Econometrics 120A: Discussion Section

Week 9

Natalia Madrid & Lapo Bini

Department of Economics

Chapter 8: Point Estimation

Point Estimation

Point estimation involves estimating the mean (μ) of random variables $X_1, X_2, \ldots, X_n$. We use a function of the observed data $x_1, x_2, \ldots, x_n$ to derive our estimate.

The goal is to find the best estimate for μ based on the distribution of the random variables.

A good estimator will depend on:

- ⇨ Unbiasedness
- ⇨ Efficiency
- ⇨ Mean Squared Error
- ⇨ Consistency

Unbiasedness and Efficiency

Unbiasedness:

- An estimator $\hat{\theta}$ is **unbiased** for a parameter θ if: $E[\hat{\theta}] = \theta$
- This means that on average, the estimator hits the true parameter value across many samples.

Efficiency:

- An estimator is **efficient** if it has the smallest variance among all unbiased estimators.
- Formally, it is defined as: $\text{Var}(\hat{\theta})$ is minimized
- Efficient estimators provide the most precise estimates.

Mean Squared Error and Consistency

Mean Squared Error (MSE):

- The **MSE** of an estimator $\hat{\theta}$ is defined as:
  $\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$
- It measures the average squared difference between the estimated values and the true parameter value.

Consistency:

- An estimator $\hat{\theta}_n$ is **consistent** for θ if: $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0 \quad \text{for all } \epsilon > 0$
- This means that as the sample size increases, the estimator converges in probability to the true parameter value.

Kind of PS5 Question 1

Suppose you are analyzing the monthly income data of a random sample of individuals in a country to estimate the average income (μ) for the population. To evaluate the central tendency of income, you decide to minimize the **Mean Squared Error (MSE)**, defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$

where y_i is the observed income for individual i, and n is the sample size.

(a) Write down the first-order condition for minimizing the MSE and solve for μ as a function of the data points.

(b) Interpret your result and explain why this estimator is suitable for estimating the average income in the population.

(c) Consider now that the distribution of income is highly skewed (e.g., due to very high incomes of a few individuals). What alternative estimator of central tendency might be more appropriate in this case? Explain your reasoning.

Kind of PS5 Question 1

Suppose you are analyzing the monthly income data of a random sample of individuals in a country to estimate the average income (μ) for the population. To evaluate the central tendency of income, you decide to minimize the **Mean Squared Error (MSE)**, defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$

where y_i is the observed income for individual i, and n is the sample size.

(a) Write down the first-order condition for minimizing the MSE and solve for μ as a function of the data points.

Kind of PS5 Question 1

Suppose you are analyzing the monthly income data of a random sample of individuals in a country to estimate the average income (μ) for the population. To evaluate the central tendency of income, you decide to minimize the **Mean Squared Error (MSE)**, defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$

where y_i is the observed income for individual i, and n is the sample size.

(b) Interpret your result and explain why this estimator is suitable for estimating the average income in the population.

Kind of PS5 Question 1

Suppose you are analyzing the monthly income data of a random sample of individuals in a country to estimate the average income (μ) for the population. To evaluate the central tendency of income, you decide to minimize the **Mean Squared Error (MSE)**, defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$

where y_i is the observed income for individual i, and n is the sample size.

(c) Consider now that the distribution of income is highly skewed (e.g., due to very high incomes of a few individuals). What alternative estimator of central tendency might be more appropriate in this case? Explain your reasoning.

Why Do We Use (n-1) for the Variance?

Simple: because $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ is bias. Let's compute the expected value and the bias of σ^2

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right] =$$

Chapter 8: Hypothesis Testing

Hypothesis Testing

Hypothesis testing is a statistical method used to make inferences or draw conclusions about a population based on sample data. The process involves several key steps:

- **Formulating Hypotheses:** Two competing hypotheses are established:

    Null Hypothesis (H0): For example, this is the hypothesis that the **unobserved** population mean is equal to 10, that is $\mu = 10$.

    Alternative Hypothesis (H1 or Ha): This represents the alternative theory against the null hypothesis $\mu = 10$. For example, an alternative hypothesis would be $\mu > 10$, there is a bigger population mean.

- **Choosing a Significance Level (α):** This is the cutoff level to determine whether to reject the null hypothesis, commonly set at 5% or 1%. It also represents the probability of making a Type I error (rejecting a true null hypothesis).

Hypothesis Testing - Graphical Representation

Test Statistic

⇨ **Calculating a Test Statistic:** A test statistic is calculated Based on the data. For large samples (CLT applies), we compute the **test statistic**, or Z-score, as:

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

where:

$\bar{x}$ = sample mean
μ_0 = hypothesized population mean under the null hypothesis
σ = population standard deviation (known)
n = sample size

Two Approaches

- **Classical Approach**: Given the size of the test α, if the test statistic t is larger or smaller than the critical value z_α, then the null hypothesis is rejected in favor of the alternative hypothesis.

- **P-Value**: The p-value is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming the null hypothesis is true. It is calculated based on the null and alternative hypothesis:

  (1) For a one-tailed test (upper tail): p-value $= P(Z > t)$.
  (2) For a one-tailed test (lower tail): p-value $= P(Z < t)$.
  (3) For a two-tailed test: p-value $= P(|Z| > |t|)$.

  where Z is a standard normal random variable. Then we compare the p-value with different values of α: if p-value $< \alpha$ we have enough evidence to reject the null hypothesis.

Kind of PS5 Question 2

A global survey conducted in 2023 found that in a random sample of 850 adults, 58% believed that transitioning to renewable energy sources is essential to combating climate change. Is this strong evidence that more than half of the global population holds this belief? Suppose you want to test whether the survey results support your belief.

(a) Write down your null and alternative hypotheses for this scenario.

(b) At the 5% significance level, determine if the survey results refute the null hypothesis using the p-value approach.

Kind of PS5 Question 2

A global survey conducted in 2023 found that in a random sample of 850 adults, 58% believed that transitioning to renewable energy sources is essential to combating climate change. Is this strong evidence that more than half of the global population holds this belief? Suppose you want to test whether the survey results support your belief.

(a) Write down your null and alternative hypotheses for this scenario.

Kind of PS5 Question 2

A global survey conducted in 2023 found that in a random sample of 850 adults, 58% believed that transitioning to renewable energy sources is essential to combating climate change. Is this strong evidence that more than half of the global population holds this belief? Suppose you want to test whether the survey results support your belief.

(b) At the 5% significance level, determine if the survey results refute the null hypothesis using the p-value approach.

PS5 Question 3

A company selling glasses claims that the average weight of a load of stones is 40 pounds and that the variance of a load of stones is 100.

Lapo and Natalia claim that the true population average is not 40 pounds.

Then, a random sample of 25 glasses is examined, yielding an estimate of 37 pounds.

(a) Set up the null and alternative hypotheses that would test that their claim is true.

(b) Suppose that the loads have a normal distribution. Determine the critical values for the test if the size of the test is set at 10%.

(c) Undertake the test. Does the data disagree with the null hypothesis?

(d) Compute the p-value for the test. What is the size of the test that allow us to reject the null?

PS5 Question 3

A company selling glasses claims that the average weight of a load of stones is 40 pounds and that the variance of a load of stones is 100.

Lapo and Natalia claim that the true population average is not 40 pounds.

Then, a random sample of 25 glasses is examined, yielding an estimate of 37 pounds.

(a) Set up the null and alternative hypotheses that would test that their claim is true.

PS5 Question 3

A company selling glasses claims that the average weight of a load of stones is 40 pounds and that the variance of a load of stones is 100.

Lapo and Natalia claim that the true population average is not 40 pounds.

Then, a random sample of 25 glasses is examined, yielding an estimate of 37 pounds.

(b) Suppose that the loads have a normal distribution. Determine the critical values for the test if the size of the test is set at 10%.

PS5 Question 3

A company selling glasses claims that the average weight of a load of stones is 40 pounds and that the variance of a load of stones is 100.

Lapo and Natalia claim that the true population average is not 40 pounds.

Then, a random sample of 25 glasses is examined, yielding an estimate of 37 pounds.

(c) Undertake the test. Does the data disagree with the null hypothesis?

PS5 Question 3

A company selling glasses claims that the average weight of a load of stones is 40 pounds and that the variance of a load of stones is 100.

Lapo and Natalia claim that the true population average is not 40 pounds.

Then, a random sample of 25 glasses is examined, yielding an estimate of 37 pounds.

(d) Calculate the p-value for the test. What is the size of the test that allows you to reject the null?