This is from webpage: http://www.fharrell.com/post/introduction/

Introduction

Last updated on Jul 22, 2018   ·   3 min read   ·   3 Comments

(http://fharrell.com/post/introduction/#disqus_thread)

🐦

(https://twitter.com/intent/tweet?

text=Introduction&url=http%3a%2f%2ffharrell.com%2fpost%2fintroduction%2f)

𝐟

(https://www.facebook.com/sharer.php?u=http%3a%2f%2ffharrell.com%2fpost%2fintroduction%2f)

in

(https://www.linkedin.com/shareArticle?

mini=true&url=http%3a%2f%2ffharrell.com%2fpost%2fintroduction%2f&title=Introduction)

🅦

(http://service.weibo.com/share/share.php?

url=http%3a%2f%2ffharrell.com%2fpost%2fintroduction%2f&title=Introduction)

✉

(mailto:?subject=Introduction&body=http%3a%2f%2ffharrell.com%2fpost%2fintroduction%2f)

Statistics is a field that is a science unto itself and that benefits all other fields and everyday life. What is unique about statistics is its proven tools for decision making in the face of uncertainty, understanding sources of variation and bias, and most importantly, *statistical thinking*. Statistical thinking is a different way of thinking that is part detective, skeptical, and involves alternate takes on a problem. An excellent example of statistical thinking is statistician Abraham Wald's analysis (https://en.wikipedia.org/wiki/Abraham_Wald) of British bombers surviving to return to their base in World War II: his conclusion was to reinforce bombers in areas in which no damage was observed. For other great examples watch my colleague Chris Fonnesbeck's Statistical Thinking for Data Science (https://www.youtube.com/watch?v=TGGGDpb04Yc).

Some of my personal philosophy of statistics can be summed up in the list below:

- Statistics needs to be fully integrated into research; experimental design is all important
- Don't be afraid of using modern methods

- Preserve all the information in the data; Avoid categorizing continuous variables and predicted values at all costs
- Don't assume that anything operates linearly
- Account for model uncertainty and avoid it when possible by using subject matter knowledge
- Use the bootstrap routinely, when not using Bayesian methods
- Make the sample size a random variable (https://stats.stackexchange.com/questions/256623) when possible
- Use Bayesian methods whenever possible
- Use excellent graphics, liberally
- To be trustworthy research must be reproducible
- All data manipulation and statistical analysis **must** be reproducible (one ramification being that I advise against the use of point and click software in most cases)

Statistics has multiple challenges today, which I break down into three major sources:

1. Statistics has been and continues to be taught in a traditional way, leading to statisticians believing that our historical approach to estimation, prediction, and inference was good enough.
2. Statisticians do not receive sufficient training in computer science and computational methods, too often leaving those areas to others who get so good at dealing with vast quantities of data that they assume they can be self-sufficient in statistical analysis and not seek involvement of statisticians. Many persons who analyze data do not have sufficient training in statistics.
3. Subject matter experts (e.g., clinical researchers and epidemiologists) try to avoid statistical complexity by "dumbing down" the problem using dichotomization, and statisticians, always trying to be helpful, fail to argue the case that dichotomization of continuous or ordinal variables is almost never an appropriate way to view or analyze data. Statisticians in general do not sufficiently involve themselves in measurement issues.

I will be discussing several of the issues in future blogs, especially item 1 above and items 2 and 4 below. Complacency in the field of statistics and in statistical education has resulted in

1. reliance on large-sample theory so that inaccurate normal distribution-based tools can be used, as opposed to tailoring the analyses to data characteristics using Bayesian methods, the bootstrap, and semiparametric models
2. belief that null hypothesis significance testing ever answered the scientific question and the p-values are useful
3. avoidance of the likelihood school of inference (relative likelihood, likelihood support intervals, likelihood ratios, etc.)

4. avoidance of Bayesian methods (posterior distributions, credible intervals, predictive distributions, etc.)

I was interviewed by Kevin Gray in July, 2017 where more of my opinions about statistics may be found (http://www.greenbookblog.org/2017/08/02/vital-statistics-you-never-learned-because-theyre-never-taught/).

See this article (/post/improve-research) for related thoughts.

2017 (/tags/2017)