distributions: One Variable, One Sample" | 587 |
| Chapter 8, "The Difference Between Two Means" | 591 |
| Chapter 9, "Comparing Many Means: One-Way Analysis of Variance" | 593 |
| Chapter 10, "Fitting Curves through Points: Regression" | 597 |
| Chapter 11, "Categorical Distributions" | 600 |
| Chapter 12, "Categorical Models" | 601 |
| Chapter 13, "Multiple Regression" | 603 |
| Chapter 14, "Fitting Linear Models" | 604 |
| Chapter 15, "Design of Experiments" | 605 |
| Chapter 16, "Bivariate and Multivariate Relationships" | 606 |
| Chapter 17, "Exploratory Modeling" | 607 |
| Chapter 18, "Control Charts and Capability" | 607 |
References and Data Sources  609
Technology License Notices  615
Index  617