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

In-class test _____/80 points *Take home test* _____/20 points

1. (12 points) A study investigated the relationship between positive and negative coaching and performance on a memory test. The researchers took a random sample of 60 college students then randomly placed them into three groups of 20 students each. One group listened to a positive coach ("You can do this!"), one group listened to a negative coach ("You can't do this"), and the control group had no coaching. Each of the subjects in the groups then took a brief test on memory that had a score from 0 to 10.

This study is (circle one) A RANDOMIZED EXPERIMENT AN OBSERVATIONAL STUDY

The variables in this study are "Type of Coaching" and "Test Score"

What kind of variable is "Type of Coaching"? CATEGORICAL NUMERICAL

What kind of variable is "Test Score"? CATEGORICAL NUMERICAL

Which of these variables is the Treatment (Factor)? Type of Coaching

Which of these variable is the Outcome (Response)? Memory Test Score

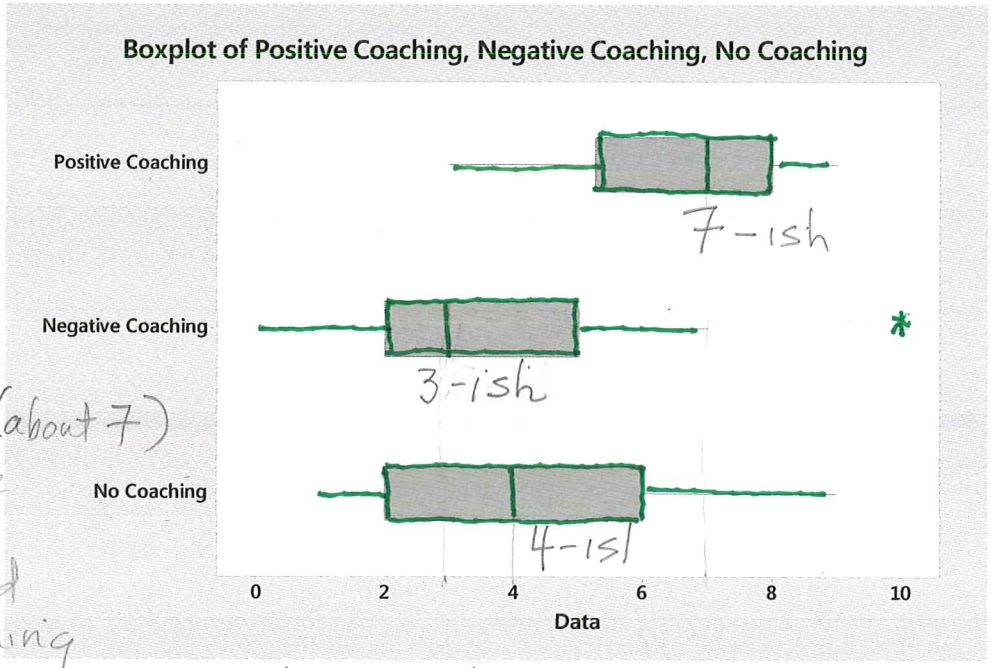
2. (9 points) The boxplots below show the distribution of the memory test scores for the study described in problem 1, with the 3 groups being "Positive Coaching", "Negative Coaching", and "No Coaching".

Which group's memory test scores showed the most variability? No Coaching

Which group's memory test scores were left skewed? Positive Coaching

Which group had an outlier and what is the value of the outlier? Negative Coaching, 10

Did the type of coaching seem to have an impact on memory according to these results? Include the "typical value" of each group in your explanation.



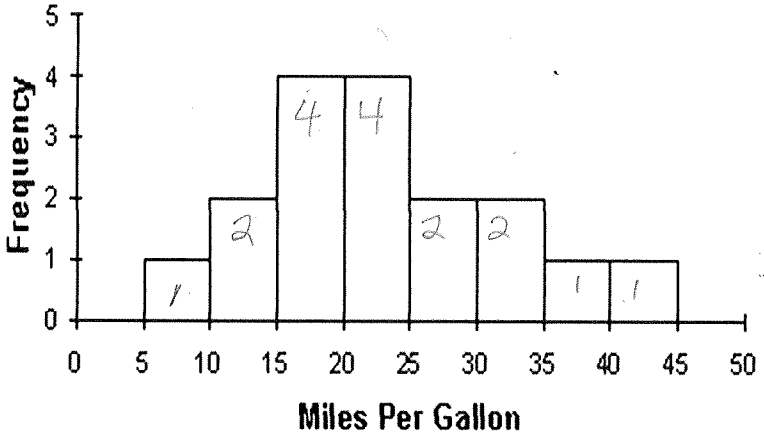
Yes, type of coaching appears to have had an impact. The typical value for the positive group (about 7) was higher than the others (3 for the negative group and 4 for the no coaching group.) Since this is a randomized experiment we can conclude the coaching caused the differences.

3. (3 pts) Why is the median of a data set considered "resistant" and the mean considered "not resistant"?

The median is not pulled away from the center of the data by extreme values while the mean is pulled away from the center by skewing or outliers.

4. (10 pts) The distribution of gas mileage (mpg) for the top selling cars in 2015 are shown below.

Use the histogram to answer the following questions.



(a) How many cars were in this study?

17 cars

(b) How many cars had gas mileage under 20 mpg? 7 cars

(c) What is the relative frequency (express as a percent) of the cars that had a gas mileage of 35 mpg or better? 2/17 =

(d) Is the data skewed left, skewed right, or approximately symmetric?

skewed right

(e) Which of the following could be the median and mean of the data? (circle the best answer)

(i) Median =15, Mean = 20

(ii) Median = 22.5, Mean = 25

(iii) Median = 25, Mean = 22.5

(iv) Median = 25, Mean = 30

5. (4 pts) A group of 10 male and female students were asked their weekly work hours. Unstack the data.

Stacked

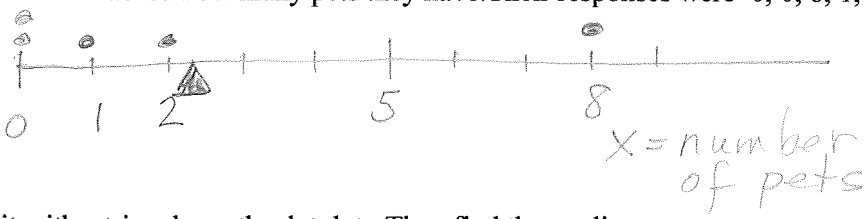
Male	16
Female	15
Female	25
Female	0
Male	10
Male	30
Female	10
Male	25
Female	0
Female	20

Unstacked

Male	Female
16	15
10	25
30	0
25	10
	0
	20

6. (20 pts) A random sample of 5 students were asked how many pets they have. Their responses were 0, 0, 8, 1, 2

(a) Construct a dot plot of this data.



(b) Find the mean of the data and mark it with a triangle on the dotplot. Then find the median.

$\bar{x} = \frac{11}{5} = 2.2$ pets

0 0 1 2 8

↑

median = 1 pet

(c) Which is a more "typical value" for this data set, the mean or the median? MEAN MEDIAN

(d) By hand, find the standard deviation of the data. $n = 5$

x	$x - \bar{x}$	$(x - \bar{x})^2$
0	−2.2	4.84
0	−2.2	4.84
1	−1.2	1.44
2	−.2	.04
8	5.8	33.64

$\sum (x - \bar{x})^2 = 44.80$

✓ $\sum x - \bar{x} = 0$

$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$

$= \sqrt{\frac{44.8}{4}}$

$= \sqrt{11.2}$

$s = 3.3466$ pets

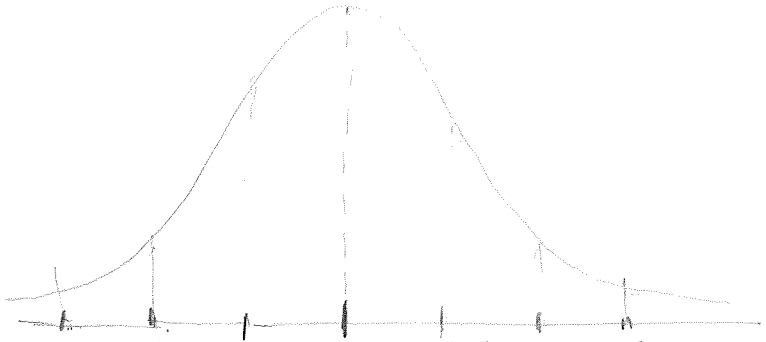
(e) Is the standard deviation "resistant"? Explain.

No. The extreme value of 8 inflated the value of the standard deviation. Most of the data is clustered around 0 to 2

7. (10 pts) The average head circumference of 1 week old female infants is 35 centimeters with a standard deviation is 1.5 centimeters. Assume head circcumferences are symmetrically distributed.

(a) Sketch a curve, with the x-axis labeled appropriately, showing the distribution of head circumferences.

$\bar{x} = 35$ cm
$s = 1.5$ cm



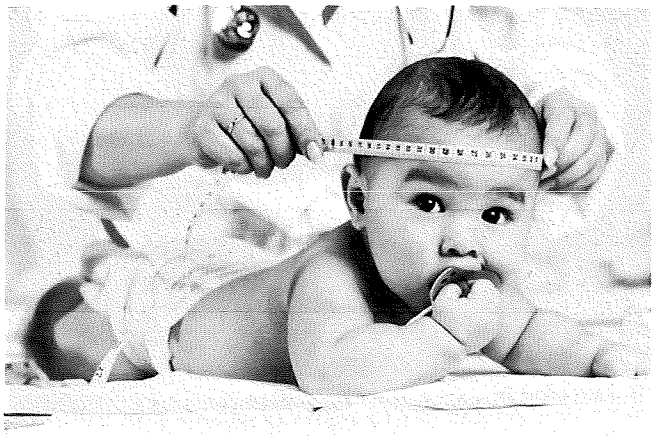
(b) Between what two values should about 68% of the head circumferences fall?

Between 33.5 cm and 36.5 cm.
($\bar{x} \pm 1s$, or within 1s of the mean, $\bar{x}$)

(c) Find the z-score for the baby girl with a head circumference of 31.5 cm.

$$z = \frac{x - \bar{x}}{s} = \frac{31.5 - 35}{1.5}$$

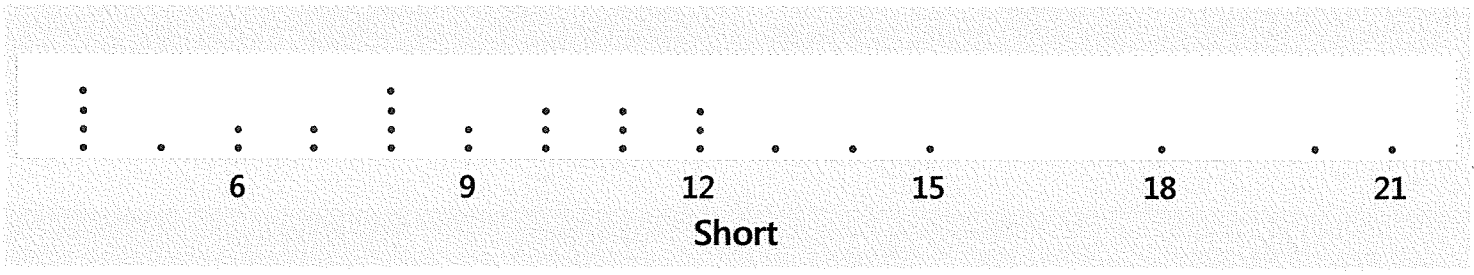
$$= \frac{-3.5}{1.5} = -2.333$$



(d) "Microcephaly" is a defined as a baby having a head circumference of more than 2 standard deviations from the mean. Does the baby from part (c) have microcephaly? Include the meaning of the z-score you found in part (c) in your answer.

Yes, this baby's head circumference is 2.333 standard deviations from the mean, so over 2 standard deviations.

8. (12 points) A student asked 30 students who ride a short board how many days per month they surf. The data is displayed in a dotplot below, with the descriptive statistics for the group shown below the dotplot.



Descriptive Statistics: Short

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Short	30	0	9.900	0.821	4.498	4.000	6.750	9.500	12.000	21.000

What is the five number summary for this data set? 4, 6.75, 9.5, 12, 21

Find the IQR. $IQR = Q_3 - Q_1 = 12 - 6.75 = 5.25$

IQR = 5.25 days

Find the Lower Outlier and Upper Outlier Limits.

$Lower = Q_1 - 1.5IQR$
$= 6.75 - 1.5(5.25)$
$= -1.125$

$Upper = Q_3 + 1.5IQR$
$= 12 + 1.5(5.25)$
$= 19.875$

How many outliers are there in this data set and what are they?

There are 2 outliers:
20 days and 21 days of surfing per month