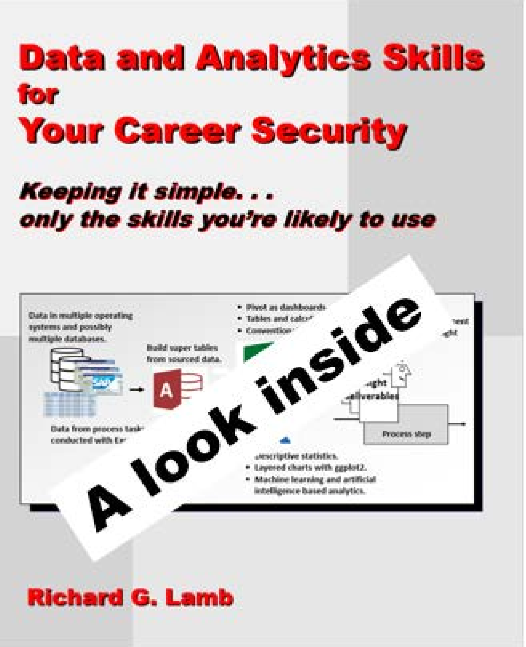

Data and Analytics Skills for Your Career Security

Keeping it simple. . .
only the skills you're likely to use

Data in multiple operating systems and possibly multiple databases.
Build super tables from sourced data.
Data from process task conducted with Ex
• Pivot as dashboards
• Tables and calc
• Convention
ight
deliverables
Process step
descriptive statistics.
• Layered charts with ggplot2.
• Machine learning and artificial intelligence based analytics.
A look inside

Richard G. Lamb

Excerpt:
Table of Contents, Preface and Index

Additional "Look Inside" at https://analytics4strategy.com/book-look-inside
Book is available from Amazon
Paperback, 508 pages, 300 figures and outputs, 7.5 x 1.15 x 9.25 inches, 2.37 pounds

Contents

Preface xiii

CHAPTER 1 Operational Availability is the Purpose 1

1.1. Attributes of Availability Performance..2

1.1.1. Availability as Probability..2

1.1.2. Constituent Probabilities 2

1.2. Availability as Subtypes 4

1.2.1. Maintenance Tasks and Activities..4

1.2.2. Subtypes of Availability..6

1.2.3. Restating the Equation..7

1.2.4. Interrelationship of Subtypes..9

1.3. Top factors of Operational Availability..9

1.4. Cost-Effectiveness in the Definition..11

1.5. Data-Driven as the Means..13

Bibliography 14

Chapter 2 Data, Analytics and Software to be Data Driven..15

2.1. Big Picture 15

2.1.1. What It Looks Like..15

2.2.2. Critical Mass and Grass Root..16

2.2.3. Essential Definitions..19

2.2. What and Why of Access and "R"..22

2.2.1. MS Access for Super Tables 22

2.2.2. "R" for the Analytic Core..26

2.2.3. Layered Charting..29

2.3. Insight Deliverables 31

2.3.1. System Reported and Recountive Insight 31

2.3.2. Know-Thy-Data Insight 32

2.3.3. Modeled Insight..33

Bibliography 41

Chapter 3 Super Tables from Operational Data 43

3.1. Perspective 44

3.2. Scenarios for Demonstration..45

3.3. Case 1: Foundation Super Table..47

3.3.1. Identify, Extract and Import or Link 47

3.3.2. Translation Tables 49

3.3.3. Overview of the Query Process....50
3.3.4. Joins and Their Ramifications....53
3.3.5. Coding the Design Grid....56
3.3.6. Generate the Super Table....61
3.4. Case 2: Seek Outlier Work Orders....63
3.4.1. Measurement for Outliers....63
3.4.2. Activating Aggregation....65
3.4.3. Planning the Aggregation....66
3.4.4. Arriving at the Z-Score for Hours....67
3.4.5. Student’s T-Score as Alternative....70
3.4.6. Expression and Where Aggregations....71
3.5. Case 3: Seek Outliers with Lead Craft....72
3.5.1. Planning the Lead Craft Aggregation....72
3.5.2. Arriving at Z-Score with Craft Lead....72
3.6. Comparison of Findings....78
3.7. SQL in the Background....79
3.8. Administration of Tables....80
Bibliography....81
Chapter 4 The “R” Software in Action....83
4.1. Approach to Reach Critical Mass....83
4.2. “R” Installation and Session....84
4.3. Basics of “R” Demonstrated....87
4.3.1. Demonstration Case....88
4.3.2. Import Data to the Session....89
4.3.3. Know-Thy-Data Analysis....90
4.3.4. Correlation Analysis....95
4.3.5. Partial Correlation Analysis....100
4.3.6. Correlation Compared to Regression....103
Bibliography....108
Chapter 5 Layered Charting to Know Thy Data....109
5.1. Inspect for Missing Data....110
5.1.1. From Super Table to R....111
5.1.2. Find the Missing Data....113
5.2. Visual and Statistical Inspection....117
5.2.1. Load and Survey the Data....119
5.2.2. Test for Normal Distribution....122
5.2.3. Inspect Correlation Between Variables....131
5.2.4. Inspect Centrality and Spread....139

5.2.5. Inspect Categorical Variables 148
5.2.6. Inspect Variables Over Time 151
5.3. Save and Disseminate 157
Bibliography 158
Chapter 6 Unearth and Rectify Bad Data 159
6.1. Bad Data and Rectification 159
6.2. Strategies for Numeric Data 161
6.2.1. Spotting Outliers and Influencers 161
6.2.2. Models for Numeric Imputation 163
6.3. Strategies for Categorical Data 164
6.3.1. Finding and Rectifying Misclassifications 164
6.3.2. Models to Rectify Misclassifications 165
6.4. Subsets and Multilevel Models 166
6.5. The Models in Other Purposes 168
Bibliography 169
Chapter 7 Relate Operational Variables to Outcomes 171
7.1. The General Linear Model 172
7.1.1. Systematic Component 172
7.2.2. Linking Function and Solution Equation 174
7.2. Process of Machine Learning 177
7.2.1. Load, Inspect and Prepare Data 179
7.2.2. Dummy Variables 182
7.2.3. Select the Variables 185
7.2.4. Transforming Variables for Linearity 193
7.2.5. Test the Model for Generalization 200
7.3. Extreme Cases Cleansed by AI 207
7.3.1. Flag Cases that Change Parameters 208
7.3.2. Flag Outliers Cases to the Mainstream 211
7.3.3. Create Datasets for Imputation 213
7.4.4. AI to Impute Cleansed Data 215
7.4. Individual Predictions and Intervals 217
Bibliography 219
Chapter 8 Achieve Entirety of Data 221
8.1. Layered Structure 222
8.2. Current at Capture 225
8.3. Access Upon Demand 230
Bibliography 236

Chapter 9 Workload-Based Budget and Variance 237
9.1. Framework for Budget and Variance 238
9.1.1. Mission, Structure and Derivation 238
9.1.2. The Two Dimensions 240
9.1.3. Budget 242
9.1.4. Variance 243
9.2. From History to Budget to Variance 245
9.2.1. The Case 246
9.2.2. From History to Budget 249
9.2.3. Table of Actuals 253
9.2.4. Variance table 255
9.2.5. Variance Report and Analysis 258
9.3. Variance as Systemic or by Outliers 262
9.3.1. The Case 263
9.3.2. Year to Date Aggregation 263
9.3.3. Flag Outliers for Investigation 266
9.3.4. Pivot Report of Outlier Orders 268
9.3.5. Systemic Variance After Outliers 269
9.4. Maintenance Craft Capacity 270
9.4.1. The Case 271
9.4.2. Capacity Report and Analyses 272
9.4.3. Tables to Queries to Capacity Report 274
Bibliography 278
Chapter 10 Through the Lens of Time Series 279
10.1. Working with Time Series 281
10.1.1. Time Series as an R Object 281
10.1.2. Transforming Data to a TS object 284
10.1.3. Windows to Time Series 285
10.1.4. Intersection of Series 287
10.2. Analytics with Time Series 289
10.2.1. Notation, Models and Smoothing 290
10.2.2. Decomposition of the Series 291
10.2.3. Autocorrelation and Correlograms 295
10.2.4. Lead-Lag Variables 299
10.2.5. Seven-Day Cycles 302
10.3. Forecasting into the Future 305
10.3.1. Holt-Winters 305
10.3.2. ARIMA 311

10.3.3. Test as Deterministic or Stochastic 314
Bibliography 318
Chapter 11 Budget-Based Schedule and Craft Capacity 319
11.1. Scheduling Framed Top-Down....................319
11.2. Budget-Based Schedule Cycle....................321
11.2.1. Establish Baseline Schedule and Craft Capacity....................321
11.2.2. Conduct Weekly Scheduling and Execution....................324
11.2.3. Datasets, Measures and Reports Stage....................327
11.3. Tables to the Schedule Cycle....................329
11.3.1. System of Tables to Build and Maintain....................329
11.3.2. Weekly Schedule as Table 330
11.3.3. Day Schedule as Table 331
11.3.4. Craft Hours Table....................333
11.3.5. Craft Classification Translation Table 334
11.3.6. Super Table to Measurement....................336
11.4. Measures and Methods 341
11.4.1. Confirm Sustainment....................341
11.4.2. Planned and Scheduled Execution 343
11.4.3. Job Plan Accuracy and Variance....................346
11.4.4. Measures as Time Series 349
Bibliography 351
Chapter 12 Elapsed Time Through Stages 353
12.1. The Insight We Need 354
12.1.2. Retention in the Pipeline 355
12.1.2. Exit Event from the Pipeline 357
12.2. How it Works....................358
12.2.1. Life Data, Event and Window....................358
12.2.2. Construct of Retention and Event 360
12.2.3. Equations to Retention-Event....................362
12.3. Retention and Event Models....................364
12.3.1. Data to the Analytic....................364
12.3.2. Empirical Modeling....................366
12.3.3. Parametric Modeling....................368
12.3.4. Multiple Predictor Variables 370
12.4. Weibull Plot 373
Bibliography 375

Chapter 13 Prove There is a Difference 377
13.1. Effects Distinguished as Models..378
13.1.1. Single Factor, Two Conditions..379
13.1.2. Single Factor, Three or More Conditions.......................................379
13.1.3. Covariant Variable in ANOVA..381
13.1.4. Two or More Factorial Variables 382
13.1.5. Measures are Repeated...383
13.1.6. Mixed Models...384
13.1.7. Robust Models..384
13.2. Test by Post Hoc and Contrasts 385
13.2.1. Post Hoc Adjusted Confidence Limits 385
13.2.2. Contrasts as Effect...386
13.3. Effects Analysis Cases..389
13.3.1. Two-Mean T-Test..389
13.3.2. One-Way ANOVA 394
13.3.3. Analysis of Covariance, ANCOVA 404
13.3.4. Factorial ANOVA 413
13.3.5. One-Way Repeated Measures 422
13.3.6. Repeated-Measures Factorial 427
13.3.7. Mixed Models...435
Bibliography 436
Chapter 14 Recover Lost Classifications 437
14.1. Generalized Procedure..438
14.2. Logistic Regression as Classifier..440
14.2.1. Concept of Logistic Regression 441
14.2.2. Case, Dataset and Preparation 443
14.2.3. Build and Interpret Model 445
14.2.4. Classification as Probability...449
14.2.5. Analysis of Residuals 451
14.2.6. Classify Cases...451
14.2.7. Test for Multicollinearity and Linearity..454
14.2.8: Logistic Regression as Predictive Insight 457
14.3. Naïve Bayes Probability as Classifier..459
14.3.1. Concept of Naïve Bayes 459
14.3.2. Case, Dataset and Exploration..460
14.3.3. Cleansing and Standardizing Text..462
14.3.4. Visualize Text as Word Clouds..464
14.3.5. Create the Document Term Matrix 466

14.3.6. Create Training and Test Datasets..467
14.3.7. Create Indicators of Frequent Words 468
14.3.8. Train and Evaluate Model..469
14.3.9. Classify Cases..472
Bibliography 473
Index 475

Preface

Does the steady drumbeat for data and analytics in your newsfeed make you uneasy? Are you worried that data and analytics will be the new Excel for which skills went from optional to mandatary for everyone's career? Are you searching for a way in but have not yet found a door other than some prohibitively huge and long undertaking to reach critical mass?

In other words, should you be worried for your career health? Yes, if your role engages with operating systems and Excel. And yes again, if your role in a managerial position entails decision-making upon the interpretation of standard reports from the operating system or Excel deliverables from other role holders.

You should be worried. As a few of these role holders progressively bring data and analytics skills into their roles, they will set a new standard that is immediate, visible and significant to the enterprise. Without the same skills you will not be able to meet the standard. Better yet, the best assurance of career security is to be one of those who sets the standard.

Maybe this is not new bad news to you. You see it coming. Your problem is that you have not yet stumbled across a way into the skills, short of awful. However, there is good news. Because I decided to take the long and awful path, this book is your serendipitous good luck. My bad luck was that a book such as this had never been written.

Why the book is your good luck is evident by my experience up to writing it. As an operational role holder, rather than data scientist, I had to learn much of everything there is to know about data and analytics before I could reduce the explanation of data and analytics to what you, in your operational role, need to know.

It took me six years to learn much of what there is to know. It then took me another more than a year to reduce what I had to learn to what you need to know to vaccinate your career against obsolescence and decline.

Philosophy and Approach

The book is written to explain and demonstrate data and analytics as methods that are relevant to all enterprise operations rather than the universe of all human endeavors. The strategy works because all operations must deal with the same issues and sub-operations. Every operation has, in some form, requirements for setting direction, aggregate planning, budget and variance control, planning, scheduling and executing the core task, staffing and optimizing the retention time of the core task in the stages through the operation.

It follows that all operations will thrive on the same data and analytics skills. Regardless of the type of operation, it is obviously impactful that its role holders be able to extract, explore, cleanse and mold to purpose the data captured in their operating systems. Regardless of the type of operation, it is obviously impactful that role holders be able to augment their experience and judgement with the insight that data and analytics make possible. Furthermore, the best augmenting insights are not and never will be available from the operating systems their roles depend upon.

Accordingly, a full explanation by demonstration of what you need to know, must necessarily be in the context of an operational domain. Therefore, as a context to all explanations by demonstration, the book has selected a domain in the management of manufacturing plants as a production asset. The type of operation is known as asset management. For those in the field of asset management, the book is a doubly serendipitous good fortune.

Chapter 1, Operational Availability is the Purpose, introduces asset management with respect to how it plays in the enterprise. It then establishes the measures of performance and the driving aspects to the measures. Because data and analytics should ride on purpose, you are advised to think out the equivalent framing of the operation in which you hold a role and the operations around it.

Of course, data and especially analytics are not, nor can be made, easy subjects. That is even after being bound within only what you need to know rather than all there is to know. As someone who read the book said, “You have to shut the door and turn off the TV.”

However, you will not need to go far before you begin experiencing the payback of your concentration. At each chapter you will return to work with an actionable idea to upgrade the output of your role and the roles of others.

Let’s talk about approaching the book to maximally bring about immediate payback to your organization and you as well. The book is written as the sequence of bringing the practices of asset management to be data driven. Asset management practitioners have got to love that. However, there is alternative path through the chapters for maximal “next-day” payback upon each stepwise accumulation of skills.

You should begin with Chapter 2, Data, Analytics and Software to be Data Driven. You need to know the territory. Just as important, you will discover that you already have at hand everything you need without first engaging management for permission to act on your payback idea or plead with them to buy you specialized software. Also important, the chapter will establish the definitions and demarcations with which you will be able calm management down by explaining why it is that data drivenness entails learned skills with immediate operational ramifications rather than capital and costs for exotic software and endeavors.

Here is one thing I discovered while learning much of what there is to know about data and analytics. Most of the power of data-drivenness is achievable with only the skills of working with data. Only a small number of ideas, less than 20 percent, call for analytics. In other words, if the forces of the universe commanded that I was to be forever limited to what I could make happen with data, I would be disappointed but still excited about what I have been left with.

Data Skills and Methods

When you know how to extract, explore and mold data into a dataset for purpose, your operation will immediately jump forward in its ramification for the enterprise. As a thought exercise, while you work through Chapter 3, Super Table from Operational Data, imagine what your operation would look like and perform like if every role holder were able to skillfully work with data as normal to their role. In fact, it is a readily achievable vision.

Working through Chapters 3, 8, 9 and 11, you will see in action almost every technique of building what the book calls super tables. The techniques are demonstrated in MS Access because the software is available to everyone by virtue of their organization's MS Office license. Furthermore, the principles of data that are demonstrated in Access travel easily to other software such as Tableau and Power BI.

Chapter 3, Super Tables from Operational Data, explains how data is extracted from operating systems, pulled into Access and joined together in a super table of all related variables. This is in contrast to Excel which leaves us divided and conquered because our data, although related, is present to us in disconnected tables. The chapter then explains how to create summary variables in the super table, a technique which loom huge in the ability to build insight from the data.

Chapter 8, Achieve Entirety of Data, is an important chapter because we must make all of our operational data available to building super tables. The chapter presents three frequently observed obstacles to entirety. To confront them, you will learn what changes must be made to operational roles and systems as necessary to achieve entirety.

Chapter 9, Workload-Based Budget and Variance, and Chapter 11, Budget-Based Schedule and Craft Capacity, demonstrate data skills in the context of expanding super tables to perform tedious, complex operational tasks. The value of data skills is highlighted by demonstrating a collection of related organizational tasks that are not otherwise possible in asset management and their absence has been a big impediment to enterprise performance. It will jump out at you that complex laborious reporting and control tasks can be reduced to merely updating the source tables at the input side of the super table.

Analytics Skills and Methods

What will be your career health at this juncture of acquired skills? It will be excellent. However, reaching preeminence would be even more excellent. More enticing is that preeminence is within your grasp because now "you know data."

Some time ago, I met a person in a large corporation whose meal ticket was his ability to produce Excel pivot charts and tables. The tables were not even super tables, nor did he know how to build them. Furthermore, the graphic capability of Excel to the fellow's claim to fame was based on single-perspective charting that was invented over 100 years ago, some as far back as the 1600's. Dashboards are an attempt to get past the limitation by placing multiple single-perspective charts in a single display. This is in contrast with layered perspectives in a single chart.

Thumb through Chapter 5, Layered Charting to Know Thy Data. You will see in action what is ggplot2 in the R software. The differences and ramifications will jump out at you. Imaging pulling such power to visualize data and measures into your role. In contrast, my buddy with his Excel pivots would be lost in your dust.

However, there is a bit of bad news. To get to layered visualization, you must step into the R software. As you will know from Chapter 2, R is a powerful and open software available for you or your firm to download without cost, limitations or strings.

The purpose of Chapter 4, The R Software in Action, is to get you up and running in R. Once again, the chapter is used to explain real life insights by demonstration. The chapter demonstrates the procedure to inspect a dataset statistically, reveal missing data, test variables for normal distribution and inspect the correlations between variables.

The philosophy of the book is that frequently occurring code will emerge from the explanations. Additional commonly occurring coding will emerge in the code for layered charting with the previously mentioned ggplot2 as well as from all of the other chapters to explain analytics. And just as for the Access code of the previously introduced chapters, you are left with R templates to substitute in your own variables, and R skills and code you will need to know.

You will arrive at another milestone upon working through Chapters 4 and 5. You will have accumulated the chops to begin moving into analytics. This is good because there is still an elephant in the room that requires analytics. It is to find and cleanse bad data. Of course, we should think of our best solutions for bad data are action taken to prevent a future of bad data in the operating system.

If there is bad data, what will be our strategy to rectify it? Chapter 6, Unearth and Rectify Bad Data, presents the types, decisions and schemes to rectify bad data. Several straightforward methods for some types of bad data were offered by Chapter 3. All others will require analytics to find and rectify them.

When some cases to a variable are bad, the question is what should they be? The cleansing process will lean on methods to replace bad cases, if necessary. However, the analytic is a core type for much more than replacing bad data with good. It is regression analytics to determine how strongly, if at all, specific operational variables are related to an outcome variable. Stated in the context of cleansing, the "outcome" variable can be the one with bad cases and we use the regression to estimate what they should be.

The chapters that explain two types of regressions are next in line to strengthen your career security. The analytics for relationships play in advancing operational effectiveness. For an outcome measure of performance, identifying the variables that matter most, or least, is powerful insight for assuring that your role is doing the right things right.

There are three mainstream regressions: linear, logistic and Poisson. The book will demonstrate linear and logistic. Chapter 7, Relate Operational Variables to Outcome, explains linear regression. Chapter 14, Recover Lost Classifications, explains logistic regression.

Regressions are a big step forward in your skills because they provide one of the most fundamental insights in operational capability. However, regression is not as simple as it would seem. We do not push all seemingly relevant variables into the model, run it and read the answer. Instead, there are considerable steps to choose and evaluate variables, and confirm fit: defined as how well the model accurately predicts the data.

Chapter 7 explains the process step by step. The same process generally applies to all regressions. I have never seen it mentioned in traditional texts that we can use the validation processes of regression to find outliers to the regression rather than only outliers to its variables. These are cases that the model would have never predicted and cases that have too much influence on the parameters of the model. Both are important insight to an operation because outliers may be telling us were to question and improve our operational processes and controls.

The difference between linear and logistic regression is their outcome variable. Linear predicts continuous numeric variables, whereas, logistic predicts classifications. Chapter 7 explains in detail the structure of the models and the distinction between them that rests upon the mathematics of prediction.

Chapter 14, Recover Lost Classifications, explains logistic regression while demonstrating a purpose for it beyond exploring relationships and outliers to outcomes. It can be used to determine what an incorrect classification to a variable should have been. The strength of relationship of predictive variables to good classifications is used to cleanse the bad cases.

However, the purpose of the chapter is dual rather than merely to explain logistic regression. As the title suggests, the overarching purpose of the methods demonstrated is to recover lost classifications.

Accordingly, the chapter will also introduce and demonstrate the method known as naïve Bayes probability. We use naïve Bayes to determine from unstructured free-text variables, such as descriptions and notes, what a classification should be upon the probable occurrence of words in the text.

The next logical extension of your skills is to be able to prove there is a difference, or will be, as the consequence of change, improvement and enforcement of operational procedures and resources. Chapter 13, Prove There is a Difference, explains the body of analytics to make the determination vis-a-vis the operational situation. The analytics are a set of eight models from the types of two-means t test, ANOVA and multilevel.

Chapter 10, Through the Lens of Time Series, introduces another method to spot change through data and analytics skills. We look for patterns in variables presented in time series. However, our interest is not limited to change because overall we want to know what has happened over time.

Time series analytics are not to be confused with a line chart in Excel in which essential insights are lost in the simplicity. The chapter will explain by demonstration how to separate recurring cycles from the core pattern. It will explain how to measure the degree that one period reflects one or more previous periods. It will explain how the same principle is used to identify variables with a lead and lag relationship.

Just as importantly, we need to assure that a core pattern to a series is deterministic rather than a random walk. This is because a random walk can look deterministic. If not deterministic, we need to know to not devise actions or attempt forecasts as if it were.

Finally, in snaking through the chapters, we arrive at a final fundamental characteristic of any operation. It is the time it takes for each discrete core operational task to pass through the stages of the operation. Chapter 12, Elapsed Time Through Stages, explains by demonstration how to determine the statistics of retention in a stage and the chance of exit having remained in the stage for some duration. Just as important, the analytic enables us to determine which variables in the operation are most strongly related to retention and exit, thus, are the levers to reshape both characteristics of the stages that matter most.

Where to Go Next

The book has bound its scope to the de facto favorite mainstream analytics and, for them, what you need to know to build and interpret them. References are provided if you want to take the watch apart rather than merely know how to tell time. What is most exciting is that, with each chapter, your ability to reach into, understand and engage additional methods outside the book's scope will grow and solidify.

I was once told that to qualify as an academic textbook there must be assignments with each chapter. We could say that replicating the demonstrations in each chapter is your assignment. My personal experience is that a powerful solidifying exercise is to emulate each demonstration and explore its code command by command. However, there is an assignment we could make for each chapter.

First, you must ask yourself a question. For which roles of my operation do the chapter's skills and methods apply, and how and why?" Second, you must act. The assignment is to return and install new methods in your role, thus, equip your organization with actionable insights it has never had before. Third, you must share the knowledge. The assignment is to pass each of your newly acquired skills to others by identifying roles for which the chapter's methods could be impactful and helping the holders of the roles to increase their career security by bringing new value to the enterprise through their work in the role.

Welcome to the new world that opens to you as a new age worker with all career security appertaining thereto.

Data and Code to the Book

From literature and the internet, there are massive resources to explain all principles and methods using the R software and, for that matter, just about any software. The standing rule in the R community is that every provided explanation from any source must be accompanied with an example, code and dataset.

The book honors the rule by making the datasets and R code (scripts) to the chapters available to download from the webpage, https://analytics4strategy.com/data-and-code. You will need the datasets to emulate the demonstrations. The scripts are placed in the text for explanation by demonstration. However, since the code cannot be copied from the book and pasted into an R session, you have the option to copy and paste scripts into your R session.

Color and Format

The book is a complex manuscript and, thus, presents challenges to formatting and cost. The body of the book is a progression through text explanations, supporting figures, output exhibits and the R and Access code that generated the exhibits. In formatting the pages, all demand a degree of sequential rigidity for being laced together in an explanation by demonstration that can be easily followed.

One issue in formatting is that some output exhibits are visualizations that depend heavily on color to communicate the insight they are coded to give us. This is especially the case for visualizations that demonstrate the advance graphics of the ggplot2 package of the R software.

Unfortunately, the cost to produce the book in color is prohibitive to pricing. To keep the price reasonable, the book is produced in grayscale. The readers, of course, can view each graphic in full color by running the provided R code. Alternatively, the reader can view any visualization in full color at webpage, https://analytics4strategy.com/book-look-inside.

Is it said in data and analytics, "the devil is in the details." When something does not work, it is often a tiny error in typing. Given the number of code blocks and internet addresses, hyphenation has not been used in formatting the text. Accordingly, readers need not wonder if a hyphen in a line of code or an internet address are from formatting or is as it should be. However, this policy occasionally creates some ugly lines in the text that hyphenation would eliminate. I ask for your forbearance.

In a few cases, the placement of output exhibits causes an over large empty white space at the bottom of a page. It is unavoidable given that some sequences must be honored with respect to the associated Access and R code upon which they are generated. Once again, I ask for your forbearance.

So, now onward we go through the fog and the dark where skill and education win over ignorance and superstition every time.

Richard G. Lamb, PE, CPA.

Index

1R algorithm, 437
Accessible on demand, 225–30
Accessible on demand defined, 222
Achievable availability, 7, 353
Adminstration of tables, 80–81
Aggregation variables, 65–66
  Expression and Where, 72, 252
  Summary options, 66
Akaike information criteria (AIC), 190
Analytics, interpretation of
  Akaike information criteria (AIC), 449
  Akaike information criterion (AIC), 190
  Centrality and spread, 120
  Confidence limits, 99
  Exponentially weighted moving average (EWMA). *See* Time series, principles of; Exponentially weighted moving average (EWMA)
  Hypothesis test, 98
  Kurtosis, 121
  Mean, 120
  Mean absolute deviation from median, 121
  Mean trimmed of extremes, 121
  Median, 120
  Min-max, 120
  Pearson, 97
  p-value, 99
  Quartiles, 120
  Range, 121
  Skew, 121
  Spearman, 97
  Standard deviation, 121
  Standard error to the mean, 121
Analytics, types of, 95–100
  ANCOVA. *See* ANCOVA, principles of; ANCOVA, templates and interpretation; and ANCOVA, R functions
  ANOVA, factorial. *See* Factorial ANOVA, principles of; Factorial ANOVA, templates and interpretation; and Factorial ANOVA, R functions
  ANOVA, one-way. *See* One-way ANOVA, principles of; One-way ANOVA, templates and interpretation; and One-way ANOVA, R functions
ARIMA model. *See* Time series, principles of; ARIMA (Autoregressive integrated moving average)
Correlation analysis
  Between categorical and numeric variables, 132
  Correlation coefficient (R value), 95
  Correlation defined, 95
  Regression Coefficient defined, 95
Correlation compared to regression analysis, 103–8
Exponentially weighted moving average (EWMA). *See* Time series, principles of; Exponentially weighted moving average (EWMA)
Factorial repeated measures. *See* Factorial repeated measures, principles of; Factorial repeated measures, templates and interpretation; and Factorial repeated measures, R functions
General linear model (GLM). *See* general linear model
Holt Winters. *See* Time series, principles of; Holt Winters
Linear regression. *See* linear regression templates and interpretation
Logistic regression. *See* Logistic regression, principles of; Logistic regression, templates and interpretation; and Logistic regression, R functions
Missing data, 110–17
  Steps to, 110
Mixed model, 384
Multilevel model, 384
Naïve Bayes. *See* Naïve Bayes, principles of; Naïve Bayes, templates and interpretation; and Naïve Bayes regression, R functions
One-way repeated measures. *See* One-way repeated measures, principles of; One-way repeated measures, templates and interpretation; and One-way repeated measures, R functions
Partial correlation analysis, 100–103
  Correlation (R), 100
  Correlation squared (R^2), 100
  Misleading, as, 102

Robust difference model, 384
Robust linear regression, 107
Survival hazard. *See* Duration analytics, principles of, Duration analytics, templates and interpretation, and Duration Analytics, R functions
Time series, 241, *See* Time series, principles of, Time series, templates and interpretation, and Time series, R functions
Two-mean t-test, single factor, two conditions. *See* Two-mean t-test, principles of, Two-mean t-test, templates and interpretation, and Two-mean t-test, R functions

ANCOVA. *See* ANCOVA, principles of; ANCOVA, templates and interpretation; and ANCOVA, R functions

ANCOVA, principles of, 381–82, 404–13
Continuous variable in ANCOVA, 381
Covariate, 381
Mean, se and confidence limits adjusted for covariate, 408
Unexplained variance, 381

ANCOVA, R functions
Anova(), 408, 412
aov(), 407, 408, 412
effect(), 408
glht(), 411

ANCOVA, templates and interpretation
Anova() to test for homogeneity, 412
aov() test for significant difference contrasts as main effects, 408
aov() to test covariate as independent, 407
Build contrast variables and inspect, 407
by() for descriptive statistics, 406
effect() to adjust mean, se and limits for covariate, 408
factor() to set variable base condition, 405
ggplot2 to plot points and covariate to score, 410
ggplot2 to plot variable condition means and limits, 419
ggplot2 to plot variate regression to condition points, 412
glht() to adjust limits with Tukey, 411
leveneTest() test for constant variance across conditions, 406
plot() of variance and normality to residuals, 411
summary.lm() test significant difference of means, contrasts as effects, adjusted for covariate, 409
summary.lm() to inspect contrasts for difference, 420

ANOVA. *See* One-way ANOVA, principles of; One-way ANOVA, templates and interpretation; and One-way ANOVA, R functions

ANOVA, factorial. *See* Factorial ANOVA

Apparency insight, 40–41
Models, 40

Append query, 50, 56

ARIMA model. *See* Time series, principles of; ARIMA (Autogressive integrated moving average)

Artificial intelligence, 438, 441

Artificial intelligence (AI), 207–17

Autocorrelation, 37

Availability performance, 2–4
Constituent probability, as, 3
Contrast to OEE, 2
defined, 2
Depicted, 3
Equation, as, 3
Life curves, 4
Operational availability, 9–11
Cost effectiveness, 11–13
Factors of, 10–11
Probability, as, 2
Subtypes, as, 6–7
Achievable, 7
Equations, 7–8
Inherent, 6
Interrelationships between, 9
Operational, 7

Backlog, workload, 249

Bad data, imputation
Models for
Linear regresssion, 163
Logistic regresssion, 165
Multilevel (MLM), 167
Naïve Bayes, 165
Trees, 166
Trees, 163
Trees, 168
Translation tables, 164

Bad data, rectification
Eliminate, 160
Impute estimate (ML and AI), 160

Bad data, types
Inapparent without measurement, 159
Misformatted and mis mispelled, 159
Missing, 159
Outliers, Leveragers and influencers, 161
Unexpected, 159

Base condition, set, 395

Base model, 380
Bayes probabilty, 438
Benjamini-Hochberg adjustment, 385, 402
Bibliography, 14, 81, 108, 158, 169, 219, 236, 278, 318, 351, 375, 436, 473
Binary logistic equation, 442
Blanchard, Benjamin S, 14
Bolin, Jocelyn, 443, 474
Bonferroni adjustment, 385, 402
Budget and variance, principles and data methodology, 240–42, 243–45
  Actuals, table of, 253–55
  Budget, contrast one- and two-dimensional, 242
  Case as progression, 247–49
    Flowchart, 248
    Variance Output reports, 248
    Workload from history, 250
  Craft capacity as workforce, 270–78
  Dual dimensional
    Activity (AKA, workload), 241
  Factor of cost, 241
  Mission, structure and derivation, 238–40
    Derivation, 239
    Mission, 238
    Structure, 238
  Objectives, three, 237
  Template tables, queries and pivots, 249–53
  Variance supertable and pivot, 255–57
  Variance, contrast one- and two-dimensional, 239
  Variance, two-dimensional computations, 243, 257
Budget-based schedule cycle. *See* Schedule cycle, stages to
Budget-based schedule measures and methods. *See* Schedule cycle, measures of
Budget-based schedule tables. *See* Schedule cycle, tables to
Budgeting, types of
  Predictive-based, 239
  Zero-based, 239
Categorical variables, interpret, 188
Censored event, 358, 360
Centality and spread, 120
Classification insight, 39–40
  Models, 39
Classification, Procedure to recover, 438–40
  Artificial intellegence, 438
  Machine learning, 438
  X, 438
Classifiers for categories
  Aggregation, upon, 72–78
Cleanse data, steps to, 207
Concatenation operator, 58
Confidence interval to regression coefficient, 189
Confidence limits, 99
Contrasts as effects, 386–89
  Contrast variables, build, 387
  Family error, 386
Cooks distance, test of cases, 209
Corpus, text dataset, 462
Correlation (R), 100
Correlation analysis
  Correlation defined, 95
  Regression Coefficient defined, 95
Correlation coefficient (R value), 95
Correlation compared to regression analysis, 103–8
Correlation defined, 95
Correlation squared (R^2), 100
Criteria row, Access code
  Conditional expressions, 60
  Contains, 59
  Dates, working with, 60
  Equals, 59
  Numeric, for, 60
  OR and AND, 58
  Text strings, for, 59
Critical mass and grass roots, 16–19
  Characteristics, qualifying, 17
  Critical mass, 16
  Grass root, 16
  Triad of software, 22
Cross correlation, 37
Crosstab query, 50, 233
Cumulative hazard, 358
Cumulative hazard function, 361, 364
Current to capture, 225–30
Current to capture defined, 221
Data analytics, principles of
  Multiple predictor model, 370–73
  Parametric model
    Weibull distribution, eras of, 369
Data and report as layers, 223
  First layer, 224
  Second layer, 224
Data and variance, principles and data methodology
  Reports and interaction analysis, 258–62, 262–70
Data driven, 15–16
  depicted, 16
  Essential definitions, 19–22
    Alogrithm, 21

    - Artificial intelligence, 21
    - Data and big data, 20
    - Machine learning, 21
  - Insight deliverables. *See* Insight deliverables
- Data frame, 98
- Data table compared to Spreadsheet, 223
- Data, entirety of
  - Accessible on demand, 230–36
    - Relational database, 230
  - Accessible on demand defned, 222
  - Current to capture, 225–30
  - Current to capture defned, 221
  - Data and report as layers, 223
    - First layer, 224
    - Second layer, 224
  - Fused data and report, 223
  - Layered structure, 222–25
  - Layered structure defned, 221
  - Spreadsheet compared to data table, 223
  - Table as format, 22, 222
- Data, reshape, 428
- Decision tree model, 40
- Decision trees, 437
- Delete query, 50
- Design grid rows, Access, 56–61, 56–61, 56
  - Criteria. *See* Criteria row code, Access
  - Field. *See* Field row code, Access
  - Show, 57
  - Sort, 57
  - Table, 57
  - Total, 62, 66
    - Summary options, 62
- Deviance as measure of fit, 448
- DFBeta, 451
- Difference analytics, 383–84, *See* also one-way repeated and factorial repeated measures and respective principles of and templates and interpretation
- Difference insight, 35–36
  - Family error, 36
  - Models, 36
- Document term matrix, 466
- Dual dimensional budget and variance, 240
- Dummy variables, 182–85
- Dunnett adjustment, 385, 403
- Duration analytics, principles of
  - Baseline model, defined, 366
  - Empircal construct
    - Censored event, 360
    - Cumulative hazard function, 361
    - Hazard function, 361
    - Kaplan-Meir estimator, 360
    - Nelson-Aalen estimator, 360
    - Survival function, 362
  - Empircal construct of, 360–62
  - Empirical model, 366–68
  - Hazard defined, 357, 359
  - Left truncated, 359
  - Life data, event and window, 358–60
    - Censored event, 358
    - Cumulative hazard, 358
    - Event, defined, 358
    - Life as hazard, 358
  - Mathematical construct, 362–64
    - Beta parameter, slope, 362
    - Cumulative hazard function (H(t)), 364
    - Eta parameter, scale, 362
    - Failure function (F(t)), 363
    - Hazard function (h(t)), 364
    - Mean time to specified event (MTTF), 364
    - Survival function (R(t)), 363
  - Median rank regression method (MRR), 374
  - Most likely estimate method (MLE), 374
  - Parametric model, 368–70
  - Survival defined, 355
  - Weibull plot, 373–75
- Duration analytics, R functions
  - coxph(), 366
  - MLEW2p(), 374
  - MRRW2p(), 374
  - phreg(), 368, 370, 372
  - strata(), 370
  - Surv(), 366, 368, 370, 372
  - survfit(), 366
- Duration analytics, templates and interpretation
  - coxph() for empirical model of data, 366
  - phreg() baseline parametric model, 368
  - phreg() parametric model with one or more variables, 372
  - phreg() parametric model with strata variable, 370
  - Weibull model with Median rank regression method (MRR), 374
  - Weibull Model with most likely estimate method (MLE), 374
  - Weibull prepare data for model, 373
- Duration insight, 38–39
  - Models, 39
  - Plots to insight
    - Cummulative hazard, 38
    - Hazard, 38

survival, 38
Durbin Watson test, 200
Entirety of data. *See* Data, entirety of
Event history analysis, 38
Exponentially weighted moving average (EWMA). *See* Time series, principles of; Exponentially weighted moving average (EWMA)
Factorial ANOVA, principles of, 382–83, 413–21
  Post hoc, method for, 421
  Visualize variance and normality to convert to one-way ANOVA, 421
Factorial ANOVA, R functions
  Anova
    (), 418
  factor(), 414
  summary.lm(), 420
Factorial ANOVA, templates and interpretation
  aov() fit and test for presence of differences, 418
  by() descriptive statistics, individual and interactive, 416
  Contrast() to build contrast variables and inspect, 418
  factor() t set base conditions to variables, 414
  ggplot2 to visualize relationship of factorial variables, 415
  leveneTest() for constant variance across conditions, 417
  plot() to visualize variance and normalcy to factorial converted to one-way ANOVA, 421
Factorial repeated measures, principles of, 427–34
  Reshape data, 428
  Robust model, 434
Factorial repeated measures, R functions
  anova(), 431
  lme(), 431
Factorial repeated measures, templates and interpretation
  Build contrast variables and inspect, 430
  by() for descriptive statistics, 430
  ggplot2 to compare conditions and interaction, 433
  glht() post hoc of multilevel model with Tukey, 434
  lme() one variable subjected to multiple measures and progressive predictor variables, 431
  melt() make wide data long, 429
Failure function, 363
Family error, 36, 379, 385, 386
Field row, Access code
  Aliases, 69, 78
  Calculated, 59
    IIF, 60, 71
    Switch, 60
  Combined with concatenation, 58
  Concatenation operator, 58
  Names common to multiple subtables, 58
Field, Andy, 473
Finch, Holmes, 443, 473, 474
F-statistic, 187
Fused data and report, 223
General linear model (GLM), 172–74
  Defined, 172
  Dummy variables, 182–85
  Interaction variable, 173
  Main effects, 173
  Null model, 172
  Select variables, 185–93
  Three components
    Linking function and solution, 174–77
  Three regressions, 171
  Transformation of variables, 173, 193–94
ggplot2 geoms
  geom_bar(), 149, 150
  geom_boxplot(), 141, 143, 157
  geom_contour(), 138
  geom_count(), 138, 139
  geom_density(), 147, 205, 206
  geom_freqpoly(), 144, 146
  geom_hex(), 138, 139
  geom_histogram(), 144, 145, 147, 204
  geom_jitter(), 138, 141, 142
  geom_line(), 152, 156, 157
  geom_path(), 154
  geom_point(), 135, 137, 139, 141, 142, 143, 154, 157, 194, 195, 198, 199, 200, 202, 209, 212
  geom_raster(), 138
  geom_smooth(), 135, 137, 156, 194, 195, 198, 199, 200, 202, 209, 212
  geom_violin(), 142
ggplot2 templates
  Barchart as mean, 150
  Barchart with dodge, 149
  Boxplot with bins for continuous variable, 143
  Boxplot with geom_jitter(), 141
  Comparison of confidence limits of mean effects, 396
  Density plot with geom_density, 205, 206
  Density with geom_histogram, 204
  ggplot2 with geom_boxplot, geom_point for mean and points, and geom_jittter, 391

Histogram and polygon overlay with geom_histogram and geom_freqpoly, 144
Histogram as density, 147
Histogram subsetted categories with geom_histogram, 145
Line chart for groups with line fit to overall, 156
Line chart with boxplot and mean, 157
Line chart with geom_line, 152
Path chart over geom_point with geom_path, 154
Polygon as density, 147
Polygon with categories overlaid by geom_freqpoly, 146
Q-Q plot, 122, 127, 204
Q-Q plot with facet_grid(), 129
Q-Q plot with facet_wrap, 125, 128
Scatter chart with geom_point, 154
Scatter plot of multiple datasets with geom_point and geom_smooth, 209
Scatter plot with geom_count() for overplotting, 139
Scatter plot with geom_hex() for overplotting, 139
Scatter plot with geom_point, geom_smooth and facet_grid, 194, 195, 198, 199, 200, 202
Scatter plot with geom_point, geom_smooth and facet-wrap, 135
Scatter plot with geom_smooth(), 134
Scatter plot with multiple datasets, geom_point and geom_smooth, 137, 212
Violin chart with geom_jitter for overlapping, 142
Hat test of cases, 208
Hazard function, 361, 364
Heap, Howard F, 14
Holt Winters. *See* Time series, principles of; Holt Winters
Independence of errors, 200
Industrial internet of things (IIoT), 18
Feasibility of, 19
Inherent availability, 6, 353
Inner join, 53
Insight deliverables, 31–41
Know-thy-data, *32–33*
Modeled, 33–41
Apparency. *See* Apparency insight
Classification. *See* Classification insight
Difference. *See* Difference insight
Duration. *See* Duration insight
Relationship. *See* Relationship insight
Time series. *See* Time series insight
Recountive, 32
System reports, 31
Institute for Digital Research and education, 474
Interaction of variables, 197
JoinProperties window, 54
Joins complex
Concatenated join, 69, 78, 251
Multiple variables join, 69, 78
Joins in query
Join properties window, 54
Type
Inner, 53
Left, 53
Outer, 54
Right, 53
Jones, James V, 14
Kaplan-Mier estimator, 360
Kelly, Ken, 443, 474
K-means, 437
K-Means model, 40
K-NN, 437
Know-thy-data, 90–95
Layered charting, 94
Normal distribution, 92
Q-Q graphic, 93
Normal distribution, tests of
Shapiro-Wilk test, 93
Pairs panel, 92
Kurtosis, 121
Lamb, Richard G, 14
Lantz, Brett, 437, 474
Layered charting, 29–31, 94
Defined, 29
Traditional, compared to, 29, 32
Layered structure, 222–25
Layered structure defined, 221
Left join, 53
Left truncated, 359
Leverage, 451
Life curves, 4
Linear regression, 34, 163
Linear regression templates and interpretation
lm() with main effect and interaction variables, 212
lm() with main effect and transformed variables, 195
lm() with main effect variables, 104, 189, 191
lm() with main effect, interaction and transformed variables, 215, 217
p-value, 187
Residual standard error, 190, 192, 197
Test for generalization

Variance inflation factor (VIF), 201
Linear regression, principles of
Akaike information criterion (AIC), 190
Artificial intelligence (AI), 207–17
Categorical variables, interpret, 188
Cleanse data, steps to, 207
Confidence interval to regression coefficient, 189
Cooks distance, test of cases, 208, 209
Dummy variables, 182–85
F-statistic, 187
Hat test of cases, 208
Imputation by estimate, 213
Interaction of variables, 197
Machine learning, 177–78
Predict individual cases, 217–18
Residual standard error, 187
R-squared, 187, 192
Select variables, 185–93
Standardized effect, 185
Standardized residuals, 194
Test for generalization, 200–207
Durbin Watson test, 200
Independence of errors, 200
Normal distribution of residuals, 203
Transformation of variables, 193–94
Linear regression, templates and interpretation
cooks.distance() to each case in data set, 209
lm() with main effect variables, 185, 192
lm() with main effect, interaction and transformed variables, 196
lm()with main effect variables, 187
Linking function, 441, 442
Logistic regression, 34, 165, 438
Logistic regression, principles of, 457–59
Artificial intelligence, 441
Binary logistic equation, 442
Build and interpret, 445–49
Classification as probability, 450–51
Classify cases, 452–54
Confidence limits to probability, computation, 453
Deviance as measure of fit, 448
DFBeta, 451
Leverage, 451
Linearity, method of, 456
Linearity, test for, 454–57
Linking function, 441
Logistic regression as classifier, 440–59
Machine learning, 441
Mulinomial logistic equation, 442
Multicolinearity, test for, 454–57
Multilevel model, 443
Multinomial logistic regression, 445
Null model, 459
Odd ratio, 446
Odds ratio, upper and lower limits, 447
Ordinal logistic regression, 443
Regression coefficients, effects of, 443, 446
Residuals, evaluate, 451
Solution equation, 442
Standardized residuals, 451
Studentized residuals, 451
Systematic component, 443
Logistic regression, R functions
(), 451
dfbeta(), 451
dffits(), 451
glm(), 445, 455, 456, 457
rstandard(), 451
rstudent(), 451
Logistic regression, templates and interpretation
Cases to be classified, 452
Compare deviance to null model, 448, 449
Evaluate residuals to logistic model, 451
Expand cases table to include probabilities and confidence limits, 453
glm() test for linearity, 456
glm() to build logistic model, 445, 455
Logistic regression to classify sample data, 457
Predict() of probability of the base classification, 452
relevel() set baseline category to model, 444
Table of classification as probabilities, 450
Upper, lower limits to odds ratio, 447
vif() to test for multicolinearity, 455
Machine learning, 177–78, 438, 441
Main effects, 173
Maintenance tasks from FMECA, 5–6
Make table query, 74
Make Table query, 50
Matrix, 98
Mean, 120, 263
Mean absolute deviation from median, 121
Mean time to specified event, 364
Mean trimmed of extremes, 121
Median, 120
Min-Max, 120
Missing data, 110–17
Mixed model, 384, 435
Moubray, John, 14

Multilevel model, 381, 384, 389, 422, 425, 426, 427, 431, 434, 435, 443
Multinomial logistic equation, 442
Multinomial logistic regression, 445
Naïve Bayes, principles of, 459–73
  Classify cases, 472–73
  Cleanse and standardize text data, 462–64
  Cleanse and standardize text, steps, 463
  Concept of Naïve Bayes, 460
  Corpus, text dataset, 462
  Document term matrix, 466
  Frequent words, indicators of, 468–69
  Rectify misclassifications, 165
  Train and evaluate model, 469–72
  Train and test datasets, create, 467–68
  Word clouds, visualize text data, 464–66
Naïve Bayes, R functions
  DocumentTermMatrix(), 466
  findFreqTerms (), 468
  inspect(), 462
  naiveBayes(), 469, 471
  tm_map(), 463
  VCorpus(), 462
  wordcloud(), 464, 465
Naïve Bayes, templates and interpretation
  Convert document term matrix to categorical data, 469
  CrossTable() to evaluate model trained with Laplace, 471
  CrossTable()to evaluate model error, 470
  DocumentTermMatrix() to generate matrix of terms and freq, 466
  Download text data, 461
  findFreqTerms() to reduce document term matrix to frequency threshold, 468
  gsub() to remove one or more sequential items in text, 462
  Improve Bayes model with Laplace, 471
  Inspect text in text corpus, 462
  naiveBayes() to created trained model, 469
  predict() to estimate unclassified cases with model, 473
  predict() to test model with text dataset, 470
  Subset document term matrix, 467
  tm_map() in steps to cleanse and standardize text dataset, 463
  VCorpus() for corpus of text dataset, 462
  wordcloud() of cleansed and standardized dataset, 464
  wordcloud() subset on classified text dataset, 465
Nelson-Aalen estimator, 360
Normal distribution, test of
  Q-Q plots, 204
Normal distribution, tests of
  Histograms, 64
  Q-Q plots, 93
  Q-Q plots, 64
  Q-Q plots, 122
  Q-Q plots, 123
  Q-Q plots, 127
  Shapiro-Wilk, 64, 93, 124, 126
Nowlan, F. Stanley, 14
Null model, 172, 448, 459
Objects
  Prefixes to object, 51
odds ratio, 446
Odds ratio, upper and lower limit, 447
One-way ANOVA, principles of, 379–81, 394–404
  Base condition, set, 395
  Base model, 380
  Benjamini-Hochberg adjustment, 402
  Bonferroni adjustment, 402
  Concept of ANOVA, 380
  Dependent one-way ANOVA, 381
  Dunnett, 403
  Family error, 379
  Independent one-way ANOVA, 381
  Post hoc adjustment, 381
  Robust molde for one-way ANOVA, 404
  Tukey, 403
  Welch's F ANOVA, 399
  XContrasts as effects, 381
One-way ANOVA, R functions
  aov(), 398, 401
  glht(), 403
  leveneTest(), 397
  oneway.test(), 399
One-way ANOVA, templates and interpretation
  aov() for significant difference to contrasts as main effects, 398, 401
  Build contrast variables and inspect, 400
  by() for descriptive statistics, 397
  factor() to set variable base condition, 395
  ggplot2() to compare confidence limits to effects, 396
  glht() to adjust confidence intervals with Tukey and Dunnett, 403
  leveneTest() test for constant variance across conditions, 397

  - oneway.test() for difference adjusted for unequal variance, 399
  - pairwise.t.test() to adjust confidence intervals with Benjamini-Hochberg, 403
  - pairwise.t.test() to adjust confidence intervals with Bonferroni, 402
  - plot() of variance and normality to residuals, 399
- One-way repeated measures, principles of, 422–27
  - Multilevel model, 422
  - Multilevel model, lme(), 425
- One-way repeated measures, R functions
  - glht(), 427
  - lme(), 426
- One-way repeated measures, templates and interpretation
  - anova() to test multilevel model with predictor against baseline model, 426
  - Build contrast variables and inspect, 425
  - by() for descriptive statistics, 425
  - ggplot2 to compare conditions, 424
  - glht() post hoc of multilevel with Tukey, 427
  - lme() baseline multilevel model, single repeated measure variable, 426
  - lme() model with predictor variable and single variable repeated measure, 426
  - melt() make wide data long, 423
- Operational availability, 7, 353
- Or row, Access code, 58
- Ordinal logistic regression, 443
- Outer join, 54
- Outliers, seek
  - Confidence limits
    - Contrast Z- to T-Score, 70
  - Normal distribution, assumptions of, 63
  - Student’s T-Score
    - Excel functions for T_Score, 71
    - Translation table to measure for outliers, 71
  - Student's T-Score, 70–71
  - Z-Score, 63–65, 67–70
- Overall equipment effectiveness OEE, 2
- Pairs panel, 92, 131
- Partial correlation analysis, 100–103
  - Correlation (R), 100
  - Correlation squared (R^2), 100
  - Misleading, as, 102
- Pearson, 97
- Poisson regression, 35
- Post hoc adjusted confidence limits, 385–86
  - Benjamini-Hochberg, 385
  - Bonferroni adjustment, 385
  - Dunnett, 385
  - Family error, 385
  - Tukey, 385
- Power BI, 26
- Predictive-based budgeting, 239
- p-value, 99, 187
- Q-Q graphic, 93
- Quartiles, 120
- Query process, overview, 50–53
- Query types
  - Append, 50, 340
  - Crosstab, 50, 233
  - Delete, 50
  - Make Table, 50
  - Make-table, 340
  - Select, 50
  - Update, 50
- Query, table and pivot templates
  - Budget and variance
    - pvtActualVsBudget, 258
    - qryActualGrp2020, 254
    - qryActualVsBudget, 256
    - qryBudget2020, 251
    - tblActual2020, 254
    - tblHistory2019, 251
  - Craft capacity
    - pvtCraftCapacity, 272
    - qryAggCraftBaseline, 274
    - qryAggCraftSample, 274
    - qrySampleVsBaseline, 276
    - tblCraftBaseline, 274
    - tblCraftSample, 274
  - Schedule week and day
    - ctqStatusCount, 345
    - qryScheduleVsActual020320, 339
    - tblCraftHrs, 334
    - tblSchedVsActual020520, 336
  - Status History
    - ctqStatusHistoryWide, 233
    - qryStageElapsed, 235
  - Super table, classifier and outlier
    - qryCase1_Foundation, 56
    - qryCraftGroupHours, 72
    - qryCraftGrpStats, 76
    - qryHoursOrder, 65
    - qryHrsLeadCraft, 75
    - qryLeadCraftClassifier, 74
    - qryOrderGrpStats, 67
    - qryZScoreOrder, 69
    - qryZScoreOrderCraft, 77

  - Variance as outlier or systemic
    - pvtOutlierOrders, 268
    - qryOutlierAggregate, 263
    - qryOutlierAggWithTest, 266
    - qryOutlierOrders, 266

- R coding, 447, *See* also R functions
  - [] as filter, 96, 113, 127, 132, 180, 208, 209, 211, 212, 214, 215, 216, 287, 295, 368, 373, 429, 455
  - '~' as '=', 104
  - ’|’ as OR, 132
  - <- to assign, 89
  - Case sensitivity, 86
  - confint() for confidence interval, 104
  - Data frame, 98, 202
  - Hash mark (#), 85
  - List, 202
  - Matrix, 98, 202
  - na.rm=TRUE to ignore missing data, 117
  - table$variable syntax, 95, 298, 308, 314, 368, 393, 429
  - Vector, 98
- R functions
  - (contrasts(), 425
  - abline(), 300
  - acf(), 295, 298, 314
  - aggregate(), 284, 302
  - AIC(), 190, 191, 192, 196, 199
  - anova(), 196, 199, 431
  - Anova(), 408, 412, 418
  - aov(), 398, 401, 407, 408, 412
  - as.character(), 120, 462
  - as.data.frame(), 473
  - as.factor, 452
  - as.factor(), 120, 181
  - as.matrix() convert object to matrix, 98
  - as.numeric(), 120, 304
  - as.vector(), 373
  - attributes(), 182
  - auto.arima(), 298, 311
  - autoplot(), 311
  - boxplot(), 282, 302, 406
  - by(), 397, 406, 416, 425, 430
  - c(), 218, 312
  - c() combine function, 94, 96
  - cbind(), 183, 288, 294, 310, 312, 400, 407, 418, 430, 452
  - ccf(), 301
  - class(), 282
  - confint(), 187, 189, 403, 411, 427, 447
  - contrasts(), 182, 183, 418, 430
  - cor(), 455
  - cor() for correlation analysis, 97
  - cor.test() for correlation analysis, 99, 133
  - coxph(), 366
  - CrossTable(), 470, 471
  - data(), 289
  - Data(), 295
  - data.frame, 452
  - data.frame(), 218, 304, 310, 312
  - date_decimal(), 312
  - decompose(), 292, 295, 301, 304
  - describe(), 91, 121
  - dev.off() to close output device, 114, 158
  - dfbeta(), 451
  - dffits(), 451
  - diff(), 314, 318
  - distinct(), 214
  - DocumentTermMatrix(), 466
  - durbinwatsontest(), 200
  - effect(), 408
  - end(), 282, 288, 289
  - exp(), 368, 447
  - factor(), 395, 405, 414, 461
  - findFreqTerms(), 468
  - fitted(), 450
  - forecast(), 311
  - format(), 310, 312
  - frequency(), 282
  - ggarrange(), 136, 206
  - gl(), 429
  - glht(), 403, 411, 427, 434
  - glm(), 445, 455, 456, 457
  - gsub(), 462
  - Guidance, 96
  - hatvalues(), 208, 451
  - head(), 90, 119
  - help(), 96
  - Holt Winters(), 308
  - ifelse(), 133, 181
  - inspect(), 462
  - install.packages(), 86, 179, 281
  - lapply(), 462, 463
  - layout(), 282, 284, 286, 289, 295, 300, 301, 302, 314
  - length(), 208, 211, 212, 295
  - leveneTest(), 397, 406, 417
  - library(), 87, 179
  - lines(), 313
  - lm() for linear regression, 104

lm.beta, 186
lme(), 426, 431
log(), 456
md.pattern(), 113, 121, 180
mean(), 318
melt(), 423, 429
MLEW2p(), 374
MRRW2p(), 374
naiveBayes(), 469, 471
names(), 429
oneway.test(), 399
pairs.panel(), 92, 131, 180
par(), 399, 406, 411, 465
par() for parameters, 94
pchisq(), 448, 449
pcor() for partial correlation, 102
pcor.test() for partial correlation, 103
pdf() output device to pdf document, 158
phreg(), 368, 370, 372
plot(), 282, 284, 288, 289, 295, 366, 399
png() to output .png file, 114
predict(), 215, 218, 309, 452, 470, 471, 473
prop.table(), 467
qqline(), 94
qqnorm(), 94
rbind(), 216
rcorr() for correlation analysis, 98
read.csv(), 89, 112
read.delim() to read txt. and dat. file, 89
read.xlsx(), 89, 113, 119, 286, 299
relevel(), 184
rlm() for robust linear regression, 107
rstandard(), 194, 195, 198, 212, 451
rstudent(), 451
seq(), 135
shapiro.test(), 93, 124
sqldf(), also see SQL in background, 90
start(), 282, 288
str(), 90
Strata(), 370
subset(), 465
summary(), 91, 120, 366, 398, 401, 427
summary.lm(), 420
Surv(), 366, 368, 370, 372
survfit(), 366
t.test(), 392, 393
table(), 467
time(), 312
tm_map(), 463
ts(), 284, 286, 287, 300, 302, 314
ts.intersect(), 287
ts.plot(), 294, 300, 309
update(), 431
VCorpus(), 462
vif(), 201, 455
window(), 285, 301, 304
within(), 452
wordcloud(), 464, 465
write() export data from R, 90
write.xlsx(), 284, 287
xlsx(), 179

R packages
- car, 405, 444
- car(), 394
- e1071, 461
- effects, 405
- forecast, 281
- ggfortify, 281
- ggm, 86, 110
- ggplot2, 86, 110, 179, 389, 405, 422, 428
- ggplot2(), 394
- ggplotr, 179
- ggpubr, 110, 179
- gmodels, 461
- Hmisc, 86, 110
- lubricate, 281
- MASS, 86, 110, 179
- mice, 110, 179
- mlogit, 444
- multicomp, 405, 422, 428
- multicomp(), 394
- nlme, 110, 179, 422, 428
- pastec, 389, 428
- pastecs, 86, 405, 422
- pastecs(), 394
- polycor, 86, 110
- psych, 86, 110, 179
- qqplotr, 110
- QuantPsyc, 179
- RColorBrewer, 461
- reshape, 422, 428
- rlm, 86, 110
- Snowballc, 461
- wordcloud, 461
- X, 461
- xlsx, 110, 112, 179, 281, 405, 422, 428

R session, 84–87
- Console, 84
- Download and install, 84
- Export data, 90

Import data, 89–90
Install packages, 85–87
Know-thy-data, 90–95
Layered charting, 94
Normal distribution, 92
Pairs panel, 92
Scipt window, 84–85
Working directory, 89
R software, 26–29
Download and install, 26
Layered charting. See Layered Charting
Power BI, 26
Script (AKA, code), 28
Tableau, 26
Windows of, 28
Range, 121
Regression Coefficient defined, 95
Regression coefficients, effects of, 443, 446
Relational database, 230
Relationship insight, 34–35
Linear regression, 34
Logistic regression, 34
Poisson regression, 35
Reshape data, 428
Residual standard error, 187, 190, 192, 197
Right join, 53
Robust difference model, 384
Robust linear regression, 107
R-squared, 187, 192
Schedule cycle, measurement of, 341–51
Accuracy and variance, job plans, 346–49
Flowchart, queries, 347
qryPlanToActualVar, 349
Execution, planned and scheduled, 343–46
ctqStatusCount, 345
Flowchart, queries, 345
Sustainment, confirm, 341–43
Flowchart, queries, 341
Time series, measured by, 349–51
ARIMA (Autoregressive integrated moving average), 350
Holt Winters, 350
Lead-lag relationship, 350
Seven-day cycle, 350
Schedule cycle, stages to, 321–29
Datasets, measures and reports, 327–29
Flowchart, stage, 327
Goals, two, 321
Schedule and craft capacity, baseline, 321–24
Flowchart, stage, 321
Stages, three, 321
Weekly schedule and execution, conduct, 324–26
Flowchart, stage, 325
Schedule cycle, tables to, 329–40, 333–34
Craft classification table, 334–36
Craft hours
tblCraftHrs, 334
Day schedule, 331–33
tblDaySchedule, 332
Super table to measurement, 336–40
Flowchart, queries, 337
qryScheduleVsActual020320, 339
tblSchedVsActual020520, 336
System of tables, 329
Weekly schedule, 330
Schedule measures and methods. *See* Schedule cycle, measures of
Select query, 48, 50
Shapiro-Wilk test, 64, 93, 203
Skew, 121
Solution equation, 442
Spreadsheet compared to data table, 223
SQL in background, 25–26, 79
Standard deviation, 121, 263
Standard error to the mean, 121
Standardized, 451
Standardized effect, 185
Standardized effects of varibles, 186
Standardized residuals, 194
Studentized residuals, 451
Super table with Access, 22–26
Generate, 61–62
Pictorial view of, 22
Process to build, 23
SQL in background, 25–26
Steps to build, 46
Table as format, 22, 222
Survival function, 362, 363
Survival hazard analysis, 38
Systematic component, 443
Table as format, 22, 222
Tableau, 26
Three models, 163
Time series analytics, 241, *See* Time series, principles of, Time series, templates and interpretation, and Time series, R functions
Time series insight, 36–38, *See also* header, Time series, principles of
Autocorrelation, 37
Components, 36

Cycles, 36
Trend, type of, 37
Cross-correlation, 37
Lead-lag relationship, 37
Models, 38
Prediction compared to forecast, 37
Time series, principles of
Aggregate months to year, 284
ARIMA (Autoregressive integrated moving average), 350
ARIMA model (Autoregressive integrated moving average), 311–13
ARIMA(Autoregressive integrated moving average), 298
Autocorrelation, 295–99
Correlogram, 295
Cross correlation, 301
Cycle, 291
Decompose and Holt Winters, compared, 308
Decomposition, 291
Drift in random walk, test for, 317
Exponentially weighted moving average (EWMA), 305
Holt Winters, 305–11, 350
Four-year forecast, 309
Smoothing equation and prediction, 306
Lead, lag relationship, 299–302
Lead-lag relationship, 350
Moving 12-month average, 290
Notation, of time series, 290
Questions of a series, 280
Seven-day cycle, 302–5, 350
Test as deterministic or stochastic, 314–17
Time series object, 282
Trend, 291
Time series, R functions
acf(), 295, 298, 314
aggregate(), 284, 302
auto.arima(), 311
autoplot (), 311
ccf(), 301
class(), 282
data(), 289
Data(), 295
decompose(), 292, 301, 304
diff(), 314, 318
end(), 282, 288, 289
forecast(), 311
frequency(), 282
Holt Winters(), 308
layout(), 314
start(), 282, 288
ts(), 284, 286, 287, 300, 302
ts.intersect(), 287
ts.plot(), 294, 300, 309
window(), 285, 301, 304
Time series, templates and interpretation
ARIMA forecasted 4-years out, 311
ARIMA plot of forecasted mean and limits, 313
ARIMA, table of forecast mean, and upper and lower limits, 312
Comparative correlograms to observed and random series, 301
Correlogram of random series, 314
Correlogram of random series and difference series, 314
Correlogram of random series in decomposition, 295
Correlogram of residuals to an ARIMA model, 298
Correlogram of time series, 295
Extract factors of daily cycle, 304
Extract random series from time series, 301
Extract windows from time series object, 285
Format to time series object, 284
Holt Winters forecast, plot and data table, 310
Holt Winters, 4-year forecast, 309
Holt Winters, smoothing parameters and coefficient, multiplicative, 308
Lead-lag correlogram, multiple datasets, 300
Multiple variables synchronized by date, 287
Plot multiple time series dataset, different axis, 286
Series cycle plotted over trend, 294
Test random walk for drift, 318
Test series as random walk, 316
Time series decomposed, additive and multiplicative, 292
TS object inspect, 282
Weekly pattern, observations, aggregate and day range, 302
Totals row, Access code
Summary Option, 62
Translation tables, 49–50, 164, 264
Trees, 166, 168
Triad of software, 17
T-Score, 70–71, 262
T-Score, table of, 265
Tukey adjustment, 385, 403
Two-mean t-test, principles of, 379, 389–94
Data, long and wide, 390

Data. Descriptive statistics of, 392
Dependent t-test (AKA, paired, 379
Independent t-test, 379
Paired two-mean t-test, 393
Two-mean t-test, R functions
t.test(), 392, 393
Two-mean t-test, templates and interpretation
by() for descriptive statitics, 392
t.test() for significance difference in means, 392
t.test() to test for significant of paired observations, 393
Update query, 50
Variance inflation factor (VIF), 201
Variance reports, 248
Vector, 98
Weibull plot, 373
Work schedule, stages to. *See* Schedule cycle, stages to
Work schedule, tables to. *See* Schedule cycle, tables to
Workload from history, 250
Workload-based budget and variance. *See* Budget and variance, principles and data methodology
Zero-based budgeting, 239
Z-Score, 63–65, 262

Data and Analytics/Operational Effectiveness

Data and Analytics Skills for Your Career Security

Keeping it simple, only the skills you're likely to need

Richard G. Lamb

For those of us who are role holders in enterprise functioning, the personal purpose of acquiring practical working skills in data and analytics is to be able to better do what we already do and find new ways to do better yet. It follows that if you are a role holder who brings and incorporates data and analytics methods in your thoughts and tasks, your career outlook will be more secure and exciting. The book is written to be your gateway to the skills and to be the templates with which you will install the methods in your operational roles.

We all know that the field of data and analytics is huge and intimidating. It is a long slog to becoming comfortable. During the author's own long slog until arriving at the book, something exciting bubbled to the surface. There is a big difference between what we need to know and everything there is to know. We need to know what is possible as insight for decisions and functioning, we need to know how to get to the insight and, finally, we need to be able to interpret the insight. Just as the book does, we can leave the rest to the data scientists.

About the Author: In 2003, Richard Lamb, while struggling to get at the history captured in the databases of operational systems, found the skills to extract datasets of related history and join them in a super table of variables to make possible what was being envisioned for operational effectiveness. In 2014, Richard realized that, with statistical analytics and free enabling powerful pc-level software, an enterprise could ask and answer questions of operational effectiveness that are otherwise not possible. His activism to bring the epiphanies into the careers of role holders in the mainstream of operations has arrived at this book to explain data and analytics through the demonstration of methods.

Richard is a Registered Professional Engineer and Certified Public Accountant. He has previously authored two books: Availability Engineering and Management for Manufacturing Plant Performance, and Maintenance Reinvented for Business Performance. He has a BSCE, BBA and MBA from the University of Houston and a graduate-level Applied Statistics Certificate from the Texas A&M University.

https://analytics4strategy.com/data-and-code

ISBN 978-1-7343947-0-2

90000>

9 781734 394702

Analytics4Strategy