Survey Project of University of Maryland Students

Raymond Ng, Ann Hoover

October 4, 2017

Objective:

The research question we wanted to answer was the following: Does mean time spent on certain types of weekend activities of UMD undergraduate students differ significantly between those in CMNS majors and those in other technical (or somewhat technical) majors?

Target Population:

Our target population was undergraduate UMD students with technical or somewhat technical majors. That is, we were not including any UMD students with liberal arts or fine arts majors in the target population because they do not fit into the categories described in our research question. As long as a student had at least one CMNS major, the student was classified as a CMNS major. Majors that are classified as being CMNS include Astronomy, Atmospheric and Oceanic Science, Biochemistry, Biological Sciences, Chemistry, Computer Science, Environmental Sciences, Geology, Mathematics, Physics, and Physical Sciences. All other students were put in the non-CMNS category. Any students with liberal or fine arts majors were disregarded. Undecided students were also not included in the sample or target population.

Sample and Administration:

Sampling Population

Our final dataset includes 30 undergraduate students at the University of Maryland, College Park. While we did receive a response from a student who was Undecided, we disregarded that response and sampled another student instead. Thus, our data only include students who are CMNS majors or in other non-CMNS, technical (or somewhat technical) majors.

Sampling Method: Getting Units and Administering Sample

Due to time and resource constraints, we conducted a convenience sample, but we first attempted to use a version of cluster sampling by selecting building on campus where we expected to find CMNS majors. The primary sampling units were buildings on campus, but instead of using an SRS of all the buildings on campus, we used our judgment to select buildings where we expected to find CMNS majors. During regular class-time hours on a weekday, these buildings were the Math building and the Armory. During Math Success hours, we also sampled students in Oakland Hall. Thus, the secondary sampling units were students who were not otherwise engaged in those buildings; that is, we selected students who were not in class or tutoring at the time. We printed out copies of the survey questions, approached students we found in the buildings, and then gave them a copy of the survey questions if they were willing to participate. We collected the completed surveys from them when they finished. However, we were not able to obtain 30 responses using this method, so we also contacted friends or acquaintances, whom we knew were not liberal arts majors, to complete our survey. These participants were much more likely to respond, but we would be concerned that they are not representative of our target population because of their biased selection. Furthermore, we would also be concerned that our selection of buildings and individuals in those buildings was also biased because we did not randomly select either.

Possible Problems

Since this is a convenience sample, we would be concerned that the people we were able to easily reach might differ from students that we were not able to reach. Specifically, our sampling population would likely exclude any students who happened to not be taking technical (or specifically math) classes this semester. Nonresponse from some individuals creates some concern that students who refused to respond to the survey might differ significantly from the people who were willing to respond. Also, it is unlikely that we obtained a representative sample because we did not use a probability sampling method.

Furthermore, some respondents incorrectly answered some of the questions. For example, some students left out their number of cumulatively earned credits or number of hours worked on average each week. In addition, a few respondents entered a total number of hours greater than 54 for weekend activities even though there were only 54 hours allotted in the designated time period. In this case, we disregarded these responses and surveyed other individuals to ensure that we had at least 30 usable responses. Some respondents also indicated that they had difficulty answering the last question about how many hours they spent on certain types of activities over a typical weekend period. Therefore, it's probable that we have measurement error in the number of hours spent on weekend activities since the answers are self-reported and not objectively measured.

We would also be concerned about measurement bias. For example, respondents might have a tendency to report the number of hours spent on academic activities as higher than the actual number of hours spent on academic activities. We might also expect students to underreport the number of hours spent on leisure activities because they don't want to be perceived as slackers. Some students might also be motivated to report a higher number of hours spent on social activities than is actually the case to avoid being perceived as a loner.

Sample with Unlimited Resources:

In our ideal sample, we would have been able to obtain list of all undergraduate students at UMD along with their corresponding majors and contact information. From that list, we would have classified all students with at least 1 CMNS major in the CMNS major category. Students with only liberal arts majors would be removed from the sampling frame. All remaining students would have been classified as having technical (or somewhat technical) majors, referred to as non-CMNS majors. From the sampling frame, we would have a better idea of how many students are in the CMNS major and non-CMNS major categories. Stratum 1 would consist of the CMNS majors; Stratum 2 would consist of the non-CMNS majors. We would use proportional allocation to determine the number of students we needed to survey from each stratum with the condition that the smaller sample size had to be at least 30. To determine which individuals to survey from the strata, we would use a random number generation to select an SRS of the appropriate sample size in each stratum. We would then contact each of the randomly selected students individually by email before the start of a randomly selected weekend to request their participation in our survey. The initial instructions would be for them to log the hours they spent on different types of activities so that they could accurately allocate their weekend hours. On Monday, we would remind them to send in their responses to the survey and their log. Since we have unlimited resources in this scenario, one way we could incentivize participation is by offering a gift card for participants who complete all parts of the survey.

Survey Questions:

1. List your major(s).

2. What year of college are you in?

   1 2 3 4 5 6+

3. What is your biological gender? Male / Female

4. Do you live on campus or off campus? On / Off

5. How many credits are you taking this semester? _____

6. How many classes are you taking this semester? _____

 For question 7, include college credits that were not earned at University of Maryland.

7. What is your cumulative number of earned college credits? _____

8. Do you work or volunteer in a professional setting? Yes / No

 Only answer question 9 if you answered "Yes" to question 8.

9. How many hours do you work each week, on average? _____

10. Do you participate in a club at the University of Maryland? Yes / No

11. Do you participate in a sports team at the University of Maryland? Yes / No

12. Are you currently in a romantic relationship? Yes / No

13. On a typical weekend (the 54-hour period starting from 6pm Friday and ending at 12am Monday), please provide your best estimate of the number of hours on average that you spend on the following activities.

 A _____ # Hours Spent on Academic Activities (studying, homework, etc.)

 B _____ # Hours Spent Working or Volunteering in a Professional Setting

 C _____ # Hours Spent Sleeping

 D _____ # Hours Spent on Social Activities

 E _____ # Hours Spent on Leisure Activities (activities you do for fun when alone)

 Please note that your total number of hours (A+B+C+D+E) should be a number between 0 and 54.

Analysis:

Before starting our analysis, we entered the survey responses in Excel. After entering all the responses exactly as they were listed on the survey, we coded categorical variables as numbers using Excel formulas. For example, CMNS Majors were coded as 1 for the CMNSMajor variable; the other respondents were coded as 0. Furthermore, responses of "Yes" for coded as 1 and "No" as 0. The gender variable was represented by the Male variable, which was coded as 1 for males and 0 for females. We then inputted the recoded spreadsheet into Stata; all Stata output is included in the final section of this document.
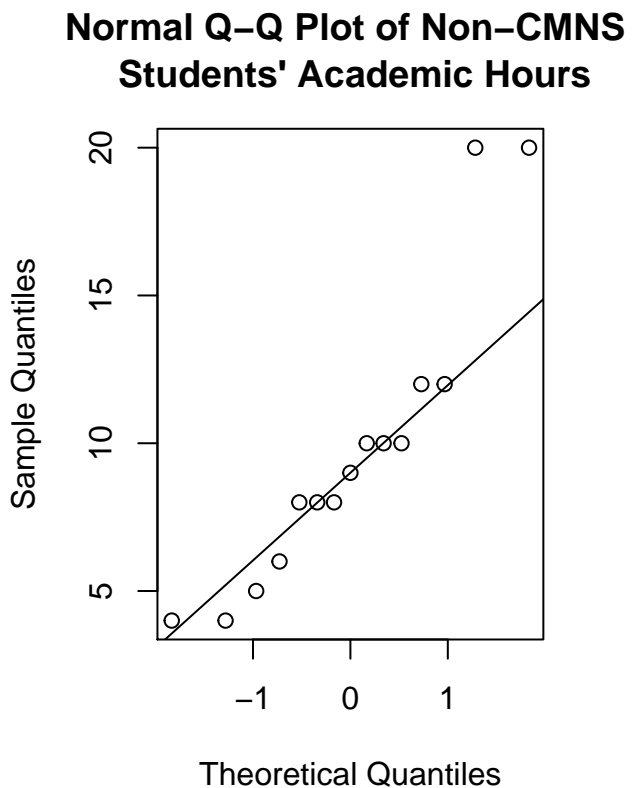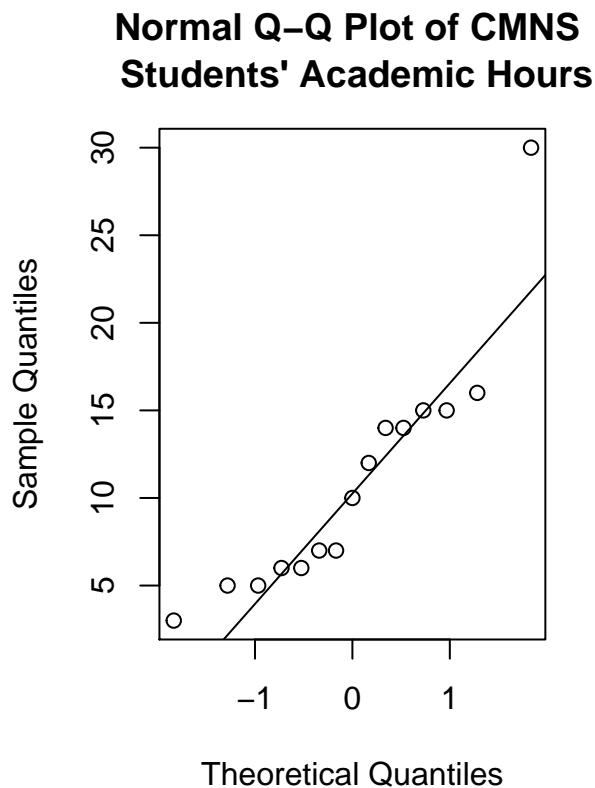
To answer our research question, we will conduct a difference of means test for the following types of activities during the weekend: Academic, Sleeping, Social, and Leisure. CMNS majors will be treated as one population while all other types of majors will be treated as the second population. The null hypothesis in all cases will be that the mean number of weekend hours allocated to one category is equal for CMNS majors and non-CMNS majors. The alternate hypothesis is that the mean number of weekend hours allocated to one category is not equal for CMNS majors and non-CMNS majors. The significance level for our tests will be 5%.
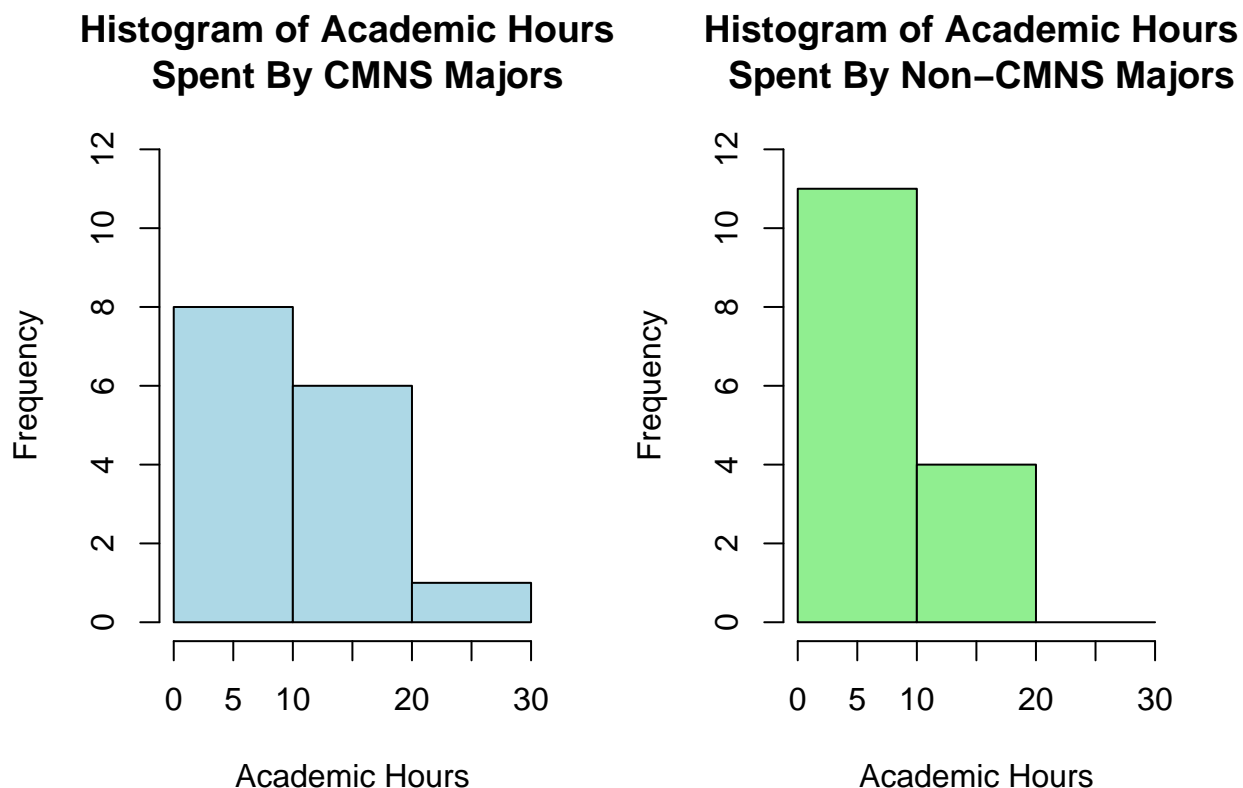
We will also try to develop regression equations to explain the allocation of hours for certain activities over the weekend. The dependent variable in these equations will be the number of weekend hours spent on a certain category of activities. The independent variables that we will include in the model are CMNS/non-CMNS major, gender, and number of current credits.

Academic Activities

The normal probability plot and histogram for weekend hours spent on academic activities indicates right skewness for both CMNS majors and non-CMNS majors. The distributions do not appear to be at all symmetric, so we should not use a two-sample t-test for this category. Instead, we conducted a nonparametric test: the two-sample Wilcoxon rank-sum test. This test does not require that the sampling distribution be approximately normal. The conditions for this test require that the observations from both groups are independent of each other and that the responses are ordinal; since we have no reason to believe our observations are dependent and the variable for the number of hours clearly has ordinal properties, we proceed with the analysis by using this test. The mean number of hours spent on academic activities was 11 for CMNS majors and 9.7 for non-CMNS majors. The p-value for this test is .7080. This indicates that the probability of obtaining a difference of means at least as large as 1.3 hours by chance is 70.80%. Thus, we fail to reject the null hypothesis that the mean number of weekend hours allocated to academic activities by CMNS majors and non-CMNS majors is the same.

In the regression model to predict number of weekend hours spent on academic activities, none of the coefficients on the chosen variables were statistically significant at the 5% significance level. The residual plot also does not show any pattern in the residuals that would indicate that we should use a transformation. Consequently, we cannot present a regression equation to explain the impact on number of weekend hours spent on academic activities of major type, gender, and number of current credits.

Normal Q–Q Plot of CMNS Students' Academic Hours

Normal Q–Q Plot of Non–CMNS Students' Academic Hours

Histogram of Academic Hours Spent By CMNS Majors

Histogram of Academic Hours Spent By Non–CMNS Majors
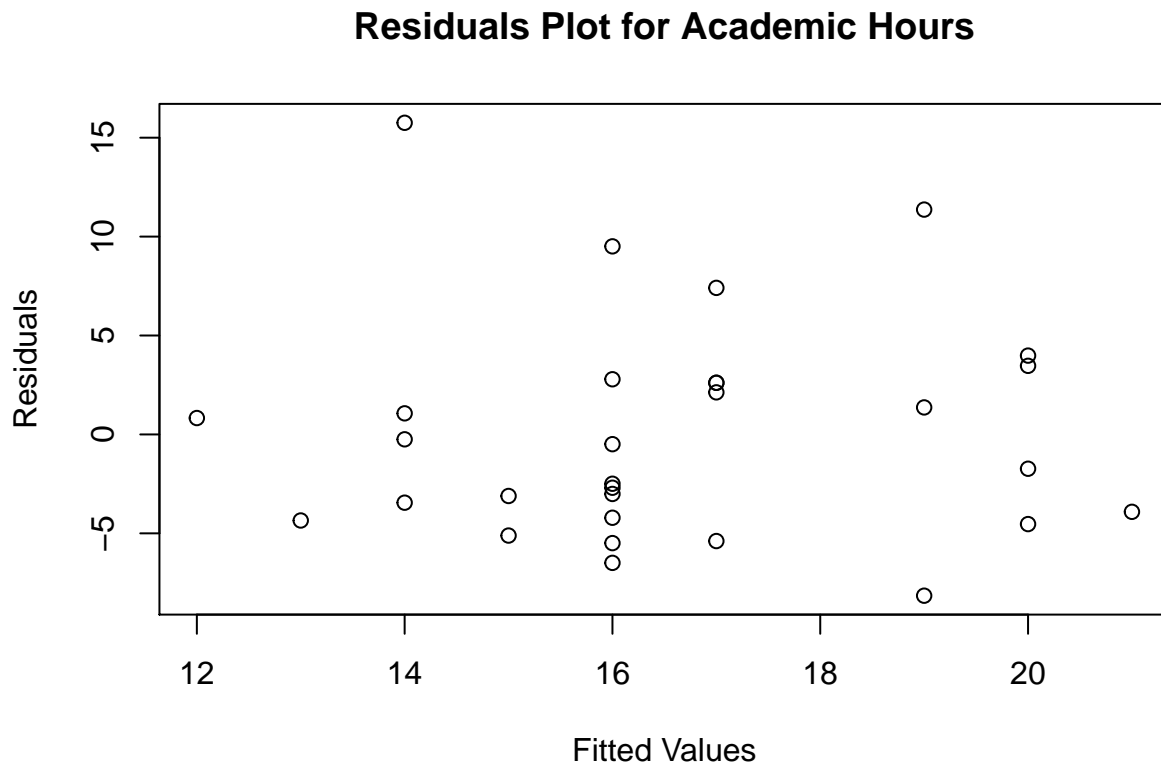
Two-Sample Wilcoxon Rank-Sum Test

```
##
##   Wilcoxon rank sum test
##
## data:  CMNS$`Academic Hours` and NonCMNS$`Academic Hours`
## W = 121.5, p-value = 0.708
## alternative hypothesis: true location shift is not equal to 0
```

Regression Model for Academic Hours Based on Major Type, Gender, and Current Credits

```
##
## Call:
## lm(formula = survey$`Academic Hours` ~ survey$"CMNS Major" +
##     survey$Male + survey$"Current Credits")
##
## Residuals:
##    Min     1Q Median    3Q    Max
## -8.151 -4.137 -1.114  2.610 15.753
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           16.6029     8.0251   2.069   0.0486 *
## survey$"CMNS Major"    1.8969     2.1762   0.872   0.3914
## survey$Male            3.1784     2.4666   1.289   0.2089
```

```
## survey$"Current Credits"  -0.6192     0.5177  -1.196    0.2425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.864 on 26 degrees of freedom
## Multiple R-squared:  0.1069, Adjusted R-squared:  0.003839
## F-statistic: 1.037 on 3 and 26 DF,  p-value: 0.3925
```

**Residuals Plot for Academic Hours**



**Residuals Plot for Academic Hours**

## Sleeping

The normal probability plot and histogram for weekend hours spent sleeping do not appear symmetric. As such, we perform the two-sample Wilcoxon rank-sum test. The mean number of ours spent sleeping was 21.2 for CMNS majors and 17.8 for non-CMNS majors. The p-value for this test is .1674. This indicates that the probability of obtaining a difference of means at least as large as 3.8 hours by chance is 16.74%. Therefore, we fail to reject the null hypothesis that the mean number of weekend hours allocated to sleeping by CMNS majors and non-CMNS majors is the same.

In the regression model to predict number of weekend hours spent sleeping, none of the coefficients on the chosen variables were statistically significant at the 5% significance level. The residual plot also does not show any pattern in the residuals that would indicate that we should use a transformation. Consequently, we cannot present a regression equation to explain the impact on number of weekend hours spent on sleeping of major type, gender, and number of current credits.

6

**Normal Q–Q Plot of CMNS Students' Sleeping Hours**
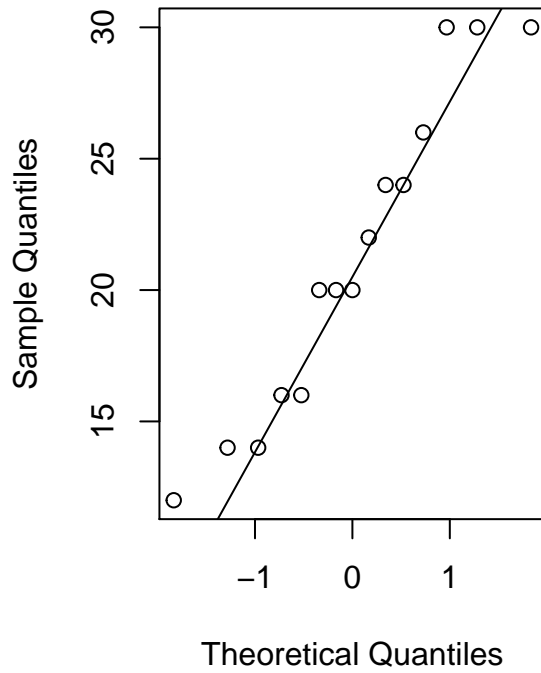
**Normal Q–Q Plot of Non–CMNS Students' Sleeping Hours**

## Histogram of Sleeping Hours Spent by CMNS Majors

## Histogram of Sleeping Hours Spent by Non−CMNS Majors
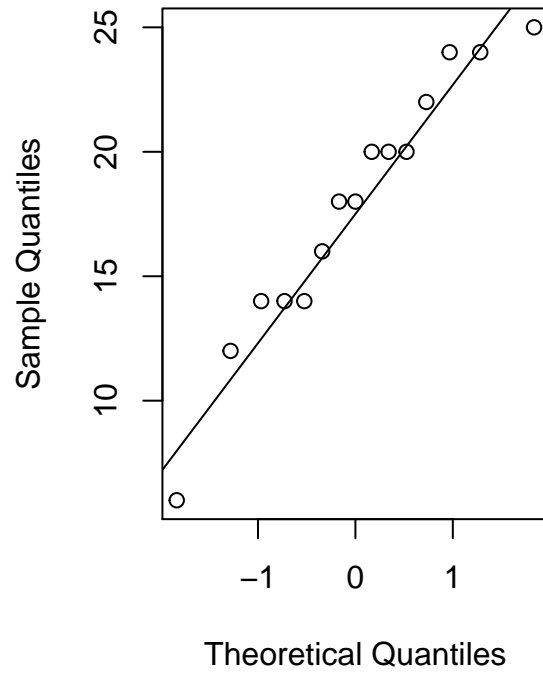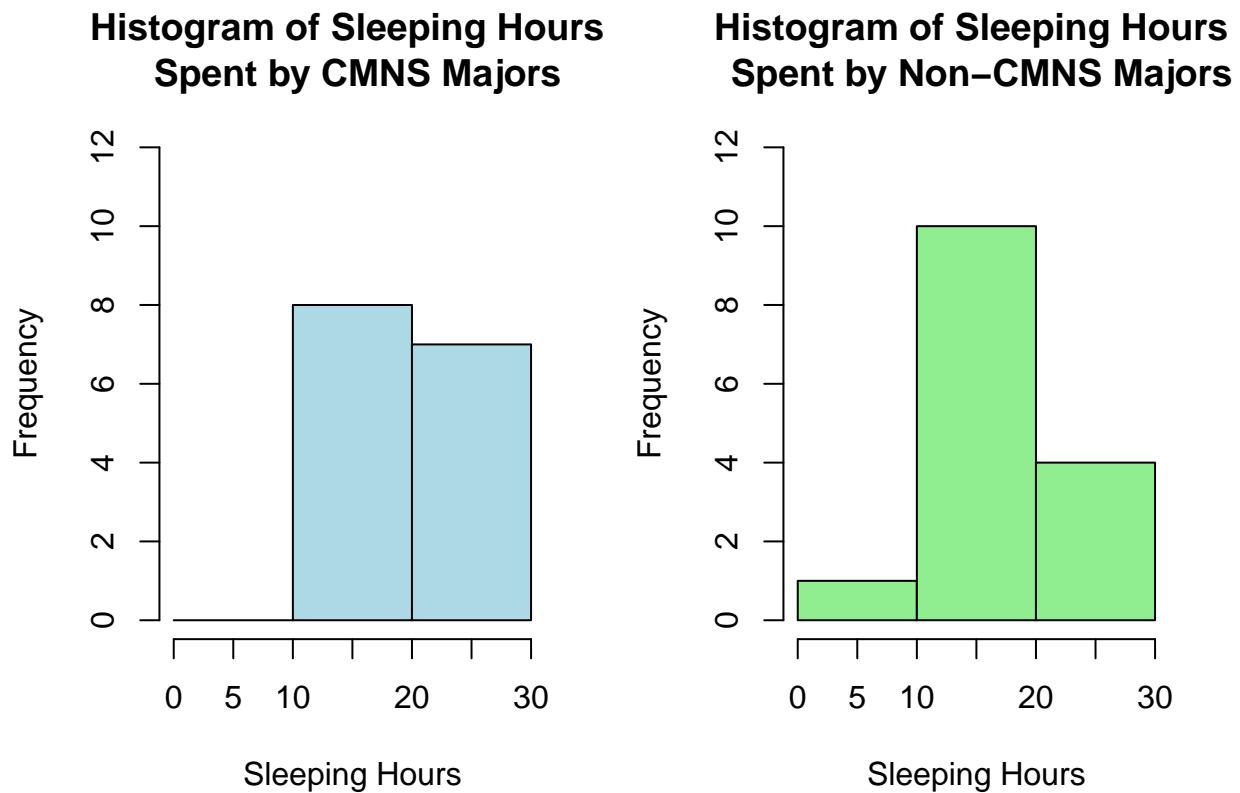


**Two-Sample Wilcoxon Rank-Sum Test**

```
##
Wilcoxon rank sum test
##
data: CMNS$`Sleeping Hours` and NonCMNS$`Sleeping Hours`
W = 145.5, p-value = 0.1674
alternative hypothesis: true location shift is not equal to 0
```

**Regression Model for Sleeping Hours Based on Major Type, Gender, and Current Credits**

```
##
## Call:
## lm(formula = survey$`Sleeping Hours` ~ survey$"CMNS Major" +
##     survey$Male + survey$"Current Credits")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.7464 -3.4351 -0.2288  4.1536  8.8019
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          31.8035     7.5280   4.225  0.00026 ***
## survey$"CMNS Major"   3.5012     2.0414   1.715  0.09823 .
## survey$Male          -1.3648     2.3138  -0.590  0.56037
```

```
## survey$"Current Credits"  -0.8494     0.4857  -1.749  0.09208 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.501 on 26 degrees of freedom
## Multiple R-squared:  0.2018, Adjusted R-squared:  0.1097
## F-statistic: 2.191 on 3 and 26 DF,  p-value: 0.1132
```

Residuals Plot for Sleeping Hours


**Residuals Plot for Sleeping Hours**
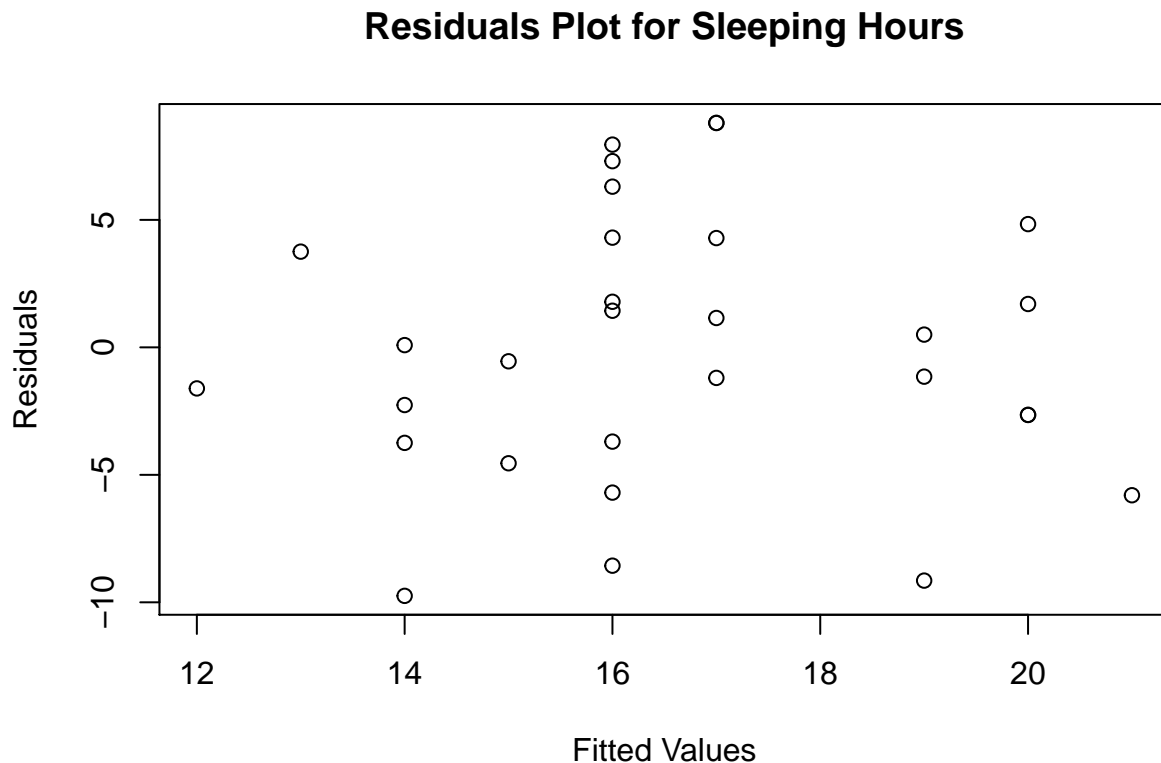
Social Activities

The normal probability plot and histogram for weekend hours spent on social activities appears non-normal for both CMNS majors and non-CMNS majors. In this case, we will use a two-sample Wilcoxon rank-sum test. The mean number of hours spent on social activities was 8.7 for CMNS majors and 9.2 for non-CMNS majors. The p-value for this test is .4637. This indicates that the probability of obtaining a difference of means at least as large as .5 hours by chance is 46.37%. Thus, we fail to reject the null hypothesis that the mean number of weekend hours allocated to social activities by CMNS majors and non-CMNS majors is the same.

In the regression model to predict number of weekend hours spent on social activities, none of the coefficients on the chosen variables were statistically significant at the 5% significance level. The residual plot also does not show any pattern in the residuals that would indicate that we should use a transformation. Consequently, we cannot present a regression equation to explain the impact on number of weekend hours spent on social activities of major type, gender, and number of current credits.

9

Normal Q-Q Plot of CMNS Students' Social Hours
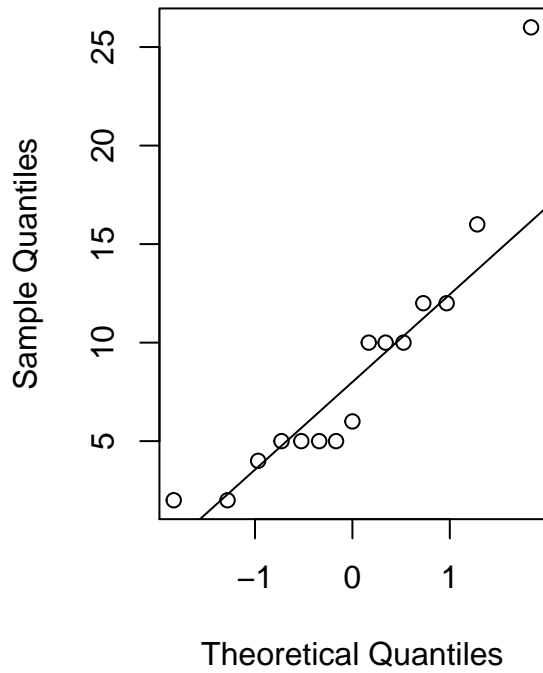
Normal Q-Q Plot of Non-CMNS Students' Social Hours

Histogram of Social Hours Spent by CMNS Majors

Histogram of Social Hours Spent by Non–CMNS Majors
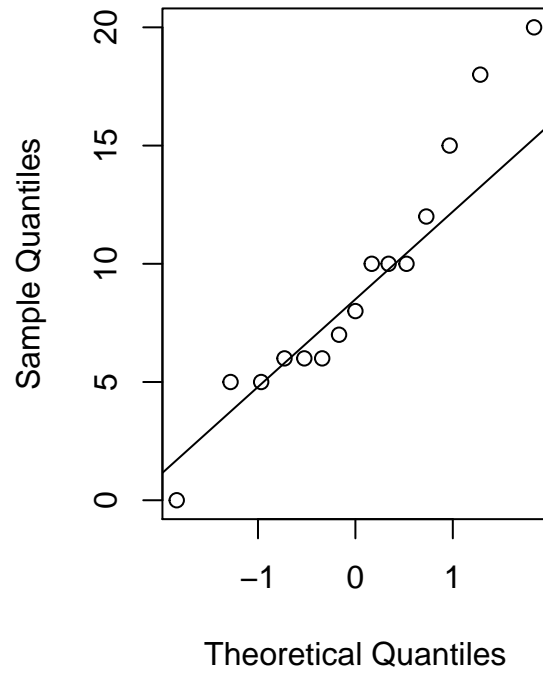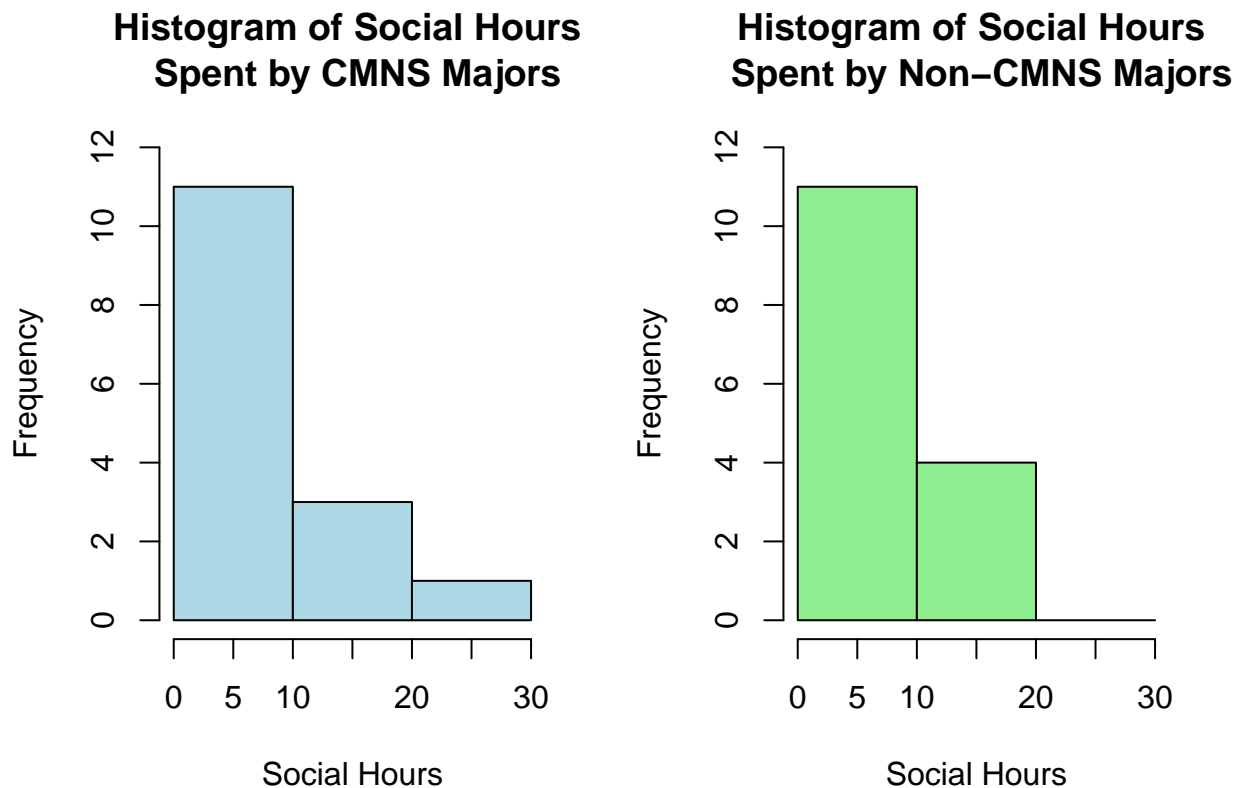
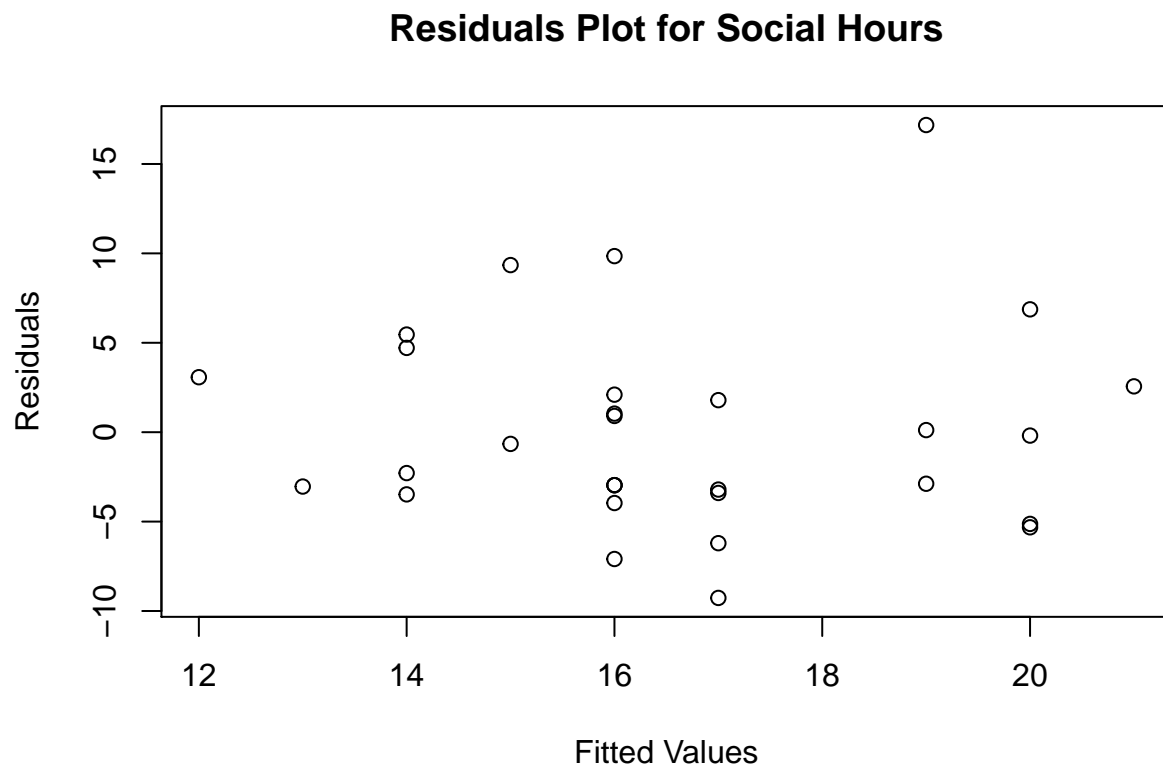**Two-Sample Wilcoxon Rank-Sum Test**

```
##
##  Wilcoxon rank sum test
##
## data:  CMNS$`Social Hours` and NonCMNS$`Social Hours`
## W = 95, p-value = 0.4637
## alternative hypothesis: true location shift is not equal to 0
```

Regression Model for Social Hours Based on Major Type, Gender, and Current Credits

```
##
## Call:
## lm(formula = survey$`Social Hours` ~ survey$"CMNS Major" + survey$Male +
##     survey$"Current Credits")
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.269 -3.352 -1.471  2.447 17.178
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            5.2405     8.1973   0.639    0.528
## survey$"CMNS Major"   -0.7536     2.2229  -0.339    0.737
## survey$Male           -0.8842     2.5195  -0.351    0.728
```

```
## survey$"Current Credits"   0.3070     0.5288   0.581     0.567
##
## Residual standard error: 5.99 on 26 degrees of freedom
## Multiple R-squared:  0.01801,    Adjusted R-squared:  -0.09529
## F-statistic: 0.159 on 3 and 26 DF,  p-value: 0.9229
```

Residuals Plot for Social Hours
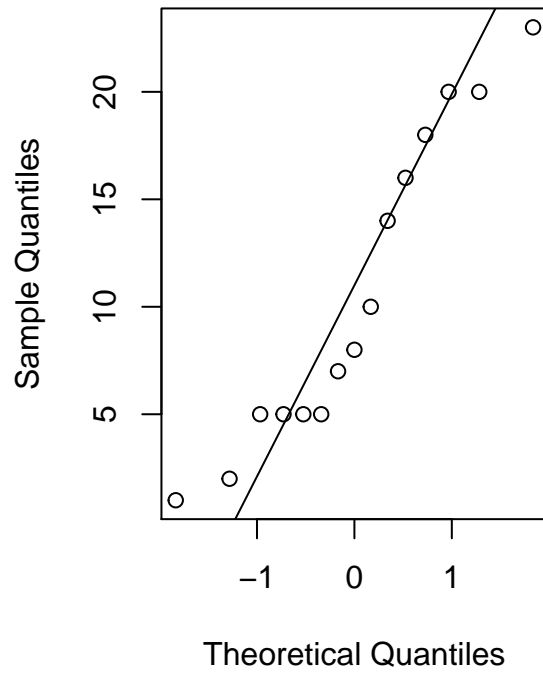
Residuals Plot for Social Hours
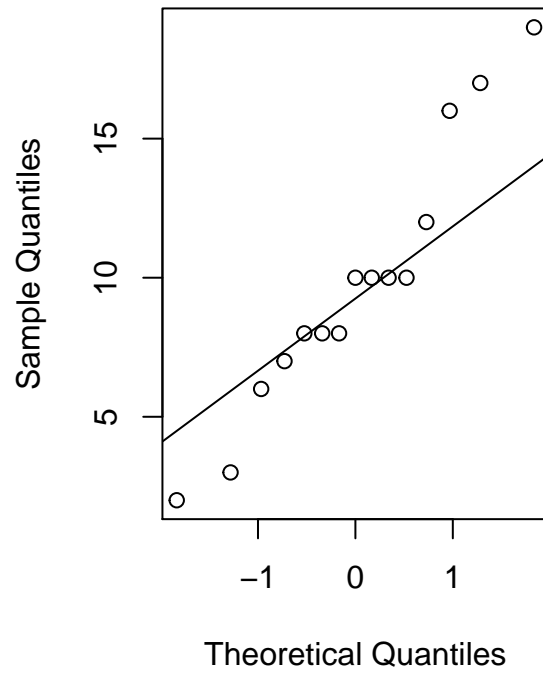

Leisure Activities

The normal probability plot and histogram for weekend hours spent on leisure activities appears non-normal. In this case, we will use a two-sample Wilcoxon rank-sum test. The mean number of hours spent on leisure activities was 10.6 for CMNS majors and 9.7 for non-CMNS majors. The p-value for this test is .9170. This indicates that the probability of obtaining a difference of means at least as large as .9 hours by chance is 91.70%. Thus, we fail to reject the null hypothesis that the mean number of weekend hours allocated to leisure activities by CMNS majors and non-CMNS majors is the same.
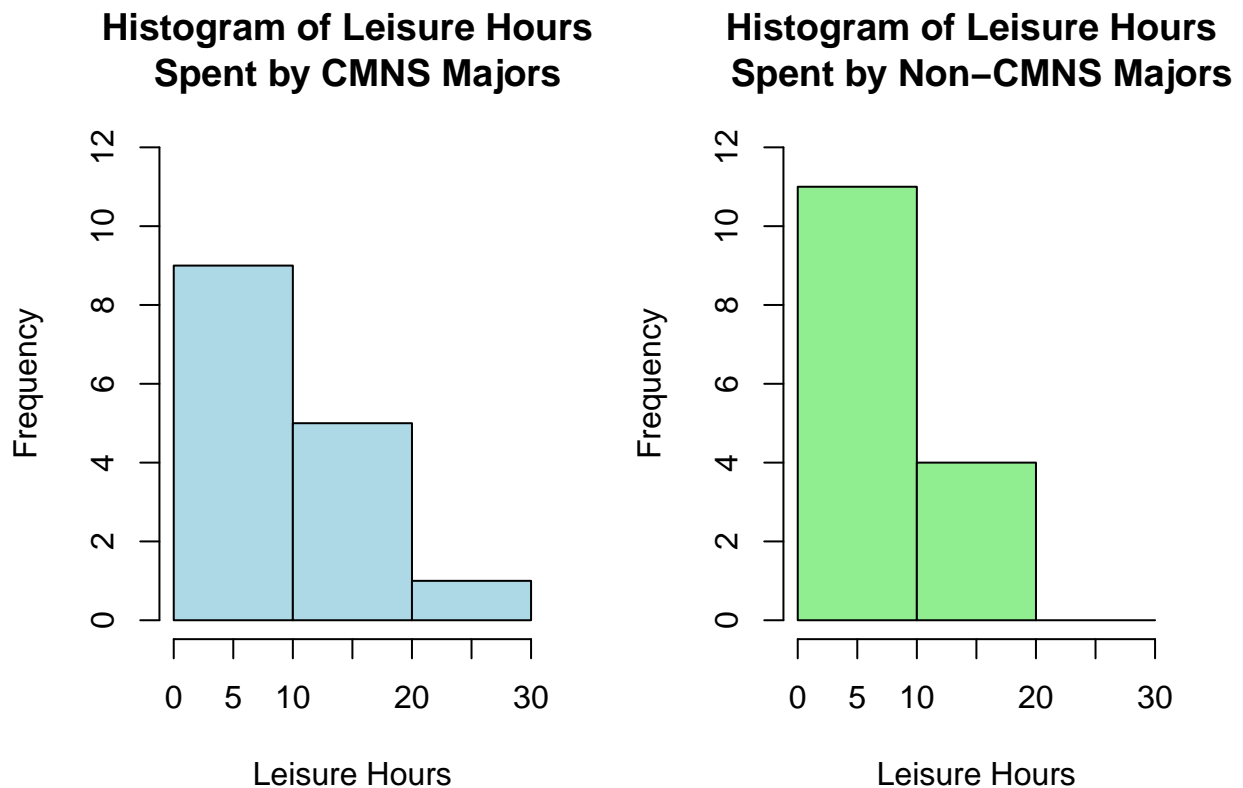
In the regression model to predict number of weekend hours spent on leisure activities, none of the coefficients on the chosen variables were statistically significant at the 5% significance level. The residual plot also does not show any pattern in the residuals that would indicate that we should use a transformation. Consequently, we cannot present a regression equation to explain the impact on number of weekend hours spent on leisure activities of major type, gender, and number of current credits.

Normal Q–Q Plot of CMNS Students' Leisure Hours

Sample Quantiles

Theoretical Quantiles

Normal Q–Q Plot of Non–CMNS Students' Leisure Hours

Sample Quantiles

Theoretical Quantiles

Histogram of Leisure Hours Spent by CMNS Majors

Histogram of Leisure Hours Spent by Non−CMNS Majors

**Two-Sample Wilcoxon Rank-Sum Test**

```
##
##  Wilcoxon rank sum test
##
## data:  CMNS$`Leisure Hours` and NonCMNS$`Leisure Hours`
## W = 110, p-value = 0.917
## alternative hypothesis: true location shift is not equal to 0
```

Regression Model for Leisure Hours Based on Major Type, Gender, and Current Credits

```
##
## Call:
## lm(formula = survey$`Leisure Hours` ~ survey$"CMNS Major" + survey$Male +
##     survey$"Current Credits")
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -9.806 -4.127 -1.207  4.207 10.994
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.55242    8.63471   0.411    0.684
## survey$"CMNS Major"  0.72764    2.34154   0.311    0.758
## survey$Male         -0.02308    2.65397  -0.009    0.993
```

```
## survey$"Current Credits"   0.40785     0.55706    0.732     0.471
##
## Residual standard error: 6.309 on 26 degrees of freedom
## Multiple R-squared:  0.02563,    Adjusted R-squared:  -0.0868
## F-statistic: 0.228 on 3 and 26 DF,  p-value: 0.8761
```

Residuals Plot for Leisure Hours

Residuals Plot for Leisure Hours
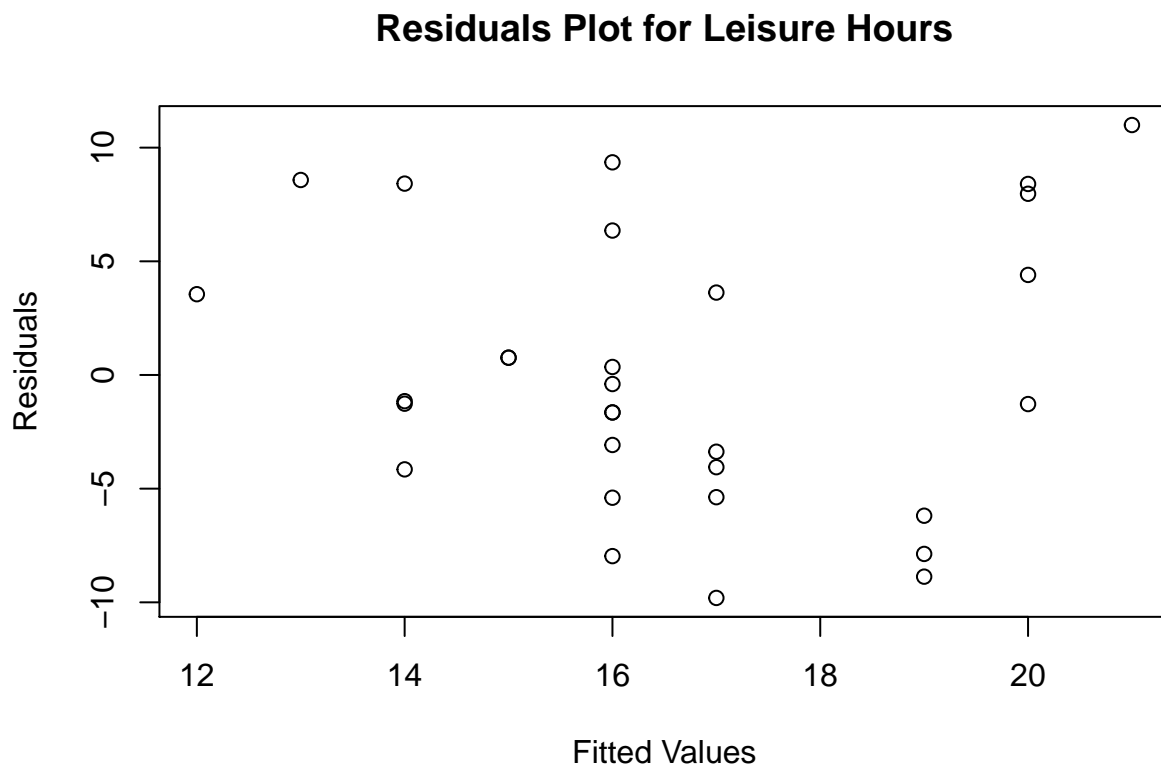
Conclusions:

We failed to find significant evidence of differences in the mean number of weekend hours allocated to the chosen categories by conducting two-sample difference of means tests in all categories. In retrospect, for the difference of means analysis, we should have sampled at least 30 students in each category because we had to check normality in the sampling distributions for each population separately to meet the conditions for the two-sample difference of means t-test. However, time and resource constraints made obtaining a total of 60 students overall very difficult. We also failed to find a significant relationship between the number of hours spent on certain types of weekend activities and gender, number of current credits, or type of major. This would suggest that the answer to our research question is that there is no evidence that undergraduate UMD students taking CMNS majors differ significantly from undergraduate UMD students taking non-liberal arts, non-CMNS majors. However, since we conducted a convenience sample, our analysis cannot be considered very reliable.