MAT 135, Discussion Questions 2.05

1. The formula for the mean is

$$\bar{x} = \frac{\sum x_i}{n}$$

Describe what this formula means in your own words.

add up the values in data set then divide by the # of things in set

2. How does the median differ from the mean? When are the two values going to be the same (or nearly so)?

median has 50% of data above/below it
not affected by outliers

they will be the same if graph is symmetric

3. What is the mode? Can there be more than one?

value that appears most often

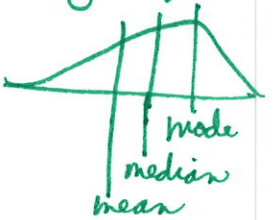
yes, if there are multiple values that appear the most

4. When should each measure of center be used as a "typical" value?

Symmetric - use mean
very skewed - use median
mode best for small data sets or when above situations are multimodal

5. How do the mean, median and mode relate to each other in a 1) symmetric distribution? 2) skewed left distribution? 3) skewed right distribution?

1) symmetric - all roughly the same
2) skewed left



3) skewed right

6. Consider the data set for presidential ages at inauguration we saw earlier. I've dropped the decimals (for number of days), so that we can look at their ages in a more typical fashion.

President	AGE (in years)	President	AGE (in years)
George Washington	57	Benjamin Harrison	55
John Adams	61	Grover Cleveland	55
Thomas Jefferson	57	William McKinley	54
James Madison	58	Theodore Roosevelt	42
James Monroe	58	William Howard Taft	51
John Quincy Adams	57	Woodrow Wilson	56
Andrew Jackson	61	Warren G. Harding	55
Martin Van Buren	54	Calvin Coolidge	51
William Henry Harrison	68	Herbert Hoover	54
John Tyler	51	Franklin D. Roosevelt	51
James K. Polk	49	Harry S. Truman	60
Zachary Taylor	64	Dwight D. Eisenhower	62
Millard Fillmore	50	John F. Kennedy	43
Franklin Pierce	48	Lyndon B. Johnson	55
James Buchanan	65	Richard Nixon	56
Abraham Lincoln	52	Gerald Ford	61
Andrew Johnson	56	Jimmy Carter	52
Ulysses S. Grant	46	Ronald Reagan	69
Rutherford B. Hayes	54	George H. W. Bush	64
James A. Garfield	49	Bill Clinton	46
Chester A. Arthur	51	George W. Bush	54
Grover Cleveland	47	Barack Obama	47

Find the mean and the median (in your calculator if you can). How can you use the calculator to help you find the mode?

mean $\bar{x} = 54.68$

median $\tilde{x} = 54.5$

mode:

Sort data to easily see which is most common

51 & 54 both occur 5 times

7. How do we calculate the range of a data set?

Max - Min

8. Describe the procedure for calculating a standard deviation by hand (i.e. $\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}}$ or $s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$, unpack this formula. What is it telling you to do, in order?).

(1) find mean.
(2) subtract mean from each data point
(3) square each result.
(4) add up these results
(5) divide by N or n-1 (population vs. sample — usually n-1)
(6) take square root

9. What is another formula we can use to find the standard deviation?

$E(X^2) - [E(X)]^2$

or $\frac{\sum X^2 - (\frac{1}{n}\sum X)^2}{n-1}$

10. How do you find these values in your calculator?

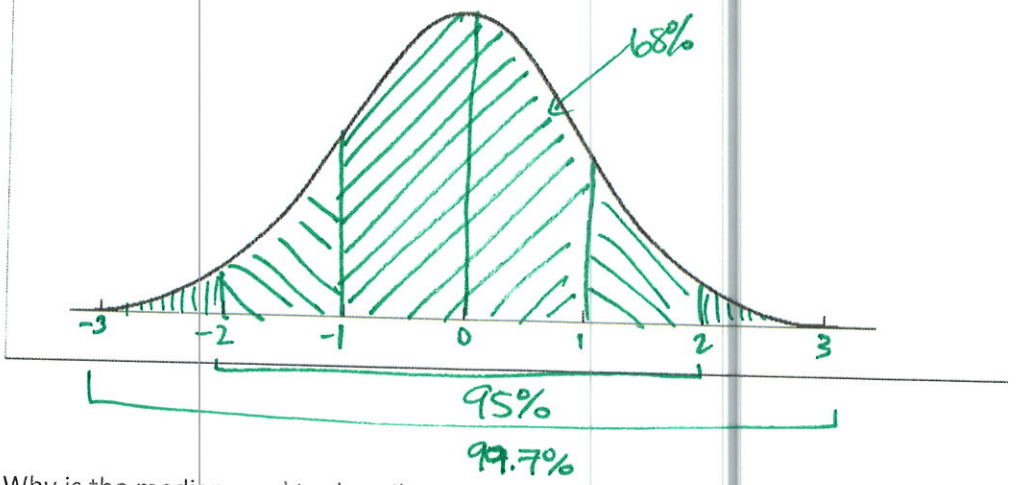
1-Var Stats

(handout available)

11. One interpretation of the standard deviation is that it is the average distance from the mean. What does this measure in statistical terms?

Spread of distribution

12. Sketch the Empirical Rule on the graph below.



13. Why is the median used to describe income and not the mean?

income is very skewed

14. Visit the graphic http://www.statslife.org.uk/images/pdf/timeline-of-statistics.pdf and choose a point in the history of statistics that seems really important to you. Be prepared to explain why.

answers will vary