# THE STATSWHISPERER

The StatsWhisperer Newsletter is published by staff at StatsWhisperer.™ Visit us at: www.statswhisperer.com

## Introduction to this Issue

Data analysis has a clear and logical process. In other words, there is a step 1, step 2, step 3, etc., regarding what needs to be done to produce a clear and comprehensive data analysis. This may not make immediate sense to many people as this process is rarely discussed definitely. Typically, our learning focuses on certain aspects of the data analysis process (e.g., learning specific statistical tests like t-tests/chi-squares). However, we are often not provided with a description of how these pieces fit together to create the overall process of analysis.

To address this issue, we will discuss the 6-steps involved in the process of data analysis in the current newsletter issue. Additionally, because StatsWhisperer strives to make learning clear and engaging, we will parallel the process of data analysis with another process readers might be more familiar with, fixing people up on dates.

Thus, through this newsletter you will see that if you can understand the process of fixing people up on dates (which we call dater analysis), you can understand the process of relating variables to one another (which we call data analysis).

These materials are from our new seminar and book "The Logistics of Statistics: The Process of Data Analysis." It is our pleasure to provide this newsletter as the book is on pre-order and many will not make the seminar. Subsequently, it is our intention that this newsletter will provide at least a glimpse into the process of data analysis.

## You Have Learned the Materials, Now What?

For those who have seen the film *Finding Nemo*, for our purposes here, you might recall the final scene (of course, please note a spoiler alert here).

A group of fish, a starfish, and a shrimp, had successfully implemented a clever escape plan where they managed to dirty their tank, be put into baggies outside of their fish tank, and then roll themselves out a window into the harbor.

When the final member rolled her baggie into the harbor to join the rest, the group cheered at their

success, as they bobbed up and down in the harbor still sealed in the baggies. After the cheers subsided and they realized they did not have a plan to get out of the baggies, one fish commented "Now what?"

Many of us in graduate school that have completed our coursework on research and statistics and then sit down to compose a data analysis project know this feeling well. For example, as doctoral students, we learn about hypotheses, statistical

# You Have Learned the Materials, Now What? (continued)

tests (chi-squares, correlations, regression, etc.), how to write-up results in APA or AMA format, and so on. Many feel as if they have learned a great deal and have a firm grasp on the subject matter. But the minute they sit down to translate everything they learned into a dissertation study they also get that same felling of "Now what?"

The truth is that learning the materials is not enough. You also need to know how to organize what you learned before you can apply it! The best tool I know of to organize what you have learned so you can translate the information into a study is to apply the process of data analysis.

The process of data analysis consists of 6 steps. Every research method and statistical procedure has its place within one of the steps.

But if you don't know the process, you don't know how each piece relates to one another to function in concert to create a data analysis study. For example, what of the statistical test known as the t-test? Is it incorporated in step 1? Is it incorporated in step 2? You cannot wonder you must know!

Thus, let's respond to the question "now what?" "Now" you must learn and follow the process of data analysis!

# The Process of Data vs. Dater Analysis

Before you learn something new, it may benefit you to think about how something you already know is like the new topic you are learning.

For example, many people are familiar with computer dating programs such as e-harmony, as well as reality dating shown such as The Bachelor or The Bachelorette. Even if you have no exposure to either, when each is explained the concepts are not difficult to grasp.

Let's take the e-harmony dating service. What does this service do? It implements steps involved in a process that is ultimately supposed to identify the strongest relationship between two people looking for a date.

This by the way is why we call this process dater analysis. You have all these potential daters and you are doing an analysis to determine which dater
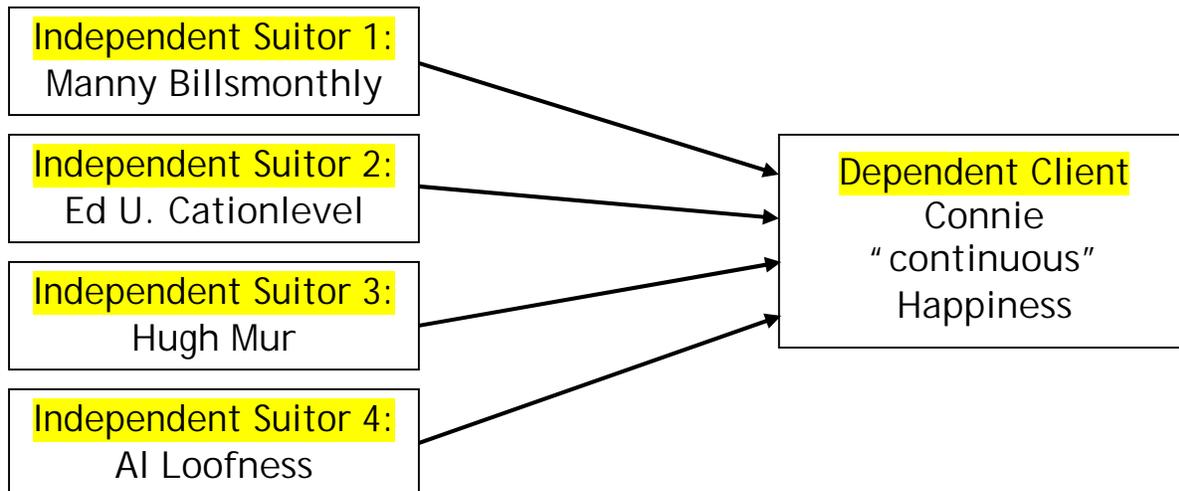
has the strongest relationship with another, so you can recommend them to one another as a match.

In our seminar, our computer dating recognizes that we have a number of clients dependent on our matching them up with the right suitor for a date. Now the suitors are not clients of our agency, they are people who need dates that sent us in their information toward getting a date. Thus, the suitors are independent of our agency, so we call them independent suitors. Our clients are dependent on our finding the right people for them, so we call them dependent clients.

In end, dater analysis involves identifying which of a number of independent suitors has the strongest relationship with a dependent client.

# The Process of Data vs. Dater Analysis (continued)

## Figure 1. Dater Analysis Outline (Goal=define relationship arrows)

Independent Suitor 1:
Manny Billsmonthly

Independent Suitor 2:
Ed U. Cationlevel

Independent Suitor 3:
Hugh Mur

Independent Suitor 4:
Al Loofness

Dependent Client
Connie
"continuous"
Happiness

Again, ultimately dater analysis involves identifying which of a number of independent suitors has the strongest relationship with a dependent client. Figure 1 presents the outline for an example of a dater analysis study.

We see on the left hand side that this dater analysis study involves our examining which of four independent suitors (**Manny Billsmonthly, Ed U. Cationlevel, Hugh Mur, and Al Loofness**) has the strongest relationship with the dependent client (**Connnie "continuous" Happiness**).

Put plainly, **Connie "continuous" Happiness** needs a date. It is your job as a computer dating agent to figure out via computer analysis which independent suitor has the strongest relationship with her, so you can send them out together.

How will you do this? You will follow the 6-step process of dater analysis! If you follow this 6-step process, you can efficiently identify which of these suitors has the strongest relationship with Connie and should go out on a date with her.

IMPORTANT: The arrows in Figure 1 that stretch from each independent suitor to the dependent client represent relationships. Thus, I like to refer to them as "relationship arrows."
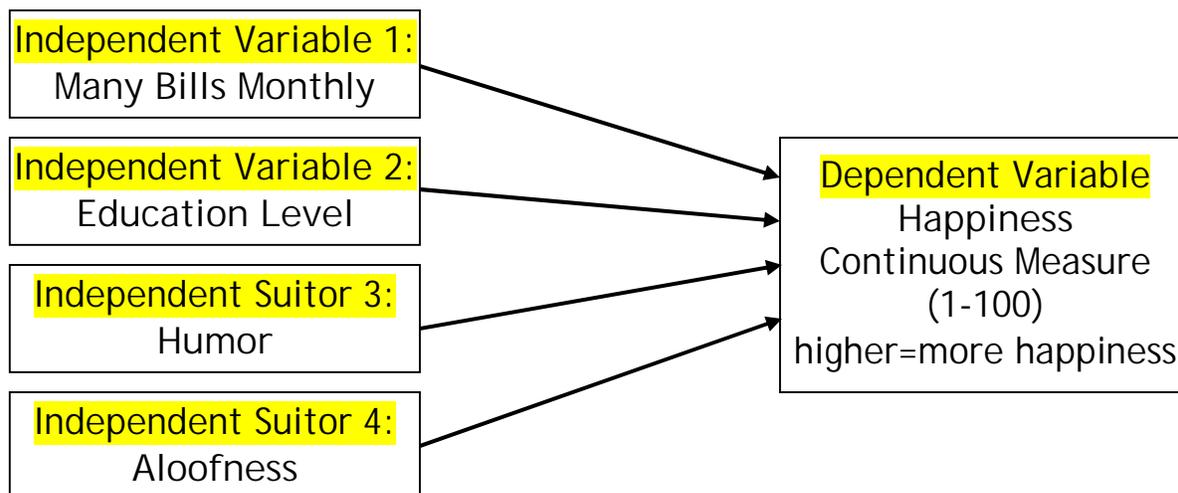
For example, the arrow between independent suitor 1 and the dependent client asks, "What is the relationship between **Manny Billsmonthly** and **Connie "continuous" Happiness**?"

These relationships are defined via analysis on a computer software program before they ever meet. That is your job as a dater analyst!

The ultimate aim of dater analysis is to define the relationship arrows between the independent suitors and the dependent client.

## The Process of Data vs. Dater Analysis (continued)

### Figure 2. Data Analysis Outline (Goal=define relationship arrows)

| Independent Variable 1: Many Bills Monthly | |
| Independent Variable 2: Education Level | Dependent Variable Happiness Continuous Measure (1-100) higher=more happiness |
| Independent Suitor 3: Humor | |
| Independent Suitor 4: Aloofness | |

Similar to dater analysis, data analysis often involves identifying a strongest relationship. However, data analysis involves identifying which of a number of independent variables has the strongest relationship with a dependent variable. Figure 2 presents the outline for an example of a data analysis study.

We see on the left hand side that this data analysis study involves our examining which of four independent variables (**Many Bills Monthly, Education Level, Humor, and Aloofness**) has the strongest relationship with the dependent variable (**Happiness** measured on a continuous 1-100 scale where higher scores indicating more happiness).

Put plainly, our continuous measure of **Happiness** will be impacted by one of our four independent variables more than all the others. That independent variable will be the one the strongest relationship with **Happiness**. Determining which independent variable has the strongest relationship with the dependent variable is the goal of our data analysis.

Now, how will you determine which independent variable has the strongest relationship with the dependent variable? Simple! You will follow the 6-step process of data analysis!

Just as in dater analysis, our concern is defining the relationship arrows. Again, the arrows in Figure 2 that stretch from each independent variable to the dependent variable represent relationships.

For example, the line between independent variable 1 and the dependent variable asks, "What is the relationship between having **Many Bills Monthly** and your level of **Happiness** in life?"

These relationships are defined via analysis on a computer software program. That is your job as a data analyst! The ultimate aim of this data analysis is to define the relationship arrows between the independent variables and the dependent variable.

# The 6 Step Process of Analysis

Figure 3 below presents the 6-step process of analysis. We will look at them here step by step.
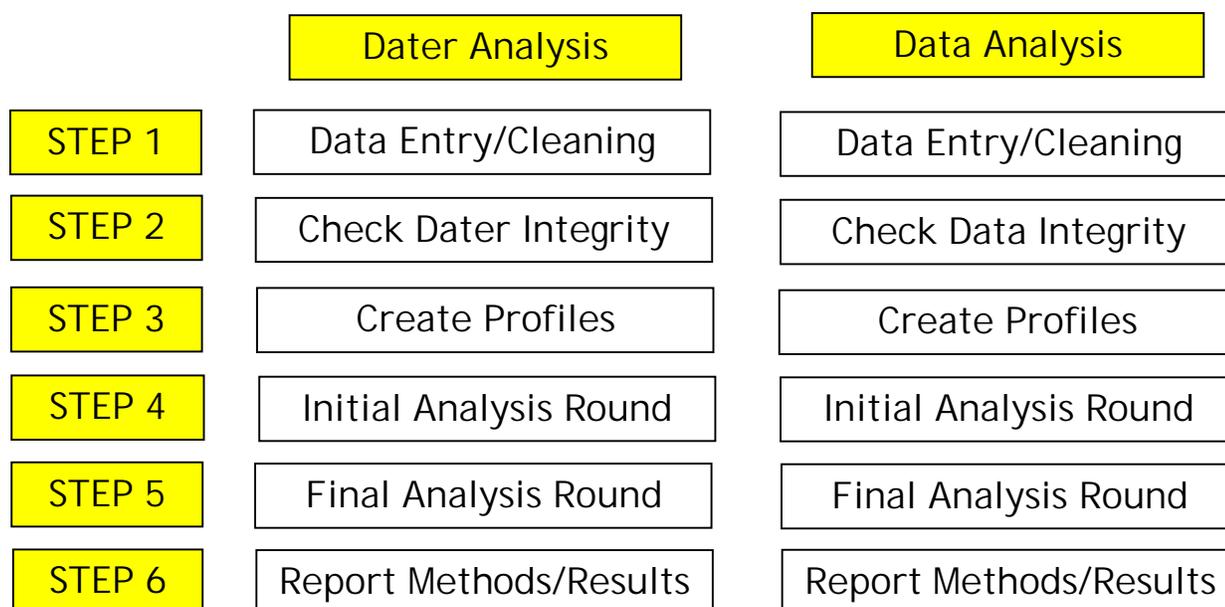
### Step 1:  Data Entry/Cleaning

Of course for both dater and data analysis, the first step in to enter and clean the data. For dater analysis, this is similar to the process employed by e-harmony.  The first act of an e-harmony client is to log onto the e-harmony site enter personal information (i.e., personal data). The data entered are stored in a database for analysis.

Cleaning the data includes making sure data are entered correctly. This is important! What if a client is looking for one sex, but enters another. Surely, their experience may not go as planned.

Regarding data analysis the procedure is quite similar. Data are entered into a database for analysis. Many times the data are entered from physical paper surveys into the database. Nowadays, data analysis study data are also entered by study participants online (just like e-harmony) and then downloaded into a database.

These data are also cleaned to make sure the data are entered correctly. In both cases, cleaning the data also includes looking for missing data. Missing data is a huge issue. For example, imagine if you are examining levels of happiness by education level, but the person's level of education is not reported (i.e., the data are missing). Your analysis and peace of mind will likely suffer!

# Figure 3. The 6-Step Process of Analysis

| | Dater Analysis | Data Analysis |
|---|---|---|
| STEP 1 | Data Entry/Cleaning | Data Entry/Cleaning |
| STEP 2 | Check Dater Integrity | Check Data Integrity |
| STEP 3 | Create Profiles | Create Profiles |
| STEP 4 | Initial Analysis Round | Initial Analysis Round |
| STEP 5 | Final Analysis Round | Final Analysis Round |
| STEP 6 | Report Methods/Results | Report Methods/Results |

# The 6 Step Process of Analysis (continued)

### Step 2: Checking Dater/Data Integrity

In dater analysis, checking dater integrity just makes sense! You need to check the integrity of daters to be sure they are appropriate for dater analysis. If you set two daters up who each have questionable integrity, they may go out on a date only to trash a hotel room and tell the press they met through your agency! No good!

In data analysis, checking data integrity also means examining the data to be sure they are appropriate for analysis. However, these checks for data integrity often involve checks to be sure the data meet the assumptions of statistical tests. For example, you may need to examine the distribution of scores to be sure they approximate a normal distribution.

### Step 3: Create Profiles

Of course in dater analysis, once you have gathered a dater's information (Step 1) and made sure their level of integrity makes them appropriate for dater analysis (step 2), you need to create a profile for them. In other words, you report their qualities as an individual. What type of music do they like? What are they looking for in another dater?

In data analysis, you create profiles here of individual variables (instead of individual daters). For example, if we are looking at the variable **education level**, we would describe how many people in the sample have a highest level of education at the High School/GED level, at the college graduate level, and beyond.

### Step 4: Initial Analysis Round

For both dater and data analysis, the initial analysis round involves determining which relationships are significant. If a relationship is significant, it is carried forward into Step 5 (Final Analysis). If it is not significant, it is eliminated from consideration at this step.

Again, in both dater and data analysis, we are looking for the strongest relationship. In Step 4 (Initial Analysis Round), we determine which relationships are significant and in Step 5 (Final Analysis) we determine of the significant relationships, which is the strongest.

Regarding dater analysis, this may sound similar to the reality show The Bachelorette. What is the process there? Well, the central figure (the bachelorette, who in your analysis would be your client) starts off with about 50 suitors.

Then there is an initial elimination process where she does what? She eliminates the suitors that she does not consider significant to her. She then carries the significant suitors forward to final analysis (Step 5), where she identifies which has the strongest relationship to her. I believe she gives that one a rose?

This is almost identical to the process in dater and data analysis, but the process of indicating a significant relationship is done through testing on computer software, which I believe differs from method used on The Bachelorette show. But what was the bachelorette doing in this show? She was defining relationship arrows between her and the suitors.

# The 6 Step Process of Analysis (continued)

### Step 4: Initial Analysis Round (continued)

Now in the initial round of analysis, the relationship arrows are defined on a one-one-one level.

For dater analysis, we are looking at if an independent suitor has a significant relationship with a dependent client on a one-on-one basis. This is done with a series of bi-dater (a test between 2 daters) statistical tests, such as chi-squares, t-tests, ANOVAS, and correlations. A significant relationship is indicated by a probability level below .05 (p<.05).

If the relationship is significant, the independent suitor is included in Step 5: Final Analysis. If the relationship is not significant, the independent suitor is eliminated from analysis here.

For data analysis, we are looking at if an independent variable has a significant relationship with a dependent variable on a one-on-one basis. This is done with a series of bivariate (a test between 2 variabels) statistical tests, such as chi-squares, t-tests, ANOVAS, and correlations. A significant relationship is indicated by a probability level below .05 (p<.05).

If the relationship is significant, the independent variable is included in Step 5: Final Analysis. If the relationship is not significant, the independent variable is eliminated from analysis here.

### Step 5: Final Analysis Round

For both dater and data analysis, the final analysis round involves considering of all the significant relationships identified in Step 4, which has the strongest relationship with the dependent client

or the dependent variable.

Again, this is similar to The Bachelorette show, where in the final analysis she is alone in the room with just a few men (presumable the most significant of the initial group) and then names one as having the strongest relationship with her.

Again, in dater and data analysis, this decision is made through statistical testing, not through an emotional inkling, as it seems to be on the referenced television show.

In data analysis, Step 5: Final Analysis is often done through the statistical procedure known as linear regression (we use linear regression here because the dependent variable **happiness** in this instance is continuous). Regression is a multivariate test, as you can consider the relationship between multiple (multi) independent variables with one dependent variable at once. Then the statistical procedure will indicate which independent variable, among a group of independent variables, has the strongest relationship with the dependent variable.

A like procedure is used for dater analysis at this level.

### Step 6: Report Methods/Results

There is a very precise manner you must report your findings, as well as the methods used to produce them via tables and text. This is the final step of the data analysis process. Once this is complete, you will have a finished analysis project! Congratulations!

# Final Comments

Of course, what is presented in this newsletter is a remarkably superficial description of the process of data analysis.

To boot, it is largely conceptual. In this newsletter we largely discussed how to think about the process of data analysis rather than the specific tests and methods used to conduct each step.

In the seminar and book we present not only what needs to be done, but actually how to do each step. For example, we describe what statistical tests to use within each step and how to select the right ones based on the daters/variables being considered. We even give a description of how to write-up the methods and results at step 6.

In the seminar and book we give you materials that are designed to act as a template for conducting a data analysis project. Specifically, if you have a data analysis study, you can plug in your variables in the studies we present, toward completing your own study.

One of the most inconvenient facts about conducting a data analysis study is that you need to cross-reference many books and sources to complete the project. What we have done, is to centralize all the facts and information in one source to minimize the amount of cross-referencing.

Our goal is to provide a guide that you can follow that tells you the steps that need to be taken from the very beginning to the very end of a data analysis project. A guide that can lead you from the very point of data entry, through statistical analysis, and finishing the write-up, without ever having to get up and reference another source.

We would favor providing more information toward that in this publication. However, there is just not enough room for that sort of detail in a newsletter. However, perhaps there will be some value in what we have gone over.

The truth is that not only do most people not realize what the process of data analysis is, but most do not realize that such a process even exists. However, knowing the process of data analysis is an absolute necessity to being an effective researcher.

Furthermore, I have noticed that knowing this process often makes learning the individual parts of the process clearer and more meaningful. And that's a good thing no matter how you slice it!

Feel free to visit statswhisperer on the web at:

**www.statswhisperer.com**

Dr. William M. O'Bannon, Jr., the founder and CEO can be contacted at:

**wb@statswhisperer.com**

Information on our upcoming seminars can be viewed at:

**http://www.statswhisperer.com/seminars/**

Thanks for your interest in our newsletter!