UNIVERSITY OF
WATERLOO

Faculty of Mathematics
Waterloo, Ontario N2L 3G1

Centre for Education in
Mathematics and Computing

Grade 7/8 Math Circles

February 24, 2021

Confidence Intervals

Introduction

Confidence intervals are an important tool that help statisticians communicate their estimate of a true value based on collected data. While it would be handy to have a complete set of accurate data about the population of interest, this is rarely possible. Instead, confidence intervals help us to identify a set of plausible values that the number we're looking for could be.

This lesson gives an overview of the concept of a confidence interval rather than delving into how different calculations work, but it's worth noting that what those calculations are can be very different depending on what kind of information is available.

Review

Here's a quick recap of some concepts from last week about random sampling: `https://www.geogebra.org/m/jw4u2sxd`

Point Estimators

Example: Let's say we want to know the mean weight of the strawberries grown at the local field. It wouldn't be practical to pick and weigh every single strawberry, so we'll take a sample. We decide on a sample of 50 strawberries, which together weigh in at 500 grams, giving us a mean (average) weight of $\frac{500}{50} = 10$g each. This number is our statistic, the *sample mean*.

- **sample mean**—a statistic; the mean value of a characteristic within a sample of the population, used to estimate the population mean
- **sample proportion**—a statistic; the fraction of individuals within a sample of the population with a certain characteristic, used to estimate the population proportion

- **population mean**—a parameter; the mean value of a characterstic within the population
- **population proportion**—a parameter; the fraction of individuals in a population with a certain characterstic

Since we picked a random, representative sample, we can now use this statistic to estimate the parameter, the *population mean* of the weights of all strawberries in the field. What number makes sense to choose as our best guess?

Since we have a statistic, the sample mean, we can use it as a *point estimator* for the population mean. That gives us 10g as our *point estimate*—an exact value that we estimate for the parameter.

- **point estimator**—a statistic that provides a reasonable, unbiased guess for a parameter
- **point estimate**—the value of a point estimator

The sample mean and sample proportion are two common *point estimators*.

Question: Three of the 20 students in your class are vegetarian, and there are 100 students in your grade in total. What is the sample proportion of vegetarians? What are the point estimator and point estimate for the population proportion of vegetarians in your grade?

Solution: The sample proportion is $\frac{3}{20}$. The point estimator is the sample proportion, and the point estimate is $\frac{3}{20} = \frac{15}{100}$.

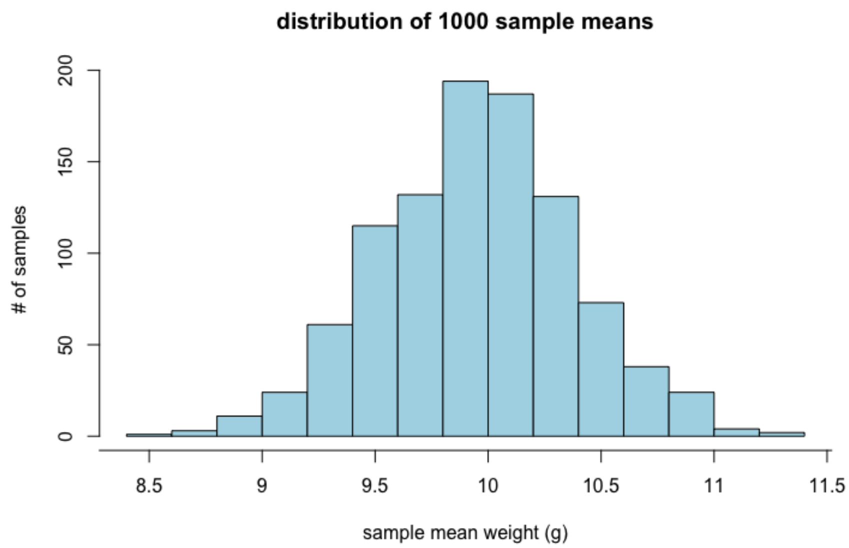
Estimating With Intervals

In the case of our example, we know that the mean weight of strawberries in our field is probably somewhere around 10g. But how close to 10g is the unknown population mean? And how sure are we about how close it is?

We'll answer those questions by asking a third one: Using the same point estimator (the sample mean), how would our point estimate change if we had taken more samples?

Example: You can explore the results of taking multiple samples using this program: `https://repl.it/@cemc/strawberry-field#main.r`. (If the link doesn't work, try copy-pasting into your browser!) Each time you click "run," a new random sample is generated. The samples are all based off of our unknown parameter, the population mean.

When we take multiple samples, the individual units included in our sample will change, meaning that the values of the statistics we measure will change. If we were to take many, many samples, we would find that a graph of sample means would look something like this:



Examining the graph above, which is called a *sampling distribution*, we notice that most of the sample means look like they sit in the middle section of the distribution. The population mean might also be somewhere around the middle!

The sample mean from our example is 10g, which is within 1g of nearly all of the other sample statistics in the sampling distribution we've generated. If our sample mean, 10g, is in fact within 1g of the population mean, then the population mean is also within 1g of the sample mean! This means that we could take an interval of 9–11g, or 10±1g, and it's *plausible* that the population mean is somewhere in that interval.

This line of reasoning applies to all possible samples and corresponding sample means from the entire population of strawberries. Whenever the sample mean is within 1g of the population mean, the vice versa is also true—so if we were to take an interval of 1g on each side of the sample mean, in most cases, the population mean would fall somewhere in that interval! This is the core principle behind the idea of *confidence intervals*: choosing an interval based on sample statistics to provide a range of values to estimate the parameter.

We can now revisit our first question: **How close is our statistic to the parameter?**

Example: We now have a confidence interval of 10±1g. The ± here is the key to answering this question—we are saying that we believe that it is plausible that the parameter value is within 1g of our sample mean, 10g. In other words, we are stating that 1g is our *margin of error* for a certain level of confidence.

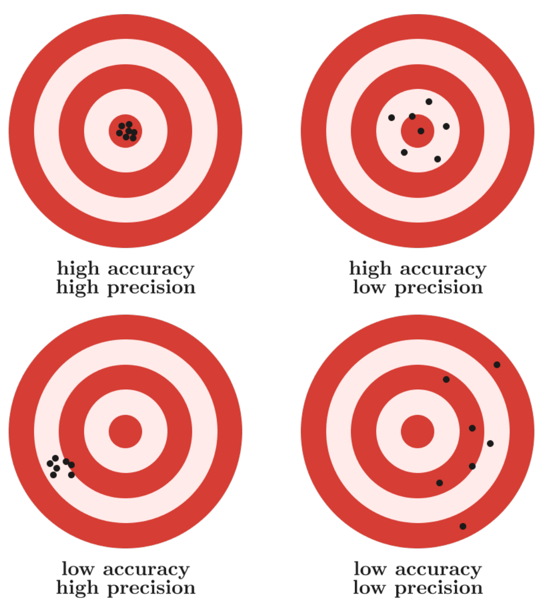
Generally, confidence intervals can be interpreted in the form: ***point estimate*** ± ***margin of error***. The point estimate is our best guess for the unknown parameter value, and the margin of error communicates how close we believe that guess to be, based on other information we have, such as the size of the sample.

Precision & Accuracy

Precision vs. accuracy is an important distinction to understand when learning about how and why confidence intervals work.

- **accuracy**—how close a measurement/estimate is to the true value
- **precision**—how close measurements/estimates are to each other

A classic example is throwing darts at a target: accuracy is landing close to the bullseye, whereas precision is landing close to each other!



We can compare using sample statistics to estimate population parameters to throwing darts at a target hidden behind a curtain. We can see where a dart lands, but we don't know for sure how close it is to the bullseye. Producing a statistic is like throwing one dart. Choosing an interval is like throwing a dart, and then drawing a circle around it to include a bunch of points around the one you hit. This makes the guess less precise by including more points, but also more accurate: the bullseye is definitely more likely to be in the circle you drew rather than at the exact point you hit!

Example: Continuing with our interval of of 9–11g, we can think about how choosing this range of values rather than a single point estimate changes the precision and accuracy of our guess. Choosing a range of values like this decreases the *precision* of the guess but increases the *accuracy*—after all, the true mean of the weight of the strawberries is certainly more likely to be between 9–11g rather than exactly 10g.

Question: What would happen to the precision and accuracy of our interval if we were to decrease our margin of error?

Solution: Decreasing the margin of error would make our confidence interval narrower. This decreases the accuracy, but also increases the precision.

Confidence Levels

When we construct confidence intervals, we find *C% confidence intervals* which have *confidence level* C%. For example, we might construct a 95% confidence interval for an unknown parameter, which has a confidence level of 95%.

The confidence level of a confidence interval communicates the success rate of the process that lead to calculating the confidence interval. So, if we were to calculate many C% confidence intervals with different same-size random samples from a given population, approximately C% of those confidence intervals would capture the unknown parameter value.

Example: After taking a random sample of 100 plants, a farmer constructs a 90% confidence interval of $\frac{85}{100}$ to $\frac{95}{100}$ for the proportion of plants that yield a harvest. This does *not* mean that, given the sample, there is a 90% chance that the confidence interval captures the parameter value—once a sample is selected, the confidence interval produced either is or is not successful. Rather, it means that overall, 90% of all the samples that could have been selected would yield successful confidence intervals.

This answers our second question: **How sure are we about how close the statistic is to the parameter?**

Because the parameter is unknown, we have no way of ever knowing whether the individual confidence interval we've constructed really does contain the true parameter value. We only know that for C% of the possible samples we could have drawn from the population, the C% confidence interval produced captures the unknown parameter.

Question: A farmer takes a random sample of their crops to estimate the proportion of plants that yield a harvest. They produce a 90% confidence interval for their proportion. If they were to take 500 samples and construct a 90% confidence interval for each, approximately how many of those confidence intervals would capture the unknown parameter value, the population proportion?

Solution: Approximately $0.90 \times 500 = 450$ of the intervals would capture the unknown parameter value.

What Confidence Intervals Mean

Once we have a confidence interval, we communicate our findings by saying that "We are C% confident that the interval constructed captures the unknown parameter value for the population."

The values that we include inside of a confidence interval are what we then consider to be *plausible* values, which are reasonable estimates for the parameter given the information we have.

Question: A polling company reports that from their survey of a sample of voters, the Purple candidate is leading the race with 52%±3 of voters casting their ballots for purple, 19 times out of 20. Purple must receive more than 50% of the votes to win. Is it plausible for Purple to lose the election? What confidence level is being used?

Solution: It is still plausible. The 95% confidence interval constructed goes from 49–55%, which includes values under 50% that could still plausibly be the true proportion of people who voted for Purple. The confidence level is $\frac{19}{20} = 0.95$, meaning a 95% confidence interval from 49–55%.

Confidence intervals, like the rest of statistics, don't give us perfect answers. We've seen that we can only be C% sure that a given interval captures the parameter; $(100 - C)\%$ of the time, we'll miss it completely! It's important to stay aware of both the advantages and drawbacks of relying on confidence intervals and use and interpret them wisely.