Interval Estimation

Ryan Miller

Sampling and Statistical Inference

- A *fundamental goal* of statisticians is to use information from a sample to make *reliable* statements about a population
  - This idea is called **statistical inference**

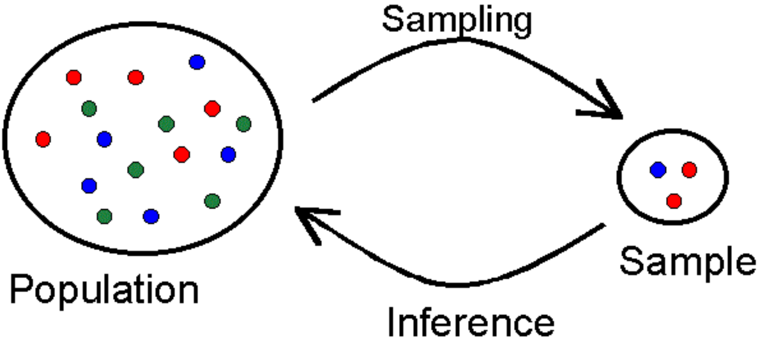

Image credit: http://testofhypothesis.blogspot.com/2014/09/the-sample.html

Statistical Inference - Notation

Statisticians use different notation to distinguish *population parameters* (things we want to know) from *estimates* (things derived from a sample). For a few common measures, this notation is summarized below:

	Population Parameter	Estimate (from sample)
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Proportion	p	$\hat{p}$
Correlation	ρ	r
Regression	β_0, β_1	b_0, b_1

For example, μ is the mean of the target population, while $\bar{x}$ is the mean of the cases that ended up in the sample

Point Estimation

- If a sampling protocol is *unbiased*, the sample average is a sensible estimate of the population mean
    - This is called a **point estimate**, referring to the fact that it is a single value

Point Estimation

- If a sampling protocol is *unbiased*, the sample average is a sensible estimate of the population mean
  - This is called a **point estimate**, referring to the fact that it is a single value
- From our study of *sampling distributions*, we know that the existence of sampling variability means a point estimate is almost certainly wrong (at least to some degree)
  - This suggests that we can more appropriately describe what we think is true of the population by reporting an **interval estimate** that accounts for *sampling variability*

Point vs. Interval Estimation

To summarize:

- **Point estimation** uses sample data to produce a *single "most likely" estimate* of a population characteristic, which will almost always miss the target (at least by some degree)
- **Interval estimation** uses sample data to produce a *range of plausible estimates* of a population characteristic, an approach that has a much better chance at capturing the truth

Point vs. Interval Estimation

To summarize:

- **Point estimation** uses sample data to produce a *single "most likely" estimate* of a population characteristic, which will almost always miss the target (at least by some degree)
- **Interval estimation** uses sample data to produce a *range of plausible estimates* of a population characteristic, an approach that has a much better chance at capturing the truth

A popular analogy:

> *Using only a point estimate is like fishing in a murky lake with a spear. We can throw a spear where we saw a fish, but we will probably miss. On the other hand, if we toss a net in that area, we have a good chance of catching the fish.*

Margin of Error

Most interval estimates have the form:

$$\text{Point Estimate} \pm \text{Margin of Error}$$

We often report these intervals using only their endpoints:

$$(\text{Est} - \text{MOE}, \text{Est} + \text{MOE})$$

Margin of Error

Most interval estimates have the form:

$$\text{Point Estimate} \pm \text{Margin of Error}$$

We often report these intervals using only their endpoints:

$$(\text{Est} - \text{MOE}, \text{Est} + \text{MOE})$$

- We'd like the *margin of error* to be constructed in way that carries a *quantifiable* claim of precision
  - ie: 80% of the time an interval with this type of margin of error will contain the population characteristic
  - Without an accompanying claim regarding precision, reporting a margin of error is not particularly useful

Statistical Inference

So, what can we say about a population proportion, p, based upon an observed sample proportion, $\hat{p}$? Consider a representative sample of 100 infants used to estimate the proportion of all babies who are born prematurely

- True or false? "We observed $\hat{p} = 0.14$, so we know that 14% of all babies are born prematurely"

Statistical Inference

So, what can we say about a population proportion, p, based upon an observed sample proportion, $\hat{p}$? Consider a representative sample of 100 infants used to estimate the proportion of all babies who are born prematurely

- True or false? "We observed $\hat{p} = 0.14$, so we know that 14% of all babies are born prematurely"
  - False - point estimates have variability

Statistical Inference

So, what can we say about a population proportion, p, based upon an observed sample proportion, $\hat{p}$? Consider a representative sample of 100 infants used to estimate the proportion of all babies who are born prematurely

- True or false? "We observed $\hat{p} = 0.14$, so we know that 14% of all babies are born prematurely"
  - False - point estimates have variability
- True or false? "We observed $\hat{p} = 0.14$, it's probably true 14% of all babies are born prematurely"

Statistical Inference

So, what can we say about a population proportion, p, based upon an observed sample proportion, $\hat{p}$? Consider a representative sample of 100 infants used to estimate the proportion of all babies who are born prematurely

- True or false? "We observed $\hat{p} = 0.14$, so we know that 14% of all babies are born prematurely"
  - False - point estimates have variability
- True or false? "We observed $\hat{p} = 0.14$, it's probably true 14% of all babies are born prematurely"
  - False - sampling distributions show the point estimate is likely off by some degree

Statistical Inference

So, what can we say about a population proportion, p, based upon an observed sample proportion, $\hat{p}$? Consider a representative sample of 100 infants used to estimate the proportion of all babies who are born prematurely

- True or false? "We observed $\hat{p} = 0.14$, so we know that 14% of all babies are born prematurely"
  - False - point estimates have variability
- True or false? "We observed $\hat{p} = 0.14$, it's probably true 14% of all babies are born prematurely"
  - False - sampling distributions show the point estimate is likely off by some degree
- True or false? "Although we don't know p, if we attach a large margin error to our point estimate,the interval estimate $14\% \pm 10\% = (4\%, 24\%)$ probably contains p"

Statistical Inference

So, what can we say about a population proportion, p, based upon an observed sample proportion, $\hat{p}$? Consider a representative sample of 100 infants used to estimate the proportion of all babies who are born prematurely

- True or false? "We observed $\hat{p} = 0.14$, so we know that 14% of all babies are born prematurely"
  - False - point estimates have variability
- True or false? "We observed $\hat{p} = 0.14$, it's probably true 14% of all babies are born prematurely"
  - False - sampling distributions show the point estimate is likely off by some degree
- True or false? "Although we don't know p, if we attach a large margin error to our point estimate,the interval estimate $14\% \pm 10\% = (4\%, 24\%)$ probably contains p"
  - False - we don't know how reliable this margin of error is, perhaps an MOE of 10% is not wide enough

Conclusion

- This presentation introduces the idea of interval estimation
  - The key concept is that point estimates are almost always off, but by attaching a margin of error we can more reliably describe the population of interest
- In class this week, we'll further explore this concept and learn how to use sampling distributions to come up with interval estimates that have *meaningful margins of error*