Pharmaceutical
Experimental Design
and Interpretation

N. ANTHONY ARMSTRONG, B. Pharm., Ph.D., F.R.Pharm.S., MCPP.

Welsh School of Pharmacy, University of Wales, Cardiff, UK
Contents

1 Introduction to experimental design
   1.1 The experimental process 1
   1.2 Computers and experimental design 2
   1.3 Overview of experimental design and interpretation 4

2 Comparison of mean values
   2.1 Comparison of two means when the variance of the whole population is known 9
   2.2 Comparison of two means when the variance of the whole population is not known 11
   2.3 Comparison of means among more than two groups of data 13
   2.4 Analysis of variance 14
   2.5 Least significant difference 16
   2.6 Two-way analysis of variance 17

3 Non-parametric treatments
   3.1 Non-parametric tests for paired data 21
      3.1.1 The sign test 21
      3.1.2 The Wilcoxon signed rank test 24
   3.2 Non-parametric tests for unpaired data 25
      3.2.1 The Wilcoxon two-sample test 25

4 Correlation and regression
   4.1 Introduction 29
   4.2 Correlation 29
      4.2.1 Linear correlation 30
         4.2.1.1 Constitutional properties 30
         4.2.1.2 Resultant properties 30
   4.3 Linear regression 30
      4.3.1 The number of pairs of variables (n) 32
      4.3.2 The correlation coefficient (r) 33
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.3 The standard error of the estimate (s)</td>
<td>36</td>
</tr>
<tr>
<td>4.3.4 The standard error of the coefficient</td>
<td>37</td>
</tr>
<tr>
<td>4.3.5 The $F$ value or variance ratio</td>
<td>37</td>
</tr>
<tr>
<td>4.4 Inverse regression analysis</td>
<td>37</td>
</tr>
<tr>
<td>4.5 Multiple regression analysis</td>
<td>38</td>
</tr>
<tr>
<td>4.5.1 Correlation coefficients</td>
<td>40</td>
</tr>
<tr>
<td>4.5.2 Standard error of the estimate</td>
<td>42</td>
</tr>
<tr>
<td>4.5.3 Standard errors of the coefficients and the intercept</td>
<td>42</td>
</tr>
<tr>
<td>4.5.4 $F$ value</td>
<td>42</td>
</tr>
<tr>
<td>4.5.5 Interaction between independent variables</td>
<td>43</td>
</tr>
<tr>
<td>4.6 Stepwise regression</td>
<td>43</td>
</tr>
<tr>
<td>4.7 Categorical data</td>
<td>44</td>
</tr>
<tr>
<td>4.8 Curve fitting of non-linear relationships</td>
<td>44</td>
</tr>
<tr>
<td>4.8.1 The power series</td>
<td>44</td>
</tr>
<tr>
<td>4.8.1.1 Quadratic relationships</td>
<td>45</td>
</tr>
<tr>
<td>4.8.1.2 Cubic equations</td>
<td>50</td>
</tr>
<tr>
<td>4.8.2 Curve fitting with models</td>
<td>50</td>
</tr>
<tr>
<td>4.8.3 Curve fitting without models</td>
<td>52</td>
</tr>
<tr>
<td>4.8.3.1 Exponential plots</td>
<td>52</td>
</tr>
<tr>
<td>4.8.3.2 Geometric plots</td>
<td>53</td>
</tr>
<tr>
<td>4.8.3.3 Hyperbolic plots</td>
<td>53</td>
</tr>
<tr>
<td>4.8.3.4 Rectangular hyperbolic plots</td>
<td>54</td>
</tr>
<tr>
<td>4.8.4 Extrapolation</td>
<td>55</td>
</tr>
<tr>
<td>4.9 Free–Wilson analysis</td>
<td>57</td>
</tr>
<tr>
<td>5 Multivariate methods</td>
<td>61</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>61</td>
</tr>
<tr>
<td>5.2 Distance matrix</td>
<td>61</td>
</tr>
<tr>
<td>5.3 Covariance matrix</td>
<td>63</td>
</tr>
<tr>
<td>5.4 Correlation matrix</td>
<td>66</td>
</tr>
<tr>
<td>5.5 Eigenvalues and eigenvectors</td>
<td>68</td>
</tr>
<tr>
<td>6 Cluster and discrimination analysis</td>
<td>73</td>
</tr>
<tr>
<td>6.1 Cluster analysis</td>
<td>73</td>
</tr>
<tr>
<td>6.1.1 Cartesian plots</td>
<td>73</td>
</tr>
<tr>
<td>6.1.2 Andrews' plots</td>
<td>76</td>
</tr>
<tr>
<td>6.1.3 Dendrograms</td>
<td>78</td>
</tr>
<tr>
<td>6.1.3.1 Hierarchic or agglomerative methods</td>
<td>78</td>
</tr>
<tr>
<td>6.1.3.2 Partitioning methods</td>
<td>80</td>
</tr>
<tr>
<td>6.2 Discrimination</td>
<td>83</td>
</tr>
<tr>
<td>7 Principal components and factor analysis</td>
<td>89</td>
</tr>
<tr>
<td>7.1 Principal components analysis</td>
<td>89</td>
</tr>
<tr>
<td>7.2 Factor analysis</td>
<td>92</td>
</tr>
<tr>
<td>7.3 Rotation</td>
<td>96</td>
</tr>
<tr>
<td>8 Sequential analysis</td>
<td>105</td>
</tr>
<tr>
<td>8.1 Introduction</td>
<td>105</td>
</tr>
<tr>
<td>8.2 Wald plots</td>
<td>107</td>
</tr>
</tbody>
</table>
Contents

8.2.1 The sign test 107
8.2.2 The sequential procedure 107
8.2.3 Construction of barrier lines 108

8.3 Bross plots 112
8.3.1 Construction of barrier lines using the binomial theorem 112
8.3.2 Confidence levels 115
8.3.3 Prior distribution 117

8.4 Triangular plots 120
8.4.1 Calculation of barriers for triangular plots 122

8.5 Truncation procedures 123
8.5.1 Truncation using a vertical barrier 124
8.5.2 Truncation using angled stopping lines 124
8.5.3 Changing the confidence limits 125
8.5.4 Truncation procedure for triangular plots 127

9 Factorial design of experiments 131
9.1 Two-factor, two-level experimental designs 132
9.1.1 Notation in factorially designed experiments 132
9.1.2 Factorial designs with interaction between factors 134
9.2 Factorial designs with three factors 137
9.3 Factorial designs and ANOVA 140
9.3.1 Yates' treatment 140
9.3.2 Linear regression 143
9.4 Factorial designs with replication 144
9.5 Factorial designs with three levels 146
9.6 Three-factor, three-level factorial designs 151
9.7 Blocks and fractional designs 155
9.7.1 Blocked designs 155
9.7.2 Fractional factorial designs 158
9.7.3 Plackett–Burman designs 160
9.7.4 Central composite and other designs 161
9.8 General comments on factorial design 162

10 Model-dependent optimization and response surface methodology 169
10.1 Optimization 169
10.2 Model-dependent optimization 170
10.2.1 Validation of the design and the regression equations 178
10.3 Optimization when interaction occurs between the independent variables 178
10.3.1 Use of coded data 180
10.4 Second-order relationships between independent and dependent variables 181
10.5 Optimization with three or more independent variables 183
10.6 Optimization using the Pareto-optimality technique 188

11 Model-independent optimization 193
11.1 Optimization by simplex search 193
11.2 Comparison of model-independent and model-dependent methods 199
Contents

12 Experimental designs for mixtures 205
   12.1 Three component systems 206
   12.2 Mixtures with more than three components 210
   12.3 Optimization in experiments with mixtures 210
   12.4 Model-dependent methods 211
      12.4.1 Linear relationships between composition and response 211
      12.4.2 Higher order relationships between composition and response 212
      12.4.3 Derivation of contour plots 213
   12.5 Pareto-optimality and mixtures 219
   12.6 Process variables in mixture experiments 220

A1 Statistical tables 225
   A1.1 Cumulative normal distribution (Gaussian distribution) 225
   A1.2 Student's $t$ distribution 225
   A1.3 Analysis of variance 226

A2 Computer programs in BASIC and MINITAB commands 229
   A2.1 Calculation of mean, standard deviation etc. 229
      A2.1.1 Insertion of data and instructions 231
      A2.1.2 Calculation of mean etc. 232
      A2.1.3 Standardization of data 232
   A2.2 Linear regression
      A2.2.1 Insertion of data and instructions 233
      A2.2.2 Calculation of the regression equation etc. 236
   A2.3 Parabolic curve fit
      A2.3.1 Insertion of data and instructions 237
      A2.3.2 Calculation of the regression equation etc. 240
   A2.4 Three-variable regression
      A2.4.1 Insertion of data and instructions 241
      A2.4.2 Calculation of the regression equation etc. 245
   A2.5 The determinant of a (3 x 3) matrix 246
   A2.6 The determinant of a (4 x 4) matrix 248
   A2.7 Determination of matrix parameters using MINITAB 249
      A2.7.1 Insertion of data and instructions 250
      A2.7.2 Standardization of data 250
      A2.7.3 Calculation of covariance matrix 251
      A2.7.4 Calculation of correlation matrix 251
      A2.7.5 Calculation of eigenvalues and eigenvectors 251
   A2.8 Three-factor, two-level factorial design
      A2.8.1 Insertion of data and instructions 252
      A2.8.2 Analysis of variance 255

A3 Sequential analysis grids 257
   A3.1 A Wald grid for a probability level of $2P = 0.05$ 257
   A3.2 A Wald grid for a probability level of $2P = 0.10$ 258
   A3.3 A Bross grid for a probability level of $2P = 0.01$ 258
   A3.4 A Bross grid for a probability level of $2P = 0.10$ 258
Contents

A4 Matrices
A4.1 Introduction 261
A4.2 Addition and subtraction 263
A4.3 Multiplication 264
   A4.3.1 Multiplying a matrix by a constant 264
   A4.3.2 Multiplying a matrix by a column vector 264
   A4.3.3 Multiplication of one matrix by another 264
   A4.3.4 Multiplication by a unit matrix 265
   A4.3.5 Multiplication by a null matrix 266
A4.4 Determinants 266