

# Descriptive Statistics II

## Histograms

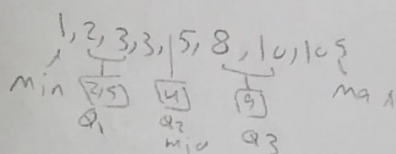
Histogram: The most common visual for quantitative data



## 5 Number Summary

- Gives values for calculating the range and interquartile range

Max  
Q<sub>3</sub> Third quartile  
Q<sub>2</sub> Second = median  
Q<sub>1</sub> First  
Min



$$\text{Range} = \text{max} - \text{min}$$

$$\text{IQR} = Q_3 - Q_1$$

## Standard Deviation

- on average, how much each point varies from the mean of the points

To calculate STD

$$1. \bar{x} = \frac{\sum x_i}{n}$$

$$2. x_i - \bar{x}$$

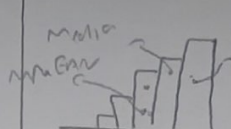
$$3. (x_i - \bar{x})^2$$

$$\text{Variance} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Average squared difference of each observation from the mean

$$\text{STD} = \sqrt{\text{variance}}$$
$$= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

## Distributions



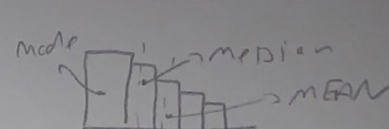
LEFT  
SKEWED

mean < median



SYMMETRIC

mean = median = mode



RIGHT  
SKEWED

mean > median

## Outliers

- Data points that fall very far from the rest of the values in our dataset

① they exist and the impact on summary statistics

② if type, remove or fix

③ understand why exist and its impact on questions we are trying to answer

④ reporting the 5-number summary is better than mean and standard deviation when outliers are present

⑤ be careful in reporting know how to ask the right questions

## Guidelines to analysis data

1. Plot the data

2. make a decision on outlier

## Bell-shaped data

- we can learn a lot with the mean and standard deviation in normally distributed data

## Skewed data

- Five number summary is best for skewed data

## Inferential Statistics

- Drawing conclusions about a population based on data collected from a sample of individuals from that population