FE 243 Introduction to Statistics

- 3 hr/week
- 70 % attendance
- Instructor: Dr.Medeni MASKAN

Exams

- Midterm I : 30 %
- Midterm II: 30 %
- Final Exam: 40 %

Course Contents

Basic Terms
Measure of Central Location
Measure of Variation
The Normal Distribution
Sampling Distribution
Test of Hypothesis About a Single Mean - One-tailed - Two-tailed
F-Distribution
Comparing Two Population Means
Comparing Two Population Means - Test of Hypothesis For Paired Observations
Chi-Square Distribution - Parametric - Nonparametric
Linear Correlation
Linear Regression
Analysis of Variance (ANOVA)

FE 243 Introduction to Statistics

What is statistics ?

- **A branch of mathematics dealing with the collection, analysis, interpretation and presentation of numerical data.**
- **Basically a tool to facilitate decision making.**
- **Any numerical value describing a characteristic of a sample.**

Population: A group of individual persons, or items from which samples are taken for statistical measurement.

Sample: A finite part of a statistical population whose properties are studied to gain information about the whole.

The sample;

1) should be representative of population.
2) " " taken from all available population.

Example: The students of Food Eng. Dept. can be assumed as population.

=> The first year students can be a sample.

Simple Random Sample: A simple random sample of "n" observations is a sample that is chosen in such a way that every subset of "n" observations of the population has the same probability of being selected (if no bias).

e.g., Game cards => 52 cards.

Probability of every card to be selected is 1/52 in a population of size 52.

Types of Statistical Applications

Statistics involves two different processes:

1) **Describing sets of data**
2) **Drawing conclusions (e.g., making estimates, decisions, predictions etc.) about the sets of data on the basis of sampling.**

So, the applications of statistics can be devided into two broad areas;

a) **Descriptive Statistics: It utilizes numerical and graphical methods to look for patterns in a data set, to summurize the information revealed in a data set and to present that information in a convenient form.**

It is an expensive and time consuming method.

b) **Inferential Statistics: It utilizes sample data to make estimates, decisions, predictions or other generalizations about a larger set of data (i.e., population).**

Data: Any recorded event (e.g., # of students of Food Eng. Dept.)

Information: A collection of numbers (data)

Knowledge: Useful data.

Our purpose is to transform data into knowledge.

Data → Information → Knowledge.

Variables: Variables refer to something being observed that exhibits variation. e.g., height, weight of babies with time.

1) Independent variables: Those that can be controlled and act upon the system. e.g., T → independent variable.

2) Dependent variables: Those that are result or yield (MC, PV etc).

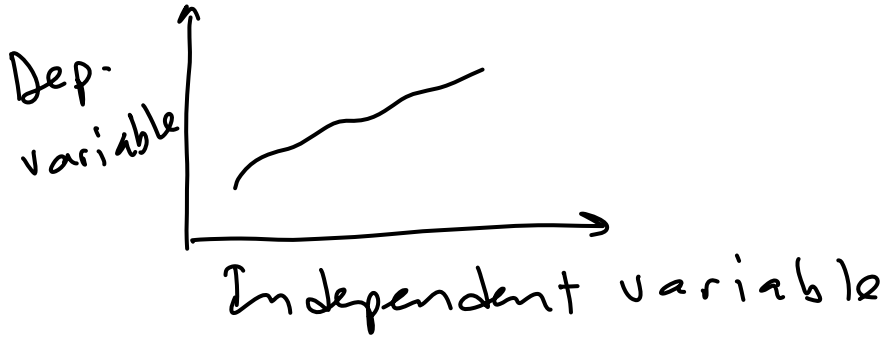
e.g., absorption of moisture from a humid atmosphere by a dry food at constant T ⇒

- MC is dependent variable
- T ∴ independent ∴

e.g., Oxidation of oil at different T's.

T ⋮ PV ⋮ ⇒ T ↗ ⇒ PV ↗

T: independent, PV: dependent variable.



Parameter: Any numerical value describing a characteristic of a population is called a parameter.
It is a constant value describing the population.
e.g., mean of a population (μ) or st. dev. (σ) etc.

Basic Steps in Statistical Analysis

1) Statement of the problem.
2) Organization of ideas and theories.
3) Statement of a hypothesis

4) Testing the stated hypothesis
5) Confirmation or denial (reject) of the stated hypothesis leading to a decision.

Decisions can be;
a) Reject the hypothesis

OR

b) Fail to reject (i.e., do not reject) the hypothesis.

6) Conclusion derived from analysis. It must be consistent with the steps 3 and 5.

Measures of Central Locations (MCL)

Any measure indicating the center of a set of data, arranged in an increasing or decreasing order of magnitude, is called a measure of central location.

The most commonly used MCL are the mean, median and mode.

Population mean (M): If the set of data $x_1, x_2, \dots, x_N$, not necessarily all distinct, represents a finite population of size N, then the population mean is

$$M = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

N → population size.

Example: The number of employees at 5 different drugstores are 3, 5, 6, 4 and 6. Treating the data as a population, find the mean number of employees for the 5 stores.

Solution:

$$\mu = \frac{3+5+6+4+6}{5} = 4.8$$

Sample mean ($\bar{x}$): If the set of data $x_1, x_2, \cdots, x_n$, not necessarily all distinct, represents a finite sample size of n, then the sample mean is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$n \rightarrow$ sample size.

Example: The data recorded from 7 cans of a food to determine % foreign impurities are 1.8, 2.1, 1.7, 1.6, 0.9, 2.7 and 1.8. Compute the sample mean.

Solution:

$$\bar{x} = \frac{1.8+2.1+\cdots+1.8}{7} = 1.8\%$$

Median : The median of a set of observations arranged in an increasing or decreasing order of magnitude is the middle value when the number of observations is odd. OR the arithmetic mean of the two middle values when the number of observations is even.

Example: On 5 term tests in a course a student has made grades of 82, 93, 86, 92 and 79. Find the median for this populations of grades.

Solution:

Arrange the grades in an increasing order $\Rightarrow$

1 2 3 4 5

79, 82, 86, 92, 93 $\Rightarrow$ odd observations.

Median = 86

Example: Nicotines in cigarettes of a certain brand are 2.3, 2.7, 2.5, 2.9, 3.1 and 1.9 mg for a random sample of 6 cigarettes. Find the median.

Solution:

Arrange data $\Rightarrow$

1 2 3 4 5 6

1.9, 2.3, 2.5, 2.7, 2.9, 3.1 $\Rightarrow$ even observations.

$$\text{Median} = \frac{2.5 + 2.7}{2} = 2.6 \text{ mg}$$

Mode: The mode of a set of observations is that value which occurs most often or with the greatest frequency.

⊛ The mode does not always exist.

Example: Consider the following data: 9, 10, 5, 9, 9, 7, 8, 6, 10 and 11. What is the mode ?

Solution:

The value with the greatest frequency is 9 (repeating 3 times) ⇒ The mode is 9.

Example: Given the data 2, 0, 3, 1, 2, 4, 2, 5, 4, 0, 1, 4. What is the mode ?

Solution:

In this case, there are two modes: 2 and 4.

The distribution is said to be bimodel.

Geometric Mean (G): The G of k positive #s, $X_1, X_2, \cdots, X_k$ is the k^{th} root of their product.

⊛ If any observation is zero ⇒ G cannot be calculated.

$$G = \sqrt[k]{X_1 \cdot X_2 \cdots X_k}$$

Example: For a sample of 3, 5, 6, 4, 6 ⇒ k=5

$$G = \sqrt[5]{3 \times 5 \times 6 \times 4 \times 6} \simeq 4.6$$