

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.

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

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.

What is the research question?

Describe the sample:

What is the (implied) Population?

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

What are the variables in this study? \_\_\_\_\_ and \_\_\_\_\_

Which of these variables is the Treatment (Factor)? \_\_\_\_\_

Which of these variable is the Outcome (Response)? \_\_\_\_\_

Can we say that taking creatine supplements CAUSED the soccer players to improve their skills? Why or why not?

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.

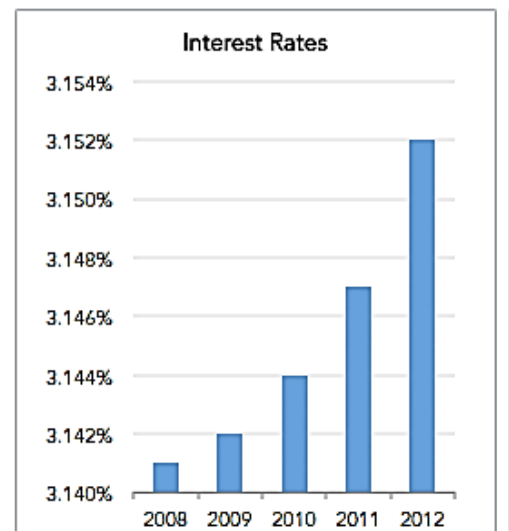
5. (3 pts) Suppose instead of designing an experiment about vitamin C and colds, you find 100 students who don't take vitamin C and 100 students who do take vitamin C are compare whether or not they get a cold over a 6-week period. You find that those who do take vitamin C get fewer colds.

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

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

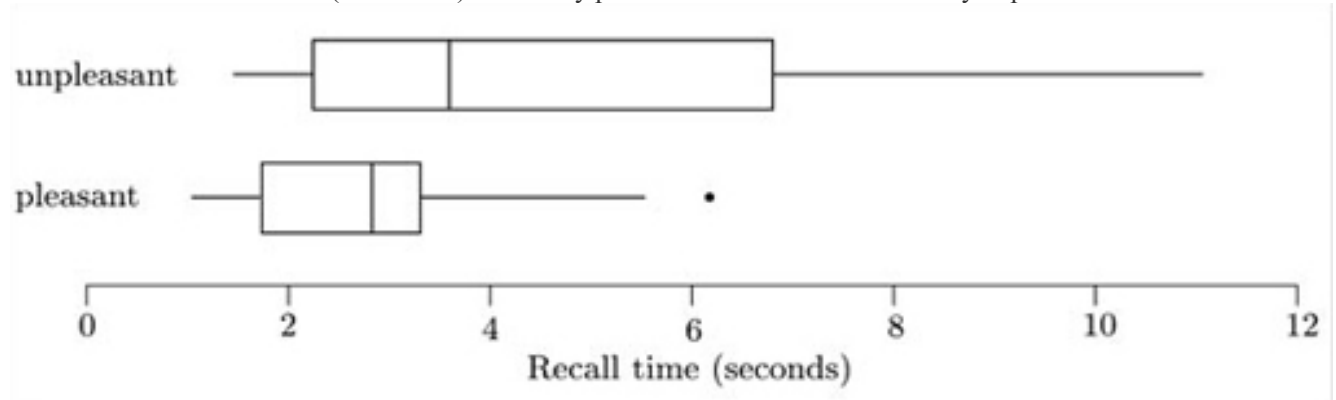
6. (3 pts) The given graph shows interest rates over several years and implies there has been a shocking increase.

Explain why this graph is deceptive.



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 = \_\_\_\_\_

Median time for pleasant memory = \_\_\_\_\_

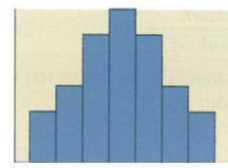
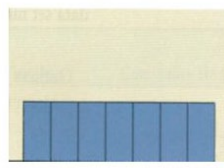
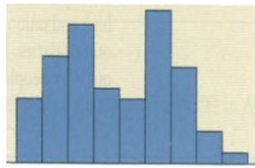
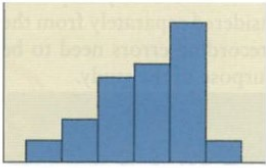
Based on these graphs, did subjects typically have an easier or harder time recalling an unpleasant memory?

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

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

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”.

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



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.

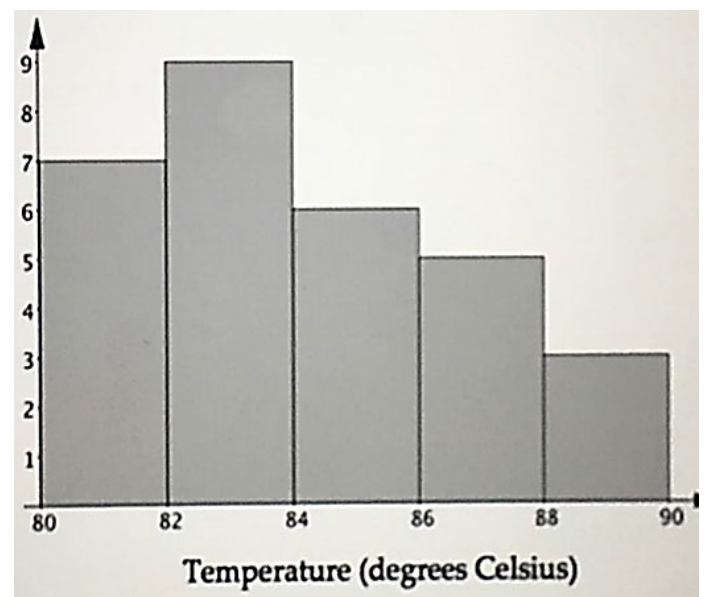
Use the histogram to answer the following questions.

(a) What is the shape of the distribution?

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

(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.



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

(e) What percentage of days had a high over 88 degrees (hint: relative frequency)?

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

(a) Construct a dot plot of this data.

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

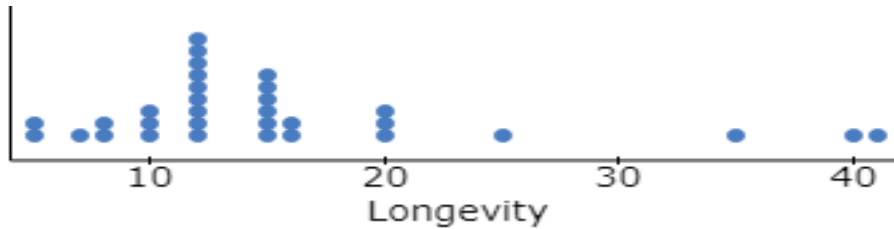
(b) Find the mean of the data and mark it with a triangle on the dotplot. Then find the median.

(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. Organize your work in a table.

(e) What is the “Sum of the Squared Error” for this data set? \_\_\_\_\_

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

- (a) How many data values are there?
- (b) How many mammals had a life span of 20 years or more?
- (c) What proportion (relative frequency) of mammals had a lifespan of 20 years or more?
- (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?

- (e) What is the five-number summary for this data set?
- (f) Find the IQR.

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

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

**Math 247: Test 1 (Take Home)**  
(20 points)

**Name:** \_\_\_\_\_

Score: \_\_\_\_\_/20

**Due date:** Tuesday, 9/10/2019

- *I encourage you to work with other people in the class but your final work must be your own. Be sure that all answers are written in your own words; i.e., do not write verbatim the same answer as another student.*
- *Scoring will be based on organization of your work, accuracy, and thoughtful, well-written answers.*
- *Use complete sentences in your answers!*

A student gathered data from 43 people, asking each to give the number of hours they sleep per night (on average), the number of hours they exercise per week, and how happy they are on a scale of 1 to 5, where 1 = Very Unhappy, 2 = Somewhat Unhappy, 3 = Neutral, 4 = Somewhat Happy, 5 = Very Happy.

The data is on [www.wrightmath.info](http://www.wrightmath.info) , Math 247, StatCrunch.

Research question: Is sleep or exercise a better predictor of happiness?

**Part 1: Explore the association between sleep and happiness.**

1. Use StatCrunch to perform a Simple Linear Regression on the sleep/happiness data. Use Sleep as the explanatory variable (independent) and Happiness as the predicted variable (dependent).

Print out the the graph and the linear regression output and include them in your response paper.

2. Describe the scatterplot including Shape, Scatter, Trend, Strength, and potential Outliers.

3. What is the correlation coefficient for Sleep and Happiness? \_\_\_\_\_

What does this tell you about the association between Sleep and Happiness?

Does this mean that more sleep causes people generally to be happier? Explain why or why not.

4. What is the equation of the regression line? Write the equation using  $x$  and  $\hat{y}$  and also using the words “Sleep” and “Happiness”.

Use the equation to predict the happiness level of someone who gets 7.5 hours of sleep, on average. Show work! Mark this point on the scatter plot.

5. What is the slope of the regression line and what does it mean in terms of Sleep and Happiness? Be specific in your answer and use units!

**Part 2: Explore the association between exercise and happiness and compare.**

1. Use StatCrunch to perform a Simple Linear Regression on the exercise/happiness data. Use Exercise as the explanatory variable (independent) and Happiness as the predicted variable (dependent).

Print out the graph and the linear regression output and include them in your response paper.

2. What is the correlation coefficient for Exercise and Happiness? \_\_\_\_\_

Compare the correlation between Sleep and Happiness vs. Exercise and Happiness. Which variable, Sleep or Exercise, is a better predictor of Happiness? How can you tell?



3. What is the slope of the regression line for Exercise and Happiness?

Since the slope is very close to \_\_\_\_\_, we can see that there isn't a significant link between Exercise and Happiness for the people in this study.

4. What is the equation of the regression line? Write the equation using  $x$  and  $\hat{y}$  and also using the words "Exercise" and "Happiness".

Use the equation to predict the happiness level of someone who gets 0 hours, 10 hours, and 20 hours of exercise, on average. Show work! Mark these points on the scatter plot.

Was there much difference in the happiness values you found for 0, 10, and 20 hours of exercise? \_\_\_\_\_  
Explain how this relates to the slope of the regression line.