

---

# CONTEMPORARY STATISTICAL MODELS

for the Plant and Soil Sciences

---

**Oliver Schabenberger**  
**Francis J. Pierce**

Technische Universität Darmstadt  
FACHBEREICH 10 — BIOLOGIE  
— Bibliothek —  
Schnittspahnstraße 10  
D-64287 Darmstadt

Inv.-Nr. 15583  
.....



**CRC PRESS**

---

Boca Raton London New York Washington, D.C.

# Contents

Preface.....	xv
About the Authors.....	xix
The CD-ROM.....	xxi
<b>1 Statistical Models</b>	<b>1</b>
1.1 Mathematical and Statistical Models.....	2
1.2 Functional Aspects of Models.....	4
1.3 The Inferential Steps — Estimation and Testing.....	10
1.4 <i>t</i> -Tests in Terms of Statistical Models.....	14
1.5 Embedding Hypotheses.....	18
1.6 Hypothesis and Significance Testing — Interpretation of the <i>p</i> -Value.....	21
1.7 Classes of Statistical Models.....	25
1.7.1 The Basic Component Equation.....	25
1.7.2 Linear and Nonlinear Models.....	26
1.7.3 Regression and Analysis of Variance Models.....	28
1.7.4 Univariate and Multivariate Models.....	29
1.7.5 Fixed, Random, and Mixed Effects Models.....	30
1.7.6 Generalized Linear Models.....	32
1.7.7 Errors in Variable Models.....	34
<b>2 Data Structures</b>	<b>35</b>
2.1 Introduction.....	36
2.2 Classification by Response Type.....	37
2.3 Classification by Study Type.....	39
2.4 Clustered Data.....	41
2.4.1 Clustering through Hierarchical Random Processes.....	42
2.4.2 Clustering through Repeated Measurements.....	43
2.5 Autocorrelated Data.....	48
2.5.1 The Autocorrelation Function.....	48
2.5.2 Consequences of Ignoring Autocorrelation.....	51
2.5.3 Autocorrelation in Designed Experiments.....	53
2.6 From Independent to Spatial Data — a Progression of Clustering.....	54

<b>3</b>	<b>Linear Algebra Tools</b>	<b>59</b>
3.1	Introduction.....	60
3.2	Matrices and Vectors.....	61
3.3	Basic Matrix Operations.....	63
3.4	Matrix Inversion — Regular and Generalized Inverse.....	67
3.5	Mean, Variance, and Covariance of Random Vectors.....	70
3.6	The Trace and Expectation of Quadratic Forms.....	72
3.7	The Multivariate Gaussian Distribution.....	74
3.8	Matrix and Vector Differentiation.....	76
3.9	Using Matrix Algebra to Specify Models.....	77
3.9.1	Linear Models.....	77
3.9.2	Nonlinear Models.....	81
3.9.3	Variance-Covariance Matrices and Clustering.....	81
<b>4</b>	<b>The Classical Linear Model: Least Squares and Alternatives</b>	<b>85</b>
4.1	Introduction.....	87
4.2	Least Squares Estimation and Partitioning of Variation.....	95
4.2.1	The Principle.....	95
4.2.2	Partitioning Variability through Sums of Squares.....	97
4.2.3	Sequential and Partial Sums of Squares and the Sum of Squares Reduction Test.....	98
4.3	Factorial Classification.....	104
4.3.1	The Means and Effects Model.....	105
4.3.2	Effect Types in Factorial Designs.....	108
4.3.3	Sum of Squares Partitioning through Contrasts.....	110
4.3.4	Effects and Contrasts in The SAS <sup>®</sup> System.....	112
4.4	Diagnosing Regression Models.....	119
4.4.1	Residual Analysis.....	119
4.4.2	Recursive and Linearly Recovered Errors.....	122
4.4.3	Case Deletion Diagnostics.....	126
4.4.4	Collinearity Diagnostics.....	130
4.4.5	Ridge Regression to Combat Collinearity.....	133
4.5	Diagnosing Classification Models.....	137
4.5.1	What Matters?.....	137
4.5.2	Diagnosing and Combating Heteroscedasticity.....	139
4.5.3	Median Polishing of Two-Way Layouts.....	144
4.6	Robust Estimation.....	152
4.6.1	$L_1$ -Estimation.....	153
4.6.2	M-Estimation.....	155
4.6.3	Robust Regression for Prediction Efficiency Data.....	158
4.6.4	M-Estimation in Classification Models.....	164
4.7	Nonparametric Regression .....	172
4.7.1	Local Averaging and Local Regression.....	174
4.7.2	Choosing the Smoothing Parameter.....	178

## Appendix A on CD-ROM

### A4.8 Mathematical Details

A4.8.1	Least Squares.....	A-3
A4.8.2	Hypothesis Testing in the Classical Linear Model.....	A-5
A4.8.3	Diagnostics in Regression Models.....	A-8
A4.8.4	Ridge Regression.....	A-11
A4.8.5	$L_1$ -Estimation.....	A-11
A4.8.6	M-Estimation.....	A-13
A4.8.7	Nonparametric Regression.....	A-16

<b>5</b>	<b>Nonlinear Models</b>	<b>183</b>
5.1	Introduction.....	185
5.2	Models as Laws or Tools.....	189
5.3	Linear Polynomials Approximate Nonlinear Models.....	193
5.4	Fitting a Nonlinear Model to Data.....	195
5.4.1	Estimating the Parameters.....	195
5.4.2	Tracking Convergence.....	201
5.4.3	Starting Values.....	204
5.4.4	Goodness-of-Fit.....	211
5.5	Hypothesis Tests and Confidence Intervals.....	213
5.5.1	Testing the Linear Hypothesis.....	213
5.5.2	Confidence and Prediction Intervals.....	221
5.6	Transformations.....	223
5.6.1	Transformation to Linearity.....	223
5.6.2	Transformation to Stabilize the Variance.....	226
5.7	Parameterization of Nonlinear Models.....	228
5.7.1	Intrinsic and Parameter-Effects Curvature.....	229
5.7.2	Reparameterization through Defining Relationships.....	234
5.8	Applications.....	236
5.8.1	Basic Nonlinear Analysis with The SAS® System — Mitscherlich's Yield Equation.....	238
5.8.2	The Sampling Distribution of Nonlinear Estimators — the Mitscherlich Equation Revisited.....	248
5.8.3	Linear-Plateau Models and Their Relatives — a Study of Corn Yields from Tennessee.....	252
5.8.4	Critical $NO_3$ Concentrations as a Function of Sampling Depth — Comparing Join-Points in Plateau Models.....	259
5.8.5	Factorial Treatment Structure with Nonlinear Response.....	266
5.8.6	Modeling Hormetic Dose Response through Switching Functions.....	273
5.8.7	Modeling a Yield-Density Relationship.....	285
5.8.8	Weighted Nonlinear Least Squares Analysis with Heteroscedastic Errors.....	293

**Appendix A on CD-ROM****A5.9 Forms of Nonlinear Models**A5.9.1 *Concave and Convex Models, Yield-Density Models*..... A-20A5.9.2 *Models with Sigmoidal Shape, Growth Models*..... A-24**A5.10 Mathematical Details**A5.10.1 *Taylor Series Involving Vectors*..... A-27A5.10.2 *Nonlinear Least Squares and the Gauss-Newton Algorithm*..... A-29A5.10.3 *Nonlinear Generalized Least Squares*..... A-31A5.10.4 *The Newton-Raphson Algorithm*..... A-32A5.10.5 *Convergence Criteria*..... A-33A5.10.6 *Hypothesis Testing, Confidence and Prediction Intervals*..... A-34**6 Generalized Linear Models 299**

6.1 Introduction..... 301

6.2 Components of a Generalized Linear Model..... 304

6.2.1 Random Component..... 305

6.2.2 Systematic Component and Link Function ..... 312

6.2.3 Generalized Linear Models in The SAS® System..... 320

6.3 Grouped and Ungrouped Data..... 328

6.4 Parameter Estimation and Inference..... 331

6.4.1 Solving the Likelihood Problem..... 331

6.4.2 Testing Hypotheses about Parameters and Their Functions..... 333

6.4.3 Deviance and Pearson's  $X^2$  Statistic..... 336

6.4.4 Testing Hypotheses through Deviance Partitioning..... 338

6.4.5 Generalized  $R^2$  Measures of Goodness-of-Fit..... 343

6.5 Modeling an Ordinal Response..... 344

6.5.1 Cumulative Link Models..... 346

6.5.2 Software Implementation and Example..... 349

6.6 Overdispersion..... 356

6.7 Applications..... 358

6.7.1 Dose-Response and LD<sub>50</sub> Estimation in a Logistic Regression Model..... 359

6.7.2 Binomial Proportions in a Randomized Block Design — the Hessian Fly Experiment..... 365

6.7.3 Gamma Regression and Yield Density Models..... 370

6.7.4 Effects of Judges' Experience on Bean Canning Quality Ratings..... 375

6.7.5 Ordinal Ratings in a Designed Experiment with Factorial Treatment Structure and Repeated Measures..... 379

6.7.6 Log-Linear Modeling of Rater Agreement..... 383

6.7.7 Modeling the Sample Variance of Scab Infection..... 393

6.7.8 A Poisson/Gamma Mixing Model for Overdispersed Poppy Counts.... 397

**Appendix A on CD-ROM****A6.8 Mathematical Details and Special Topics**A6.8.1 *Exponential Family of Distributions*..... A-36A6.8.2 *Maximum Likelihood Estimation*..... A-36A6.8.3 *Iteratively Reweighted Least Squares*..... A-39A6.8.4 *Hypothesis Testing*..... A-40

6.8.5	<i>Fieller's Theorem and the Variance of a Ratio</i> .....	A-42
6.8.6	<i>Overdispersion Mechanisms for Counts</i> .....	A-44

<b>7</b>	<b>Linear Mixed Models for Clustered Data</b> .....	<b>403</b>
7.1	Introduction.....	405
7.2	The Laird-Ware Model.....	412
7.2.1	Rationale.....	412
7.2.2	The Two-Stage Concept.....	415
7.2.3	Fixed or Random Effects.....	422
7.3	Choosing the Inference Space.....	425
7.4	Estimation and Inference.....	430
7.4.1	Maximum and Restricted Maximum Likelihood.....	432
7.4.2	Estimated Generalized Least Squares.....	437
7.4.3	Hypothesis Testing.....	438
7.5	Correlations in Mixed Models.....	446
7.5.1	Induced Correlations and the Direct Approach.....	446
7.5.2	Within-Cluster Correlation Models.....	450
7.5.3	Split-Plots, Repeated Measures, and the Huynh-Feldt Conditions.....	461
7.6	Applications.....	465
7.6.1	Two-Stage Modeling of Apple Growth over Time.....	466
7.6.2	On-Farm Experimentation with Randomly Selected Farms.....	474
7.6.3	Nested Errors through Subsampling.....	479
7.6.4	Recovery of Inter-Block Information in Incomplete Block Designs.....	488
7.6.5	A Split-Strip-Plot Experiment for Soybean Yield.....	493
7.6.6	Repeated Measures in a Completely Randomized Design.....	504
7.6.7	A Longitudinal Study of Water Usage in Horticultural Trees.....	512
7.6.8	Cumulative Growth of Muskmelons in Subsampling Design.....	520

### **Appendix A on CD-ROM**

A7.7	<i>Mathematical Details and Special Topics</i>	
A7.7.1	<i>Henderson's Mixed Model Equations</i> .....	A-52
A7.7.2	<i>Solutions to the Mixed Model Equations</i> .....	A-53
A7.7.3	<i>Likelihood Based Estimation</i> .....	A-54
A7.7.4	<i>Estimated Generalized Least Squares Estimation</i> .....	A-58
A7.7.5	<i>Hypothesis Testing</i> .....	A-60
A7.7.6	<i>The First-Order Autoregressive Model</i> .....	A-62

<b>8</b>	<b>Nonlinear Models for Clustered Data</b> .....	<b>525</b>
8.1	Introduction.....	526
8.2	Nonlinear and Generalized Linear Mixed Models.....	528
8.3	Toward an Approximate Objective Function.....	529
8.3.1	Three Linearizations.....	531
8.3.2	Linearization in Generalized Linear Mixed Models.....	534
8.3.3	Integral Approximation Methods.....	535
8.4	Applications.....	537
8.4.1	A Nonlinear Mixed Model for Cumulative Tree Bole Volume.....	539
8.4.2	Poppy Counts Revisited — a Generalized Linear Mixed Model for Overdispersed Count Data.....	545

8.4.3	Repeated Measures with an Ordinal Response.....	551
-------	---	-----

### **Appendix A on CD-ROM**

8.5	<i>Mathematical Details and Special Topics</i>	
8.5.1	<i>PA and SS Linearizations.....</i>	<i>A-64</i>
8.5.2	<i>Generalized Estimating Equations.....</i>	<i>A-65</i>
8.5.3	<i>Linearization in Generalized Linear Mixed Models.....</i>	<i>A-68</i>
8.5.4	<i>Gaussian Quadrature.....</i>	<i>A-69</i>

<b>9</b>	<b>Statistical Models for Spatial Data</b>	<b>561</b>
9.1	Changing the Mindset.....	563
9.1.1	Samples of Size One.....	563
9.1.2	Random Functions and Random Fields.....	565
9.1.3	Types of Spatial Data.....	567
9.1.4	Stationarity and Isotropy — the Built-in Replication Mechanism of Random Fields.....	572
9.2	Semivariogram Analysis and Estimation.....	577
9.2.1	Elements of the Semivariogram.....	577
9.2.2	Parametric Isotropic Semivariogram Models.....	581
9.2.3	The Degree of Spatial Continuity (Structure).....	585
9.2.4	Semivariogram Estimation and Fitting.....	587
9.3	The Spatial Model.....	599
9.4	Spatial Prediction and the Kriging Paradigm.....	603
9.4.1	Motivation of the Prediction Problem.....	603
9.4.2	The Concept of Optimal Prediction.....	607
9.4.3	Ordinary and Universal Kriging.....	609
9.4.4	Some Notes on Kriging.....	613
9.4.5	Extensions to Multiple Attributes.....	619
9.5	Spatial Regression and Classification Models.....	625
9.5.1	Random Field Linear Models.....	625
9.5.2	Some Philosophical Considerations.....	628
9.5.3	Parameter Estimation.....	629
9.6	Autoregressive Models for Lattice Data.....	632
9.6.1	The Neighborhood Structure.....	632
9.6.2	First-Order Simultaneous and Conditional Models.....	634
9.6.3	Parameter Estimation.....	637
9.6.4	Choosing the Neighborhood Structure.....	637
9.7	Analyzing Mapped Spatial Point Patterns.....	638
9.7.1	Introduction.....	638
9.7.2	Random, Aggregated, and Regular Patterns — the Notion of Complete Spatial Randomness.....	640
9.7.3	Testing the CSR Hypothesis in Mapped Point Patterns.....	642
9.7.4	Second-Order Properties of Point Patterns.....	648
9.8	Applications.....	650
9.8.1	Exploratory Tools for Spatial Data — Diagnosing Spatial Autocorrelation with Moran's I.....	651
9.8.2	Modeling the Semivariogram of Soil Carbon.....	658
9.8.3	Spatial Prediction — Kriging of Lead Concentrations.....	669

9.8.4 Spatial Random Field Models — Comparing C/N Ratios among Tillage Treatments..... 673

9.8.5 Spatial Random Field Models — Spatial Regression of Soil Carbon on Soil N..... 679

9.8.6 Spatial Generalized Linear Models — Spatial Trends in the Hessian Fly Experiment..... 684

9.8.7 Simultaneous Spatial Autoregression — Modeling Wiebe's Wheat Yield Data..... 693

9.8.7 Point Patterns — First- and Second-Order Properties of a Mapped Pattern?..... 697

**Appendix A on CD-ROM**

**A9.9 Mathematical Details and Special Topics**

*Geostatistical Data*

A9.9.1 *Estimating the Empirical Semivariogram*..... A-72

A9.9.2 *Parametric Fitting of the Semivariogram*..... A-75

A9.9.3 *Nonparametric Fitting of the Semivariogram*..... A-79

A9.9.4 *Solutions to Kriging Equations*..... A-81

A9.9.5 *Is Kriging Perfect Interpolation?*..... A-85

A9.9.6 *Block and Indicator Kriging*..... A-91

*Spatial Random Field Models*

A9.9.7 *Composite Likelihood in Spatial Generalized Linear Models*..... A-95

*Lattice Data*

A9.9.8 *Maximum Likelihood Estimation in Lattice Models*..... A-97

A9.9.9 *Are Autoregressive Models Stationary?*..... A-98

*Point Patterns*

A9.9.10 *Estimating First- and Second-Order Properties of Point Patterns*..... A-100

A9.9.11 *Point Process Models*..... A-104

A9.9.12 *Spectral Analysis of Spatial Point Patterns*..... A-109

*Supplementary Application*

A9.9.13 *Point Patterns — Spectral Analysis of Seedling Counts*..... A-112

**Bibliography**..... 703

**Author Index**..... 721

**Subject Index**..... 725