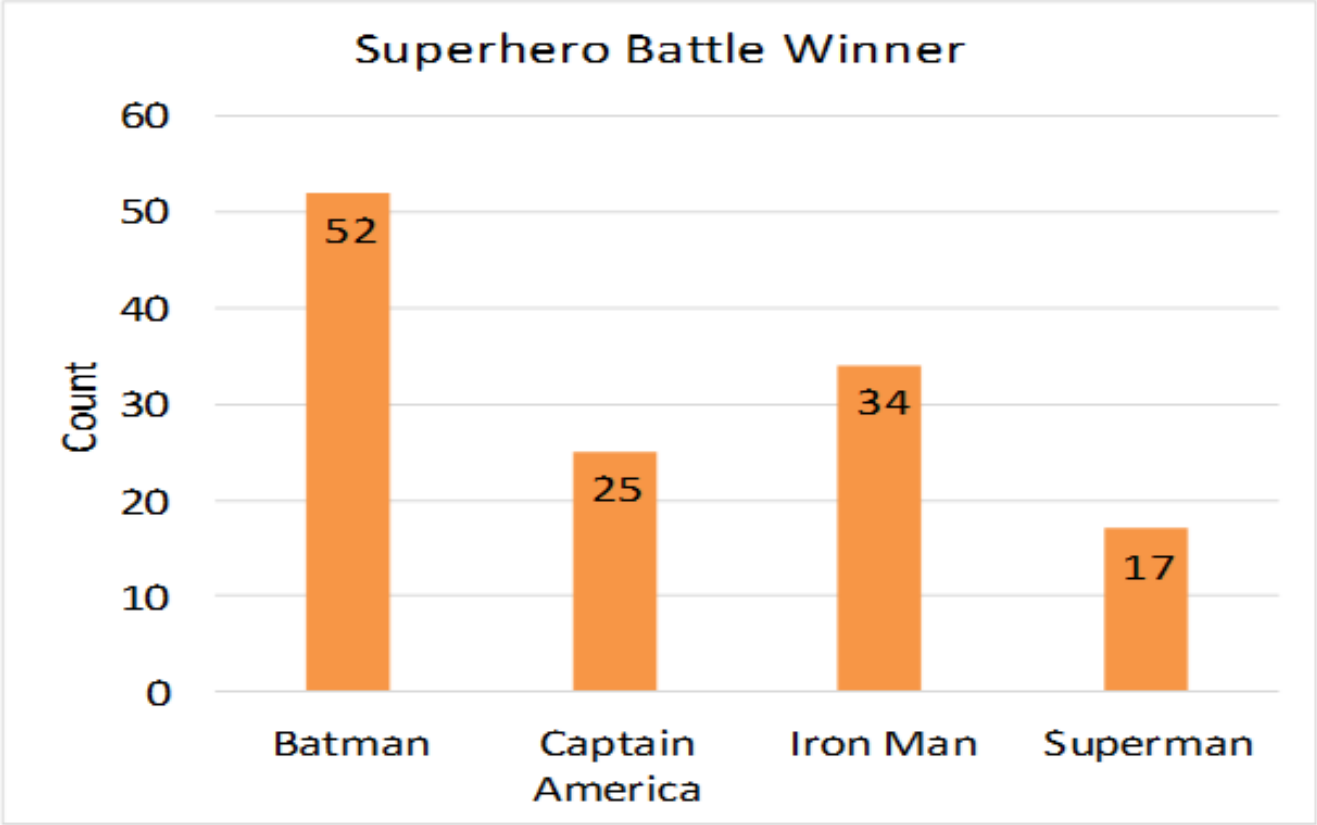
Describing Data

Introduction

- Descriptive Statistics allow to characterize data based on its properties.
- There are four major types of descriptive statistics:
  - Measures of Frequency
  - Measures of Central Tendency
  - Measures of Dispersion or Variation
  - Measures of Position

Measures of Frequency

- Count, Percent, Frequency
- Shows how often something occurs
- Use this when you want to show how often a response is given



Measures of Central Tendency

- Mean, Median, and Mode
- Locates the distribution by various points
- Use this when you want to show how an average or most commonly indicated response

Measures of Dispersion or Variation

- Describe how similar or varied the set of observed values are.
- Include the
  - range,
  - quartiles and the interquartile range,
  - variance and
  - standard deviation
- Can be measured for quantitative data.
- Use this when
  - You want to show how "spread out" the data are.
- It is helpful to know when your data are so spread out that it affects the mean.

Why do we measure spread?

- Summarizing the dataset can help us understand the data.
- The mode, median, and mean summarize the data into a single value. This is only part of the 'picture' that summarizes a dataset.
- Measures of spread summarize the data in a way that shows how scattered the values are and how much they differ from the mean value.

Variance and Standard Deviation

- Variance: The average of the squared differences from the mean.
- Standard Deviation: The square root of the variance and is used to measure distance from the mean.

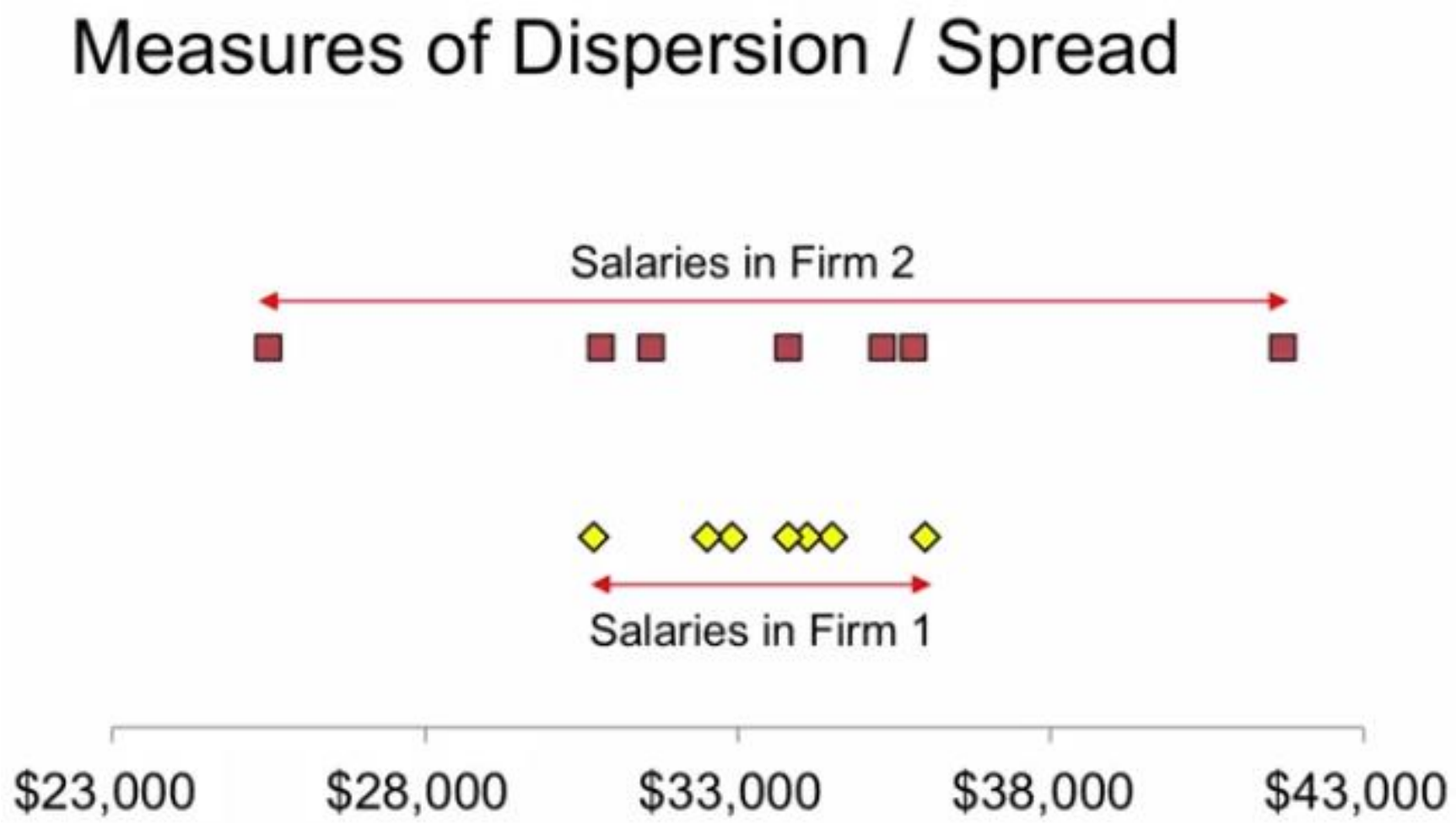
Example

Measures of Dispersion / Spread

Firm 1	Firm 2
$34,500	$35,800
$30,700	$25,500
$32,900	$31,600
$36,000	$41,700
$34,100	$35,300
$33,800	$33,800
$32,500	$30,800
Mean = $33,500	**Mean = $33,500**
Median = $33,800	**Median = $33,800**

Mean/Median is same for both.
How to define that both datasets are different?

Example

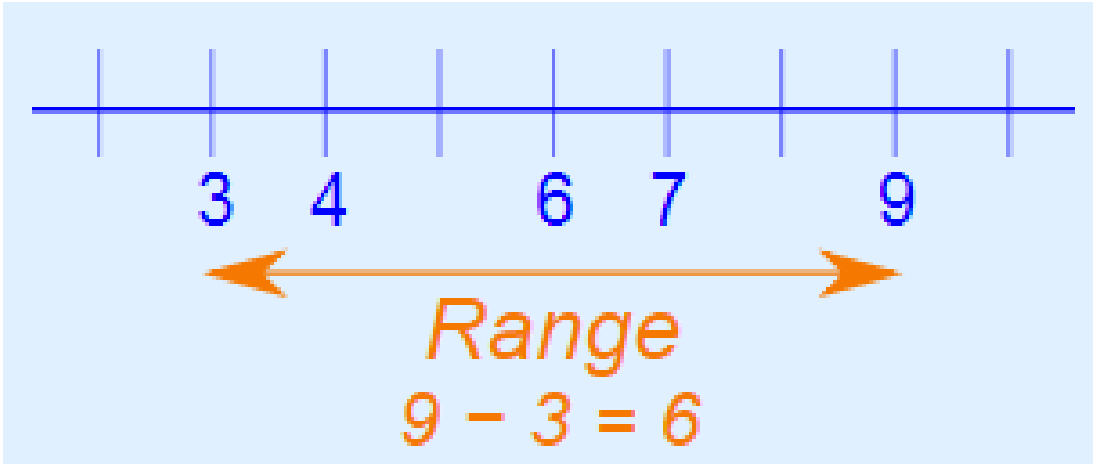

Measures of Dispersion / Spread
Salaries in Firm 2
Salaries in Firm 1
$23,000
$28,000
$33,000
$38,000
$43,000

Range

- Difference between Maximum and Minimum
- Range = Max Value – Min Values

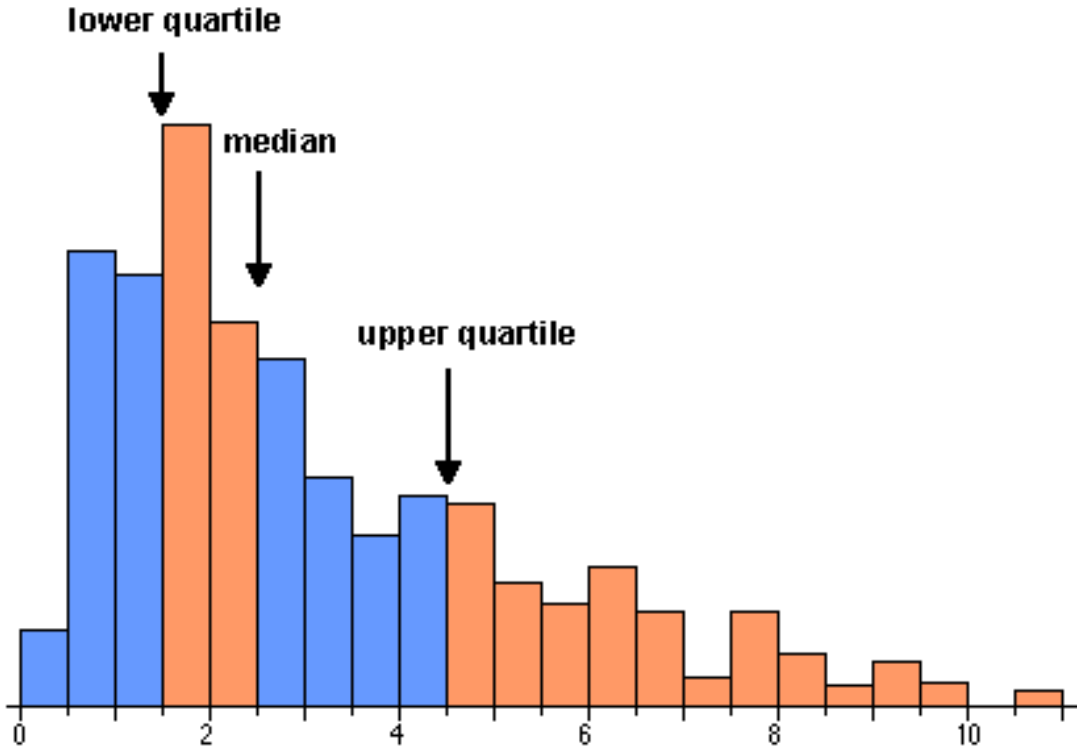
- Range of salaries in firm 1
- =36000 – 30700 = 5300

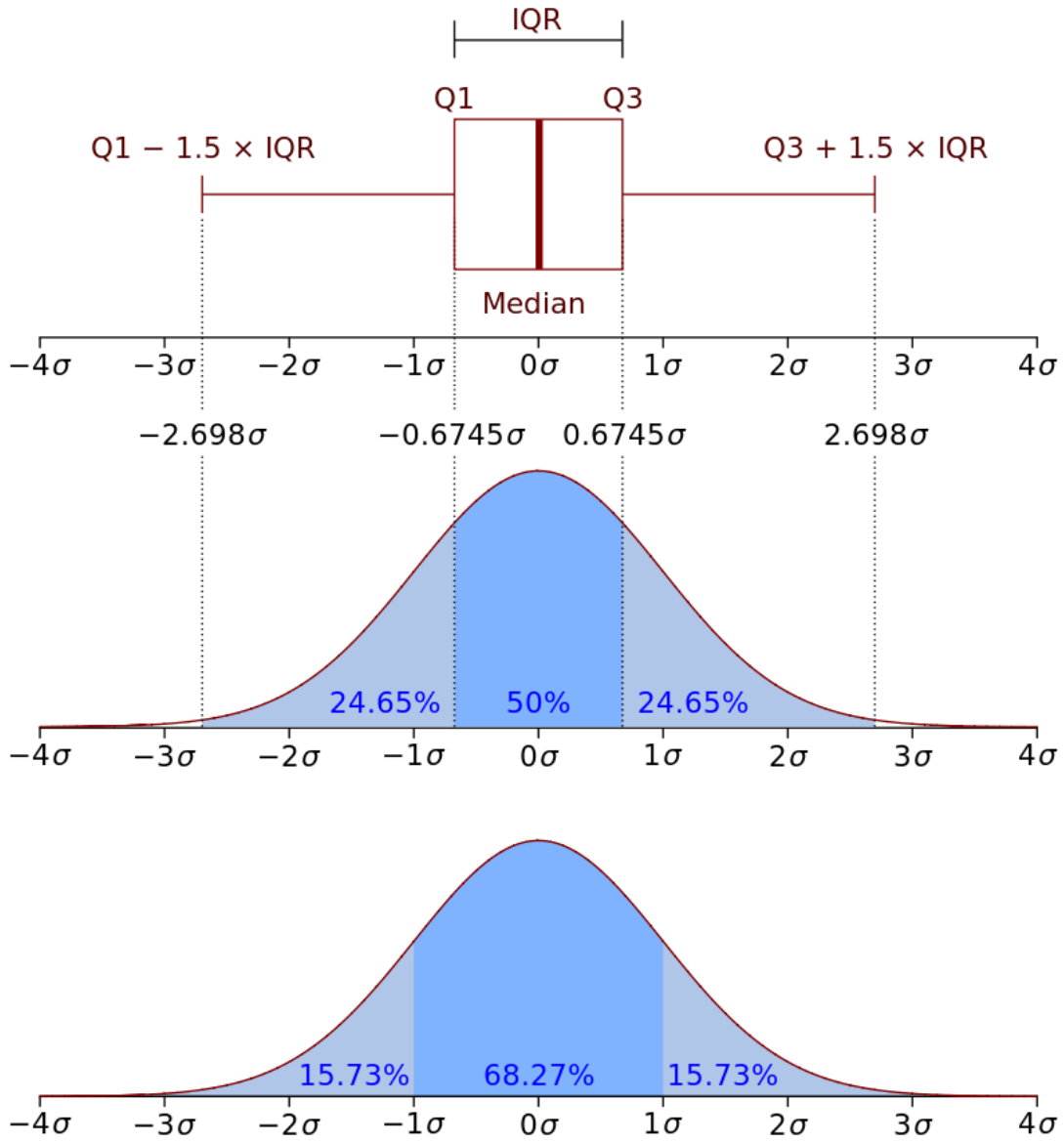
- Range of salaries in firm 2
- =41700– 25500= 16200



Measures of Position

- Percentile Ranks, Quartile Ranks



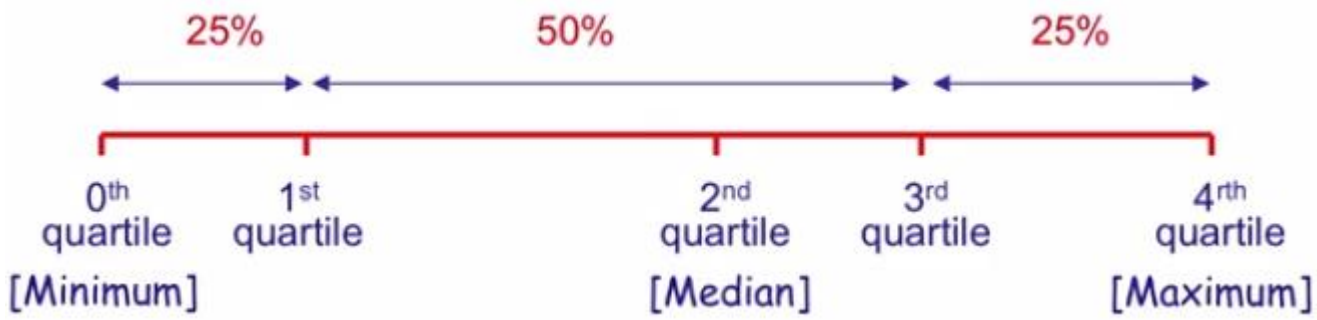


Inter Quartile Range (IQR)

- Defines middle 50% of data.
- 25% data of the left and 25% of the right will be discarded.

Inter Quartile Range (IQR)

- 1st Quartile: Is the number such that 25% of the observations are <= to this number.
- 3rd Quartile: Is the number such that 75% of the observations are <= to this number.
- 2nd Quartile: Is the median.
- Minimum number in the range is at 0th Quartile.
- Maximum number in the range is at 4th Quartile.



Thanks