BUS 310 – Lesson 1 Notes

Sampling and Estimation

1. Introduction to Sampling and Estimation

In statistics, we are often interested in learning about a population, but studying the entire population is usually impractical due to time, cost, or accessibility constraints. Instead, we collect data from a sample and use it to make inferences about the population.

The goal of sampling is to use information from a subset of the population to estimate unknown characteristics of the population accurately.

2. Key Definitions

Population

- The **entire group** of individuals, objects, or measurements of interest.
- Example: All registered voters in Virginia, all GMU students, or all manufactured light bulbs from a factory.

Sample

- A **subset** of the population selected for analysis.
- Example: 200 randomly selected GMU students.

Parameter

- A **numerical summary** that describes a population.
- Usually unknown.
- Examples:
  - Population mean (μ)
  - Population proportion (π)

Statistic

- A **numerical summary** calculated from a sample.
- Used to estimate population parameters.
- Examples:
  - Sample mean ($\bar{x}$)
  - Sample proportion ($\hat{p}$)

3. Purpose of Statistical Inference

In a typical statistical inference problem:

1. Identify the population of interest.
2. Take a sample from that population.
3. Use sample statistics to estimate population parameters.
4. Evaluate whether observed sample results are reasonable given assumptions about the population.

Example:

- Assume the population mean age is $\mu = 50$.
- A sample produces a mean $\bar{x} = 20$.
- Statistical inference helps determine whether this result is likely or unusual.

4. Types of Sampling Methods

Sampling methods are divided into **probability** and **non-probability** sampling.

5. Non-Probability Sampling

Judgment (Subjective) Sampling

- Items are chosen based on expert judgment rather than chance.
- Probability of selection is unknown.

Advantages

- Easy and inexpensive
- Quick to conduct

Disadvantages

- High risk of bias
- Cannot measure sampling error
- Results may not represent the population accurately

6. Probability Sampling

In **probability sampling**, selection is based on chance, and every element has a **known probability** of being selected.

Types of Probability Samples

- Simple Random Sample
- Systematic Sample
- Stratified Sample
- Cluster Sample

7. Simple Random Sampling

Definition

- Every population element has an **equal chance** of being selected.
- Selection of one element does not affect others.

Methods

- Random number tables
- Computer-generated random numbers
- Lottery or "fishbowl" method

Advantages

- Easy to understand
- Minimizes selection bias

Disadvantages

- May not represent important subgroups well
- Requires a complete list of the population

8. Systematic Sampling

Procedure

1. Determine sample size (n)

2. Determine population size (N)
3. Calculate k = N / n
4. Randomly select a starting point
5. Select every kth element thereafter

Example

- N = 40, n = 4 → k = 10
- Commonly used in telephone surveys

Advantages

- Simple and fast
- Cost-effective

Disadvantages

- Bias may occur if there is **periodicity** in the population list

9. Stratified Sampling

Definition

- Population is divided into **mutually exclusive and exhaustive subgroups (strata)** based on a common characteristic.
- A simple random sample is taken from each stratum.

Examples of Stratification

- Age groups
- Gender
- Income level
- Commuters vs residents

Advantages

- Ensures representation across the entire population
- Produces more precise estimates

Disadvantages

- More complex
- Requires detailed population information

10. Cluster Sampling

Definition

- Population is divided into clusters that each represent the population.
- A random sample of clusters is selected.
- All elements or a random subset within clusters are studied.

Examples

- Election exit polls using voting districts
- Surveying selected schools instead of all students

Advantages

- More cost-effective
- Useful when population is geographically dispersed

Disadvantages

- Less precise
- Requires larger samples to achieve the same accuracy

11. Comparing Sampling Methods

Method	Precision	Cost
Simple Random	Moderate	Moderate
Systematic	Moderate	Low
Stratified	High	Higher
Cluster	Lower	Low

12. Introduction to Estimation

The purpose of sampling is to **estimate unknown population parameters** using sample data.

Estimators

- Sample mean

- Sample variance
- Sample proportion

Point Estimate

- A single value used to estimate a population parameter.
- Example: $\bar{x} = 50$ estimates μ

13. Sampling Error

- Sampling error is the difference between the sample estimate and the true population parameter.
- It exists because samples are only a subset of the population.
- Sampling error **cannot be eliminated**, but it can be reduced.

Effect of Sample Size

- Larger sample sizes → smaller sampling error

14. Sampling Distribution of the Mean

- The distribution of sample means from all possible samples of size n.
- Mean of the sampling distribution equals the population mean (μ).

15. Standard Error of the Mean

- The standard deviation of the sampling distribution.
- Measures how much sample means vary from sample to sample.
- As sample size increases, standard error decreases.

16. Central Limit Theorem (CLT)

Key Results

1. If sample size is large enough, the sampling distribution of the mean is approximately normal, regardless of population shape.

2. If the population is normally distributed, the sampling distribution is normal for any sample size.

The CLT allows us to apply normal probability methods to sample means.

17. Applications of the Sampling Distribution

- Used to calculate probabilities involving sample means
- Used for:
    - Confidence interval estimation
    - Hypothesis testing

Important distinction:

- Individual observations → use population standard deviation
- Sample means → use **standard error**

18. Types of Statistical Inference

1. **Confidence Interval Estimation**
    - Provides a range of values around a point estimate
2. **Hypothesis Testing**
    - Determines whether sample evidence supports a specific claim