

## Why Statistics for Lawyers?

The brunt of this course will be devoted to statistics and exploratory data analysis.

*Exploratory data analysis* is just looking at data to see what you see. We will spend some time, for example, looking at how to see the shape of data and what that can tell you about the world, how to shape investigations with it.

But the heavy stuff all falls under the rubric of *statistics*—and my hope is to convince you by the end of the course that it needn't be quite so heavy as lawyers have traditionally feared.

So what's the point of statistics? There are really two points to statistics: first, to allow us to quantify our confidence in our understanding of reality, and second, to help us tease out cause and effect. Let's motivate our exploration of statistics with a few examples.

### You Be the Scientist

First, suppose you're a scientist and you think you've invented a cure for the common cold. How do you test it and make sure it works?

You might start by experimenting on yourself. That has a pretty long tradition in science, actually—Marie and Pierre Curie, for example, experimented on themselves to learn about how radiation worked (and suffered very severe radiation poisoning as a result).

But, if you're going to take on the risk of hurting yourself, you should be sure that you're learning something as a result! Suppose you take the drug, and then you get well the next day. You might conclude that the drug cured you, but it could also just be that the cold was about to run its natural course. The drug could have done nothing... or it could have even made it worse—maybe if you hadn't taken it, you would have gotten better that evening.

The problem is that you'd really like to know how quickly you would have gotten better if you hadn't taken the medicine. But you can never know that, because you did take the medicine. We can never see the *counterfactual* case of what would have happened if we had done something different. That inability has spawned countless speculative takes on history: what would have happened if Japan had never attacked Pearl Harbor and brought America into the war? Would we all be speaking German? Nobody can know.

So you decide to try the next best thing: your old college roommate also has a cold, and, as you learned from your undergraduate days, you're pretty good at talking your poor roommate into doing dumb things. So you decide, he'll take the medicine, and you'll go untreated, and you'll see who gets well faster. This is a scientific experiment, and you're the control group.

Alas, it's a pretty lousy scientific experiment. How do you know that you and your friend don't heal at different paces? Or maybe s/he took some other medicine beforehand and, not wanting to disappoint you, didn't report that fact? Maybe being the kind of person who is willing to participate in someone's wild-west-style science experiment also makes you more likely to heal faster (something about stress and cortisol levels, perhaps). What you should have done was recruit a bunch of other participants, and then randomly assign some of them to a treatment group and some to a control group.

The discipline of statistics will guide our hypothetical scientist in designing that experiment and evaluating its results. For example:

- Recruiting a large number of experimental participants and delivering treatment to them is difficult, expensive and potentially risky (what if our drug turns out to have nasty side effects). We'd like to recruit as few people as possible in order to get reliable results. The discipline of *statistical power analysis* can tell us how many people we need to recruit in order to be able to detect the effect we think we have.
- Suppose we see an effect. People who take the medicine get well, on average, a day faster than people who don't. How do we tell if that's a real effect, as opposed to just a coincidence based on idiosyncratic properties of the group of people who we happened to recruit? We use statistical tools like *hypothesis testing* in order to figure out how confident we can be that the results we see represent a real biological phenomenon (although we can never be completely confident).

## Experiments in the Law

It turns out that experimental research can be involved in building a legal case as well. The example I like best (partly because of my own history as a civil rights lawyer) is discrimination testing. This is a technique that investigators use in a number of contexts, perhaps most prominently in the housing context. The Department of Justice as well as nonprofit fair housing organizations often use testing to investigate claims of housing discrimination. (You may have been introduced to the concept of discrimination testers in your Professional Responsibility class, in the course of discussing the controversial disciplinary case *In re Gatti*, 8 P.3d 966 (Ore. 2000) (under Oregon professional responsibility rules, lawyer may not misrepresent identity for the purposes of conducting investigation into alleged wrongdoing), which was seen as a potential threat to the practice of discrimination testing.)

Here's roughly how housing discrimination testing works: suppose there's reason to believe that an apartment complex discriminates against African-American rental applicants, say, by telling every African-American tenant that there are no vacancies. The investigating organization sends a white couple and an African-American couple, who are trained to behave similarly in other relevant respects,

to inquire about apartments. If the two couples are given different information about the existence of vacancies, the organization concludes that it must be because of their race—and has concrete evidence of illegal discrimination.

One thing that you may have noticed is that I did not describe a large-scale study in the last paragraph. Indeed, these testing organizations typically use numbers of testers that would be unlikely to support a statistically generalizable result (they tend to operate on tight budgets). Yet (as we will discuss in more detail later on) that does not mean that this evidence is not relevant or persuasive. For example, the court in *United States v. Garden Homes Mgmt., Corp.*, 156 F. Supp. 2d 413 (D.N.J. 2001) rejected a motion for summary judgment, holding that the evidence of three paired tests like those described above was sufficient to create an issue of fact on whether or not the defendant maintained a practice of discrimination. As the court said:

While, in the abstract, three fair housing tests might not necessarily reflect Defendants' rental practices, the tests at issue are particularly probative. They all occurred within one month and yielded virtually identical results. The timing and uniform results of these tests give rise to a permissible inference that racial discrimination is not a sporadic event but, instead, is part of the way that Defendants do business.

There are a number of ways we might interpret this. We might think that juries are entitled to make their own inferences about the evidence independent of the way that a statistician would think about it, or we might think—as the court seemed to suggest in *Southern California Housing Rights Center v. Krug*, 564 F.Supp.2d 1138 (C.D., Cal., 2007)—that the behavior exhibited by landlords in these tests are themselves acts of discrimination in violation of the Fair Housing Act, independent of the fact that no tester actually intended to rent. See generally *Havens Realty Corp. v. Coleman*, 455 U.S. 363 (1982) (tester has independent right to accurate information, sufficient to support standing under Fair Housing Act). At any rate, we will discuss this further down the road.

## From Experiments to Observations

Statistical evidence of discrimination will actually be a big theme of this course. (It's one of the most prominent and interesting ways in which statistics is used in the law.) However, most of the statistical evidence we see in the law is what's known as *observational*—rather than randomly assigning people to treatment and control groups, or otherwise controlling the delivery of the treatment to people. (Indeed, even the housing tests probably can't be seen as randomized controlled trials in the fullest sense as the ideal from experimental science, as race cannot be randomly assigned to people—although we do have scientific studies where researchers have managed to come close to that, such as a famous study

where resumes with stereotypically racialized names were sent to employers.)

For example, a number of cities have filed lawsuits against mortgage lenders, alleging that they engaged in discrimination by steering borrowers of color to higher interest-rate loans. These cities have not (as far as I know) conducted tests by, for example, sending applicants of different races in to apply for loans. Rather, they've conducted a statistical analysis of the terms of the loans offered in their communities, and filed lawsuits on the basis of a disparity across races—finding that borrowers of color typically ended up with worse loans.

The worry about such statistical evidence is that it might be *confounded* by *omitted variables*. This is just a fancy stats-person way of saying what we said before about the medical test. Suppose that the city filing the suit has a lot of racial segregation and concentrated poverty—it might be the case that the victims of that segregation get worse loan terms because they've suffered financial disadvantage affecting their creditworthiness, not because of intentional discrimination by the lender.

As alternative to randomization (also known as “experimental controls”), researchers use “statistical controls” to try to solve this problem. We will spend a lot of time on statistical controls in this course, so here's just a taste now.

Imagine an Age Discrimination in Employment Act lawsuit, in which the plaintiff alleges that during layoffs at a firm, older workers had a substantially higher chance of being let go. Just such a lawsuit has recently been filed against IBM. In such a lawsuit, the employer will invariably claim that the older workers aren't being fired because of their ages—rather, they'll claim that, for example, the company choose to retain workers with more relevant skills, and if that happened to fall harder on older workers, it's just because they failed to keep their skills up to date. (This seems to be IBM's basic defense in the noted suit.) How should the plaintiff's lawyer use statistics to try to prove their case?

Here's one tool they might try. You'll probably remember the equation for a line from your high school geometry classes:

$$y = ax + b$$

Just to refresh your memory: you can graph a line by starting at  $b$ , the intercept, and then just getting the rest of the points from that formula, i.e., when the point is at 1 on the  $x$  axis, it's at  $a + b$  on the  $y$  axis, when it's at 2 on the  $x$  axis, it's at  $2a + b$  on the  $y$  axis, and so forth.

In statistics, we can go the other way: we can take a bunch of data and try, using mathematical techniques that we'll look at later, to find the line that best describes it. This is easier to think about when the  $y$  axis takes on more than two values, so let's think of it as salary (for a moment) rather than whether someone was fired or not. Also, let's follow the conventions of statistics and put the intercept first, plus sneak some Greek letters in there.

So, if plaintiff is claiming that the older an employee is, the lower a raise they got last year, then we can collect a dataset of all the employees and their raises, and then set the percentage raise last year to be  $y$  and the employee's age to be  $x$ . Then we can use those mathematical techniques that I just hinted at (patience!) to find the *regression line* that best fits the data. That means finding estimates for the *coefficients* alpha and beta in:

$$y = \alpha + \beta x$$

If beta is negative, that suggests that older workers are indeed getting smaller raises. Then we can use more statistical techniques which we will learn to discern whether it is negative enough to be *significant*—to give us confidence that lower raises are genuinely due to age rather than to chance.

But wait! The employer still has that argument about skills and such. It turns out that regression analysis allows us to take account of those arguments. Let's say that an employee's level of skill is represented by how many training courses they took last year. What we really want to do is compare apples to apples—to compare older employees who took, say, 10 training courses last year to younger employees who took 10 courses, and older employees who took 9 courses to younger employees who took 9 courses, and so on.

It turns out that you can extend our equation to draw a line to include more  $x$  variables. So let's say  $x_1$  is the employee's age, and  $x_2$  is the number of training courses the employee took last year. Then we can find the best values of  $\beta_1$ ,  $\beta_2$ , and  $\alpha$  in the following:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

It turns out that (subject to some caveats and pitfalls on which we shall spend some time), the value of  $\beta_1$  will be a good estimate of the effect of age on pay raises *holding the number of training courses constant*—so it gives us the apples-to-apples comparison we're looking for. And, within some limits, we can keep adding variables as the employer comes up with more and more excuses for the lower salary.

Anyway, we'll be spending quite a bit of time on regression analysis, this is just a taste for now. You will hopefully be able to see that being able to make sense of these concepts will help you in your practice. This is just an introduction, we'll dig in much deeper in the weeks to come.