

# Lecture 4: Applications and Real-World Use of the Central Limit Theorem (CLT)

Duration: 50 minutes

## Learning Objectives:

By the end of this lecture, students will be able to:

- Understand how the Central Limit Theorem (CLT) underpins the construction of confidence intervals for population means.
- Explain the role of CLT in hypothesis testing for population means.
- Identify and discuss real-life scenarios where the CLT is applied.
- Apply the CLT to estimate a population mean from sample data.

## 1 Introduction: Bridging Theory to Practice (5 minutes)

Good morning, everyone! In our previous lectures, we rigorously defined the Central Limit Theorem and built an intuitive understanding of why the distribution of sample means tends to be normal, regardless of the population's shape. We also learned about the mean of sample means ( $\mu_{\bar{x}} = \mu$ ) and the standard error of the mean ( $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ ).

Today, we're moving from theory to application. The CLT isn't just a fascinating mathematical concept; it's one of the most powerful tools in a statistician's arsenal. It allows us to make reliable inferences about large populations based on data collected from much smaller samples. We'll explore its direct use in constructing confidence intervals and conducting hypothesis tests, and then look at some compelling real-world examples.

## 2 Application in Confidence Intervals (15 minutes)

Recall that a **confidence interval** is a range of values, derived from sample statistics, that is likely to contain the true value of an unknown population parameter. The CLT is absolutely crucial for constructing confidence intervals for the population mean ( $\mu$ ) when the population standard deviation ( $\sigma$ ) is known or when the sample size is large enough to approximate it.

### Why CLT is essential for Confidence Intervals:

1. **Normality Assumption:** Most confidence interval formulas (especially those involving z-scores) assume that the sampling distribution of the sample mean is normal. Thanks to the CLT, we know this assumption holds true for large sample sizes ( $n \geq 30$ ), even if the original population is not normal.
2. **Standard Error:** The CLT provides the formula for the standard error of the mean ( $\sigma/\sqrt{n}$ ), which is a key component in determining the width of the confidence interval.

### Formula for a Confidence Interval for the Population Mean ( $\mu$ ):

When  $\sigma$  is known (or  $n$  is large and we use  $s$  as an estimate for  $\sigma$ ):

$$\bar{x} \pm z^* \left( \frac{\sigma}{\sqrt{n}} \right)$$

Let's break this down:

- $\bar{x}$ : This is your **sample mean**, the best point estimate for the population mean.
- $z^*$ : This is the **critical z-value** corresponding to your desired confidence level. For example:
  - For a 90% confidence interval,  $z^* \approx 1.645$

- For a 95% confidence interval,  $z^* \approx 1.96$
- For a 99% confidence interval,  $z^* \approx 2.576$
- $\sigma$ : This is the **population standard deviation**. If unknown but  $n \geq 30$ , we can often use the sample standard deviation ( $s$ ) as a good estimate for  $\sigma$ .
- $\sqrt{n}$ : The square root of your **sample size**.
- $\sigma/\sqrt{n}$ : This is the **standard error of the mean** ( $\sigma_{\bar{x}}$ ), which quantifies the typical distance between a sample mean and the true population mean.

The term  $z^*(\sigma/\sqrt{n})$  is called the **margin of error**. It tells us how much our sample mean ( $\bar{x}$ ) might reasonably deviate from the true population mean ( $\mu$ ).

### 3 Application in Hypothesis Testing of the Mean (10 minutes)

**Hypothesis testing** is a statistical method used to determine if there is enough evidence in a sample data to infer that a certain condition is true for the entire population. The CLT is fundamental here because it allows us to calculate the probability of observing our sample mean (or something more extreme) *if* the null hypothesis were true.

#### The Logic:

1. **Formulate Hypotheses:** We set up a null hypothesis ( $H_0$ ) (e.g.,  $\mu = \text{some value}$ ) and an alternative hypothesis ( $H_a$ ) (e.g.,  $\mu \neq \text{some value}$ ).
2. **Assume  $H_0$  is True:** We assume, for the sake of argument, that the population mean is exactly what  $H_0$  states.
3. **CLT Kicks In:** Because of the CLT, we know that if  $H_0$  were true and we took many samples, the distribution of their means would be approximately normal, centered at the value specified by  $H_0$ , with a standard deviation equal to the standard error ( $\sigma/\sqrt{n}$ ).
4. **Calculate Test Statistic:** We then calculate a z-score (our test statistic) for our *observed sample mean* ( $\bar{x}$ ), based on the assumption that  $H_0$  is true:

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

where  $\mu_0$  is the hypothesized population mean under  $H_0$ .

5. **Find P-value:** Using this z-score and the standard normal distribution (thanks to CLT!), we find the p-value. The p-value is the probability of observing a sample mean as extreme as, or more extreme than, our observed  $\bar{x}$ , *if the null hypothesis were true*.
6. **Make Decision:** If the p-value is very small (e.g., less than our significance level  $\alpha$ ), it means our observed sample mean is very unlikely to occur by chance if  $H_0$  were true. This provides evidence to reject  $H_0$  in favor of  $H_a$ .

Without the CLT guaranteeing the normality of sample means, we wouldn't be able to use the standard normal distribution to calculate these probabilities and make valid inferences.

### 4 Real-Life Examples (10 minutes)

The CLT is silently working behind the scenes in countless everyday applications.

#### 1. Polling Data / Opinion Surveys:

- When you see a news report stating "Candidate A has 52% support with a margin of error of  $\pm 3\%$ , 19 times out of 20," that's the CLT at work.

- Pollsters take a sample (e.g., 1000 people) and calculate the proportion (which is a type of mean). The CLT ensures that the distribution of these sample proportions would be approximately normal, allowing them to calculate the margin of error and confidence interval.

## 2. Quality Control in Manufacturing:

- A company producing light bulbs wants to ensure their average lifespan is at least 1000 hours. They can't test every bulb.
- Instead, they take regular samples (e.g., 50 bulbs) from the production line, measure their lifespan, and calculate the sample mean.
- Using CLT, they can set up control charts or conduct hypothesis tests to see if the average lifespan of their sampled bulbs deviates significantly from the target, indicating a problem in the production process.

## 3. Election Predictions (Exit Polls):

- On election day, news channels conduct exit polls, surveying a sample of voters as they leave polling stations.
- They use the collected sample data to predict the outcome of the election with a certain confidence level and margin of error. The CLT provides the statistical justification for these predictions, treating the proportion of votes for each candidate as a sample mean.

## 4. Medical Research / Drug Trials:

- To test the effectiveness of a new drug, researchers might compare the average blood pressure reduction in a sample of patients taking the new drug versus a control group.
- CLT allows them to determine if any observed difference in sample means is statistically significant or merely due to random chance, helping them draw conclusions about the drug's efficacy for the broader population.

# 5 Practical Example: Estimating Population Mean Height (8 minutes)

Let's work through an example to solidify our understanding.

**Scenario:** Suppose the true population mean height of all university students is unknown. We want to estimate it.

**Data Collection:** We take a random sample of  $n = 50$  students. We measure their heights and calculate the sample mean ( $\bar{x}$ ) and sample standard deviation ( $s$ ). Let's say our sample results are:

- Sample Mean ( $\bar{x}$ ) = 170 cm
- Sample Standard Deviation ( $s$ ) = 10 cm

**Applying CLT to Estimate the Population Mean:** Since our sample size ( $n = 50$ ) is greater than 30, we can invoke the Central Limit Theorem. This means:

1. The sampling distribution of sample means will be approximately normal.
2. We can use our sample standard deviation ( $s$ ) as a good estimate for the population standard deviation ( $\sigma$ ). So, we'll use  $s = 10$  cm for  $\sigma$ .
3. We can calculate the standard error of the mean:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{50}} \approx \frac{10}{7.071} \approx 1.414 \text{ cm}$$

Now, let's construct a **95% confidence interval** for the true population mean height.

- For a 95% confidence level, our critical z-value ( $z^*$ ) is 1.96.

Using the confidence interval formula:

$$\begin{aligned}\bar{x} \pm z^* \left( \frac{\sigma}{\sqrt{n}} \right) \\ 170 \pm 1.96(1.414) \\ 170 \pm 2.771\end{aligned}$$

So, our 95% confidence interval is:

$$\begin{aligned}(170 - 2.771, 170 + 2.771) \\ (167.229, 172.771)\end{aligned}$$

**Interpretation:** We are 95% confident that the true population mean height of all university students lies between 167.23 cm and 172.77 cm.

This example clearly demonstrates how CLT allows us to generalize findings from a small sample to a much larger, unknown population.

## 6 Mini-Project (Optional, to be discussed/assigned) (2 minutes)

To really drive home the intuition of the CLT, I highly recommend a hands-on simulation.

### Mini-Project Idea: Simulating Rolling a Die

1. **Population:** Imagine a fair six-sided die. The outcomes are [1, 2, 3, 4, 5, 6]. This population is *uniform*, not normal.
  - Population Mean ( $\mu$ ) = 3.5
  - Population Standard Deviation ( $\sigma$ ) =  $\sqrt{((1 - 3.5)^2 + \dots + (6 - 3.5)^2)/6} \approx 1.7078$
2. **Simulation Steps:**
  - Use a programming language (Python is excellent for this with `numpy` and `matplotlib`) or a spreadsheet program (Excel).
  - Simulate rolling a single die 1000 times and plot its histogram. (It should look flat/uniform).
  - Now, simulate taking 1000 samples, where each sample consists of rolling the die  $n = 30$  times.
  - For each of these 1000 samples, calculate the mean of the 30 rolls.
  - Plot a histogram of these 1000 sample means.
3. **Expected Outcome:** You will observe that the histogram of the 1000 sample means will be approximately bell-shaped (normal), centered around the population mean of 3.5, and its spread will be much narrower than the original population (approximately  $\sigma/\sqrt{30} \approx 1.7078/\sqrt{30} \approx 0.312$ ). This visually confirms the power of the CLT.

## 7 Conclusion (5 minutes)

Today, we've seen how the Central Limit Theorem is not just a theoretical concept but a practical powerhouse in statistics. It liberates us from needing to know the population's exact distribution, allowing us to:

- Construct reliable **confidence intervals** for population means, providing a range of plausible values for our unknown parameter.
- Perform robust **hypothesis tests** for population means, enabling us to make data-driven decisions.

From political polling to quality control, the CLT's widespread application underpins much of what we consider "data-driven insights" in the modern world. Understanding it is a crucial step towards becoming proficient in data analysis.

**Any questions before we conclude?**

# Central Limit Theorem Applications - Practice Questions

**Instructions:** Please answer the following questions based on the lecture material.

## Questions:

- 1. Role in Confidence Intervals:** Explain why the Central Limit Theorem is crucial for constructing confidence intervals for the population mean. What specific assumption does it help us meet?
- 2. Confidence Interval Formula Breakdown:** In the confidence interval formula for the population mean,  $\bar{x} \pm z^* \left( \frac{\sigma}{\sqrt{n}} \right)$ :
  - a) What does  $\bar{x}$  represent?
  - b) What does  $z^*$  represent?
  - c) What is the term  $\left( \frac{\sigma}{\sqrt{n}} \right)$  called, and what does it measure?
- 3. Role in Hypothesis Testing:** How does the Central Limit Theorem allow us to calculate probabilities (like p-values) in hypothesis testing for the population mean?
- 4. Scenario: Quality Control:** A company manufactures bags of potato chips. They claim each bag contains, on average, 150 grams of chips. A quality control manager takes a random sample of 49 bags and finds a sample mean weight of 148 grams with a sample standard deviation of 7 grams.
  - a) Can the CLT be applied in this scenario? Why or why not?
  - b) Calculate the standard error of the mean for this sample.
  - c) Construct a 90% confidence interval for the true mean weight of chip bags. (Use  $z^* = 1.645$ )
  - d) Based on your confidence interval, what can you conclude about the company's claim that the average weight is 150 grams?
- 5. Real-Life Application:** Describe one real-life example of the CLT's application discussed in the lecture, and briefly explain how the CLT is used in that context.
- 6. Impact of Confidence Level:** If you wanted to increase your confidence level from 95% to 99% for a confidence interval (keeping sample size and standard deviation the same), how would that affect the width of the interval? Why?

# Central Limit Theorem Applications - Answer Key

## Answers:

- 1. Role in Confidence Intervals:** The Central Limit Theorem (CLT) is crucial for constructing confidence intervals for the population mean because it ensures that the **sampling distribution of the sample mean is approximately normal**, even if the original population distribution is not normal. This normality assumption is vital because many confidence interval formulas (especially those using z-scores) are based on the properties of the normal distribution.
- 2. Confidence Interval Formula Breakdown:**
  - a)  $\bar{x}$  represents the **sample mean**, which is our point estimate for the unknown population mean.
  - b)  $z^*$  represents the **critical z-value** corresponding to the desired confidence level. It determines how many standard errors away from the mean we need to go to capture a certain percentage of the distribution.
  - c) The term  $\left(\frac{\sigma}{\sqrt{n}}\right)$  is called the **Standard Error of the Mean**. It measures the typical amount that a sample mean deviates from the true population mean. It quantifies the precision of our sample mean as an estimate of the population mean.
- 3. Role in Hypothesis Testing:** In hypothesis testing for the population mean, the Central Limit Theorem allows us to assume that, if the null hypothesis ( $H_0$ ) is true, the sampling distribution of the sample mean will be approximately normal and centered at the hypothesized population mean ( $\mu_0$ ). This enables us to calculate a z-score (test statistic) for our observed sample mean. With this z-score and the knowledge that the sampling distribution is normal, we can then use the standard normal distribution (z-table or calculator) to find the p-value, which is the probability of observing our sample result (or something more extreme) if  $H_0$  were true.
- 4. Scenario: Quality Control:**
  - a) **Can the CLT be applied?** Yes. The sample size ( $n = 49$ ) is greater than 30, which is generally considered sufficiently large for the CLT to apply, even if the population distribution of chip bag weights is not perfectly normal. We can also use the sample standard deviation ( $s = 7$  grams) as a good estimate for the population standard deviation ( $\sigma$ ).
  - b) **Calculate the standard error of the mean:**  $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{7}{\sqrt{49}} = \frac{7}{7} = 1$  gram
  - c) **Construct a 90% confidence interval:**  $\bar{x} \pm z^* \left(\frac{s}{\sqrt{n}}\right)$   $148 \pm 1.645(1)$   $148 \pm 1.645$  Confidence Interval: **(146.355 grams, 149.645 grams)**
  - d) **Conclusion about the company's claim:** The company claims the average weight is 150 grams. Our 90% confidence interval for the true mean weight is (146.355, 149.645) grams. Since 150 grams falls *outside* this confidence interval, we can conclude with 90% confidence that the true average weight of the chip bags is likely *not* 150 grams. It appears to be lower.
- 5. Real-Life Application:** (One example chosen from the lecture's real-life applications) **Polling Data/Opinion Surveys:** Pollsters take a sample of voters/respondents and calculate a sample proportion (which acts as a mean). The CLT ensures that the distribution of these sample proportions would be approximately normal. This normality allows pollsters to calculate a margin of error and a confidence interval around their reported percentage, giving the public a sense of the precision and reliability of the survey results (e.g., "Candidate A has 52% support with a  $\pm 3\%$  margin of error").
- 6. Impact of Confidence Level:** If you increase your confidence level from 95% to 99% (keeping sample size and standard deviation the same), the **width of the confidence interval would increase (become wider)**. This is because a higher confidence level requires a larger critical z-value ( $z^*$  is 1.96 for 95% and 2.576 for 99%). To be more confident that our interval captures the true population mean, we need to make the interval wider to encompass a larger range of possible values.