Stat 1450, Exam #1, Fall 2013 Name KEY

Instructions: Show work. Provide complete explanations. Round calculations according to the procedures outlined in class.

1. Describe the difference between a **parameter** and a **statistic**. (8 points)

a parameter is a feature of a population, while a statistic is a feature of a sample

2. How is with a **nominal** level of measurement different from data with an **ordinal** level of measurement? (8 points)

a nominal level of measurement categorizes data but has no natural ranking, while ordinal level of measurement is naturally ranked (and is typically numerical)

3. How is a **stratified** sampling different from a **cluster** sampling? (8 points)

a stratified sampling breaks population into groups and chooses random individuals from within each group, whereas a cluster sampling divides population into groups and then chooses random groups.

Stratified sampling also divides populations into unlike categories, while cluster sampling divides population into like clusters.

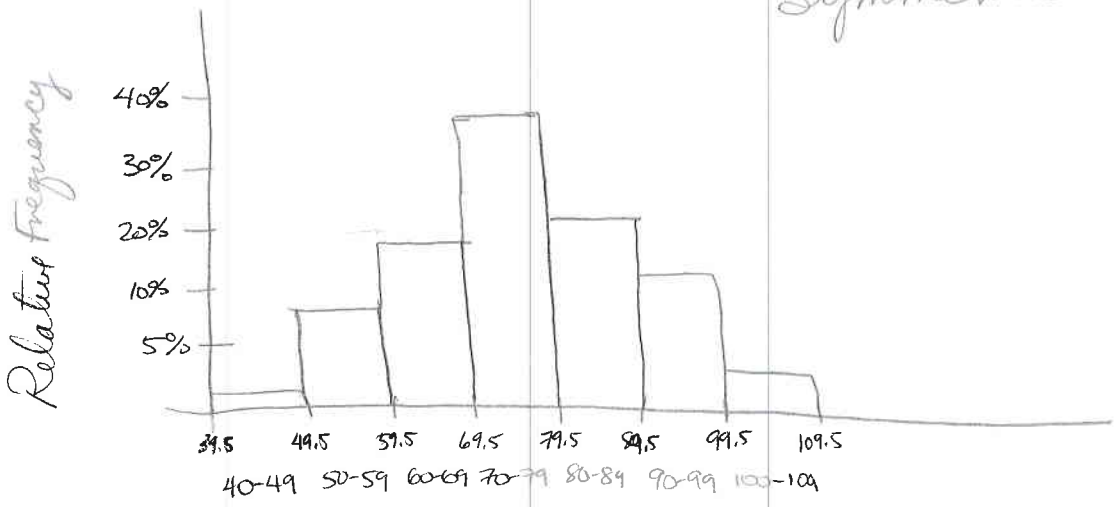
4. Describe a situation in which you might prefer a **retrospective** study to a **prospective** one. (10 points)

a retrospective study looks into the past using records or surveys, whereas a prospective study chooses a group to study and then tracks them forward in time.

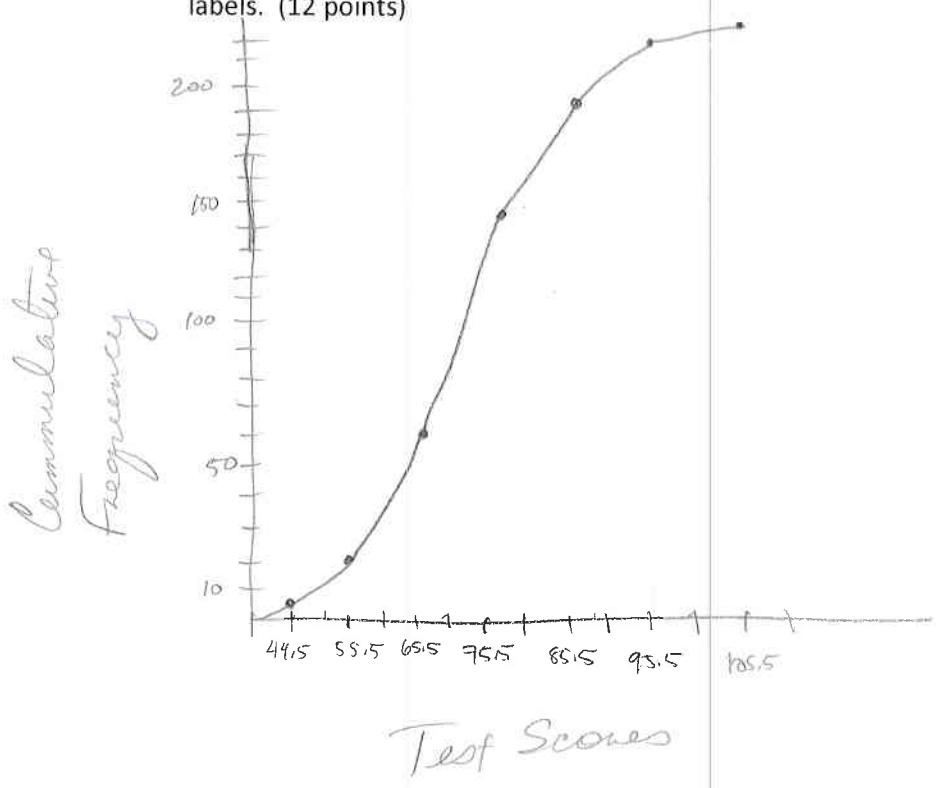
5. In a study of 223 students in a high school math program, the following frequency table was created from test scores on a particular math test. Use this date to construct a relative frequency table and the corresponding histogram. Be sure to label your axes correctly. From the graph you draw, does the graph look approximately symmetric? Or does it appear to be skewed (and in which direction)? (28 points)

Class	Frequency	Relative Freq.
40-49	3	1.3%
50-59	18	8%
60-69	41	18.3%
70-79	87	39%
80-89	45	20.2%
90-99	23	10.3%
100-~~110~~ 109	6	2.7%

Symmetric

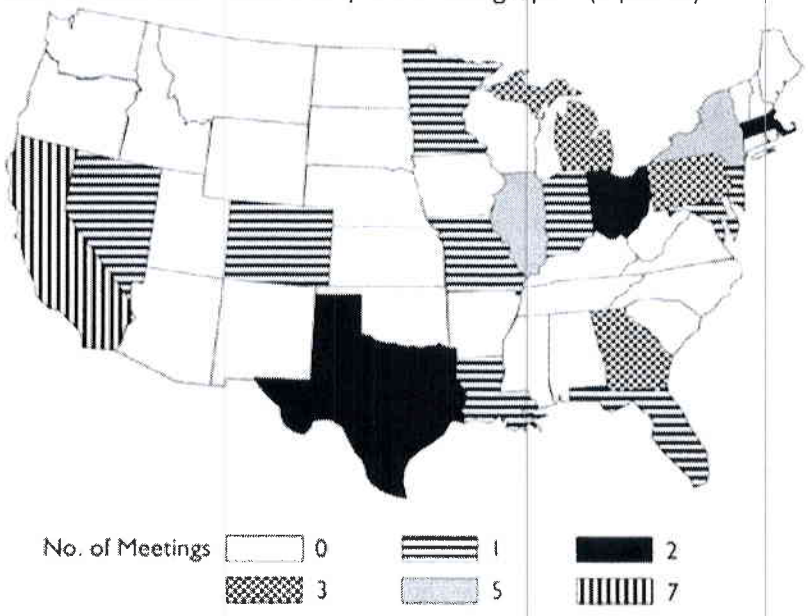


6. Use the data in Problem #5 to construct an ogive graph of the data. Be sure to use correct axis labels. (12 points)



Class	Cum. Freq.
40-49	3
50-59	21
60-69	62
70-79	149
80-89	194
90-99	217
100-109	223

7. I googled for "bad statistical graphs" and among the graphs that came up was the one below. Give at least two reasons why it's a "bad graph". (8 points)



- Has no header to say what graph is about
- Cross-lakes create optical illusions that make looking at graph difficult

8. The data from Problem #5 is reproduced below. Calculate the mean of the data set. (12 points)

Class	Frequency
40-49	3
50-59	18
60-69	41
70-79	87
80-89	45
90-99	23
100-~~110~~109	6

$$\bar{x} = \frac{(45.5)(3) + (55.5)(18) + (65.5)(41) + (75.5)(87) + (85.5)(45) + (95.5)(23) + 105.5(6)}{223}$$

$$= 76.5$$

9. Explain the difference between the **median**, the **mean**, and the **mode**. (12 points)

the median is the "middle value" in an ordered data set

the mean is the arithmetic mean, the sum of all values divided by # of values

the mode is the most frequent data value

10. If the lowest score in the math class (from problems #5, 6, 8) was 42, and the highest was 106, approximate the standard deviation of the data set. (6 points)

$$s \approx \frac{106-42}{4} = 16$$

11. According to the Empirical rule, how much of the data must be within 3 standard deviations of the mean in a normal distribution? How much does Chebyshev's Theorem guarantee for any distribution? (12 points)

Empirical Rule $\Rightarrow$ 99.7%

Chebyshev's $\Rightarrow$ $1 - 1/3^2 = \frac{8}{9} \approx 89\%$

12. The SAT has a mean of 1518 and a standard deviation of 325. Your friend scored a 2160. Calculate the z-score for this data point. Is this an unusual value? (8 points)

$z = \frac{2160 - 1518}{325} = 1.98$

no, its not unusual

13. Suppose that a student scored exactly 80 on the math test in problem #5. What percentile is that student at? (9 points)

According to cumulative frequency table from #6
149 scores are below 80 so

percentile $= \frac{149}{223} \times 100 = 66.8 \Rightarrow$ 67th percentile

14. Using the data below for FICO scores, give the five-number summary of the data. Then use that data to draw a box-and-whisper plot. (18 points)

664, 693, 698, 714, 751, 753, 779, 789, 802, 818, 834, 836

min = 664

Q_1 = 698 (706 using median method)

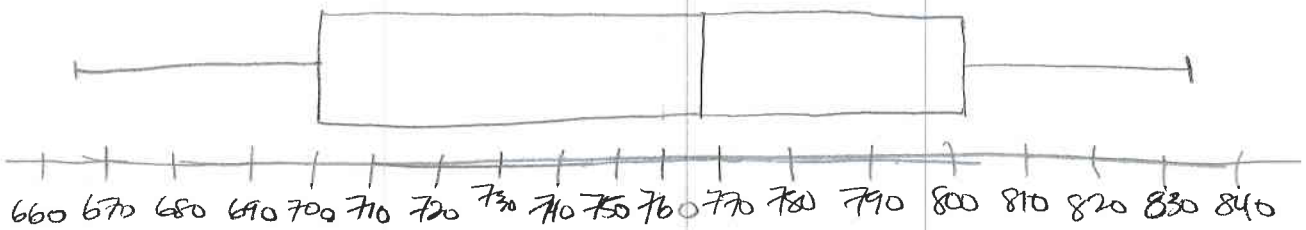
$L = \frac{25}{100} \times 12 = 3$

$L = \frac{75}{100} \times 12 = 9$

median = 766

Q_3 = 802 (810 using median method)

max = 836



15. If we wanted to draw a modified boxplot, are there any outliers that would make the graph look different from the graph in #14? (8 points)

$IQR = 802 - 698 = 104$

$1.5 \times IQR = 156$

$Q_1 - 156 = 542$

$Q_2 + 156 = 958$

$IQR = 810 - 706 = 104$

$Q_1 - 156 = 550$

$Q_3 + 156 = 966$

There are no outliers
The modified box plot would look the same as in #14.