Math 247: Test 1 (Wright-~~Spring~~ Fall 19) **Name:** KEY

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

1. (2 pts) Suppose you gather data from your classmates by asking how many miles they live from Cuesta and whether they live with their parents or not.

   List the variables and state whether each is categorical or numerical.

   Miles from Cuesta : numerical
   Live with parents : categorical

2. (4 pts) Write what each of the following symbols stands for: n, s, $\sum$, $\bar{y}$

   n = Sample size
   s = Sample standard deviation
   $\sum$ = sum
   $\bar{y}$ = mean (average) of y-values

3. (10) A study was conducted to see whether supplementation with creatine improved soccer skills in young soccer players. Twenty male soccer players (15 – 19 years old) participated in the study. They were randomly assigned to two groups of 10 each. Group 1 took a creatine-monohydrate supplement, and Group 2 took a placebo, each for 7 days. Before and after the supplementation protocol, each subject underwent a series of soccer skill tests. The researchers found that the creatine group significantly improved more in in all soccer skills compared to the placebo group.

1 What is the research question? Will young men taking creatine develop better soccer skills?

1 Describe the sample: 20 male soccer players (15 - 19 years old)

1 What is the (implied) Population? Young men who play soccer.

1 This study is (circle one) [✓ A RANDOMIZED EXPERIMENT] AN OBSERVATIONAL STUDY

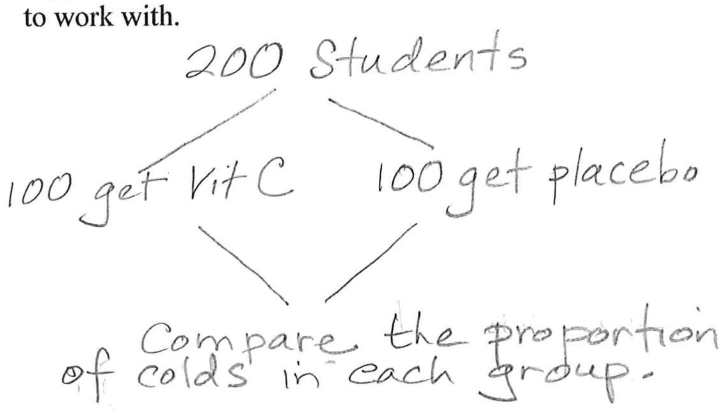
2 What are the variables in this study? Creatine and Soccer skills

1 Which of these variables is the Treatment (Factor)? Creatine

1 Which of these variable is the Outcome (Response)? Soccer skills

2 Can we say that taking creatine supplements CAUSED the soccer players to improve their skills? Why or why not? Yes, this is a Controlled, randomized experiment so there shouldn't be any confounding.

4. (4 pts) Briefly describe the design of a controlled experiment to determine whether the use of vitamin C supplements reduces the chance of getting a cold for college students. Assume you have 200 college students to work with.



1. Randomly assign 100 students to the Vit C group and 100 to the placebo group.
2. Blind the researchers as to who is in which group.
3. Apply treatment and placebo
4. Compare outcome of colds.

5. (3 pts) Suppose instead of designing an experiment about vitamin C and colds, you find 100 students who don't take vitamin C and 100 students who do take vitamin C are compare whether or not they get a cold over a 6-week period. You find that those who do take vitamin C get fewer colds.

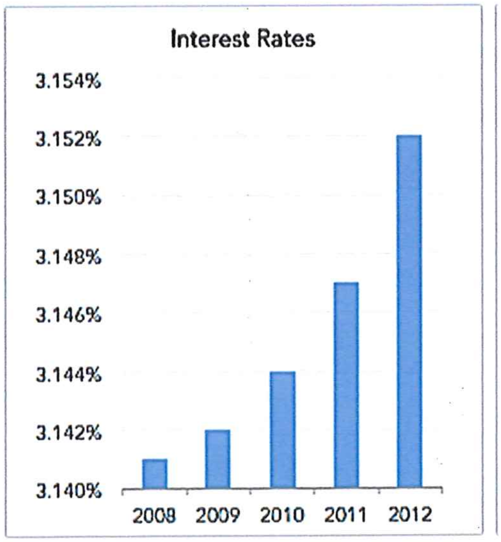
Would it be correct to state that your study shows that vitamin C CAUSES people to get fewer colds? Why or why not?

No! This is an observational study so we can't conclude cause-and-effect. The participants chose their own groups – the researcher did not assign them to groups.

Describe one potential confounder in this situation. Describe how the confounder ties the Treatment variable to the Response variable.

People who take Vitamin C might have healthier habits in general so healthy habits would link choosing to take Vit C to getting fewer colds.

6. (3 pts) The given graph shows interest rates over several years and implies there has been a shocking increase.

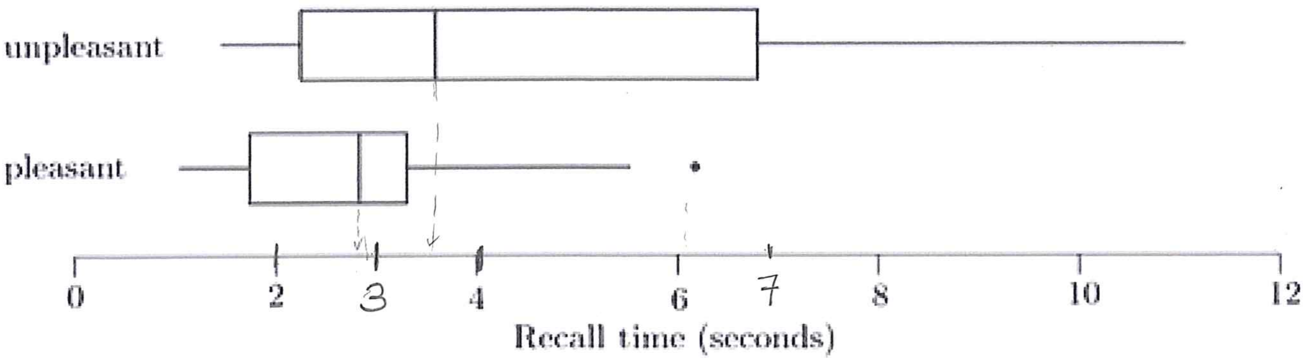


Explain why this graph is deceptive.

The y-axis does not begin at 0! There is actually very little difference in these interest rates.

7. (8 pts) **Memory recall times** In a study of memory recall times, a series of words was shown to a subject on a computer screen. For each word, the subject was instructed to recall either a pleasant or an unpleasant memory associated with that word. (Example: word = "ocean"; round 1, recall a pleasant memory; round 2, recall an unpleasant memory).

When the subject was able to recall a memory, they pressed a bar on the computer keyboard. The boxplots below show the recall times (in seconds) for twenty pleasant memories and for twenty unpleasant memories.



Estimate the median for both groups:

Median time for unpleasant memory = 3.6-ish seconds

Median time for pleasant memory = 2.8-ish seconds

many possible answers - must be reasonable

Based on these graphs, did subjects typically have an easier or harder time recalling an unpleasant memory?

Harder time - it took them longer to recall an unpleasant memory. (Apparently our brain doesn't want to remember unpleasant things!)

Which set of recall times (type of memory) showed the most variability?

Unpleasant memory recall times had the most variability.

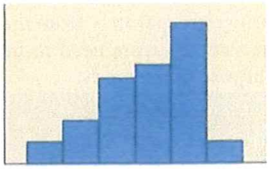
Which data set has an outlier and what is the approximate value of the outlier?

Pleasant memory recall times had an outlier of about 6.2 seconds.

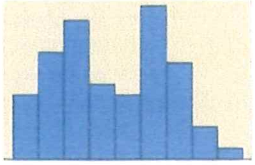
8. (2 pts) True or [✓ False]: If a data set has outliers, it's better to use the mean as a "typical value" since the mean is "resistant".

The mean is not resistant to outliers!

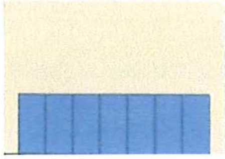
9. (4 pts) Describe the shape of each distribution (just use a word or two):



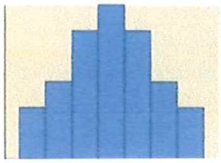
Skewed left



Bimodal

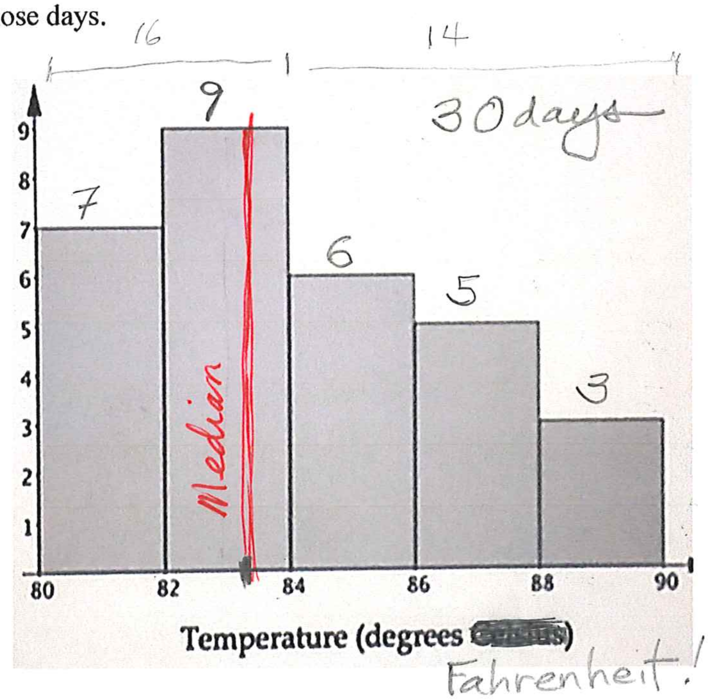


Uniform



Symmetric

10. (7 pts) The daily high temperatures (in degrees Fahrenheit) were recorded in SLO over 30 days in the summer. The histogram shows the distribution of temperatures over those days.



Use the histogram to answer the following questions.

1 (a) What is the shape of the distribution?

Skewed right

(b) Show (approximately) where the median would be for this data set.

Median between 82 and 84

1 (c) Would the mean be greater than the median or smaller than the median for this data? (Circle one)

Mean is [✓ LARGER] SMALLER than the median.

because of the right skew.

2 (d) How many days had a high that was less than 84 degrees?

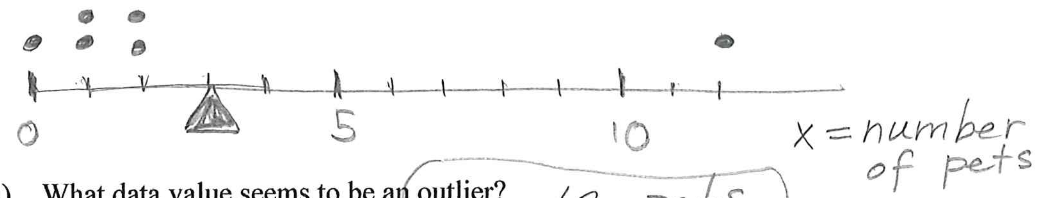
16 days

2 (e) What percentage of days had a high over 88 degrees (hint: relative frequency)? over or at 88° (typo)

$\frac{3}{30} = .10 = 10\%$

11. (18 pts) A random sample of 6 students were asked how many pets they have.
Their responses were 2, 0, 1, 12, 1, 2

2 (a) Construct a dot plot of this data.



2 (a) What data value seems to be an outlier? 12 pets

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

$$\bar{x} = \frac{\Sigma x}{n} = \frac{18}{6} = 3 \text{ pets}$$

Med = 1.5 pets   0, 1, 1 | 2, 2, 12

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

1.5 pets is more typical

6 (d) By hand, find the standard deviation of the data. Organize your work in a table.

x	$x-\bar{x}$	$(x-\bar{x})^2$
0	-3	9
1	-2	4
1	-2	4
2	-1	1
2	-1	1
12	9	81

✓ $\Sigma x-\bar{x} = 0$ $\Sigma(x-\bar{x})^2 = 100$

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

$$= \sqrt{\frac{100}{6-1}}$$

$$= \sqrt{20}$$

$S = 4.472$

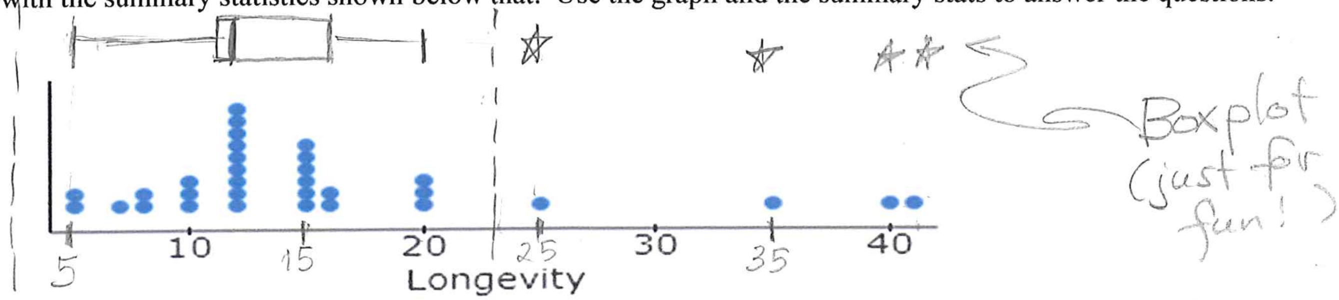
about 4.5 pets

2 (e) What is the "Sum of the Squared Error" for this data set? $\Sigma(x-\bar{x})^2 = 100$

SSE = 100

Typo

12. (13 pts) The lifespan (in years) for a number of different mammals ~~in San Luis Obispo~~ is graphed below, with the summary statistics shown below that. Use the graph and the summary stats to answer the questions.



Summary statistics:

Column	n	Mean	Variance	Std. dev.	Std. err.	Median	Range	Min	Max	Q1	Q3
Longevity	32	15.4	77.09	8.78	1.55	12	36	5	41	11	16

1 (a) How many data values are there? 32

1 (b) How many mammals had a life span of 20 years or more? 7

1 (c) What proportion (relative frequency) of mammals had a lifespan of 20 years or more? $\frac{7}{32} = .219 = 21.9\%$

1 (d) 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]

Why? The data has outliers.

2 (e) What is the five-number summary for this data set? Min, Q_1, Med, Q_3, Max
5, 11, 12, 16, 41

2 (f) Find the IQR.

$IQR = Q_3 - Q_1 = 16 - 11 = 5$ years

3 (g) Find the Lower Outlier and Upper Outlier Limits.

Lower $= Q_1 - 1.5 * IQR$
$= 11 - 1.5(5)$
$= 3.5$

Upper $= Q_3 + 1.5 * IQR$
$= 16 + 1.5(5)$
$= 23.5$

(h) Is the data value of 25 years an outlier? Explain how you can tell based on the Outlier Limits you found in (g).

Yes., Since 25 is outside of the upper limit, it is considered an outlier.