

Statistics—a particularly significant skill for early-career researchers

June 1 2016, by Megan Yu



Credit: PLOS Blogs

I first encountered statistics in my AP Statistics class when I was 16 years old; I remember a particular lecture where we learned the importance of distinguishing the parametric t-test and the nonparametric [Mann-Whitney U test](#). As a chemistry undergraduate student, I continue using the same basic statistical principles to analyze whether certain molecules could serve as potential biomarkers for cancer, or why children with feeding disorders impose greater stress on the family than children without these disabilities.

Unfortunately, not everyone has the opportunity to learn [statistics](#) at an early age, and, in my experience, few early career researchers (ECRs) have made the time to learn it on their own so that they can use it in their research. It is common to find ECRs to be unsure of why they should choose one statistical test over another, or to have only a surface-level understanding of what statistics is and how it could best benefit their science.

The essence of statistics

Simply put, statistics is a branch of mathematics that covers a range of procedures for gathering, organizing, analyzing, and presenting quantitative [data](#). It has two main branches—descriptive and inferential statistics. [Descriptive statistics](#) is primarily concerned with describing [quantitative data](#). [Inferential statistics](#) is used to make inferences about the population being studied by analyzing data collected from individual samples. Data analysis of the individual samples is followed by modeling patterns in the data to account for randomness and uncertainty.

The prevalence and importance of statistics

Statistics is used in most areas of science. For instance, in a recent [PLOS ONE paper](#), Otero-Losada et al. used biostatistics to show that moderate running was beneficial to pancreatic morphology in cola-drinking rats. In another [PLOS ONE paper](#) published this month, Young and Gobler used a one-way ANOVA test and discovered acidification can promote macroalgae overgrowth in eutrophic estuaries. This small sample of papers shows that, from mammals to microbes, statistics is necessary to make sense of results.

Since statistics can be applied to a diverse array of scientific disciplines, it has evolved into different branches. For example, [astrostatistics](#)

applies statistical principles to the understanding of astronomical data, while [econometrics](#) uses statistical methods in the empirical study of economic theories and relationships. [Biostatistics](#) uses statistical principles to understand biological phenomena, and [environmental statistics](#) uses statistical methods to understand and evaluate the environmental conditions around us. These are just some of the examples of specialized statistics.

Given that society cannot run effectively without a standardized system that allows everyone to summarize data, it is important for every researcher to have the principles of statistics ready in their toolbox. Not only will researchers need statistics to be able to present and communicate their findings more effectively, but they will also need to be able to understand and evaluate the credibility of other academic papers in their fields. Statistics also helps researchers control for sources of variation, detect outliers, visualize their data, and design effective experiments that help answer their research questions.

Common problems in statistical communication

Despite the clear value of statistics for scientific endeavors, it is common to find improper use of statistics in research. Researchers may inadvertently alter their scales to change the distribution of their data or ignore outliers in order to present their data more coherently. Other common problems in statistical analysis include presenting correlation as causation, misreporting the estimated error in the data, and overgeneralizing the results.

Also, the pressure to publish in order to advance in science can lead to researchers collecting or selecting more data samples until nonsignificant results become significant. This phenomenon, otherwise known as "p-hacking," challenges the traditional scientific model of publishing [significant data](#), or data that yield a p-value less than 0.05. According to

a [2012 *Psychological Science* paper](#) more than half of the the 2,000 psychologists surveyed admitted to "failing to report all of a study's dependent measures," and "deciding whether to collect more data after looking to see whether the results were significant."

Another common scientific issue, which is also fueled by researchers' desires to advance their scientific careers, is the irreproducibility of results. [According to a 2015 *Science* paper](#), among replications of 100 experimental and correlational studies published in three psychology journals, 97% of original studies had reported statistically significant results, but only 36% of the replications had statistically significant results.

What can be done?

Researchers have not yet determined how statistical training can be improved for basic science and translational researchers. [An April 2016 *PLOS Biology* paper](#) outlines methods to improve statistical education, which include: encouraging departments to require statistics training, tailoring coursework to the students' fields of research, and developing tools and strategies to promote education and dissemination of statistical knowledge. Furthermore, in a heavily cited *PLOS Biology* paper, Megan Head and colleagues find that p-hacking, while rampant in evolutionary biology, does not seem to impact the final result. Head et al. suggest that researchers clearly adhere to common analysis standards, use sufficient sample sizes, perform data analyses blind whenever possible, and assess the quality of research methods separately from the results. These recommendations should help address the issues of p-hacking and irreproducibility plaguing modern science. In addition to these recommendations, I think journals should do three things to prevent p-hacking: (1) provide clear and detailed guidelines for the complete reporting of data analyses and results, (2) encourage the specification of methods, and (3) facilitate open access to raw data.

While statistical education may not be readily available for all scientists, I encourage early career researcher to consider taking a statistics class in their own field of study, or consult an online statistics guide/statistician when performing research. A deeper understanding of fundamental statistical principles will not only enhance the mission of science to produce robust scientific findings that could improve our understanding of the world, but also encourage researchers to engage in ethical scientific conduct.

More information: Otero-Losada M, González J, Müller A, Ottaviano G, Cao G, Azzato F, et al. (2016). Exercise Ameliorates Endocrine Pancreas Damage Induced by Chronic Cola Drinking in Rats. *PLoS ONE* 11(5): e0155630.

Young CS, Gobler CJ. (2016). Ocean Acidification Accelerates the Growth of Two Bloom-Forming Macroalgae. *PLoS ONE* 11(5): e0155152.

John LK, Loewenstein G, Prelec D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 23(5): 524–532.

Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. (2015). The Extent and Consequences of P-Hacking in Science. *PLoS Biol* 13(3): e1002106.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716-aac4716.

This story is republished courtesy of PLOS Blogs: blogs.plos.org.

Provided by Public Library of Science

Citation: Statistics—a particularly significant skill for early-career researchers (2016, June 1)
retrieved 26 April 2024 from

<https://phys.org/news/2016-06-statisticsa-significant-skill-early-career.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.