Standard Error

Every statistic has a standard error associated with it.

- Not always reported and not always easy to calculate.
- Example: Waiting times

A measure of the (in)accuracy of the statistic.

- A standard error of 0 means that the statistic has *no* random error.
- The bigger the standard error, the less accurate the statistic.

Implicit in this the idea that anything we calculate in a sample of data is subject to random errors.

- The mean we calculated for the waiting times is not the *true* mean, but only an estimate of the true mean.
- Even if we could perfectly replicate our study, we would get a different value for the mean.

What are the sources of error?

- Classic approach in statistics: our data set may be only a random sample from some larger population.
- We may make errors of measurement.
- There are lots of other random factors affecting our outcome that we can't control.

The standard error of a statistic is the standard deviation of that statistic across hypothetical repeated samples.

- Example: 100 replications of waiting time study.
- In theory, need to replicate an infinite number of times.

The standard errors that are reported in computer output are only estimates of the true standard errors.

- Remarkably, we can estimate the variability across repeated samples by using the variability between samples.
- The more variability within the sample, the more variability between samples.
- The formula for the standard error of the mean is $\frac{s}{\sqrt{n}}$, i.e., the standard deviation divided by the square root of the sample size.

In general, the bigger the sample, the smaller the standard error.

- Why? Big samples give us more information to estimate the quantity we're interested in.
- The standard error generally goes down with the square root of the sample size. Thus, if you quadruple the sample size, you cut the standard error in half.

Confidence Intervals

The standard error is often used to construct confidence intervals.

- To construct a 95 percent confidence interval around the mean, add two standard errors and subtract two standard errors.
- E.g., for the waiting time example, the mean was approx. 20 and its standard error was 1. Then the upper confidence limit is 22 and the lower confidence limit is 18.
- Interpretation: we can be 95 percent confident that the true mean is somewhere between 18 and 22.
- Further interpretation: Suppose we could replicate our study many times. For each replication we could construct a 95 percent confidence interval by adding and subtracting 2 standard errors from the mean. Then 95 percent of those confidence intervals would contain the true mean.

Why two standard errors? Remember our rule for normal distributions: 95% of the cases fall within two standard deviations of the mean.

- Even though the original distribution of waiting times was not well approximated by a normal distribution, the distribution of means across repeated samples *is* approximately normal.

- Why? *Central limit theorem*: Whenever you average a bunch of things together, the resulting average tends to be approximately normally distributed. The more things you add together, the closer the approximation.

- In large samples, most statistics have approximately a normal distribution across repeated samples.