Big data! Machine learning! Cloud computing! There's been a lot of buzz over the last few years about these promising "new" features of data science. But behind-the-scenes, statistics and data science are already embedded in our daily life. They alert us to unusual outbreaks of E. coli in romaine lettuce, determine unsafe levels of smoke and pollutants in the air after a wildfire, inform product placement in grocery stores and malls, advise treatment decisions for medical procedures, and recommend new words and spelling on smart phones. Unfortunately, statistics has been treated like a powerful secret weapon; we know it's extremely useful, but not many people understand why or how to use its powers. And even fewer know how to become a (bio)statistician. In this post, I'll demystify some important statistical concepts and share my journey to studying biostatistics.

With the increased storage capacity of computers, we can now save more information than was once thought possible, which is why some people call it the era of "big data". But with all this information, it's trickier to analyze it correctly without misinterpreting or misrepresenting what we have. Training in statistics allows us to design a study to best address the questions we have and to make reasonable judgment calls when analyzing imperfect data. And imperfections in data collection are inevitable, whether a participant skips a question on a survey, a machine malfunctions and doesn't accurately measure the air quality on a certain day, or a researcher forgets to collect people's background characteristics. In fact, if we did have all the information we wanted, we wouldn't need to use statistics! We'd have all the answers in front of us. Statistics help us make educated guesses when we don't have all the relevant information.

The field of biostatistics applies statistical tools to public health and medical settings (hence the "bio" part). It answers questions like "which types of foods are better for you if you're worried about heart disease and are trying to lower your LDL ('bad') cholesterol levels?" Before we can address this question, though, we need to consider which comparisons truly interest us. Do we want to summarize people overall, or are we interested in certain populations, such as comparing people of the same age group, education level, or health status? Do we think this health relationship occurs differently in men and women, or in our youth, adults, and elders? How do we take into account that some people are taking cholesterol medication, like statins? Thus, when designing a study, it's crucial to make sure you're collecting the appropriate information so that you can answer the relevant questions.

My interest in patterns and statistics led me to pursue a mathematics degree, and my drive to help others motivated me to search for an applied subfield. Though one-on-one connections through community service have great value to me, I thought that I could make a bigger impact if I approached problems on a broader scale. That's when people started suggesting I look at public health. While the medical professions excel at improving

the health of one person at a time, public health investigates whole populations at once to inform best-practices and policy change. I attended several conferences such as Infinite Possibilities and the Society for Advancement of Chicanos/Hispanics and Native Americans in Science, which highlighted the variety of career options available for math students: being a science advisor for national grants, determining safe dosage levels of new medications at the FDA, predicting consumer behaviors for tech companies, finding hotspots of cancer and infectious diseases as an epidemiologist, or even starting a consulting firm.

Interacting with other STEM undergrads and professionals of diverse backgrounds inspired me to pursue a PhD in biostatistics, since this field bridges my passions of data analysis and health. Now, during collaborations with scientists or health professionals, my biostatistics training allows me to better understand the context of a problem, choose reasonable modeling assumptions, and clearly convey what can and cannot be deduced from a particular dataset. I've been involved with projects ranging from forestry (assessing the impact of severe fires on pine tree growth), to cardiology (determining whether exercise decreases the strength of the relationship between depression and heart disease), to psychology (evaluating whether a resilience class helped undergraduates reduce their stress at the end of the quarter). In my dissertation, I am focusing on foods that contribute to a heart-healthy diet, using data from the Multi-Ethnic Study of Atherosclerosis (MESA). Not only do I get to advance the nutritional field, I am expanding current data analysis models so that other statisticians can use them on similarly structured datasets from other sciences.

One of the major beauties of being a biostatistician is getting to work on problems from a variety of fields. I've already had the pleasure of learning about global health, reproductive health, social work, health services, and more. If you too enjoy uncovering various facets of people's health with statistical models, learn more about biostatistics below!

www.biostat.washington.edu/about/biostatistics
www.healthcare-management-degree.net/faq/what-exactly-is-biostatistics/
simplystatistics.org/2015/11/09/biostatistics-its-not-what-you-think-it-is/
www.environmentalscience.org/career/biostatistician

Biography

As a PhD student in Biostatistics, Natalie Gasca focuses on identifying food patterns that are related to heart disease prevention. She is deciding how to create the most informative patterns, expanding the use of statistical methods in nutritional science, and improving upon those methods to better answer her scientific aims.