

Table of Contents

Preface xiii

The Software xiii

JMP Start Statistics, Fourth 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 2

Typographical Conventions 4

2 JMP Right In 7

Hello! 7

First Session 8

Open a JMP Data Table 9

Launch an Analysis Platform 12

Interact with the Surface of the Report 13

Special Tools 16

Modeling Type 17

Analyze and Graph 18

The Analyze Menu 18

The Graph Menu 20

Navigating Platforms and Building Context 22

Contexts for a Histogram 22

Contexts for the t-Test 22

Contexts for a Scatterplot 23

Contexts for Nonparametric Statistics 23

The Personality of JMP 24

3 Data Tables, Reports, and Scripts 27

Overview	27
The Ins and Outs of a JMP Data Table	28
<i>Selecting and Deselecting Rows and Columns</i>	28
<i>Mousing Around a Spreadsheet: Cursor Forms</i>	29
Creating a New JMP Table	31
<i>Define Rows and Columns</i>	31
<i>Enter Data</i>	34
<i>The New Column Command</i>	35
<i>Plot the Data</i>	36
<i>Importing Data</i>	38
<i>Importing Text Files</i>	40
<i>Importing Microsoft Excel Files</i>	41
<i>Using ODBC</i>	42
<i>Opening Other File Types</i>	43
<i>Copy, Paste, and Drag Data</i>	44
Moving Data Out of JMP	45
Working with Graphs and Reports	48
<i>Copy and Paste</i>	48
<i>Drag Report Elements</i>	49
<i>Context Menu Commands</i>	49
Juggling Data Tables	50
<i>Data Management</i>	50
<i>Give New Shape to a Table: Stack Columns</i>	52
The Summary Command	54
<i>Create a Table of Summary Statistics</i>	54
Working with Scripts	57

4 Formula Editor Adventures 61

Overview	61
The Formula Editor Window	62
A Quick Example	63
Formula Editor: Pieces and Parts	66
<i>Terminology</i>	66
<i>The Formula Editor Control Panel</i>	67
The Keypad Functions	69
The Formula Display Area	70
Function Browser Definitions	71
<i>Row Function Examples</i>	72
<i>Conditional Expressions and Comparison Operators</i>	75
<i>Summarize Down Columns or Across Rows</i>	78
<i>Random Number Functions</i>	84

Tips on Building Formulas	89
<i>Examining Expression Values</i>	89
<i>Cutting, Dragging, and Pasting Formulas</i>	89
<i>Selecting Expressions</i>	90
<i>Tips on Editing a Formula</i>	90
Exercises	91

5 What Are Statistics? 95

Overview	95
Ponderings	96
<i>The Business of Statistics</i>	96
<i>The Yin and Yang of Statistics</i>	96
<i>The Faces of Statistics</i>	97
<i>Don't Panic</i>	98
Preparations	99
<i>Three Levels of Uncertainty</i>	99
<i>Probability and Randomness</i>	100
<i>Assumptions</i>	100
<i>Data Mining?</i>	101
Statistical Terms	102

6 Simulations 107

Overview	107
Rolling Dice	108
<i>Rolling Several Dice</i>	110
<i>Flipping Coins, Sampling Candy, or Drawing Marbles</i>	111
Probability of Making a Triangle	112
Confidence Intervals	117

7 Univariate Distributions: One Variable, One Sample 119

Overview	119
Looking at Distributions	120
<i>Probability Distributions</i>	122
<i>True Distribution Function or Real-World Sample Distribution</i>	123
<i>The Normal Distribution</i>	124
Describing Distributions of Values	126
<i>Generating Random Data</i>	126
<i>Histograms</i>	127
<i>Stem-and-Leaf Plots</i>	128
<i>Outlier and Quantile Box Plots</i>	130
<i>Mean and Standard Deviation</i>	132

- Median and Other Quantiles* 133
- Mean versus Median* 133
- Higher Moments: Skewness and Kurtosis* 134
- Extremes, Tail Detail* 134
- Statistical Inference on the Mean 135
 - Standard Error of the Mean* 135
 - Confidence Intervals for the Mean* 135
 - Testing Hypotheses: Terminology* 138
 - The Normal z-Test for the Mean* 139
 - Case Study: The Earth's Ecliptic* 140
 - Student's t-Test* 142
 - Comparing the Normal and Student's t Distributions* 143
 - Testing the Mean* 144
 - The p-Value Animation* 145
 - Power of the t-Test* 148
- Practical Significance vs. Statistical Significance 149
- Examining for Normality 152
 - Normal Quantile Plots* 152
 - Statistical Tests for Normality* 155
- Special Topic: Practical Difference 158
- Special Topic: Simulating the Central Limit Theorem 160
- Seeing Kernel Density Estimates 161
- Exercises 162

8 The Difference between Two Means 167

- Overview 167
- Two Independent Groups 168
 - When the Difference Isn't Significant* 168
 - Check the Data* 168
 - Launch the Fit Y by X Platform* 170
 - Examine the Plot* 171
 - Display and Compare the Means* 171
 - Inside the Student's t-Test* 173
 - Equal or Unequal Variances?* 174
 - One-Sided Version of the Test* 176
 - Analysis of Variance and the All-Purpose F-Test* 177
 - How Sensitive Is the Test?*
 - How Many More Observations Are Needed?* 180
 - When the Difference Is Significant* 182
- Normality and Normal Quantile Plots 184
- Testing Means for Matched Pairs 186
 - Thermometer Tests* 187
 - Look at the Data* 188

<i>Look at the Distribution of the Difference</i>	188
<i>Student's t-Test</i>	189
<i>The Matched Pairs Platform for a Paired t-Test</i>	190
<i>Optional Topic:</i>	
<i>An Equivalent Test for Stacked Data</i>	193
The Normality Assumption	195
Two Extremes of Neglecting the Pairing Situation: A Dramatization	197
A Nonparametric Approach	202
<i>Introduction to Nonparametric Methods</i>	202
<i>Paired Means: The Wilcoxon Signed-Rank Test</i>	202
<i>Independent Means: The Wilcoxon Rank Sum Test</i>	205
Exercises	205

9 Comparing Many Means: One-Way Analysis of Variance 209

Overview	209
What Is a One-Way Layout?	210
Comparing and Testing Means	211
Means Diamonds: A Graphical Description of Group Means	213
Statistical Tests to Compare Means	214
Means Comparisons for Balanced Data	217
Means Comparisons for Unbalanced Data	217
Adjusting for Multiple Comparisons	222
Are the Variances Equal Across the Groups?	224
<i>Testing Means with Unequal Variances</i>	228
Nonparametric Methods	228
<i>Review of Rank-Based Nonparametric Methods</i>	228
<i>The Three Rank Tests in JMP</i>	229
Exercises	231

10 Fitting Curves through Points: Regression 235

Overview	235
Regression	236
<i>Least Squares</i>	236
<i>Seeing Least Squares</i>	237
<i>Fitting a Line and Testing the Slope</i>	238
<i>Testing the Slope by Comparing Models</i>	240
<i>The Distribution of the Parameter Estimates</i>	242
<i>Confidence Intervals on the Estimates</i>	243
<i>Examine Residuals</i>	246
<i>Exclusion of Rows</i>	246

Time to Clean Up 247
Polynomial Models 248
 Look at the Residuals 248
 Higher-Order Polynomials 248
 Distribution of Residuals 249
Transformed Fits 250
 Spline Fit 251
Are Graphics Important? 252
Why It's Called Regression 254
What Happens When X and Y Are Switched? 256
Curiosities 259
 Sometimes It's the Picture That Fools You 259
 High-Order Polynomial Pitfall 260
 The Pappus Mystery on the Obliquity of the Ecliptic 261
Exercises 262

11 Categorical Distributions 265

Overview 265
Categorical Situations 266
Categorical Responses and Count Data: Two Outlooks 266
A Simulated Categorical Response 269
 Simulating Some Categorical Response Data 269
 Variability in the Estimates 271
 Larger Sample Sizes 272
 Monte Carlo Simulations for the Estimators 273
 Distribution of the Estimates 274
The X^2 Pearson Chi-Square Test Statistic 275
The G^2 Likelihood-Ratio Chi-Square Test Statistic 276
 Likelihood Ratio Tests 277
 The G^2 Likelihood Ratio Chi-Square Test 277
Univariate Categorical Chi-Square Tests 278
 Comparing Univariate Distributions 278
 Charting to Compare Results 280
Exercises 281

12 Categorical Models 283

Overview 283
Fitting Categorical Responses to Categorical Factors: Contingency Tables 284
 Testing with G^2 and X^2 284
 Looking at Survey Data 285

<i>Car Brand by Marital Status</i>	288
<i>Car Brand by Size of Vehicle</i>	289
Two-Way Tables: Entering Count Data	289
<i>Expected Values Under Independence</i>	290
<i>Entering Two-Way Data into JMP</i>	291
<i>Testing for Independence</i>	291
If You Have a Perfect Fit	293
Special Topic: Correspondence Analysis— Looking at Data with Many Levels	295
Continuous Factors with Categorical Responses: Logistic Regression	297
<i>Fitting a Logistic Model</i>	298
<i>Degrees of Fit</i>	301
<i>A Discriminant Alternative</i>	302
<i>Inverse Prediction</i>	303
<i>Polytomous Responses: More Than Two Levels</i>	305
<i>Ordinal Responses: Cumulative Ordinal Logistic Regression</i>	306
Surprise: Simpson's Paradox: Aggregate Data versus Grouped Data	310
Generalized Linear Models	313
Exercises	317

13 Multiple Regression 319

Overview	319
Parts of a Regression Model	320
A Multiple Regression Example	321
<i>Residuals and Predicted Values</i>	323
<i>The Analysis of Variance Table</i>	325
<i>The Whole Model F-Test</i>	325
<i>Whole-Model Leverage Plot</i>	326
<i>Details on Effect Tests</i>	326
<i>Effect Leverage Plots</i>	327
Collinearity	328
<i>Exact Collinearity, Singularity, Linear Dependency</i>	332
The Longley Data: An Example of Collinearity	334
The Case of the Hidden Leverage Point	335
Mining Data with Stepwise Regression	337
Exercises	341

14 Fitting Linear Models 345

Overview	345
The General Linear Model	346
<i>Kinds of Effects in Linear Models</i>	347
<i>Coding Scheme to Fit a One-Way ANOVA as a Linear Model</i>	349

<i>Regressor Construction</i>	352
<i>Interpretation of Parameters</i>	353
<i>Predictions Are the Means</i>	353
<i>Parameters and Means</i>	353
<i>Analysis of Covariance: Putting Continuous and Classification Terms into the Same Model</i>	354
<i>The Prediction Equation</i>	357
<i>The Whole-Model Test and Leverage Plot</i>	357
<i>Effect Tests and Leverage Plots</i>	358
<i>Least Squares Means</i>	360
<i>Lack of Fit</i>	362
<i>Separate Slopes: When the Covariate Interacts with the Classification Effect</i>	363
Two-Way Analysis of Variance and Interactions	367
Optional Topic: Random Effects and Nested Effects	373
<i>Nesting</i>	374
<i>Repeated Measures</i>	376
<i>Method 1: Random Effects-Mixed Model</i>	377
<i>Method 2: Reduction to the Experimental Unit</i>	380
<i>Method 3: Correlated Measurements-Multivariate Model</i>	382
<i>Varieties of Analysis</i>	384
<i>Summary</i>	385
Exercises	385

15 Bivariate and Multivariate Relationships 387

Overview	387
Bivariate Distributions	388
Density Estimation	388
<i>Bivariate Density Estimation</i>	389
<i>Mixtures, Modes, and Clusters</i>	391
<i>The Elliptical Contours of the Normal Distribution</i>	392
Correlations and the Bivariate Normal	393
<i>Simulation Exercise</i>	393
<i>Correlations Across Many Variables</i>	396
<i>Bivariate Outliers</i>	398
Three and More Dimensions	399
<i>Principal Components</i>	400
<i>Principal Components for Six Variables</i>	402
<i>Correlation Patterns in Biplots</i>	404
<i>Outliers in Six Dimensions</i>	404
Summary	407
Exercises	408

16 Design of Experiments 411

Overview 411

Introduction 412

Experimentation Is Learning 412

Controlling Experimental Conditions Is Essential 412

Experiments Manage Random Variation within A Statistical Framework 412

JMP DOE 413

A Simple Design 413

The Experiment 413

The Response 413

The Factors 414

The Budget 414

Enter and Name the Factors 414

Define the Model 416

Is the Design Balanced? 419

Perform Experiment and Enter Data 420

Analyze the Model 421

Details of the Design 425

Using the Custom Designer 426

Using the Screening Platform 427

Screening for Interactions: The Reactor Data 429

Response Surface Designs 436

The Experiment 436

Response Surface Designs in JMP 436

Plotting Surface Effects 440

Designating RSM Designs Manually 441

The Prediction Variance Profiler 442

Design Issues 446

Routine Screening Examples 450

Design Strategies Glossary 453

17 Exploratory Modeling 457

Overview 457

The Partition Platform 458

Modeling with Recursive Trees 459

Viewing Large Trees 464

Saving Results 466

Neural Networks 467

Modeling with Neural Networks 469

Profiles in Neural Nets 470

Using Cross-Validation 474

Saving Columns 474

Exercises 475

18 Discriminant and Cluster Analysis 477

Overview 477

Discriminant Analysis 478

Canonical Plot 479

Discriminant Scores 479

Cluster Analysis 482

A Real-World Example 486

Exercises 488

19 Statistical Quality Control 489

Overview 489

Control Charts and Shewhart Charts 490

Variables Charts 491

Attributes Charts 491

The Control Chart Launch Dialog 491

Process Information 492

Chart Type Information 493

Limits Specification Panel 493

Using Known Statistics 494

Types of Control Charts for Variables 494

Types of Control Charts for Attributes 499

Moving Average Charts 500

Levey-Jennings Plots 503

Tailoring the Horizontal Axis 504

Tests for Special Causes 505

Westgard Rules 507

Multivariate Control Charts 509

20 Time Series 511

Overview 511

Introduction 512

Lagged Values 512

Testing for Autocorrelation 516

White Noise 518

Autoregressive Processes 519

Correlation Plots of AR Series 522

Estimating the Parameters of an Autoregressive Process 522

Moving Average Processes 524

Correlation Plots of MA Series 525

Example of Diagnosing a Time Series	526
ARMA Models and the Model Comparison Table	528
Stationarity and Differencing	530
Seasonal Models	532
Spectral Density	536
Forecasting	537
Exercises	539

21 Machines of Fit 541

Overview	541
Springs for Continuous Responses	542
<i>Fitting a Mean</i>	542
<i>Testing a Hypothesis</i>	543
<i>One-Way Layout</i>	543
<i>Effect of Sample Size Significance</i>	544
<i>Effect of Error Variance on Significance</i>	545
<i>Experimental Design's Effect on Significance</i>	546
<i>Simple Regression</i>	547
<i>Leverage</i>	548
<i>Multiple Regression</i>	549
<i>Summary: Significance and Power</i>	549
Machine of Fit for Categorical Responses	549
<i>How Do Pressure Cylinders Behave?</i>	549
<i>Estimating Probabilities</i>	551
<i>One-Way Layout for Categorical Data</i>	552
<i>Logistic Regression</i>	554

References and Data Sources 557

Answers to Selected Exercises 561

Chapter 4, "Formula Editor Adventures"	561
Chapter 7, "Univariate Distributions: One Variable, One Sample"	565
Chapter 8, "The Difference between Two Means"	572
Chapter 9, "Comparing Many Means: One-Way Analysis of Variance"	577
Chapter 10, "Fitting Curves through Points: Regression"	584
Chapter 11, "Categorical Distributions"	586
Chapter 12, "Categorical Models"	587
Chapter 13, "Multiple Regression"	590
Chapter 14, "Fitting Linear Models"	591
Chapter 15, "Bivariate and Multivariate Relationships"	593
Chapter 17, "Exploratory Modeling"	594
Chapter 18, "Discriminant and Cluster Analysis"	594

Chapter 20, "Time Series" 595

Technology License Notices 597

Index 599