Chapter 7 Study Guide: The Central Limit Theorem

Introduction

Why are we so concerned with means? Two reasons are that they give us a middle ground for comparison and they are easy to calculate. In this chapter, you will study means and the Central Limit Theorem.

The Central Limit Theorem (CLT for short) is one of the most powerful and useful ideas in all of statistics. Both alternatives are concerned with drawing finite samples of size $n$ from a population with a known mean, $\mu$, and a known standard deviation, $\sigma$.

The first alternative says that if we collect samples of size $n$ and $n$ is "large enough," calculate each sample's mean, and create a histogram of those means, then the resulting histogram will tend to have an approximate normal bell shape.

The second alternative says that if we again collect samples of size $n$ that are "large enough," calculate the sum of each sample and create a histogram, then the resulting histogram will again tend to have a normal bell-shape.

In either case, it does not matter what the distribution of the original population is, or whether you even need to know it. The important fact is that the sample means and the sums tend to follow the normal distribution. And, the rest you will learn in this chapter.

The size of the sample, $n$, that is required in order to be 'large enough' depends on the original population from which the samples are drawn. If the original population is far from normal then more observations are needed for the sample means or the sample sums to be normal. Sampling is done with replacement.

Collaborative Classroom Activity

Part 1:

1. Roll two dice and find the mean of the numbers that you get.

<table>
<thead>
<tr>
<th>Die 1</th>
<th>Die 2</th>
<th>Mean</th>
</tr>
</thead>
</table>

2. Place a dot for each mean value that you obtained on the class dot plot on the white board.

3. Reproduce the dot plot here

4. Find the mean and the standard deviation of the class means using your TI84 calculator.
Part 2:

1. Roll five dice and find the mean of the numbers that you get. Repeat the experiment five times. You should have five means in total. Write them in the table below.

<table>
<thead>
<tr>
<th>Die 1</th>
<th>Die 2</th>
<th>Die 3</th>
<th>Die 3</th>
<th>Die 4</th>
<th>Mean</th>
</tr>
</thead>
</table>

2. Place a dot for each mean value that you obtained on the class dot plot on the white board.

3. Reproduce the dot plot here

4. Find the mean and the standard deviation of the class means using your TI84 calculator.

As the number of dice rolled increases from 1 to 2 to 5 to 10, the following is happening:

1. The mean of the sample means remains approximately the same.
2. The spread of the sample means (the standard deviation of the sample means) gets smaller.
3. The graph appears steeper and thinner.

You have just demonstrated the Central Limit Theorem (CLT).

The Central Limit Theorem tells you that as you increase the number of dice, the sample means tend toward a normal distribution (the sampling distribution).
The Central Limit for Sample Means (Averages)

Suppose $X$ is a random variable with a distribution that may be known or unknown (it can be any distribution). Using a subscript that matches the random variable, suppose:

- **a.** $\mu_X = \text{the mean of } X$
- **b.** $\sigma_X = \text{the standard deviation of } X$

If you draw random samples of size $n$, then as $n$ increases, the random variable $\bar{X}$ which consists of sample means, tends to be normally distributed and

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

**The Central Limit Theorem** for Sample Means says that if you keep drawing larger and larger samples (like rolling 1, 2, 5, and, finally, 10 dice) and calculating their means the sample means form their own normal distribution (the sampling distribution).

The normal distribution has the same mean as the original distribution and a variance that equals the original variance divided by $n$, the sample size. $n$ is the number of values that are averaged together not the number of times the experiment is done.

To put it more formally, if you draw random samples of size $n$, the distribution of the random variable $\bar{X}$, which consists of sample means, is called the sampling distribution of the mean. The sampling distribution of the mean approaches a normal distribution as $n$, the sample size, increases.

The random variable $\bar{X}$ has a different $z$-score associated with it than the random variable $X$.

$$z = \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}}$$

$\mu_X$ is both the average of $X$ and of $\bar{X}$

$\bar{x}$ is the value of $\bar{X}$ in one sample

$\sigma_X = \frac{\sigma}{\sqrt{n}}$ = standard deviation of $\bar{X}$ and is called the standard error of the mean.

**Law of Large Numbers**

The Law of Large Numbers says that if you take samples of larger and larger size from any population, then the mean $\bar{x}$ of the sample tends to get closer and closer to $\mu$.

From the Central Limit Theorem, we know that as $n$ gets larger and larger, the sample means follow a normal distribution. The larger $n$ gets, the smaller the standard deviation gets. (Remember that the standard deviation for $\bar{X}$ is $\frac{\sigma}{\sqrt{n}}$)

This means that the sample mean $\bar{x}$ must be close to the population mean $\mu$. We can say that $\mu$ is the value that the sample means approach as $n$ gets larger. The Central Limit Theorem illustrates the Law of Large Numbers.
EXAMPLE 1
An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size \( n = 25 \) are drawn randomly from the population.

a. Find the probability that the sample mean is between 85 and 92.

b. Find the value that is 2 standard deviations above the expected value (it is 90) of the sample mean.

EXAMPLE 2
The length of time, in hours, it takes an "over 40" group of people to play one soccer match is normally distributed with a mean of 2 hours and a standard deviation of 0.5 hours. A sample of size \( n = 50 \) is drawn randomly from the population.

Find the probability that the sample mean is between 1.8 hours and 2.3 hours.

Using the Central Limit Theorem

It is important for you to understand when to use the CLT. If you are being asked to find the probability of the mean, use the CLT for the mean.

If you are being asked to find the probability of an individual value, do not use the CLT. Use the distribution of its random variable.

Examples of the Central Limit Theorem

EXAMPLE 3
A study involving stress is done on a college campus among the students. The stress scores follow a uniform distribution with the lowest stress score equal to 1 and the highest equal to 5. Using a sample of 75 students, find:

a. The probability that the mean stress score for the 75 students is less than 2.

b. The 90th percentile for the mean stress score for the 75 students.

c. The probability that the total of the 75 stress scores is less than 200.

d. The 90th percentile for the total stress score for the 75 students.
EXAMPLE 4
Yoonie is a personnel manager in a large corporation. Each month she must review 16 of the employees. From past experience, she has found that the reviews take her approximately 4 hours each to do with a population standard deviation of 1.2 hours.

Let $X$ be the random variable representing the time it takes her to complete one review. Assume $X$ is normally distributed.

Let $\bar{X}$ be the random variable representing the mean time to complete the 16 reviews. Assume that the 16 reviews represent a random set of reviews.

A. Complete the distributions.
   1. $X \sim \quad$
   2. $\bar{X} \sim \quad$

B. Find the probability that one review will take Yoonie from 3.5 to 4.25 hours.

C. Find the probability that the mean of a month’s reviews will take Yoonie from 3.5 to 4.25 hrs.

D. Find the 95th percentile for the mean time to complete one month’s reviews.

E. Find the probability that the sum of the month’s reviews takes Yoonie from 60 to 65 hours.

F. Find the 95th percentile for the sum of the month’s reviews.

G. What causes the probabilities in Exercise B and Exercise C to differ?
EXAMPLE 5
A manufacturer makes screws with a diameter that falls in a range of 0.15 cm to 0.25 cm. Within that range, the distribution is uniform.

a. In words, \(X=\)

b. \(X\sim\)

c. Find \(\mu_X\) and \(\sigma_X\)

d. In words, \(\bar{X}=\)

e. \(\bar{X}\sim\)

f. Find the probability that an individual screw is between 0.20 cm and 0.22 cm. Graph the situation and shade in the area to be determined.

g. The manufacturer chooses 25 screws at random to measure their lengths. Find the probability that the average length of the 25 screws is between 0.20 cm and 0.22 cm. Graph the situation and shade in the area to be determined.

h. Explain the why there is a difference in (f) and (g).

EXAMPLE 5
The average length of a maternity stay in a U.S. hospital is said to be 2.4 days with a standard deviation of 0.9 days. Assume that the length of hospital maternity stay is normally distributed. We randomly survey 80 women who recently bore children in a U.S. hospital.

a. In words, \(X=\)

b. In words, \(\bar{X}=\)

c. \(X\sim\)

d. \(\bar{X}\sim\)

e. Is it likely that an individual stayed more than five days in the hospital? Why or why not?

f. Is it likely that the average stay for the 80 women was more than five days? Why or why not?

g. Which is more likely:
   An individual stayed more than five days OR the average stay of 80 women was more than five days.