Historical Background of Statistics
Chris Spatz

WE BEGIN OUR exploration of statistics with a trip to London. The year is 1900.

Walking into an office at University College London, we meet a tall, well-dressed man about 40 years old. He is Karl Pearson, Professor of Applied Mathematics and Mechanics. I ask him to tell us a little about himself and why he is an important person. He seems authoritative, glad to talk about himself. As a young man, he says, he wrote essays, a play, and a novel, and he also worked for women's suffrage. These days, he is excited about this new branch of biology called genetics. He says he supervises lots of data gathering.

Pearson, warming to our group, lectures us about the major problem in science—there is no agreement on how to decide among competing theories. Fortunately, he just published a new statistical method that provides an objective way to decide among competing theories, regardless of the discipline. The method is called chi square.[1] Pearson says, "Now, arguments will be much fewer. Gather a thousand data points and calculate a chi square test. The result gives everyone an objective way to determine whether or not the data fit the theory."

Exploration Notes from a student: Exploration off to good start. Hit on a nice, easy-to-remember date to start with, visited a founder of statistics, and had a statistic called chi square described as a big deal.

Our next stop is Rothamsted Experiment Station just north of London. Now the year is 1925. There are fields all around the agricultural research facility, each divided into many smaller plots. The growth in the fields seems quite variable.

Arriving at the office, the atmosphere is congenial. The staff is having tea. There are two topics—a new baby and a new book. We get introduced to Ronald Fisher, the chief statistician. Fisher is a small man with thick glasses and red hair.

He tells us about his new child[2] and then motions to a book on the table. Sneaking a peek, we read the title: *Statistical Methods for Research Workers*. Fisher becomes focused on his book, holding forth in an authoritative way. He says the book explains how to conduct experiments and that an experiment is just a comparison of two or more conditions. He tells us we don't need a thousand data points. He says that small samples, randomly selected, are the way for science to progress. "With an experiment and my technique of analysis of variance," he exclaims, "you can determine why that field out there"—here he waves toward the window—"is so variable. We can find out what makes some plots lush and some mimsy." *Analysis of variance*,[3] he says, works in any discipline, not just agriculture.

---

[1] Chi square, which is explained in this book in Chapter 14, has been called one of the 20 most important inventions in the 20th century (Hacking, 1984).
[2] (in what will become a family with eight children).
[3] explained in Chapters 11-13

Exploration Notes: Looks like statistics had some controversy in it.[4] Also looks like progress. Statistics is used for experiments, too, and not just for testing theories. And Fisher says experiments can be used to compare anything. If that's right, I can use statistics no matter what I major in.

Next we go to Poland to visit Jerzy Neyman at his office at the University of Warsaw. It is 1933. As we walk in, he smiles, seems happy we've arrived, and makes us feel completely welcome.

Motioning to an envelope on his desk, he tells us it holds a manuscript that he and Egon Pearson[5] wrote. "The problem with Fisher's analysis of variance test is that it focuses exclusively on finding a difference between groups. Suppose the statistical test doesn't detect a difference. Does that prove there is no difference? No, of course not. It may be that the test was just not sensitive enough to detect the difference. Right?" At his question, a few of us nod in agreement. Seeing uncertainty, he notes, "Maybe a larger sample is needed to find the difference, you see? Anyway, what we've done is expand statistics to cover not just finding a difference, but also what it means when the test doesn't find a difference. Our approach is what you people in your time will call *null hypothesis significance testing.*"

Exploration Notes: Statistics seems like a work in progress. Changing. Now it is not just about finding a difference but also about what it means not to find a difference. Also, looks like null hypothesis significance testing is a phrase that might turn up on tests.

Our next trip is to libraries, say, anytime between 1940 and 2000. For this exploration, the task is to examine articles in professional journals published in various disciplines. The disciplines include anthropology, biology, chemistry, defense strategy, education, forestry, geology, health, immunology, jurisprudence, manufacturing, medicine, neurology, ophthalmology, political science, psychology, sociology, zoology, and others. I'm sure you get the idea—the whole range of disciplines that use quantitative measures in their research. What this exploration produces is the discovery that all of these disciplines rely on a data analysis technique called *null hypothesis significance testing* (NHST).[6] Many different statistical tests are employed. However, for all the tests in all the disciplines, the phrase, "$p < .05$" turns up frequently.

Exploration Notes: It seems that all that earlier controversy has subsided and scientists in all sorts of disciplines have agreed that NHST is the way to analyze quantitative data. All of them seem to think that if there is a comparison to be made, applying NHST is a necessary step to get correct conclusions. All of them use "$p < .05$," so I'll have to be sure to find out exactly what that means.

Our next excursion is a 1962 visit with Jacob Cohen at New York University in New York City. He is holding his article about studies published in the *Journal of Abnormal and Social Psychology*, a leading psychology journal. He tells us that the NHST technique has

---

[4] The slight sniping I've built into this story is just a hint of the strong animosity between Fisher and Pearson.
[5] Egon Pearson was Karl Pearson's son.
[6] Null hypothesis significance testing is first explained in Chapters 9 and 10.

problems. Also, he says we should be calculating an effect size statistic, which will show whether the differences observed in our experiments are large or small.

Exploration Notes: The idea of an effect size index makes a lot of sense. Just knowing there is a difference isn't enough. How big is the difference? Wonder what "problems with NHST" is all about.

Back to the library for a final excursion to check out recent events. We come across a 2014 article by Geoff Cumming on the "new statistics." We find things like, "avoid NHST and use better techniques" (p. 26) and "we should not trust any p value" (p. 13). This seems like awfully strong advice. Are researchers taking this advice? Looking through more of today's research in journals in several fields, we find that most statistical analyses use NHST and there are many instances of "$p < .05$."

Exploration Notes, Conclusion: These days, it looks like statistics is in transition again. There's a lot of controversy out there about how to analyze data from experiments. The NHST approach is still very common, though, so it's clear I must learn it. But I want to be prepared for changes. I hope knowing NHST will be helpful for the future.[7]

Welcome to statistics at a time when the discipline is once again in transition. A well-established tradition (null hypothesis significance testing) has been in place for almost a century but is now under attack. New ways of thinking about data analysis are emerging, and along with them, a collection of statistics that do not include the traditional NHST approach. As for the immediate future, though, NHST remains the method most widely used by researchers in many fields. In addition, much of the thinking required for NHST is required for other approaches.

Our exploration tour is over, so I'll quit supplying notes; they are your responsibility now. As your own experience probably shows, making up your own summary notes improves retention of what you read. In addition, I have a suggestion. Adopt a mindset that thinks *growth*. A student with a growth mindset expects to learn new things. When challenges arise, as they inevitably do, acknowledge them and figure out how to meet the challenge. A growth mindset treats ability as something to be developed (see Dweck, 2016). If you engage yourself in this course, you can expect to use what you learn for the rest of your life.

The main title of this book is "Exploring Statistics." *Exploring* conveys the idea of uncovering something that was not apparent before. An attitude of *searching, wondering, checking*, and so forth is what I want to encourage. (Those who object to traditional NHST procedures are driven by this exploration motivation.) As for this book's subtitle, "Tales of Distributions," I'll have more to say about it as we go along.

Statistics: Then and Now

Statistics began with counting, which, of course, was prehistory. The origin of the mean is almost as obscure. It was in use by the early 1700s, but no one is credited with its discovery. Graphs, however, began when J. H. Lambert, a Swiss-German scientist and mathematician, and

[7] Not only helpful, but necessary, I would say.

William Playfair, an English political economist, invented and improved graphs in the period 1765 to 1800 (Tufte, 2001).

The Royal Statistical Society was established in 1834 by a group of Englishmen in London. Just 5 years later, on November 27, 1839, at 15 Cornhill in Boston, a group of Americans founded the American Statistical Society. Less than 3 months later, for a reason that you can probably figure out, the group changed its name to the American Statistical Association, which continues today (*www.amstat.org*).

According to Walker (1929), the first university course in statistics in the United States was probably “Social Science and Statistics,” taught at Columbia University in 1880. The professor was a political scientist, and the course was offered in the economics department. In 1887, at the University of Pennsylvania, separate courses in statistics were offered by the departments of psychology and economics. By 1891, Clark University, the University of Michigan, and Yale had been added to the list of schools that taught statistics, and anthropology had been added to the list of departments. Biology was added in 1899 (Harvard) and education in 1900 (Columbia).

You might be interested in when statistics was first taught at your school and in what department. College catalogs are probably your most accessible source of information.