

**Instructions:** Show all work. Use exact answers unless specifically asked to round. Explain thoroughly using complete sentences. If you use your calculator to perform statistical tasks, say which command/operations you are using and what you entered into your calculator, and what you got back to show work. If you do not show work and the answer is incorrect, no credit will be awarded.

1. What is a meta-analysis? Give an example. (5 points)

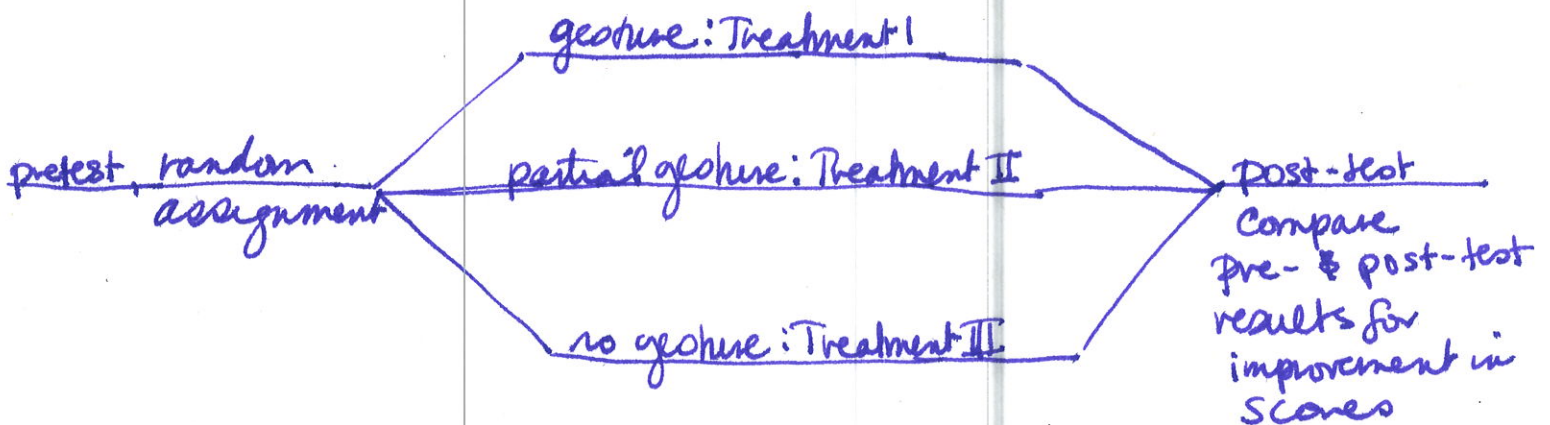
a meta-analysis is an analysis of a bunch of past analyses on the same subject to look at what the state of the field is, what the consensus is, if there is one and if there are any gaps or questions to spur new research

2. A state decides to study teaching effectiveness in public schools. The state randomly selects 50 public schools and then surveys every teacher at each of those 50 schools. What kind of sampling method is this? Explain your reasoning. (5 points)

Cluster Sampling

each school is a "cluster" and then everyone in the randomly selected cluster is selected.

3. A study was conducted to test whether gestures can enhance math performance. Students were divided into demographic groups and given a pre-test to assess their prior knowledge. The groups were then randomly divided into three groups for the test conditions of the lesson: gesture, no gesture, and partial gesture. After the lesson, the students were given a post-test to assess their learning. Pre-test and post-test scores were compared. Diagram the experiment described. Be sure your diagram clearly states all treatment groups, and how the groups were determined, and what assessments were made. (10 points)

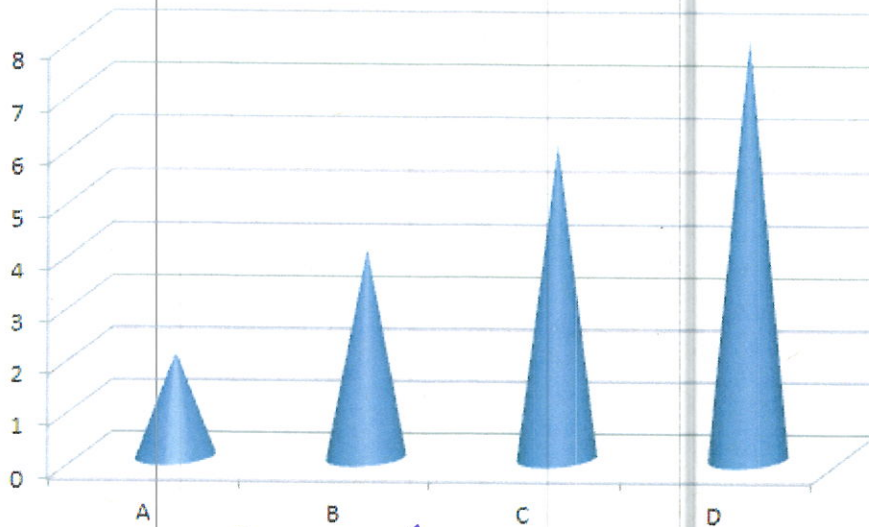


it may be that treatments were also block designed within each treatment carrying forward the ethnic groups, however, this is unclear and should be managed w/ it by comparing individual pre- & post results





6. An example of a bad graph is shown below. List (at least) 4 things that are wrong/misleading about the graph. (8 points)

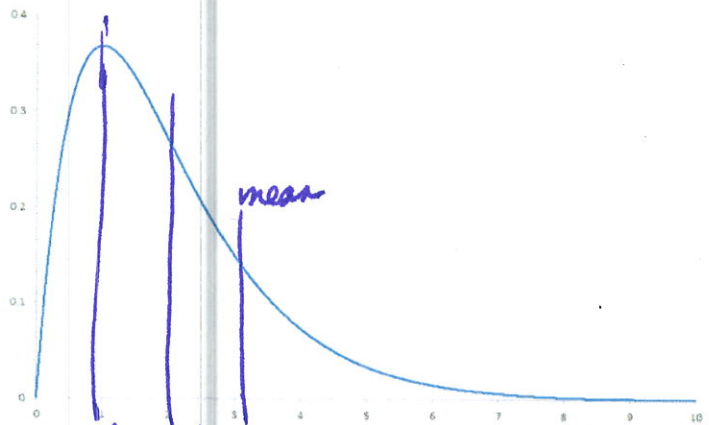


- 1) don't use cones for bar graphs. Our eyes see area and the area does not scale like height of cones.
- 2) don't use 3D effects. They make the graph more difficult to read
- 3) the axes are not labeled.
- 4) The graph has no title - what is being graphed here?

7. A distribution curve is shown to the right. Use it to answer the questions that follow.

- a. What is the shape of the distribution? (2 points)

right skewed

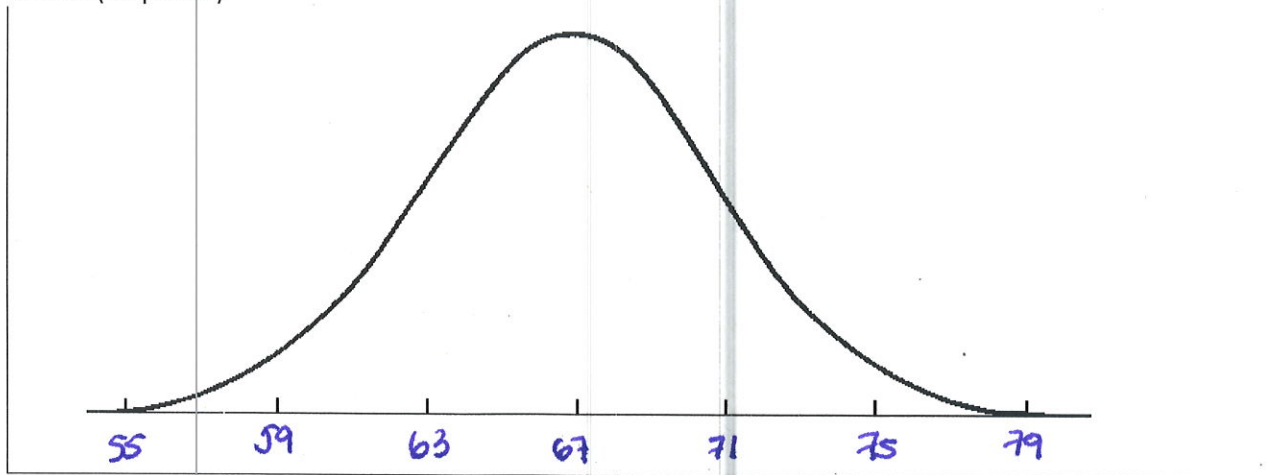


- b. On the graph, label the approximate (relative) positions of the mean, median and mode. (6 points)

- c. Which measure of center should be used to describe a "typical" value in the distribution? (2 points)

median

8. The speed of automobiles on a particular stretch of highway is 67 mph with a standard deviation of 4 mph. Use the graph below to draw the distribution, and use it to answer the questions that follow. (12 points)



- a. What percentage of cars are driving at speeds between 59 and 75 mph?

95% (between -2 & 2 st. dev)

- b. What percentage of cars are driving faster than 79 mph?

0.15%

(above 3<sup>rd</sup>  
 $\frac{100 - 99.7}{2} = .15$ )

- c. What percentile does a vehicle represent if it is driving at 63 mph?

16%

(below -1  
 $\frac{100 - 84}{2} = 16\%$ )

9. Mrs. Baker teaches second grade, and is speaking to the parents of one of her students about their child's scores. Siobhan received an 80% on two tests in two different classes, but one test of was marked as a C, and one was marked as a B. On one test the mean was 75%, with a standard deviation of 9%, and the other one had a mean of 69% with a standard deviation of 7%. Use z-scores to try to explain to the parents why the grades are different. (8 points)

Test I

$$\frac{80 - 75}{9} = .56$$

Test II

$$\frac{80 - 69}{7} = 1.57$$

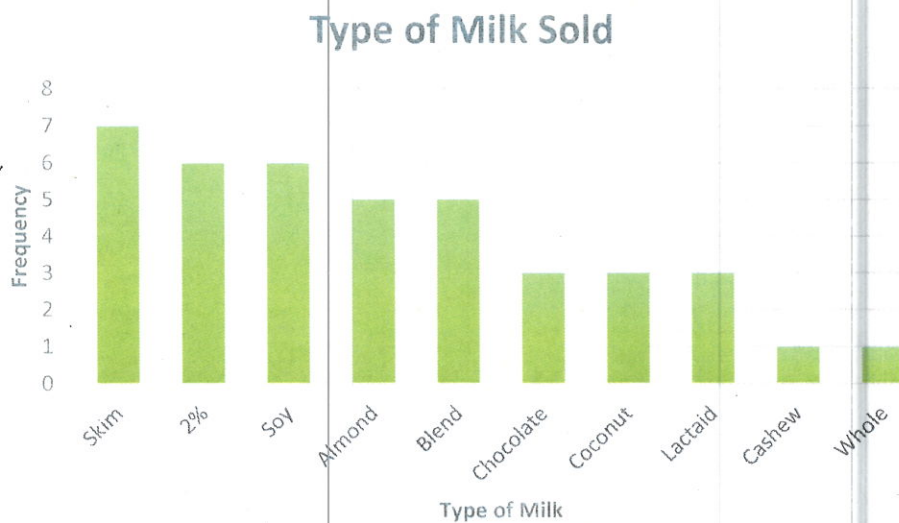
The first test was "easier" and Siobhan did very similar to the bulk of the class. Students near the middle (the middle 68%) receive a C. The second test was "harder" with a lower mean. Siobhan did much better than most of the students and so earned the higher grade.

The following problems are based on the data that was provided to you in advance. Submit all the analysis you produced. Indicate where the answers can be found in your attached sheets (if it can be found there), and record your answers to the questions below. If you have not previously calculated the answer to the question, the data is provided so that you can do it now.

10. Use the data set below to construct a frequency table. Then build a Pareto graph from the table. (15 points)

Whole	Chocolate	Almond	Soy	Lactaid	Coconut	Blend	Skim	2%	Skim
Almond	Coconut	Skim	Soy	Blend	Skim	2%	2%	Almond	Cashew
Blend	Chocolate	Skim	Lactaid	2%	Soy	Almond	Blend	Skim	2%
Chocolate	Soy	Lactaid	Blend	Skim	2%	Soy	Soy	Almond	Coconut

Type of Milk	Frequency
2%	6
Almond	5
Blend	5
Cashew	1
Chocolate	3
Coconut	3
Lactaid	3
Skim	7
Soy	6
Whole	1



11. Use the data set below to answer the questions that follow.

20	0	72	64	69	60	0	14	27	71
7	5	67	65	27	58	10	14	19	5
50	26	71	9	26	22	30	56	6	16
30	29	37	22	7	24	17	36	15	21
27	7	15	9	8	12	6	98	23	8

a. Construct a frequency table using 8 classes. (7 points)

Class	Frequency	Cumulative Frequency

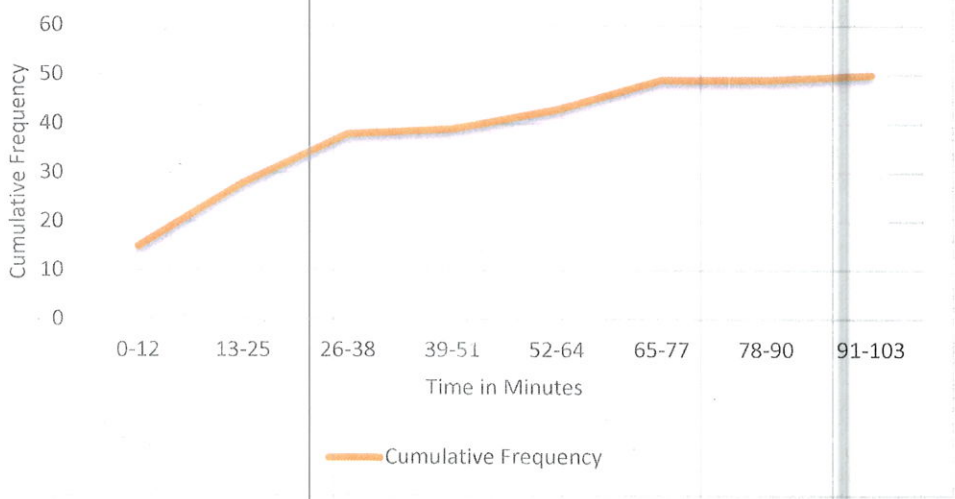
b. Draw an Ogive graph of the data. (5 points)

*See next page*

c. Find the five-number summary of the data. (5 points)

Commuting Time	Frequency	Cumulative Frequency
0-12	15	15
13-25	13	28
26-38	10	38
39-51	1	39
52-64	4	43
65-77	6	49
78-90	0	49
91-103	1	50

Ogive of Commuting Times



Min	0
Q1	9
Median	22
Q3	37
Max	98



- d. Draw a boxplot with outliers. (10 points)

See next page

- e. What percentile does a 30-minute commute represent? (4 points)

$$\frac{36}{50} = 72^{\text{nd}} \text{ percentile}$$

- f. What value represents the 60<sup>th</sup> percentile? (4 points)

$$30^{\text{th}} \text{ value} = 26 \text{ minutes}$$

- g. What is the mean and standard deviation of the data? (6 points)

$$\text{mean} = 28.7$$

$$\text{st dev} = 23.93$$

- h. Describe the shape of the distribution. Based on that, which measure of central tendency would be best to use for this data? (4 points)

Skewed right or bimodal  
use median



100  
98 ●

70  
72

40  
40.25

30  
28.74

20  
22

10  
9

0  
1  
0



12. Use the data in the table below to answer the questions that follow.

Time of daily physical exercise (min)	Cholesterol (mg/dl)	Time of daily physical exercise (min)	Cholesterol (mg/dl)
5	224	20	176
5	221	20	165
5	210	20	163
5	195	20	173
5	224	20	169
10	209	25	166
10	194	25	156
10	182	25	145
10	195	25	165
10	177	25	159
15	183		
15	175		
15	166		
15	193		
15	178		

a. Sketch a scatterplot of the data. (7 points)

*See next page*

b. Find the linear regression equation that best fits the data. (6 points)

$$y = -2.708x + 223.14$$

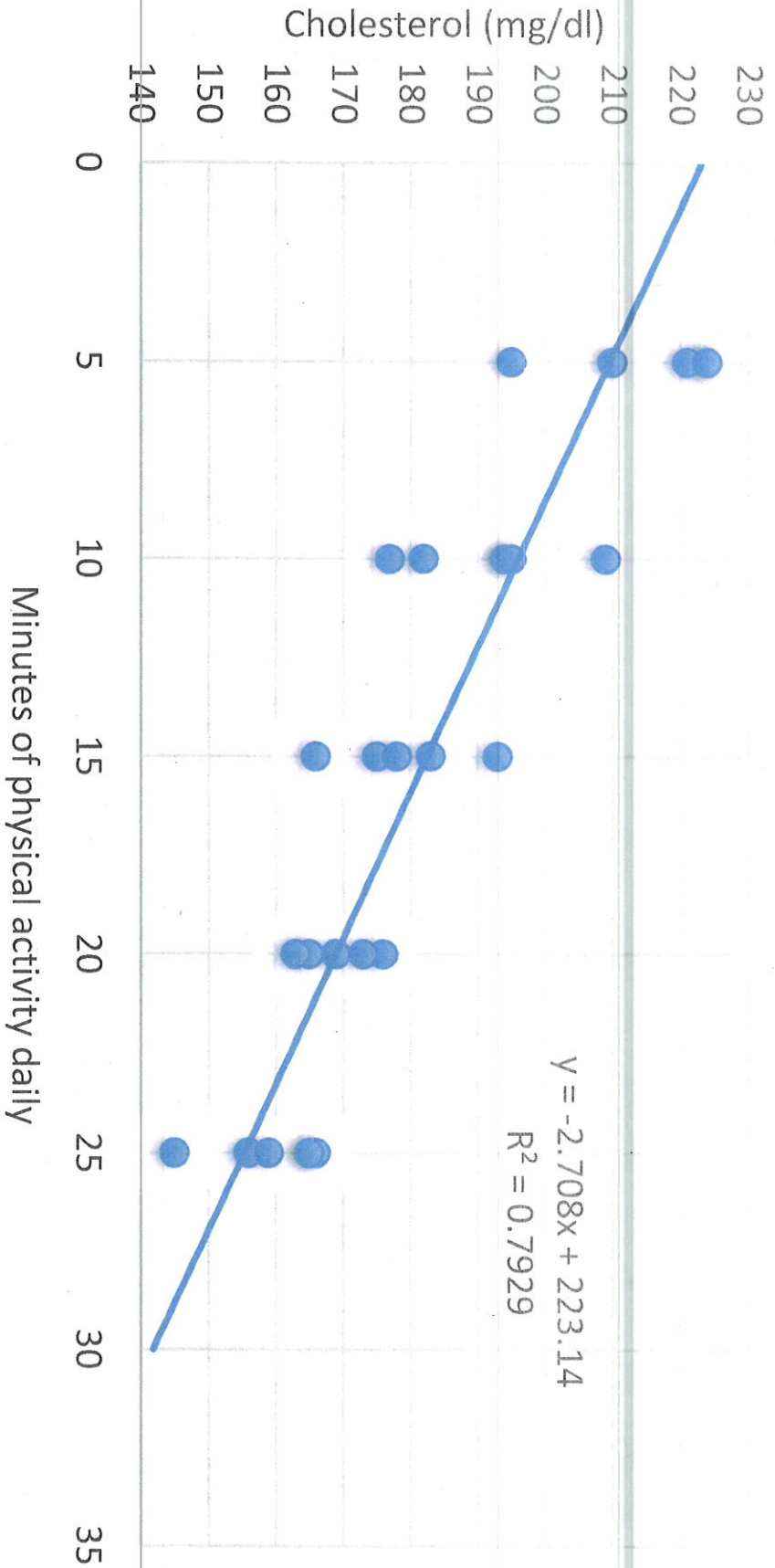
c. What is the correlation? (3 points)

$$r = -.8904$$

d. What is the coefficient of determination? (3 points)

$$r^2 = .7929$$

# Daily Physical Exercise vs. Cholesterol



- e. Interpret the slope in the context of the problem. (5 points)

for each additional minute of physical activity cholesterol will/should decrease by 2.7 mg/dl.

- f. What percent of the change in cholesterol can be explained by a change in time spent on physical activity? (3 points)

79.29%

- g. What does the y-intercept mean in the context of the problem? (5 points)

w/o any physical activity, the average cholesterol level is 223 mg/dl.