# Exam I

## STAT 105, Section B
## FALL 2015

### Instructions

- The exam is scheduled for 80 minutes, from 8:00 to 9:20 AM. At 9:20 AM the exam will end.
- A forumula 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**).

**Solution:** 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

**Solution:**

$$\begin{aligned}\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i &= \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) \\ &= \frac{1}{5}(12 + 15 + 18 + 19 + 26) \\ &= \frac{1}{5}(90) \\ &= 18\end{aligned}$$

   (b) (2 points) the median

**Solution:** The median is $Q(0.5)$.

Since $np + 0.5 = 5 \cdot 0.5 + 0.5 = 3$ then $Q(.5) = x_3 = 18$ is the median.

Alternatively, since $n = 5$ is odd, we know that the median is the middle of the ordered values.

   (c) (2 points) the variance

**Solution:** Since this is a sample, we must $s^2$:

$$
\begin{aligned}
s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= \frac{1}{5-1}\sum_{i=1}^{5}(x_i-\bar{x})^2 \\
&= \frac{1}{4}\left((x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2\right) \\
&= \frac{1}{4}\left((12-18)^2+(15-18)^2+(18-18)^2+(19-18)^2+(26-18)^2\right) \\
&= \frac{1}{4}\left((-6)^2+(-3)^2+(0)^2+(1)^2+(8)^2\right) \\
&= \frac{1}{4}\left(36+9+0+1+64\right) \\
&= \frac{1}{4}\left(110\right) \\
&= 27.5
\end{aligned}
$$

(d) (2 points) the standard deviation

**Solution:** We must use the sample standard deviation, $s$:

$$s = \sqrt{s^2} = \sqrt{27.5} = 5.2440442$$

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

**Solution:** We will need to use the quantile function.

In this case, $i = \lfloor np + 0.5 \rfloor = \lfloor 5 \cdot 0.25 + 0.5 \rfloor = \lfloor 1.75 \rfloor = 1$. Since $i \neq np + 0.5$, we must use the second form of the function:

$$
\begin{aligned}
Q(.25) &= x_i + (np - i + 0.5) \cdot (x_{i+1} - x_i) \\
&= x_1 + (5 \cdot .25 - 1 + 0.5) \cdot (x_2 - x_1) \\
&= 12 + (0.75) \cdot (15 - 12) \\
&= 12 + (0.75) \cdot (3) \\
&= 12 + 2.25 \\
&= 14.25
\end{aligned}
$$

(f) (2 points) the Interquartile Range

**Solution:** Since IQR $= Q(.75) - Q(.25)$, we will need to use the quantile function to find $Q(.75)$.

In this case, $i = \lfloor np + 0.5 \rfloor = \lfloor 5 \cdot 0.75 + 0.5 \rfloor = \lfloor 4.25 \rfloor = 4$. Since $i \neq np + 0.5$, we must use the second form of the function:

$$\begin{aligned} Q(.75) &= x_i + (np - i + 0.5) \cdot (x_{i+1} - x_i) \\ &= x_4 + (5 \cdot .75 - 1 + 0.5) \cdot (x_5 - x_4) \\ &= 19 + (3.25) \cdot (26 - 19) \\ &= 19 + (3.25) \cdot (7) \\ &= 19 + 22.75 \\ &= 41.75 \end{aligned}$$

So we get IQR $= Q(.75) - Q(.25) = 41.75 - 14.25 = 27.5$

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 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.

**Solution:** This is an experiment. The methods are being applied by the engineer and which method is used is decided by the engineer.

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

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

**Solution:** The change in concentration.

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

**Solution:** The method used.

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

**Solution:** The lake the sample was taken from is a treated as a block in this example.

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

**Solution:** Replication was only used in the sense that each method was used more than once. However, inside the blocks (the sample from the lakes) each treatment is only applied once.

4. Recently my teenage niece had an opportunity to upgrade her smart phone. She narrowed her choices down to two phones (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, 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?

**Solution:** This is an observational study. My niece takes no active role in change anything about the individuals she interviews or their phones.

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

**Solution:** There are two responses being gathered about each interviewee: (1) whether or not they would prefer the other phone and (2) how satisfied they are with their current phone. variables

   (c) 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.

**Solution:** This is quantitative and continuous.

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

**Solution:** This is qualitative. The levels are overly critical, appropriately critical, or not critical enough.

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

**Solution:** This is qualitative. The levels are "yes" and "no".

   iv. the type of phone the interviewee currently owns.

**Solution:** This is qualitative. The levels are "phone 1" and "phone 2".

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
```

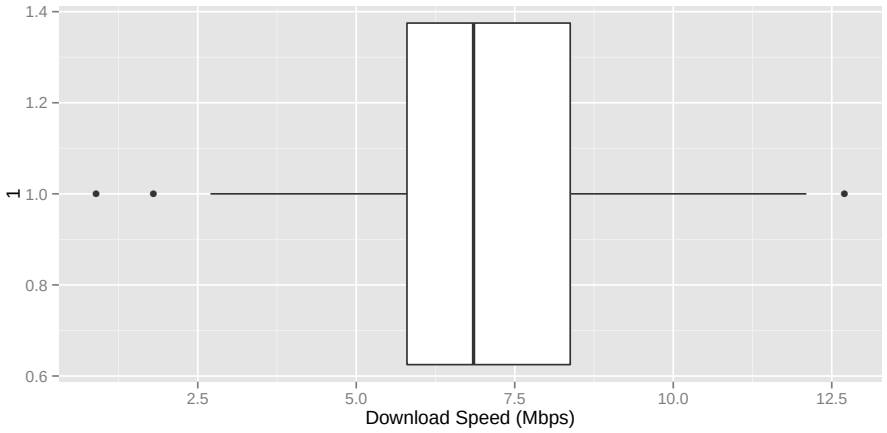
Note that `0 | 9` represents 0.9. In this case, the first quartile is $Q(.25) = 5.7$, the median is 6.85, and the IQR is 2.7.

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

| **Value Range** | **Frequency** | **Relative Frequency** | **Cumulative Relative Frequency** |
|---|---|---|---|
| 0.00 - 2.00 | 2 | 0.07 | 0.07 |
| 2.01 - 4.00 | 2 | 0.07 | 0.14 |
| 4.01 - 6.00 | 4 | 0.13 | 0.27 |
| 6.01 - 8.00 | 12 | 0.4 | 0.67 |
| 8.01 - 10.00 | 6 | 0.2 | 0.87 |
| 10.01 - 12.00 | 2 | 0.07 | 0.94 |
| 12.01 - 14.00 | 2 | 0.07 | 1.01 |

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

**Solution:** The boxplot is below:




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

**Solution:** Yes, there are two unusually low observations as indicated by the box plot. They are 0.9 and 1.8.

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

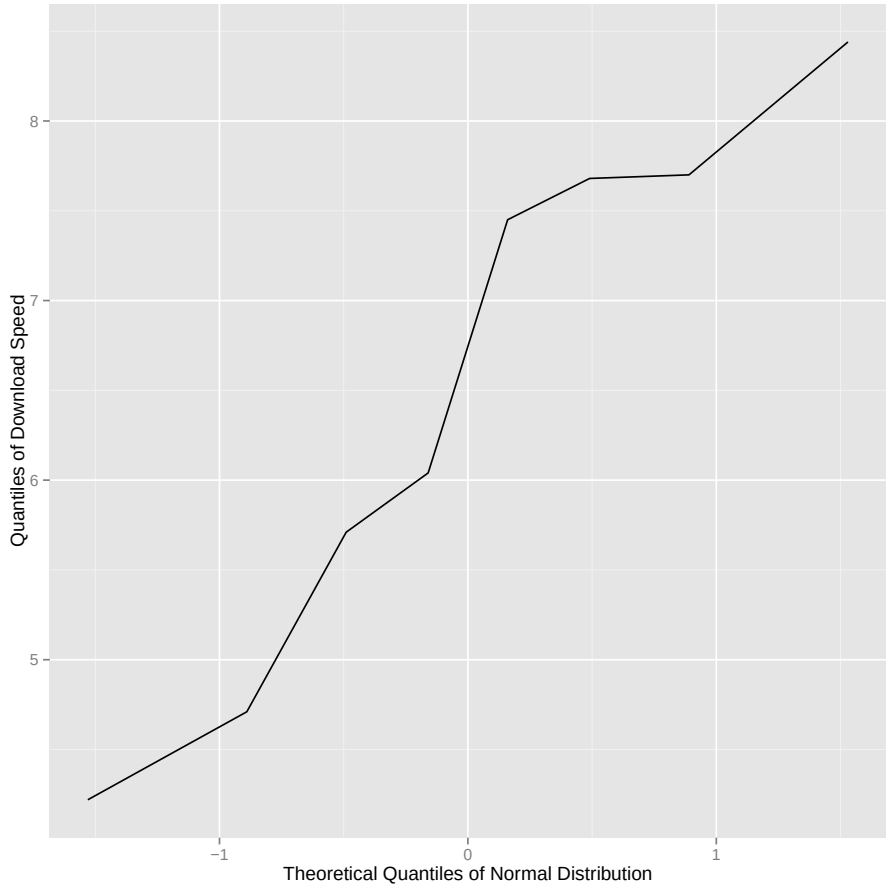
7.45, 4.22, 7.7, 6.04, 7.68, 5.71, 4.71, 8.44

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 |

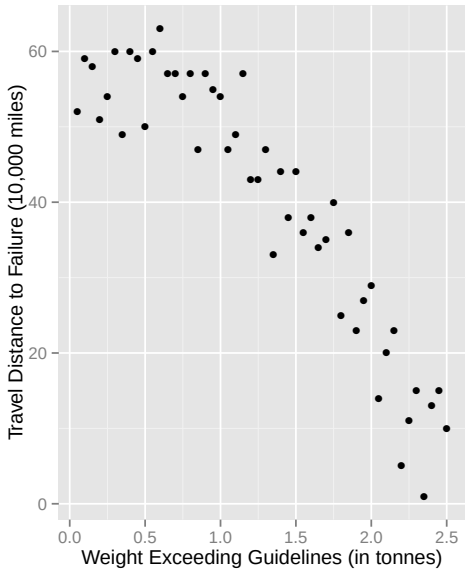
**Solution:** We get the QQ-plot by plotting the ordered values of our sample against the ordered quantiles from the normal distribution (as given above):




The points do seem to somewhat linear - an argument could be made that because of this the upload speed is normally distributed.

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:

$$\sum_{i=1}^{50} x_i = 64 \qquad \sum_{i=1}^{50} x_i^2 = 107$$

$$\sum_{i=1}^{50} y_i = 2008 \qquad \sum_{i=1}^{50} y_i^2 = 95182$$

$$\sum_{i=1}^{50} x_i y_i = 1999$$

   (a) Using the summaries above, fit a linear relationship between **weight exceeding guidelines** (x) and **travel distance to failure** (y).

      i. (5 points) Write the equation of the fitted linear relationship.

**Solution:** We need to get the estimates of $b_0$ and $b_1$ to write the equation of the fitted line:

$$\begin{aligned} b_1 &= \frac{\sum_{i=1}^{50} x_i y_i - 50\bar{x}\bar{y}}{\sum_{i=1}^{50} x_i^2 - 50\bar{x}^2} \\ &= \frac{1999 - 50\left(\frac{64}{50}\right)\left(\frac{2008}{50}\right)}{107 - 50\left(\frac{64}{50}\right)^2} \\ &= \frac{1999 - 50(1.27)(40.16)}{107 - 50(1.27)^2} \\ &= \frac{1999 - 2550.16}{107 - 80.645} \\ &= \frac{-551.16}{26.355} \\ &= -20.91 \end{aligned}$$

and thus

$$\begin{aligned} b_0 &= \bar{y} - b_1\bar{x} \\ &= \left(\frac{2008}{50}\right) - (-20.91)\left(\frac{64}{50}\right) \\ &= (40.16) - (-20.91)(1.28) \\ &= 66.9248 \end{aligned}$$

So the equation of the fitted line is:

$$\hat{y} = 66.9248 - 20.91x$$

ii. (5 points) Find and interpret the value of $R^2$ for the fitted linear relationship.

**Solution:** Since this is a linear model, we can first find the correlation coeffecient and then find the value of $R^2$:

$$\begin{aligned} r &= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right)\left(y_i - n\bar{y}^2\right)}} \\ &= \frac{1999 - 50(1.27)(40.16)}{\sqrt{(107 - (50)(1.27)^2)\,(95182 - (50)(40.16)^2)}} \\ &= \frac{-551.16}{\sqrt{(25.5)\,(14540.5)}} \\ &= \frac{-551.16}{608.9193296} \\ &= -0.91 \end{aligned}$$

So $R^2 = (-0.91)^2 = 0.8281$ meaning that 82.8% of the variability in failure time can be explained by the linear relationship of failure time with weight.

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.

**Solution:**

$$\begin{aligned}\hat{y} &= 66.9248 - 20.91(3.4)\\ &= -4.1692\end{aligned}$$

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

**Response Distance to Failure**

**Summary of Fit**

| | |
|---|---|
| RSquare | 0.909834 |
| RSquare Adj | 0.905998 |
| Root Mean Square Error | 5.281589 |
| Mean of Response | 0.16 |
| Observations (or Sum Wgts) | 50 |

**Analysis of Variance**

| Source | DF | Sum of 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.

**Solution:** From the JMP output:

$$\hat{y} = 16.27 + 4.66x - 10.2x^2$$

ii. (5 points) Find and interpret the value of $R^2$ for the fitted quadratic relationship.

**Solution:**

$$R^2 = \frac{\text{SSTO} - \text{SSE}}{\text{SSTO}} = \frac{13229.647 - 1311.073}{13229.647} = 0.9008989$$

Using the quadratic relationship, approximately 90.09% of the variability in failure time can be explained.

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.

**Solution:**

$$\begin{aligned}\hat{y} &= 16.27 + 4.66x - 10.2x^2\\ &= 16.27 + 4.66(3.4) - 10.2(3.4)^2\\ &= -85.798\end{aligned}$$