Contents

1 Introduction to Measurement and Analysis ......................... 1
  1.1 Key Questions Regarding Health Systems Engineering .............. 1
    1.1.1 What’s Wrong with Healthcare? .............................. 1
    1.1.2 Why Should “I” Fix It? ...................................... 3
    1.1.3 What Is Health Systems Engineering? .......................... 3
    1.1.4 What Is “Analytics”? .......................................... 4
    1.1.5 Is This an Individual Activity or an Organizational Activity? ............................. 5
  1.2 The Organization of This Book .................................... 5
  1.3 Review of Introductory Concepts .................................. 8
References .............................................................. 10

Part I Managing Data

2 Data and Types of Data .................................................. 13
  2.1 What Are Data ...................................................... 13
  2.2 How Are Data Structured .......................................... 15
  2.3 Visualizing Data ................................................... 16
References .............................................................. 17

3 Software for Analytics .................................................. 19
  3.1 Initial Overview of Software ........................................ 19
    3.1.1 Excel ........................................................... 20
    3.1.2 Menu-Driven Statistical Software ................................ 21
    3.1.3 Advanced Analytical Software .................................. 21
    3.1.4 Middleware to Make R Usable .................................. 22
  3.2 Excel (and Concepts Essential to Using Excel) ..................... 22
    3.2.1 Excel as “HOME” ................................................ 23
    3.2.2 Some Useful Cell Functions in Excel ............................ 23
3.3 Menu Methods in Excel
   3.3.1 HOME ............................ 26
   3.3.2 INSERT ........................... 28
   3.3.3 DATA ............................. 29
   3.3.4 Add-Ins .......................... 32
   3.3.5 Excel Summary ..................... 33
3.4 Returning to Data Analysis in Excel: The “Workhorse” ............ 33
   3.4.1 Descriptive Statistics ............. 34
3.5 Minitab .............................. 46
3.6 R, RExcel, and RCommander .................. 56
3.7 Plotting ............................... 59
References .................................. 62

4 Measurement and Uncertainty .............................. 63
  4.1 The Stochastic Nature of Data ..................... 63
     4.1.1 The Distribution Functions of Data .... 64
     4.1.2 Understanding Uncertainty Requires
          Understanding DIKW ..................... 65
  4.2 Databases, Datasets, and Dataframes .................. 66
  4.3 Missing Values ........................... 67
  4.4 Probability, Likelihood, and Odds .................. 68
  4.5 Logit and Logistic .......................... 68
  4.6 Induction Versus Deduction ....................... 69
  4.7 Bias .................................... 70
  4.8 Residuals ................................ 70
  4.9 Central Tendency ................................ 72
  4.10 Variability ................................ 72
  4.11 Centering and Standardizing ..................... 73
  4.12 Homogeneity of Variances ....................... 74
  4.13 Quantile (QQ) and Other Useful Plots ............ 74
  4.14 Likert Scales and Likert Plots .................. 75
  4.15 A Summary for Preparing to Analyze a Dataset .......... 76
Reference ..................................... 77

Part II Applying a Scientific Approach to Data Analysis

5 Mathematical and Statistical Concepts in Data Analysis ............. 81
  5.1 A Brief Introduction to Mathematics .................. 81
  5.2 Vectors and Matrices ........................... 82
  5.3 Arrays and Data Frames ........................... 83
  5.4 Summary of Vector and Matrix Rules ................ 83
  5.5 The Curse of Dimensionality and the Bias-Variance Trade-Off ... 84
  5.6 Factors That Degrade Models ....................... 85
     5.6.1 Cause and Effect ...................... 85
     5.6.2 Collinearity .......................... 86
     5.6.3 Proximity, Distances, and Similarities .... 87
Part III  Techniques in Analysis

6 Analysis by Modeling Data  .................................................. 93
   6.1 Preparing to Build Models  ....................................... 93
       6.1.1 The Model in Light of the Hypothesis ............. 94
       6.1.2 Planning to Validate a Model ....................... 94
       6.1.3 Validation Methods When Partitioning Isn’t Feasible . 95
       6.1.4 Planning to Combine Models: The Ensemble Approach ... 97
   6.2 Building Models: Supervised and Unsupervised
       Machine Learning ................................................. 98
       6.2.1 Types of Models ........................................ 98
       6.2.2 The Problem of Unbalanced Datasets .............. 98
   References ......................................................... 100

7 Principles of Supervised Learning ......................................... 101
   7.1 An Introduction to Supervised Machine Learning:
       Regression and Classification ............................... 101
   7.2 Setting the Stage for Regression: An Example ........... 103
       7.2.1 Principles of Regression .............................. 104
       7.2.2 Assumptions of Regression ......................... 104
       7.2.3 Limitations of a Regression Model ............... 105
       7.2.4 Alternative Approaches to Regression .......... 106
   7.3 Supervised Learning with a Categorical Outcome:
       Classification ................................................. 114
       7.3.1 Setting the Stage ....................................... 114
       7.3.2 Hierarchical Classification ......................... 114
       7.3.3 Models Based on Bayesian (Conditional) Probability . 116
       7.3.4 Models Based on a Measure of Proximity
               (Distance or Similarity) ............................ 116
       7.3.5 Model-Based Classification Methods ........... 117
       7.3.6 Other Classification Methods ..................... 119
   References ......................................................... 119

8 Unsupervised Machine Learning: Datasets Without Outcomes .......... 121
   8.1 Unsupervised Learning: Models Without Outcomes ....... 121
   8.2 Unsupervised Learning with Nominal Data:
       Association Analysis ......................................... 122
       8.2.1 Setting the Stage ..................................... 122
       8.2.2 Terminology in Association Analysis ............. 122
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3</td>
<td>Unsupervised Learning with Quantitative Data: Clustering</td>
<td>125</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Setting the Stage</td>
<td>125</td>
</tr>
<tr>
<td>8.3.2</td>
<td>Approaches to Clustering</td>
<td>126</td>
</tr>
<tr>
<td>8.3.3</td>
<td>Validating Clusters</td>
<td>127</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>129</td>
</tr>
<tr>
<td>9</td>
<td>Survival Analysis</td>
<td>131</td>
</tr>
<tr>
<td>9.1</td>
<td>Concepts of Survival and Hazard</td>
<td>131</td>
</tr>
<tr>
<td>9.2</td>
<td>Censoring of Data</td>
<td>133</td>
</tr>
<tr>
<td>9.3</td>
<td>Biases Inherent in Survival Analysis: Lead-Time, Length, and Overdiagnosis Bias</td>
<td>133</td>
</tr>
<tr>
<td>9.4</td>
<td>Analytical Methods Based on Survival</td>
<td>134</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Kaplan-Meier Survival Analysis</td>
<td>134</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Comparing Survival Curves Using Log-Rank</td>
<td>136</td>
</tr>
<tr>
<td>9.4.3</td>
<td>Evaluating Possible Contributors to Survival</td>
<td>136</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>137</td>
</tr>
<tr>
<td>10</td>
<td>Interpreting Outcomes: Cause and Effect and ( p )-Values</td>
<td>139</td>
</tr>
<tr>
<td>10.1</td>
<td>The Problem with Assuming Cause and Effect</td>
<td>139</td>
</tr>
<tr>
<td>10.2</td>
<td>Simpson’s Paradox and Cause-Effect</td>
<td>140</td>
</tr>
<tr>
<td>10.3</td>
<td>( p )-Values and Types of Error</td>
<td>142</td>
</tr>
<tr>
<td>10.4</td>
<td>Type I and Type II Errors in Hypothesis Testing</td>
<td>142</td>
</tr>
<tr>
<td>10.5</td>
<td>Reinterpreting Sensitivity and Specificity</td>
<td>143</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>144</td>
</tr>
<tr>
<td>11</td>
<td>Useful Tools</td>
<td>145</td>
</tr>
<tr>
<td>11.1</td>
<td>Tools for Understanding and Applying Analytical Methods</td>
<td>145</td>
</tr>
<tr>
<td>11.2</td>
<td>Outlier/Anomaly Detection</td>
<td>145</td>
</tr>
<tr>
<td>11.3</td>
<td>Statistical Process Control</td>
<td>147</td>
</tr>
<tr>
<td>11.4</td>
<td>Failure Mode and Effect Analysis</td>
<td>149</td>
</tr>
<tr>
<td>11.4.1</td>
<td>Summary</td>
<td>151</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>152</td>
</tr>
<tr>
<td>Appendix</td>
<td>Using R</td>
<td>153</td>
</tr>
<tr>
<td>Appendix A</td>
<td>A Few Basic Rules</td>
<td>153</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Interpreting the Documentation</td>
<td>154</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Linear Regression</td>
<td>155</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Selecting and Installing a Package</td>
<td>157</td>
</tr>
<tr>
<td>Appendix E</td>
<td>Another Summary for Actually Using R</td>
<td>158</td>
</tr>
<tr>
<td>Glossary</td>
<td></td>
<td>161</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
<td>175</td>
</tr>
<tr>
<td>Index</td>
<td></td>
<td>177</td>
</tr>
</tbody>
</table>
Measurement and Analysis in Transforming Healthcare Delivery
Volume 1: Quantitative Approaches in Health Systems Engineering
Fabri, P.J.
2016, XIV, 181 p. 32 illus., 31 illus. in color., Hardcover
ISBN: 978-3-319-40810-1