

In-class test _____/70 points

Take home test _____/30 points

1. (2 pts) Suppose you gather data from your classmates by asking their eye color and age.

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

2. (5 pts) Write what each of the following symbols stands for: \bar{x} , x , n , s , Σ

3. (9 pts) The September 2011 issue of the “Berkeley Wellness Letter” said that coffee reduces the chance of prostate cancer. A study of 48,000 male health care professionals showed that those consuming the most coffee (six or more cups per day) had a 60% reduced risk of developing advanced prostate cancer.

What was the research question for the study?

What was the sample for the study?

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)? _____

Was the reporter correct in saying coffee REDUCES* the chance of prostate cancer (*which implies drinking more coffee CAUSED the reduction in prostate cancer)? Briefly explain your answer:

4. (4 pts) (a) 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.

(b) If you found that the students who took vitamin got fewer colds in your experiment, would it be correct to state that vitamin C **CAUSED** the students to get fewer colds? Assume you've done a perfect experiment! Briefly explain your answer.

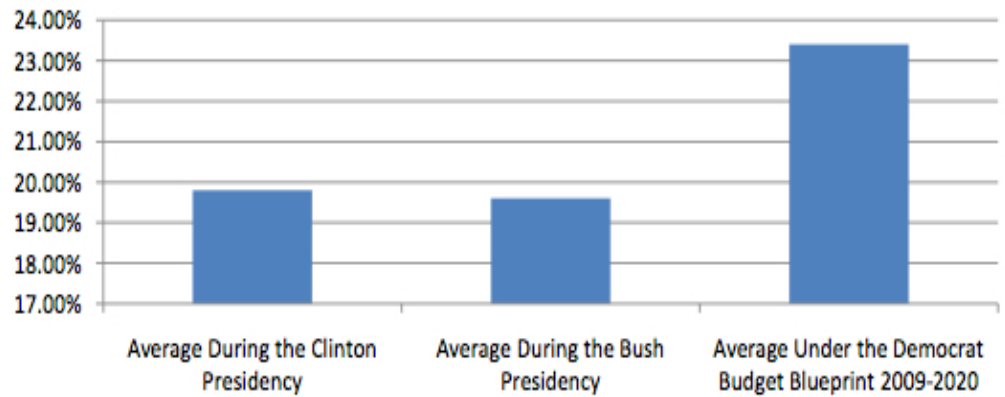
5. (3 pts) Suppose instead of designing an experiment about vitamin C and colds, you find 100 students who don't take vitamin C and 100 students who do take vitamin C are compare whether or not they get a cold over a 6-week period. You find that those who do take vitamin C get fewer colds.

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.

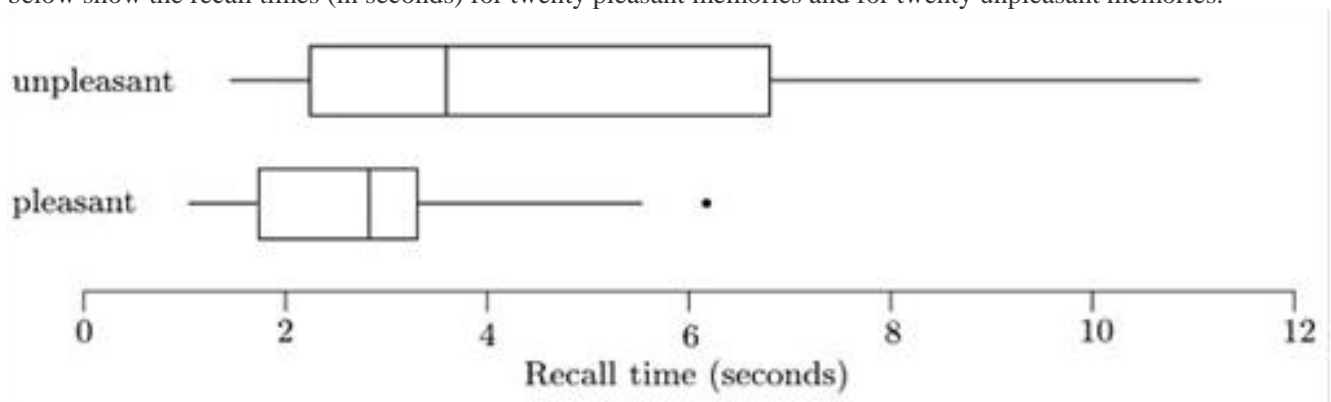
6. (3 pts) The given graph shows Federal spending as a share of the economy.

Explain why this graph is deceptive. (Write NOTHING about politics, strictly make an observation about how the graph is set up!)



7. (5 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 = _____

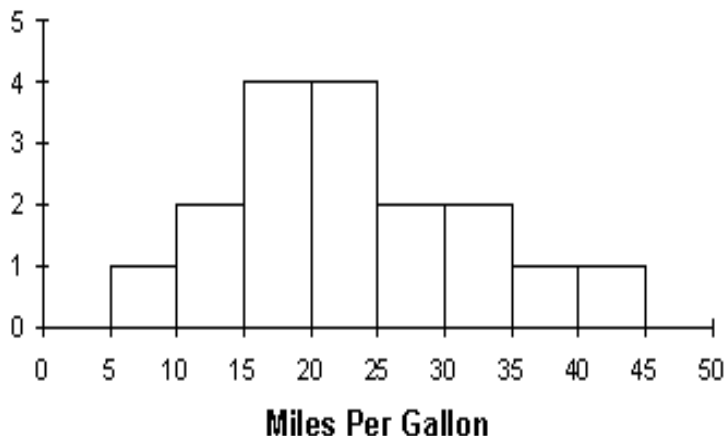
Median time for pleasant memory = _____

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. (6 pts) The distribution of gas mileage (mpg) for the top selling cars in 2015 are shown in the graph.



Use the histogram to answer the following questions.

(a) How many cars were in this study?

(b) How many cars had a gas mileage under 15 mpg? _____

(c) What is the relative frequency (express as a fraction and a decimal) of the cars that had a gas mileage under 15 mpg? _____

(d) What is the shape of the distribution? _____

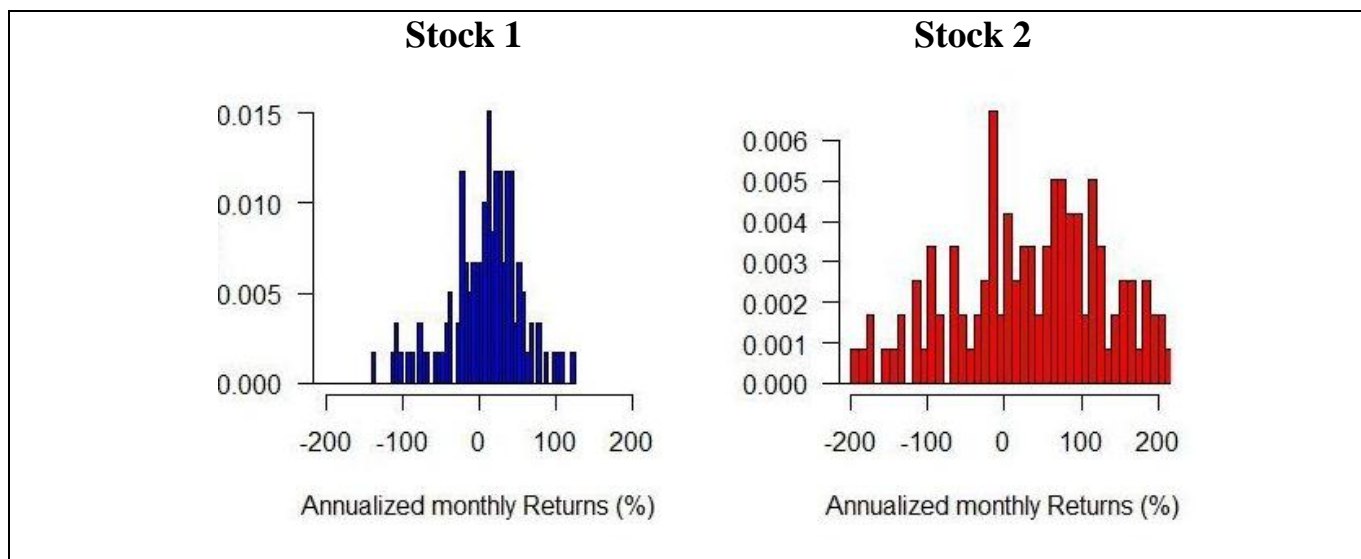
(e) Estimate the median value for the distribution: _____

(f) Based on the shape we know what about the mean relative to the median?

- (i) The mean is greater than the median
- (ii) The mean is less than the median
- (iii) The mean is exactly equal to the median
- (iv) Can't tell.

9. (2 pts) The two histograms below show the monthly returns (interest) for stocks for the S and for GM over a 10 year period. Variability is a measure of risk in stock investment (more variability means more risk). Which was the riskier stock?

- (a) Stock 1 is riskier
- (b) Stock 2 is riskier
- (c) They have equal risk
- (d) Can't tell



10. (18 pts) A random sample of 6 students were asked how many pets they have. Their responses were 2, 0, 1, 14, 1, 2

(a) Construct a dot plot of this data.

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

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

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

(e) How did the outlier affect the mean?

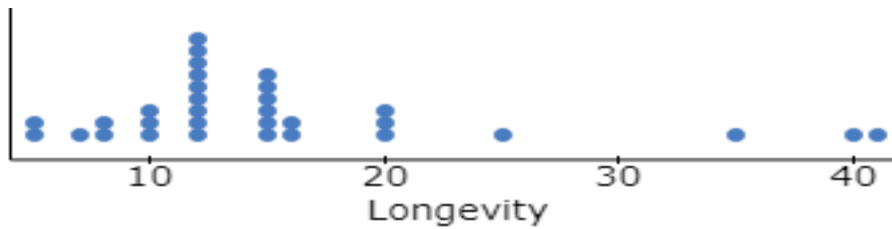
(f) Because of this effect, we say that the mean is not _____

(g) By hand, find the standard deviation of the data.

(h) How did the outlier affect the standard deviation?

(i) Is the standard deviation resistant?

11. (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 over 30 years?
- (c) What proportion (relative frequency) of mammals had a lifespan over 30 years?
- (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)
(30 points)

Name: _____

Score: _____/30

- **Due dates:**
 - **MW class:** Wednesday 2/20/19.
 - **TR class:** Thursday, 2/21/19
- *I encourage you to work with other people in the class. You're welcome to ask for help from Alysha or other tutors 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.*
- *Your work can be typed or NEATLY handwritten.*
- *Your work should incorporate all of the technology work mentioned below; i.e., **copy and paste the StatCrunch work into your write-up.***
- *Scoring will be based on organization of your work, accuracy, and thoughtful, well-written answers.*
- *Answer all of the following questions on another piece of paper and attach to this cover sheet.*
- *Use complete sentences in your answers!*

You are going to do an observational study about sleep habits of students vs. non-students.

1. Poll 25 students and 25 non-students about how many hours each person sleeps on a usual weeknight. Describe how you gathered your data.
2. Use StatCrunch to graph a dotplot, a histogram, and a boxplot for each data set.
3. Use StatCrunch to find the Summary Statistics" for each data set.
4. Write up a description of each data set, **using complete sentences!** Include shape, center, spread, potential outliers in your description.
 - For center, give both the mean and median and state which gives a better idea of what is "typical" for each group.
 - For spread, give both the standard deviation and the IQR (you'll have to find IQR by hand...show work!). State which gives the better description of spread.
5. Write a summary that compares the sleep habits of the two groups. What did you find out in comparing the groups. Is this what you expected might be true?