

**Dec 1, 2022**

# **Engineering Statistics**

**Week 6: Sampling distributions**

**©Mustafa Cavus, Ph.D.**

# Introduction

# The concepts of statistical inference

## Population

A population is the complete set of all items that interest an investigator.

## Sample

A sample is an observed subset of a population.

# The concepts of statistical inference

## Sample

A sample is an observed subset of a population.

# The concepts of statistical inference

**Parameter:** A numerical feature of a population is called a parameter.

- The true value of a population parameter is an unknown constant.
- It can be correctly determined only by a complete study of the population.
- However, studying on a population is not practical because of time and cost restrictions.
- Therefore, a **proper sample** from the population should be drawn and inferences about a parameter should be based on this sample.

# The concepts of sampling distribution

# Sampling from a population

The observations  $X_1, X_2, \dots, X_n$  are a **random sample** of size  $n$  from the population distribution if they result from **independent selections** and **each observation has the same distribution as the population**.

# Sampling from a population

Therefore, a random sample is a collection of independent and identically distributed (iid) random variables.



# Sampling from a population

A statistic is a numerical valued function of the sample observations.

# Sampling from a population

Let  $X_1, X_2, \dots, X_n$  be a sample from a population having a parameter  $\theta$ . Any function of the sample is called **statistic or point estimator\*** of  $\theta$ , i.e.,

$$\hat{\theta} = T(X_1, X_2, \dots, X_n)$$

# Sampling from a population

- Statistic or point estimator are used in same manner.
- A statistic or point estimator is also a random variable.
- A value that the point estimator takes called as point estimate and denoted by lower case letters, i.e. while  $T$  is point estimator of the parameter  $\theta$ ,  $t$  is point estimate.

# Example: Normal distribution case

Let  $X_1, X_2, \dots, X_n$  be a sample from  $N(\mu, \sigma^2)$  distribution,

- The statistic  $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  is a point estimator of population mean  $\mu$ .
- The statistic  $\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  is a point estimator of population variance  $\sigma^2$ .

# Example: Binomial distribution case

Let  $X_1, X_2, \dots, X_n$  be a sample from  $\text{Binom}(n, p)$  distribution. The statistic,

$$\hat{p} = \frac{\text{number of success in the sample}}{n}$$

is a point estimator of population proportion  $p$ .

# List of population parameters and their point estimators

	Population parameter	Point estimator
Mean	$\mu$	$\bar{X}$
Variance	$\sigma^2$	$s^2$
Standard deviation	$\sigma$	$s$
Proportion	$p$	$\hat{p}$

# Sampling distribution

- Since any statistic (point estimator) is a random variable, it has its own probability distribution.
- The distribution of a statistic is called its **sampling distribution**.
- The sampling distribution of a statistic is determined from the distribution of the population, and it also depends on the sample size  $n$ .

# Sampling distribution

- Populations for various statistical studies are modeled as random variables whose probability distributions have a mean and variance, which are generally not known as we conduct our statistical sampling and analysis.
- We will select a sample of observations-realizations of a random variable from our population and compute **sample statistics** that will be used to obtain inferences about the population, such as the population mean and variance.
- To make inferences we need to know the **sampling distribution** of the observations and the computed sample statistics.



# Example

In a single toss of a fair coin, let  $x$  equal the number of heads observed. Now consider a sample of  $n = 2$  tosses. Find the sampling distribution of  $\bar{x}$ , the sample mean.

# Example

Outcome	Probability	$\bar{x}$
HH	1/4	1
HT	1/4	0.5
TH	1/4	0.5
TT	1/4	0

Take H = 1 and T = 0

# Example

$\bar{x}$	0	0.5	1
$p(\bar{x})$	0.25	0.50	0.25

# Properties of sampling distributions

## Unbiasedness

If the expected value of a sample statistic is equal to the population parameter, the statistic is said to be an unbiased estimator of the parameter, i.e.

$$E(T) = \theta \rightarrow T \text{ is unbiased estimator of } \theta$$

**If the expected value of the sampling distribution is not equal to the parameter, the statistic is said to be a biased estimator of the parameter.**

# Properties of sampling distributions

## Standard error

The standard deviation of a statistic is called standard error (SE) of the statistic.

$$SE(T) = \sqrt{Var(T)}$$

# Example

In a single toss of a fair coin, let  $x$  equal the number of heads observed. Now consider a sample of  $n = 2$  tosses. **Find the standard errors of the sample mean.**

Sampling distribution of  $\bar{X}$

# Sampling from a normal distribution

Let the random variables  $X_1, X_2, \dots, X_n$  denote a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Then, sampling distribution of the  $\bar{X}$  is normal with mean  $\mu$  and variance  $\frac{\sigma^2}{n}$ , i.e.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$E(\bar{X}) = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}, \quad SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$



# Sampling from a normal distribution

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

# Example

The weight of a pepperoni and cheese pizza from a local provider is a random variable whose distribution is normal with mean 16 ounces and standard deviation 1 ounce. You intend to purchase four pepperoni and cheese pizzas. What is the probability that:

- (a) The average weight of the four pizzas will be greater than 17.1 ounces?
- (b) The total weight of the four pizzas will not exceed 61 ounces?

# Example

(a) The average weight of the four pizzas will be greater than 17.1 ounces?

$$P(\bar{X} > 17.1) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{17.1 - 16}{1/\sqrt{4}}\right) = P(Z > 2.2)$$

# Example

(b) The total weight of the four pizzas will not exceed 61 ounces?

$$P(X_1 + X_2 + X_3 + X_4 \leq 61) = P(\bar{X} \leq 15.25)$$

$$P(\bar{X} \leq 15.25) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{15.25 - 16}{1/\sqrt{4}}\right) = P(Z \leq -1.5)$$

# Example

A manufacturer of automobile batteries claims that the distribution of the lengths of life of its best battery has a mean of 54 months and a standard deviation of 6 months. Suppose a consumer group decides to check the claim by purchasing a sample of 50 batteries and subjecting them to test that estimate the battery's life.

- (a) Assuming that the manufacturer's claim is true, describe the sampling distribution of the mean lifetime of a sample of 50 batteries.
- (b) Assuming that the manufacturer's claim is true, what is the probability that the consumer group's sample has a mean life of 52 or fewer months?

# Example

(a) Assuming that the manufacturer's claim is true, describe the sampling distribution of the mean lifetime of a sample of 50 batteries.

$X$ : lifetime of a battery

$\bar{X}$ : mean lifetime of the batteries

$$\bar{X} \sim N(54, \frac{6^2}{50})$$

# Example

(b) Assuming that the manufacturer's claim is true, what is the probability that the consumer group's sample has a mean life of 52 or fewer months?

$$P(\bar{X} \leq 52) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{52 - 54}{6/\sqrt{50}}\right) = P(Z \leq -2.357023)$$

# Course materials

You can download the notes and codes from:

[https://github.com/mcavs/ESTUMatse\\_2022Fall\\_EngineeringStatistics](https://github.com/mcavs/ESTUMatse_2022Fall_EngineeringStatistics)





# Contact

Do not hesitate to contact me on:



[https://twitter.com/mustafa\\_cavus](https://twitter.com/mustafa_cavus)



<https://www.linkedin.com/in/mustafacavusphd/>



[mustafacavus@eskisehir.edu.tr](mailto:mustafacavus@eskisehir.edu.tr)