Mention rounding

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

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

1. Most breast cancer patients are over the age of 50 at diagnosis. A researcher at a particular New York cancer center believes that her patients are even older than the norm, typically older than 65 years at diagnosis. To investigate, she reviews the ages of a random sample of 100 of her patients diagnosed with breast cancer. Identify the following:

2 (a) Population : (ok) Breast cancer patients (of a researcher)

3 (b) Sample : The group selected by the researcher

(c) Sample size $n = 100$

2 (d) Variable of interest: Age at diagnosis

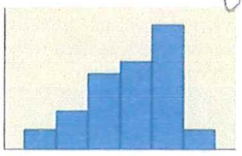
2 (e) The variable of interest is (circle one) CATEGORICAL (NUMERICAL)

2. Choose from the list below to best describe the shape of each distribution:

4 3 ½

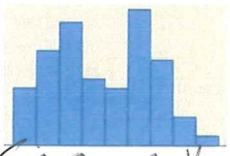
skewed left, skewed right, uniform, symmetric, unimodal (may apply to more than 1 graph), bimodal

skewed left



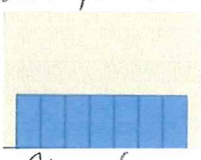
unimodal

bimodal



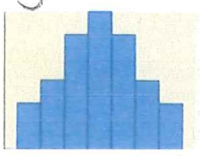
(is somewhat symmetric)

uniform



(is also symmetric)

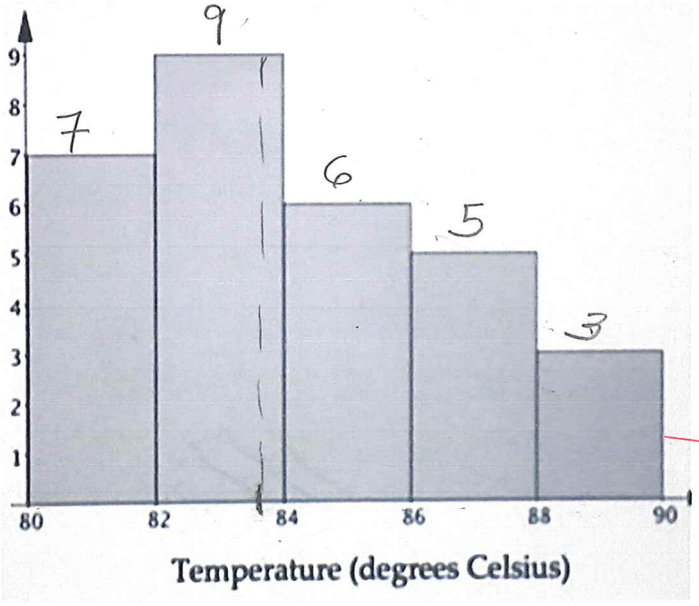
Symmetric



unimodal

8

3. (12 pts) The daily high temperatures (in degrees Fahrenheit) were recorded in SLO over a period of time in the summer. The histogram shows the distribution of temperatures over those days.



Use the histogram to answer the following questions.

2 (a) How many days were in this study? 30

2 (b) How many days had a high that was 86 degrees or more? 8

2 (c) What is the relative frequency (express as a percent) of the days that had a high that was 86 degrees or more.

$\frac{8}{30} = .267 = 26.7\%$

2 (d) Which of the following could be the median and mean of the data? (circle the best answer…no work needed)

(i) Median = 83.5, Mean = 83.5

(ii) Median = 83.5, Mean = 81.5

(iii) Median = 81.5, Mean = 83.5

(iv) Median = 83.5, Mean = 84.5

4. A study was done in 2014 to investigate the relationship between playing college football and the volume (in microliters) of the hippocampus (the part of the brain that is the center of memory, emotion, and the autonomic nervous system). The study included three groups of 25 men: healthy controls who had never played football, football players with no history of concussions, and football players with a history of concussions.

20

2 (a) This study is (circle one) a controlled experiment (an observational study)

3 (b) What is the "Treatment" (explanatory) variable? Playing College football

Is this variable categorical (qualitative) or numerical (quantitative)? (CATEGORICAL) NUMERICAL

3 (c) What is the "Response" (outcome) variable? Volume of hippocampus

Is this variable categorical (qualitative) or numerical (quantitative)? CATEGORICAL (NUMERICAL)

The figure below shows side-by-side boxplots for total hippocampus volume, in microliters.

2 (d) Which group's hippocampus data showed the most variability?

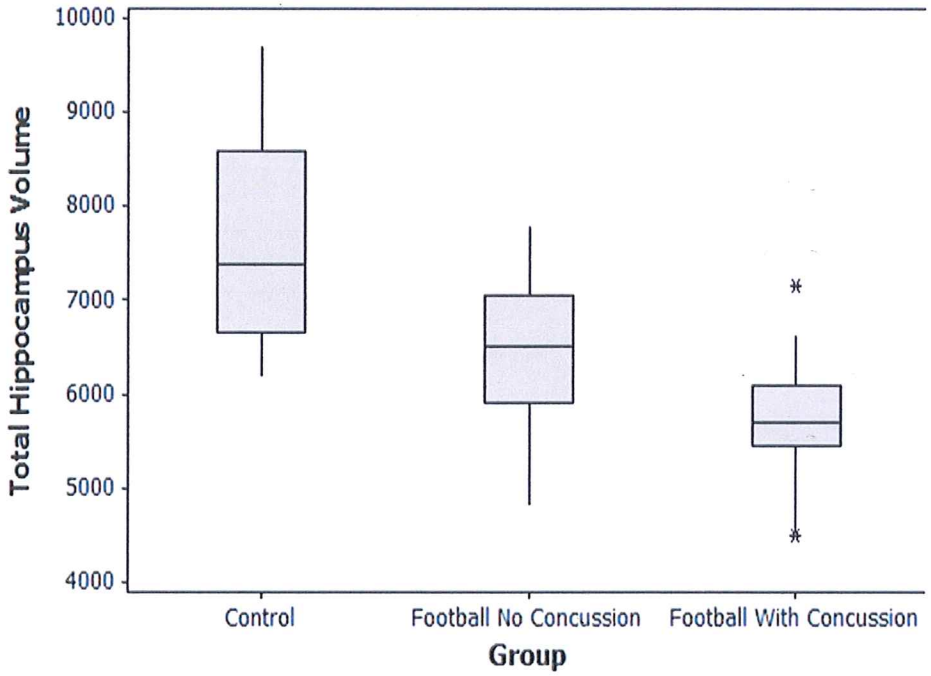
The control group

2 (e) Which group's hippocampus data were the most symmetrically distributed?

Football, no concussion

2 (f) Which group had outliers?

Football with concussion



2 (g) (True) or false: Typically, the group that didn't play football had greater hippocampus volume than those who did play football.

2 (h) True or (false): All of the subjects who didn't play football had greater hippocampus volume greater than all of the subjects who did play football.

2 (i) (True) or false: There is an association between playing college football and hippocampus size.

2 (j) Can we conclude that playing college football causes a reduction in the size of the hippocampus? Why or why not?

No we can't conclude causation because this is an observational study, not a controlled experiment.

16

5. Suppose you select a sample of five students and ask them how many texts they sent during class that day. The data values are 0, 1, 0, 12, 2 0, 0, 1, 2, 12

1 a) Construct a dotplot for this data

2 b) Find the sample mean for the data. Plot it on the dotplot.

$$\bar{x} = \frac{0+0+1+2+12}{5} = 3 \text{ texts}$$

2 c) Find the sample median for the data. Plot it on the dotplot.

Med = 1 text

1 d) Which statistic is a more "typical" value for this data: the mean (the median)

1 e) Which value appears to be an outlier? 12 texts

2 f) What effect did the outlier have on the mean?

Discuss The outlier pulled the mean up and away from the median, so away from the "typical" value"

2 g) Because of this effect, we say that the mean is not RESISTANT

5 h) Find the sample standard deviation by hand. Show work!

n=5

x	$x-\bar{x}$	$(x-\bar{x})^2$
0	−3	9
0	−3	9
1	−2	4
2	−1	1
12	9	81
Σ	0 check!	$\Sigma(x-\bar{x})^2 = 104$

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

$$= \sqrt{\frac{104}{5-1}} = \sqrt{\frac{104}{4}} = \sqrt{26}$$

$s = 5.099$ texts

6. (9 pts) Three-year-old boys have a distribution of heights which is unimodal and symmetric, with mean of 38 inches and a standard deviation of 2 inches.

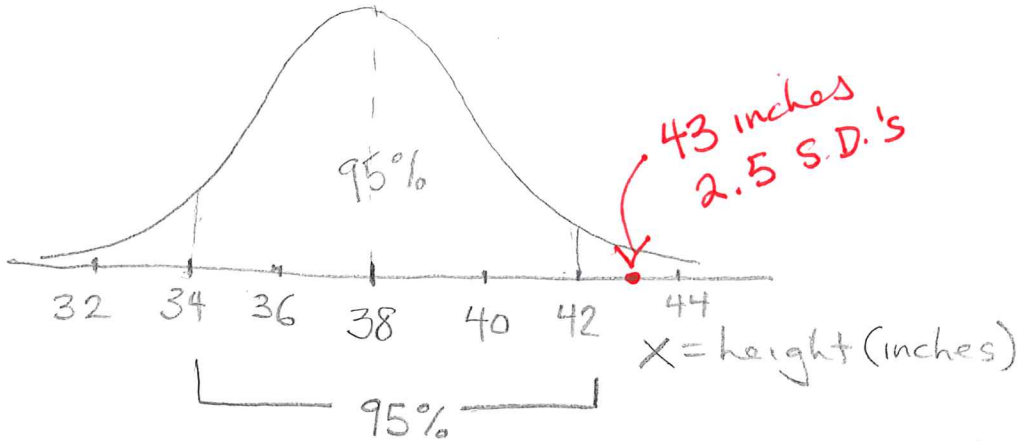
2 (a) Sketch a curve, with the x-axis labeled appropriately, showing the distribution of heights, with the standard deviation included as part of the sketch.

Parking lot

$\bar{x} = 38$ in

$s = 2$ in

Explain: "labeling axis"



2 (b) Between what two values should about 95% of the heights fall? Include units in your answer.

95% of 3-year-old boys should be between 34 in and 42 in

2 (c) Find the z-score for a three-year-old boy's height of 43 inches.

$x = 43$ in $\quad z = \frac{x - \bar{x}}{s}$

$z = 2.5$

$= \frac{43 - 38}{2} = \frac{5}{2} = 2.5$

3 (d) Suppose there are two children in a family, a three-year-old boy who is 43 inches tall and a ten-year-old girl who is 61 inches tall. If the mean and standard deviation for the heights of ten-year-old girls is 54.5 inches and 2.5 inches, determine which child is unusually tall for his/her age.

Girl's z-score:

$x = 61$

$\bar{x} = 54.5$

$s = 2.5$

$z = \frac{61 - 54.5}{2.5}$

$= \frac{6.5}{2.5}$

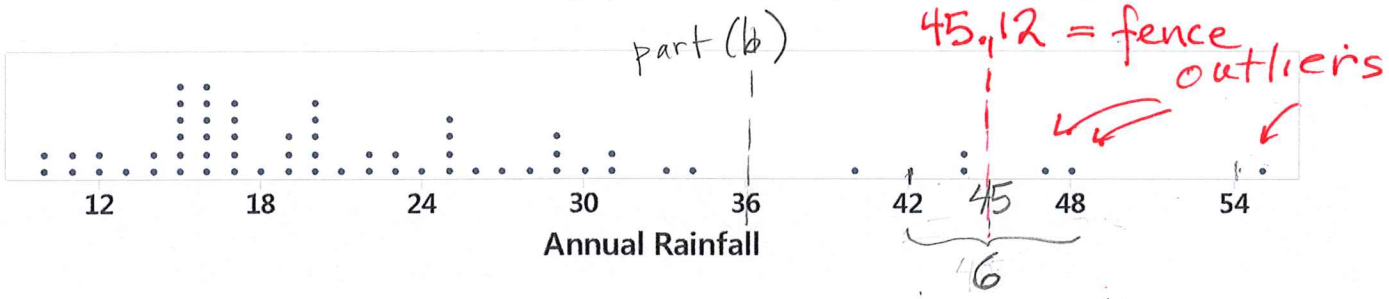
$= 2.6$

$z_{boy} = 2.5$ (part (c))

$z_{girl} = 2.6$

Since the boy's height has a z-score of 2.5 while the girl's height has $z = 2.6$, both children are unusually tall, with the girl's height being slightly more unusual.

12 7. The annual rainfall (in inches) for a number of years in San Luis Obispo is graphed below, with the summary statistics from Minitab shown below that. Use the graph and the summary stats to answer the questions.



Descriptive Statistics: Annual Rainfall

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Annual Rainfall	62	0	22.47	1.27	9.97	10.35	15.52	19.73	27.36	54.53

1 (a) How many data values are there? $n = 62$

2 (b) What percentage of the years had an annual rainfall total that was over 36 inches?

$6/62 = .097 = 9.7\%$ of the years

1 (c) Which would it be more appropriate to describe the center and variation of this data set: (circle one)

the mean and standard deviation

(the median and IQR)

1 (d) What is the five number summary for this data set? 10.35, 15.52, 19.73, 27.36, 54.33

1 (e) Find the IQR.

$IQR = Q_3 - Q_1$

$= 27.36 - 15.52 = 11.84$

$IQR = 11.84$

4 (f) Find the Lower Outlier and Upper Outlier Limits.

Lower Limit $= Q_1 - 1.5IQR$

$= 15.52 - 1.5(11.84)$

$= -2.24$ inches

Upper Limit $= Q_3 + 1.5IQR$

$= 27.36 + 1.5(11.84)$

$= 45.12$ inches

2 (g) What values are outliers? Include units!

The outliers appear to be approximately 47, 48, and 55 inches of rain.

↑ note: we can deduce this value is 54.53 from the minitab output (max) but can only approximate the others.