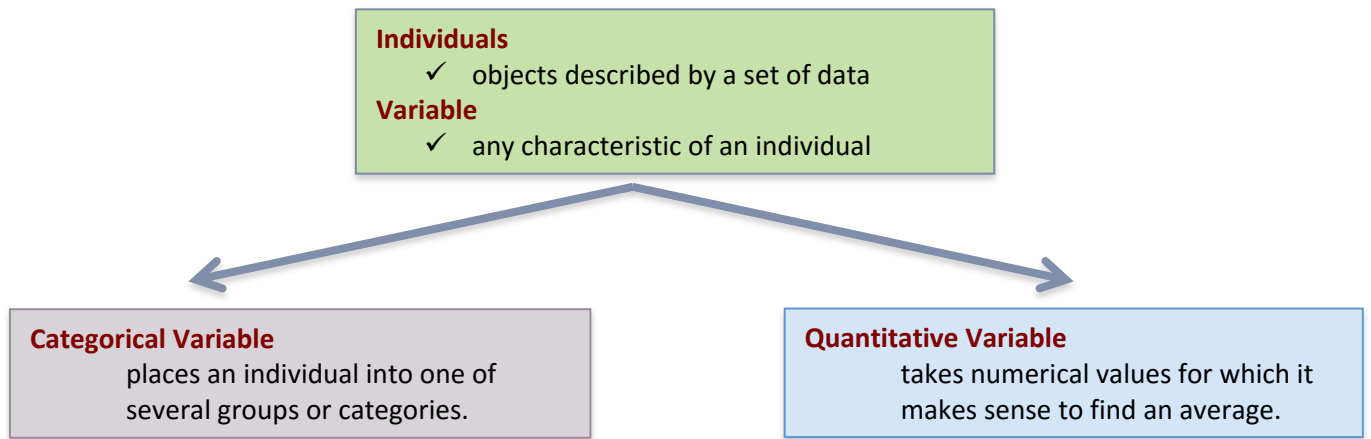


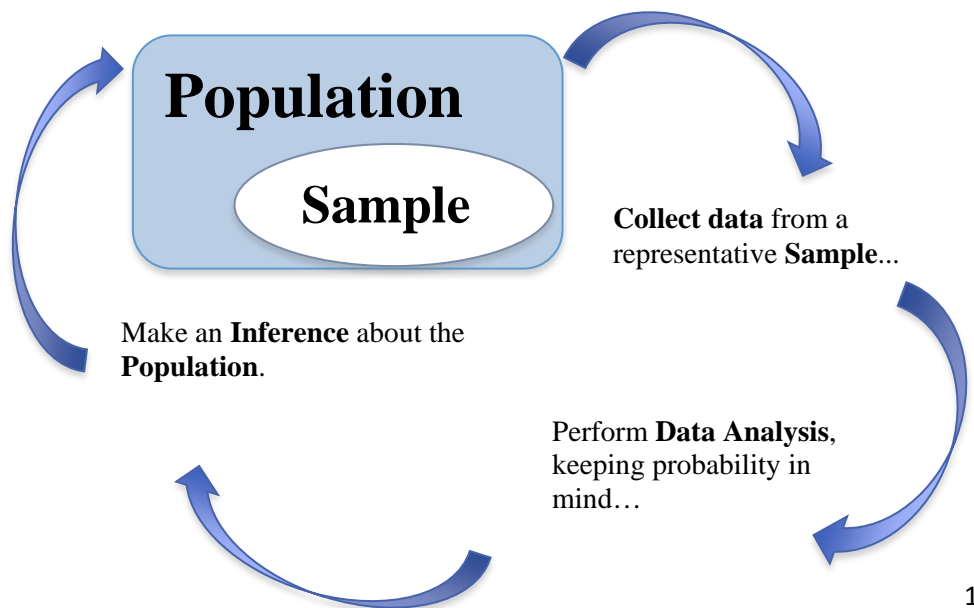
Section 1.0 – Making Sense of Data

Statistics:

Data Analysis:



Distribution:



Section 1.1 – Analyzing Categorical Data

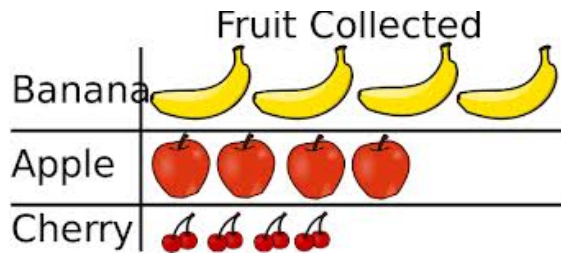
Counts v. Percents

Bar Graphs

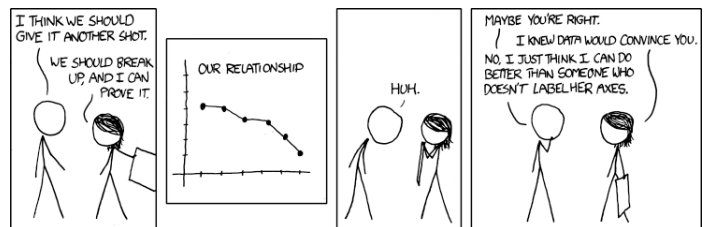
Pie Charts

Scales

Pictographs



Class/Bins, Labels & Consistent Units



Two-Way Tables:

Example on page 12

Young adults by gender and chance of getting rich			
	Female	Male	Total
Almost no chance	96	98	194
Some chance, but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

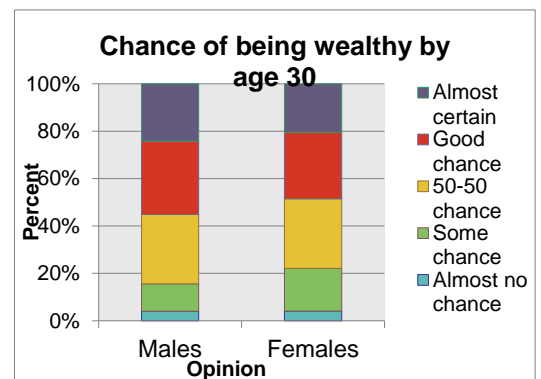
Response	Percent
Almost no chance	
Some chance	
A 50-50 chance	
A good chance	
Almost certain	



Write a short description of your graph:

Side-by-side bar graph

Segmented/stacked bar graph



Marginal Distribution:

Round Off Error:

Conditional Distributions:

Response	Male
Almost no chance	
Some chance	
A 50-50 chance	
A good chance	
Almost certain	

Notes: When Describing Categorical Data: Percents are often more useful than counts

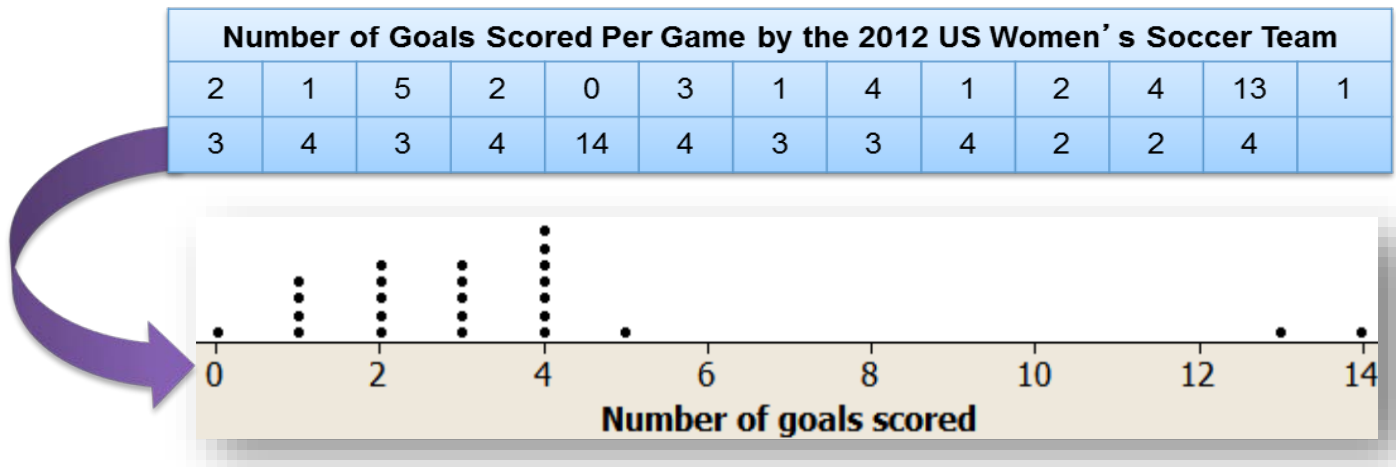
No single graph can show all the data from a two-way table easily

Caution!

Even a strong association between two categorical variables can be influenced by other variables lurking in the background.

Section 1.2 – Displaying Quantitative Data with Graphs

Dotplots:



Describing the Pattern of a Distribution

- 1) In any graph, look for the **overall pattern** and for striking **departures** from that pattern.
- 2) Describe the overall pattern of a distribution by its:
 - **Shape**
 - **Center**
 - **Spread**

Shape:

Approximately Symmetric

Skewed Right

Skewed Left

AP Statistics Notes for Chapter 1

Modality

Unimodal

Bimodal

Describe the shape of the dotplot on page 5 of your notes:

Center:

Spread:

Unique Characteristics:

Outliers:

Deviations from Pattern

**For more details see the examples starting on page 26 of your textbook.*

Stem & Leaf Plot (Stemplot)

Regular Stemplot: 2, 8, 20, 21, 21, 22, 32, 33, 34, 35, 36, 49, 49, 50

Generally when making a stemplot you want to have at least 4 stems. If there are too few stems it is difficult to see the shape; but the same is also true for too many stems!

Split Stems: 19, 20, 21, 22, 22, 24, 25, 25, 26, 28, 30, 30, 33, 39

Stemplot Key

Back-to-back Stemplots: A: 11, 12, 20, 34

B: 11, 12, 13, 23, 28, 29, 36, 37, 38

Histogram

How to make a histogram:

- 1) Divide the range of data into classes of equal width.
- 2) Find the count (frequency) or percent (relative frequency) of individuals in each class.
- 3) Label and scale your axes and draw the histogram. The height of the bar equals its frequency. Adjacent bars should touch, unless a class contains no individuals.

Calculator:

Using frequency list to plot histograms when there are many repeated numbers:

Class

Frequency

Relative Frequency

Cumulative Frequency Graphs & Charts

Cautions!

- 1) Don't confuse histograms and bar graphs.
- 2) Don't use counts (in a frequency table) or percents (in a relative frequency table) as data.
- 3) Use percents instead of counts on the vertical axis when comparing distributions with different numbers of observations.
- 4) Just because a graph looks nice, it's not necessarily a meaningful display of data.

Section 1.3 – Describing Quantitative Data with Numbers

Measures of Center: Mean v. Median

Mean:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Non-Resistant Measure

Median (M)

Method of Calculation:

If odd number of numbers

If even number of numbers

Resistant Measure

What does comparing the Mean and Median tell us?

Excerpt from: **“America's middle class: Poorer than you think”** By Tami Luhby (CNNMoney)

*Rich Americans. That's our global reputation. The numbers seem to back it up. Americans' average wealth tops \$301,000 per adult, enough to rank us fourth on the latest Credit Suisse Global Wealth report. But that figure doesn't tell you how the middle class American is doing. Americans' median wealth is a mere \$44,900 per adult -- half have more, half have less. That's only good enough for 19th place, below Japan, Canada, Australia and much of Western Europe. "Americans tend to think of their middle class as being the richest in the world, but it turns out, in terms of wealth, they rank fairly low among major industrialized countries," said Edward Wolff, a New York University economics professor who studies net worth. **Why is there such a big difference between the two measures?***

AP Statistics Notes for Chapter 1

Use the data below to calculate the mean and median of the commuting times (in minutes) of 20 randomly selected New York workers.

10	30	5	25	40	20	10	15	30	20	15	20	85	15	65	15	60	60	40	45
----	----	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Commute Times

```

0 | 5
1 | 005555
2 | 0005
3 | 00
4 | 005
5 |
6 | 005
7 |
8 | 5
    
```

Key: 4|5 represents a New York worker who reported a 45-minute travel time to work.

Find the median on the stemplot:

Find the mean and median using your calculator:

A measure of center alone can be misleading.

A useful numerical description of a distribution requires both a measure of center and a measure of spread.

Measures of Spread:

Range: the range is a single number that represents the difference between the max and the min.

In everyday language we sometimes say “the data values range from 5 to 35”. But the correct way to say this is “the Range = 30 (from 5 to 35)”.

To calculate the **quartiles**, arrange the observations in increasing order and locate the median.

1. The **second quartile Q_2** is the median of the data set. It is also the 50th percentile.
2. The **first quartile Q_1** is the median of the observations located to the left of the median in the ordered list. It is also the 25th percentile.
3. The **third quartile Q_3** is the median of the observations located to the right of the median in the ordered list. It is also the 75th percentile.
4. The **fourth quartile Q_4** is the maximum of the data set. It is also the 100th percentile.

The **interquartile range (IQR)** is defined as:

Examples to find quartiles:

1) 1,1,3,5,9,10,11

2) 4,6,6,7,8,9

3) 3,8,9,12,14,18,20,30

In addition to serving as a measure of spread, the interquartile range (*IQR*) is used as part of a rule of thumb for identifying **outliers**.

The 1.5 x *IQR* Rule for Outliers

An observation is an outlier if it falls more than 1.5 x *IQR* above the third quartile or below the first quartile. Using the fence/gateway/cut-off to find outliers.

Practice: Use the 'Commute Time' data on page 10 of your notes packet to determine if there are outliers.

The 5-Number Summary

Minimum, Q_1 , M , Q_3 , Maximum

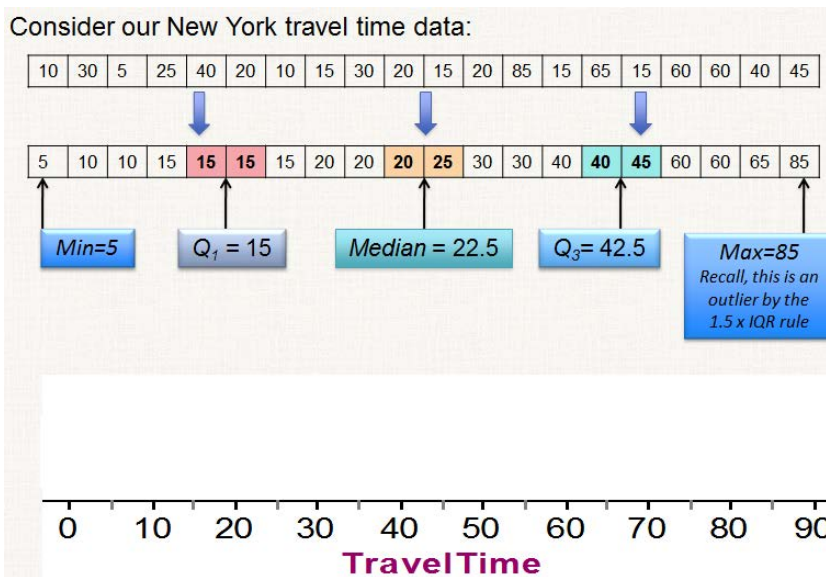
Calculator: 1-Var Stats

Boxplot (Box & Whisker Plot)

How To Make A Boxplot:

- A central box is drawn from the first quartile (Q_1) to the third quartile (Q_3).
- A line in the box marks the median.
- Lines (called whiskers) extend from the box out to the smallest and largest observations that are not outliers.
- Outliers are marked with a special symbol such as an asterisk (*).

Modified Boxplot vs. Regular Boxplot

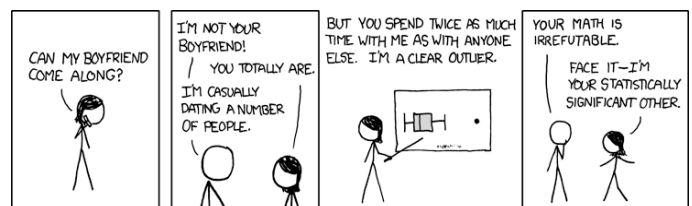


Side-by-side boxplots

Calculator:

(always select the modified boxplot)

Use 'trace' to find values on the boxplot



Standard Deviation (s_x) measures the average distance of the sample observations from their mean. It is calculated by finding an average of the squared distances and then taking the square root.

The average squared distance is called the **variance**.

Formulas:

$$\text{standard deviation} = s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

$$\text{variance} = s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

For more on standard deviation go to page 60 of your book

Choosing Measures of Center and Spread

- Numbers can lie! (or at the very least be misleading) Numerical summaries do not fully describe the shape of a distribution. **ALWAYS PLOT YOUR DATA!**
- The median and *IQR* are usually better than the mean and standard deviation for describing a skewed distribution or a distribution with outliers.
- Use mean and standard deviation only for reasonably symmetric distributions that don't have outliers.

Properties of s_x :

s_x is standard deviation of a **sample** σ_x is standard deviation of a **population**

How to Organize a Statistical Problem: A Four-Step Process

- **State:** What's the question that you're trying to answer?
- **Plan:** How will you go about answering the question? What statistical techniques does this problem call for?
- **Do:** Make graphs and carry out needed calculations.
- **Conclude:** Give your conclusion in the setting of the real-world problem.

Homework Assignments & Section Summaries

1.0 And 1.1 Homework Exercises #1 – 31 odd, and # 28, 30, 32 (skip # 9) ... 18 questions

Section 1.0 Summary

- ✓ A dataset contains information on individuals.
- ✓ For each individual, data give values for one or more variables.
- ✓ Variables can be categorical or quantitative.
- ✓ The distribution of a variable describes what values it takes and how often it takes them.
- ✓ Inference is the process of making a conclusion about a population based on a sample set of data.

Section 1.1 Summary

- ✓ DISPLAY categorical data with a bar graph
- ✓ IDENTIFY what makes some graphs of categorical data deceptive
- ✓ CALCULATE and DISPLAY the marginal distribution of a categorical variable from a two-way table
- ✓ CALCULATE and DISPLAY the conditional distribution of a categorical variable for a particular value of the other categorical variable in a two-way table
- ✓ DESCRIBE the association between two categorical variables

1.2 Homework Exercises #37 - 73 odd and # 60, 70, 72, 74, (skip # 49, 51, 57, 61, 63, 67) ... 17 questions

Section 1.2 Summary

- ✓ MAKE and INTERPRET dotplots and stemplots of quantitative data
- ✓ DESCRIBE the overall pattern of a distribution
- ✓ IDENTIFY the shape of a distribution
- ✓ MAKE and INTERPRET histograms of quantitative data
- ✓ COMPARE distributions of quantitative data

1.3 Homework Exercises #79 – 109 odd, and # 108, 110 (skip # 85, 101) ... 16 questions

Section 1.3 Summary

- ✓ CALCULATE measures of center (mean, median).
- ✓ CALCULATE and INTERPRET measures of spread (range, *IQR*, standard deviation).
- ✓ CHOOSE the most appropriate measure of center and spread in a given setting.
- ✓ IDENTIFY outliers using the $1.5 \times IQR$ rule.
- ✓ MAKE and INTERPRET boxplots of quantitative data.
- ✓ USE appropriate graphs and numerical summaries to compare distributions of quantitative variables.