STAT 350 Homework #1

Name:\_\_\_Tuo Wang\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**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** |
| 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)

(Title): Distribution of Births per Week



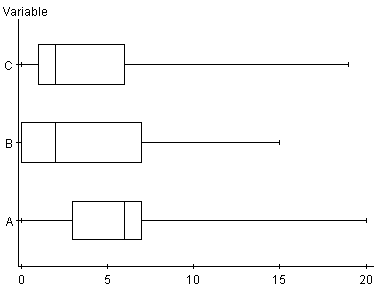
1. Suggest some reasons why there are fewer births on weekends. (2 points)

**Induced births and C-sections**

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

Box plot are great for comparing multiple data set



**Days in Hospital**

a) 25% of all stays in Hospital C are longer than this many days? **1 day / 6 days**

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

**Hospital A, the box length indicate the middle 50% of data’s variability or central tendency. Box plot of Hospital A has the shortest length of Box, so the least variability.**

**Also,**

**IQR for Hospital A: 4 ; IQR for Hospital B: 7; IQR for Hospital C: 5**

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

**Find the IQR and \* 1.5 ;**

**(Q1-1.5IQR, Q3+1.5IQR),**

**anything outside the range is considered a outlier.**

**Yes, even there is no outlier indicated beyond the whisker; IQR for Hospital A is 4**

**4\*1.5=6; Q3 + 6 =13 ; 20 is the outlier**

**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(lbs) | Weight After(lbs) | Difference (b4.after) |
| 1: Abdul | 174 | 165 | **9** |
| 2: Ed | 191 | 195 | **4 -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 -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)



**Yes, the data line follow the trend line closely, we accepted that this data set is 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)**

**The researcher selected simple random sample of the weight difference of men participant after the diet program.**

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)

**\_4.875/\_3.625\_**\_\_, sd = \_\_**3.603/**\_5.0125\_\_, = \_\_\_**3.63/sqrt 8= 1.283** /1.7722



df = \_\_ n-1 = **7\_**\_\_\_, t\* = **95% Confidence interval, a = 5% t = a/2 = 2.5% = 2.365**

margin of error = \_\_\_\_\_\_**1.772\_**\_\_\_4.1913\_\_\_\_\_



Confidence Interval : **4.875+/- 2.365\*1.283 =4.875+/-3.034 = (1.841, 7.909) /** (-0.5663, 7.8163)

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

**By using this diet program, there is 95% chance that people will lose weight from 1.841 lbs to 7.909 lbs , I would suggest that people try this program because they start to lose weight after the program.**

At confidence level of 95%, it is not clear whether the this diet program can help people lose weight or gain weight

4. 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)

**μc=average work performance of control group**

**μrude=average work performance of rudeness group**

**μc-μrude=the mean difference of**

**Ho=rudeness doesn't matter in work place ; μc=μrude**

**Ha=rudeness matter in work place ; μc>μrude**

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

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

**both samples are simple random samples**

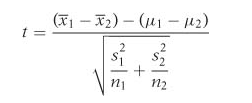
**both samples are normally distributed**

**two samples are independent form each other**

The groups are independent and both randomly selected and

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

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

**= 2.806**

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

d) Determine the p-value (2 pts)

**p=0.0037(~0.004)**

| **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)

**We reject the null because p is so small**

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.

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

**There is a relations between rudeness and work performance**

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

error. (3 pts)

**Type 1 Error, Reject the null but the null is correct**

**Type II error, Fail to reject the null but the null is false**

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. ☺

5 &6. 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.

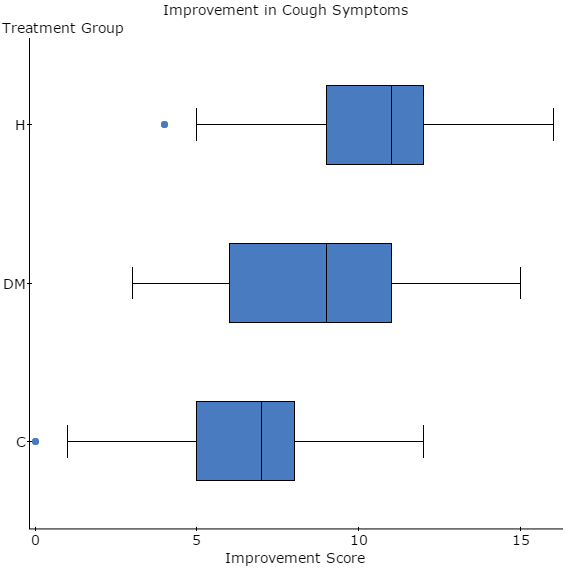
Data is shown below and is also available in Stat Crunch.

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

**Qualitative and Ordinal**

**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)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Treatment | Mean | Median | Standard Deviation | Sample Size |
| DM | 8.33 | 9 | 3.256 | 33 |
| Honey | 10.714 | 11 | 2.855 | 35 |
| Control | 6.513 | 7 | 2.94 | 37 |

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

**The sample size of each control group is different. We should make sure the sample size is equal so the sample size itself doesn't become a variable.**

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.

6. 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)

**Ho=the coughing improvement scores doesn't differ for DM and Honey ; σ2H=σ2DM**

**Ha=the coughing improvement scores differ for DM and Honey;**

**σ2H ≠ σ2DM**

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)

**The sample is simple random sample**

**The sample is normally distributed**

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

**sd DM – 3.26**

**sd Honey – 2.86**

**(sd DM)square/(sd Honey)square = 1.30=F**

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

d) Determine the p-value (3 pts)

**0.226**

| **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 because the p is small**

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 difference of the cough improvement from Honey and DM**

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

7. 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)

**Ho=there is no difference between delivery time**

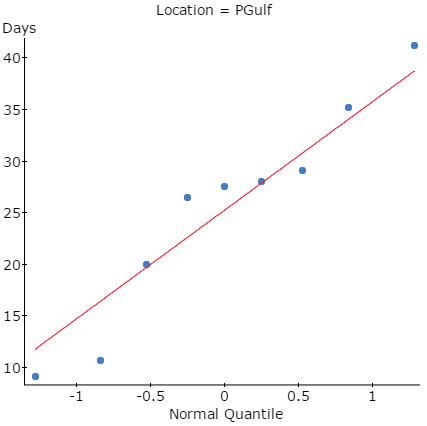
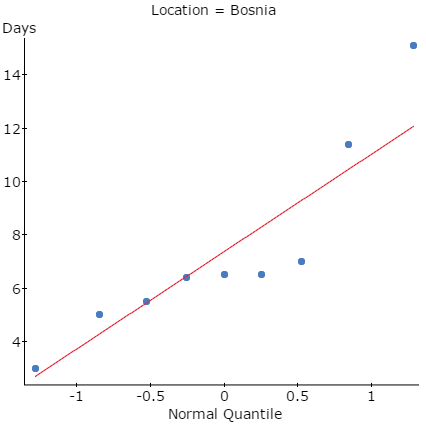
**Ha=there is a difference between delivery time**

Ho: and Ha: ≠

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

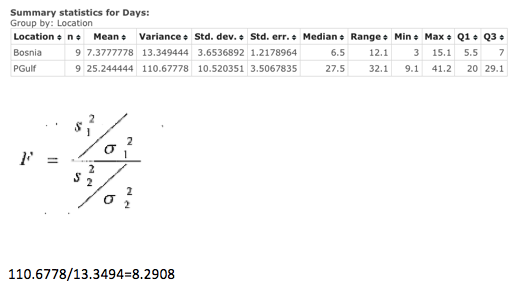
**The sample is simple random sample**

**The sample is normally distributed**



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

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



d) Determine the p-value (3 pts)

**P=0.0072**

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

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

**we reject the null the p value is greater than 0.05**

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

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

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

**so there is a difference between the delivery time**