

1 Lecture 1 Notes

Definitions:

- Population - _____
- **Population** is defined as a complete collection of all elements that are of interest.
- Census - _____
- A **census** is defined as a list of data from every member of the population. A census is not always possible, imagine studying all individuals in the United There are people that are homeless or are here momentarily.
- Sample - _____
- A **sample** defined as a sub-collection of members from part of the population. Normally collected using a random method. Extra: Think of ways of generating a random sample.

Information:

- Data - _____
- **Data** is defined as observations, measurements, of elements of the sample that is of interest
- Statistics - _____

- **Statistics** is collection of methods for planning experiments, obtaining data; then organizing, summarizing, analyzing, interpreting, presenting and drawing conclusions based on data.

Information from the Population and Sample:

- Parameter - _____
- **Parameter** is a numerical measurement describing some characteristic of the population
- Statistic - _____
Statistic is a numerical measurement describing some characteristic of sample that is a proxy of the parameter. Since we cannot always know the parameter we use the statistic

Table 1: My caption

Information	Population	Sample
Mean	μ	\bar{x}
Standard Deviation		
Variance		
Proportion		

Table 2: My caption

Information	Population	Sample
Mean	μ	\bar{x}
Standard Deviation	σ	s
Variance	σ^2	s^2
Proportion	p	\hat{p}

Types of Information:

- Qualitative (categorical) - _____
 - Nominal Level - _____
 - Nominal Level - Categorical data only. Data can not be arranged in order. (i.e., YES/NO/NA or iPhone/Android/NA)
 - Ordinal Level - _____
 - Ordinal Level - Data can be arranged in order. Differences are meaningless. (i.e., A/B/C/D or Like it/Love it/Got to Have it)
- Quantitative (numerical) - _____
- Quantitative (numerical) - Counts or Measurements (i.e., # of red cars or weight)
 - Interval Level - _____
 - Interval Level - Like Ordinal but differences are meaningful. There is no natural zero starting point.(i.e., 1988/1989)
 - Ratio Level - _____
 - Ratio Level - Similar to interval level with a zero starting point. The zero starting point makes ratios meaningful. (i.e., weight, height)

Collection of Data

1. Various ways of collecting data:

- (a) Experiments _____

- (b) Experiments - apply a treatment to determine if it has an effect on individuals
- (c) Observational Study _____
- (d) Observational Study - observe and measure characteristics of a sample, we do nothing to the sample

2. Design of experiment:

- (a) Replication _____
- (b) Replication is to repeat an experiment with many individuals to measure a particular effect (i.e., multiple individuals receive treatment and multiple individuals don't)
- (c) Blinding _____
- (d) Blinding the individual does not know treatment type. This is a way to remove the placebo effect. Double blind is when both the subject and researcher does not know the treatment type (i.e., give treatment but patient does not know which drug was given)
- (e) Randomization _____
- (f) Randomization individuals are randomly assigned to treatment type, this is important to remove confounding factors (i.e., assign random number between 1 and number of individuals, use random number generator to select individuals to be part of each treatment)
- (g) Confounding _____
- (h) Confounding when a effect is seen but we do not know why it is present (i.e., every one at the the Dominican Hospital is older than 75)

3. Observational Studies:

- (a) Cross-Sectional Study _____
- (b) Cross-Sectional Study is when data observed, measured, and collected at one time (i.e., analyze medical records and record tonometry tomorrow)
- (c) Retrospective (or Case-Control) Study _____
- (d) Retrospective (or Case-Control) Study is when data is collected from a past time period (i.e., analyze medical records 10 years ago)

(e) Prospective (or Longitudinal or Cohort) Study _____

(f) Prospective (or Longitudinal or Cohort) Study is when data is collected in the future (i.e., design analysis for medical records 10 years from now)

4. Sampling Techniques:

(a) Simple Random Sample (SRS) _____

(b) Simple Random Sample is when everyone in the population has the same chance of being in the the sample

(c) Stratified Sampling _____

(d) Stratified Sampling is when g natural groups (strata) are in the population and a SRS is selected from each group (strata)

(e) Systematic Sampling _____

(f) Systematic Sampling is when a list of the population is created and every k^{th} individual is selected to be in sample

(g) Convenience Sampling _____

(h) Convenience Sampling is when the closest/easiest to obtain group from the population is the sample

(i) Cluster Sampling _____

(j) Cluster Sampling is when the population is divided into clusters and randomly selected in which the entire population is selected to be in the sample

(k) Volunteering Sampling _____

(l) Volunteering Sampling is when individuals volunteer to be a part of the sample