MATH 1150 Test 1

Name: Solutions

Term: Spring 2018

Instructor: Alex Rice

You have 65 minutes to complete seven multi-part problems, scored out of a total of 100 points. The only things allowed on your desk are your copy of the test, scratch paper, writing utensils, a calculator, and one 8.5×11 handwritten sheet of paper that you prepared yourself. Correct final answers without properly shown work and appropriate notation will not receive full credit. Conversely, properly shown and executed steps can earn substantial partial credit, even in the absence of a correct final answer. I am, of course, assuming full knowledge and strict compliance with the Millsaps College Honor Code. Please provide your signature after the following pledge.

I hereby certify that I have neither given nor received unauthorized aid on this assignment.

Signature: (1 point)

Data Table A

16 adult residents of downtown Atlanta were randomly selected. For each person, the researcher recorded their gender (M/F), the square footage of their current home, and whether or not they own a vehicle (Y/N).

Person	Gender	Square Footage	Own Vehicle?
A	M	1075	Y
B	M	750	N
C	F	940	Y
D	M	770	Y
E	F	640	N
F	F	1220	N
G	F	800	N
H	M	1800	Y
I	M	450	N
J	F	2950	Y
K	M	680	Y
L	F	920	N
M	F	1125	Y
N	M	780	N
O	F	~~1290~~ 1310	Y
P	M	1900	N

Problems 1 through 4 concern Data Table A

Problem 1 (5 points each)

a) How many cases are there? How many variables?

Cases: 16

variables: 3

b) Which of the variables are categorical? Which are quantitative?

cat: gender, own vehicle?

quant: square footage

Problem 2 (5 points each)

a) In the collected sample, what proportion of individuals own a vehicle? Use the appropriate notation in your answer.

$$\hat{p} = \frac{8}{16} = 50\%$$

b) Create a two-way table for the gender and vehicle ownership variables.

Gender \ Vehicle?	Y	N	Total
M	4	4	8
F	4	4	8
Total	8	8	

c) Based on this data, is there an association between gender and vehicle ownership? Explain your reasoning.

No, as the proportions of vehicle ownership amongst men and women are the same.

Problem 3: Suppose we attempt to use Data Table A to address the question: What portion of U.S. adults own a vehicle? (5 points each)

a) What is the population? What is the sample?

pop.: all US adults

sample: these 16 Atlanta residents

b) Do you believe that this a good sample to address this question? If so, explain why, and if not, identify a problem with the sample and discuss how it might impact the data.

No, there is huge bias caused by selecting all cases from the same city. In particular, vehicle ownership is likely less common in dense urban areas.

Problem 4: Amongst individuals in the survey who own a vehicle, the home square footages in increasing order are as follows:

680, 770, 940, 1075, 1125, 1310, 1800, 2950

Meanwhile, amongst individuals in the survey who do not own a vehicle, the home square footages in increasing order are as follows:

640, 750, 780, 800, 920, 1110, 1220, 1900

a) Compute the mean, standard deviation, five number summary, and inner quartile range (using the appropriate notation for everything) for the home square footages of the 8 individuals who DO own a vehicle. Only use your calculator for arithmetic, do not use any statistics functionality to automatically generate this information, and show your work. (10 points)

mean: $\bar{x} = \frac{680 + 770 + \cdots + 1800 + 2950}{8} = 1331.25$

std dev: $s = \sqrt{\frac{(680 - 1331.25)^2 + \cdots + (2950 - 1331.25)^2}{7}} \approx 740.75$

min: 680

max: 2950

median: avg of 1075, 1125 = 1100

Q_1: median of first 4 = 855

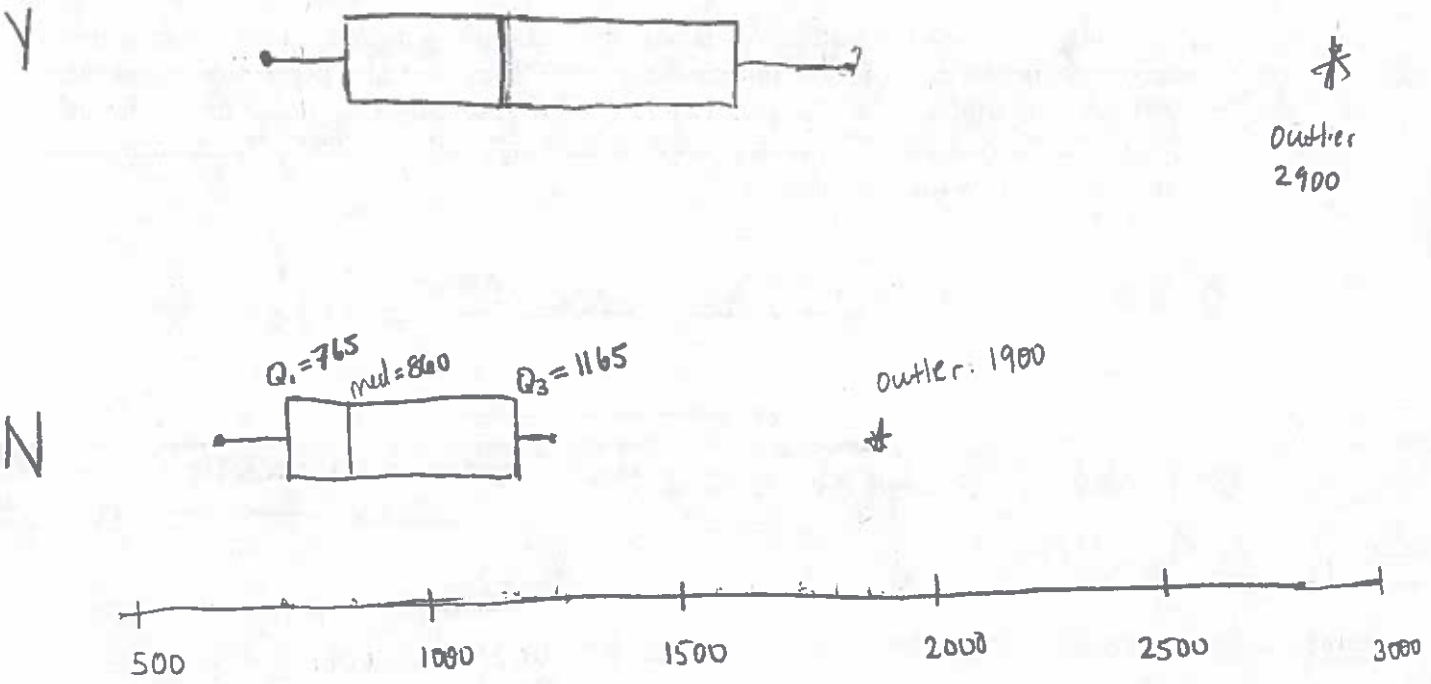
Q_3: median of last 4 = 1555

$IQR = Q_3 - Q_1 = 700$

b) Would you describe the data you analyzed in part a) as symmetric, skewed left, or skewed right? Explain your reasoning. (5 points)

Skewed right, as evidenced by the mean being larger than the median, and the large values 1800 and 2950 represent the "tail" going right

c) Create a side-by-side graph with two boxplots: one for the square footages of individuals who DO own a ~~home~~ vehicle, and one for the square footages of individuals who do NOT own a ~~home~~ vehicle. (10 points)



d) There appears to be an association between the vehicle ownership variable and the home square footage variable. Do you think this association is a causal association? If yes, explain your reasoning. If no, identify a potential confounding variable. (5 points)

Likely not causal, as wealth/income are major confounding variables.

Problem 5: (6 points each) Suppose we know that for the weight, in pounds, of adult male American pit bull terriers, the mean is $\mu = 48$, while the standard deviation is $\sigma = 6$.

Suppose we also know that for the height, in feet, of adult female giraffes, the mean is $\mu = 14$, while the standard deviation is $\sigma = 2$.

Finally, suppose that the data for both of these quantitative variables is symmetric and bell-shaped.

a) Provide a range within which the weights of approximately 95% of adult male American pit bull terriers lie.

$48 - 2(6) = 36$

$48 + 2(6) = 60$

95% between 36 lbs and 60 lbs.

b) Based on the provided information, which of the following cases is more extreme: an adult male American pit bull terrier that weighs a whopping 70 pounds, or an adult female giraffe that stands only 6 feet tall?

70 lb pit bull

$$z = \frac{70-48}{6} = \frac{22}{6} \approx 3.67$$

6 ft giraffe

$$z = \frac{6-14}{2} = -4$$

farther from zero

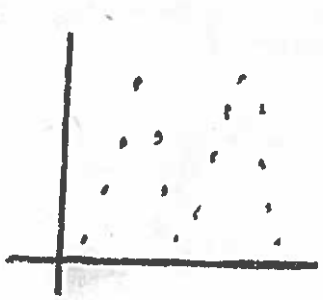
so the 6 ft giraffe is more extreme

Problem 6: (2 points each) Match the following scatterplots to their correlation values.

a) $r = .03$
b) $r = -.989$
c) $r = .83$
d) $r = -.53$
e) $r = .39$

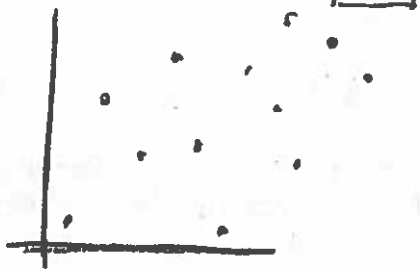
d

a

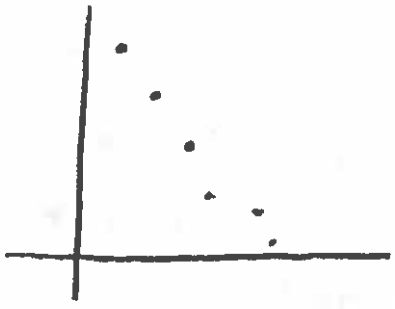


c

e



b



Problem 7: (3 points each) Suppose we collect data on the cost (in dollars), and distance (in miles) of one-way flights. Based on the data, we apply linear regression and obtain the regression line

$$\widehat{\text{Price}} = 97 + 0.15(\text{Distance})$$

a) Based on the notation above, what is the explanatory variable, and what is the response variable?

exp: distance

resp: price

b) What price would the regression line predict for a one-way ticket for a 1000 mile flight?

$$\hat{P}rice = 97 + 0.15(1000)$$

$$= 247 \text{ dollars}$$

c) If one of the actual collected data points was a 2500 mile one-way flight that cost $315, compute the residual for this data point.

$$\widehat{Price} = 97 + 0.15(2500)$$
$$= 472$$

$$Price = 315$$

$$residual = Price - \widehat{Price}$$
$$= 315 - 472 = \boxed{-157}$$

d) How can we interpret the slope of the regression line, 0.15, in this context?

The regression line predicts that for each additional mile of your trip, the price will increase 15¢.