Math 247: Test 1 **Name:** KEY

Test _____/90 points Review _____/ 10 points
(Each problem is worth 15 points)

15 1. Medicaid Explansion. Medicaid (a health program for low-income residents) was expanded in 3 states in the year 2000. A sample of of adults from these states was compared a sample of adults from 6 neighboring states where there was not a Medicaid expansion; these adults were observed 5 years before and 5 years after the Medicaid expansion. The primary outcome studied was all-cause mortality rates. Medicaid expansions were associated with a significant reduction of about 2% in all-cause mortality.

3 (a) Is this an observational study or a controlled experiment? (Obervational) Controlled Experiment

3 (b) What was the treatment variable? Medicaid expansion

3 (c) What was the response variable? All-cause mortality rates.

2 (d) What was the comparison group? 6 neighboring states, no expansion

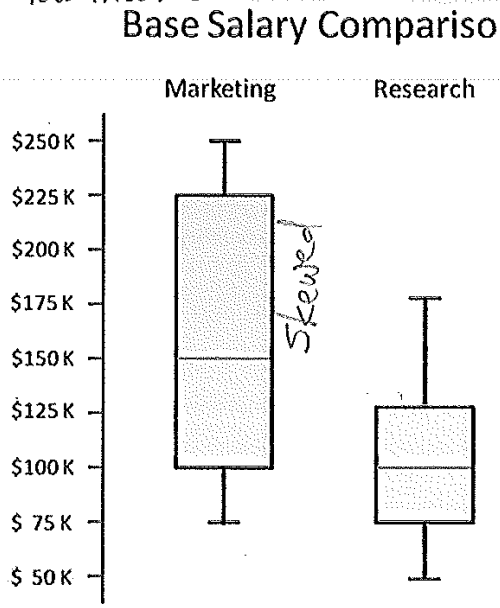
2 (e) Can we conclude that Medicaid expansion CAUSED the reduction in mortality? Explain.

Yes, no question No, this was not a controlled experiment – the States were not randomly assigned to the treatment groups of "Expansion", "No Expansion"

2 (f) Describe a possible confounder for this study.

The States with exparded Medicare may have also increased health education, improved food programs, etc. Definition of "low income" varies state-to-state.

15 2. The salaries for 27 employees working in marketing and 35 employees working in research were gathered from a large corporation. The boxplot illustrates the distribution of salaries in each group.



3 (a) Is employee-type a categorical or numerical variable?

3 (b) Is salary a categorical or numerical variable?

3 (c) Which group's data had the most variability?
Marketing

3 (d) Estimate the "typical" salary for the each group:
marketing employees: $150K
research employees: $100K

3 (e) Which distribution of salaries shows the most skewing? Explain or indicate on the graph how you can tell.
Marketing shows the most skewing - asymmetrical box.

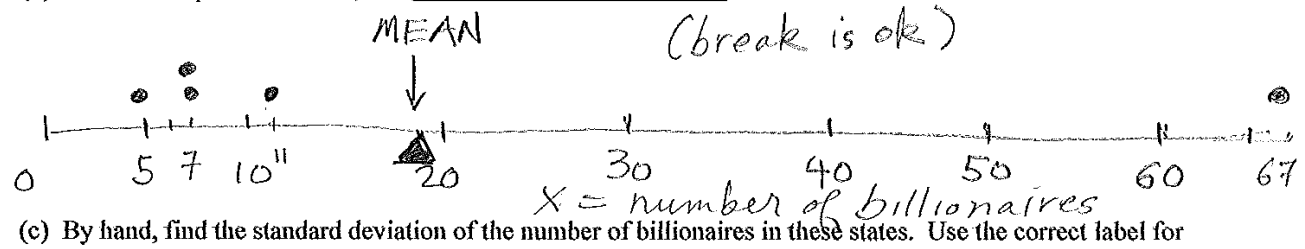
3. According to Forbes.com, the numbers of billionaires in the five states in the Northeast with the most billionaires are given in the table:

New York	67
Connecticut	11
Pennsylvania	7
Massachussetts	7
New Jersey	5

1 (a) Find the mean number of billionaires. Use the correct label for this value.

$\bar{x} = \frac{67+11+7+7+5}{5} = \frac{97}{5} = 19.4$ billionaires

2 (b) Sketch a dotplot of the data, and mark the location of the mean. Scale the axis so all the data fits!



6 (c) By hand, find the standard deviation of the number of billionaires in these states. Use the correct label for this value.

x	$x-\bar{x}$	$(x-\bar{x})^2$ (always positive!)
5	$5-19.4=-14.4$	$(-14.4)^2=207.36$
7	$7-19.4=-12.4$	$(-12.4)^2=153.76$
7	$7-19.4=-12.4$	$(-12.4)^2=153.76$
11	$11-19.4=-8.4$	$(-8.4)^2=70.56$
67	$67-19.4=47.6$	$(47.6)^2=2265.76$
		2851.2

$s = 26.7$ bill.

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{2851.2}{5-1}} =$ 26.7 billionaires

2 (d) The data value from which state appears to be an outlier? (No calculation necessary!)

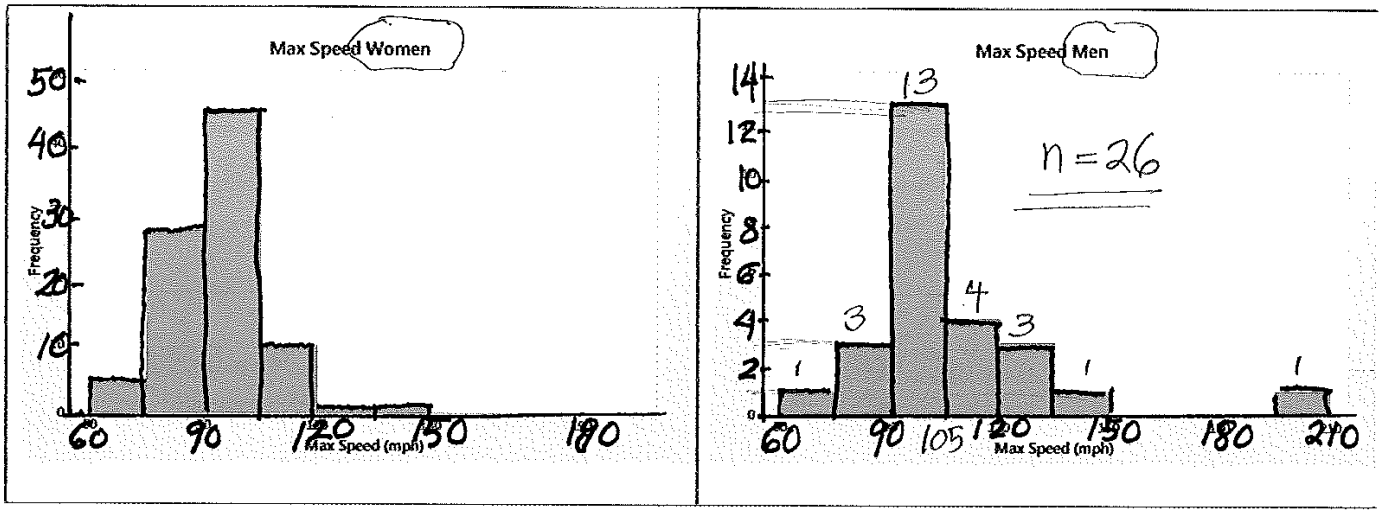
New York.

2 (e) Describe the effect the outlier had on the mean and on the standard deviation. Include the words "typical value" and "spread" in your description.

The mean has been increased so much by the outlier that it is NOT a typical value anymore. The standard deviation also has been made too large and exaggerates the spread of the data.

2 (f) Because the outlier affected the value of the mean and standard deviation so much, we would say that neither the mean nor the standard deviation is RESISTANT.

/15 4. The histograms shown below give the maximum speed ever driven by a sample of college men and a sample of college women.



3 (a) Which sample was larger, the sample of men or the sample of women? Sample of women

3 (b) What is the frequency of men who have driven a maximum speed between 90 and ~~115~~ 105 miles per hour? 13 count both bars -2

3 (c) What is the relative frequency of this group? 13/26 = .50

3 (d) Which group's data, men or women, had the most variability? Men

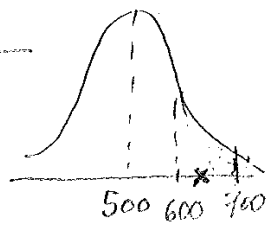
3 (e) Did the typical woman in this sample drive faster, slower, or about the same as the typical man? About the same.
accept slower as answer

/15 5. Quantitative SAT scores have a mean of 500 and a standard deviation of 100, while ACT scores have a mean of 21 and a standard deviation of 5. Assuming both types of scores have distriubitons that are unimodal and symmetric, which is more unusual: a quantitative SAT score of 650 or and ACT score of 33?

Show the calculation of the **z-scores** for the SAT and the ACT data, and include the interpretation of the z-score values in your answer. Also include the Empirical Rule in your answer.

SAT

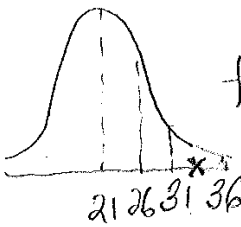
$\bar{x} = 500$
$s = 100$
$x = 650$



$Z = \frac{650-500}{100} = 1.5$

ACT

$\bar{x} = 21$
$s = 5$
$x = 33$



$Z = \frac{33-21}{5} = 2.4$

The ACT score is 2.4 S.D.'s away from the mean which the Empirical Rule tells us is more unusual than the SAT score which is only 1.5 S.D.'s from the mean.*

15 6. The table below shows the number of prisoners on death row in 2013 for the western states in the U.S.

	State	Prisoners
1	Alaska	0
2	Hawaii	0
3	Wyoming	1
4	Montana	2
5	New Mexico	2
6	Colorado	4
7	Washington	8
8	Utah	9
9	Idaho	13
10	Oregon	37
11	Nevada	79
12	Arizona	127
13	California	727

2 (a) Find the median of the data. Label!

Med = 8 prisoners (pr.)

2 (b) Find the first and third quartiles. Label!

$Q_1 = 2$ pr. $Q_3 = 37$ pr.

($Q_1 = 1.5$ $Q_3 = 58$) Book

2 (c) Find the IQR. Label!

IQR = 35 pr.

(IQR = 56.5) Book

2 (d) Find the Right Outlier Limit and the Lower Outlier Limit. Is Arizona an outlier? Is California an outlier?

Left Limit $= Q_1 - 1.5\,IQR$
$= 2 - 1.5(35)$
$= -50.5$ pr.

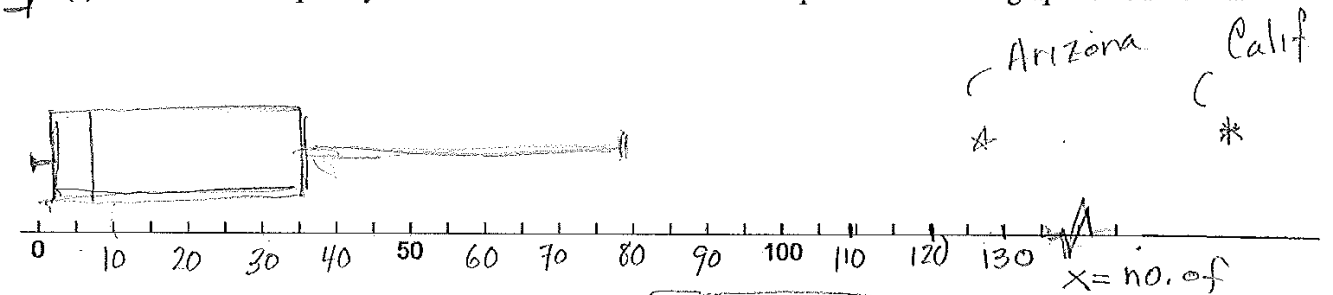
(Left = −83.25 Book)

Right Limit $= Q_3 + 1.5\,IQR$
$= 37 + 1.5(35)$
$= 89.5$ pr.

Arizona + Cali outliers

(Right = 142.75 Book)
only Cali outlier

4 (e) Construct a boxplot by hand for the data. You will have to put a break in the graph for California.



1 (f) This distribution is SKEWED LEFT (SKEWED RIGHT) SYMMETRIC

2 (g) The mean for this data set is 77.6 prisoners and the standard deviation is 198.8 prisoners. Would it be better to use the mean and standard deviation or the median and IQR to summarize the data? Explain your choice.

Use the median = 8 and the IQR = 35 to summarize. Since these values better represent the "typical" value and the spread. The mean and SD would be influenced by the outliers