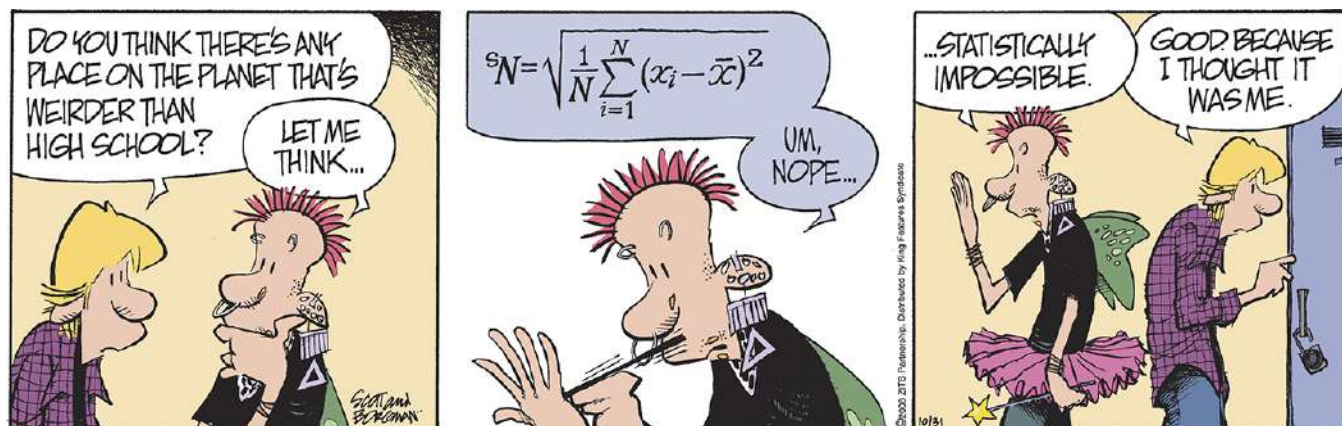# Chapter 1: Exploring Data

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 <u>ACTUALLY DO</u> 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!**
- **"Check Your Understanding" problems are optional (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)
- outlier
- shape
- symmetric distribution
- Skewed to the right
- Skewed to the left
- Unimodal distribution
- Bimodal distribution
- Uniform distribution

- resistant
- outlier
- Σ
- center
- $\overline{x}$
- μ
- median
- spread, variability
- quartiles
- Q_1, , M, Q_3

- IQR
- five-number summary
- minimum
- maximum
- range
- boxplot
- standard deviation
- variance

Introduction   **Data Analysis: Making Sense of Data**

1. Individuals are…

2. A *variable (IT IS NOT X)* is …

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

        Give examples of categorical variables: _____

        Give examples of quantitative variables: _____

4. Define *distribution*:

5. Explain what *inference* is*:*

        In your own words, explain the example discussed on page 5:

1.1   Analyzing Categorical Data

1. What is the difference between a *data table, frequency* table, and *relative frequency table*?

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 is a *two-way table (tip: sketch an example and discuss rows and columns)*?

5. Define *marginal distribution*:

6. Define *conditional distribution*:

   - Describe how you decide which *conditional distribution to* compare (pg17, *think about it: explanatory vs. response*).

7. Use the following two-*way table*,

	18–34	35–54	55+	Total
iPhone	169	171	127	467
Android	214	189	100	503
Other	134	277	643	1054
Total	517	637	870	2024

a) To explain the difference between a *marginal distribution* and a *conditional distribution*.

b) To give an example demonstrating how to calculate a *marginal distribution*:

c) To give an example demonstrating how to calculate a *conditional distribution*:

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

9. 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).

10. Explain what it meant by an *association* between two variables; Give an example.

SKIP SIMPSON'S PARODOX.

1.2   Analyzing Categorical Data

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.   AND if you compare 2or more distributions, you must

    write a sentence for each of the following 4 components. Each sentence iscusses a component and

    compared the distributuions.

    S_____   O_____   C_____   S_____

    - **I call this CUSS and BS. You can use either mnemonic.**
    - **Center, Unusual, Shape, Spread and Be Specific,**
    - **Make sure you understand how to compare distributions in context.**

3. <u>Describe Shape</u>
   a) Describe and illustrate 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)*

Describe Shape (continued)

b) Describe and illustrate the following distributions:

Unimodal (do NOT use NORMAL!)	Bimodal (Don't worry about little bumps)	Uniform

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

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

a) When is it beneficial to *split the stems* on a stemplot?

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

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. Why is it advantageous to use a relative frequency histogram instead of a frequency histogram?

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

1.3   Analyzing Categorical Data

<u>Describe Center</u>

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

2. For *mean,* $\bar{\boldsymbol{X}}$ (Xbar):
   - give the formula and explain how to use it

   - explain where to find the mean on the calculator

3. Explain the difference between $\bar{\boldsymbol{X}}$ and $\boldsymbol{\mu}$ (mu).

4. Define <u>*resistant*</u> measure:

5. Explain why the mean is <u>**not a resistant**</u> measure of center.

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

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

   - Explain where to find the median on the calculator

7. Explain why the median is a <u>**resistant measure**</u> of center?

Describe Center (continued)

8. How does the shape of the distribution affect the mean and median?

- Shape is symmetric

- Shape is skewed right

- Shape is skewed left

Describe Spread and Outliers

9. What is the *range*? **IMPORTANT:** *range* **is a single number!**

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

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

\

Describe Spread and Outliers (continued)

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

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

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

    • Large *outliers*

    • Small *outliers*

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

16. Explain how to use the five-number summary to make a *boxplot*. How do you identify outliers?

17. 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) change window settings, (3) graph box plots <u>with outliers identified;</u> and (4) graph side-by side box plots.

Describe Spread and Outliers (continued)

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

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

    - give the formula and explain how to use it

    - explain where to find the standard deviation on the calculator

19. *Variance (s_x^2 or s^2)*

    - What does the *variance (s_x^2 or s^2)* measure?

    - give the formula for variance

    - explain where to find the variance on the calculator

    - Why do we prefer to use standard deviation and NOT variance?

20. What are the *4 properties* of the standard deviation explained on page 64?

21. Do *Technology Corner (*page 65) problem to find summary statistics. 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.

22. How should one go about choosing measures of center and spread?