

# Preface

---

Welcome to **LogXact**, the statistical package for regression procedures featuring exact methods. This manual describes the procedures and facilities available in the LogXact software.

## *New in LogXact 6*

Among the new features in LogXact 6 are:

1. **Polytomous Regression for both Unordered and Ordered Responses** (See Chapter 13)  
Now you can model multivariate categorical response data:
  - Unordered Responses
    - Asymptotic and Exact Estimates for
      - \* Equal Beta
      - \* Unequal Beta
  - Ordinal Responses
    - Proportional Odds Model (Asymptotic Estimates)
    - Adjacent Category Model (Asymptotic and Exact Estimates)
    - Continuation Ratio Model (Asymptotic and Exact Estimates)
2. **Poisson Regression for Stratified data**  
Asymptotic and Exact Estimates for Stratified data. (See Chapter 14)
3. **Logistic Regression: Exact Prediction**  
For the first time in LogXact, methods are provided to carry out exact prediction (See Chapter 7).
4. **Probit Regression** (Asymptotic Estimates) (See Chapter 11)
5. **Complementary Log-Log Model** (Asymptotic Estimates) (See Chapter 12)
6. **Classification Analyses** for Binary Logistic, Probit and Cloglog Models (See Chapters 7, 8, 11 and 12)
7. **Best Subset Selection** for Multiple Linear Regression (See Chapter 6)
8. **Cytel Studio** - A Common Base for LogXact and StatXact (See Cytel Studio User Manual)
  - New Software Environment
    - Now you can have StatXact and LogXact installed as modules in the same software environment and move seamlessly between StatXact and LogXact using a common interface
  - New interface -workspace organizer  
Lets you work on multiple data sets simultaneously, organize their display and save the various outputs.

# Preface

---

- New statistics module (univariate and multivariate) featuring t-tests, ANOVA, Multiple Linear Regression, Canonical Correlation, Discriminant Analysis, Factor Analysis, Multivariate ANOVA, and Principal Components Analysis.
- Plots include Box, Bubble, Density, Scatter, Histogram, Stem and Leaf etc.
- Improved importing and exporting features

## *Acknowledgements*

The development of this version of LogXact was a collaborative effort involving

**Nitin Patel, Pralay Senchaudhuri, V.P.Chandran, Anthiyur Kannappan, Vipul Suru, Aniruddha Deshmukh, Suresh Sonawane, Ajey Godbole, Shailesh Kulthe, Vivek Pradhan, Aneesh Patel, Janhavi Sampgaonkar, Anil Gore, and Cyrus Mehta.**

Other collaborators in the development of present and previous versions of LogXact include: Rajesh Mehta, Rajendra Lagu, Yogesh Gajjar, Chris Corcoran, Brent Coull, Philip O'Neil and Ajay Sathe.

The  $\LaTeX$  typesetting macros for the user manual were developed by Amy Hendrickson of  $\TeX$ nology Inc. The manual cover and software splash screen were designed by Jennifer Brackett.

We owe a special debt to Professor Marvin Zelen for creating an exciting intellectual environment at the Harvard School of Public Health. He originally encouraged us to work on a number of challenging research problems in computational statistics. This research eventually led to the development of StatXact. His generous support of all aspects of our work, spanning more than a decade, cannot be repaid.

We are grateful to the National Institutes of Health for funding the initial development of StatXact and LogXact through a Small Business Innovation Research Grant.

Finally we express our gratitude to Swami Sarvagatananda.

# Table of Contents

---

<b>PREFACE</b>	<b>i</b>
<b>New in LogXact 6</b> . . . . .	<b>i</b>
<b>Acknowledgements</b> . . . . .	<b>ii</b>
<b>Part I</b> <b>REGRESSION MODULE-PRELIMINARIES</b>	
<b>1</b> <b>INTRODUCTION TO LOGXACT</b>	<b>1</b>
<b>1.1</b> <b>What is LogXact</b> . . . . .	<b>1</b>
<b>1.2</b> <b>Types of Inference</b> . . . . .	<b>1</b>
<b>1.3</b> <b>How to Use this Manual</b> . . . . .	<b>3</b>
<b>1.3.1</b> <b>Font Selection</b> . . . . .	<b>5</b>
<b>1.4</b> <b>Before You Begin</b> . . . . .	<b>5</b>
<b>2</b> <b>WHAT IS EXACT LOGISTIC REGRESSION?</b>	<b>7</b>
<b>2.1</b> <b>Osteogenic Sarcoma Data</b> . . . . .	<b>7</b>
<b>2.2</b> <b>Model, Likelihood and Sufficient Statistics</b> . . . . .	<b>8</b>
<b>2.3</b> <b>Large-Sample Likelihood Ratio Test</b> . . . . .	<b>8</b>
<b>2.4</b> <b>Exact Likelihood Ratio Test</b> . . . . .	<b>9</b>
<b>3</b> <b>A LOGXACT TUTORIAL</b>	<b>13</b>
<b>3.1</b> <b>Activating LogXact</b> . . . . .	<b>13</b>
<b>3.2</b> <b>Bringing Data into LogXact</b> . . . . .	<b>13</b>
<b>Description of the Data Set</b> . . . . .	<b>13</b>
<b>Opening a LogXact Data File</b> . . . . .	<b>14</b>
<b>Examining Ungrouped Case Data in the Data Editor</b> . . . . .	<b>15</b>
<b>Converting to Group Data</b> . . . . .	<b>15</b>
<b>3.3</b> <b>Saving Your Data</b> . . . . .	<b>17</b>
<b>As a LogXact file</b> . . . . .	<b>17</b>
<b>3.4</b> <b>Model Specification, Estimation and Testing</b> . . . . .	<b>18</b>

# Table of Contents

---

	Model Specification . . . . .	19
	Parameter Estimation . . . . .	20
3.4.1	Factor Variables . . . . .	23
	Maximum Likelihood Can Fail . . . . .	24
	Exact Estimates Exist . . . . .	24
	Testing Subsets of Regression Parameters . . . . .	25
3.4.2	Exploration Mode . . . . .	27
	Obtainable Output . . . . .	27
3.5	The LogXact Command Language . . . . .	27
	Command Bar . . . . .	27
	Submit Command File . . . . .	29
3.6	Getting On-Line Help . . . . .	30
3.7	Exit . . . . .	30
4	<b>THE REGRESSION MENU AND DIALOG BOXES</b>	<b>33</b>
4.1	Continuous Response . . . . .	33
	4.1.1 Multiple Linear Regression dialog box . . . . .	33
4.2	Binary Response . . . . .	40
	4.2.1 Logistic Model dialog box . . . . .	40
	4.2.2 Probit Model dialog box . . . . .	54
	4.2.3 Complementary Log Log Model dialog box . . . . .	60
4.3	Polytomous Response . . . . .	66
	4.3.1 Unordered Response dialog box . . . . .	66
	4.3.2 Proportional Odds Model dialog box . . . . .	69
	4.3.3 Adjacent Category Model dialog box . . . . .	70
	4.3.4 Continuation Ratio Model dialog box . . . . .	71
4.4	Count Data . . . . .	72
	4.4.1 Poisson Regression dialog box . . . . .	72
4.5	Regression Settings . . . . .	77
	Why Use Options . . . . .	77
	General Options . . . . .	78
	Monte Carlo Options . . . . .	81

<b>5</b>	<b>THE LOGXACT RESULTS</b>	<b>85</b>
5.1	Continuous Response Output . . . . .	85
5.1.1	Estimates for Multiple Linear Regression . . . . .	85
5.1.2	Test Results for Multiple Linear Regression . . . . .	86
5.2	Binary Response Output . . . . .	87
5.2.1	Estimates for Logistic Model . . . . .	87
5.2.2	Test Results for Logistic Model . . . . .	91
5.2.3	Estimates for Probit Model . . . . .	96
5.2.4	Test Results for Probit Model . . . . .	97
5.2.5	Estimates for Complementary Log Log Model . . . . .	98
5.2.6	Test Results for Complementary Log Log Model . . . . .	99
5.3	Polytomous Response . . . . .	100
5.3.1	Estimates for Unordered Response . . . . .	100
5.3.2	Test Results for Unordered Response . . . . .	101
5.3.3	Estimates for Proportional Odds Model . . . . .	103
5.3.4	Estimates for Adjacent Category Model . . . . .	104
5.3.5	Test Results for Adjacent Category Model . . . . .	105
5.3.6	Estimates for Continuation Ratio Model . . . . .	107
5.3.7	Test Results for Continuation Ratio Model . . . . .	108
5.4	Count Data Output . . . . .	111
5.4.1	Estimates for Poisson Regression . . . . .	111
5.5	Distribution worksheet . . . . .	115
5.6	Postfit worksheet . . . . .	115
5.7	Log file . . . . .	116
5.8	Codes Used in the Display of Results . . . . .	117

**Part II CONTINUOUS RESPONSE REGRESSION**

<b>6</b>	<b>MULTIPLE LINEAR REGRESSION (WITH ADVANCED OPTIONS)</b>	<b>121</b>
6.1	Available procedures . . . . .	121
6.1.1	Statistical terms explained . . . . .	122
6.1.2	Collinearity Diagnostics . . . . .	122
	Parameters for Collinearity Diagnostics . . . . .	122

# Table of Contents

---

6.1.3	Residuals . . . . .	123
6.1.4	Influence Statistics . . . . .	123
6.1.5	Subset Selection in Linear Regression . . . . .	124
	Dropping Irrelevant Variables . . . . .	124
	Dropping Independent Variables With Small Coefficient Values . . . . .	125
	Algorithms for Subset Selection . . . . .	127
6.1.6	Examples . . . . .	129
	Example: Werner data . . . . .	129
	Example: Supervisor Performance Data . . . . .	135
Part III	<b>BINARY RESPONSE REGRESSION</b>	
7	<b>LOGISTIC REGRESSION FOR UNSTRATIFIED DATA</b>	<b>145</b>
7.1	Purpose of this Chapter . . . . .	145
7.2	Summary of Results . . . . .	146
7.3	Animal Toxicology Data from the National Toxicology Program . . . . .	147
	The Results Window . . . . .	152
	Test Results . . . . .	153
	Monte Carlo and Markov Chain Monte Carlo results: . . . . .	157
	Estimation Results . . . . .	158
	Factor Variables . . . . .	161
	Monte Carlo Results . . . . .	163
7.4	Predictors of Disease Free Survival for Osteogenic Sarcoma . . . . .	165
	Monte Carlo Results . . . . .	172
7.5	Acute Myeloblastic Leukemia Data . . . . .	173
	Fitting the Model . . . . .	174
	Post-Fit Analysis . . . . .	177
	Exact Estimates . . . . .	182
7.6	Receiver Operating Characteristic (ROC) Curve . . . . .	184
	7.6.1 Terms Explained . . . . .	184
	7.6.2 Example: ROC curve . . . . .	185
7.7	ROC Curve vs Classification Table . . . . .	193

7.8	Classification Table . . . . .	194
7.9	Exact Prediction . . . . .	197
	7.9.1 How Asymptotic Prediction method differs from Exact Prediction Method . . . . .	200
<b>8</b>	<b>LOGISTIC REGRESSION FOR STRATIFIED DATA</b>	<b>203</b>
8.1	Purpose of this Chapter . . . . .	203
8.2	Summary of LogXact Results . . . . .	204
8.3	Running a Statistical Test . . . . .	204
8.4	Birth Complications and Schizophrenia . . . . .	206
	Monte Carlo Results . . . . .	213
8.5	Cross-Over Clinical Trial of Analgesic Efficacy . . . . .	214
	Monte Carlo Results . . . . .	221
8.6	Radiation and Lung Cancer . . . . .	224
	Monte Carlo Results . . . . .	227
8.7	General $m_i$ to $n_i - m_i$ Matching . . . . .	228
<b>9</b>	<b>HANDLING LARGE DATA SETS USING FACTORY DEFAULTS</b>	<b>229</b>
9.1	How LogXact Handles Large Data Sets . . . . .	229
9.2	Example with UTI Data . . . . .	232
	9.2.1 Default Settings . . . . .	235
	Important default settings for Exact and Direct Monte Carlo . . . . .	235
	9.2.2 Adjusting Time and Memory Settings . . . . .	237
	9.2.3 The Work in Progress Message Box . . . . .	241
	Upper Message Area: . . . . .	242
	Lower Message Area: . . . . .	243
	The Skip and Stop Buttons . . . . .	244
	9.2.4 Continuing with the UTI Data Example . . . . .	244
	9.2.5 Direct Monte Carlo Sampling . . . . .	245
9.3	Diarrhea Study . . . . .	251
	Risk Factors Associated with Acute Diarrhea in Hospitalized Patients. . . . .	251
	Opening a LogXact Data File . . . . .	251

# Table of Contents

---

	Model Specification . . . . .	252
	Estimating using the Exact Algorithm . . . . .	253
9.4	Gene-Environment Interaction Study . . . . .	259
	Study Details . . . . .	259
	Opening a LogXact Data File . . . . .	259
	Model Specification, Estimation and Testing . . . . .	260
9.5	Markov Chain Monte Carlo . . . . .	268
<b>10</b>	<b>HANDLING LARGE DATA SETS USING EXPLORATION MODE</b>	<b>275</b>
10.1	Direct Monte Carlo Sampling with Advanced Options . . . . .	275
10.2	User Activity Flow Chart . . . . .	276
10.3	Death Penalty Study . . . . .	278
	Study details . . . . .	278
	Opening the Data File . . . . .	278
	Model Specification . . . . .	279
	Testing of Hypothesis by Monte Carlo . . . . .	281
10.4	Social Behavior Study . . . . .	289
	Study Details . . . . .	289
	Opening the Data File . . . . .	290
	Model Specification and Testing . . . . .	290
	Testing of Hypothesis by Monte Carlo . . . . .	291
10.5	Feline Upper Respiratory Disease . . . . .	297
	Study Details . . . . .	297
	Opening the Data File . . . . .	297
	Model Specification . . . . .	298
10.6	A Word of Caution . . . . .	303
10.7	Practical Approaches when there is Insufficient Memory for an Exact Solution . . . . .	304
10.8	Emergency Job Interruption . . . . .	309
<b>11</b>	<b>PROBIT MODEL</b>	<b>311</b>
11.1	Purpose of this Chapter . . . . .	311
11.2	Probit Model Results . . . . .	311

11.3	<b>Rat Toxicology Data from the National Toxicology Program</b> . . . . .	312
	Probit Estimate Results . . . . .	315
	Test Results . . . . .	316
	Estimation Results in Odds Ratio . . . . .	318
	Post-Fit Analysis . . . . .	319
11.4	<b>ROC Curve and Classification Table</b> . . . . .	321
12	<b>COMPLEMENTARY LOG LOG MODEL</b>	<b>323</b>
12.1	<b>Purpose of this Chapter</b> . . . . .	323
12.2	<b>Summary of LogXact Results</b> . . . . .	323
12.3	<b>Serological Malaria Data</b> . . . . .	324
	The Results . . . . .	327
	Test Results . . . . .	328
	Estimation Results . . . . .	330
	Post-Fit Analysis . . . . .	331
12.4	<b>ROC Curve and Classification Table</b> . . . . .	333
<b>Part IV POLYTOMOUS RESPONSE REGRESSION</b>		
13	<b>POLYTOMOUS REGRESSION</b>	<b>337</b>
13.1	<b>Available Procedures</b> . . . . .	337
13.2	<b>When to Use Each Model</b> . . . . .	338
13.3	<b>Theory and Methods</b> . . . . .	339
	13.3.1 Models . . . . .	339
13.4	<b>Example: Alligator data</b> . . . . .	343
	13.4.1 Unordered Response Model . . . . .	343
	Equal Beta . . . . .	343
	Unequal Beta . . . . .	346
13.5	<b>Example: Wallet Data</b> . . . . .	347
13.6	<b>Examples: Job Satisfaction Data</b> . . . . .	352
	13.6.1 Proportional Odds Model . . . . .	352
	13.6.2 Adjacent Categories Logit Model . . . . .	355

# Table of Contents

---

13.7	Examples: Statistics Grades Example . . . . .	358
13.8	Examples: Happiness and Marital Status . . . . .	362
13.9	Examples: Developmental Toxicity Example . . . . .	367
Part V	COUNT DATA REGRESSION	
14	POISSON REGRESSION	373
14.1	Purpose of this Chapter . . . . .	373
14.2	Defining a Poisson Regression Model . . . . .	374
14.2.1	The Event . . . . .	374
14.2.2	Stratum . . . . .	374
14.2.3	Model Terms . . . . .	374
14.2.4	The Rate Multiplier . . . . .	374
14.3	Summary of LogXact Output for Poisson regression . . . . .	375
14.4	Unstratified Poisson Regression . . . . .	376
14.4.1	Example: Horseshoe Crab Data . . . . .	376
14.5	Stratified Poisson Regression . . . . .	385
14.6	Example . . . . .	386
14.6.1	Example: Welsh Nickel Workers Data . . . . .	386
Part VI	CONTINGENCY TABLES	
15	USING LOGXACT FOR EXACT INFERENCE IN CONTINGENCY TABLES	393
15.1	The Single $2 \times 2$ Table . . . . .	393
15.2	Stratified $2 \times 2$ Tables . . . . .	394
15.3	The Single $2 \times c$ Table . . . . .	396
15.4	Stratified $2 \times c$ Tables . . . . .	397
15.5	Interaction in Stratified $2 \times c$ Tables . . . . .	398
15.6	McNemar's Test . . . . .	398

---

**Part VII    EXERCISES AND EXAMPLES**

<b>16</b>	<b>EXERCISES AND EXAMPLES</b>	<b>403</b>
16.1	Exercise 1 . . . . .	403
16.2	Exercise 2 . . . . .	404
16.3	Exercise 3 . . . . .	404
16.4	Exercise 4 . . . . .	405
16.5	Exercise 5 . . . . .	405
16.6	Exercise 6 . . . . .	406
16.7	Exercise 7 . . . . .	406
16.8	Exercises 8, 9 and 10 . . . . .	407
16.9	Exercise 11 . . . . .	408
16.10	Exercise 12 . . . . .	408
16.11	Hiring of Female Assistant Professors . . . . .	409
16.12	Rasch Model . . . . .	410
16.13	Experimental Conditions for Ingot Rolling . . . . .	411
16.14	Erythrocyte Sedimentation Rate . . . . .	412
16.15	Antibiotics and Diarrhea . . . . .	413
16.16	New Jersey Death Penalty Data . . . . .	415
16.17	Meningococcol Disease Outbreak . . . . .	417
16.18	Condom Failure . . . . .	419
16.19	Outbreak in South Carolina . . . . .	420
16.20	Feline Upper Respiratory Disease . . . . .	421
16.21	Fraudulent Automobile Insurance Claims . . . . .	422
16.22	Auditor Litigation Risk study . . . . .	425
16.23	Alligator Food Choice Example . . . . .	426
16.24	Example 24 . . . . .	427
16.25	Job Satisfaction Example . . . . .	428
16.26	Example 26 . . . . .	429
16.27	Example 27 . . . . .	429
16.28	Example 28 . . . . .	430
16.29	Example 29 . . . . .	430
16.30	Example 30 . . . . .	431

# Table of Contents

---

## Part VIII APPENDICES

<b>A</b>	<b>HYPOTHESIS TESTING AND PARAMETER ESTIMATION IN LOGXACT</b>	<b>437</b>
A.1	Generalized Linear Models . . . . .	437
A.2	Logistic Regression as a GLM . . . . .	439
A.2.1	The Unstratified Model . . . . .	439
A.2.2	The Stratified Model . . . . .	440
A.2.3	Unified Treatment of Both Models . . . . .	440
A.3	Poisson Regression as a GLM . . . . .	441
A.3.1	The Unstratified Model . . . . .	442
A.3.2	The Stratified Model . . . . .	442
A.3.3	Unified Treatment of Both Models . . . . .	443
A.4	Unconditional Maximum Likelihood Inference . . . . .	444
A.4.1	Hypothesis Tests . . . . .	444
A.4.2	Parameter Estimation . . . . .	445
A.5	Conditional Maximum Likelihood Inference . . . . .	447
A.5.1	Hypothesis Testing . . . . .	449
A.5.2	Parameter Estimation . . . . .	450
A.6	Conditional Exact Inference . . . . .	452
A.6.1	Exact Hypothesis Tests . . . . .	452
A.6.2	Displaying the Exact Null Distribution . . . . .	455
A.6.3	Exact Parameter Estimation . . . . .	455
A.7	Conditional Inference for Polytomous Data . . . . .	458
A.7.1	Test Statistics . . . . .	459
A.7.2	Exact Hypothesis Tests . . . . .	460
<b>B</b>	<b>MONTE CARLO METHODS FOR LOGISTIC REGRESSION IN LOGXACT</b>	<b>463</b>
B.1	Direct Monte Carlo . . . . .	463
B.2	Markov Chain Monte Carlo . . . . .	464
B.3	Strategies for Choosing Monte Carlo Method . . . . .	464
<b>C</b>	<b>EXACT PREDICTION FOR LOGISTIC REGRESSION</b>	<b>467</b>

<b>D</b>	<b>GOODNESS OF FIT IN LOGXACT</b>	<b>469</b>
D.1	Model-specific and Data-specific Grouping . . . . .	469
D.2	Overall Model Significance: The Likelihood Ratio Statistic . . . . .	477
D.3	Overall Goodness of Fit: The Deviance . . . . .	478
	GLIM Deviance . . . . .	479
	LogXact Deviance . . . . .	483
	Summing Up the Difference . . . . .	486
D.4	Logistic Regression Diagnostics . . . . .	487
	Residuals . . . . .	488
	Hosmer-Lemeshow Goodness-of-fit Test . . . . .	490
	Influential Groups . . . . .	491
D.5	Goodness of Fit Statistics Provided by LogXact . . . . .	491
<b>E</b>	<b>CLASSIFICATION TABLE</b>	<b>497</b>
<b>F</b>	<b>ADVANTAGES AND DISADVANTAGES OF CONDITIONAL INFERENCE</b>	<b>499</b>
F.1	Advantages of Conditioning . . . . .	499
	Consistency . . . . .	499
	Exact Inference . . . . .	499
	Superior Small-Sample . . . . .	500
	Logical Coherence . . . . .	500
F.2	Obstacles to Conditional Inference . . . . .	500
	Computational Complexity . . . . .	500
	Overconditioning . . . . .	501
<b>G</b>	<b>SIZE LIMITATIONS</b>	<b>503</b>
G.1	Limits for Data Entry . . . . .	503
G.2	Limits for Statistical Procedures . . . . .	503
G.3	Memory and Disk Space Requirement . . . . .	504
<b>H</b>	<b>LOGXACT - OUTPUT OBTAINABLE UNDER BINARY, POLYTOMOUS AND POISSON REGRESSION</b>	<b>505</b>

# *Table of Contents*

---

<b>I</b>	<b>CODES USED IN THE DISPLAY OF RESULTS</b>	<b>509</b>
<b>J</b>	<b>LIST OF ALL LOGXACT COMMAND LANGUAGE COMMANDS</b>	<b>511</b>
<b>J.1</b>	<b>Set Commands . . . . .</b>	<b>511</b>
<b>J.2</b>	<b>Statistical Commands . . . . .</b>	<b>512</b>
<b>K</b>	<b>LOGXACT DATA SETS</b>	<b>517</b>
<b>L</b>	<b>VALIDATION OF STATISTICAL RESULTS</b>	<b>519</b>
<b>M</b>	<b>LIST OF LOGXACT BETA TESTERS</b>	<b>521</b>
	<b>REFERENCES</b>	<b>525</b>
	<b>INDEX</b>	<b>531</b>