

Chapter 1

Logan | Nick

1.1 Populations, Samples, and Processes

Statistics provides methods for organizing and summarizing data and for drawing conclusions from that data

Def

Data : a collection of facts

Def

Population : A well defined collection of objects for which we wish to obtain info

Def

Census : When desired info is obtained from every member of the population

- problems : Time, money, practical

Def

Sample : A subset of the population

1.

You want the home price in Edwardsville

- Fewer well trained appraisers gives better results than many poorly trained

2. Tree Age Study

Testing is destructive, so a sample is better

Def

variable : any characteristic whose value may differ from one subject to another.

- denote with low letters

Note

- Don't say \$McDonald's = 10\$

- Do say \$x = \$ the length of the tibia bone in 10 year old boys.

Def

univariate data : result from making observations of 1 variable

- these variable can be qualitative / quantitative

Def

Bivariate data : when observations are made on each of 2 variables for each individual

- (weight.mpg) of cars

Def

Multivariate data : observations made on many variables

- patient data

Ex

Labor force, sample 60,000, find population + sample

- population = labor force, sample size = 60,000 households

Branches of Stats

1. Descriptive Stats : data are collected and you wish to summarize and describe features of the data (graphs, numerical summaries)
2. Inferential stats : data is collected from a sample and used to draw a conclusion about the population
 - confidence intervals, hypothesis test, prediction, etc...

Types of sampling

- Simple random sampling : random choice / draw of the hat sampling
- Systematic sampling : selecting every k^{th} member of the population
- Cluster sampling : divide population into groups, then select some of these groups @ random
- Stratified sampling : divide population into groups. Find subgroups of groups (strata) and then draw random sample in strata
- Convenience sampling : sampling in the most convenient way
 - best to avoid , but a good starter

Notate

sample size : n

- For a dataset with n observations on some variable x , the individual observations will be denoted as x_1, x_2, \dots, x_n .

#1.2

stem and leaf plots

Ex

(54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22)

During these problems it helps to first organize the numebrs in the list first

2 | 2, 5

3 | 4, 5

4 | 1, 1, 6, 6, 6, 7, 9

5 | 4, 4, 9

6 | 0