

THE STATS WHISPERER

The StatsWhisperer Newsletter is published by staff at StatsWhisperer.™ For many more free resources in learning statistics, including webinars and subscribing to this newsletter, visit us on the web at: www.statswhisperer.com

Introduction to this Issue: How to Think Like a Statistician

This current newsletter issue will reveal a secret to learning and conducting data analysis that is very rarely presented. This newsletter issue will discuss how to think like a statistician.

Mastery in any field involves not only learning tools and methods, but also developing the mindset of the master in applying them. One could suggest that it might be ideal to begin to develop the mindset of the master while learning the tools of the trade. This line of thinking is particularly germane to data analysis. The general perception of how data analysis is conducted is radically different from the actual approach. This disparity is the source of a great deal of confusion and frustration.

Subsequently, we will discuss what one should expect when conducting a data analysis project, as well as how to navigate the process. We will

Data Analysis is an Art, as Well as a Science

Before being exposed to learning or conducting data analysis, many individuals regard the process as pure science. This is understandable, as data analysis certainly has many scientifically based steps, rules, and procedures that must be observed.

However, if one approaches data analysis as a pure science, one will quickly become stuck somewhere along the process of analysis. Why? Because within data analysis there are many areas where there are soft rules of few guidelines at all.

INSIDE THIS ISSUE

Introduction to this Issue: How to Think Like a Statistician	1
Data Analysis in an Art, as Well as a Science	1
How Data Analysis is A-Maze-ing	4
Final Comments	5

also draw attention to the difference between what is typically thought about data analysis before and after one is exposed to the process.

This piece is part of our new book, *The Steps of Data Analysis*. This type of inside information, along with guidance on the practical application of knowledge, is a hallmark of the text.

You can reserve a 1st edition printing of the text at:

<http://www.statswhisperer.com/products/>

Electronic copies of the text will be made available in August and printed copies in September.

Many people believe data analysis incorporates hard rules that will always guide one's decisions. This is not so. Thus, it is important to realize that **sometimes in data analysis there is no right or wrong answer, but a best answer.**

Furthermore, arriving at this best answer is a process directly impacted by the judgment, creativity, skill, and intuition of the analyst. Thus, selecting the best answer becomes as much of an art, based on creativity and intuition, as a science.

Data Analysis is an Art, as Well as a Science (Continued)

Many people not acquainted with data analysis are *shocked* to learn that there is not always a right and wrong response to a challenge presented in a data analysis project! This is where the perception of those not exposed to data analysis meets the reality of the actual approach.

The Data Analysis Project is the Work of Art of the Data Analyst

When a data analyst or statistician begins to use their personal judgment, creativity, skill, and intuition to make decisions while conducting a data analysis project, what is happening?

Simply put, **the data analysis project is becoming (at least in part) an expression of the personal judgment, creativity, skill, and intuition of that individual data analyst.** In other words, just as an artist expresses his or her skill, intuition, and self on a canvas with paints, the data analyst does so through the data analysis project with statistical tests and procedures.

If Data Analysis is Personal to the Individual, Could Two Data Analysts Produce Conflicting Findings Using the Exact Same Data?

Yes, absolutely! This is a fact of data analysis that is almost never mentioned, but is always at work.

People used to ask me, what do you think of marriage? To which I would reply, it depends who you are married to. If you are married to someone great, chances are marriage is great or the other way around. Why? It is because a marriage is always a personal expression of the people in that specific marriage. There is no getting around that fact.

Similarly, because the data analyst must rely on his or her personal decision making when conducting data analysis, the data analysis project becomes a personal expression of him or her.

Subsequently, because opinions differ between data analysts, even among top statisticians, one data analyst might consider one method valid in an analysis where another may not. Just one small decision like this could result in data producing significant findings versus findings that are not significant, while using the exact same data. Thus, as I said, the results of a data analysis project may indeed vary by the personal qualities (e.g., thought, judgment) of the specific analyst conducting the study.

An Example

There are several examples of how the findings of a data analysis project might vary according to the personal judgment of the data analyst. One blatant example, might concern the issue of missing data values within a dataset.

Specifically, many times in a dataset one will have study participants that did not provide responses to all survey items, which would result in missing data values within a dataset. The data analyst must decide how to deal with these missing data values.

The simplest solution (though often not the most sophisticated or appropriate) is to do a complete case analysis. Using this method, the study participants with missing values would be removed from the analysis and only the cases with full data would be used.

The question then becomes, what percentage of cases can be removed from the dataset before becoming problematic? As you could imagine, the percentage recommended varies even among experts. For example, for a large sample, recommendations for this cutoff point have varied between 5% (Schafer, 1999), 10% (Bennett, 2001), and even 20% (Peng et al., 2006).

Data Analysis is an Art, as Well as a Science (Continued)

Now, I present this question to you:

If you had a sample of 1000 study participants, do you think that the results of a data analysis might be different if you removed from your sample:

5% ($n=50$) of Study Participants

10% ($n=100$) of Study Participants

Or

20% ($n=200$) Study Participants

Absolutely! For example, if you removed 20% ($n=200$) of study participants from the analysis, these might have been the very cases that might have driven some significant finding within the study. Hence, if you left the 20% in the analysis there may have been a significant finding, but if you followed the recommendation where 20% might be removed, there might have been an absence of significant findings.

Disconcerting?

Many people find the idea that the results of a data analysis study might vary by the data analyst and not just the data. People enjoy a sense of certainty, especially when it comes to topics like the results of a data analysis study that might be used to guide policy, treatment, and future research. Assuredly, when we look at the results of a study we would all like to think that the results are the results and would have been so no matter who was analyzing the data. However, many times, this may not be the case.

What to do?

When you look at a painting completed by Claude Monet, you may recognize that he is from the school of impressionism. Through his work, you can tell what school of art he is from. You can then compare how your style, beliefs, judgment, creativity, and intuition match up with his. In other words, you can relate your art to his art.

Similarly, when you examine someone else's data analysis, you can do the same thing. You can read their research, their methods of statistical analysis, results, and conclusions, and judge if you are from the same school of art as them. In other words, are their thoughts methods, and decisions, in line with yours? Or did they take steps that you would not have taken (such as removing 5% VS. 20% of cases due to missing data)?

Obviously, if the data analyst does not report all these essential elements of a study, it is hard to relate the methods and decisions you would have chosen to the procedures they did use. So it goes without saying that good science is dependent upon proper reporting.

Personal Judgment

Thus, in summation, to conduct your own data analysis study, you need more than just knowledge of essential statistical procedures and rules. You also need a level of artistic personal judgment that will guide you in making important data analysis decisions where there are no hard rules.

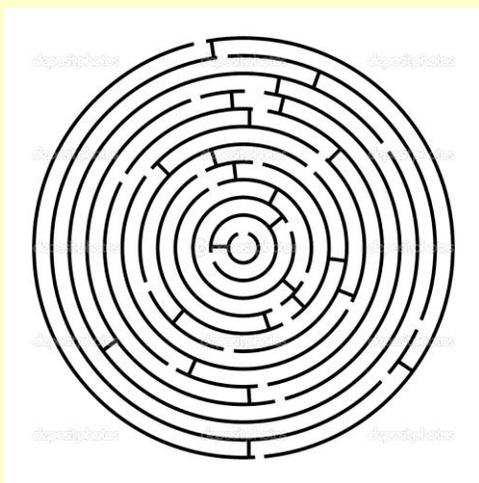
Furthermore, when you read the works of other data analysts, statisticians, and researchers, you must consider more than the statistical analysis. You must consider the decisions they made in data analysis and assess their level of artistic personal judgment. Do you agree with what appears to be their ideology? How alike or disparate are you? Does it appear that these similarities or differences might have impacted the results of the study?

To be an effective data analyst or statistician, you must know what kind of artist you are. You must understand the role your skill, creativity, and intuition are playing in your data analysis.

How Data Analysis in A-Maze-ing

Data analysis can seem a little extensive at times, so I often like to break the whole process down into understandable and manageable terms and figures. Subsequently, I would like to explain how data analysis is like a maze.

When one thinks about a maze one might envision and expect a consistent layout, such as the circularly shaped maze below.



However, one would not expect (or even desire, because it would be redundant) one maze to have the same inner patterns and demands as another, even if both had the same circular layout.

Specifically, each maze may have a similar layout, but the challenges presented by the patterns, the *twists and turns* within the maze would be specific to that individual maze.

So how is data analysis like a maze?

Just like a maze, the layout of a data analysis study may be consistent. For example, I have released several materials that illustrate that each data analysis study has a consistent layout, namely the 7 Steps of Data Analysis.

Furthermore, just like a maze, each data analysis study presents a unique set of patterns, demands, and paths, or *twists and turns* within that consistent 7 Step overall structure.

For example, earlier we described the scenario where a data analyst would need to make decisions regarding the treatment of missing data values. This is a classic *twist and turn* in a data analysis project. It may be a minor twist and turn where only a few cases are missing data or a major twist and turn where half that sample have missing data values. As I said, the twists and turns will be specific to the inner workings of that study.

However, as anyone who has traveled a maze can tell you, one navigates best when informed about the possibilities at every twist and turn.

Therefore, in addition to describing the 7 Steps of Data Analysis (the maze outlay), our materials, including the new *Steps of Data Analysis* text, also conveys the sense of judgment, attitude, and skill that helps analysts identify that “best direction” to take when presented with a twist or turn within that maze-like structure.

You can reserve a 1st edition printing of the text at: <http://www.statswhisperer.com/products/>

Electronic copies of the text will be made available in August and printed copies in September.

Final Comments

It is critical to realize that most artists, like painters, became great with experience and skill.

This is also true regarding data analysis. You do not need to be Monet of the first day.

Developing skill in the art of data analysis is like developing skill in any other art. The skill increases with practice and application.

Many times in my lectures I will use my earlier published works and point out what I would have done differently today.

As they say, where you start is not important, but that you start is!

REFERENCES

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25, 464-469.

Peng, C.-Y. J., Harwell, M., Liou, S.-M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), *Real data analysis* (pp. 31-78). Greenwich, CT: Information Age.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8, 3-15.

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

wb@statswhisperer.com

Thanks for your interest in our newsletter!

Personal Editor/Tutor/Coach

Published writer and experienced editor offers professional and creative editing, tutoring, and coaching assistance.

Credentials include:

- MFA in Creative Writing from The New School University
- Over 10 years experience in private editing, tutoring, and coaching creative and academic writers
- Personal editor to Stacey Patton, Simon & Schuster author of memoir, "That Mean Old Yesterday" (2007)
- Other editing clients include The New Yorker, The Feminist Press, Four Way Books, Salonika Magazine, etc.
- Taught writing instruction at The New School University, 14th Street Y in New York City, Berkeley College
- Certified proofreader
- Published poet

Rates are flexible/References available upon request

email Kathleen at kekky@mindspring.com