# When Predictions Fail: Crash Course Statistics #43

Crash Course: Statistics
https://youtube.com/watch?v=5l60dMUhY5o
https://nerdfighteria.info/v/5l60dMUhY5o

[Complexly theme]

Hi, I'm Adriene Hill, and welcome back to Crash Course Statistics. We've learned a lot about how statistics can help us understand the world better, make better decisions, and guess what'll happen in the future.

Prediction is a big part of how modern statistical analysis is used, and it's helped us make improvements to our lives big and small. But predictions are just educated guesses. We use the information that we have to build up a model of how the world works. A lot of the examples we talked about earlier in the series were making predictions about the present–things like, which coffee shop has better coffee, or how much does an increase in cigarette smoking decrease heart health?

But in this episode, we're gonna focus on using statistics to make predictions about the future. Like, who will win the next World Series, or what stock's gonna do well next month?

Looking back at times when we've failed to make accurate predictions can help us understand more about how to get it right, or whether we just don't have enough information. Today, we're gonna talk about three areas of prediction: markets, earthquakes, and elections. We'll look at why predicting these events can be tricky; why we get it wrong.

[Crash Course theme]

Banks were influential in creating the perfect storm that lead to the 2008 financial crisis. If you've seen "The Big Short" or read the book it's based on, you know that. You also know that Steve Carell should never go blonde again.

The financial crisis is really complicated, and we're gonna simplify it a lot here, but if you're interested, you can check out episode 12 of our Economics series. For now, we're gonna focus on two prediction issues related to the crisis: 1) Overestimating the independence of loan failures, and 2) Economists who didn't see the crisis coming.

So before the crisis, banks were giving out mortgages to pretty much anyone. Normally, banks and lenders in general are choosy about who they lend to. If you give someone a million-dollar loan, and they can't pay it back, you lose out. But banks weren't holding on to that debt; they were selling it to others.

They combined mortgages into groups and sold shares of the loans as mortgage backed securities. The banks knew that some people wouldn't pay their loan in full, but when the mortgages were packaged together, the risk was supposedly mitigated.

Say, there's a 10% chance that each borrower will default on or fail to repay their loan. While not totally risky, it's not ideal for investors. But if you packaged even five similar loans together, the probability that all of them will default is now only 0.001%, because the probability of all of them failing–if each loan failing is independent of another loan failing–is 0.1 to the 5th power.

But we just made a prediction mistake. Many investors overestimated the independence of loan failures. They didn't take into account that if the then-overhauled housing market, and subsequently the economy, began to crumble, the probability of loans going unpaid would shoot way up.

They also had bad estimates of just how risky some of these loans were. Families were losing their homes, and the unemployment rate in the US steadily increased from around 5% to as high as 10% in just a couple of years. There was a global recession that most economists' models hadn't predicted. And to this day, they're still debating exactly why.

Economist John T Harvey claims: "Economics is skewed towards rewarding people for building complex mathematical models, not for explaining how the actual economy works." Others theorize that we need to focus more on people and their sometimes irrational behavior. Wharton finance professor Franklin Allen partly attributes our inability to predict the financial crisis to models that underplayed the impact of banks–the same banks that were involved in the lending practices that helped create, and then deflate, the housing bubble. He claims: "That's a large part of the issue. They simply didn't believe the banks were important."

But they were. Prediction depends a lot in whether or not you have enough data available, but it also depends on what your model deems as "important." You can collect a huge amount of data predicting the rates of diabetes in each country, but if your model only considers hair color, whether or not the person drives a hybrid, and the number of raccoons they think can fight, it probably won't be a good model.

When we create a model to predict things, we're trying to use data, math, and statistics in order to approximate how the world works. We're never going to get it perfect, but if we include most of the important things, we can usually get pretty close.

Even if we can tell what features will be important, it might be hard to get enough data. Earthquakes are particularly difficult to predict. The United States Geological survey even has a webpage dedicated to telling the public that currently, earthquakes just aren't predictable. Clusters of smaller earthquakes often happen before larger ones, but these pre-quakes aren't that helpful in predicting when a big earthquake will hit, because they're almost just as often followed by nothing.

In order to accurately predict an earthquake, you would need three pieces of information: its location, magnitude, and time. It can be relatively easy to get two out of three of those. For example, I predict that there will be an earthquake in the future in Los Angeles, California. And I'd be right, but unless I can also specify an exact time, no one's going to be handing me any honorary degrees in seismology.

We're not bad at earthquake forecasting even if we struggle with accurate earthquake prediction. Earthquake forecasting focuses on the probabilities of earthquakes, usually over longer periods of time. It can also help predict likely effects and damage. This forecasting work is incredibly important for mitigating the sometimes devastating effects of larger earthquakes. For example, scientists might look at the likelihood of severe earthquakes along the San Andreas fault. Their estimates can help inform building codes, disaster plans, and even earthquake insurance rates.

And earthquakes are not without some kind of pattern. They do tend to occur in clusters, with aftershocks following quakes in a pretty predictable pattern. But in his book, "The Signal and the Noise," Nate Silver warns about looking so hard at the data that we see noise–random variation with no pattern–as signal. The causes of earthquakes are incredibly complex, and the truth is, we're not in a place where we can currently predict when, where, and how they'll occur.

Especially the larger, particularly destructive earthquakes. To predict a magnitude 9 earthquake, we'd need to look at data on other similar earthquakes. But there just isn't that much out there. Realistically, it could be centuries before we have enough to make solid predictions.

Even for more common magnitude earthquakes, it could take a lot of data before we have enough to see the pattern amidst all the randomness. Some experts have written off the possibility of accurate earthquake prediction almost entirely, but others hold on to the hope that with enough data and time, we'll figure it out.

Speaking of earthquakes, the 2016 US presidential election results have been described as a political earthquake. Many experts didn't predict the election of Donald Trump. It's easy to forget that predictions are not certain. If we could be 100% certain about anything, we wouldn't really need predictions.

In the past, we've talked about the fact that when predicting percentages, like how many people will vote for one candidate versus the other, there are margins of error. If Candidate A is predicted to get a 54 +/- 2% of the vote, that means that experts predict that Candidate A will get 54% of the vote, but wouldn't be surprised by 52% or 55%. These margins help communicate uncertainty.

But when predictions are discreet–like "will win" or "won't win"– it can be easier to misunderstand this uncertainty. It's possible for predictions to fail without models being bad. Nate Silver discusses the fact that many predictions put Trump's chance of winning the 2016 presidential election at about 1 in 100. Silver's prediction on his website, FiveThirtyEight, put Trump at a much higher chance of about 3 in 10. If you had for statisticians to predict a winner, the smart choice according to these numbers would have been Hillary Clinton.

But here's te problem: Many people see 1 in 100 odds against an event, and take it to mean that the event is essentially impossible. By the numbers, a 1 in 100 chance–even though low– still says the event will happen 1 every 100 times. There's been a lot of debate about how these polls and predictions got it wrong. But one thing we should take away from the election prediction is that low probabilities don't equal impossible events.

If a meticulously curated prediction gives a 1 in 100 chance for a candidate to win, and that candidate wins, it doesn't mean that the prediction was wrong. Unlikely things do happen, and we need to take that into account.

But we should still keep striving to make our polls better. Many who have done post-mortems on the 2016 election polls and predictions attribute some blame to biases in the polls themselves. According to the New York Times: "Well-educated voters are much likelier to take surveys than less educated ones." That means we had a non-response bias from those less educated voters.

Because of that, Nate silver argues that pollsters put too much emphasis on the responses of college-educated voters, who were more likely to vote for Clinton. By improperly weighting them, they overestimated her chance of winning. Prediction isn't easy. Well, making bad predictions is easy. I predict that at the end of this episode, Brandon will bring me ten German chocolate cakes, and I will eat them with my raccoons.

But making good predictions is hard. And even good predictions can be hard to interpret. In order to make accurate predictions, a lot of things need to go right. First, we need good, accurate, and unbiased data, and lots of it. And second, we need a good model. One that takes into account all the important variables.

There's a quote attributed to Confucius that I'm not really sure he said, that goes something like: "To know what you know and what you do not know, that is true knowledge." For example, I know that I don't know that he said that, so I am quite knowledgeable.

There's great value in knowing what we can and can't predict. While we shouldn't stop trying to make good predictions, there's wisdom in recognizing that we won't always get it right. Knowing what we can't accurately predict may be just as important as making accurate predictions. Thanks for watching, I'll see you next time.

[Crash Course credit music]

Crash Course Statistics is filmed in the Chad and Stacy Emigholz Studio in Indianapolis, Indiana, and it's made with the help of all these nice people. Our animation team is Thought Café. If you'd like to keep Crash Course free for everyone forever, you can support the series at Patreon, a crowdfunding platform that allows you to support the content you love. Thank you to all our patrons for your continuing support.

Crash Course is a production of Complexly. If you like content designed to get you thinking, check out some of our other channels at complexly.com. Thanks for watching.