The 7 Steps of Data Analysis:

A Manual for Conducting a Quantitative Research Study

First Edition

WILLIAM M. BANNON, JR.

StatsWhisperer Press

New York

Copyright © 2013 by StatsWhisperer Press.
A Division of William Bannon Associates, Inc.
1669 85th Street, Brooklyn, N.Y. 11214

All rights reserved. No part of this book protected by this copyright notice may be reproduced or utilized in any manner whatsoever or by any means, electronic or mechanical, including photocopying, recording, or by any informational storage and retrieval system, without the permission of the copyright owner, except in the case of brief quotations embodied in critical articles or reviews.

Senior Editor: Tamara A. Bannon
Assistant Editor: Lori Bruno-Lee
Cover Design: Tamara A. Bannon

Library of Congress Cataloging-in-Publication Data

Bannon, Jr., William M.,

The 7 Steps of Data Analysis / William M. Bannon, Jr. –1st ed. p. cm.

Includes bibliographical references

ISBN 978-0-615-85729-9

Library of Congress Control Number 2013949397

1. Steps of Data Analysis (Statistics)—Handbooks, manuals, etc.

I. Authorship I. Bannon, Jr., William M. II. Title

ISBN 978-0-615-85729-9

Printed in the United States of America

Dedication:

To My Wife Tammy and Children Reese, Autumn, Luciana, and Austin
Without Whose Support This Book Would Not Be Possible.

And My Father William and Mother Susanne
Without Whom Life Would Not Be Possible.

Acknowledgments

I would like to extend my thanks to all of the authors, educators, and colleagues who contributed their effort, talent, and good will to the learning experience that helped form the principles presented in this book. In particular, I would like to thank my original research mentor Dr. Chaya Piotrkowski for not only introducing me to the principles of quantitative and qualitative research, but also for illustrating the marvelous things possible through each. Also, I would like to thank my long time mentor Dr. Mary McKay, who worked closely with me for many years as I developed this system through analyzing the rich data she collected through her bevy of studies. I would like to thank Dr. Irwin Epstein for all his superb mentorship over these many years. Lastly, I would like to thank my colleagues at Pace University for their contributions, especially Dr. Joanne Singleton and Dr. Lillie Shortridge-Bagget, who exemplify the cutting edge mentality, intelligence, and work ethic, that move a field forward.

Contents

Preface **xi**

Part One: Introduction **1**

1.1 The Background of this Textbook 1
1.1.1 Who Was this Book Written for? *1*
1.1.2 The Data Analysis Cycle of Inner-Knowledge *2*
1.2 The Key is Knowing the Essentials 3
1.3 Creating a Foundation to Build Upon 4
1.4 Layout of the Text 4
1.5 What is Data Analysis? 5
1.6 The Components of Data Analysis 5
1.6.1 The Cake Recipe & The 7 Steps Of Data Analysis *6*
1.6.2 The Cake Ingredients & The Study Data *7*
1.6.3 The Cooking Utensils & Statistical Tests *8*
1.7 Why Statistics Are Awesome 9
1.7.1 Applying Statistics to Home and Work Life *9*
1.7.2 Curing the Who's The Boss Syndrome *11*
1.8 Applying the Materials 12

Part Two: A Model for Conducting Statistical Research **13**

2.1 What Will This Section Tell Us? 13
2.2 The Evolution of Professional Models: Taking It to the Streets 13
2.3 The Evolution of Statistical Research: Taking It to the Streets 15
2.4 What Elements Should a Model of Data Analysis Reflect? 17
2.5 Shared Elements of Successful Models 19
2.5.1 Fundamentally Sound *19*
2.5.2 Conceptually Clear and Understandable *20*
2.5.3 A Clear Presentation of Ordered Steps *28*
2.6 A Game Plan for Success 33

Part Three: The Data Analysis Concepts You Need To Know **35**

3.1 What Will This Section Tell Us? 35
3.2 Parametric and Non-Parametric Statistics 35

3.3 Descriptive and Inferential Statistics 36
3.4 Level of Variable Measurement 37
3.4.1 Categorical Study Variables *37*
3.4.2 Continuous Study Variables *38*
3.5 Single Item VS Composite Item Scores 39
3.5.1 The Single Item Score *39*
3.5.2 The Composite Item Score *39*
3.6 Study Variable Type 40
3.6.1 Independent Variable *40*
3.6.2 Dependent Variable *41*
3.6.3 Covariate Variable *41*
3.7 The Clarity of the Study Hypothesis 46
3.7.1 Succinct *48*
3.7.2 Clarity *48*
3.8 The Three Dimensions of a Relationship 50
3.8.1 Aspect 1: The significance of the relationship *50*
3.8.2 Aspect 2: The directionality of the relationship *54*
3.8.3 Aspect 3: The magnitude of the relationship *55*
3.9 The Russian Doll Effect in Analysis 58
3.10 Haste or Paste: Save the Syntax 59

Part Four: A Quantitative Study with a Continuous Dependent Variable 61

4.1 What Will this Section Tell Us? 61
4.2 Step 1: Study Map 62
4.2.1 The Study Map in Text *62*
4.2.2 The Study Map in a Diagram *62*
4.3 Step 2: Data Entry 62
4.3.1 Coding the Data *63*
4.3.1.1 The Data Dictionary *63*
4.3.2 Entering the Data *68*
4.3.2.1 Creating the Software Database *68*
4.3.2.2 Entering Survey Data *72*
4.3.3 Cleaning the Data *73*
4.3.3.1 Referencing Survey Hard Copies *74*
4.3.3.2 Examining the Variables *75*
4.3.3.3 Violations in Logic *77*
4.3.3.4 Recoding Variables *79*
4.4 Step 3: Checks of Data Integrity 85
4.4.1 Statistical Power *86*

4.1.1.1 Power Analysis *86*
4.4.2 Test Assumptions *89*
4.4.2.1 Normal Distribution *90*
4.4.2.2 Multicollinearity *119*
4.4.2.3 Homoscedasticity *126*
4.4.2.4 Linearity *133*
4.4.2.5 No Undue Influence of Outlier Scores *138*
4.4.2.6 Other Test Assumptions *142*
4.4.3 Missing Data *144*
4.4.3.1 Defining Missing Data in a Study *145*
4.4.3.2 Amount of Missing Data *148*
4.4.3.3 Patterns of Missing Data *158*
4.4.3.4 Treatment of Missing Data *166*
4.4.4 Measurement Tools *171*
4.4.4.1 Scale Reliability *172*
4.4.4.2 Scale Validity *180*
4.5 Step 4: Univariate Analysis 182
4.5.1 Categorical Variables *184*
4.5.2 Continuous Variables *185*
4.6 Step 5: Bivariate Analysis 187
4.6.1 One-Way ANOVA *188*
4.6.2 Correlation *195*
4.6.3 Independent-Samples T-Test *197*
4.6.4 What Did Bivariate Analysis Tell Us? *202*
4.7 Step 6: Multivariate Analysis 203
4.7.1 Dummy-Coding Variables *204*
4.7.2 Conducting Multiple Linear Regression Analysis *212*
4.7.3 What Did Multivariate Analysis Tell Us? *219*
4.8 Step 7: Write-up & Report 220
4.8.1 Body of the Paper *220*
4.8.1.1 Abstract *221*
4.8.1.2 Introduction *221*
4.8.1.3 Literature Review *221*
4.8.1.4 Methods *223*
4.8.1.5 Results *225*
4.8.1.6 Discussion *225*
4.8.1.7 Conclusion *225*
4.8.2 Sample Manuscript for Sample Study One *226*

Part Five: A Quantitative Study with a Categorical Dependent Variable 241

5.1 What Will this Section Tell Us? 241
5.2 Step 1: Study Map 243
5.2.1 The Study Map in Text 243
5.2.2 The Study Map in a Diagram 243
5.3 Step 2: Data Entry 244
5.3.1 Recoding: Making Happiness Happy 244
5.4 Step 3: Checks of Data Integrity 248
5.4.1 Statistical Power 249
5.4.1.1 Power Analysis 249
5.4.2 Test Assumptions 253
5.4.2.1 Multicollinearity 253
5.4.3 Missing Data 253
5.4.4 Measurement Tools 253
5.5 Step 4: Univariate Analysis 254
5.6 Step 5: Bivariate Analysis 254
5.6.1 Independent Samples T-Test 254
5.6.2 Chi-Square 259
5.6.3 What Did Bivariate Analysis Tell Us? 267
5.7 Step 6: Multivariate Analysis 268
5.7.1 Dummy-Coding Variables 269
5.7.2 Conducting Binary Logistic Regression 270
5.7.3 What Did Multivariate Analysis Tell Us? 277
5.8 Step 7: Write-up & Report 278
5.8.1 Changes to the Body of the Paper 278
5.8.2 The Sample Manuscript for Sample Study Two 279

Part Six: Assessing Published Quantitative Research Studies 293

6.1 What Will This Section Tell Us? 293
6.2 Predetermined Criteria: The Key to Assessing a Quantitative Study 293
6.3 The Mind's Tendency to Make Sense of Findings 295
6.4 Small Errors and Big Problems 296
6.5 You Don't Need To Be a Statistician to Evaluate Statistical Research 299
6.6 A Strengths-Based Perspective 301
6.7 Assessing a Quantitative Study 302
6.7.1 Applying Step 1: Study Map 302
6.7.2 Applying Step 2: Data Entry 305
6.7.3 Applying Step 3: Checks of Data Integrity 305
6.7.3.1 Statistical Power 305

6.7.3.2 Test Assumptions *306*
6.7.3.3 Missing Data *307*
6.7.3.4 Measurement Tools *308*
6.7.4 Applying Step 4: Univariate Analysis *310*
6.7.5 Applying Step 5: Bivariate Analysis *311*
6.7.6 Applying Step 6: Multivariate Analysis *313*
6.7.7 Applying Step 7: Write-up and Report *314*
6.7.8 Summing Up the Published Article *323*

Part Seven: Conclusion **325**

7.1 A Standard Model of Data Analysis 325
7.2 The Next Generation of Researchers 327

Glossary **328**

Appendix **337**

References **345**

Index **348**

Preface

The primary purpose of this book is to empower the user to conduct a legitimate and effective data analysis project (a quantitative research study) from the point of conceiving of a research question, through statistical analysis, and to the completion of the project, in a professionally formatted (APA, AMA) manuscript that presents the study results. This book is also written to illustrate an effective method of interpreting published quantitative research in a manner that will allow the reader to understand, assess, and critique the statistical procedures used in a study. In short, this book is written to enable the reader, who may be a complete stranger to research and statistics, to effectively conduct, read, understand, and intelligently evaluate research studies that employ data analysis methods.

The function of this book is clear and simple. There is no pretense of literary greatness, nor does the book exhibit any unusual scholarship on my part. Rather, it simply presents data analysis in a clear and comprehensible format. Specifically, data analysis is presented as a series of seven logical, simple, and understandable steps. Many learning techniques and examples are presented that make data analysis conceptually accessible and relatable to the reader, while demonstrating that he or she does not need to be intimidated by statistical research.

Moreover, this book does not only describe the process of data analysis, but applies it. Specifically, *The 7 Steps of Data Analysis* model is applied to complete two data analysis studies for two reasons. First, these studies are presented to illustrate the many steps, decisions, and challenges encountered when conducing a data analysis study. Second, these studies act as templates for the reader to follow when implementing his or her own research activities. Primarily, the reader can plug in variables from his or her own research project and follow the process of analysis as illustrated by the studies herein. This book also applies *The 7 Steps of Data Analysis* model in assessing already published quantitative research. Essentially, every effort is made to facilitate the reader in gaining an aptitude in not only understanding these materials, but also in applying them in real world situations.

As you read this book carefully, make an effort to absorb the materials, and persistently practice the principles and formulas within, and you will be amazed at your grasp of statistical procedures and how they are applied together to conduct a data analysis study. How can I be so sure that the practice of these materials will produce

such results? The answer is, for many years I have taught these techniques to graduate students in Master's and Doctoral level statistics and quantitative research courses, as well as a great many private clients, while carefully noting the operation of these methods in their learning and quantitative research skills. I found that the principles presented in this book worked efficiently over a long period of time on a consistent basis. Of course there were revisions over the years, however, the techniques outlined in this book are the finished version of the system.

In my writings, including my regular statistical newsletter *The StatsWhisperer*™, in my Webinar programs, and in lectures presented across the country, I have taught these same scientific yet simple principles of data analysis. Many have read, listened, and practiced these methods to gain proficiency in quantitative analysis. Because several of these individuals have requested that these techniques be put into book form so that the system can be better studied and practiced, I am publishing these materials in the current book titled *The 7 Steps of Data Analysis*. I would like to point out that the techniques contained in this book are not my invention. They are techniques that we, as data analysts, have for years worked hard to cultivate, test, practice and perfect. However, this is the first statistics book that presents these procedures in a practical direct-action step-by-step format as a means of facilitating effective learning and encouraging profound success.

William M. Bannon, Jr.

PART 1

INTRODUCTION

1.1 The Background of this Textbook

While this text is written as a general presentation of the methods needed to produce a quantitative study, it may be helpful to understand the initial motivation for producing the materials. Subsequently, in this section we will briefly describe the needs, ideas, and decisions that inspired this text. We will also describe the components of data analysis, as well as how the materials are organized in this text to teach you this process.

1.1.1 Who Was this Book Written for?

This book was originally written for people who now find themselves in a somewhat peculiar situation, the same one I found myself in many years ago. The situation to which I am referring is graduate school. Often a graduate student enrolls in a program to study a subject other than statistics, but finds that in order to graduate his/her chosen program he/she must learn, understand, and even conduct data analysis. This is a common occurrence in clinically-based advanced degree programs, including programs in medicine, nursing, and social work. Understanding and properly using statistics becomes especially important when the program requires the completion of a thesis or dissertation based on a data analysis project.

Regarding my own situation, I enrolled in a Masters Degree in Social Work (MSW) program with the intention of becoming a clinician. The first semester of this program included a required foundational course in Social Work Research and Statistics. Within

the first month of beginning this course I noticed that I was keenly interested in the subject matter. The professor of the course told me I had been bitten by the "research bug." Soon after realizing that the research bug had bitten me, I noticed that most other students had *not* been bitten, but were generally annoyed by the research bug.

I noticed within the program at large, not every graduate student was getting a great deal of fulfillment or enjoyment from their study of research and statistics. Certainly, most were not grasping how the statistical procedures being taught could be organized toward completing a data analysis study. This left students anxious and frustrated, with a negative attitude toward research and statistics, which then prompted me to ask the question: *How can I make data analysis more understandable, meaningful, and enjoyable for advanced degree students in clinically-based programs?*

1.1.2 The Data Analysis Cycle of Inner-Knowledge

After years of learning data analysis and pondering the experience of my fellow students, I came up with an answer: **Break the Data Analysis Cycle of Inner-Knowledge**. You may wonder not only what the *Data Analysis Cycle of Inner-Knowledge* is, but also how I arrived at it. Let me explain the process. Early on, even though I graduated from a clinically focused program, I decided to become a statistician. Years later, after working on many studies, I realized something extraordinary.

It seemed that every successful statistician and data analyst had to figure out the same necessary steps and procedures needed in order to produce a professional level data analysis study. This deduction was largely achieved by becoming an efficient data analyst myself, through a combination of advice from mentors, learned coursework, and trial and error. Soon after developing this understanding, I recognized that there was a peculiar cycle where this knowledge was kept in a sort of inner circle of statisticians and data analysts. Specifically, what I recognized is that:

Once a statistician or data analyst figures out how to conduct an effective data analysis, he/she largely continues to apply this knowledge in his/her craft, but almost never thinks to share this knowledge with anyone else. Thus, I recognized what I termed *The Data Analysis Cycle of Inner-Knowledge*.

I realized that although not intentional, there was a key piece of information privy to a relatively small circle of data analysts and statisticians, but not to the mass of professionals in the research and practice world. It was as if one had figured out the secret to becoming rich, but was too busy managing and applying his/her money to remember to share the secret with people who were not yet rich.

I also realized that if this knowledge was published and related efficiently, it could help a great many professionals become effective data analysts. Essentially, the years of mentor advice, coursework, and trial and error needed to become an effective data analyst could be delivered to these professionals rather quickly and effectively. Furthermore, the mass of graduate students who view statistics as a conglomeration of confusing procedures would realize that statistical tests are a useful part of a larger process of data analysis. Additionally, those students who need to complete a data analysis project for a thesis or dissertation program requirement will have an instructional roadmap for doing so.

Thus, the goal of this book is to present, review, and share the seemingly arcane stepwise process of data analysis in a new and interesting manner that is easy to understand, and even more importantly, easy to implement. The key to this goal can be expressed in one word: *essentials*.

1.2 The Key is Knowing the Essentials

The key to conducting data analysis effectively and efficiently rests with the realization that *the mind retains a limited number of essential facts*. Perhaps it is just human nature or a quality of the mind, but the brain seems to break large concepts down only to retain a small number of essential facts, no matter what the topic. For example, a car has hundreds of working parts. Most people are unfamiliar with most of them, but they can tell you to use a car you must first insert the key into the ignition and turn it to start the engine. Most people may not be able to describe the structure of the US government, in terms of the number of senators and congressmen therein. However, most people can tell you the US government includes only one president. Even if an individual learned each moving piece of a car and a detailed description of the US government, several years later, he or she would likely have forgotten most of these specifics. However, he or she would likely still remember that you need a key to start a car and that the US government has a single president. Thus, the mind retains a few essential facts upon a topic.

If we accept the premise that the mind will most likely retain a small number of essential rules, then perhaps the most effective method of learning data analysis would be to specify and review the *essential guidelines of data analysis*. Certainly in my experience, knowing these guidelines has been the key to making data analysis understandable, implementable, and enjoyable. Thus, in this text we will identify, review, and apply *The 7 Steps to Data Analysis*, which will provide a foundation for learning, interpreting, and conducting quantitative research.

1.3 Creating a Foundation to Build Upon

Obviously, data analysis is an intricate process that takes years to master. The purpose of this book is not to turn you into a master of data analysis overnight, but to lay a foundation for gaining an ever increasing mastery of the subject. The text is predicated upon the belief that if one develops an essential foundational knowledge of data analysis, one can always add to that knowledge. However, if one does not develop an understanding of the fundamentals, then there is nothing to build upon.

Therefore, this text does not go into great detail about every small nuance regarding each topic mentioned. I have found that this approach makes a text exceedingly dense and overshadows the central message being presented. Instead, this text covers the major topics and concepts associated with data analysis. Readers are certainly encouraged to gain a greater depth of knowledge regarding these topics after completing this text. In short: *I try not to focus too much on the details, so the essentials can be clearly presented.*

One of the biggest challenges in learning data analysis is not that there is an unavailability of materials on a given topic. Today, most topics are covered rather well by online sources. The challenge is that most people are unaware of what topics they need to gather information upon. For example, in order to determine how many participants are needed for a data analysis study, you would conduct a procedure known as a power analysis. If you know this fact, you can identify many resources that will explain this procedure.

However, if you do not know the term for the procedure (i.e., power analysis), you may have difficulty gathering information on the topic. Worse yet, if you are not informed of the procedure and related concepts, you may not even realize that there is a requisite number of study participants for a data analysis study. Through presenting the major facets and facts of data analysis, this text will provide you with the essential terms, concepts, and basics you need to know to understand and/or conduct a data analysis study. Furthermore, this book was printed with wide margins to give the reader the opportunity to enter by hand any important information pertinent to their learning and research activities that may not be mentioned in the text.

1.4 Layout of the Text

You might notice that the layout of this book is dramatically different from any other statistical textbook, which is done intentionally. The materials are laid out in such a way that they can be applied sequentially as you conduct a data analysis study. Along

with the conceptual information, the text includes instruction on how to conduct each procedure in statistical software, as well as how to effectively interpret the statistical output. The statistical software program SPSS is used, as that seems to be the most popular and widely used data analysis program.

I have attended many lectures where I felt the information presented was spectacular and I was excited to apply the information once I got back to my desk. However, I often found that once I sat down to apply the newly learned information, I could not figure out how to do so. Subsequently in this text, information is not only presented, but laid out in a manner for direct application.

1.5 What is Data Analysis?

This book has been written with the intention of instructing the reader on how to learn, understand, interpret, and conduct data analysis. That may prompt you to ask, what is data analysis? If you ask the average person what data analysis is, you are likely to get the response "statistics." While statistics are the primary tool used in data analysis, statistical tests and procedures actually only comprise a segment of data analysis. Essentially, data analysis is applying a series of statistical tests and procedures in a specific stepwise progression in order to examine a dataset. In other words, data analysis is taking a set of tools (statistical tests/procedures) that when applied in a certain order (*The 7 Steps of Data Analysis*) reveal the message(s), lesson(s), and answer(s) the data have to tell us. However, the foundational nature of this textbook suggests data analysis might be foreign to some readers. Therefore, perhaps a more effective discussion might focus on how the components of data analysis may be similar to another process already familiar to the reader, such as making a cake.

1.6 The Components of Data Analysis

In short, the components involved in making a cake are almost identical to those in conducting a data analysis study. Therefore, to make the lesson easier, we will present how the components of making a cake are related to the components of conducting a data analysis study. My hope is that this parallel will make the concept of data analysis much more "digestible." Subsequently, let's first talk about the components of making a cake. Just to keep it simple, let's say you need three things to make a cake. Specifically, you need the:

1) Cake Recipe: An *outline of the steps to follow* toward making that cake, which is simply a "to do" list presenting a series of steps that must be taken in a precise order.

2) Cake Ingredients: The *materials* necessary to produce the cake.

3) Cooking Utensils: The *tools* to transform the ingredients into the product (cake).

To conduct a data analysis, you need the equivalent three things, which are the:

1) 7 Steps of Data Analysis: An *outline of the steps to follow* toward conducting a data analysis study. Like a recipe, I suggest *The 7 Steps of Data Analysis* is also simply a "to do" list presenting a series of steps in a precise order (a **recipe** for a study).

2) Study Data: The *materials* necessary to produce a data analysis study (**ingredients** for a study).

3) Statistical Tests and Procedures: The *tools* to transform the data into the product, which is the completed data analysis study (**utensils** for a study).

We will clarify these relationships and parallels in the next part of this section.

1.6.1 The Cake Recipe & The 7 Steps Of Data Analysis

As we suggested, you could think of a completed data analysis study like a completed baked good, like a cake. The data analysis study and cake are both finished products that were created by following a "to do" list. For example, in the box labeled **To Do List: Making a Cake** within **Figure 1.1**, the **Cake Recipe** is presented, which is a list of things that must be done to produce the product (a cake). In the box labeled **To Do List: Data Analysis Study**, **The 7 Steps of Data Analysis** is presented, which is a list of things that must be done to produce the final product (a quantitative study). Within **Figure 1.1**, you will see a side by side comparison of how the **Cake Recipe** and **The 7 Steps of Data Analysis** are each a list of steps that must be followed and completed in a certain order to produce the respective final product.

Figure 1.1 The *to do* lists: The cake recipe and The 7 Steps of Data Analysis model

TO DO LIST: MAKING A CAKE

Cake Recipe

1) Preheat oven to 350 degrees
2) Grease and flour a 9x9 inch pan
3) Cream together sugar and butter
4) Beat in the eggs
5) Stir in the vanilla
6) Add flour and baking powder
7) Stir in the milk until smooth
8) Bake for 30 to 40 minutes

TO DO LIST: DATA ANALYSIS STUDY

The 7 Steps of Data Analysis

1) Create a Study Map
2) Data Entry
3) Check Data Integrity
4) Univariate Analysis
5) Bivariate Analysis
6) Multivariate Analysis
7) Write-up & Report

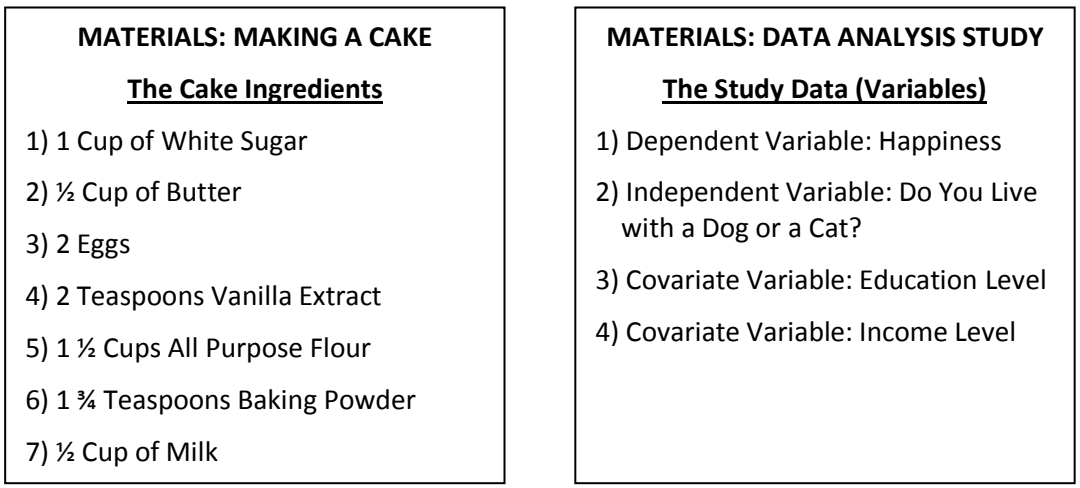
Of course each step within **The 7 Steps of Data Analysis** model involves many other steps and considerations. In other words, each of the seven steps is a representation of many other smaller and more detailed steps. However, every data analysis study can be approached by categorizing these smaller and more detailed steps into this *to do* style list toward completing an efficient statistical analysis.

1.6.2 The Cake Ingredients & The Study Data

The cake is created using materials, known as ingredients. The word "ingredients" is really just a catch-all phrase. The materials known as ingredients have specific names and qualities, such as butter and flour. Furthermore, different cakes require different types, amounts, and combinations of these materials. In other words, each cake has specific needs regarding the necessary ingredients used to make that cake. For example, within **Figure 1.2** the box labeled **Materials: Making a cake**, describes the necessary ingredients for making this specific cake, such as butter, eggs, sugar, etc. However, a different cake recipe might call for replacing these ingredients with alternatives.

The data analysis study is conducted using materials, known as data. The word "data" is really just a catch-all phrase. The materials known as data are structured as variables **(**for a fuller discussion of variables see section **3.4 Study Variable Type)**. Different quantitative studies require different types, amounts, and combinations of these variables. For example, as listed in the box labeled **Materials: Data Analysis Study** within **Figure 1.2**, a study might include one dependent variable (*Happiness*), one independent variable (*Do You Live with a Dog or a Cat?*), and two covariate variables (*Education Level* and *Income Level*).

Figure 1.2 The needed project materials: Cake ingredients and study data (variables)



However, like cake ingredients, these factors will vary from project to project. For example, another study may require more or less independent or covariate variables. Additionally, a study may require different types of variables. For example, the variable *Happiness* might be replaced with the variable *Depression* as the dependent variable incorporated in the study.

1.6.3 The Cooking Utensils & Statistical Tests

After the cake recipe and ingredients have been selected, specific tools known as cooking utensils are used to relate the ingredients to one another. Relating the ingredients successfully will result in a sound final product (a cake). The rules for relating the ingredients are clear. For example, within the box labeled **Tools Relating: Ingredients** in **Figure 1.3**, the first line instructs to relate (i.e., *cream*) the ingredients *sugar and butter* you would use a cooking utensil known as a *Hand Mixer*. The next line informs to relate (i.e., *beat*) the ingredient *Eggs Into The Overall Mixture*, you would use a cooking utensil know as a *Whisk*. Of course, there are many other types of cooking utensils not mentioned, but you get the idea.

Simply put, just as there are cooking utensils to relate the cake ingredients, there are statistical tests to relate study variables to one another. For example, the first line within the box labeled **Tools Relating: Study Variables** in **Figure 1.3**, describes to relate the variables *Happiness* and *Education Level*, you would use a statistical test known as an *ANOVA*. The second line describes to relate the study variables *Happiness* and *Do You Live With a Dog or a Cat?* you would use a statistical test known as a *T-test*. Relating the variables with the correct statistical test will result in a sound final product (a quantitative study).

Figure 1.3 The needed project tools: Cooking utensils and statistical tests

TOOLS RELATING: INGREDIENTS

Ingredients Related	Cooking Utensil Used
Cream Together Sugar And Butter	Hand Mixer
Beat The Eggs Into The Mixture	Whisk
Stir Vanilla In With Other Ingredients	Large Spoon

TOOLS RELATING: STUDY VARIABLES

Study Variables Related	Statistical Test Used
Happiness & Education Level	ANOVA
Happiness & Do You Live With a Dog or a Cat?	T-test
Happiness & Income Level	Correlation

1.7 Why Statistics Are Awesome

In the last section, we described how statistical tests and procedures are the primary tools used in data analysis. But, just what are statistics and why are they awesome, as the title of this section suggests? The response to both questions is: statistics are tools that help us **understand the reality around us in a way we could not otherwise!** Statistics offer us a means of unearthing knowledge about the world we live in that otherwise would remain hidden. Specifically, statistical procedures are methods of measuring the world around us as variables (e.g., *Happiness* or *Education Level*) and then testing the association(s) between those measured variables (e.g., *Is Happiness Associated With Your Highest Level Of Education?*). These measured variables are largely undetectable by the five senses of touch, taste, smell, sight, or hearing. However, we can measure these variables empirically by producing a portrait of these unseen aspects of the world around us via statistical methods.

For example, let's say you are among a crowd of 100 people at a restaurant. While you are waiting for a table, you begin to wonder if among this group, the people *With Graduate School Degrees* have a higher level of *Happiness* relative to all other patrons (those who *Do Not Have A Graduate School Degree*). Could you use your sense of touch, taste, smell, sight, or hearing to effectively observe if the relationship between these variables exists? Not really, as you cannot touch, see, hear, taste, or smell one's *Level of Education* or *Happiness*. However, if you provided each table with an empirical measure of *Happiness* and level of *Education* you certainly could identify these factors! In this way statistics are like a sixth sense!

1.7.1 Applying Statistics to Home and Work Life

Statistics can also be used to understand the reality of your home life, as well as your work life. Taking your home life as an example first, let's say you are married and you notice that each time you compliment your spouse in the morning he/she cleans the house that evening. Soon you get tired of wondering if there is an association between this complimenting and cleaning scenario, so you begin to gather data toward producing a better picture of what is really going on. Over the next 100 days you compliment your spouse once in the morning for 50 days and do not compliment him/her on the other 50 days. You then count the number of times your spouse has cleaned the house on the days that you did and did not compliment him/ her.

You table these data in a manner similar to **Table 1.1**, which describes on the mornings you gave a compliment to your spouse he/she cleaned the house 80% of the time. On the mornings you did not give a compliment, your spouse cleaned the house 5% of the

time. Now you have actual values, or statistics, that better reflect the reality of your home life. The data reflect that your spouse seemed to clean the house on a much higher percentage of the days you paid them a compliment (80%) relative to the days you did not (5%), which suggest a significant relationship between the two occurrences.

Table 1.1 The relationship between *gave a compliment* and *cleaned the house*

		Cleaned the House		
		Yes	**No**	**Total**
Gave a Compliment	**Yes**	80%	20%	100%
	No	5%	95%	100%

An example like this could just as easily be generalized into one's professional role. For example, if you were a nurse and had a patient that did not like to take his/her daily medication, you could set up the same experiment to determine if your complimenting the patient had an association with the percentage of time he/she took his/her medications (please see **Table 1.2** below).

Table 1.2 The relationship between *gave a compliment* and *patient took medication*

		Patient Took Medication		
		Yes	**No**	**Total**
Gave a Compliment	**Yes**	80%	20%	100%
	No	5%	95%	100%

Again, these same numbers indicate the patient took his or her medication on a much higher percentage of the days you paid them a compliment (80%) relative to the days you did not pay them a compliment (5%). This would again suggest a relationship between complimenting the patient and his or her taking medication. So you can see that without exaggeration, statistics allow us to see reality in a deeper and more meaningful way than we could otherwise.

Many of us have heard and agree with the statement by Socrates "The unexamined life is not worth living." However, while the statement is widely circulated, the knowledge of *how* to examine one's life is not, which may be a cause of concern not only in your personal life, but your professional life too. For example, suppose you are a physician who believes it is not worth treating patients unless you can examine how well your treatments work. In that case, determining *how* to examine this facet of your professional life would be a critical challenge. Statistics and data analysis are one of the few methods available that can provide this *how*.

1.7.2 Curing the *Who's The Boss* Syndrome

Meaningful research questions often surround us, but frequently these important potential areas of study are left unaddressed. To remind us that these topics of study exist, we might benefit from a catchy phrase that may potentially stick in our minds to help us keep on the lookout for these unaddressed research questions. Therefore, I have ascribed these potential areas of study to be part of the *Who's The Boss* Syndrome.

In the 1980s, the sitcom *Who's the Boss* (1984-1992) debuted. The program aired almost every Thursday night for eight years. Each time the program aired, the opening credits would present the question *Who's the Boss?* as the title of the show was revealed. This question was presented weekly for the eight year duration of the show (many more times if you count syndication). However, when the show ended in 1992, the viewer was never told of the cast characters, just who the boss was. Thus, the same question was posed for nearly a decade, but never answered. In fact, many viewers seemed to even forget a question was being asked and did not expect an answer. Although this development may not seem very consequential, this very thing often occurs in many professional areas of study and must be diligently guarded against.

The *Who's the Boss Syndrome* occurs when we are faced with important questions regularly (e.g., every Thursday night on the ABC television network) that seem to perpetually go unaddressed and unanswered. Some of these questions might be light, but some of these questions are profound and essential. For example, a certain medical treatment might be widely used within a certain population for decades, but the effectiveness of this treatment might never have been examined. Thus, there may be a perpetual question of *Is this treatment effective?* that might be pondered year after year without ever being answered. This would be the more serious side of the *Who's the Boss Syndrome*.

In light of such situations, we might suggest that rolling questions are generally not acceptable. Specifically, if an important question exists, we must find an effective means of answering it. In other words, perhaps it is not permissible to continuously ask the same important question regularly and not produce a meaningful answer. I found that statistics and data analysis are a superior means of producing meaningful answers to such questions. Subsequently, a viable solution to the *Who's The Boss* Syndrome is applying statistics and data analysis toward answering the important questions with which we are faced.

1.8 Applying the Materials

Learning new and interesting information can be a worthwhile pleasure. However, most people have had the experience of learning seemingly useful information, only to discover that he or she can't figure out how to apply it! It seems that at some point we all realize that there is often a huge gap between *knowing great information* and *knowing how to apply that great information*. Therefore, in this textbook, we will move beyond the presentation of knowledge and into the application of knowledge! Specifically, we will apply *The 7 Steps of Data Analysis* in three instances.

First, in **Part 4**, we will conduct sample study one, which is a quantitative research study examining a continuous dependent variable. Specifically, sample study one examines if the continuous dependent variable *Happiness* is related to the independent variable *Do You Live with a Dog or a Cat?*, as well as the covariate variables *Education Level* and *Income Level* at a statistically significant level. Next, in **Part 5**, we will conduct sample study two, which is a quantitative research study examining a dichotomous (a categorical variable with two response categories) dependent variable. Specifically, sample study two examines if the dichotomous dependent variable *Happy* (Yes or No) is related to the independent variable *Do You Live with a Dog or a Cat?*, as well as the covariate variables *Education Level* and *Income Level* at a statistically significant level. Lastly, in **Part 6**, *The 7 Steps of Data Analysis* model is applied to assessing the quality of a published quantitative research study.

Please note that **Parts 4**, **5**, and **6** are meant to serve as templates, which can be generalized to data analysis studies of one's own. For example, if an individual had a data analysis study to complete that incorporated a continuous dependent variable, he or she could plug in that variable within **Part 4** (our data analysis study examining a continuous dependent variable) and follow all the steps thereafter toward completing his or her own study.

However, before applying *The 7 Steps of Data Analysis*, it is essential to discuss how these steps compose a needed and timely model of data analysis. Therefore, in **Part 2** we will discuss the concept and utility of a model of data analysis, as well as the qualities such a model should reflect. Additionally, in **Part 3**, we will share some fundamental concepts that the reader should be aware of prior to approaching data analysis.