

**Introduction to Statistics**  
Practice for Test One

Show all work clearly. You may use a calculator. You have 80 minutes.

**SECTION ONE: Multiple Choice** (5 points for each answer: 50 points total)

1. For the list

$$20, 20, 0, 5, 10, -5, 10, 15, 15, 20, -40,$$

circle the answers corresponding to the approximate average, approximate standard deviation, and median.

a) Average: The answer is **A** since  $\frac{20+20+0+\dots+(-40)}{11} \approx 6.36$

b) Standard Deviation: The answer is **B** by calculating

$$\sqrt{\frac{20^2 + 20^2 + 0^2 + \dots + (-40)^2}{11}} - (6.36)^2.$$

c) Median: The answer is **B**. You must first *order* the data to get

$$-40, -5, 0, 5, 10, 10, 15, 15, 20, 20, 20$$

and then take the middle number (the sixth from the left).

2. Consider the number of cars each family in Syracuse owns. Suppose the distribution is

cars	0	1	2-3	4-7	More than 7
families (in 10,000s)	4	3	6	2	0

a) In the histogram of this data, if the bar over the interval 4–7 cars is  $\frac{1}{2}$  inch tall, how tall should the bar over 2–3 cars be?

The answer is **C**. The rectangle above the class interval 4-7 has area 2 ( $= 4 \times (1/2)$ ). Hence, each family accounts for an area of 1. Thus, the area above the class interval 2–3 should be 6. Hence, the height  $h$  solves  $2h = 6$ , or  $h = 3$ .

b) The average of the distribution above is about 2. The SD is closest to which of the following.

The answer is **B**. Within 1 standard deviation of the average, we should have roughly  $2/3$  of the data. The only answer that could possibly be correct is 1.3.

3. It is found that the scatter plot of hours of tv viewed per week and hours of sleep is football shaped with a concentration of data towards the middle. We have found that the average number of hours of tv viewed is 15 hours with a standard deviation of 5 hours; and the average number of hours of sleep is 7.5 hours with a standard deviation of 1.5 hours. The correlation coefficient was calculated to be  $-.4$ .

How many hours of sleep, with error estimate, would you predict that a person who watches 20 hours of tv per week would receive?

The answer is **A**. To get 6.9 we use the regression line. Remember, we always put the variable we are predicting on the left side. Thus,

$$\text{sleep} = (-.4) \left( \frac{1.5}{15} \right) (20 - 15) + 7.5$$

and we get sleep = 6.9.

4. Below is a (fictional) four week study of outside temperature versus the number of crimes for that night. Let  $X$  be the average nightly temperature for the week and  $Y$  be the average number of crimes per night for that same week.

$X$	$Y$	$XY$	
62.3	220.1	13712.23	
60.0	210.3	12618.00	
66.7	239.9	16001.33	
58.2	199.8	11628.36	
61.8	217.535	13489.98	← averages

The correlation coefficient for the average nightly temperature for the week versus the average number of violent crimes per night is approximately

The answer is **B**. We create the xy column above and calculate the averages of each column as well as  $SD_x$  and  $SD_y$ . Using

$$r = \frac{13489.98 - (61.8)(217.535)}{SD_x SD_y}$$

completes the problem.

**SECTION TWO: Short Answer (5 points for each answer: 30 points total)**

5. Consider the data from question 5.
- a) Interpret the meaning of your answer from question 5.
- $r \approx 1$  so there is almost a perfect linear correlation between weekly average temperature at night and weekly average number of crimes. Roughly, for a 1 degree increase in temperature, the number of crimes increases by  $\approx 5$ .*
- b) If we had the temperature and crime data for each of the 28 days, do you suspect that its correlation coefficient would be the same, higher, or lower? Explain your reasoning.
- I suspect it would be lower since we are looking at averaged data. Since averages smooth away variation, the correlation tends to be higher when comparing averaged data than it would be with the raw data.*
6. Consider the dataset of odometer readings for cars at Colgate (how many miles the car has been driven). Would you expect the median or average to be larger? Briefly justify your answer.

*I expect the average to be larger. To support this, I believe that most cars at Colgate are new and have low mileage. Hence, the histogram would be left-sided (tail on the right). Since the average follows the tail, the average would be higher.*

7. Actual studies show that the accident and fatality rates for minivans are lower than for cars. (It doesn't seem to matter how this is specifically measured, the data show lower rates for minivans.) Yet many minivans don't have many of the safety features required of cars. Explain whether or not and why this data shows that minivans are safer to drive than other vehicles.

*This does not show that minivans are safer. Correlation is not causation. Perhaps, people with minivans are better drivers.*

8. Data is collected for two colleges on student car ownership. The data is divided by where students live: dorms, group housing (specialty houses and Greek system houses), and off campus. The following facts are found:

- A greater number of off campus students own cars at Mental State than at Ilove U.
- A greater number of dorm dwellers own cars at Mental State than at Ilove U.
- A greater number of group house students own cars at Mental State than at Ilove U.
- 20% of the dorm dwelling students at Mental State own cars compared to 10% of the dorm dwelling students at Ilove U.
- 80% of the group house students at Mental State own cars compared to 60% of the group house students at Ilove U.
- 95% of the off campus students at Mental State own cars compared to 90% of the off campus students at Ilove U.

Based on these facts, answer the following questions.

- a) Does this imply that more students at Mental State own cars than students at Ilove U.? Briefly explain.

*Yes. More students in each category own cars, so more total students must own cars.*

- b) Does this imply that a higher percentage of students at Mental State own cars than students at Ilove U.? Briefly explain.

*No. Simpson's paradox could come into play here, i.e. the trends of subsets could be reversed when viewed in whole. This is because Mental State's overall average is between 20% and 95% Ilove U's is between 10% and 90%, clearly overlapping intervals. Consider the following example (note that the actual number of students in each category is higher for Mental State than at Ilove U).*

	Mental	Ilove
dorm	20/100	1/10
group	8/10	6/10
off	95/100	90/100
Total	123/210 $\approx$ 58.6%	97/120 $\approx$ 80.8%