towardsdatascience.com

The Inspection Paradox is Everywhere

Allen Downey

10-13 minuter

The inspection paradox is a statistical illusion you've probably never heard of. It's a common source of confusion, an occasional cause of error, and an opportunity for clever experimental design.

And once you know about it, you see it everywhere.

One of my favorite examples is the apparent paradox of class sizes. Suppose you ask college students how big their classes are and average the responses. The result might be 90. But if you ask the college for the average class size, they might say 35. It sounds like someone is lying, but they could both be right.

When you survey students, you oversample large classes: If there are 10 students in a class, you have 10 chances to sample that class; if there are 100 students, you have 100 chances. In general, if the class size is x, it will be overrepresented in the sample by a factor of x.

That's not necessarily a mistake. If you want to quantify student experience, the average across students might be a more meaningful statistic than the average across classes. But you have to be clear about what you are measuring and how you report it.

By the way, I didn't make up the numbers in this example. They come from data reported by Purdue University for undergraduate class sizes in the 2013–14 academic year.

From their report I estimate the actual distribution of class sizes (with guesses to fill missing data). Then I compute the "biased" distribution you would get by sampling students. Figure 1 shows the
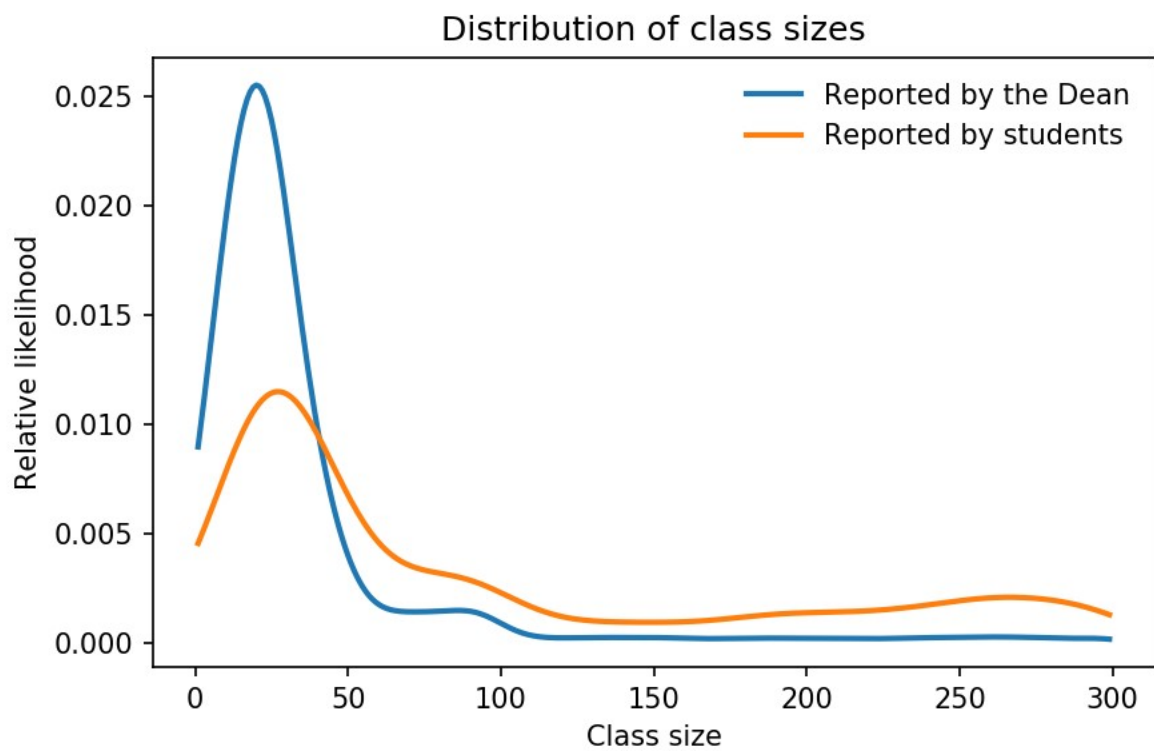
results.

Figure 1: Undergraduate class sizes at Purdue University, 2013–14 academic year: estimated distribution as reported by the University and biased view as seen by students.

The student sample is less likely to contain classes smaller than 40 and more likely to contain larger classes.

Going the other way, if you are given the biased distribution, you can invert the process to estimate the actual distribution. You could use this strategy if the actual distribution is not available, or if it is easier to run the biased sampling process.

The same effect applies to passenger planes. Airlines complain that they are losing money because so many flights are nearly empty. At the same time, passengers complain that flying is miserable because planes are too full. They could both be right. When a flight is nearly empty, only a few passengers enjoy the extra space. But when a flight is full, many passengers feel the crunch.

The inspection paradox also happens when you are waiting for public transportation. Buses and trains are supposed to arrive at constant intervals, but in practice some intervals are longer than others.

With your luck, you might think you are more likely to arrive during a long interval. And you're right: a random arrival is more likely to fall in a long interval because, well, it's longer.

To quantify this effect, I collected data from the Red Line in Boston. Using the MBTA's real-time data service, I recorded the arrival times for 70 trains between 4pm and 5pm over several days.

The shortest gap between trains was less than 3 minutes; the longest was more than 15. Figure 2 shows the actual distribution of time between trains, and the biased distribution that would be observed by passengers.
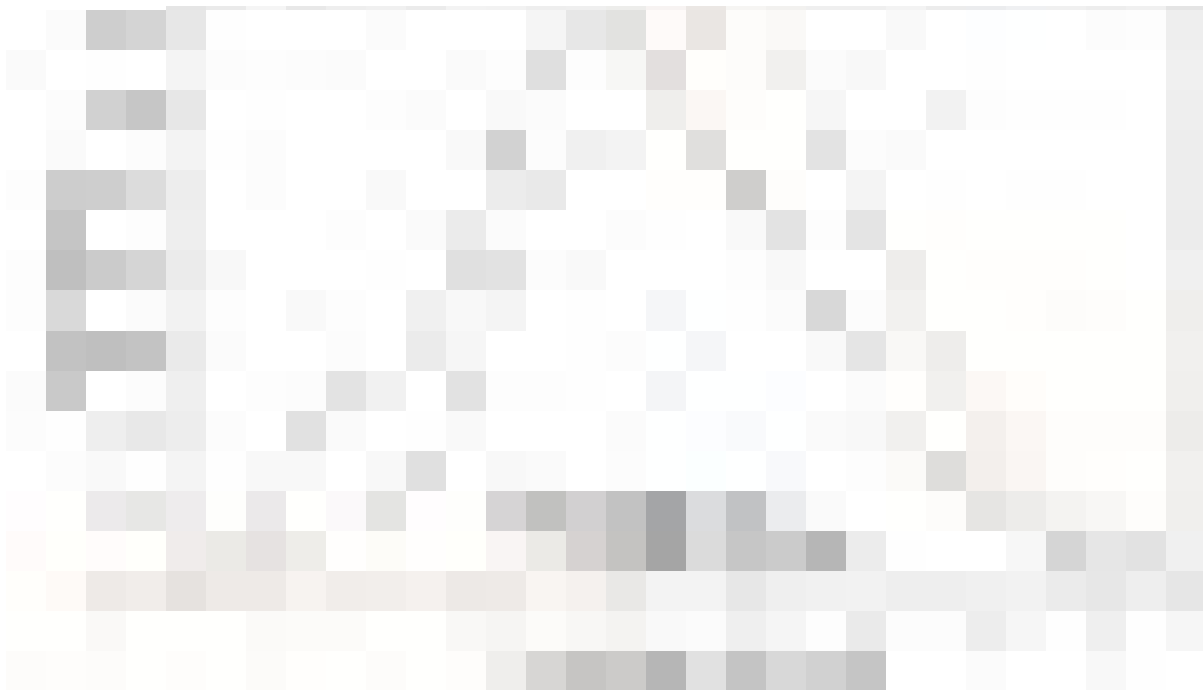
Figure 2: Distribution of time between trains on the Red Line in Boston between 4pm and 5pm; actual distribution as seen by the train operator and biased distribution seen by passengers.

The average time between trains is 7.8 minutes, but the average of the biased distribution is 9.2 minutes, almost 20% longer.

In this case the difference between the two distributions is moderate because the variance of the actual distribution is moderate. When variance is higher, as in the next example, the effect of the inspection paradox can be much bigger.

In 1991, Scott Feld presented the "friendship paradox": the observation that most people have fewer friends than their friends have. He studied real-life social networks, but the same effect appears in online networks: if you choose a random Facebook user and then choose one of their friends at random, the chance is about 80% that the friend has more friends.

The friendship paradox is a form of the inspection paradox. When you choose a random user, every user is equally likely. But when

you choose one of their friends, you are more likely to choose someone with a lot of friends. Specifically, someone with x friends is overrepresented by a factor of x.

Figure 3: Number of online friends for Facebook users: actual distribution and biased distribution seen by sampling friends.

To demonstrate the effect, I use data from the Stanford Large Network Dataset Collection, which includes a sample of about 4000 Facebook users. I compute the number of friends each user has, and the number of friends their friends have. Figure 3 shows both distributions.

The difference is substantial: In this dataset, the average user has 44 friends; the average friend has 104, more than twice as many. And the probability that your friend is more popular than you is 76%.

Some examples of the inspection paradox are more subtle. One of

them occurred to me when I ran a 209-mile relay race in New Hampshire. I ran the sixth leg for my team, so when I started running, I jumped into the middle of the race. After a few miles I noticed something unusual: when I overtook slower runners, they were usually much slower; and when faster runners passed me, they were usually much faster.

At first I thought the distribution of runners was bimodal, with many slow runners, many fast runners, and few runners like me in the middle. Then I realized I was being fooled by the inspection paradox.

In long relay races, runners at different speeds end up spread over the course; if you stand at a random spot and watch runners go by, you see a representative sample of speeds. But if you jump into the race in the middle, the sample you see depends on your speed.

Whatever speed you run, you are more likely to pass slow runners, more likely to be overtaken by fast runners, and unlikely to see anyone running at the same speed as you. Specifically, the chance of seeing another runner is proportional to the difference between your speed and theirs.

We can simulate this effect using data from a conventional road race. Figure 4 shows the actual distribution of speeds from the James Joyce Ramble, a 10K race in Massachusetts. It also shows the biased distribution that would be seen by a runner at 7 mph.

Figure 4: Distribution of speed for runners in a 10K race, and biased distribution as seen by a runner at 7 mph.

In the actual distribution, there are many runners near 7 mph. But if you run at that speed, you are unlikely to see them. The observed distribution has two modes, with fast and slow runners oversampled and fewer runners in the middle.

Even if you are not a runner, you might have noticed the same effect on the highway. You are more likely to see drivers who go too fast or too slow, and less likely to see safe, reasonable drivers like yourself.

A final example of the inspection paradox occurred to me when I read *Orange is the New Black*, a memoir by Piper Kerman, who spent 13 months in a federal prison. Kerman expresses surprise at the length of the sentences her fellow prisoners are serving. She is right to be surprised, but it turns out that she is the victim of not just an inhumane prison system, but also the inspection paradox.

If you arrive at a prison at a random time and choose a random prisoner, you are more likely to choose a prisoner with a long sentence. Once again, a prisoner with sentence x is oversampled by a factor of x.

But what happens if you observe a prison over an interval like 13 months? It works out that if your sentence is y, the chance of overlapping with a prisoner whose sentence is x is proportional to x + y.

Using data from the U.S. Federal Bureau of Prisons, I estimate the actual distribution of sentences for federal prisoners, as would be seen by a judge, the biased distribution as seen by a one-time visitor, and the partially-biased distribution seen by a prisoner with a 13-month sentence. Figure 5 shows the three distributions.

Figure 5: Distribution of federal prison sentences as seen when sentenced, when observed by a random visitor, and when observed by a prisoner, like Piper Kerman, with a 13-month sentence.

In the unbiased distribution, almost 50% of prisoners serve less than one year. But short-timers are less likely to be observed than lifers. To a one-time visitor, fewer than 5% of prisoners have

sentences less than a year.

The distribution seen by a short-time prisoner is only modestly less biased than the view of a one-time visitor. The mean of the actual distribution is 3.6 years; the mean of the biased distribution is almost 13 years. To a 13-month observer, the mean is about 10 years.

The dataset I used for this example is a snapshot of the prison population in July 2019. So the reported distribution is biased; I had to "unbias" it to estimate the actual distribution.

Also, federal prisoners typically serve 85% of their nominal sentence. I took that into account in the calculations.

In summary, the inspection paradox appears in many domains, sometimes in subtle ways. If you are not aware of it, it can cause statistical errors and lead to invalid inferences. But in many cases it can be avoided, or even used deliberately as part of an experimental design.

You can read more about these examples in my books, which are available as free PDFs from Green Tea Press and also published by O'Reilly Media.

- I discuss the class size example in *Think Stats, 2nd Edition* (affiliate link),

- The Red Line example in *Think Bayes* (affiliate link),

- And the friendship paradox in *Think Complexity, 2nd Edition* (affiliate link).

I wrote about relay races and prison sentences in my blog, *Probably Overthinking It*.

The code I used to analyze these examples and generate the figures is in this Jupyter notebook, which is in this repository on GitHub. And you can run the code yourself on Binder.

The figures and numbers in this article are based on random sampling and some assumptions I had to make about the distributions, so they should be considered approximately correct.

John Allen Paulos presents the friendship paradox in Scientific American, "Why You're Probably Less Popular Than Your Friends", 18 January 2011.

The original paper on the topic might be Scott Feld, "Why Your Friends Have More Friends Than You Do", American Journal of Sociology, Vol. 96, №6 (May, 1991), pp. 1464–1477.

Amir Aczel discusses some of these examples, and a few different ones, in a Discover Magazine blog article, "On the Persistence of Bad Luck (and Good)", 4 September 4, 2013.

In *Pythonic Perambulations*, Jake Vanderplas presents Python code to explore the eternal question "Why Is My Bus Always Late?", September 2018.

Allen Downey is a Professor of Computer Science at Olin College in Massachusetts. He is a runner with a maximum 10K speed of 8.7 mph.