Chapter 6 Normal Probability Distributions
The Central Limit Theorem (6.4)

Objectives

In this lesson, we will learn how to

- state the central limit theorem;
- find the mean and standard deviation of sample means;
- decide when a normal distribution can be used for sample means;
- distinguish between an individual value and a sample mean;
- apply the central limit theorem to real applications;
- connect the central limit theorem with the rare event rule.

1 Why the Central Limit Theorem Matters

The central limit theorem, often called the CLT, is one of the most important ideas in statistics.

It tells us that even if the original population is not normal, the distribution of sample means often becomes approximately normal when the sample size is large enough.

Key Concept

The central limit theorem allows us to use a normal distribution for many important problems involving sample means.

2 The Central Limit Theorem

Suppose a population has mean μ and standard deviation σ. If all simple random samples of the same size n are selected from this population, then the sample means form a sampling distribution.

Central Limit Theorem

For all samples of the same size n, with $n > 30$, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution with

$$\mu_{\bar{x}} = \mu$$

and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

The value $\sigma_{\bar{x}}$ is called the standard error of the mean. It is sometimes abbreviated as SEM.

3 Practical Rules for Sample Means

When working with a sample mean $\bar{x}$, first check whether the normal distribution is appropriate.

When Can We Use a Normal Distribution for $\bar{x}$?

We can use a normal distribution for sample means if either condition is satisfied:

- the original population is normally distributed; or
- the sample size is large, usually $n > 30$.

When the original population is not normal and $n \leq 30$, the methods in this section generally do not apply.

Formulas for Sample Means

For the sampling distribution of sample means:

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

4 Individual Value or Sample Mean?

One common mistake is using the wrong standard deviation. Always decide whether the problem involves an individual value x or a sample mean $\bar{x}$.

Individual Value vs. Sample Mean

Individual value x:

$$z = \frac{x - \mu}{\sigma}$$

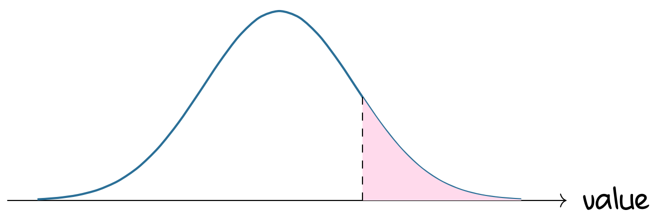
Sample mean $\bar{x}$ from n values:

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

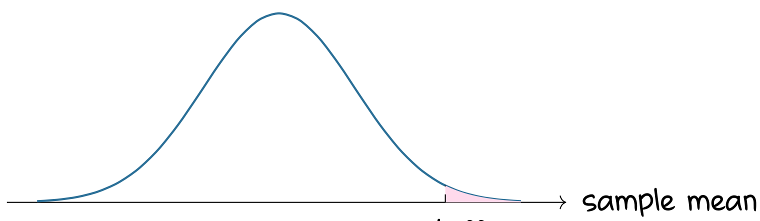
Important: For a sample mean, the standard deviation is smaller because sample means vary less than individual data values.

5 Visual Idea

A single adult male's hip width might vary quite a bit from person to person. But the mean hip width of a sample of 126 adult males varies much less.



Individual values: wider spread



Sample means: less variation

Example 1: American Airlines uses Boeing 737 jets with 126 seats in the main cabin. An engineer is considering reducing the seat width from 16.6 inches to 16.0 inches.
Adult male hip widths are normally distributed with

$$\mu = 14.3 \text{ in.}, \qquad \sigma = 0.9 \text{ in.}$$

(a) Find the probability that a randomly selected adult male has a hip width greater than 16.0 inches.

(b) Find the probability that 126 adult males have a mean hip width greater than 16.0 inches.

(c) Which result is more relevant for designing individual aircraft seats?

Solution

Part (a): Individual value
Because this part asks about one randomly selected adult male, we

use

$$z = \frac{x - \mu}{\sigma}.$$

Substitute $x = 16.0$, $\mu = 14.3$, and $\sigma = 0.9$:

$$z = \frac{16.0 - 14.3}{0.9}$$

$$z = \frac{1.7}{0.9} \approx 1.89.$$

Now we want

$$P(x > 16.0) = P(z > 1.89).$$

Using technology,

$$P(z > 1.89) \approx 0.0295.$$

So about 2.95% of adult males have hip widths greater than 16.0 inches.

Solution

Part (b): Sample mean
Now we are dealing with the mean hip width of $n = 126$ adult males.
We use the central limit theorem.
The mean of the sample means is

$$\mu_{\bar{x}} = \mu = 14.3.$$

The standard error is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.9}{\sqrt{126}}.$$

$$\sigma_{\bar{x}} \approx 0.0802.$$

Convert $\bar{x} = 16.0$ to a z score:

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

$$z = \frac{16.0 - 14.3}{0.9/\sqrt{126}}$$
$$z \approx 21.20.$$

This z score is extremely large, so

$$P(\bar{x} > 16.0) \approx 0.$$

It is extremely unlikely that a random group of 126 adult males would have a mean hip width greater than 16.0 inches.

Solution

Part (c): Interpretation
The result from part (a) is more relevant for seat design, because individual seats are occupied by individual passengers.
Part (a) shows that about 3% of adult males may have hip widths greater than 16.0 inches. Even though 3% may look small, it could mean several passengers on a full flight might not fit comfortably.
Therefore, reducing the seat width to 16.0 inches does not appear to be a good design choice.

6 Rare Event Rule and Hypothesis Testing

The central limit theorem is also useful for the beginning idea of hypothesis testing.

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is very small, and the observed event is significantly less than or significantly greater than what we expect, then we conclude that the assumption is probably not correct.

This rule helps us decide whether a sample result is unusual enough to question an assumption about the population.

Example 2: Assume that the population of human body temperatures has a mean of $98.6°$F, as commonly believed. Also assume that the population standard deviation is

$$\sigma = 0.62°\text{F}.$$

A simple random sample of $n = 106$ subjects has sample mean

$$\bar{x} = 98.2°\text{F}.$$

If the true mean body temperature is really $98.6°$F, find the probability of getting a sample mean of $98.2°$F or lower. Is $98.2°$F significantly low? What does this suggest?

Solution

Because this problem involves a sample mean, we use

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

We are given

$$\mu = 98.6, \qquad \sigma = 0.62, \qquad n = 106, \qquad \bar{x} = 98.2.$$

Substitute into the formula:

$$z = \frac{98.2 - 98.6}{0.62/\sqrt{106}}.$$

$$z = \frac{-0.4}{0.62/\sqrt{106}}.$$

$$z \approx -6.64.$$

So we need

$$P(\bar{x} \leq 98.2) = P(z \leq -6.64).$$

This probability is extremely small. Using a table, this is essentially

$$0.0001 \text{ or smaller.}$$

Solution

Interpretation
If the true population mean were really $98.6°$F, then getting a sample mean of $98.2°$F or lower from 106 subjects would be extremely unlikely. Because this probability is so small, the sample mean $98.2°$F is significantly low.
This suggests that the common belief that the population mean body temperature is $98.6°$F may not be correct. The true mean may be closer to $98.2°$F.

7 Finite Population Correction

Most of the time, the formula

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

works well. However, if we sample without replacement from a finite population and the sample is a large part of the population, we should adjust the standard error.

Finite Population Correction

When sampling without replacement and

$$n > 0.05N,$$

where N is the population size, multiply the standard error by

$$\sqrt{\frac{N-n}{N-1}}.$$

So the adjusted standard error is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

If the sample size is no more than 5% of the population size, the correction is usually not needed.

8 Common Mistakes to Avoid

Be Careful

- Do not use σ when the problem involves $\bar{x}$. Use $\sigma/\sqrt{n}$.
- Do not apply the central limit theorem when the population is not normal and $n \leq 30$.
- Do not confuse an individual value x with a sample mean $\bar{x}$.
- Remember that the distribution of sample means has less variation than the original population.

Summary

Key Ideas

- The central limit theorem allows us to use a normal distribution for sample means when $n > 30$, or when the original population is normal.
- For sample means, the mean is

$$\mu_{\bar{x}} = \mu.$$

- For sample means, the standard deviation is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

- For an individual value, use

$$z = \frac{x - \mu}{\sigma}.$$

- For a sample mean, use

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

- The rare event rule helps us decide whether a sample result is unusual enough to question an assumption.