STAT 350 Homework #1

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**1. Suppose that statistics professor records the following each student enrolled in the class. Classify each of these variables as one of the following four types;**

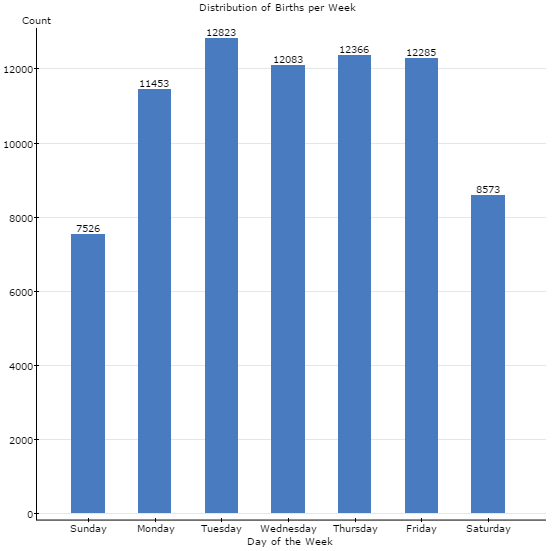
**Qualitative and Nominal, Qualitative and Ordinal, Quantitative and Discrete or Quantitative and Continuous. (1 point each)**

|  |  |
| --- | --- |
| **Variable** | **Type** |
| Gender | **Qualitative and Nominal** |
| Major | **Qualitative and Nominal** |
| Number of quizzes taken  (a measure of class attendance) | **Quantitative and Discrete** |
| Time spent sleeping the previous night | **Quantitative and Continuous** |
| Handedness (left- or right-handed) | **Qualitative and Nominal or Ordinal** |
| Height of a desk | **Quantitative and Continuous** |
| Final Letter Grade in a Course | **Qualitative and Ordinal** |

2. Births are not, as you might think, evenly distributed across the days of the week. Here are the average numbers of babies born on each day of the week in 2002.

|  |  |
| --- | --- |
| Day | Births |
| Sunday | 7,526 |
| Monday | 11,453 |
| Tuesday | 12,823 |
| Wednesday | 12,083 |
| Thursday | 12,366 |
| Friday | 12,285 |
| Saturday | 8,573 |

1. Use StatCrunch to construct a well-labeled bar chart using the data from the table above. (5 points)

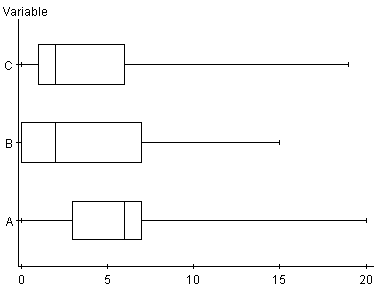


1. Suggest some reasons why there are fewer births on weekends. (2 points)

Reasons can vary but may include suggestions such as: doctors schedule emergency births during the week when the hospital is better staffed.

3. Below are the 5-number summaries **and** the plots that describe the length of a stay following total knee replacement surgery for patients in three different hospitals (A, B and C). (2 points each)

Hospital A: 0, 3, 6, 7, 20 Hospital B: 0, 0, 2, 7,15 Hospital C: 0,1, 2, 6, 19 days



**Days in Hospital**

a) 25% of all stays in Hospital C are longer than this many days?

6 days

b) Which hospital has the least variability in the middle 50% of the stays? Explain.

IQR Hospital A = 4 days

IQR Hospital B = 7 days

IQR Hospital C = 5 days

c) Does hospital A contain any outliers in their data? **Verify your answer by using the outlier test.**

IQR \* 1.5 = 4\*1.5 = 6

Q1-6 = -3 so no low outliers but

Q3+6 = 13 so yes, the maximum value of 20 days is an outlier as it is greater than 13 days

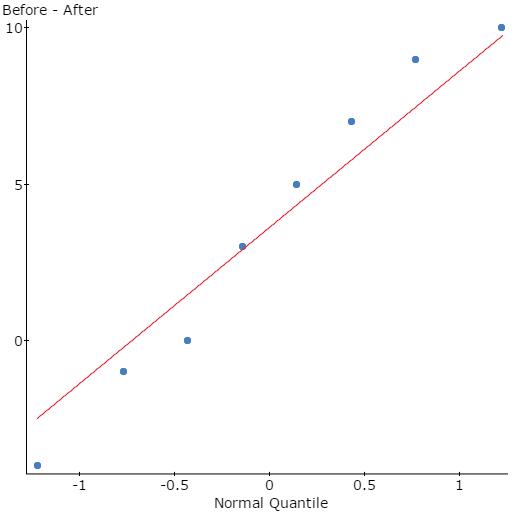
**Extended response questions #4-#7 – 20 points each**

4. Eight randomly selected male students went on a new diet program in an attempt to lose weight. The results are listed below.

|  |  |  |  |
| --- | --- | --- | --- |
| Subject | Weight Before  (in lbs) | Weight After  (in lbs) | Difference  Before-After |
| 1: Abdul | 174 | 165 | 9 |
| 2: Ed | 191 | 195 | -4 |
| 3: Jim | 188 | 183 | 5 |
| 4: Max | 182 | 179 | 3 |
| 5: Phil | 201 | 201 | 0 |
| 6: Ray | 188 | 181 | 7 |
| 7: David | 195 | 185 | 10 |
| 8. Mike | 170 | 171 | -1 |

a) Complete the chart above. (1 point)

b) Construct a QQ plot (as called a normal probability plot) for the difference between weight before and after the diet program. Use this plot to comment if it is appropriate to analyze the mean difference in weight using the Normal model. (4 points)



As the values fall reasonably close to a straight line we can state that the weight loss difference is approximately normally distributed.

c) State any other assumptions/conditions that must be satisfied before constructing a 95% confidence **interval for the population mean difference in weight (before-after).**

**(4 points)**

**Other assumptions: The data are paired as we are using before and after weight loss on eight independent. The sample was randomly selected men.**

d) Calculate the mean, standard deviation, standard error, df, t\* value and margin of error to construct a 95% confidence interval for the population mean change in weight. (7 points)

 3.625 , sd = 5.0125 , = 1.7722,

df = 7 , t\* = 2.365 , margin of error = 4.1913

Confidence Interval (-0.5663, 7.8163)

e) Interpret your findings using the confidence interval in part (d). Would you recommend this diet program? (4 points)

At the 95% level of confidence it is unclear if the diet is successful as the confidence interval includes zero. This indicates that some people on this diet could gain weight and other lose weight.

5. Textbook page 427 #9.26. Use the data set provided to construct and interpret the appropriate statistical hypothesis test **using StatCrunch**. Remember to check that all assumptions for inference are met.

a) State the null and alternative hypothesis and state the parameter of interest. (3 pts)

b) Check conditions for inference have been satisfied (3 pts)

c) Compute summary statistics needed to calculate the test statistic (3 pts)

d) Determine the p-value (2 pts)

e) Make a decision based on the p-value (3 pts)

f) Interpret your decision in the context of the question. (3 pts)

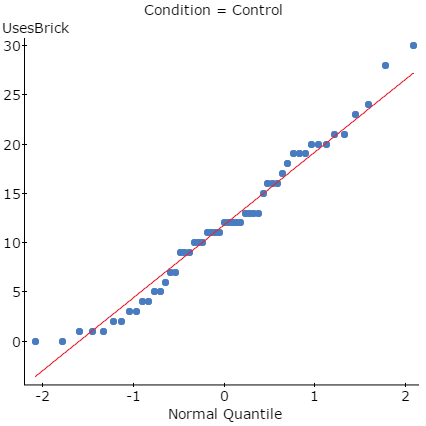
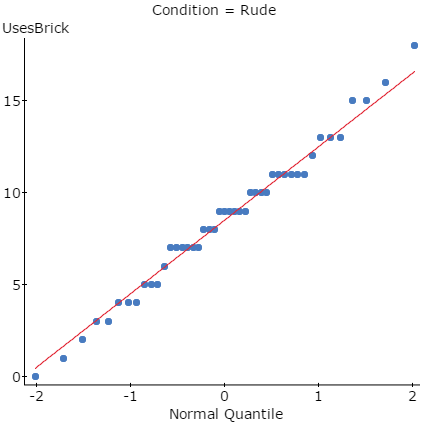
g) Could you have made a Type I or Type II error? Give a consequence of this potential

error. (3 pts)

μ1 : Mean of Uses for Brick where Condition=Control  
μ2 : Mean of Uses for Brick where Condition=Rude  
μ1 - μ2 : Mean difference between two means  
H0 : μ1 - μ2 = 0  
HA : μ1 - μ2 > 0

The groups are independent and both randomly selected and

BOTH samples comes from a normal populations as seen in the QQ plots

(without pooled variances)

| **Difference** | **Sample Diff.** | **Std. Err.** | **DF** | **T-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| μ1 - μ2 | 3.3002096 | 1.1758005 | 82.43068 | 2.8067769 | 0.0031 |

As the p-value of 0.0031 is less than the significance level of 0.01 we will reject the null hypothesis and conclude that there is evidence to suggest that the true mean performance level for students in the rudeness condition is lower than the true mean performance level for students in the control group.

We could have made a Type I error as we rejected the null hypothesis. This would suggests that while we stated that rudeness lower mean performance levels for students in fact did not. A consequence is thinking that by not being rude you can increase mean performance when it does not – however it is hard to find a serious problem with telling people to be nicer. ☺

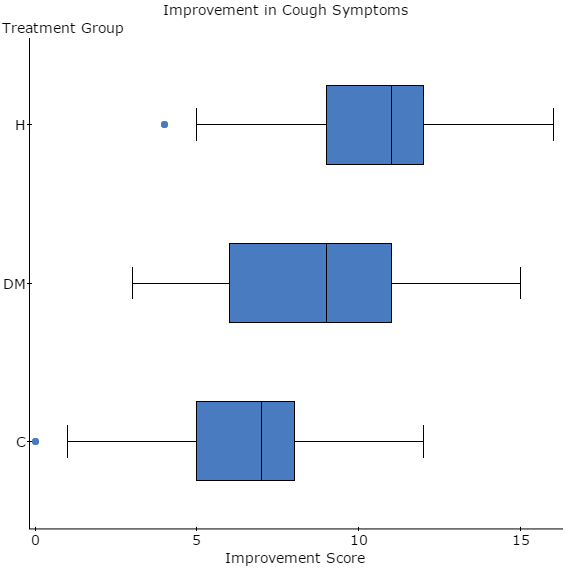
6 & 7. Textbook page 45 #2.32. Does a teaspoon of honey before bed really calm a child’s cough? To test the folk remedy, pediatric researchers at Penn State University carried out a designed study involving a sample of 105 children who were ill with an upper respiratory tract infection. One the first night, parent’s rated their children’s cough on a scale from 0 (no problems at all) to (30 extremely severe). On the second night, the parents were instructed to give this sick child a dosage of liquid “medicine” prior to bedtime. Unknown to the parents, some were given a dosage of an over the counter cough medicine (DM) while others were given a similar does of honey. Also a third group gave their sick children no dosage at all. Again the parent’s rated their child’s cough symptoms and the improvement in total cough symptoms score was determined for each child.

Data is shown below and is also available in Stat Crunch.

a) What type of data is being collected? **Qualitative and Nominal, Qualitative and Ordinal, Quantitative and Discrete or Quantitative and Continuous? (2 pts)**

**Quantitative and Discrete**

**b) Construct a well-labeled boxplot for each of the three treatment groups improvement scores on one graph. (8 pts)**



**c) Use StatCrunch to obtain the following descriptive statistics to compare the three treatment groups. (6 pts)**

**Summary statistics for Improve Score:**  
Group by: Treatment

| **Treatment** | **Mean** | **Median** | **Std. dev.** | **n** |
| --- | --- | --- | --- | --- |
| Control | 6.5135135 | 7 | 2.9403479 | 37 |
| DM | 8.3333333 | 9 | 3.2564039 | 33 |
| Honey | 10.714286 | 11 | 2.8550412 | 35 |

d) Do you think there might be a problem with using this particular control group? (4pts)

Answers could vary but could include: It might have been a better idea to give the parents a placebo medication rather than no medicine since a parent might find it difficult to no help their child if their coughing symptoms are severe.

7. Use the data from the previous exercise and textbook page 458 #9.103. The researchers want to know if the variability in coughing improvement scores differs for the DM and Honey groups. Conduct the appropriate hypothesis test using =0.05.

a) State the null and alternative hypothesis and state the parameter of interest. (4pts)

σ12 : Variance of DM  
σ22 : Variance of Honey  
σ12/σ22 : Ratio of two variances  
H0 : σ12/σ22 = 1  
HA : σ12/σ22 ≠ 1

b) Check conditions for inference have been satisfied (3 pts)

Both samples are random and independent

Both populations are normally distributed

c) Compute summary statistics needed to calculate the test statistic (3 pts)

| **Treatment** | **Variance** | **n** |
| --- | --- | --- |
| DM | 10.604167 | 33 |
| Honey | 8.1512605 | 35 |

d) Determine the p-value (3 pts)

| **Ratio** | **Num. DF** | **Den. DF** | **Sample Ratio** | **F-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| σ12/σ22 | 32 | 34 | 1.3009235 | 1.3009235 | 0.4514 |

e) Make a decision based on the p-value (3 pts)

Fail to Reject the null hypothesis as the p-value is large, greater than 0.05.

f) Interpret your decision in the context of the question. (4 pts)

There is no evidence to suggest that there is any variability in coughing improvement scores between the honey and DM group.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Honey DM Control

12 4 5

11 6 8

15 9 6

11 4 1

10 7 0

13 7 8

10 7 12

4 9 8

15 12 7

16 10 7

9 11 1

14 6 6

10 3 7

6 4 7

10 9 12

8 12 7

11 7 9

12 6 7

12 8 9

8 12 5

12 12 11

9 4 9

11 12 5

15 13 6

10 7 8

15 10 8

9 13 6

13 9 7

8 4 10

12 4 9

10 10 4

8 15 8

9 9 7

5 "" 3

12 "" 1

"" "" 4

"" "" 3

8. Textbook page 459 #9.105. Use the data set provided to construct and interpret the appropriate statistical hypothesis **test using StatCrunch**. Remember to show all steps and check that all assumptions for inference are met.

a) State the null and alternative hypothesis and state the parameter of interest. (4pts)

b) Check conditions for inference have been satisfied (3 pts)

c) Compute summary statistics needed to calculate the test statistic (3 pts)

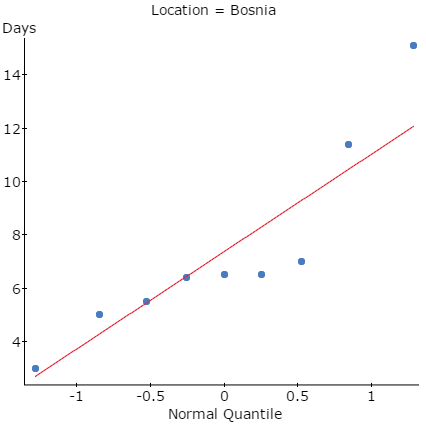
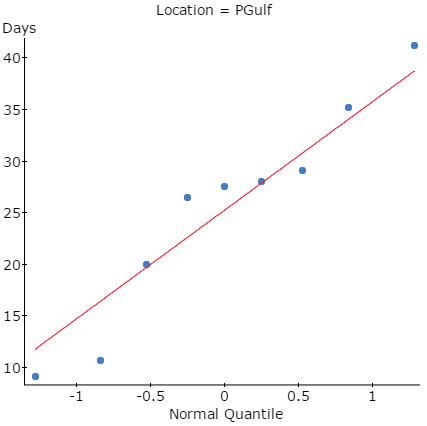
d) Determine the p-value (3 pts)

e) Make a decision based on the p-value (3 pts)

f) Interpret your decision in the context of the question. (4 pts)

a) Ho: and Ha: ≠

b) Samples must be independent and randomly selected both from normally distributed populations.

Note: the QQ plot of Bosnia is a little concerning!

**c) Summary statistics for Days:**   
Group by: Location

| **Location** | **n** | **Variance** |
| --- | --- | --- |
| Bosnia | 9 | 13.349444 |
| PGulf | 9 | 110.67778 |

**d) Hypothesis test results:**  
σ12 : Variance of Days where Location = PGulf  
σ22 : Variance of Days where Location = Bosnia  
σ12/σ22 : Ratio of two variances  
H0 : σ12/σ22 = 1  
HA : σ12/σ22 ≠ 1

| **Ratio** | **Num. DF** | **Den. DF** | **Sample Ratio** | **F-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| σ12/σ22 | 8 | 8 | 8.2908153 | 8.2908153 | 0.0072 |

e) As the p-value < significance level, we reject the null hypothesis.

f) There is evidence of a difference in the shipment variances between these two locations.