Vanderbilt University Political Science - Stats I
Fall 2024 - Prof. Jim Bisbee

Lecture 8

8.1 The Intuitive Estimator Is Not Always the Unbiased Estimator!

- You are probably not surprised to learn that $\overline{Y}$ is an unbiased estimator for μ while $\overline{Y}_B$ is a biased estimator. After all, $\overline{Y}$ is the mean of the sample and so intuitively it seems like it should be unbiased estimator for the population mean. The same for $\overline{Y_1} - \overline{Y_2}$ as an estimator of $\mu_1 - \mu_2$. These are *intuitive* estimators.

- HOWEVER, it is not always the case that the intuitive estimator is the unbiased estimator. We'll illustrate this in a way that explains the distinction statisticians draw between sample variance and population variance.

- It would seem natural to estimate the variance of a population, σ^2, with the sample variance $S^2 = \frac{\sum_i \left(Y_i - \overline{Y}\right)^2}{n}$. But we can show that this is in fact a *biased* estimator for σ^2:

$$
\begin{aligned}
E\left(S^2\right) &= E\left(\frac{\sum_i \left(Y_i - \overline{Y}\right)^2}{n}\right) \\
&= E\left[\frac{\sum_i (Y_i)^2}{n} - \left(\overline{Y}\right)^2\right] \text{ [remember this from PS 1?]} \\
&= \frac{1}{n} E\left[\sum_i (Y_i)^2 - n\left(\overline{Y}\right)^2\right] \text{ [multiplying terms by } \frac{n}{n}\text{]} \\
&= \frac{1}{n}\left\{\left(\sum_i E\left[(Y_i)^2\right]\right) - nE\left[\left(\overline{Y}\right)^2\right]\right\} \text{ [distributing expectations]} \\
&= \frac{1}{n}\left\{\left(\sum_i VAR(Y) + [E(Y)]^2\right) - n\left(VAR(\overline{Y}) + \left[E\left(\overline{Y}\right)\right]^2\right\}
\end{aligned}
$$

[using the identity $VAR(Y) = E(Y^2) - [E(Y)]^2$, and identicality of Y_i]

$$
\begin{aligned}
&= \frac{1}{n}\left\{\left(\sum_i \sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\} \text{ [substituting identities]} \\
&= \frac{1}{n}\left\{n\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\} \\
&= \frac{1}{n}\left\{n\sigma^2 - \sigma^2\right\} = \frac{n-1}{n}\sigma^2 \neq \sigma^2.
\end{aligned}
$$

- Now consider an alternate estimator, $S_U^2 = \frac{\sum_i \left(Y_i - \overline{Y}\right)^2}{n-1}$. Well, $E\left(S_U^2\right) = \frac{1}{n-1} E\left[\sum_i \left(Y_i - \overline{Y}\right)^2\right] = \frac{n-1}{n-1}\sigma^2 = \sigma^2$ [will be on homework] and hence is unbiased.

- When describing the variance of a sample, we will write S^2 and use the formula we've been using so far this semester. But when describing the unbiased estimator of a population variance, we will write S_U^2 and use this new formula. Note the difference between this practice and your text.

- But we are probably making a mountain out of a molehill. Because what is $B(S^2)$? What is $\lim_{n\to\infty} B(S^2)$? What is the implication of this?

$$\begin{aligned}
B(S^2) &= E\left(S^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\sigma^2 - \sigma^2 \\
&= \left(\frac{n-1}{n} - 1\right)\sigma^2 \\
&= -\sigma^2/n. \\
\text{So} &: \lim_{n\to\infty} B(S^2) = \lim_{n\to\infty} -\sigma^2/n = 0.
\end{aligned}$$

- Make sure to talk notation: we'll use the notation S_U^2 to specify the unbiased estimator for the population variance, where $S_U^2 \equiv \frac{\sum_i \left(Y_i - \overline{Y}\right)^2}{n-1}$.

8.2 Confidence Intervals

- Last time we talked about point estimators and two desirable properties: unbiasedness and relatively small variance. Today we'll discuss another kind of estimator: interval estimators.

- An **interval estimator** is:

  - a rule
  - specifying how we use the sample to calculate two numbers
  - that form the endpoints of an interval
  - containing/trapping/enclosing a parameter of interest, θ.

- Intervals have two desirable properties. We want them to:
  - Contain the parameter of interest, θ
  - Be relatively narrow (sound familiar?)
- As with point estimators, the length and location of the interval are random quantitites, so our goal is to find an interval estimator that generates narrow intervals with a high probability of trapping θ.
- [Distribute handout about here.]
- Interval estimators are commonly called **confidence intervals (CIs)**.
  - CIs are constructed of two quantities called the **upper** and **lower confidence limits** (or **upper** and **lower bounds).**
  - The probability that a random CI will enclose θ is called the **confidence coefficient:** it is:
    * the fraction of the time,
    * in repeated sampling,
    * that the CI will contain θ.
  - We thus like confidence coefficient associated with our CI to be high.
  - The confidence coefficient is written $(1-\alpha)$. If $\widehat{\theta}_L$ and $\widehat{\theta}_H$ are the random lower and upper confidence limits, then

$$P(\widehat{\theta}_L \leq \theta \leq \widehat{\theta}_U) = (1-\alpha).$$

- We typically call a CI with confidence coefficient $(1-\alpha)$ a "$[100 \cdot (1-\alpha)]-$percent confidence interval." Sometimes we also say a "CI with alpha=[α]."
- Here we discuss how to construct a confidence interval for a sample statistic $\widehat{\theta}$ that is Normally distributed with mean μ and standard error $\sigma_{\widehat{\theta}}$. (What would be an example of such a statistic? The CLT tells us that one example would be Y-bar constructed from a large sample.)

- Let's standardize the statistic as follows:

$$Z = \frac{\widehat{\theta} - \theta}{\sigma_{\widehat{\theta}}},$$

  Now it is distributed approxmately standard Normal, as we have substracted the estimator's hypothesized mean and divided by its standard deviation.

- To construct a confidence interval for $\widehat{\theta}$, pick two values $-z_{\alpha/2}, z_{\alpha/2}$ such that

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = \int_{-z_{\alpha/2}}^{z_{\alpha/2}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1 - \alpha.$$

- Substituting for Z, we have

$$\begin{aligned} P(-z_{\alpha/2} &\leq \frac{\widehat{\theta} - \theta}{\sigma_{\widehat{\theta}}} \leq z_{\alpha/2}) = 1 - \alpha \\ P(-z_{\alpha/2}\sigma_{\widehat{\theta}} &\leq \widehat{\theta} - \theta \leq z_{\alpha/2}\sigma_{\widehat{\theta}}) = 1 - \alpha \\ P(\widehat{\theta} - z_{\alpha/2}\sigma_{\widehat{\theta}} &\leq \theta \leq \widehat{\theta} + z_{\alpha/2}\sigma_{\widehat{\theta}}) = 1 - \alpha. \end{aligned}$$

- And therefore $\widehat{\theta}_L = \widehat{\theta} - z_{\alpha/2}\sigma_{\widehat{\theta}}$, and $\widehat{\theta}_H = \widehat{\theta} + z_{\alpha/2}\sigma_{\widehat{\theta}}$.

- And how to find $-z_{\alpha/2}$ and $z_{\alpha/2}$? Well, (Draw Normal curve on board.) Help class figure out that $z_{\alpha/2}$ is the value satisfying $P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2}$.

- Interlude for a bit of notation: recall that we write the standard Normal CDF (the probability that the standardized Normal RV Z will be less than or equal to z) as $\Phi(z)$. Looking at the Normal curve, it is evident that $\Phi(-z) = 1 - \Phi(z)$. Furthermore, it is often helpful to talk about the *inverse* of the Normal CDF–that is, a function whose argument is a probability and which returns a value of Z. We write this $\Phi^{-1}(p)$,where p is the probability and $P(Z \leq \Phi^{-1}(p)) = p$. Similarly, $P(Z \geq -\Phi^{-1}(p)) = p$.

- So in this case, if we're trying to find the $z_{\alpha/2}$ that satisfies $P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2}$, we can also write $z_{\alpha/2} = -\Phi^{-1}\left(\frac{\alpha}{2}\right)$ and $-z_{\alpha/2} = \Phi^{-1}\left(\frac{\alpha}{2}\right)$.

- In `R`, we type

`qnorm`(***p***) to find $\Phi^{-1}(p)$.

- So when I type `qnorm`**(.025)**, I get **-1.959964.**
  - This is the z-score associated with a $1-2(.025) = .95$ CI. And so if $\alpha = .05$, then $z_{\alpha/2} = 1.96$.
  - and when I type `qnorm`**(.05)**, I get **-1.6448536.**
  - and when I type `qnorm`**(.005), I get -2.5758293.**
- So a 95% confidence interval for an estimator $\widehat{\theta}$ whose sampling distribution is Normally distributed is constructed as

$$[\widehat{\theta} - (1.96)\,\sigma_{\widehat{\theta}}, \widehat{\theta} + (1.96)\,\sigma_{\widehat{\theta}},$$

  Now, what is $\sigma_{\widehat{\theta}}$? Why, we've already learned that: it's $\sigma_{\widehat{\theta}} = \sqrt{VAR(\widehat{\theta})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.
- What about a 99% CI? A 90% CI?
- So let's do an example. [Do Example 8.7 on page 412, but do not specify the variance, leave it as unknown.]

8.3 What is σ^2?

- As you'll recall, the CLT tells us that if $Y_1, Y_2, ...Y_n$ be i.i.d. random variables with $E(Y_i) = \mu$ and $VAR(Y_i) = \sigma^2$,then

$$U_n \equiv \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$$

  converges in probability to the standard normal CDF.
- Note that we've been dancing around a little bit of a problem. We've been using the CLT to construct interval estiamtors for μ, but they remain unquantified because they're in terms of σ^2, the population standard variance.
- This is a problem because it is very unusual to know the value of σ^2. Lots of homework problems and exercises will supply you the value of σ^2, but in practice we almost never have any reason to know what it actually is.

- So we need to estimate it with $S_U^2 \equiv \frac{\sum_i (Y_i - \overline{Y})^2}{n-1}$, our unbiased estimator for σ^2.
- At first seems intuitive. We have an unbiased estimate of σ^2 and we should be comfortable substituting S_U^2 for σ^2. It turns out that to justify this move, we need a bit more theory.

8.4 Interlude: the property of consistency

- In order to justify the estimation, we need to learn a new property of estimators: **consistency**. An estimator $\widehat{\theta}_n$ constructed from a sample of size n (subscripted to indicate just that) is a *consistent estimator* for θ if for any positive number ε,

$$\lim_{n\to\infty} P(|\widehat{\theta}_n - \theta| > \varepsilon) = 0.$$

- That is, as the sample size used to construct a consistent estimator becomes large, the chance that the error of the estimator is non-zero converges to zero. (Let's say this again...) We sometimes equivalently say that "$\widehat{\theta}_n$ converges in probability to θ," or $\widehat{\theta}_n \xrightarrow{p} \theta$
- Are unbiased estimators consistent estimators? Often, but not necessarily. Intuitively, this is because you could have an unbiased estimator that–even as n gets large–bounces around θ in a random, unbiased fashion but never centers on θ. Mathematically, this can be expressed as the helpful result (we'll omit proof here, it's on p. 450 of your text) that an unbiased estimator $\widehat{\theta}_n$ for θ is a consistent estimator for θ if its variance converges in probability to zero, that is:

$$E(\widehat{\theta}_n) = \theta \text{ and } \lim_{n\to\infty} VAR(\widehat{\theta}_n)=0 \Rightarrow \lim_{n\to\infty} P(|\widehat{\theta}_n - \theta| > \varepsilon) = 0, \textit{i.e.}$$
$$\widehat{\theta} \textit{ unbiased} \text{ for } \theta \text{ and } \lim_{n\to\infty} VAR(\widehat{\theta}_n) = 0 \Rightarrow \widehat{\theta} \textit{ consistent} \text{ for } \theta.$$

- This result, for example, means that $\overline{Y}$ is not only an unbiased estimator for μ, but because $VAR(\overline{Y}) = \frac{\sigma^2}{n}$ and thus $\lim_{n\to\infty} VAR(\overline{Y}) = 0$, $\overline{Y}$ is a consistent estimator for μ. The fact that $\overline{Y} \xrightarrow{p} \mu$ is sometimes referred to as the *law of large numbers*. It is the formal way of saying the intuitive idea that the average of many independent measures from a population should be quite close to the true population mean with high probability.

- Note that an estimator $\widehat{\theta}$ can be *consistent* for θ but not *unbiased* for θ.
  - E.g., S^2 as an estimator for σ^2.
- An estimator can be unbiased for θ but not consistent for θ.if its variance does not become monotonically smaller as n goes to infinity.
- Some helpful results are that if we have two estimators $\widehat{\theta}$ and $\widehat{\theta}'$ (I'm now going to drop the subscripts n to keep notation cleaner) such that $\widehat{\theta} \xrightarrow{p} \theta$ and $\widehat{\theta}' \xrightarrow{p} \theta'$, then [put these on left-hand board for reference later]:

$$\widehat{\theta} + \widehat{\theta}' \xrightarrow{p} \theta + \theta'$$
$$\widehat{\theta} \times \widehat{\theta}' \xrightarrow{p} \theta \times \theta'$$
$$\frac{\widehat{\theta}}{\widehat{\theta}'} \xrightarrow{p} \frac{\theta}{\theta'}, (\theta' \neq 0).$$
$$\text{Furthermore if } g(\cdot) \text{ continuous at } \theta, \text{then } g(\widehat{\theta}) \xrightarrow{p} g(\theta).$$

8.5 Back to σ^2

- Recall that the reason we discuss consistency was to help us justify estimating the population variance with S_U^2.
- Note that if we were simply estimating σ^2 in a vacuum, we'd be perfectly comfortable using S_U^2. After all $E\left(S_U^2\right) = \sigma^2$.
- But our task is a little more complicated here. We typically wish to substitute S_U^2 for σ^2 in the ratio $U_n \equiv \frac{\overline{Y}-\mu}{\sqrt{\sigma^2/n}}$ (which the CLT tells us is distributed standard Normal as n becomes large), and be assured that the resulting ratio, $\frac{\overline{Y}-\mu}{\sqrt{S_U^2/n}}$ (or, as we usually write it, $\frac{\overline{Y}-\mu}{S_U/\sqrt{n}}$) is itself distributed standard Normal as n becomes large.
- That is, we wish to show that $F\left(\frac{\overline{Y}-\mu}{S_U/\sqrt{n}}\right) \xrightarrow{p} \Phi$. (Put box around this to keep eyes on the prize during rest of tedious derivation.)
- Intuitively, it seems like it should. But this is definitely a more complicated question, because where in the original ratio σ^2 was a parameter, S_U^2 is a random variable.

- To begin, let's first show that S_U^2 is not only unbiased for σ^2; it's also consistent for σ^2.

- Rewrite

$$\begin{aligned}
S_U^2 &\equiv \frac{\sum_i \left(Y_i - \overline{Y}\right)^2}{n-1} \\
&= \frac{1}{n-1}\left[\sum_i Y_i^2 + \sum_i \overline{Y}^2 - \sum_i 2Y_i\overline{Y}\right] \\
&= \frac{1}{n-1}\left[\left(\sum_i Y_i^2\right) + n\overline{Y}^2 - 2n\left(\overline{YY}\right)\right] \\
&= \frac{1}{n-1}\left[\left(\sum_i Y_i^2\right) - n\overline{Y}^2\right] \\
&= \frac{n}{n-1}\left(\frac{1}{n}\sum_i Y_i^2 - \overline{Y}^2\right)
\end{aligned}$$

- Now let's consider

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_i Y_i^2 - \overline{Y}^2 &= \lim_{n\to\infty}\frac{1}{n}\sum_i Y_i^2 - \lim_{n\to\infty}\frac{1}{n}\sum_i \overline{Y}^2 \\
&= \mu_{Y^2} - \left(\mu_Y\right)^2,
\end{aligned}$$

where the first result is due to the law of large numbers (as $n \to \infty$, the sample mean converges to the population mean–in this case, the population mean of Y^2, which I write μ_{Y^2}). The second result is due to $g(\widehat{\theta}) \xrightarrow{p} g(\theta)$, since the function $g(x) = x^2$ is continuous and since $\overline{Y} \xrightarrow{p} \mu$, $g(\overline{Y}) \xrightarrow{p} g(\mu)$, or $\lim\limits_{n\to\infty}\frac{1}{n}\sum_i \overline{Y}^2 = \left(\mu_Y\right)^2$.

- Note that $\mu_{Y^2} - (\mu_Y) = E(Y^2) - \mu^2$, and now we have something that should look familiar once we recall the variance decomposition formula

$$\sigma^2 = E(Y^2) - \mu^2.$$

Thus

$$\lim_{n\to\infty}\frac{1}{n}\sum_i Y_i^2 - \overline{Y}^2 = \mu_{Y^2} - \left(\mu_Y\right)^2 = \sigma^2.$$

- Now the multiplicand $\frac{n}{n-1}$. But what is $\lim\limits_{n\to\infty}\frac{n}{n-1}$? It is a sequence of numbers converging to

1. Thus

$$
\begin{aligned}
\lim_{n\to\infty} \frac{n}{n-1}\left(\frac{1}{n}\sum_i Y_i^2 - \overline{Y}^2\right) &= 1\sigma^2 = \sigma^2, \text{and} \\
& S_U^2 \xrightarrow{p} \sigma^2.
\end{aligned}
$$

- And so S_U^2 is a consistent estimator for σ^2.