Handout for Lecture 12

Good Estimators, Sample Mean Distribution, and Confidence Intervals

ECON 340: Economic Research Methods — Instructor: Div Bhagia

Good Estimators

Denote $\hat{\theta}$ as an estimator for the population parameter θ. Some desirable properties for an estimator

- *Unbiasedness*: $E(\hat{\theta}) = \theta$
- *Efficiency*: lower variance is better
- *Consistency*: as the sample size becomes infinitely large, $\hat{\theta} \to \theta$

Question 1: If some sample estimator $\hat{\theta}$ is an unbiased estimator for the true population parameter θ i.e. $E(\hat{\theta}) = \theta$. This implies that:

☐ $\hat{\theta} = \theta$ in all samples.

☒ If we take repeated samples, average of $\hat{\theta}$ is equal to θ

Question 2: We are choosing between two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, both of which are unbiased i.e. $E(\hat{\theta}_1) = \mu$ and $E(\hat{\theta}_2) = \mu$. But the variance of $\hat{\theta}_1$ is lower than that of $\hat{\theta}_2$ i.e. $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$. Which of the following is true?

☐ We are indifferent between the two estimators.

☒ We prefer $\hat{\theta}_1$ over $\hat{\theta}_2$.

☐ We prefer $\hat{\theta}_2$ over $\hat{\theta}_1$.

☐ We need more information to reach any conclusion.

Sample Mean Distribution

Let $X_1, X_2, ..., X_n$ denote independent random draws (random sample) from a population with mean μ and variance σ^2. Then the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a random variable with:

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

In addition, the distribution of the sample mean is *normal* if *either* of the following is true:

- The underlying population is normal
- The sample size is large, say $n \geq 100$

Given the variance of the sample mean as $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$, its standard deviation, commonly referred to as the *standard error*, is $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Question 3: If the average of hourly wages in the population is $\mu = \$30$ and the variance of hourly wages is $\sigma^2 = 16$. Then what is the expected value, variance, standard error, and distribution of the sample mean estimator for a sample size of 400?

$$E(\bar{X}) = 30, \quad \sigma^2_{\bar{X}} = \frac{16}{400} = 0.04, \quad \sigma_{\bar{X}} = \sqrt{0.04} = 0.2$$

Since $n \geq 100$, by Central Limit Theorem $\bar{X}$ is normally distributed.

Question 4: You are interested in the average starting salary of CSUF graduates and are considering taking a random sample of 120 students. I advise you to take as large of a sample as feasible. This is sound advice because taking an even larger sample would ensure that

- ☐ The sample average $\bar{x} = \mu$
- ☐ The sample average $\bar{x}$ is drawn from a normal distribution
- ☒ The sample average $\bar{x}$ is drawn from a distribution with lower variance

Note: I am using $\bar{x}$ to denote a realization of $\bar{X}$.

Question 5: Can you explain intuitively why the variance of the sample mean increases with σ^2 and decreases with n?

Why does the variance of the sample mean decrease with n?

Suppose you aim to determine the average test score for all students at a university. When you use smaller samples, you could encounter significant sample-to-sample variability. For example, one sample might consist of students who performed exceptionally well, while another might include students who scored poorly. In contrast, a larger sample is more likely to accurately represent the overall student population, thereby reducing the variability between different samples.

Why does the variance of the sample mean increase with σ^2?

If the range of scores is quite wide, one sample could consist of students who performed exceptionally well, leading to a high sample mean, while another might include students who scored poorly, resulting in a low sample mean. However, if there is little to no variation in the student test scores across the university, you are likely to obtain similar sample means across different samples.

<u>Confidence Intervals</u>

Let $z_{\alpha/2}$ be the z-value that leaves area $\alpha/2$ in the upper tail of the normal distribution. Then $1 - \alpha$ confidence interval is given by

$$\bar{x} \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Margin of Error}}$$

Question 6: Continuing with Question 3, say you took a sample of 400 individuals and found the average hourly wages in your sample of $\bar{x} = 26$. Create a 95% confidence interval for the true population mean.

Note that here $1 - \alpha = 0.95$, so $\alpha/2 = 0.025$. From the Standard Normal Table, $z_{0.025} = 1.96$. In which case, the 95% confidence interval is given by:

$$26 \pm 1.96 \times 0.2 = [25.6, 26.4]$$