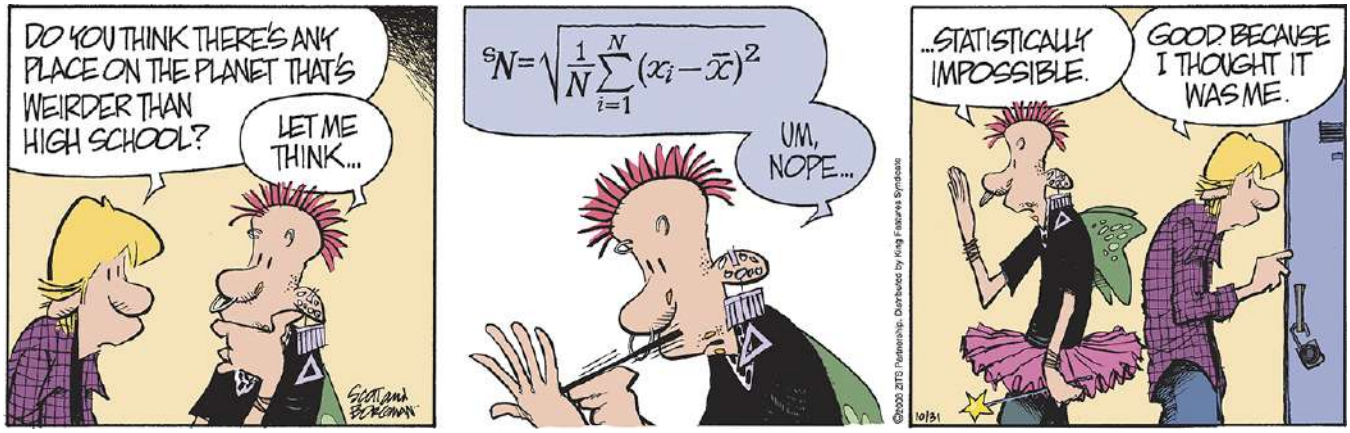
Chapter 1: Exploring Data



Reading Note:

- **Notes can be written or typed. See my website for the MSWord document for these notes.**
- **Read the "EXAMPLE" sections and make sure you understand the answers to the questions.**
- **Read "TECHNOLOGY CORNER" sections and ACTUALLY DO the examples with your calculator. Make sure you understand how to use your calculator!!!! It will be your best friend ☺**
- **You do NOT need to read the "ACTIVITY" sections!**
- **Do the "Check Your Understanding (CYU)" problems (the answers are in the back of the book).**

VOCABULARY YOU must know!

- individual
- variable
- categorical variable
- quantitative variable
- distribution
- inference
- frequency table
- relative frequency table
- pie chart
- bar graph
- segmented bar graph
- side-by-side bar graph
- two-way table (rows and columns)

- marginal distributions
- conditional distributions
- association
- dotplots
- stemplots
- histogram
- **SOCS (or CUSS & BS)**

- Resistant measures
- outlier

- **Shape**
- symmetric distribution
- Skewed to the right
- Skewed to the left
- Unimodal distribution
- Bimodal distribution
- Uniform distribution

- **Center**
- Σ
- $\bar{x}$ - sample mean
- μ - population mean
- median

- **Spread**
- variability
- s_x - sample standard deviation
- σ - population standard deviation
- variance
- range
- IQR

- **Five-number summary**
- Quartiles - Q_1, M, Q_3
- minimum
- maximum
- boxplot

Section: Introduction (pg2) - "Data Analysis: Making Sense of Data"

1. Individuals are…

2. A variable (it is **NOT X** in statistics) is …

3. Explain the difference between a *categorical* variable and a *quantitative* variable.

Give examples of *quantitative* variables:

Give examples of *categorical* variables:

- When can a *categorical* be a number and give an example:

4. Define *distribution*:

✓ CHECK YOUR UNDERSTANDING pg5 (clearly show work and write answers in sentences)

5. Explain *inference* (Use example on page 5, "From Data Analysis to Inference" to give an example of inference.)

1.I HW: Page 7 #'s 1, 3**, 7-8

These problems are the key concepts covered in this section. Your must do the **"**'d"** problems here. The other problems are for you to check your understanding and do not need to be handwritten.
Go to my web site to check your understanding and self-correct your work:
http://www.brunswick.k12.me.us/pgroves/home/ap-statistics/ap-statistics-summer-assignment/

Section: 1.1 - Analyzing Categorical Data

1. What is the difference between *frequency tables* and *relative frequency tables*?

2. What type of data are *pie charts* and *bar graphs* used for?

3. Bar graphs represent each_____ as a bar and the *bar heights* give the category _____ or _____.

4. What makes a bad graph? What should you look for?

5. What is a ***two-way table****?*

- Fill in table for Example *"I'm Gonna Be Rich"* on page 12.
  a. What are the *Rows*? _____
  b. What are the *Columns*? _____

I'm Gonna Be Rich:

Opinion	Female	%	Male	%	Total	%
Almost no chance						
Some chance						
50-50 chance						
Good chance						
Almost certain						
Total						
%						

c. Later, you will add the *marginal distributions* to the table and label them.
d. Later, you will add the *conditional distributions* to the table and label them.

6. Define ***marginal distribution***:

✓ CHECK YOUR UNDERSTANDING pg14 (clearly show work and write answers in sentences)

CYU#1 ***Marginal distributions*** *:* describe how to calculate marginal distributions and add them to the table on prior page "***I'm Gonna Be Rich" example.***

- Also calculate the marginal distributions for Opinions and add to the table.

CYU#2 Create gender graph and describe it here:

7. Define ***conditional distribution***:

    - Describe how you decide which *conditional distribution to* compare (pg17, *Think About It: explanatory vs. response*).

✓ CHECK YOUR UNDERSTANDING pg17 (clearly show work and write answers in sentences)

CYU#1 ***Conditional distributions*** *:* describe how to calculate and add ALL *conditional distributions* to the "***I'm Gonna Be Rich"*** table**.**

CYU#2 How does the Figure 1.4 (pg 16) *conditional distributions* differ from the ones calculated in #1?

- Describe the *conditional distributions* presented in the graph in Figure 1.5 (pg 17):

- Describe the *conditional distributions* presented in the graph in Figure 1.6 (pg 17):

8. It is important to understand the difference between ***marginal distributions*** and ***conditional distributions***.
   a. _____ Distributions help us compare differences in groups in our sample. Explain in your words:

   b. _____ Distributions help us describe the overall composition of our sample. Explain in your words:

9. What is the purpose of using a *segmented bar graph* and *side-by-side bar graph*?

10. Explain the difference between a *segmented bar graph* and *side-by-side bar* graph (an easy way to do this is to sketch graphs of each and show the differences).

11. Explain what it meant by an **association** between two variables;

- Give an example of **association**. Use the "***I'm Gonna Be Rich" example*** to describe association between gender and opinions.

SKIP ~~SIMPSON'S PARODOX.~~

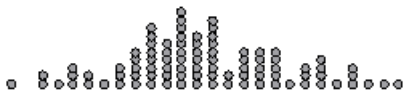
1.1 HW: Page 22 #'s 11, 15, 17**, 19**, 21, 25**, 27-31

These problems are the key concepts covered in this section. Your must do the **"**'d"** problems here. The other problems are for you to check your understanding and do not need to be handwritten.
Go to my web site to check your understanding and self-correct your work:
http://www.brunswick.k12.me.us/pgroves/home/ap-statistics/ap-statistics-summer-assignment/

Section: 1.2 - Displaying Quantitative Data with Graphs

1. Here is a sketch of a *dotplot*. What is the advantage of using this type of graph (discuss size of the data set and what the graph shows)?

   What does each **Dot** represent?
   What is missing from graph?



2. **[VERY IMPORTANT CONCEPT!!]**

- When examining a distribution, you must describe the overall pattern with these 4 components.

  S_____ O_____ C_____ S_____

  a) I call this CUSS and BS. (i.e. Center, Unusual, Shape, Spread and Be Specific). You can use either mnemonic.
  b) Make sure you understand how to **compare distributions in context. Tip: write 4 sentences.**

- When you compare 2 or more distributions, you must write a sentence for each of the above 4 components , comparing the different distributions.

3. Describe Shape Describe and sketch a graph for the following distributions:

Symmetric **(do NOT use the word NORMAL here!)**	S*kewed to the right (or positively skewed)*	S*kewed to the left (or negatively skewed)*
Unimodal (do NOT use NORMAL!)	*Bimodal (*Don't worry about little bumps)	*Uniform*

✓ CHECK YOUR UNDERSTANDING pg31 (clearly show work and write answers in sentences)

4. What is the advantage of using a *stemplot* (discuss size of the data set and what the graph shows)?

   a) Give an example of a **KEY**, which is required in a *stemplot* graph:

   b) When should you *split the stems* on a stemplot?

   c) When is it best to use a *back-to-back stemplot*?

✓ CHECK YOUR UNDERSTANDING pg34 (clearly show work and write answers in sentences)

CYU#1

CYU#2 Sketch the stem plot and use this graph to clearly explain your answers to the multiple choice questions:

5. When is a *histogram* a better choice of a graph than a *dotplot* or a *stemplot?*

6. Are *bar graphs* and *histogram* the same? **NO!**

   IMPORTANT Make sure you understand the differences!!!
   - *Bar graphs* display "categorical data" **and** the bars are NOT connected.
   - *Histograms* display "continuous numerical data", that is data that represents measured quantity, **and** the bars are connected to show the shape of the distribution

7. List the three steps involved in making a histogram.

8. When should you use a relative frequency histogram instead of a frequency histogram?

9. Do *Technology Corner (*page 38) problem.
   - State data is on page 35.
   - Make sure you understand how to (1) put data in lists, (2) change window settings to easily sketch your graph, and (3) graph histograms.

✓ CHECK YOUR UNDERSTANDING pg39 (clearly show work and write answers in sentences)

Use your calculator to sketch the histogram. For IQ scores, use: min=80; max=150, bar width=10.

CHECK YOUR UNDERSTANDING pg41 (clearly show work and write answers in sentences)

.1 HW: page 42 #'s 37, 45, 48, 49, 53**, 57, 69-74**

- These problems are the key concepts covered in this section. Your must do the **"**'d"** problems here. The other problems are for you to check your understanding and do not need to be handwritten.
- Go to my web site to check your understanding and self-correct your work:
  - **http://www.brunswick.k12.me.us/pgroves/home/ap-statistics/ap-statistics-summer-assignment/**

Section: 1.3 - Describing Quantitative Data with Numbers

Measuring Center

1. What is the meaning of Σ **(sigma)**?

2. For *mean,* $\bar{x}$ (Xbar):
   - Give the formula and explain how to use it. *Note you will not need to memorize the formula but need to understand how to use it.*

   - explain where to find the mean on the calculator

3. Explain the difference between $\bar{x}$ and μ (mu). ***Answer: IMPORTANT DEFINITIONS!!!***
   $\bar{x}$ is the mean for a sample.
   μ is the mean for the population.
4. Define ***resistant*** measure:

5. Explain why the mean is ***not a resistant*** measure of center.

6. What is the *median **(M)*** of a distribution
   - Explain how to calculate median by hand , when there is an **odd number** of data values

   - Explain how to calculate median by hand , when there is an **even number** of data values

   - Explain where to find the median on the calculator

7. Explain why the median is a **resistant measure** of center?

Measuring Center (continued)

8. How does the shape of the distribution affect the mean and median? Sketch graphs and describe the location of the mean and median.
   - Shape is symmetric

   - Shape is skewed right

   - Shape is skewed left

CHECK YOUR UNDERSTANDING pg55 (clearly show work and write answers in sentences)

Measuring Spread

9. What is the *range*? ***Answer:***

- ***Range* is *the maximum-minimum value. Range is a single number!***
- ***For example**: if the maximum age was 50 and the minimum age was 30 then the **Range=20 years**.*

10. Is the range a resistant measure of spread? Explain. ***Answer:***

- ***Range* is *is NOT a resistant measure because it is influenced by outliers.***
- ***For example**: if we surveyed adults and looked at their ages and most of the ages were from 30 to 50 but **one** respondent was 70 years old. This outlier would change the range dramatically from 20 years to 40 years.*

11. Quartiles:
    - How do you find the first quartile Q1 by hand?

    - How do you find the third quartile Q3 by hand?

    - Explain where to find the quartiles on the calculator

12. What is the *Interquartile Range* (IQR)? **IMPORTANT: IQR it is a single number!**

13. Is the IQR a resistant measure of spread? Explain.

Identifying Outliers

14. How is the IQR used to identify *outliers*?

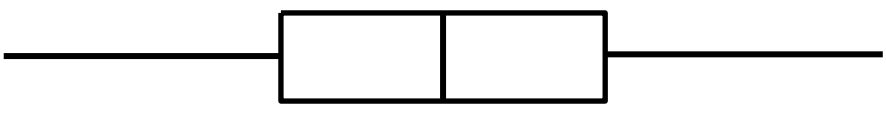
- Large *outliers*

- Small *outliers*

5-number summary

15. What is the *five-number summary* of a distribution?

16. Use the graph below to explain how to use the five-number summary to make a *boxplot*.



17. How do you identify outliers in a *boxplot*?

5-number summary (continued)

CHECK YOUR UNDERSTANDING pg61 (clearly show work and write answers in sentences)

18. Do *Technology Corner (*page 61) problem.
    - ●Data for NC is on page 56. Data for NY is on page 57.
    - ●Make sure you understand how to put
        (1) data in lists,
        (2) graph box plots with outliers identified,
        (3) graph side-by-side box plots, and
        (4) use **TRACE** to find the 5-number summary in a boxplot.

Measure Spread - Variance

19. *Variance ($s_x{}^2$ or s^2)*
    - What does the *variance ($s_x{}^2$ or s^2)* measure? ***Answer: Variance is the average squared distance.***
    - What are the units of measure for *variance ($s_x{}^2$ or s^2)*? ***Answer: Variance is measured in squared units. For example, squared feet.***
    - Give the formula for *variance*. *Note you will not need to memorize the formula but need to understand how to use it.*

20. Explain where to find the variance on the calculator.

Measure Spread – The Standard Deviation

21. *Standard deviation (s or s_x)*:

- What does the *standard deviation (s or s_x)* measure?

- Give the formula for *standard deviation*. *Note you will not need to memorize the formula but need to understand how to use it.*

- Explain, in English, how to calculate the *standard deviation.* The 3 Steps are outlined on page 64.

- If you know the *variance*, how do you find the *standard deviation*?

- Explain where to find the *standard deviation* on the calculator.

22. Why do we prefer to use *standard deviation* and NOT *variance*? ***Answer:***

- ***Variance is measured in squared units which are meaningless units to most people;***
- ***Where, standard deviation and mean are using the same units, making them easier to interpret in a problem.***

23. Explain the difference between $\boldsymbol{s_x}$ and σ (sigma). ***Answer: IMPORTANT DEFINITIONS!!!***

s_x is the standard deviation for a sample.

σ is the standard deviation for a population.

Measure Spread – The Standard Deviation *(continued)*

CHECK YOUR UNDERSTANDING pg64 (clearly show work and write answers in sentences)

24. Do *Technology Corner (*page 65) problem.
    - Data for NC is on page 56. Data for NY is on page 57.
    - Make sure you understand how to put
        (1) data in lists,
        (2) find one-variable statistics, and
        (3) read computer output given one-variable statistics.

25. How should one go about choosing measures of center and spread? ***Answer:***
    - ***If the distribution is symmetric, use mean and standard deviation.***
    - ***If the distribution is skewed, use median and IQR because these are resistant measures and not influenced by outliers.***

1.3 HW: page 70 #'s 79, 81, 83, 87, 89, 91**, 93, 97, 105**, 107-110**

These problems are the key concepts covered in this section. Your must do the **"**'d"** problems here. The other problems are for you to check your understanding and do not need to be handwritten.
Go to my web site to check your understanding and self-correct your work:
http://www.brunswick.k12.me.us/pgroves/home/ap-statistics/ap-statistics-summer-assignment/