# You may be (stuck) here! And here are some potential reasons why.

As published in Benchmarks RSS Matters, May 2015

http://web3.unt.edu/benchmarks/issues/2015/05/rss-matters

Jon Starkweather, PhD

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
**R**esearch and **S**tatistical **S**upport

http://www.unt.edu

http://www.unt.edu/rss

RSS hosts a number of "Short Courses".
A list of them is available at:
http://www.unt.edu/rss/Instructional.htm

Those interested in learning more about R, or how to use it, can find information here:
http://www.unt.edu/rss/class/Jon/R_SC

# You may be (stuck) here! And here are some potential reasons why.

I often read R-bloggers (Galili, 2015) to see new and exciting things users are doing in the wonderful world of R. Recently I came across Norm Matloff's (2014) blog post with the title "Why are we still teaching *t*-tests?" To be honest, many RSS personnel have echoed Norm's sentiments over the years. There do seem to be some fields which are perpetually stuck in decades long past — in terms of the statistical methods they teach and use. Reading Norm's post got me thinking it might be good to offer some explanations, or at least opinions, on why some fields tend to be stubbornly behind the analytic times. This month's article will offer some of my own thoughts on the matter. I offer these opinions having been academically raised in one such *Rip Van Winkle* (Washington, 1819) field and subsequently realized how much of what I was taught has very little practical utility with real world research problems and data.

## 1 The Lady, Her Tea, and the Reverend

It is extremely beneficial to review the history of statistics in order to understand why some fields seem to be slow in adopting contemporary methods and analyses. There are very few books I would consider *required* reading for anyone with a serious interest in applied statistical analysis. Two such books will be briefly discussed here. First, *The lady tasting tea: How statistics revolutionized science in the twentieth century* by David Salsburg (2001); which is a history book, not a statistics textbook. Salsburg's book provides a very good review of the creation and application, as well as the persons associated with the creation, of statistical analyses during what Salsburg refers to as the *statistical revolution*. Salsburg goes into detail about the persons and personalities behind each breakthrough in the field of statistics, such as early pioneers like Karl Pearson, Charles Spearman, Egon Pearson, Jerzy Neyman, and Sir Ronald Fisher; as well as more recent trail blazers like David Cox, George Box, Donald Rubin, and Bradley Efron; and many more between. However, Salsburg's book only covers one perspective of statistics: the *Frequentist* perspective, which includes the ubiquitous Null Hypothesis Significance Testing (NHST) and associated *p*-values. Very, very briefly, this perspective assumes that the model parameters are fixed and assumed to be known and the data are essentially random; for instance, if the null hypothesis is true, what is the probability of this data? These types of problems can be stated in the general form; what is the probability of the data given a hypothesis? In symbols, this translates to:

$$P(D|H) \tag{1}$$

The other book I consider *required* reading for anyone with a serious interest in applied statistical analysis covers the other perspective of statistics: the *Bayesian* perspective. The Bayesian perspective differs from traditional Frequentist inference by assuming that the data are fixed and model parameters are described by a probability distributions, which sets up problems in the form of; what is the probability of a hypothesis (or parameter), given the data at hand? These types of problems can be stated with symbols as:

$$P(H|D) \tag{2}$$

Sharon McGrayne's (2011) book, *The theory that would not die: How Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy* is similar to Salsburg's (2001) book in that both are history books, not statistical textbooks. McGrayne's book, obviously, begins with the Reverend Thomas Bayes' ideas from the 1740s. The book tracks the origins of Bayes' Rule as a theory and concept which for many years was only theoretical because the complex computations required to actually put it into practice were impossible. The book charts the history of the resurgence of Bayes' Rule as computers emerged in the twentieth century which allowed scientists to apply Bayes' Rule to a variety of (often top secret) complex, practical, real world problems.

# 2 The Desire to be Quantitative

The importance of the histories mentioned above is critical to understanding how some fields have been slow to adopt more modern methods and analyses. As history can show us, much of the previous 100 years of statistical analysis has been dominated by the Frequentist perspective. Most of the methods and analysis of the Frequentist perspective are designed for use in strictly experimental or quasi-experimental research designs. Therefore, as new scientific disciplines emerged and developed with a desire to be empirically grounded, the only methods available were the traditional analyses — what I refer to as the *usual suspects*. These usual suspects include all the things presented in the vast majority of first year applied statistics courses in departments such as Psychology, Sociology, Education, etc. In fact, it has been my experience that the many, many textbooks used for these classes contain the exact same content and it is often presented in the exact same order. The content begins with definitions (e.g. population, sample, the scales of measurement [Stevens, 1946], independent variable, dependent variable, etc.), then descriptive statistics are covered (e.g. measures of central tendency, variability, shape, & relationship), followed by a discussion of the normal distribution and properties of the Standard Normal Distribution (e.g. $Z$-scores, also called standard scores), then a brief discussion of NHST and statistical power, then the $Z$-test is discussed, then the $t$-tests are discussed (e.g. one-sample, independent samples, dependent samples), then oneway analysis of variance [ANOVA] with perhaps a light treatment of factorial ANOVA, then regression — mostly with only one predictor, then subsequent chapters / syllabi cover several non-parametric analogues for the methods previously discussed (e.g. Mann-Whitney $U$, Wilcoxon signed-ranks test, Kruskal-Wallis oneway ANOVA, Chi-square tests, etc.). Now, there is nothing inherently wrong with these methods, they work very well for research designs which provide the types of data they are designed to handle. Unfortunately, these usual suspect analyses each have fairly extensive assumptions which, when the analyses are applied to data which fails to meet those assumptions the resulting statistics are heavily biased or perhaps even invalid. Again, most of these methods were developed for research situations which are truly experimental (i.e. random sampling from a well-defined population of interest, random assignment of cases to conditions of an independent variable, and experimental manipulation of that independent variable while controlling all other variables as much as possible). Unfortunately, true experimental designs are not possible for most of the research done in the emerging or younger scientific disciplines (e.g. Psychology, Sociology, Education, etc.).

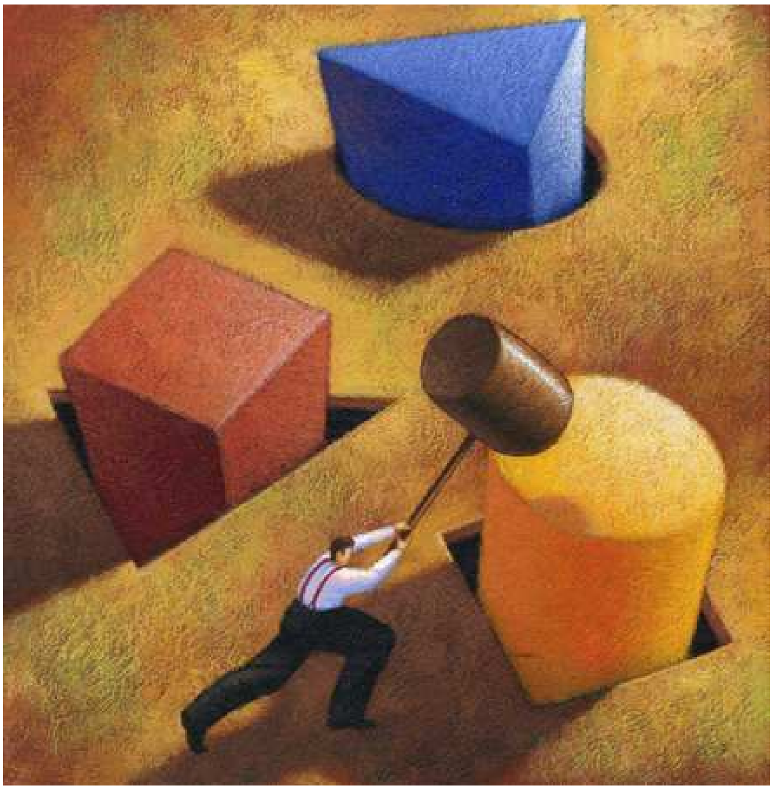
# 3 Intergenerational Momentum

The previous section hinted at what I mean by *Intergeneration Momentum*. The previous section shows how initially the younger sciences had limited options when it came to data analysis — the Frequentist perspective was the only perspective and therefore, only the usual suspects were available. However, intergenerational momentum is responsible for the fact that the vast majority of young science researchers are still using those usual suspects when more effective methods have been developed. Max Planck (1950) said, "a scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it" (p. 33 - 34). Unfortunately, even Planck's mechanism for the advancement of science fails in some fields because some mentors stubbornly stick with one or a few analyses. Worse still, some of these mentors use their authority, or power, as the gate-keepers of a successful thesis or dissertation, to pressure their graduate students to use the mentors' preferred analysis or analyses. Therefore, the *intergeneration* reliance upon outdated, and potentially inadequate, analyses continues in a self-replicating stagnation. One of the most frequent examples of an analysis stubbornly being used despite its creator and namesake attempting to enlighten researchers to its limitations is: *Cronbach's alpha coefficient* (Cronbach, 1951). "I no longer regard the alpha formula as the most appropriate way to examine most data" (Cronbach & Shavelson, 2004, p. 403). Alpha has three critical assumptions; two of which ($\tau$ equivalency & uncorrelated errors), are virtually never satisfied by data resulting from most surveys (for more on this topic, see Starkweather, 2012). Like many of the usual suspects (i.e. traditional Frequentist analyses) the assumptions are often not assessed adequately or are simply ignored — meaning, an untold number of research conclusions are likely based upon very biased or simply invalid statistical results.

## 4 Looking Toward the Future

The primary unit of analysis, for many of the newer or young sciences, is the human being or some aspect of human experience. Unfortunately, from a research perspective, human beings are extremely complex entities and they are constantly interacting with other complex entities (e.g. other humans, social / cultural systems, political systems, economic systems, etc.). Therefore, researchers whose primary units of analysis are human beings should be collecting data which will allow them to fit, compare, and revise complex statistical models capable of accurately representing the complexity of the researcher's subjects and their numerous interactions with other complex entities (e.g. other humans & other complex systems mentioned above). It is well past the time to recognize that our forbearers' General Linear Model [GLM] statistics (e.g. t-tests, ANOVAs, regressions, etc.) should no longer be the default modeling solutions. After all, how many current researchers generate their reports, or manuscripts, using a 1921 – 1940 Corona typewriter?

The above typewriter[1], beautiful as it is, also highlights another area of stagnation among many contemporary researchers. Statistical software has advanced at an incredible rate over the last two decades. Yes, my zealously R-centric eyes are looking at you SPSS and SAS users. There are two, among many, important factors for recommending R[2] over the other two software packages. First, R[3] is completely free, like the air you breathe is free. It seems to me almost irresponsible to continue using expensive software (e.g. SPSS & SAS) in this economic climate when free alternatives exist. Second, R has all the capabilities of SPSS and SAS but, the reverse is not true. R contains the most cutting edge functionality due to its regular rapid update schedule and the continued expansion of its functionality through new procedures being developed by theoretical and applied statisticians' submitted packages (for more on this topic; see Starkweather, 2013).



Lastly, the image[4] above reflects the idea that far too many research analysts are using Frequentist methods when Bayesian methods are much better suited for the types of hypotheses and data of the new or young sciences. The problems with the Frequentist perspective, and in particular NHST, have been thoroughly discussed for many years (Efron, 1986; Cohen, 1994; Krantz, 1999; Hubbard, & Bayarri, 2003; Gigerenzer, Krauss, & Vitouch, 2004; Gelman, & Stern, 2006). The bottom line is this, Bayes methods are not a cure all, but they are likely much better for the vast majority of research situations in the new or young sciences. There are many 'introduction to Bayesian statistics' text books available in a variety of fields (see Starkweather, 2011). Furthermore, there are alternatives to both Frequentists and Bayesian methods; such as machine learning techniques, computational artificial intelligence methods, soft modeling methods, and evolutionary optimization based methods (swarm algorithms, MCMC

[1]Image found at the Smith Corona virtual museum (gallery for 1st generation typewriters, specifically the Corona #3 model): `http://www.smithcorona.com/wp-content/tn3/0/1915CoronaTypewriterCompanyInc.Corona3.j`

[2]http://r-project.org/

[3]http://cran.r-project.org/

[4]Image found at the TribePad blog: `http://tribepad.com/2012/01/the-round-peg-round-hole-approach-to`

methods, genetic algorithms, ant colony optimization, etc.). Additionally, there are wrapper techniques which can be applied to most any analysis and improve the precision of estimates; such as resampling methods like the bootstrap, boosting, bagging, and model averaging (e.g. ensemble averaging). It's time to de-emphasize the usual suspects of NHST and integrate Bayesian and / or other more current methods into curricula to break the stagnation which severely limits these new or young sciences.

Until next time; here's a gentle reminder that May 4th is not *only* Star Wars Day[5]..."Tin soldiers and Nixon coming..."

## 5 References and Resources

Cohen, J. (1994). The Earth is round ($p < .05$). *American Psychologist, 49*(12), 997 - 1003.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297 - 334. doi:10.1007/bf02310555

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391 - 418.

Fisher, R. A. (1935). *The design of experiments*. New York: Hafner Press.

Galili, Tal. (2015). R-bloggers: R news and tutorials contributed by (563) R bloggers [A Word Press blog]. Available at:
`http://www.r-bloggers.com/`

Gelman, A., & Stern, H. (2006). The difference between "Significant" and "Not Significant" is not itself statistically significant. *The American Statistician*, *60*(4), 328 - 331.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences*, pp 391 - 408. Thousand Oaks, CA: Sage Publication.

Efron, B. (1986). Why isn't everyone a Bayesian. *The American Statistician, 40*(1), 1 - 5.

Hubbard, R., & Bayarri, M. J. (2003). Confusion over measures of evidence ($p$'s) versus errors ($\alpha$'s) in classical statistical testing. *The American Statistician, 57*(3), 171 - 182.

Irving, Washington. (1819). Rip Van Winkle in *The sketchbook of Geoffrey Crayon, Gent.* [Geoffrey Crayon was Irving's pseudonym]. New York: Ebenezer Washington & Henry Brevoort.

Krantz, D. H. (1999). The null hypothesis testing controversy in Psychology. *Journal of the American Statistical Association, 44*(448), 1372 - 1381.

Matloff, Norm. (2014). Mad (Data) Scientist: Musings, useful code, etc. on R and data science [A

[5]`http://en.wikipedia.org/wiki/Star_Wars_Day`

Word Press blog]. Available at:
http://matloff.wordpress.com/2014/09/15/why-are-we-still-teaching-about-t-t

McGrayne, S. B. (2011). *The theory that would not die: How Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New Haven, CT: Yale University Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*, 559 - 572.

Planck, M. (1932). *Where is science going?* New York: W. W. Norton & Company, Inc. Freely available in several formats (e.g. Adobe.pdf) at:
https://archive.org/details/whereissciencego00plan_0

Planck, M. (1950). *Scientific autobiography and other papers*. London: Williams & Norgate LTD.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. New York: W. H. Freeman & Company.

Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology, 15*(2), 201 - 292.

Starkweather, J. (2011). Go Forth and Propagate: Book Recommendations for Learning and Teaching Bayesian Statistics. Benchmarks: RSS Matters, September 2011. Available at:
http://www.unt.edu/rss/class/Jon/Benchmarks/BayesBooks_L_JDS_Sep2011.pdf

Starkweather, J. (2012). Step out of the past: Stop using coefficient alpha; there are better ways to calculate reliability. Benchmarks: RSS Matters, June 2012. Available at:
http://www.unt.edu/rss/class/Jon/Benchmarks/Omega_JDS_Jun2012.pdf

Starkweather, J. (2013). Why R; it's not a question, it's an answer. Benchmarks: RSS Matters, October 2013. Available at:
http://www.unt.edu/rss/class/Jon/Benchmarks/WhyR_L_JDS_Oct2013.pdf

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677 - 680.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, pp 1 - 49. New York: Wiley.

This article was last updated on May 15, 2015.

This document was created using LaTeX