Benchmarks Online

Research and Statistical Support
University of North Texas

RSS Matters

You can link to the last RSS article here: Equivalence Tests - Ed.

Statistics: a Clarification

By Dr. Mike Clark, Research and Statistical Support Services Consultant

Many people who ask for our help at RSS due so in part, whether consciously or not, because they are not exactly aware of what statistics is. As such I have cobbled this list together in hopes of making that at least a little clearer. The list is by no means definitive nor exhaustive, but does hit on what I perceive to be some basic misunderstandings prevalent in both the applied setting and to the uninitiated. By *statistics* I mean all parts of a statistical analysis from conceptualization through interpretation and as the chosen tool of scientific investigation. Clicking on the numbers will take you to the corresponding clarification.

Some things statistics is *not*:

1. Math

2. A collection of heuristics

3. Getting output

4. Done quickly or something for which there is a 'quick' answer to

5. Easy

6. Impossible to learn

7. Disconnected from theory

8. Expensive

1. "Statistics is a science in my opinion, and it is no more a branch of mathematics than are

physics, chemistry and economics; for if its methods fail the test of experience—not the test of logic—they are discarded."

That's from Tukey[1], and he was right.

2. While rules of thumb have their place, engaging in statistical analysis is not simply deeming 'meaningful' to anything with X value (whatever the metric), doing X analysis whenever the variables are just so, etc. In some sciences this is only a recent revelation to those on the applied side, and others still have yet to get to that understanding. No research endeavor is an island, and any analysis must be thoughtfully interpreted in light of relevant history, current context and future implications.

3. Getting statistical output has been rendered so easy by technology it could be produced by children with some programs. Just because one has some output doesn't mean one has done a statistical analysis. If this is confusing to you as to why you need more training.

4. If you're doing a proper statistical analysis, it will never be something you can do quickly if you want to do it well. See this related RSS Matters [article](#) by Dr. Herrington. Also, people ask us 'quick questions' all the time. We've yet to ever find a quick answer that would be adequate.

5. This goes with #4 to some extent. Even the simplest of techniques (e.g. calculating a 'middle') has several possibilities available (median, mode, mean, winsorized mean, M-estimators etc.), and some might be equally viable given a situation. A simple independent samples t-test could be done standard, robustly (e.g. trimmed means), nonparametrically (e.g. bootstrapped), determined via a confidence interval for the mean difference, approached purely from an effect size perspective etc. If you think something is able to be done quickly and easily, odds are you have taken a severely limited approach.

6. Learning statistical concepts can certainly be difficult but it certainly is not impossible for *anyone* to learn. It will take time though, and if one equates time with overwhelming difficulty, and many do, then perhaps it is impossible for that person, as much of life probably is.

7. I've often heard people claim to be good with theory and bad with statistics. I have yet to understand how that is possible. No scientific theory gets by without an implication of measurement, and if a theory is devised without an understanding of how it would be measured (and thus eventually analyzed) it is not a scientific one. As such from the very get go methodological and analytical considerations should be at the forefront of any research project, and be continuously considered from that beginning to the end. One of the more common and biggest problems we come across at RSS is data that has no connection to the theory a client is interested in.

8. It can be, but it doesn't have to be. There are alternatives, and one such alternative is [R](#), which is free and which we support here at RSS and use quite extensively. For much of our needs it does everything and more than the expensive stuff. If you don't like the price for something there are options. Here are [some](#).

Some things statistics *is*:

1. An essential part of scientific inquiry

2. Something that entails critical thinking

3. In a constant state of development

4. A source for new ideas

5. Exploratory as well as confirmatory

6. A means by which to understand causal structure

7. A means by which to reduce uncertainty

8. Mathematically intensive

9. Interpretive

10. A very useful tool

1. There is no science without measurement, and no understanding of the measurement without statistics applied by a keen observer. It is the analysis that enables us to extract the knowledge hiding within the data.

2. This part really gets folks that tend to think statistical analysis is many of those things listed in the 'not' column. If you don't like to think hard, you shouldn't be doing statistical analyses or trying to interpret them.

3. People seem to think that their introduction to statistics at the undergrad or graduate level is some sort of overview of the field. Something to take note of: anything in a typical stats textbook involves methods that at the very least are probably a quarter century old, contains many and perhaps most that are well older than that, and omits the vast majority of the field. Things have changed since Fisher's days, a lot.

4. The result of a good analysis is always a springboard for new ways to think about things and a fuller way of understanding previous findings from some research domain. If it does not lead to new ways of thinking you can assume something has gone wrong. As a hint: non-'significant' results are still meaningful.

5. Description and exploration are part and parcel of science. One can gain a tremendous amount of insight from an initial glimpse into a research domain as well simply examining the data to the fullest. For some reason that fell out of favor in some areas of science and/or was not deemed to be 'enough' (for theses, dissertations, publications), or somehow cheating if you explore your own data. A good exploration can be as good as any confirmatory one, and oftentimes can be more useful.

6. If one thinks causal explanations are not possible in a domain of scientific research, they have a gross misunderstanding of both science and causality. There is a continuum of confidence in causal statements, but the goal of science and its methods is to assign causal attributions to natural events.

7. There are lots of things we don't know. Science and its methods are the most viable means with which to reduce our ignorance regarding ourselves and the world around us.

8. Statistics uses a lot of math. So does everyday life, what's the big deal?

9.   Whatever the results are, different interpretations are available and perhaps equally plausible based on the experiences and knowledge of interested parties. This is a *good* thing, not a reason to throw our hands up in frustration.

10. In the end, statistical analysis is simply the tool we use to extract meaning from a great deal of information we can not possibly comprehend otherwise. Seen as such it should hardly be seen as something aversive, nor should there be an inherent negativity or suspicion associated with it. Save that for politics.

All of this may sound like just employment-related bias (though I might mention that I would not refer to myself as a statistician), and much of my experience comes from a social science realm where perhaps this clarification is more needed. However, it doesn't take much to see there are plenty out there in various domains of research that are content to play a game. Some, perhaps many, perhaps a majority of people in some areas of research take a minimum effort approach to their methods and analysis, then try to convince others about what they have supposedly discovered and concluded from those lackadaisical efforts. While there is enough good research being done to sustain progress, it is slowed considerably by such shenanigans. Furthermore, those reading others' works who have the aforementioned misunderstandings of statistics restrict themselves to having to believe whatever the authors conclude or simply distrusting whatever is presented, not a very good state of affairs, but unfortunately one the general public has to deal with.

The causes and possible solutions to these problems I will have to leave for another article. For now it is hoped that the clarification will provide some insight for those that we assist and others who might come across it. Be seeing you.

Return to top