

STP 226 ELEMENTARY STATISTICS

NOTES

PART 1 - INTRODUCTION

CHAPTER 1

THE NATURE OF STATISTICS

Definition of Statistics from the Webster's New World Dictionary:

Statistics:

1. Facts or data of a numerical kind, assembled, classified, and tabulated so as to present significant information about a given subject.
2. The science of assembling, classifying, and tabulating such facts or data.

Statisticians today, not only assemble, classify, and tabulate data; but they also analyze data to make generalizations and decisions.

1.1 Statistics Basics

Descriptive Statistics – consists of methods for organizing and summarizing information.

Example - listing of the height of all students in this STP226 class.

Inferential Statistics – consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample of the population.

Example - using the heights of all students in this STP226 class to predict/estimate the heights of all students at ASU.

Population - collection of all individuals or items under consideration in a statistical study.

Example - all students at ASU, if they are all, that one is interested in.

Sample - that part of the population from which information is collected.

Example - students in this class, if interest is in all students at ASU.

Classifying Statistical Studies

Descriptive Studies - if the purpose of the study is to examine and explore information only for its own intrinsic interest. Descriptive study may be performed either on a sample or on a population. Only if an inference is made about the population based on the sample results, does the study become inferential.

Example: The Election committee in a Math Department summarized the results of an election of a Personal Committee Members:

Johnes 235 votes
Lebioda 24 votes
Lucas 15 votes

Inferential Studies - using the information gathered in a sample to make generalization, draw a conclusion about the population of interest.

Example: In 1977 major leagues started using baseballs from Rawling Company. Pitchers started complaining that new balls are harder and bounce farther and faster. An independent testing company randomly selects 85 baseballs from 1977 supplies and measures bounce, weight and hardness of all of them. These measurements are then compared to measurements obtained from similar tests in years 1952, 1953, 1961, 1963, 1970 and 1973 when baseballs from Spalding Co. were used. The conclusion of the study was that new baseballs are livelier than other balls but not outside of the rules imposed by major league.

Observational studies and Designed Experiments.

In an **Observational Study** researchers simply observe characteristics and take measurements, as in a sample survey.

Ex. A sample of 82 adults were given a flu shots, after two months it was it turned out that 20 of them got the flu, others did not.

This study can't claim that flu shots were effective against flu, people that did not get sick may have been for ex. following a better diet than the ones that got sick and that is why they did not get a flu, not because of the flu shot. There may be an association between the shots and decreased likelihood of the flu, but causal effect is not established here.

In a **Designed Experiment** researchers impose treatments and controls and then observe characteristics and take measurements.

Ex. A sample of 82 adults was randomly divided into 2 groups of 41, one group was given a flu shots, the other group a placebo, not an active treatment. Both groups were following similar diet and exercise regime. After two months it turned out that only 9 of the adults that received flu shots got sick with the flu, as compared to 20 sick in a placebo group.

This study shows that flu shots had a positive effect on the people that took them, treatment group had much fewer flu cases than the placebo group while both groups were controlled as to their diet and and exercise regime. We may claim that flu shots indeed decreased likelihood of the flu.

Observational studies can only reveal association, whereas designed experiments can help establish causation.

1.2 Simple Random Sampling

Is a Study necessary?

There is always need for people to gather information so that they can compare. They do this by conducting studies: eg. study the gas mileage on certain types of cars, study the merits of two teaching methods, etc

Before doing a study, make sure the same study is not done already, this is a waste of time, energy, and money.

- do a literature search to see if the study has been done before
- you do not have to go through all the books in the library because there are many **information-collection agencies** who have access to such information
- eg. The Educational Resources Information Center gathers results of studies in psychology and publish them in its publication "Psychological Abstracts".
- there are these agencies in just about every subject area.

It is cheaper and easier to use the results of a similar study than to redo the whole study.

Simple Random Sampling

Gathering information through a census (from whole population) is generally time consuming, costly, frequently impractical, and sometimes impossible.

Other than a census, information can be gathered by **Sampling** and **Experimentation**.

We will concentrate mostly on Sampling.

We always want a **representative sample** - should reflect as closely as possible the relevant characteristics of the population under study/consideration

Example - average weight of a sample of football players to make inference about all adult males is **not representative**

Example - estimate median income of all California residents by sampling income of Beverly Hills residents is **not representative**

Most sampling procedures today involve use of **Probability Sampling**. In that type of sampling random numbers table, random numbers generator or simply coin toss are used to decide which member of the population will be selected to a sample. These methods still can produce sample that is not representative, but problems like unintentional bias are eliminated. Use of probability sampling allows researchers to control the chance of obtaining unrepresentative sample and allows them to apply techniques of inferential statistics.

Simple Random Sampling - a sampling procedure for which each possible sample of a given size is equally likely to be the one obtained.

- can be done by tossing a coin, rolling a die, picking from a hat

- **With replacement** - a member can be selected more than once
- **Without replacement** - a member can be selected at most once (**method of this course !!!**)

Simple Random Sample - a sample obtained by simple random sampling.

Example - population of 5: A, B, C, D, E

Sample of size 2:

{A, B} {A, C} {A, D} {A, E} {B, C} {B, D}
 {B, E} {C, D} {C, E} {D, E}

Each of these samples of size 2 is equally likely to be the sample chosen (chance of any one to be selected is 1/10)

Random Number Tables - randomly generated numbers

When population is large, picking from a hat, rolling a die, or tossing a coin is not practical.

Example - want sample of 15 from 728, random numbers is more practical to use

Numbers from 1 to 728 represent members (consider 3 digits). Close your eyes and place your finger on a starting point, circle 3 digits, move downwards and select all valid numbers until 15 numbers from 001 to 728 are selected. Invalid numbers here are 000 and numbers 729 and greater.

Calculators and computers have random numbers generating functions and make the job even easier.

For Ex. TI- 83 has a function **randInt (n,k)** that will give you numbers from possible range n to k, for ex randInt(2,10) will select numbers from 2 to10, selections are with replacement.

This function in found in MATH menu, select option PRB, randInt is function #5.