

Applied Linear Regression for Longitudinal Data

With an Emphasis on Missing
Observations

Frans E.S. Tan
Shahab Jolani



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

Preface.....	xiii
Acknowledgements	xv
Short Description of Research and Simulation Studies	xvii
1. Scientific Framework of Data Analysis	1
1.1 Validation Modelling: Comparative Studies.....	1
1.2 Predictive Modelling: Predicting a Future Outcome.....	3
1.3 Missing Observations	3
1.3.1 Missing Data Mechanisms.....	4
1.3.2 Patterns of Missing Data.....	8
1.3.3 Randomised versus Nonrandomised Studies	10
1.3.4 Concluding Remarks	10
1.4 Assignments.....	11
2. Revisiting and Shortcomings of Standard Linear	
Regression Models	13
2.1 Standard Linear Regression Modelling	13
2.1.1 Formulation of the Research Question:	
The Cross-Sectional Nutrition Study	13
2.1.2 The Concept of Pearson (Product-Moment)	
Correlation	14
2.1.3 Specification of a Standard Linear Regression Model	
for the Nutrition Study	17
2.1.4 Interpretation of the Model Parameters	21
2.2 Multiple Linear Regression (MLR) Modelling.....	23
2.2.1 Specification of a Multiple Linear Regression Model	
for the Nutrition Study	23
2.2.2 Interpretation of the Model Parameters of the	
Nutrition Study	26
2.3 MLR Conditional on the Subjects	30
2.3.1 A Longitudinal Case Study: Violent-Behaviour Study.....	30
2.3.2 Intermezzo: Dummy Variables	33
2.3.3 Analysis of Violent Data	35
2.4 Multiple Piecewise Linear Regression	37
2.4.1 A Comparative Longitudinal Study: Proximity Study.....	37
2.4.2 Specification of the Multiple Piecewise-Linear	
Regression Model.....	38

2.5	Assumptions of the MLR with Uncorrelated Errors.....	39
2.5.1	Linearity Assumption.....	39
2.5.2	Normality Assumption	40
2.5.3	The Independence Assumption.....	41
2.6	Issues of MLR for a Longitudinal Study.....	41
2.6.1	Implication of the Independence Assumption	41
2.6.2	Shortcoming of the OLS Model	44
2.7	Handling Missing Observations in Cross-Sectional Designs	44
2.7.1	Simple Methods	45
2.7.1.1	Complete-Case Analysis	45
2.7.1.2	Marginal Mean Substitution.....	45
2.7.1.3	Conditional Mean Substitution	46
2.7.1.4	Stochastic Regression Imputation.....	47
2.7.2	Advanced Methods: Direct Likelihood and Multiple Imputation	49
2.7.3	Generating Imputations.....	52
2.7.4	Comparison of Methods	56
2.7.4.1	Partly Missing Dependent Variable Related to Fully Observed Independent Variables.....	56
2.7.4.2	Partly Missing Dependent Variable Related to the Dependent Variable Itself.....	57
2.7.4.3	Partly Missing Independent Variable Unrelated to Fully Observed Dependent Variable	58
2.7.4.4	Partly Missing Independent Variable Related to the Dependent Variable	58
2.7.5	Conclusion and Recommendations.....	59
2.8	Assignments.....	61
3.	An Introduction to the Analysis of Longitudinal Data.....	71
3.1	Examples of Multilevel Designs	71
3.1.1	Cross-Sectional Multilevel Designs.....	72
3.1.1.1	Performance of Pupils in Schools	72
3.1.1.2	Cross-Country Comparison: Cross-Sectional Design	73
3.1.2	Longitudinal Designs	73
3.1.2.1	Cross Country Comparison: Longitudinal Design	73
3.1.2.2	Growth Study.....	75
3.1.2.3	Interpersonal Proximity Study	75
3.2	Comparison of the Multilevel Linear Regression Model with the Standard (OLS) Linear Regression	76
3.2.1	Age-Specific Sex Effect	79

3.2.2	Sex-Specific Age Effect	81
3.3	Accounting for the Multilevel Structure	82
3.3.1	Formulation of the Problem: A Marginal and Subject-Specific Representation	82
3.3.2	Fixed or Random Factor.....	84
3.3.3	Treating Subjects as a Random Factor.....	85
3.3.3.1	Random-Intercept Models	85
3.3.3.2	Random-Slope Models	92
3.3.3.3	Serial Correlation.....	93
3.3.3.4	Example: Random-Effect Model or Marginal Model With and Without Serial Correlation.....	94
3.4	Marginal Models	99
3.5	Handling Missing Observations in Longitudinal Designs	103
3.5.1	Patterns of Missing Data in Longitudinal Studies	103
3.5.2	Complete-Case Analysis in Longitudinal Studies	106
3.5.3	Last Observation Carried Forward in Longitudinal Studies	108
3.5.4	Advanced Missing Data Methods in Longitudinal Studies	109
3.5.4.1	Missing Observations in the Dependent Variable Only	109
3.5.4.2	Missing Observations in the Independent Variables Only.....	115
3.5.4.3	Missing Observations in the Dependent and Independent Variables.....	116
3.5.5	Conclusion and Recommendations.....	117
3.6	Assignments.....	118
4.	Model Building for Longitudinal Data Analysis	123
4.1	Basic Guidelines	123
4.1.1	Step 1. Exploratory Analysis	123
4.1.2	Step 2. Procedure for Choosing the Best Variance-Covariance Structure	124
4.1.3	Step 3. Interpretation of the Regression Parameter Estimates	124
4.2	Analysis of Some Case Studies	125
4.2.1	Best Practice for the Analysis of Proximity Data.....	125
4.2.1.1	Step 1. Exploratory Analysis.....	125
4.2.1.2	Step 2. Procedure for Choosing the Best Variance-Covariance Structure.....	128
4.2.1.3	Step 3. Interpretation of the Parameter Estimates.....	134

4.2.2	Best Practice for the Analysis of the Growth Study	137
4.2.3	Best Practice for the Analysis of the Salsolinol Study	141
4.3	Multiple Imputation of Missing Observations: Building Imputation Models	144
4.3.1	Data Format: Wide versus Long	146
4.3.2	Imputation in Wide Format: The Classical Approach	148
4.3.3	Imputation in a Long Format: The Multilevel Approach	153
4.4	Summary	155
4.5	Assignments	157
5.	Analysis of a Pre/Post Measurement Design	161
5.1	Best Practice for the Analysis of a Pre/Post Measurement Design	161
5.1.1	The Analysis of Covariance (ANCOVA) Approach	161
5.1.2	The Gain-Score Approach	164
5.1.3	A Random Intercept Representation of the Gain-Score and ANCOVA Approach	165
5.2	Case Studies	168
5.2.1	An Example of a Randomised Clinical Trial: Beating the Blues	168
5.2.2	An Example of a Nonexperimental Study: Well-Being, a Life-Event Study	172
5.3	Handling Missing Observations in Pre/Post Measurement Designs	175
5.3.1	The ANCOVA Approach	176
5.3.2	The Gain-Score Approach	183
5.4	Assignments	186
6.	Analysis of Longitudinal Life-Event Studies	189
6.1	Best Practice for the Analysis of a Longitudinal Life-Event Study	189
6.1.1	Choosing between the Gain-Score and ANCOVA Approach	189
6.1.2	Gain-Score Approach of the Well-Being Study	191
6.1.2.1	Sensitivity Analysis for Gain-Score Approach Using Multiple Imputation	195
6.1.3	ANCOVA Approach of the Well-Being Study	199
6.1.3.1	Sensitivity Analysis for ANCOVA Approach Using Multiple Imputation	200

6.2	Extension to More Than Three Time Points and Groups	202
6.2.1	Interpretation of the Regression Parameters of a Gain-Score Model	202
6.2.2	Interpretation of the Regression Parameters of an ANCOVA-Model.....	203
6.3	Assignments.....	204
7.	Analysis of Longitudinal Experimental Studies	207
7.1	Best Practice for the Analysis of Beating the Blues Trial	207
7.2	Assignments.....	212
	References	215
	Index	221