

# STATS 250 NOTES\* (condensed) Exam 1

## Chapter 1

**\*\*Report decimal places 3-4 places after the decimal point.\*\***

Stats investigation process --- ADEUFR:

1. Ask a Research Question
2. Design a study and collect data
3. Explore the data, provide graphical displays, and numerical summaries
4. Use statistical analysis methods to draw inferences from the data
5. Formulate conclusions, communicate findings and answer research questions
6. Reflect and look forward(point out limitations and suggest further studies)

When data is not collected for question --- ITEUFR:

1. Import the data
2. Tidy the data
3. Explore the data, provide graphical displays, and numerical summaries
4. Use statistical analysis methods to draw inferences from the data
5. Formulate conclusions, communicate findings and answer research questions
6. Reflect and look forward(point out limitations and suggest further studies)

Categorical Variables --- or qualitative variables

- place an individual or item into one of the several groups or categories, which are called **levels**.
- Nominal variables vs. Ordinal variables, ordinal follows specific order(small, medium, large)
- They report frequency counts
- Relative frequency - Decimals or Percent
- Graphs used to display -- bar charts, pie chart

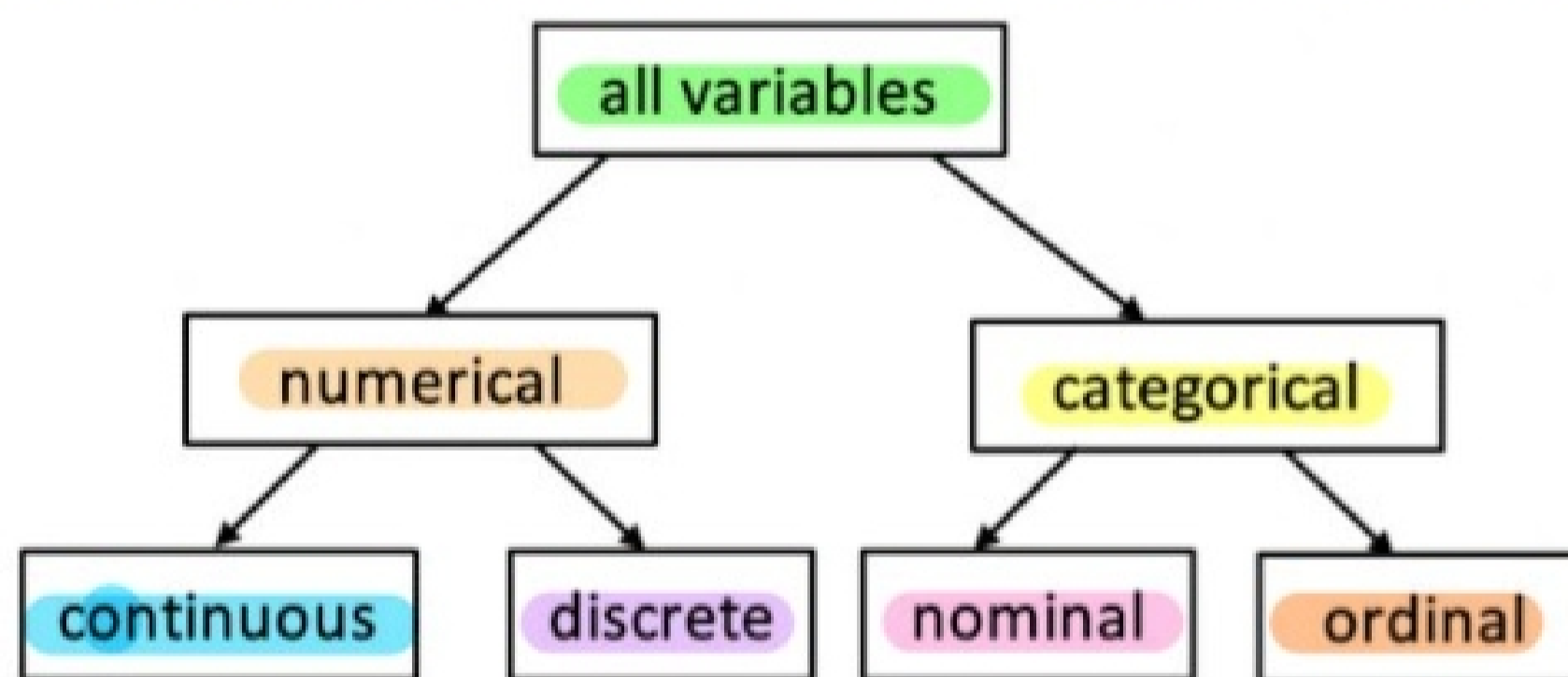
- Example: Ice cream flavors are a categorical variable, and how many flavors are the levels, and how many people like which flavor would be frequency counts, and when you report the counts in decimals or percentages its considered relative frequency

### Numerical Variables --- or quantitative or measurement

- variables that take on a wide range of numerical values, and it is sensible to do math with numerical variables.
- Discrete: numerical variables with jumps. Ex: 1, 2, 3, 4, 5
- Continuous variables: Can take any value in an interval or collection of intervals. Ex: 2.3, 6.7, 7.808, 8.9999
- Numerical variables can have subtypes. Ex: ages 18-25, 25-34, 35-44...
- It's possible to turn numerical values to categorical by grouping them, but not categorical values to numerical
- Graphs: Histograms, boxplots, scatterplots

**Contingency tables:** one-way data is summarized with tables

**Data Matrix:** Common way to organize raw, unprocessed data. They have columns and rows.



**Population:** The entire group we are interested in learning about. All undergraduate students in the US.

- We dont observe every case in a study. Would be time-consuming and costs too much, and sometimes can destroy the item in the process of measurement

**Sample:** A subset of the cases that is often a small fraction of the overall population.

### Undergraduate students un UMICH

- It provides an estimate for the overall population, less time-consuming, and less costly.
- There could be biases in a sample so the way we sample is important

- Biases: Convenience sampling, Response bias, Non-response bias

**Anecdotal Evidence:** Typically composed of unusual cases that are recalled based on their striking characteristics.

**Sampling from a Population:** To draw inferences about a population

1. The sample must be representative of the entire population
2. Use random sampling: subjects of study/experiment should be selected randomly to ensure the sample is representative

**Explanatory and Response variables**

- **Explanatory:** a variable that predicts the outcome
- **Response:** a variable that is the outcome --- responds to the explanatory variable

**Tip: Identifying explanatory & response variables**

To identify the explanatory variable in a pair of variables, identify which of the two is suspected of affecting the other.



**Two types of data collection: Observational and Experimental**

- **Observational Studies:** refer to instances where researchers collect data in a way that does not directly interfere with how the data arise. They simply observe.
  - Interested in looking at the relationship between two or more variables
  - Data is usually collected only by monitoring what occurs
  - Making causal inferences based on observational studies is difficult but not impossible
- **Experiments:** researcher directly influences the process by which data arise.
  - Subjects are usually assigned to one or more treatments RANDOMLY
  - There is usually a control variable or a placebo effect
  - Require the primary explanatory variable in a study be assigned to each subject by researchers
  - Making causal conclusions is reasonable, depending on the way the explanatory variable is assigned.