

Contents

1	Introduction	1
1.1	An Example R Session	1
1.2	A Few Useful Concepts and Commands.....	3
1.2.1	Concepts	3
1.2.2	Commands.....	4
1.3	Data Objects and Data Types	9
1.3.1	Vectors of Character Strings	10
1.3.2	Matrices, Lists, and Data Frames.....	12
1.4	Functions and Debugging	15
2	Reading and Transforming Data Format	19
2.1	Reading and Transforming Data	19
2.1.1	Data Layout.....	19
2.1.2	A Simple Questionnaire Example	19
2.1.3	Other Ways to Read in Data	25
2.1.4	Other Ways to Transform Variables	26
2.1.5	Using R to Compute Course Grades	30
2.2	Reshape and Merge Data Frames	31
2.3	Data Management with a SQL Database	33
2.4	SQL Database Considerations.....	35
3	Statistics for Comparing Means and Proportions	39
3.1	Comparing Means of Continuous Variables	39
3.2	More on Manual Checking of Data	42
3.3	Comparing Sample Proportions	43
3.4	Moderating Effect in <code>loglin()</code>	45
3.5	Assessing Change of Correlated Proportions.....	49
3.5.1	McNemar Test Across Two Samples.....	50
4	R Graphics and Trellis Plots	55
4.1	Default Behavior of Basic Commands	55
4.2	Other Graphics	56

4.3	Saving Graphics.....	56
4.4	Multiple Figures on One Screen.....	57
4.5	Other Graphics Tricks	57
4.6	Examples of Simple Graphs in Publications.....	58
4.6.1	http://journal.sjdm.org/8827/ jdm8827.pdf	60
4.6.2	http://journal.sjdm.org/8814/ jdm8814.pdf	63
4.6.3	http://journal.sjdm.org/8801/ jdm8801.pdf	64
4.6.4	http://journal.sjdm.org/8319/ jdm8319.pdf	65
4.6.5	http://journal.sjdm.org/8221/ jdm8221.pdf	66
4.6.6	http://journal.sjdm.org/8210/ jdm8210.pdf	68
4.7	Shaded Areas Under a Curve	69
4.7.1	Vectors in polygon ()	71
4.8	Lattice Graphics.....	72
5	Analysis of Variance: Repeated-Measures	79
5.1	Example 1: Two Within-Subject Factors	79
5.1.1	Unbalanced Designs	83
5.2	Example 2: Maxwell and Delaney	85
5.3	Example 3: More Than Two Within-Subject Factors	88
5.4	Example 4: A Simpler Design with Only One Within-Subject Variable	89
5.5	Example 5: One Between, Two Within	89
5.6	Other Useful Functions for ANOVA	91
5.7	Graphics with Error Bars	93
5.8	Another Way to do Error Bars Using plotCI().....	95
5.8.1	Use Error () for Repeated-Measure ANOVA	96
5.8.2	Sphericity	102
5.9	How to Estimate the Greenhouse–Geisser Epsilon?	103
5.9.1	Huynh–Feldt Correction	105
6	Linear and Logistic Regression	109
6.1	Linear Regression	109
6.2	An Application of Linear Regression on Diamond Pricing	110
6.2.1	Plotting Data Before Model Fitting	111
6.2.2	Checking Model Distributional Assumptions.....	114
6.2.3	Assessing Model Fit.....	115
6.3	Logistic Regression	118
6.4	Log–Linear Models	119
6.5	Regression in Vector–Matrix Notation	120
6.6	Caution on Model Overfit and Classification Errors	122

7	Statistical Power and Sample Size Considerations	129
7.1	A Simple Example	129
7.2	Basic Concepts on Statistical Power Estimation	130
7.3	<i>t</i> -Test with Unequal Sample Sizes.....	131
7.4	Binomial Proportions	132
7.5	Power to Declare a Study Feasible	133
7.6	Repeated-Measures ANOVA	133
7.7	Cluster-Randomized Study Design	135
8	Item Response Theory	139
8.1	Overview	139
8.2	Rasch Model for Dichotomous Item Responses.....	139
8.2.1	Fitting a <code>rasch()</code> Model.....	140
8.2.2	Graphing Item Characteristics and Item Information.....	143
8.2.3	Scoring New Item Response Data	145
8.2.4	Person Fit and Item Fit Statistics	145
8.3	Generalized Partial Credit Model for Polytomous Item Responses	146
8.3.1	Neuroticism Data.....	147
8.3.2	Category Response Curves and Item Information Curves	147
8.4	Bayesian Methods for Fitting IRT Models.....	149
8.4.1	GPCM	149
8.4.2	Explanatory IRT	152
9	Imputation of Missing Data	161
9.1	Missing Data in Smoking Cessation Study	161
9.2	Multiple Imputation with <code>aregImpute()</code>	163
9.2.1	Imputed Data	165
9.2.2	Pooling Results Over Imputed Datasets.....	166
9.3	Multiple Imputation with the <code>mi</code> Package	168
9.4	Multiple Imputation with the <code>Amelia</code> and <code>Zelig</code> Packages.....	171
9.5	Further Reading	173
10	Linear Mixed-Effects Models in Analyzing Repeated-Measures Data	177
10.1	The “Language-as-Fixed-Effect Fallacy”	177
10.2	Recall Scores Example: One Between and One Within Factor	180
10.2.1	Data Preparations.....	180
10.2.2	Data Visualizations	181
10.2.3	Initial Modeling	182
10.2.4	Model Interpretation	182
10.2.5	Alternative Models	186
10.2.6	Checking Model Fit Visually	189
10.2.7	Modeling Dependence	190

10.3	Generalized Least Squares Using <code>gls()</code>	195
10.4	Example on Random and Nested Effects	198
10.4.1	Treatment by Therapist Interaction	200
11	Linear Mixed-Effects Models in Cluster-Randomized Studies	205
11.1	The Television, School, and Family Smoking Prevention and Cessation Project	205
11.2	Data Import and Preparations	206
11.2.1	Exploratory Analyses	207
11.3	Testing Intervention Efficacy with Linear Mixed-Effects Models	210
11.4	Model Equation	213
11.5	Multiple-Level Model Equations	215
11.6	Model Equation in Matrix Notations	216
11.7	Intraclass Correlation Coefficients	220
11.8	ICCs from a Mixed-Effects Model	221
11.9	Statistical Power Considerations for a Group-Randomized Design	223
11.9.1	Calculate Statistical Power by Simulation	223
A	Data Management with a Database	229
A.1	Create Database and Database Tables	229
A.2	Enter Data	230
A.3	Using RODBC to Import Data from an ACCESS Database	232
A.3.1	Step 1: Adding an ODBC Data Source Name	233
A.3.2	Step 2: ODBC Data Source Name Points to the ACCESS File	233
A.3.3	Step 3: Run RODBC to Import Data	234
	References	237
	Index	243