Confidence intervals

Patrick Breheny

September 10

Are our results generalizable?

- Recall that in the polio vaccine study, the incidence of polio was cut by $71/28 \approx 2.5$ times
- This is what we saw in our sample, but remember that what we *want* to know is whether we can generalize these results to the rest of the 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 (at least from a frequentist perspective)

Why we would like an interval

- Once again,
  - 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 a little, maybe a lot
- Since it is highly unlikely that our estimate is exactly the same as the parameter, 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 technical details 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 does "a confidence level of 95%" mean?

What a 95% confidence level means

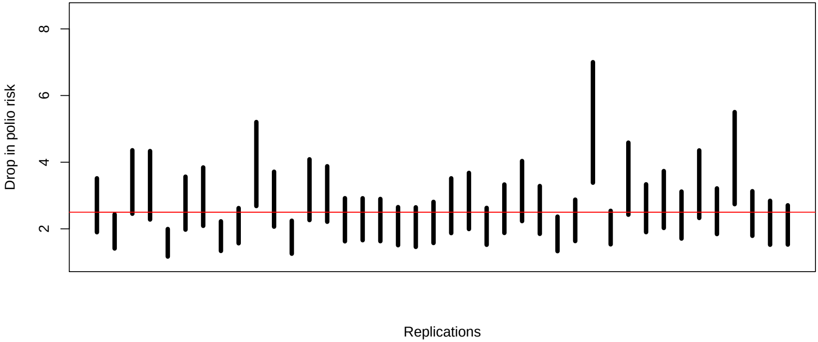
- There's actually nothing special about the interval (1.9,3.5), but there is something special about the procedure that was used to create it
- The interval (1.9,3.5) was created by a procedure that, when used repeatedly, contains the true population parameter 95% of the time
- Does (1.9,3.5) contain 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)

Coverage

- A more rigorous definition: Suppose that $L(\mathrm{Data})$ and $U(\mathrm{Data})$ are functions of the data, and consider the random interval $[L, U]$
- The *coverage probability* of the interval with respect to a parameter θ is the probability that $\theta \in [L, U]$ over random samples of data
- The interval $[L, U]$ thus forms a *confidence interval* for θ with *confidence level* $1 - \alpha$ if its coverage probability is at least $1 - \alpha$
- Note that this is very much a long-run frequency justification: we are given no guidelines for how to construct these intervals – any procedure can produce a confidence interval if it contains its target $100(1 - \alpha)\%$ of the time

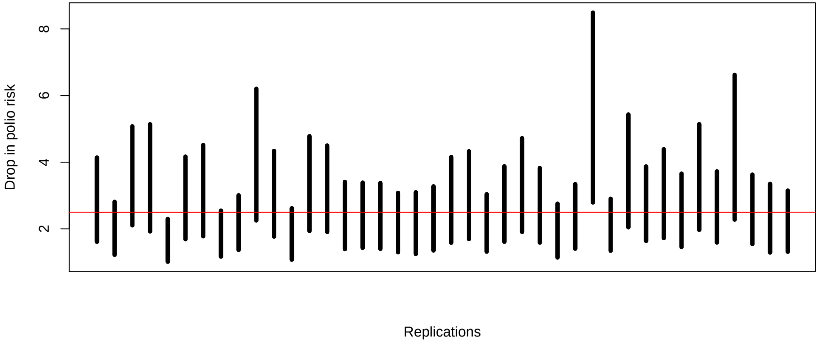
Simulated 80% confidence intervals

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 undoubtedly 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 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

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
- Note that the Bayesian interpretation doesn't run into any issues here: parameters *are* random if you don't know what they are, and thus a statement like the above is perfectly valid
- According to the long-run frequency interpretation, though, it has no meaning

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 thus it is *unlikely that the reduction 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

Summary

- There is 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