Confidence intervals

Patrick Breheny

January 28

Recap

- In our last lecture, we discussed at some length the Public Health Service study of the polio vaccine
- We discussed the careful design of the study to ensure that human perception and confounding factors could not bias the results in favor of or against the vaccine
- However, there was one factor we could not yet rule out: the role of random chance in our findings

Are our results generalizable?

- Recall that in the study, the incidence of polio was cut by $71/28 \approx 2.5$ times
- This is what we saw in our sample, but remember – this is not what we really want to know
- What we want to know is whether or not we can generalize these results to the rest of the world's population
- The two most common ways of addressing that question are:
  - Confidence intervals
  - Hypothesis testing
- Both methods address the question of generalization, but do so in different ways and provide different, and complimentary, information

Why we would like an interval

- Not to sound like a broken record, but
  - What we know: People in our sample were 2.5 times less likely to contract polio if vaccinated
  - What we want to know: How much less likely would the rest of the population be to contract polio if they were vaccinated?
- This second number is almost certainly different from 2.5 – maybe by a little, maybe by a lot
- Since it is highly unlikely that we got the exactly correct answer in our sample, it would be nice to instead have an interval that we could be reasonably confident contained the true number (the parameter)

What is a confidence interval?

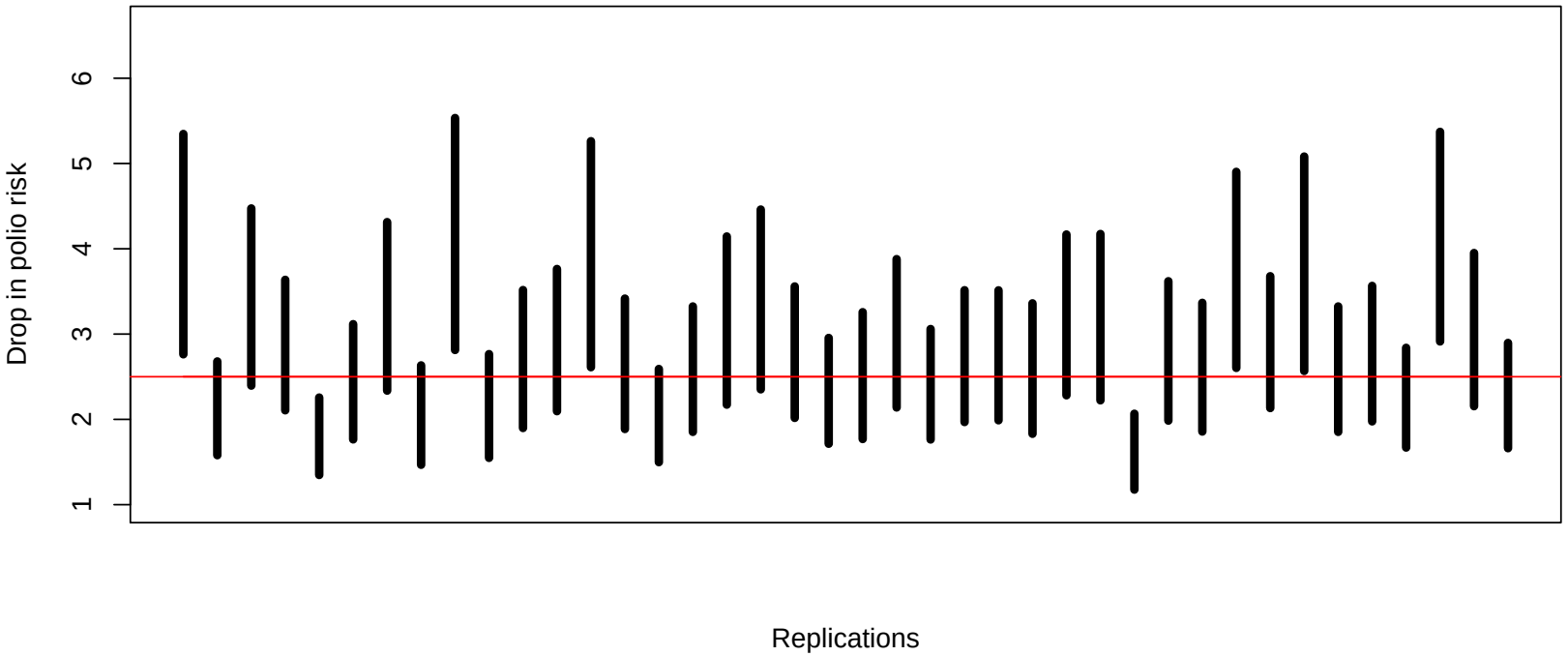
- It turns out that the interval (1.9,3.5) does this job, with a confidence level of 95%
- We will discuss the nuts and bolts of constructing confidence intervals often during the rest of the course
- First, we need to understand what a confidence interval is
- Why (1.9,3.5)? Why not (1.6,3.3)?
- And what the heck does "a confidence level of 95%" mean?

What a 95% confidence level means

- There's nothing special about the interval (1.9,3.5), but there is something special about the procedure that was used to create it
- The interval (1.9,3.5) was created by a procedure that, when used repeatedly, contains the true population parameter 95% of the time
- Does (1.9,3.5) contain the true population parameter? Who knows?
- However, in the long run, our method for creating confidence intervals will successfully do its job 95% of the time (it has to, otherwise it wouldn't be a 95% confidence interval)

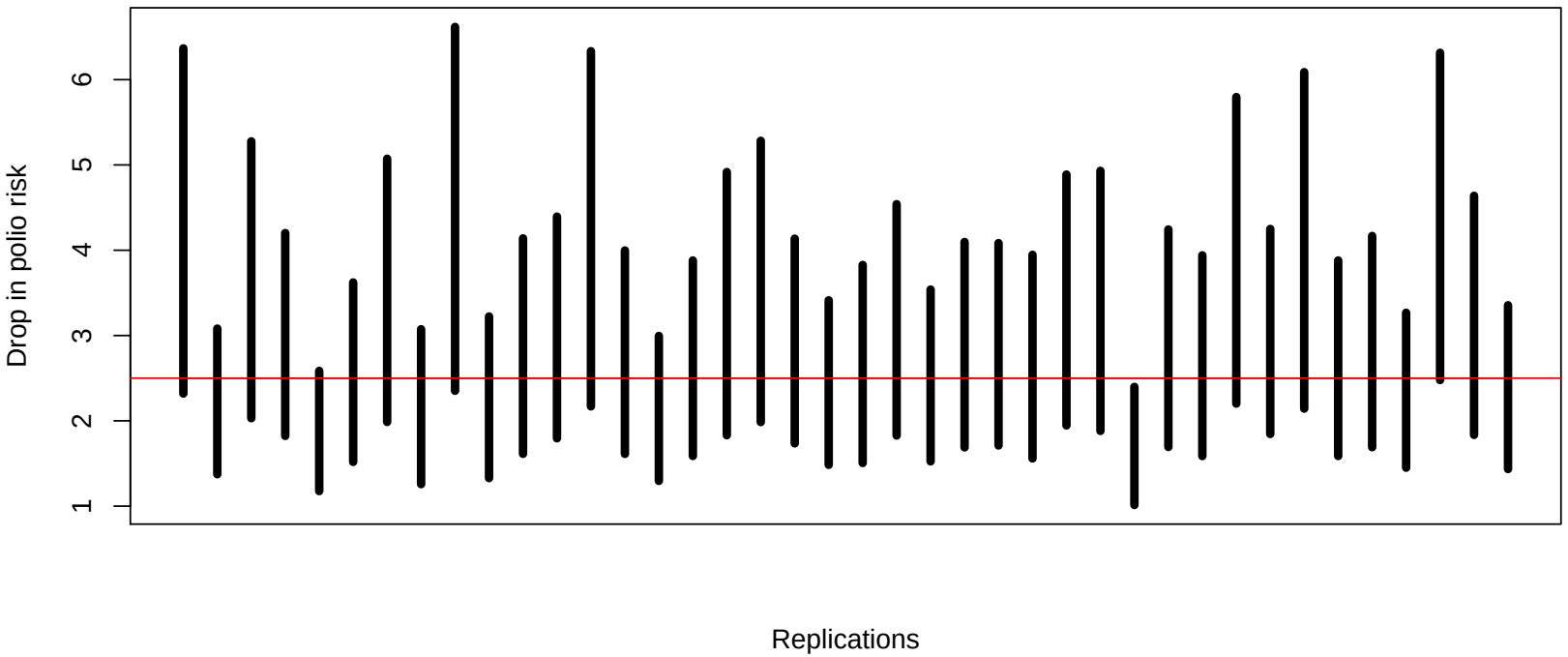
Simulated 80% confidence intervals

Imagine replicating the polio study 40 times (red line = truth):



Simulated 95% confidence intervals

Same studies, same data, difference confidence level:



What's special about 95%?

- The vast majority of confidence intervals in the world are constructed at a confidence level of 95%
- What's so special about 95%?
- Nothing
- However, it does make things easier to interpret when everyone sticks to the same confidence level, and the convention that has stuck in the scientific literature is 95%, so we will largely stick to 95% intervals in this class as well

Consequences

- Thus, if science as a whole goes about constructing these intervals, we can trust that its conclusions will be correct 95% of the time
- This is the sort of long-run guarantee that makes these intervals so appealing to the scientific community
- In reality, however, that percentage is somewhat lower than 95% due to factors such as incorrect assumptions and bias resulting from the experimental design
- For example, a 95% confidence interval for the results of the *Literary Digest* poll would be wrong nearly 100% of the time due to the fundamentally biased nature of the study

The width of a confidence interval

- The width of a confidence interval reflects the degree of our uncertainty about the truth
- Three basic factors determine the extent of this uncertainty, and the width of any confidence interval:
  - The confidence level
  - The amount of information we collect
  - The precision with which the outcome is measured

Confidence levels

- As we saw, the width of a confidence interval is affected by whether it was, say, an 80% confidence interval or a 95% confidence interval
- This percentage is called the *confidence level*
- Confidence levels closer to 100% always produce larger confidence intervals than confidence intervals closer to 0%
- If I need to contain the right answer 95% of the time, I need to give myself a lot of room for error
- On the other hand, if I only need my interval to contain the truth 10% of the time, I can afford to make it quite small

Amount of information

- It is hopefully obvious that the more information you collect, the less uncertainty you should have about the truth
- Doing this experiment on thousands of children should allow you to pin down the answer to a tighter interval than if only hundreds of children were involved
- It may be surprising that the interval is as wide as it is for the polio study: after all, hundreds of thousands of children were involved
- However, keep in mind that a very small percentage of those children actually contracted polio – the 99.9% of children in both groups who never got polio tell us very little about whether the vaccine worked or not
- Only about 200 children in the study actually contracted polio, and these are the children who tell us how effective the vaccine is (note that 200 is a lot smaller than 400,000!)

Precision of measurement

- The final factor that determines the width of a confidence interval is the precision with which things are measured
- I mentioned that the diagnosis of polio is not black and white – misdiagnoses are possible
- Every misdiagnosis increases our uncertainty about the effect of the vaccine
- As another example, consider a study of whether an intervention reduces blood pressure
- Blood pressure is quite variable, so researchers in such studies will often measure subjects' blood pressure several times at different points in the day, then take the average
- The average will be more precise than any individual measurement, and they will reduce their uncertainty about the effect of the treatment

The subtle task of inference

- Inference is a complicated business, as it requires us to think in a manner opposite than we are used to:
  - Usually, we think about what will happen, taking for granted that the laws of the universe work in a certain way
  - When we infer, we see what happens, then try to conclude something about the way that the laws of the universe must work
- This is difficult to do: as Sherlock Holmes puts it in *A Study in Scarlet*, "In solving a problem of this sort, the grand thing is to be able to reason backward."

Confidence interval subtleties

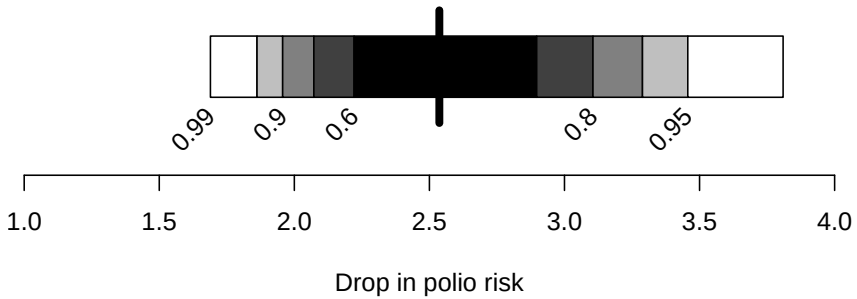
- This subtlety leads to some confusion with regard to confidence intervals – for example, is it okay to say, "There is a 95% probability that the true reduction in polio risk is between 1.9 and 3.5"?
- Well, not exactly – confidence intervals are constructed on the basis that the true reduction is some fixed value, and once we have calculated the interval (1.9,3.5), it's fixed too
- Thus, there's really nothing random anymore – the interval either contains it or it doesn't
- Now, is this an important distinction, or are we splitting hairs here? Depends on who you ask, and we'll talk about this more later in the course

What do confidence intervals tell us?

- So, in the polio study, what does the confidence interval of (1.9,3.5) tell us?
- It gives us a range of likely values by which the polio vaccine cuts the risk of contracting polio: it could cut the risk by as much as 3.5 times less risk, or as little as 1.9 times less risk
- But – and this is critical – it is *unlikely that the vaccine increases the risk, or has no effect, and that what we saw was due to chance*
- Our conclusions may be very different if our confidence interval looked like (0.5,7), in which case our study would be inconclusive

Not all values in an interval are equally likely

- It is important to note, however, that not all values in a confidence interval are equally likely
- The ones in the middle of the interval are more likely than the values toward the edges
- One way to visualize this is with a multilevel confidence bar:



Summary

- There are always a range of values of a parameter (i.e., an interval) that is consistent with the data
- A 95% confidence interval means that the procedure used to construct the interval will contain the true value 95% of the time
- The higher the desired confidence level, the wider we need to make the interval
- The width of a confidence interval is also affected by the amount of information we collect and the accuracy with which we collect it