

Exam I

STAT 305, Section D FALL 2019

Instructions

- The exam is scheduled for 80 minutes, from 8:00 to 9:20 AM. At 9:20 AM the exam will end.
- A formula sheet is attached to the end of the exam. Feel free to tear it off.
- You may use a calculator during this exam.
- Answer the questions in the space provided. If you run out of room, continue on the back of the page.
- If you have any questions about, or need clarification on the meaning of an item on this exam, please ask your instructor. No other form of external help is permitted attempting to receive help or provide help to others will be considered cheating.
- **Do not cheat on this exam.** Academic integrity demands an honest and fair testing environment. Cheating will not be tolerated and will result in an immediate score of 0 on the exam and an incident report will be submitted to the dean's office.

Name: _____

Student ID: _____

1. (2 points) Circle the **bold face** term that makes the following statement true:

A measurement device that reports the true measurement of the item on which the device is being used is (**precise** or **accurate**).

2. A sample of size 5 was drawn from a population and the resulting observations are reported below.

12, 15, 18, 19, 26

Using these observed values, report the following:

(a) (2 points) the mean

(b) (2 points) the median

(c) (2 points) the variance

(d) (2 points) the standard deviation

(e) (2 points) the value of $Q(.25)$

(f) (2 points) the interquartile range

3. An environmental engineer is testing four methods for reducing the concentration of a certain lake pollutant found in Iowa lakes. To do this he first randomly selected 20 Iowa lakes from which he took water samples, then split each of the 20 samples into 4 portions, and randomly labeled the four portions 1, 2, 3, and 4. Finally, he attempted to reduce the concentration of each of the portions labeled 1 using the first method, of each of the portions labeled 2 using the second method, of each of the portions labeled 3 using the third method, and of each of the portions labeled portion 4 using the fourth method. After the methods had been applied, he measured the change in concentration.

(a) (2 points) Is this an experiment or an observational study? Explain.

(b) Identify the following (if there was not one, simply put "not used").

i. (2 points) Response variable(s):

ii. (2 points) Experimental variable(s):

iii. (2 points) Blocking variable(s):

(c) (2 points) Was replication used in this experiment? If so, where was it applied? If not, how could we have applied it?

4. Aisha recently discovered she has the opportunity to upgrade her smart phone. She narrowed her choices down to two phones (we will call them phone A and phone B) but had a hard time making her final decision. She decided to ask people she knew who had one of the phones to rate their satisfaction from 0% to 100%. She also asked them if they would prefer to have the other phone. In order to help put their feelings in perspective, she also made note of how negative she thought they were in general (since critical people might be harsher in their criticism in general), using three descriptions: the interviewee was classified as overly critical, appropriately critical, or not critical enough.

(a) (2 points) Is this an experiment or an observational study?

(b) (2 points) What is the population under study?

(c) (2 points) Identify the response variable(s).

(d) For each of the following variables,

- Identify whether it is qualitative or quantitative variable, and
- If it is qualitative, what are the possible values it can take? If it is quantitative, is it continuous or discrete?

i. the individual's reported phone satisfaction percentage.

ii. my niece's appraisal of the interviewee's negativity.

iii. whether or not the interviewee would prefer to have the other phone.

iv. the type of phone the interviewee currently owns.

5. The strength of an internet connection is often described in terms of its download speed, measured in megabits per second (or Mbps). A systems administrator is concerned that recent changes in her company's main server framework may be having a negative impact on the local network's download speed. Every 2 minutes for an hour, she recorded the network speed at that moment and collected her results into the following stem-and-leaf plot:

The decimal point is at the |

```

0 | 9
1 | 8
2 | 7
3 | 6
4 | 134
5 | 7
6 | 1145677
7 | 01338
8 | 2346
9 | 79
10 | 45
11 |
12 | 17

```

Note that 0 | 9 represents 0.9. In this case, the first quartile is $Q(.25) = 5.7$, the median is 6.85, and the IQR is 2.7.

- (a) (10 points) Complete the following frequency table:

Value Range	Frequency	Relative Frequency	Cumulative Relative Frequency
0.00 - 2.00			
2.01 - 4.00			
4.01 - 6.00			
6.01 - 8.00			
8.01 - 10.00			
10.01 - 12.00			
12.01 - 14.00			

(b) (10 points) Create a box plot to summarize the data. Carefully label the axes.

(c) (4 points) Are there any unusually low observations? If so What were the speeds at those points?

(d) (10 points) She also measured upload speed, obtaining the following 8 values.

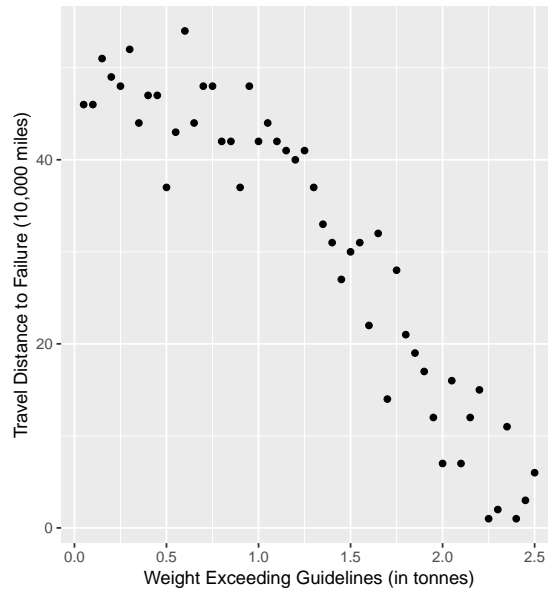
7.45, 4.22, 7.7, 6.04, 7.68, 5.71, 4.71, 8.44

Create a theoretical Q-Q plot using the following quantiles from the normal distribution as the theoretical quantiles. Carefully label your axes. What does this graph tell us about the upload speeds?

	1	2	3	4	5	6	7	8
p	0.0625	0.1875	0.3125	0.4375	0.5625	0.6875	0.8125	0.9375
$Q(p)$	-1.53	-0.89	-0.49	-0.16	0.16	0.49	0.89	1.53

6. The major cause of axel failure in freight trucks is when shippers exceed the recommended weight limits that can be handled by the axels. Issues resulting from these failures have been becoming more frequent as shippers try to cut corners, leading members of the state's Department of Transportation to ask one of their civil engineers to look into the available data and better advise them on the relationship between excessive weight and axel failure.

A company manufacturing axels provides the engineer with data gathered from conducting experiments loading axels with excessive weight and simulating traveling conditions. The data consists of two columns, **excessive weight (in tonnes)** is the amount of weight over the limit that was placed on the axel, and **distance to failure (in tens of thousands of miles)** is the simulated distance to the axel's failure.



Here are some summaries of the data:

$$\sum_{i=1}^{50} x_i = 64$$

$$\sum_{i=1}^{50} x_i^2 = 107$$

$$\sum_{i=1}^{50} y_i = 1558$$

$$\sum_{i=1}^{50} y_i^2 = 61528$$

$$\sum_{i=1}^{50} x_i y_i = 1444$$

- (a) Using the summaries above, fit a linear relationship between **weight exceeding guidelines** (x) and **travel distance to failure** (y).
- (5 points) Write the equation of the fitted linear relationship.

- ii. (5 points) Find and interpret the value of R^2 for the fitted linear relationship.
- iii. (5 points) Using the fitted line, provide a predicted value of travel distance to failure when the weight exceeding the guidelines is 3.4 tonnes.

- (b) The JMP output below comes from fitting a quadratic model using x and x^2 .

Response Distance to Failure					
Summary of Fit					
RSquare					REDACTED
RSquare Adj					REDACTED
Root Mean Square Error					5.281589
Mean of Response					0.16
Observations (or Sum Wgts)					50
Analysis of Variance					
		Sum of			
Source	DF	Squares	Mean Square	F Ratio	
Model	2	13229.647	6614.82	237.1314	
Error	47	1311.073	27.90		Prob > F
C. Total	49	14540.720			<.0001*
Parameter Estimates					
Term		Estimate	Std Error	t Ratio	Prob> t
Intercept		16.27602	2.333507	6.97	<.0001*
Weight Exceeding Limit		4.6604349	4.221593	1.10	0.2752
(Weight Exceeding Limit)^2		-10.2775	1.604983	-6.40	<.0001*

- i. (5 points) Write the equation of the fitted quadratic relationship.
- ii. (5 points) Find and interpret the value of R^2 for the fitted quadratic relationship.
- iii. (5 points) Using the fitted quadratic relationship, provide a predicted value of travel distance to failure when the weight exceeding the guidelines is 3.4 tonnes.