# Contents


Preface xi

1 Basic Concepts 1
1.1 Introduction . . . . . 1
1.2 Examples . . . . . 3
1.2.1 NHANES 2011–2014 Accelerometry Data . . . . . 3
1.2.2 COVID-19 US Mortality Data . . . . . 7
1.2.3 CD4 Counts Data . . . . . 12
1.2.4 The CONTENT Child Growth Study . . . . . 15
1.3 Notation and Methodological Challenges . . . . . 19
1.4 R Data Structures for Functional Observations . . . . . 20
1.5 Notation . . . . . 23

2 Key Methodological Concepts 25
2.1 Dimension Reduction . . . . . 25
2.1.1 The Linear Algebra of SVD . . . . . 26
2.1.2 The Link between SVD and PCA . . . . . 27
2.1.3 SVD and PCA for High-Dimensional FDA . . . . . 28
2.1.4 SVD for US Excess Mortality . . . . . 29
2.2 Gaussian Processes . . . . . 34
2.3 Semiparametric Smoothing . . . . . 36
2.3.1 Regression Splines . . . . . 37
2.3.1.1 Univariate Regression Splines . . . . . 37
2.3.1.2 Regression Splines with Multiple Covariates . . . . . 38
2.3.1.3 Multivariate Regression Splines . . . . . 39
2.3.2 Penalized Splines . . . . . 40
2.3.3 Smoothing as Mixed Effects Modeling . . . . . 43
2.3.4 Penalized Spline Smoothing in NHANES . . . . . 44
2.3.4.1 Mean PA among Deceased and Alive Individuals . . . . . 44
2.3.4.2 Regression of Mean PA . . . . . 46
2.4 Correlation and Multiplicity Adjusted (CMA) Confidence Intervals and Testing . . . . . 47
2.4.1 CMA Confidence Intervals Based on Multivariate Normality . . . . . 48
2.4.2 CMA Confidence Intervals Based on Parameter Simulations . . . . . 50
2.4.3 CMA Confidence Intervals Based on the Nonparametric Bootstrap of the Max Absolute Statistic . . . . . 51
2.4.4 Pointwise and Global Correlation and Multiplicity Adjusted (CMA) p-values . . . . . 52
2.4.5 The Origins of CMA Inference Ideas in this Book . . . . . 53
2.5 Covariance Smoothing . . . . . 54
2.5.1 Types of Covariance Smoothing . . . . . 55
2.5.1.1 Covariance Smoothing for Dense Functional Data . . . . . 56

2.5.1.2 Covariance Smoothing for Sparse Functional Data . . . . . 57
2.5.2 Covariance Smoothing in NHANES . . . . . 58
2.5.3 Covariance Smoothing for CD4 Counts . . . . . 60

**3 Functional Principal Components Analysis** **65**
3.1 Defining FPCA and Connections to PCA . . . . . 65
3.1.1 A Simulated Example . . . . . 66
3.1.1.1 Code for Generating Data . . . . . 67
3.1.1.2 Data Visualization . . . . . 68
3.1.1.3 Raw PCA versus FPCA Results . . . . . 69
3.1.1.4 Functional PCA with Missing Data . . . . . 73
3.1.2 Application to NHANES . . . . . 74
3.1.2.1 Data Description . . . . . 74
3.1.2.2 Results . . . . . 75
3.2 Generalized FPCA for Non-Gaussian Functional Data . . . . . 76
3.2.1 Conceptual Framework . . . . . 78
3.2.2 Fast GFPCA Using Local Mixed Effects . . . . . 80
3.2.3 Binary PCA Using Exact EM . . . . . 83
3.2.4 Functional Additive Mixed Models . . . . . 84
3.2.5 Comparison of Approaches . . . . . 86
3.2.6 Recommendations . . . . . 87
3.3 Sparse/Irregular FPCA . . . . . 88
3.3.1 CONTENT Child Growth Data . . . . . 88
3.3.2 Data Structure . . . . . 90
3.3.3 Implementation . . . . . 91
3.3.4 About the Methodology for Fast Sparse FPCA . . . . . 96
3.4 When PCA Fails . . . . . 96

**4 Scalar-on-Function Regression** **101**
4.1 Motivation and EDA . . . . . 102
4.2 "Simple" Linear Scalar-on-Function Regression . . . . . 106
4.2.1 Model Specification and Interpretation . . . . . 107
4.2.2 Parametric Estimation of the Coefficient Function . . . . . 108
4.2.3 Penalized Spline Estimation . . . . . 113
4.2.4 Data-Driven Basis Expansion . . . . . 118
4.3 Inference in "Simple" Linear Scalar-on-Function Regression . . . . . 123
4.3.1 Unadjusted Inference for Functional Predictors . . . . . 123
4.4 Extensions of Scalar-on-Function Regression . . . . . 126
4.4.1 Adding Scalar Covariates . . . . . 126
4.4.2 Multiple Functional Coefficients . . . . . 127
4.4.3 Exponential Family Outcomes . . . . . 129
4.4.4 Other Scalar-on-Function Regression Models . . . . . 129
4.5 Estimation and Inference Using mgcv . . . . . 130
4.5.1 Unadjusted Pointwise Inference for SoFR Using `mgcv` . . . . . 132
4.5.2 Correlation and Multiplicity Adjusted (CMA) Inference for SoFR . . 134

**5 Function-on-Scalar Regression** **143**
5.1 Motivation and Exploratory Analysis of MIMS Profiles . . . . . 144
5.1.1 Regressions Using Binned Data . . . . . 145
5.2 Linear Function-on-Scalar Regression . . . . . 151
5.2.1 Estimation of Fixed Effects . . . . . 153

5.2.1.1 Estimation Using Ordinary Least Squares . . . 153
5.2.1.2 Estimation Using Smoothness Penalties . . . 155
5.2.2 Accounting for Error Correlation . . . 159
5.2.2.1 Modeling Residuals Using FPCA . . . 161
5.2.2.2 Modeling Residuals Using Splines . . . 166
5.2.2.3 A Bayesian Perspective on Model Fitting . . . 168
5.3 A Scalable Approach Based on Epoch-Level Regressions . . . 170

**6 Function-on-Function Regression** **175**
6.1 Examples . . . 176
6.1.1 Association between Patterns of Excess Mortality . . . 176
6.1.2 Predicting Future Growth of Children from Past Observations . . . 176
6.2 Linear Function-on-Function Regression . . . 176
6.2.1 Penalized Spline Estimation of FoFR . . . 178
6.2.2 Model Fit and Prediction Using FoFR . . . 180
6.2.3 Missing and Sparse Data . . . 181
6.3 Fitting FoFR Using `pffr` in `refund` . . . 181
6.3.1 Model Fit . . . 185
6.3.2 Additional Features of `pffr` . . . 187
6.3.3 An Example of `pffr` in the CONTENT Study . . . 190
6.4 Fitting FoFR Using `mgcv` . . . 194
6.5 Inference for FoFR . . . 198
6.5.1 Unadjusted Pointwise Inference for FoFR . . . 199
6.5.2 Correlation and Multiplicity Adjusted Inference for FoFR . . . 201

**7 Survival Analysis with Functional Predictors** **211**
7.1 Introduction to Survival Analysis . . . 212
7.2 Exploratory Data Analysis of the Survival Data in NHANES . . . 213
7.2.1 Data Structure . . . 213
7.2.1.1 Traditional Survival Analysis . . . 213
7.2.1.2 Survival Analysis with Functional Predictors . . . 215
7.2.2 Kaplan-Meier Estimators . . . 217
7.2.3 Results for the Standard Cox Models . . . 218
7.3 Cox Regression with Baseline Functional Predictors . . . 220
7.3.1 Linear Functional Cox Model . . . 220
7.3.1.1 Estimation . . . 221
7.3.1.2 Inference on the Functional Coefficient . . . 224
7.3.1.3 Survival Curve Prediction . . . 232
7.3.2 Smooth Effects of Traditional and Functional Predictors . . . 234
7.3.3 Additive Functional Cox Model . . . 236
7.4 Simulating Survival Data with Functional Predictors . . . 239

**8 Multilevel Functional Data Analysis** **243**
8.1 Data Structure in NHANES . . . 244
8.2 Multilevel Functional Principal Component Analysis . . . 245
8.2.1 Two-Level Functional Principal Component Analysis . . . 245
8.2.1.1 Two-Level FPCA Model . . . 246
8.2.1.2 Estimation of the Two-Level FPCA Model . . . 246
8.2.1.3 Implementation in R . . . 248
8.2.1.4 NHANES Application Results . . . 249
8.2.2 Structured Functional PCA . . . 252

8.2.2.1 Two-Way Crossed Design . . . . . 253
8.2.2.2 Three-Way Nested Design . . . . . 254
8.3 Multilevel Functional Mixed Models . . . . . 255
8.3.1 Functional Additive Mixed Models . . . . . 258
8.3.2 Fast Univariate Inference . . . . . 259
8.3.3 NHANES Case Study . . . . . 260
8.4 Multilevel Scalar-on-Function Regression . . . . . 262
8.4.1 Generalized Multilevel Functional Regression . . . . . 262
8.4.2 Longitudinal Penalized Functional Regression . . . . . 263

**9 Clustering of Functional Data** **265**
9.1 Basic Concepts and Examples . . . . . 265
9.2 Some Clustering Approaches . . . . . 268
9.2.1 K-means . . . . . 268
9.2.1.1 Clustering States Using K-means . . . . . 268
9.2.1.2 Background on K-means . . . . . 271
9.2.2 Hierarchical Clustering . . . . . 271
9.2.2.1 Hierarchical Clustering of States . . . . . 271
9.2.2.2 Background on Hierarchical Clustering . . . . . 274
9.2.3 Distributional Clustering . . . . . 276
9.2.3.1 Distributional Clustering of States . . . . . 276
9.2.3.2 Background on Distributional Clustering . . . . . 276
9.3 Smoothing and Clustering . . . . . 280
9.3.1 FPCA Smoothing and Clustering . . . . . 280
9.3.2 FPCA Smoothing and Clustering with Noisy Data . . . . . 285
9.3.3 FPCA Smoothing and Clustering with Sparse Data . . . . . 287
9.3.4 Clustering NHANES Data . . . . . 289

**Bibliography** **291**

**Index** **313**