



Mehmet Mehmetoglu  
Tor Georg Jakobsen

# APPLIED STATISTICS USING STATA

A Guide for the  
Social Sciences

SAGE

Los Angeles | London | New Delhi  
Singapore | Washington DC | Melbourne

# EXTENDED CONTENTS

Companion Website	xiii
About the Authors	xiv
Preface	xv
<b>1 RESEARCH AND STATISTICS</b>	<b>1</b>
1.1 The methodology of statistical research	2
1.2 The statistical method	3
1.3 The logic behind statistical inference	4
1.3.1 Probability theory	5
1.3.2 Population size	6
1.3.3 Why do I need significance levels if I am investigating the whole population?	7
1.4 General laws and theories	8
1.4.1 Objectivity and critical realism	9
1.5 Quantitative research papers	9
1.6 Concluding remarks	11
Questions	12
Further reading	12
References	13
<b>2 INTRODUCTION TO STATA</b>	<b>15</b>
2.1 What is Stata?	16
2.1.1 The Stata interface	16
2.1.2 How to use Stata	18

2.2	Entering and importing data into Stata	20
2.2.1	Entering data	20
2.2.2	Importing data	20
2.3	Data management	21
2.3.1	Opening data	22
2.3.2	Examining data	23
2.3.3	Making changes to variables	25
2.3.4	Generating variables	27
2.3.5	Subsetting data	30
2.3.6	Labelling variables	30
2.4	Descriptive statistics and graphs	31
2.4.1	Frequency distributions	31
2.4.2	Summary statistics	33
2.4.3	Appending data	37
2.4.4	Merging data	37
2.4.5	Reshaping data	38
2.5	Bivariate inferential statistics	39
2.5.1	Correlation	39
2.5.2	Independent <i>t</i> -test	39
2.5.3	Analysis of variance (ANOVA)	40
2.5.4	Chi-squared test	41
2.6	Conclusion	42
	Questions	43
	Further reading	43
<b>3</b>	<b>SIMPLE (BIVARIATE) REGRESSION</b>	<b>45</b>
3.1	What is regression analysis?	46
3.2	Simple linear regression analysis	47
3.2.1	Ordinary least squares	50
3.2.2	Goodness of fit	52
3.2.3	Hypothesis test for slope coefficient	55
3.2.4	Prediction in linear regression	58
3.3	Example in Stata	59
3.4	Conclusion	63
	Questions	64
	Further reading	64
	References	64
<b>4</b>	<b>MULTIPLE REGRESSION</b>	<b>67</b>
4.1	Multiple linear regression analysis	68
4.1.1	Estimation	69
4.1.2	Goodness of fit and the <i>F</i> -test	70
4.1.3	Adjusted $R^2$	71
4.1.4	Partial slope coefficients	72
4.1.5	Prediction in multiple regression	73
4.1.6	Standardization and relative importance	74

4.2 Example in Stata	75
4.3 Conclusion	81
Questions	81
Further reading	81
References	82
<b>5 DUMMY-VARIABLE REGRESSION</b>	<b>85</b>
5.1 Why dummy-variable regression?	86
5.1.1 Creating dummy variables	86
5.1.2 The logic behind dummy-variable regression	88
5.2 Regression with one dummy variable	89
5.2.1 Example in Stata	90
5.3 Regression with one dummy variable and a covariate	91
5.3.1 Example in Stata	93
5.4 Regression with more than one dummy variable	94
5.4.1 Example in Stata	96
5.4.2 Comparing the included groups	97
5.5 Regression with more than one dummy variable and a covariate	100
5.5.1 Example in Stata	101
5.6 Regression with two separate sets of dummy variables	103
5.6.1 Example in Stata	105
5.7 Conclusion	106
Questions	107
Further reading	107
References	107
<b>6 INTERACTION/MODERATION EFFECTS USING REGRESSION</b>	<b>109</b>
6.1 Interaction/moderation effect	110
6.2 Product-term approach	112
6.2.1 Interaction between a continuous predictor and a continuous moderator	114
6.2.2 Interaction between a continuous predictor and a dummy moderator	118
6.2.3 Interaction between a dummy predictor and a dummy moderator	121
6.2.4 Interaction between a continuous predictor and a polytomous moderator	124
6.3 Conclusion	130
Questions	131
Further reading	131
References	131
<b>7 LINEAR REGRESSION ASSUMPTIONS AND DIAGNOSTICS</b>	<b>133</b>
7.1 Correct specification of the model	135
7.1.1 All <i>X</i> -variables relevant, and none irrelevant	135
7.1.2 Linearity	137

7.1.3 Additivity	146
7.1.4 Absence of multicollinearity	146
7.2 Assumptions about residuals	148
7.2.1 That the error term has a conditional mean of zero	148
7.2.2 Homoscedasticity	149
7.2.3 Uncorrelated errors	150
7.2.4 Normally distributed errors	151
7.3 Influential observations	153
7.3.1 Leverage	153
7.3.2 DFBETA	154
7.3.3 Cook's distance	155
7.4 Conclusion	157
Questions	158
Further reading	158
References	158
<b>8 LOGISTIC REGRESSION</b>	<b>161</b>
8.1 What is logistic regression?	163
8.1.1 Tests of significance	166
8.2 Assumptions of logistic regression	167
8.2.1 Example in Stata	169
8.3 Conditional effects	175
8.4 Diagnostics	178
8.5 Multinomial logistic regression	180
8.6 Ordered logistic regression	185
8.7 Conclusion	189
Questions	190
Further reading	190
References	190
<b>9 MULTILEVEL ANALYSIS</b>	<b>193</b>
9.1 Multilevel data	195
9.1.1 Statistical reasons for using multilevel analysis	198
9.2 Empty or intercept-only model	199
9.2.1 Example in Stata	201
9.3 Variance partition or intraclass correlation	203
9.4 Random intercept model	204
9.5 Level-2 explanatory variables	206
9.5.1 How much of the dependent variable is explained?	208
9.6 Logistic multilevel model	208
9.7 Random coefficient (slope) model	210
9.8 Interaction effects	213
9.9 Three-level models	216
9.9.1 Cross-classified multilevel model	219

9.10 Weighting	221
9.11 Conclusion	223
Questions	223
Further reading	223
References	224
<b>10 PANEL DATA ANALYSIS</b>	<b>227</b>
10.1 Panel data	228
10.2 Pooled OLS	231
10.3 Between effects	236
10.4 Fixed effects (within estimator)	240
10.4.1 Explaining fixed effects	241
10.4.2 Summary of fixed effects	248
10.4.3 Time-fixed effects	249
10.5 Random effects	250
10.6 Time-series cross-section methods	252
10.6.1 Testing for non-stationarity	256
10.6.2 Lag selection	258
10.6.3 The TSCS model	260
10.7 Binary dependent variables	261
10.8 Conclusion	265
Questions	266
Further reading	266
References	266
<b>11 EXPLORATORY FACTOR ANALYSIS</b>	<b>269</b>
11.1 What is factor analysis?	270
11.1.1 What is factor analysis used for?	272
11.2 The factor analysis process	272
11.2.1 Extracting the factors	273
11.2.2 Determining the number of factors	276
11.2.3 Rotating the factors	277
11.2.4 Refining and interpreting the factors	280
11.3 Composite scores and reliability test	281
11.4 Example in Stata	283
11.5 Conclusion	288
Questions	288
Further reading	289
References	289
<b>12 STRUCTURAL EQUATION MODELLING AND CONFIRMATORY FACTOR ANALYSIS</b>	<b>293</b>
12.1 What is structural equation modelling?	294
12.1.1 Types of structural equation modelling	296

12.2	Confirmatory factor analysis	297
12.2.1	Model specification	298
12.2.2	Model identification	299
12.2.3	Parameter estimation	301
12.2.4	Model assessment	302
12.2.5	Model modification	309
12.3	Latent path analysis	311
12.3.1	Specification of the LPA model	312
12.3.2	Measurement part	313
12.3.3	Structural part	317
12.4	Conclusion	319
	Questions	320
	Further reading	321
	References	321
<b>13</b>	<b>CRITICAL ISSUES</b>	<b>325</b>
13.1	Transformation of variables	326
13.1.1	Skewness and kurtosis	326
13.1.2	Transformations	329
13.2	Weighting cases	331
13.3	Robust regression	334
13.4	Missing data	338
13.4.1	Traditional methods for handling missing data	339
13.4.2	Multiple imputation	342
13.5	Conclusion	348
	Questions	348
	Further reading	348
	References	349
	Index	351