

Empirical Model-Building and Response Surfaces

GEORGE E. P. BOX

NORMAN R. DRAPER

John Wiley & Sons

New York . Chichester . Brisbane . Toronto . Singapore

Contents

1.	INTRODUCTION TO RESPONSE SURFACE METHODOLOGY	1
1.1.	Response Surface Methodology (RSM)	1
1.2.	Indeterminacy of Experimentation	4
1.3.	Iterative Nature of the Experimental Learning Process (Conjecture—Design—Experiment—Analysis)	7
1.4.	Some Classes of Problems (Which, How, Why)	10
1.5.	Need for Experimental Design	15
1.6.	Geometric Representation of Response Relationships	15
1.7.	Three Kinds of Applications	17
2.	THE USE OF GRADUATING FUNCTIONS	20
2.1.	Approximating Response Functions	20
2.2.	An Example	28
	Appendix 2A. A Theoretical Response Function	32
3.	LEAST SQUARES FOR RESPONSE SURFACE WORK	34
3.1.	The Method of Least Squares	34
3.2.	Linear Models	35
3.3.	Matrix Formulas for Least Squares	40
3.4.	Geometry of Least Squares	47
3.5.	Analysis of Variance for One Regressor	50
3.6.	Least Squares for Two Regressors	53
3.7.	Geometry of the Analysis of Variance for Two Regressors	56
3.8.	Orthogonalizing the Second Regressor. Extra Sum of Squares Principle	57
		ix

3.9.	Generalization to p Regressors	62
3.10.	Bias in Least-Squares Estimators Arising from an Inadequate Model	66
3.11.	Pure Error and Lack of Fit	70
3.12.	Confidence Intervals and Confidence Regions	74
3.13.	Robust Estimation, Maximum Likelihood, and Least Squares	79
Appendix 3A.	Iteratively Reweighted Least Squares	89
Appendix 3B.	Justification of Least Squares by the Gauss-Markov Theorem, and Robustness	90
Appendix 3C.	Matrix Theory	91
Appendix 3D.	Nonlinear Estimation	99
Appendix 3E.	Results Involving $V(\hat{y})$	102
	Exercises	103
4.	FACTORIAL DESIGNS AT TWO LEVELS	105
4.1.	The Value of Factorial Designs	105
4.2.	Two-Level Factorials	107
4.3.	A 2^6 Design Used in a Study of Dyestuffs Manufacture	115
4.4.	Diagnostic Checking of the Fitted Models, 2^6 Dyestuffs Example	119
4.5.	Response Surface Analysis of the 2^6 Design Data	123
Appendix 4A.	Yates' Method for Obtaining the Factorial Effects for a Two-Level Design	127
Appendix 4B.	Normal Plots on Probability Paper	128
Appendix 4C.	Confidence Regions for Contour Planes	134
	Exercises	135
5.	BLOCKING AND FRACTIONATING 2^k FACTORIAL DESIGNS	143
5.1.	Blocking the 2^6 Design	143
5.2.	Fractionating the 2^6 Design	148
5.3.	Resolution of a 2^{k-p} Factorial Design	154
5.4.	Construction of 2^{k-p} Designs of Resolution III and IV	154
5.5.	Designs of Resolution V and of Higher Resolution	162
5.6.	Application of Fractional Factorial Designs to Response Surface Methodology	166
5.7.	Plotting Effects from Fractional Factorials on Probability Paper	167
	Exercises	167

CONTENTS	xi
6. THE USE OF STEEPEST ASCENT TO ACHIEVE PROCESS IMPROVEMENT	182
6.1. What Steepest Ascent Does	183
6.2. An Example. Improvement of a Process Used in Drug Manufacture	184
6.3. General Comments on Steepest Ascent	188
6.4. A Confidence Region for the Direction of Steepest Ascent	190
6.5. Steepest Ascent Subject to a Constraint	194
Appendix 6A. Effects of Changes in Scales on Steepest Ascent Exercises	199 202
7. FITTING SECOND-ORDER MODELS	205
7.1. Fitting and Checking Second Degree Polynomial Graduating Functions	205
7.2. A Comprehensive Test of the Need for Transformation in the Response Variable	216
7.3. Interim Summary	218
7.4. Further Examination of Adequacy of Fit; Higher-Order Terms	219
7.5. Robust Methods in Fitting Graduating Functions	223
Appendix 7A. Fitting of First and Second Degree Models	227
Appendix 7B. Checking for Nonadditivity with Tukey's Test	230
Appendix 7C. Third-Order Biases in Estimates for Second-Order Model, Using the 3^3 Design	232
Appendix 7D. Analysis of Variance to Detect Possible Third-Order Terms	234
Appendix 7E. Orthogonalization of Effects for Three-Level Designs	236
Appendix 7F. Orthogonal Effects Associated with Individual Degrees of Freedom and Related Normal Plots	239
Appendix 7G. Yates' Algorithm for 3^k Factorial Designs Exercises	240 243
8. ADEQUACY OF ESTIMATION AND THE USE OF TRANSFORMATION	268
8.1. What Can Transformations Do?	268
8.2. A Measure of Adequacy of Function Estimation	275
8.3. Choosing the Response Metric to Validate Distribution Assumptions	280
8.4. Estimating Response Transformation from the Data	288

8.5. Transformation of the Predictor Variables	293
Appendix 8A. Proof of a Result Used in Section 8.2	300
Exercises	301
9. EXPLORATION OF MAXIMA AND RIDGE SYSTEMS WITH SECOND-ORDER RESPONSE SURFACES	304
9.1. The Need for Second-Order Surfaces	304
9.2. Example: Representation of Polymer Elasticity as a Quadratic Function of Three Operating Variables	305
9.3. Examination of the Fitted Surface	313
9.4. Investigation of Adequacy of Fit: Isolation of Residual Degrees of Freedom for the Composite Design	316
10. OCCURRENCE AND ELUCIDATION OF RIDGE SYSTEMS, I	323
10.1. Introduction	323
10.2. Factor Dependence	326
10.3. Elucidation of Stationary Regions, Maxima, and Minima by Canonical Analysis	332
Appendix 10A. A Simple Explanation of Canonical Analysis	342
11. OCCURRENCE AND ELUCIDATION OF RIDGE SYSTEMS, II	346
11.1. Ridge Systems Approximated by Limiting Cases	346
11.2. Canonical Analysis in the A Form to Characterize Ridge Phenomena	349
11.3. Example 11.3: A Consecutive Chemical System Yielding a Near Stationary Planar Ridge Maximum	355
11.4. Example 11.4: A Small Reactor Study Yielding a Rising Ridge Surface	360
11.5. Example 11.5: A Stationary Ridge in Five Variables	368
11.6. The Economic Importance of the Dimensionality of Maxima and Minima	372
11.7. A Method for Obtaining a Desirable Combination of Several Responses	373
Appendix 11A. Calculations for the Analysis of Variance, Table 11.5	375
Appendix 11B. "Ridge Analysis" of Response Surfaces	375
Exercises	381

CONTENTS	xiii
12. LINKS BETWEEN EMPIRICAL AND THEORETICAL MODELS	403
12.1. Introduction	403
12.2. A Mechanistic Model	403
12.3. Insight Provided by the Fitted Mechanistic Model	409
12.4. Ridge Properties and the Mechanism	413
12.5. Relationships of Empirical Coefficients to Mechanistic Parameters: Their Use in the Identification of Possible Theoretical Models	417
Appendix 12A. A More Detailed Analysis of the Fitted Mechanistic Model	421
13. DESIGN ASPECTS OF VARIANCE, BIAS, AND LACK OF FIT	423
13.1. The Use of Approximating Functions	423
13.2. The Competing Effects of Bias and Variance	425
13.3. Integrated Mean Squared Error	428
13.4. Region of Interest and Region of Operability in k Dimensions	430
13.5. An Extension of the Idea of a Region of Interest: The Weight Function	433
13.6. Designs that Minimize Squared Bias	437
13.7. Ensuring Good Detectability of Lack of Fit	442
13.8. Checks Based on the Possibility of Obtaining a Simple Model by Transformation of the Predictors	450
Appendix 13A. Derivation of Eq. (13.5.8)	462
Appendix 13B. Necessary and Sufficient Design Conditions to Minimize Bias	463
Appendix 13C. Minimum Bias Estimation	464
Appendix 13D. A Shrinkage Estimation Procedure	466
Appendix 13E. Conditions for Efficacy of Transformations on the Predictor Variables	467
Appendix 13F. The Third-Order Columns of the X Matrix for a Composite Design	472
Exercises	474
14. VARIANCE-OPTIMAL DESIGNS	477
14.1. Introduction	477
14.2. Orthogonal Designs	478
14.3. The Variance Function	481
14.4. The Alphabetic Optimality Approach	489
14.5. Alphabetic Optimality in the Response Surface Context	495

15. PRACTICAL CHOICE OF A RESPONSE SURFACE DESIGN	502
15.1. Introduction	502
15.2. First-Order Designs	506
15.3. Second-Order Composite Designs	508
15.4. Second-Order Designs Requiring Only Three Levels	515
15.5. Designs Requiring Only a Small Number of Runs	520
Exercises	525
ANSWERS TO EXERCISES	529
TABLES	605
A. Tail Area of the Unit Normal Distribution	606
B. Probability Points of the t Distribution	607
C. Probability Points of the χ^2 Distribution	608
D. Percentage Points of the F Distribution	610
BIBLIOGRAPHY	615
INDEX OF AUTHORS ASSOCIATED WITH EXERCISES	657
INDEX OF AUTHORS OF ARTICLES MENTIONED IN THE TEXT	659
GENERAL INDEX	663