# Descriptive Statistics Data Prosentation review 6 review 7

be adequate to describe data

- There are many more measures used
  - Measures of central tendency
    - Mean, trimmed mean, median, mode
  - Measures of dispersion
    - Standard deviation, variance, range, interquartile range, mean absolute deviation, coefficient of dispersion
  - Measures on individual data points
    - Standard deviation unit, standard score

### Data Organization

- Graph of interval vs frequency
- Bar graph for grouped data
  - Original frequency distribution table
  - X-axis: cell boundaries
  - Y-axis: frequency



#### **Frequency Distribution Table Cell Boundaries** f Xi [42 - 47) 5 44.5 [47 - 52) 6 49.5 [52 - 57) 8 54.5 [57 - 62) 59.5 11 [62 - 67) 64.5 10 69.5 [67 - 72) 7 [72 - 77) 4 74.5 [77 - 82] 2 79.5

Histogram done in Excel

### Descriptive Statistics Data Organization

- Graph of interval vs frequency
- Bar graph for grouped data
  - Original frequency distribution table
  - X-axis: cell boundaries
  - Y-axis: frequency



Frequency Distribution Table				
Cell Boundaries	f <sub>i</sub>	x <sub>i</sub>		
[42 - 47)	5	44.5		
[47 - 52)	6	49.5		
[52 - 57)	8	54.5		
[57 - 62)	11	59.5		
[62 - 67)	10	64.5		
[67 - 72)	7	69.5		
[72 - 77)	4	74.5		
[77 - 82]	2	79.5		

Histogram done in Minitab

Minitab used different cell boundaries: has 10 cells each with width of 4

Lower bound: 42 (same) Upper bound: 82 (same)

### Outline of Discussion Notesale, Signal of 67 preview from 22 of 67

- Descriptive Statistics
  - Data Organization
  - Data Analysis
  - Statistical Measures
- Case Study

## Descriptive Statistics Data Approvision of 67 Previsorplage 28 of 67

- How to interpret a boxplot (box and whisker plot)
  - Some boxplots have outliers
  - Outliers are values which are beyond the whiskers but within three interquartile ranges from the box edge
  - Outliers farther than three interquartile ranges are called extreme outliers



## Descriptive Statistics Data Approvision of 67 Previscatper Prot

- How to interpret a scatter plot
  - If the points form a diagonal line, variables are correlated
  - Perfect horizontal line means correlation is zero



#### Descriptive Statistics • Statistical Measures • Measures of Central Tendency

- Describes the tendency of sample data to cluster around a particular value
  - Mean
  - Median
  - Mode

#### Descriptive Statistics • Statistical Measures • Measures of Variability/Dispersion

- Example:
  - Determine the 64<sup>th</sup> percentile, range, and interquartile range of the following data set
  - Quiz Question

Number of hours of sleep per day					
9	7	8	6	Н	
6	5	6	9	4	

- Hint: Arrange first in increasing order: 4, 5, 6, 6, 6, 7, 8, 9, 9, 11
- Hint: X<sup>th</sup> percentile = X<sup>\*</sup>(n+1), range = max min, interquartile range = 75<sup>th</sup> percentile – 25<sup>th</sup> percentile

#### Descriptive Statistics • Statistical Measures • Measure of Linear Relationship

- Correlation coefficient
  - Determines the linearity between variables of a population

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

- Pearson's r coefficient
  - Determines the linearity between variables of a sample

$$r = \frac{n \sum XY - \sum X \sum Y}{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}$$

Nonzero correlation coefficient means there is a linear relationship

- Zero correlation coefficient means either they are independent of each other or their relationship is nonlinear
- Excel scatterplot uses r<sup>2</sup> which is more accurate and reliable

#### Outline of Discussion Notesale. preview from Notesale. preview from 60 of 67

- Descriptive Statistics
  - Data Organization
  - Data Analysis
  - Statistical Measures
- Case Study

### Summary • Data Organization previewers Distribution Table

- Histogram
- Cumulative Frequency Distribution Table
- Ogive
- Pareto Chart