

## Lesson 1.3 - February 4, 2021

---

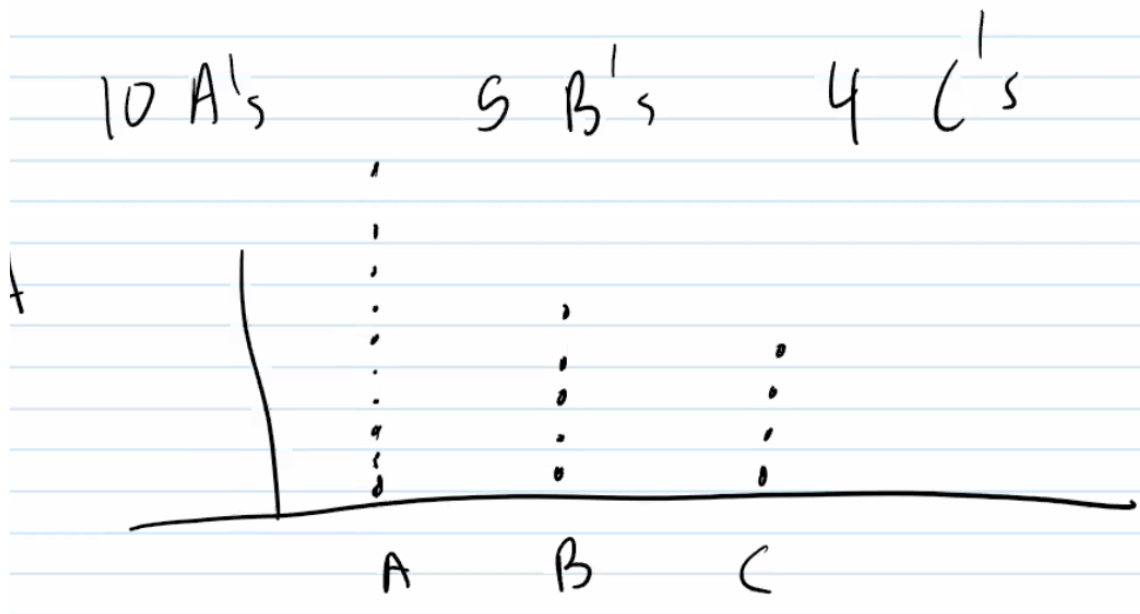
### 1.3 Extension

- If you are not directly assigning groups for control and treatment groups, then it is an observational study
  - We collect data about a sample
    - Data can either be categorical or quantitative
    - Categorical data can be numerical, but may be something like a session ID
    - Quantitative data is a numerical value, like how many minutes a person spends on a website
  - **StatCrunch**: The built-in data analysis software that can be accessed via the [MyLab](#) and [Mastering](#) tab on Canvas
  - It is important to find which questions a dataset can or cannot answer
  - Finding patterns help us interpret the world through numbers or through generic ideas
- 

### Analyzing Data Using Graphs

- All of these describe the frequency of certain values in a specific data set for each variable
- **Dot Plots**
  - Useful for a small number of values of categories
  - Can be used for categorical data

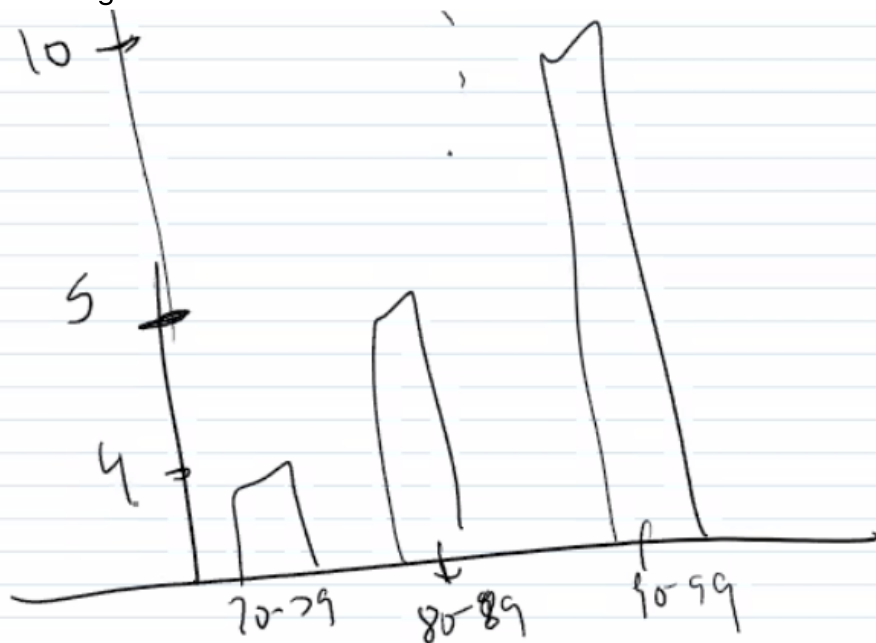
- Example Dot Plot:



- When there are too many categories, there would be too many categories and the dot plot would be too wide-spread
- To solve the above problems, we group similar values together, leading to the histogram

- **Histograms**

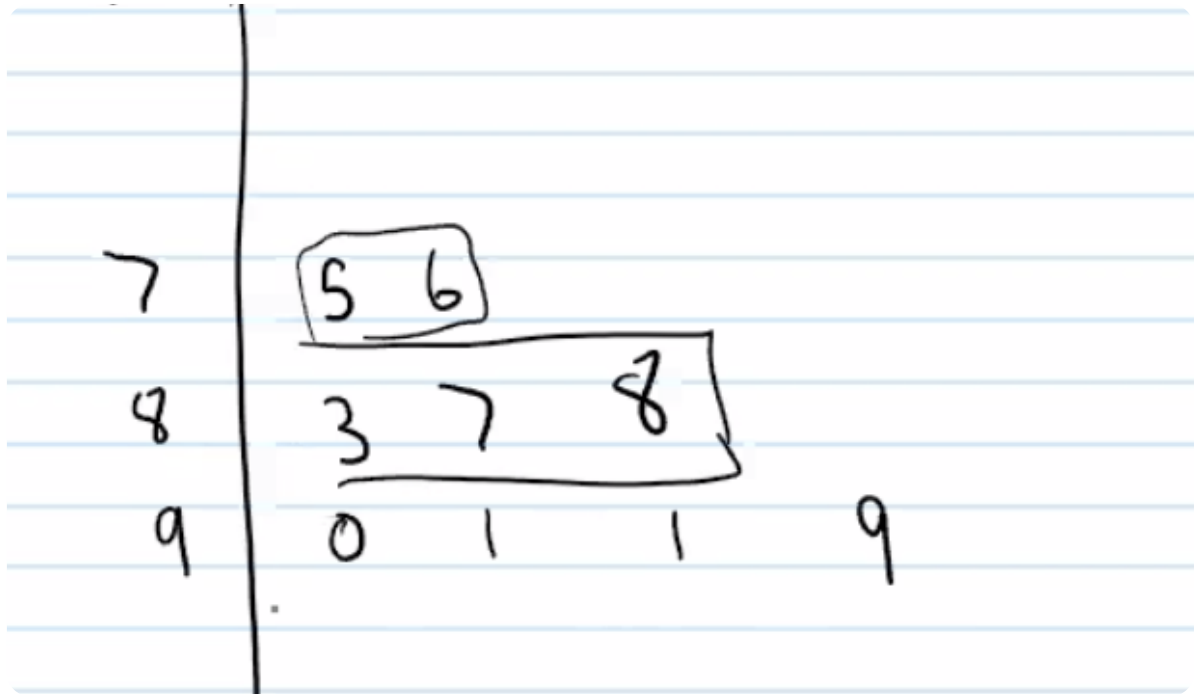
- Useful when there's not a good way to split up the values or there are many possible values in the dataset
- Groups similar values together, but based off the dot plot
- Example Histogram:



- Dots are replaced with bars with a nicely scales axis for the  $y$  axis
- Idea of "bins" breaks down the data into separate categories, by which we can graph them

- **Stem and Leaf Plots**

- Most useful when working by hand
- Otherwise, is one of the least useful types of graphs
- Sideways histogram where the bins are chosen for you
- On the left side of the  $y$  axis, put every number except the last digit
- On the right side of the  $y$  axis, put the last digit
- Example stem and leaf plot:



---

## Statcrunch Specifics

- In Statcrunch, you can graph histograms and other graphs by clicking on [Graph](#) and going into the sub-menu and clicking on the type of graph you want to make
- When graphing a histogram, you can let Statcrunch automatically create bins and decide their widths and their heights
- Any fields that are marked `--optional--` do not need to be filled out and can be calculated automatically
- If you make a graph, you can adjust the size (width and height) to see what the graph would look like if it was expanded to full length
- Statcrunch will not let you make histograms for categorical data
- Each bin's width represents the range of the variable that we are grouping the data into

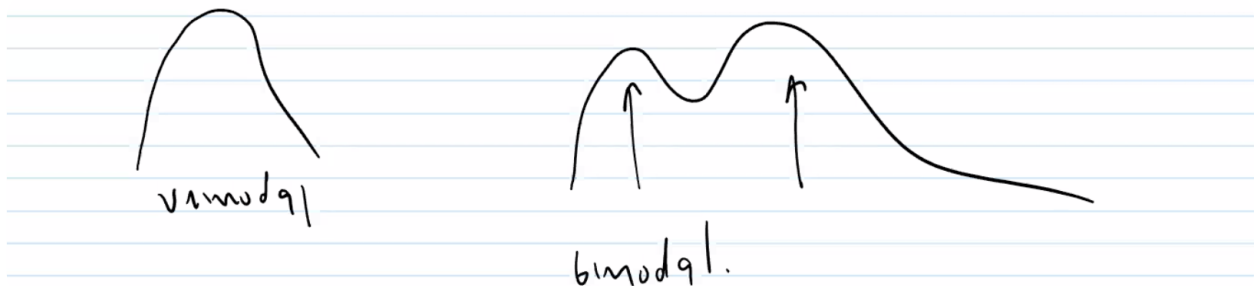
---

## Graph Patterns

- Skew says which direction the graph's "tail" is "pointed" towards, points to outliers in that direction
- Outliers are data points that are far away from most of the other points
- Sometimes outliers aren't real and could be a mistake
- Think about how much effect you want an outlier to have on the final results of your study
- Symmetric graphs have no skew whatsoever and have the maximum value directly in the middle of the graph
- The peak of the graph represents the most frequently occurring value of the entire set
- In an income graph, a certain extremely rich person (Mark Zuckerberg, for example) could "pull" the graph towards their end of the graph
- Increasing the number of bins and decreasing the width of each bin in a histogram provides a cleaner curve
- A curve is skewed to a direction if there is an outlier in that direction

## Modal Graphs

- Unimodal means the graph has a singular peak
- Bimodal means the graph has 2 or more peaks
- Example unimodal/bimodal graph:



- There is usually an explanation for bimodal graphs; always question why the bimodal graph may have been caused

noelle marty