Review of key points about estimators

- Populations can be at least partially described by population parameters
- Population parameters include: mean, proportion, variance, etc.
- Because populations are often very large (maybe infinite, like the output of a process) or otherwise hard to investigate, we often have no way to know the exact values of the parameters
- *Statistics* or *point estimators* are used to estimate population parameters
- An estimator is calculated using a function that depends on information taken from a sample from the population
- We are interested in evaluating the “goodness” of our estimator - topic of sections 8.1-8.4
- To evaluate “goodness”, it’s important to understand facts about the estimator’s sampling distribution, its mean, its variance, etc.
Different estimators are possible for same parameter

- In everyday life, people who are working with the same information arrive at different ideas/decisions based on the same information.
- Given the same sample measurements/data, people may derive different estimators for the population parameter (mean, variance, etc.).
- For this reason, we need to evaluate the estimators on some criteria (bias, etc.) to determine which is best.
- Complication: the criteria that are used to judge estimators may differ.
- Example: For estimating $\sigma^2$ (variance), which is better:
  
  $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ (sample variance) or some other estimator
  
  $s^{*2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ (which more closely resembles population variance).
Repeated estimation yields sampling distribution

- If you use an estimator once, and it works well, is that enough proof for you that you should always use that estimator for that parameter?
- Visualize calculating an estimator over and over with different samples from the same population, i.e. take a sample, calculate an estimate using that rule, then repeat
- This process yields sampling distribution for the estimator
- We look at the mean of this sampling distribution to see what value our estimates are centered around
- We look at the spread of this sampling distribution to see how much our estimates vary
bias

We may want to make sure that the estimates are centered around the parameter of interest (the population parameter that we’re trying to estimate).

One measurement of center is the mean, so may want to see how far the mean of the estimates is from the parameter of interest → bias

Assume we’re using the estimator \( \hat{\theta} \) to estimate the population parameter \( \theta \)

\[ Bias(\hat{\theta}) = E(\hat{\theta}) - \theta \]

If bias equals 0, the estimator is *unbiased*

Two common unbiased estimators are:

1. Sampling proportion \( \hat{p} \) for population proportion \( p \)
2. Sample mean \( \bar{X} \) for population mean \( \mu \)
Bias and the sample variance

What is the bias of the sample variance, $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$? Contrast this case with that of the estimator $s^*^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$, which looks more like the formula for population variance.
Variance of an estimator

- Say your considering two possible estimators for the same population parameter, and both are unbiased.

- Variance is another factor that might help you choose between them.

- It’s desirable to have the most precision possible when estimating a parameter, so you would prefer the estimator with smaller variance (given that both are unbiased).

- For two of the estimators that we have discussed so far, we have the variances:
  1. \( Var(\hat{p}) = \frac{p(1-p)}{n} \)
  2. \( Var(\bar{X}) = \frac{\sigma^2}{n} \)
Mean square error of an estimator

- If one or more of the estimators are biased, it may be harder to choose between them.

- For example, one estimator may have a very small bias and a small variance, while another is unbiased but has a very large variance. In this case, you may prefer the biased estimator over the unbiased one.

- Mean square error (MSE) is a criterion which tries to take into account concerns about both bias and variance of estimators.

- $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \rightarrow$ the expected size of the squared error, which is the difference between the estimate $\hat{\theta}$ and the actual parameter $\theta$
MSE can be re-stated

Show that the MSE of an estimate can be re-stated in terms of its variance and its bias, so that

\[ MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2 \]
Moving from one population of interest to two

- Parameters and sample statistics that have been discussed so far only apply to one population. What if we want to compare two populations?
- Example: We want to calculate the difference in the mean income in the year after graduation between economics majors and other social science majors
  \[ \mu_1 - \mu_2 \]
- Example: We want to calculate the difference in the proportion of students who go on to grad school between economics majors and other social science majors
  \[ p_1 - p_2 \]
Comparing two populations

- Try to develop a point estimate for these quantities based on estimators we already have
- For the difference between two means, \( \mu_1 - \mu_2 \), we try the estimator \( \bar{x}_1 - \bar{x}_2 \)
- For the difference between two proportions, \( p_1 - p_2 \), we try the estimator \( \hat{p}_1 - \hat{p}_2 \)

We want to evaluate the “goodness” of these estimators.

- What do we know about the sampling distributions for these estimators?
- Are they unbiased?
- What is their variance?
Show that $\bar{x}_1 - \bar{x}_2$ is an unbiased estimator for $\mu_1 - \mu_2$.
Also show that the variance of this estimator is $\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$.
Mean and variance of \( \hat{p}_1 - \hat{p}_2 \)

Show that \( \hat{p}_1 - \hat{p}_2 \) is an unbiased estimator for \( p_1 - p_2 \).
Also show that the variance of this estimator is
\[
\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}
\]
Summary of two sample estimators

- We have just shown that $\bar{x}_1 - \bar{x}_2$ and $\hat{p}_1 - \hat{p}_2$ are unbiased estimators, as were $\bar{x}$ and $\hat{p}$
- The CLT doesn’t apply to these estimators since they are not sample means - they are differences of sample means
- Other theorems do state that given at least moderate ($n \geq 30$) sample sizes, these estimators have sampling distributions that are approximately normal
Estimation errors

- Even with a good point estimate $\hat{\theta}$, there is very likely to be some error ($\hat{\theta} = \theta$ not likely)
- We can express this error of estimation, denoted $\varepsilon$, as $\varepsilon = |\hat{\theta} - \theta|
- This is the number of units that our estimate, $\hat{\theta}$, is off from $\theta$ (doesn’t take into account the direction of the error)
- We can use the sampling distribution of $\hat{\theta}$ to help place some bounds on our estimate