

Genetic Analysis

of Complex Traits Using SAS[®]

Edited by
Arnold M. Saxton

 **sas**[®]

59454

Contents

About the Authors vii

Acknowledgments xi

Chapter 1 Overview 1

1.1 Introduction 1

1.2 Book Organization 2

1.3 SAS Usage 4

1.3.1 Example of a Basic SAS DATA Step 4

1.3.2 Example of a Basic Macro 5

1.4 References 6

PART 1 Classical Quantitative Genetics

Chapter 2 Estimation of Genetic Variances and Covariances by Restricted Maximum Likelihood Using PROC MIXED 11

2.1 Introduction 11

2.2 The Nested Half-Sib and Related Designs 13

2.2.1 Example of Nested Half-Sib Design, One Trait 13

2.2.2 Example of Nested Half-Sib Design, Two Traits 19

2.2.3 Example of Clonal Design, Two Environments 23

2.3 The Diallel and Related Designs 27

2.3.1 Diallel Example 28

2.3.2 Extension to the NC2 Design 33

2.4 Pedigreed Populations 34

2.5 References 34

Chapter 3 More Estimation of Genetic Parameters 35

3.1 Introduction 35

3.2 Genetic Parameters Estimated with Regression 35

3.3 Genetic Gain and Realized Heritability 41

3.4 Inbreeding and Relationship 44

3.5 Heterosis, or Hybrid Vigor 49

3.6 References 54

Chapter 4 Genetic Selection 55

- 4.1 Introduction 55
- 4.2 Single-Trait Selection 56
 - 4.2.1 Individual Selection 56
 - 4.2.2 Selection on Relatives 57
 - 4.2.3 Indirect Selection 58
- 4.3 Independent Culling 59
- 4.4 Selection Index 62
- 4.5 Selection on BLUP 65
- 4.6 References 66

Chapter 5 Genotype-by-Environment Interaction 69

- 5.1 Introduction 69
- 5.2 Modeling Genotype-by-Environment Interaction 72
 - 5.2.1 ANOVA Model with Fixed GEI 72
 - 5.2.2 Linear-Bilinear Models with Fixed GEI 76
 - 5.2.3 Linear Mixed Model Approach to GEI Analysis 82
 - 5.2.4 Generalized Linear Models to Explore GEI 86
- 5.3 Smoothing Spline Genotype Analysis (SSGA) 90
- 5.4 References 94

Chapter 6 Growth and Lactation Curves 97

- 6.1 Introduction 97
- 6.2 Modeling Lactation and Growth Curves 98
- 6.3 Using PROC REG to Fit Lactation and Growth Data 101
- 6.4 Using PROC NLIN to Fit Non-linear Models 110
- 6.5 Repeated Measures Theory for Lactation and Growth Curves 123
- 6.6 PROC MIXED for Test Day Models 125
- 6.7 Prediction of Individual Test Day Data 138
- 6.8 The SPECTRA and ARIMA Procedures for Time Series Prediction 139
- 6.9 References 146

Chapter 7 Empirical Bayes Approaches to Mixed Model Inference in Quantitative Genetics 149

- 7.1 Introduction 149
 - 7.1.1 Hierarchical Models 151
- 7.2 An Example of Linear Mixed Model Inference 153
- 7.3 Empirical Bayes Inference for Normally Distributed Data 157
 - 7.3.1 Empirical Bayes Analysis in the Linear Mixed Model 158
- 7.4 Generalized Linear Mixed Models 160
 - 7.4.1 Inference on Fixed and Random Effects 162
 - 7.4.2 Using the %GLIMMIX Macro for Quantitative Genetic Inference 163
 - 7.4.3 Empirical Bayes Inference in GLMM 166

- 7.5 Empirical Bayes vs. MCMC 169
- 7.6 Final Comments 174
- 7.7 References 175

PART 2 Molecular Genetics

Chapter 8 Gene Frequencies and Linkage Disequilibrium 179

- 8.1 Introduction 179
- 8.2 Single-Locus Frequencies 180
- 8.3 Hardy-Weinberg Proportions 184
- 8.4 Multiple-Locus Frequencies 188
- 8.5 Marker-Trait Association Tests 191
 - 8.5.1 Analyzing Samples of Unrelated Individuals 192
 - 8.5.2 Analyzing Family Data 195
- 8.6 References 200

Chapter 9 The Extended Sib-Pair Method for Mapping QTL 201

- 9.1 Introduction 202
- 9.2 Statistical Model 203
- 9.3 Inferring the Proportion of Genes IBD Shared by Sibs at QTL 204
 - 9.3.1 Proportion of Alleles IBD at Marker Loci—Parental Genotypes Known 204
 - 9.3.2 Proportion of Alleles IBD at Marker Loci—Parental Genotypes Unknown 206
 - 9.3.3 Multipoint Estimation of Proportion of Genes IBD at QTL 207
- 9.4 Maximum Likelihood Estimation 208
 - 9.4.1 Likelihood Ratio Test 209
 - 9.4.2 Approximate Critical Value for Significance Test 209
- 9.5 Single Marker Analysis and Interval Mapping of QTL 209
- 9.6 Program Implementation 210
 - 9.6.1 Choosing the Scanning Increment 210
 - 9.6.2 Implementing the Multipoint Method 212
 - 9.6.3 Implementing the Mixed Model Methodology through PROC IML 214
 - 9.6.4 Implementing the Mixed Model Methodology through PROC MIXED 216
- 9.7 Executing the Program 216
 - 9.7.1 Data Preparation 217
 - 9.7.2 Running the Main Program 218
 - 9.7.3 Application Notes 219
 - 9.7.4 Example Demo 220
- 9.8 References 222

Chapter 10 Bayesian Mapping Methodology 225

- 10.1 Introduction 225
- 10.2 Two-Point Linkage Analysis 226
 - 10.2.1 Inference Regarding Recombination Rates 227
 - 10.2.2 Example 228
 - 10.2.3 Testing for Linkage 230
- 10.3 Three-Point Analysis 231
- 10.4 Genetic Map Construction 238
- 10.5 QTL Analysis 240
 - 10.5.1 Example 242
 - 10.5.2 Results 246
- 10.6 Final Comments 249
- 10.7 References 249

Chapter 11 Gene Expression Profiling Using Mixed Models 251

- 11.1 Introduction 251
- 11.2 Theory of Mixed Model Analysis of Microarray Data 252
- 11.3 Mixed Model Analysis of Two-Color Microarray Data Using PROC MIXED 254
 - 11.3.1 Data Preparation 254
 - 11.3.2 Computation of Relative Fluorescence Intensities 255
 - 11.3.3 Computation of Gene-Specific Significance Models 257
 - 11.3.4 Explore and Interpret the Results 258
- 11.4 ANOVA-Based Analysis of Affymetrix Data Using SAS Microarray Solution 260
 - 11.4.1 Download Data 260
 - 11.4.2 Examine Data 262
 - 11.4.3 Submit the Analysis 263
 - 11.4.4 Specifications for Mixed Model Analysis 264
 - 11.4.5 Submit the Analysis 266
- 11.5 Discussion 276
- 11.6 References 277

Additional Reading 279

Index 281