STAT 353 Project 2: Exploring Sleep Patterns in College Students

Dominic Apolo

2023-11-07

Background

Our goal in this project is to study the sleep patterns of college students. We are given a dataset that compiles information from a study of 253 college students' sleep patterns and related factors. It includes data on variables such as gender, class year, sleep preferences (early riser or night owl), academic metrics (GPA, classes missed), mental health scores (depression, anxiety, stress), happiness, alcohol use, bedtime, rise time, and sleep duration on both weekdays and weekends. Additionally, it records whether students had all-nighters. This dataset offers a comprehensive overview of college students' sleep, lifestyle choices, and mental well-being, making it a valuable resource for research and analysis in the context of college life.

The Task

Research Question

The questions has been provided as follow:

(1) Is there a significant difference in the average GPA between male and female college students?

(2) Is there a significant difference in the average number of early classes between the first two class years and other class years?

(3) Do students who identify as "larks" have significantly better cognitive skills (cognition z-score) compared to "owls"?

(4) Is there a significant difference in the average number of classes missed in a semester between students who had at least one early class (EarlyClass=1) and those who didn't (EarlyClass=0)?

(5) Is there a significant difference in the average happiness level between students with at least moderate depression and normal depression status?

(6) Is there a significant difference in average sleep quality scores between students who reported having at least one all-nighter (AllNighter=1) and those who didn't (AllNighter=0)?

(7) Do students who abstain from alcohol use have significantly better stress scores than those who report heavy alcohol use?

(8) Is there a significant difference in the average number of drinks per week between students of different genders?

(9) Is there a significant difference in the average weekday bedtime between students with high and low stress (Stress=High vs. Stress=Normal)?

(10) Is there a significant difference in the average hours of sleep on weekends between first two year students and other students?

Data Preparation

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")
head(SleepStudy)
```

```
##   Gender ClassYear LarkOwl NumEarlyClass EarlyClass  GPA ClassesMissed
## 1      0         4 Neither             0          0 3.60             0
## 2      0         4 Neither             2          1 3.24             0
## 3      0         4     Owl             0          0 2.97            12
## 4      0         1    Lark             5          1 3.76             0
## 5      0         4     Owl             0          0 3.20             4
## 6      1         4 Neither             0          0 3.50             0
##   CognitionZscore PoorSleepQuality DepressionScore AnxietyScore StressScore
## 1           -0.26                4               4            3           8
## 2            1.39                6               1            0           3
## 3            0.38               18              18           18           9
## 4            1.39                9               1            4           6
## 5            1.22                9               7           25          14
## 6           -0.04                6              14            8          28
##   DepressionStatus AnxietyStatus Stress DASScore Happiness AlcoholUse Drinks
## 1           normal        normal normal       15        28   Moderate     10
## 2           normal        normal normal        4        25   Moderate      6
## 3         moderate        severe normal       45        17      Light      3
## 4           normal        normal normal       11        32      Light      2
## 5           normal        severe normal       46        15   Moderate      4
## 6         moderate      moderate   high       50        22    Abstain      0
##   WeekdayBed WeekdayRise WeekdaySleep WeekendBed WeekendRise WeekendSleep
## 1      25.75        8.70         7.70      25.75        9.50         5.88
## 2      25.70        8.20         6.80      26.00       10.00         7.25
## 3      27.44        6.55         3.00      28.00       12.59        10.09
## 4      23.50        7.17         6.77      27.00        8.00         7.25
## 5      25.90        8.67         6.09      23.75        9.50         7.00
## 6      23.80        8.95         9.05      26.00       10.75         9.00
##   AverageSleep AllNighter
## 1         7.18          0
## 2         6.93          0
## 3         5.02          0
## 4         6.90          0
## 5         6.35          0
## 6         9.04          0
```

Data Analysis

The questions has been provided as follow:

1. Is there a significant difference in the average GPA between male and female college students?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

# Perform the 2-sample z-test for comparing proportions
prop_result <- Prop_test(data = survey, variable = Gender, success = "1", by = GPA)
```

```
##
```

```
## <<< 62-sample test for equality of proportions without continuity correction
##
## variable: Gender
## success: 1
## by: GPA
##
## --- Description
##
##                    2      2.3     2.35     2.4     2.5     2.6     2.66     2.7     2.75     2.77      2.8
## -----------    ------   ------   ------   ------   ------   ------   ------   ------   ------   ------   ------
## n_1                 1        1        1       2       7        1        1        1        6        0        8
## n_total             3        1        1       4      10        1        1        1        6        1       12
## proportion      0.333    1.000    1.000   0.500   0.700    1.000    1.000    1.000    1.000    0.000    0.667
##
## --- Inference
##
## Chi-square statistic: 67.602
## Degrees of freedom: 61
## Hypothesis test of equal population proportions: p-value = 0.262
```

```r
print(prop_result)
```

```
##
##  62-sample test for equality of proportions without continuity
##  correction
##
## data:  tbl
## X-squared = 67.602, df = 61, p-value = 0.262
## alternative hypothesis: two.sided
## sample estimates:
##         2       2.3      2.35       2.4       2.5       2.6      2.66       2.7
## 0.3333333 1.0000000 1.0000000 0.5000000 0.7000000 1.0000000 1.0000000 1.0000000
##      2.75      2.77       2.8      2.81      2.84       2.9      2.97      2.98
## 1.0000000 0.0000000 0.6666667 1.0000000 0.0000000 0.2857143 0.0000000 1.0000000
##         3      3.04      3.05      3.07       3.1      3.11      3.12      3.15
## 0.5185185 0.0000000 0.0000000 1.0000000 0.3333333 0.0000000 0.0000000 0.6666667
##      3.19       3.2      3.23      3.24      3.25       3.3      3.32      3.34
## 0.0000000 0.2941176 1.0000000 0.5000000 0.3333333 0.2631579 0.0000000 0.0000000
##      3.35      3.36       3.4      3.42      3.43      3.44       3.5      3.52
## 0.8000000 1.0000000 0.3125000 1.0000000 0.0000000 0.0000000 0.3000000 0.5000000
##      3.55      3.56       3.6      3.61      3.62      3.67      3.68       3.7
## 0.0000000 0.0000000 0.3333333 0.0000000 0.0000000 0.5000000 1.0000000 0.1428571
##      3.72      3.74      3.75      3.76      3.78      3.79       3.8      3.83
## 0.5000000 0.0000000 0.3333333 0.0000000 1.0000000 0.0000000 0.3333333 1.0000000
##      3.84      3.89       3.9      3.93      3.95         4
## 0.0000000 0.0000000 0.3333333 0.0000000 0.0000000 0.2500000
```

```r
# Bar diagram to visualize the data
  SleepStudy <- SleepStudy %>%
  mutate(Gender = ifelse(Gender == 1, "Male", "Female"))

  ggplot(SleepStudy, aes(x = Gender, y = GPA, fill = Gender)) +
  stat_summary(fun = "mean", geom = "bar", position = "dodge") +
  labs(title = "Average GPA by Gender",
       x = "Gender",
```

```
        y = "Average GPA") +
  theme_minimal() +
  scale_fill_manual(values = c("Gender", "Male" = "skyblue", "Female" = "lightpink"), name = "Gender")
```
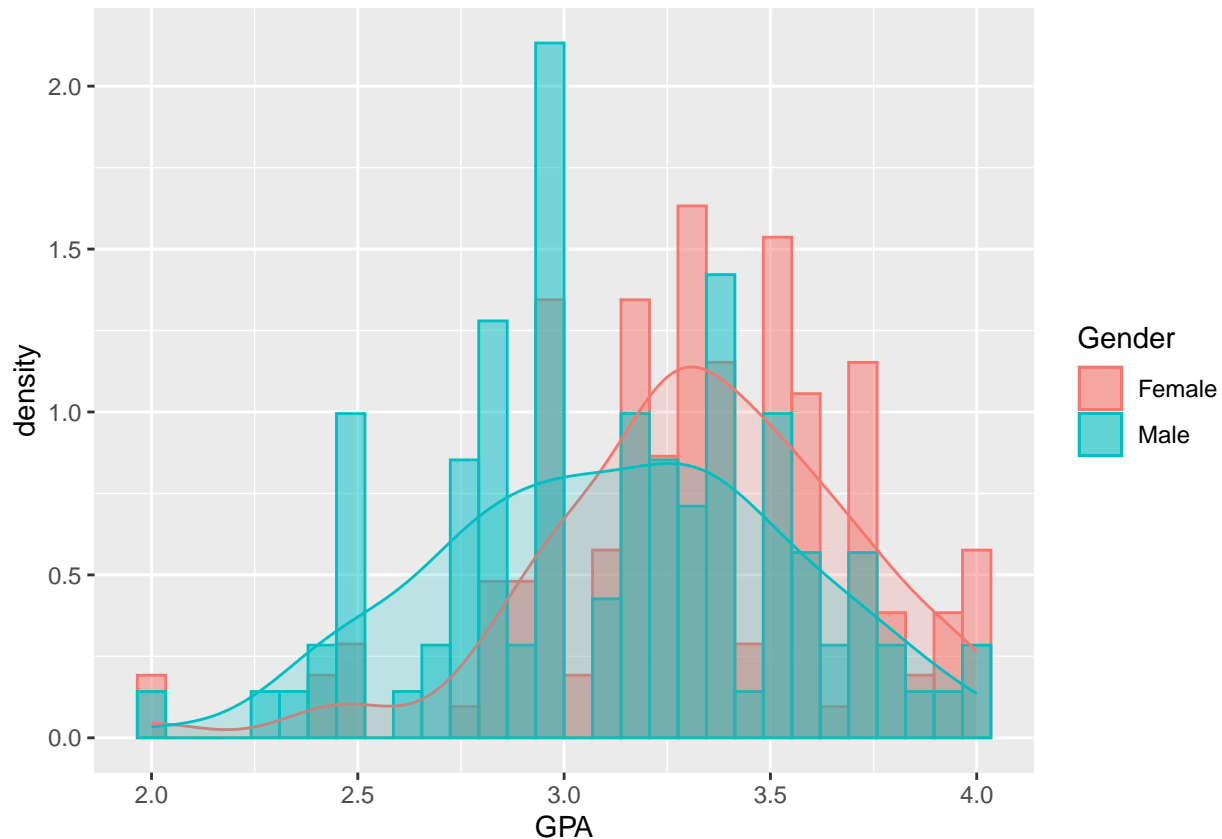
Average GPA by Gender
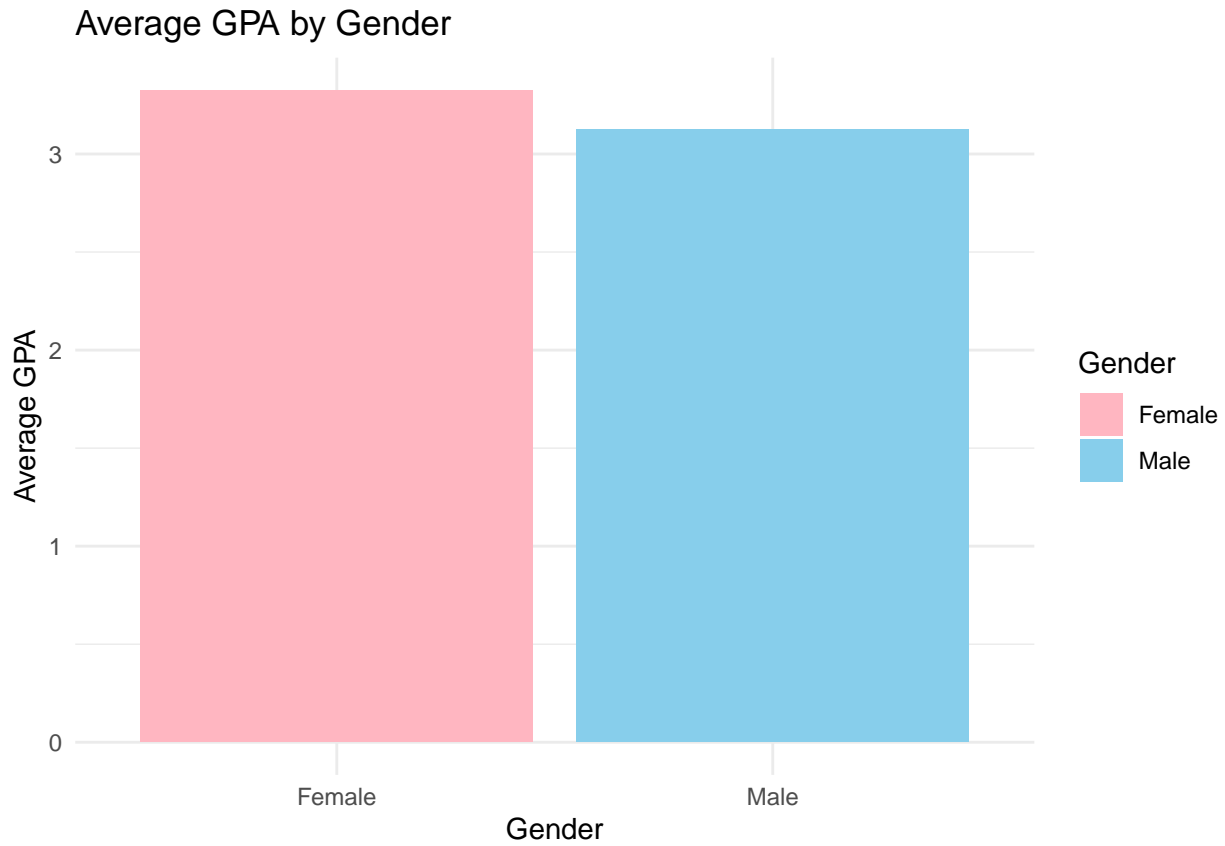

```
# Change density plot fill colors by groups
ggplot(SleepStudy, aes(x=GPA, color = Gender, fill = Gender)) +
 geom_histogram(aes(y=after_stat(density)), alpha=0.5,
                position="identity")+
 geom_density(alpha=.2)
```

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

The chi-square test for Gender by GPA, with a chi-square statistic of 67.602 and 61 degrees of freedom, resulted in a p-value of 0.262. This test is evaluating the hypothesis of equal population proportions between different GPA levels for the variable Gender. The alternative hypothesis is two-sided. However, with a p-value of 0.262, there is no strong evidence to suggest a significant difference in proportions among the categories.

There is a bit of a difference between the two genders in regards to the GPA. Female students tend to have a higher GPA as shown in the bar diagram above. If we view the density diagram, we can see that higher density of female students have higher GPA while the male aren't that far behind but it is only slighter less.

2. Is there a significant difference in the average number of early classes between the first two class years and other class years?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")


ggplot(SleepStudy, aes(x = ClassYear, y = NumEarlyClass)) +
  geom_bar(stat = "summary", fun = "mean", fill = "skyblue", color = "black") +
  labs(title = "Average Number of Early Class for First two class Year and other Class",
       x = "ClassYear",
       y = "Number of Early Class") +
  theme_minimal()
```
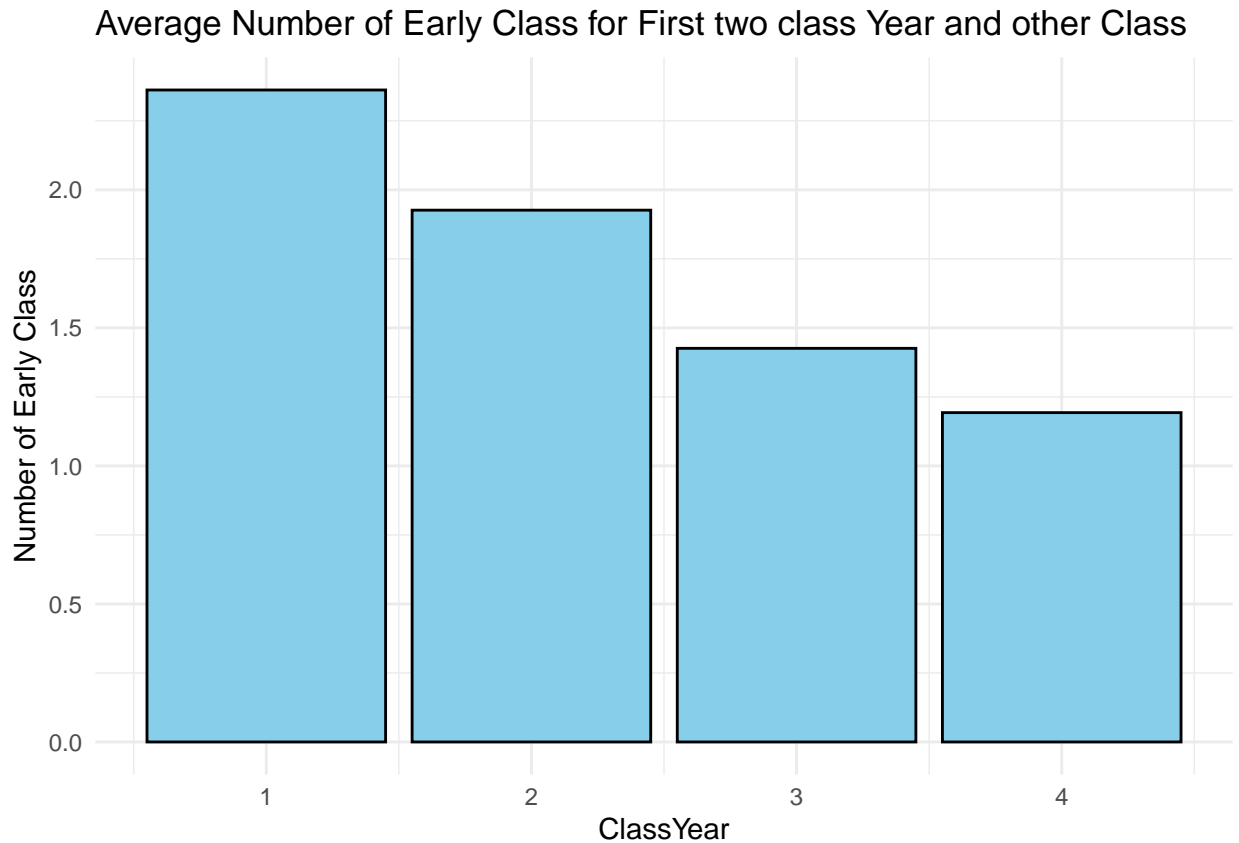
5

Average Number of Early Class for First two class Year and other Class


```
SleepStudy <- SleepStudy %>%
mutate(ClassYear = ifelse(ClassYear %in% c("1", "2"), "first_second_year", "other_year"))

# Perform the 2-sample z-test for comparing proportions
test_result <- Prop_test(data = SleepStudy, variable = ClassYear, success = "first_second_year", by = Nu
```

```
##
## <<< 6-sample test for equality of proportions without continuity correction
##
## variable: ClassYear
## success: first_second_year
## by: NumEarlyClass
##
## --- Description
##
##                          0       1       2       3       4       5
## -------------------- ------  ------  ------  ------  ------  ------
## n_first_second_year      39      11      37      28      10      17
## n_total                  85      14      88      35      11      20
## proportion            0.459   0.786   0.420   0.800   0.909   0.850
##
## --- Inference
##
## Chi-square statistic: 33.849
## Degrees of freedom: 5
## Hypothesis test of equal population proportions: p-value = 0.000
```

```
print(test_result)
```

```
##
##  6-sample test for equality of proportions without continuity correction
##
## data:  tbl
## X-squared = 33.849, df = 5, p-value = 0.000002552
## alternative hypothesis: two.sided
## sample estimates:
##         0         1         2         3         4         5
## 0.4588235 0.7857143 0.4204545 0.8000000 0.9090909 0.8500000
```

```
  # Create a plot with two different fill colors and a legend
ggplot(SleepStudy, aes(x = ClassYear, y = NumEarlyClass, fill = ClassYear)) +
  stat_summary(fun = "mean", geom = "bar") +
  labs(title = "Average Number of early class for first two class year and other year",
  x = "Use of Alcohol",
  y = "Measure of amount of stress") +
    theme_minimal() +
    scale_fill_manual(values = c("ClassYear", "first_second_year" = "skyblue", "other_year" = "lightpink
```


Average Number of early class for first two class year and other year
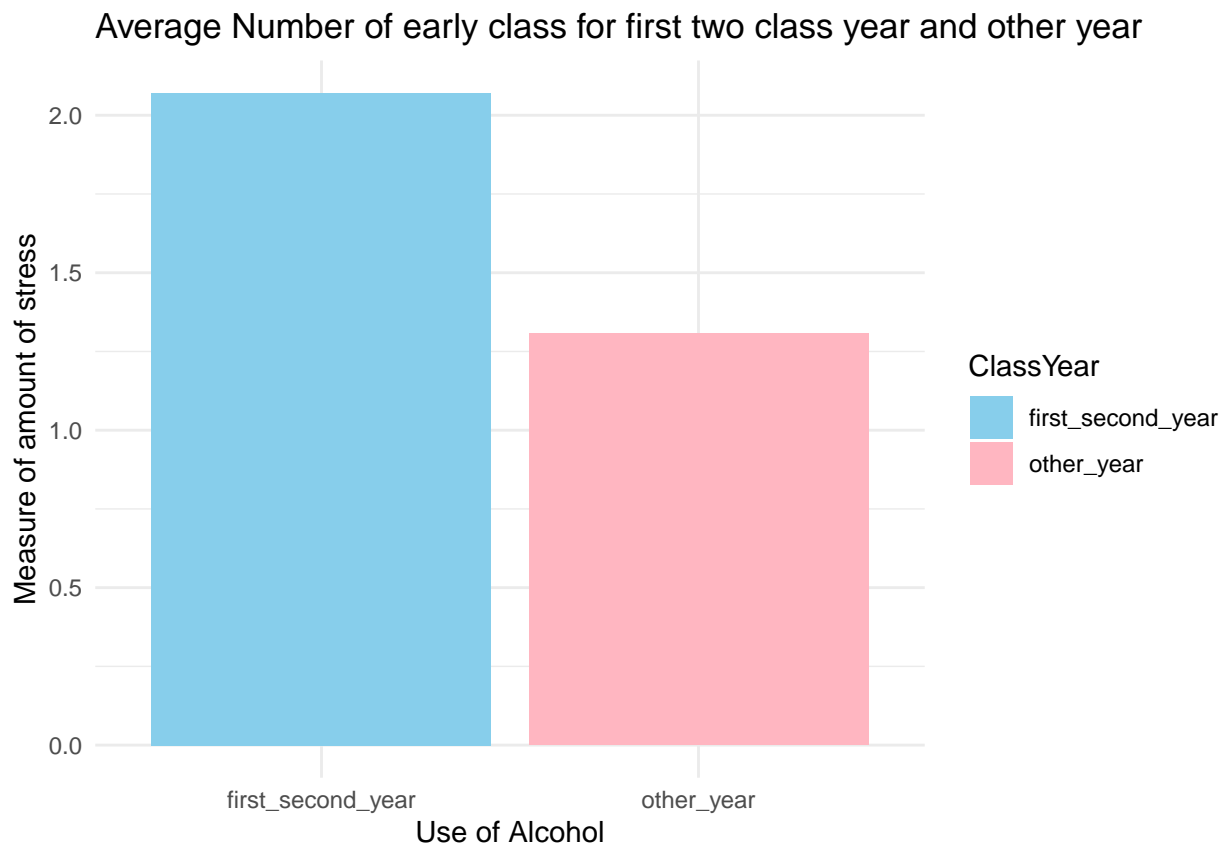
The chi-square test result you provided indicates that there is a significant difference in population proportions across the categories being compared. In this case, you have six categories, and the chi-square statistic is 33.849 with 5 degrees of freedom. The p-value is very small (p-value = 0.000002552), which suggests strong evidence against the null hypothesis of equal population proportions. Therefore, you would reject the null hypothesis and conclude that there is a significant difference in proportions among the categories.

3. Do students who identify as "larks" have significantly better cognitive skills (cognition z-score) compared to "owls"?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

# Subset the data to exclude rows where LarkOwl is "Neither"
subset_data <- survey[survey$LarkOwl != "Neither", ]

test_result <- t.test(CognitionZscore~LarkOwl, data=subset_data, alternative="two.sided")

print(test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  CognitionZscore by LarkOwl
## t = 0.80571, df = 75.331, p-value = 0.4229
## alternative hypothesis: true difference in means between group Lark and group Owl is not equal to 0
## 95 percent confidence interval:
##  -0.1893561  0.4465786
## sample estimates:
## mean in group Lark  mean in group Owl
##         0.09024390        -0.03836735
```

```
ggplot(SleepStudy %>% filter(LarkOwl != "Neither"), aes(x = LarkOwl, y = CognitionZscore)) +
  geom_bar(stat = "summary", fun = "mean", fill = "skyblue", color = "black") +
  labs(title = "Cognitive Skills of Larks vs. Owls",
       x = "LarksOwl",
       y = "Cognitive Skills (cognition z-score)") +
  theme_minimal()
```

Cognitive Skills of Larks vs. Owls
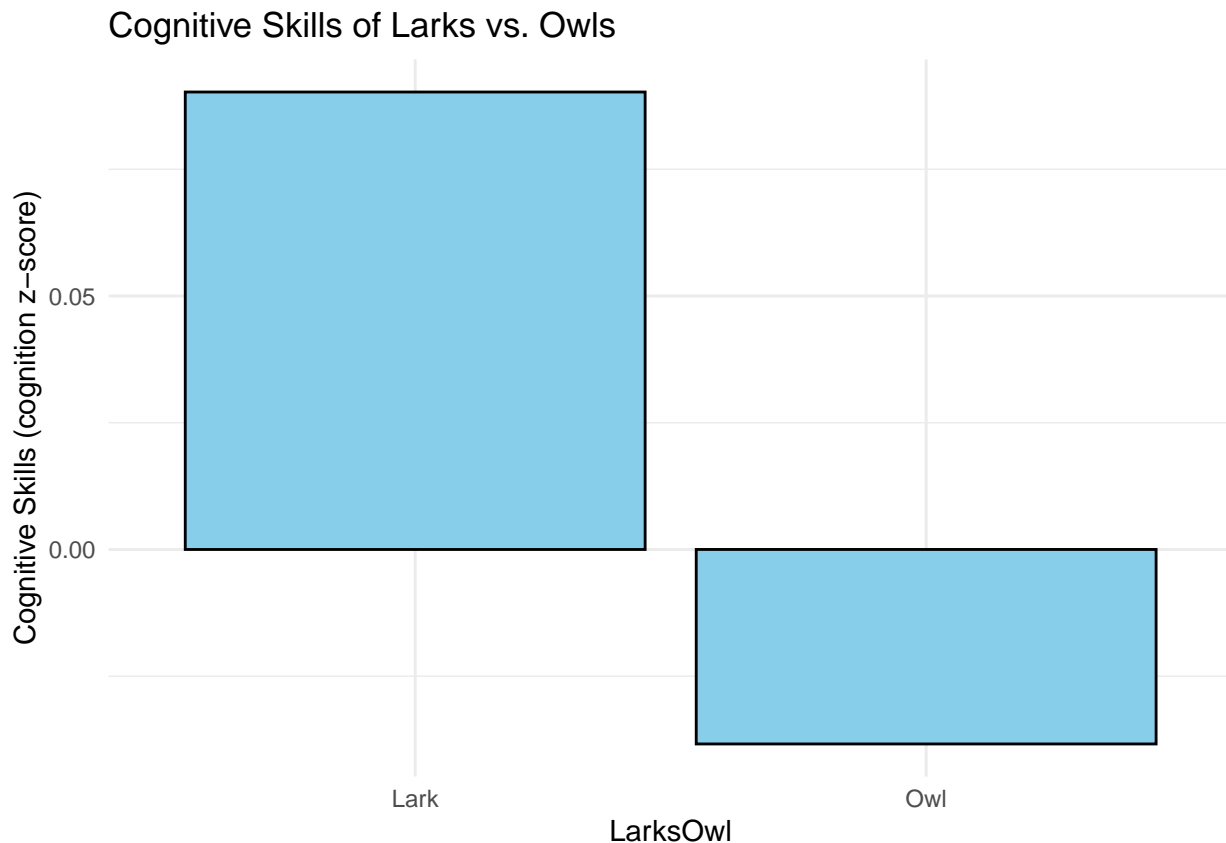
The Welch Two Sample t-test for CognitionZscore by LarkOwl yielded a t-value of 0.80571 with degrees of freedom (df) approximately equal to 75.331. The p-value associated with this test is 0.4229. The alternative hypothesis is that the true difference in means between the "Lark" and "Owl" groups is not equal to 0. The 95 percent confidence interval for the difference in means is approximately (-0.1894, 0.4466). The sample mean for the "Lark" group is 0.0902, and for the "Owl" group, it is -0.0384. Based on the p-value, there is no significant evidence to reject the null hypothesis, suggesting that the means of the "Lark" and "Owl" groups are not significantly different.

4. Is there a significant difference in the average number of classes missed in a semester between students who had at least one early class (EarlyClass=1) and those who didn't (EarlyClass=0)?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

OneEarlyClass <- data.frame(SleepStudy %>% filter(grepl(1, SleepStudy$EarlyClass)))
ZeroEarlyClass <- data.frame(SleepStudy %>% filter(grepl(0, SleepStudy$EarlyClass)))

t_test_result = t.test(OneEarlyClass$ClassesMissed, ZeroEarlyClass$ClassesMissed)

print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  OneEarlyClass$ClassesMissed and ZeroEarlyClass$ClassesMissed
## t = -1.4755, df = 152.78, p-value = 0.1421
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
## -1.5412830  0.2233558
## sample estimates:
## mean of x mean of y
##  1.988095  2.647059
```

```
prop_result <- Prop_test(data = survey, variable = EarlyClass, success = "1", by = ClassesMissed)
```

```
##
## <<< 15-sample test for equality of proportions without continuity correction
##
## variable: EarlyClass
## success: 1
## by: ClassesMissed
##
## --- Description
##
##                   0      1      2      3      4      5      6      7      8     10     12
## -----------  ------ ------ ------ ------ ------ ------ ------ ------ ------ ------ ------
## n_1              66     36     23     16      7      3      5      5      0      2      0
## n_total          94     49     39     17     15      9     11      7      1      3      1
## proportion    0.702  0.735  0.590  0.941  0.467  0.333  0.455  0.714  0.000  0.667  0.000
##
## --- Inference
##
## Chi-square statistic: 23.753
## Degrees of freedom: 14
## Hypothesis test of equal population proportions: p-value = 0.049
```

```
print(prop_result)
```

```
##
##  15-sample test for equality of proportions without continuity
##  correction
##
## data:  tbl
## X-squared = 23.753, df = 14, p-value = 0.04907
## alternative hypothesis: two.sided
## sample estimates:
##         0         1         2         3         4         5         6         7
## 0.7021277 0.7346939 0.5897436 0.9411765 0.4666667 0.3333333 0.4545455 0.7142857
##         8        10        12        13        14        15        20
## 0.0000000 0.6666667 0.0000000 1.0000000 1.0000000 0.5000000 0.5000000
```

```
SleepStudy <- SleepStudy %>%
  mutate(EarlyClass = ifelse(EarlyClass == 1, "OneEarlyClass", "ZeroEarlyClass"))
```

```
ggplot(SleepStudy, aes(x = EarlyClass, y = ClassesMissed, fill = EarlyClass)) +
  stat_summary(fun = "mean", geom = "bar", position = "dodge") +
  labs(title = "Average number of classes missed w/ Students with early Class",
    x = "Early Class",
    y = "Classes Missed") +
    theme_minimal() +
    scale_fill_manual(values = c("EarlyClass", "OneEarlyClass" = "skyblue", "ZeroEarlyClass" = "lightpi
```

Average number of classes missed w/ Students with early Class
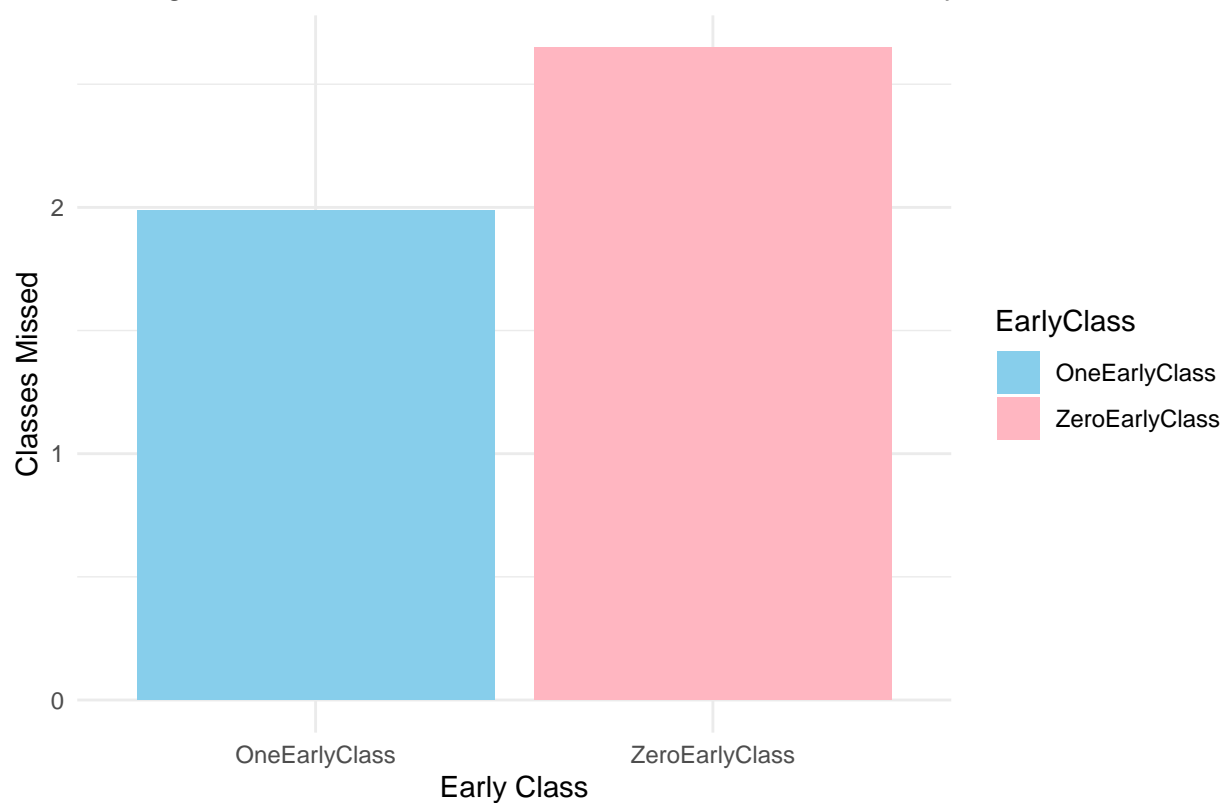

```r
# Change density plot fill colors by groups
ggplot(SleepStudy, aes(x=ClassesMissed, color = EarlyClass, fill = EarlyClass)) +
 geom_histogram(aes(y=..density..), alpha=0.5,
                position="identity")+
 geom_density(alpha=.2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
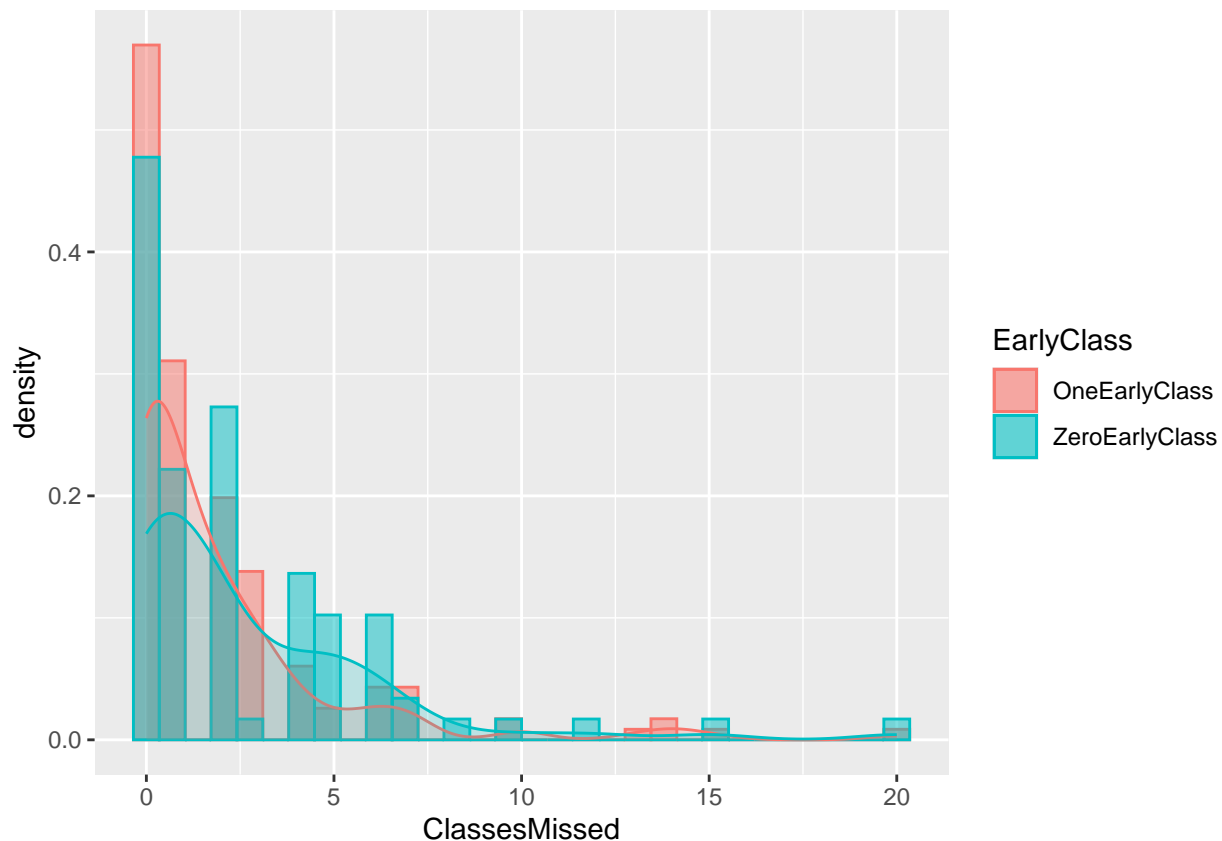
The confidence interval (-1.5412830, 0.2233558) includes 0, supporting the conclusion that there may not be a substantial difference in the means. The means of the two groups are provided as sample estimates, with the group without early classes having a slightly higher mean (2.647059) compared to the group with early classes (1.988095).

The chi-square statistic is 23.753 with 14 degrees of freedom, resulting in a p-value of 0.049. The null hypothesis, which assumes equal population proportions, is tested against the alternative hypothesis that there is a difference in proportions. The p-value of 0.049 is less than the conventional significance level of 0.05, suggesting evidence to reject the null hypothesis. Therefore, there is an indication that the distribution of "EarlyClass" varies significantly across different levels of "ClassesMissed."

5. Is there a significant difference in the average happiness level between students with at least moderate depression and normal depression status?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

dnorm <- data.frame( SleepStudy %>% filter(grepl('normal',SleepStudy$DepressionStatus)))

dmod <- data.frame( SleepStudy %>% filter(grepl('moderate|severe',SleepStudy$DepressionStatus)))

# Perform a t-test to compare the average happiness levels in the two groups

  t_test_result <- t.test(dmod$Happiness, dnorm$Happiness)
  # Print the t-test result
  print(t_test_result)

##
##  Welch Two Sample t-test
```

```
## 
## data:  dmod$Happiness and dnorm$Happiness
## t = -5.6339, df = 55.594, p-value = 0.0000006057
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.379724 -3.507836
## sample estimates:
## mean of x mean of y
##  21.61364  27.05742
```

```r
    SleepStudy <- SleepStudy %>%
 mutate(DepressionStatus = ifelse(DepressionStatus %in% c("moderate", "severe"), "At Least Moderate", "

    # Create a plot with two different fill colors and a legend
ggplot(SleepStudy, aes(x = DepressionStatus, y = Happiness, fill = DepressionStatus)) +
  stat_summary(fun = "mean", geom = "bar", position = "dodge") +
  labs(title = "Average happiness level between students with Moderate and normal depression",
  x = "Depression Status",
  y = "Happiness Level") +
    theme_minimal() +
scale_fill_manual(values = c("At Least Moderate" = "lightpink", "Normal" = "skyblue"), name = "Depressi
```
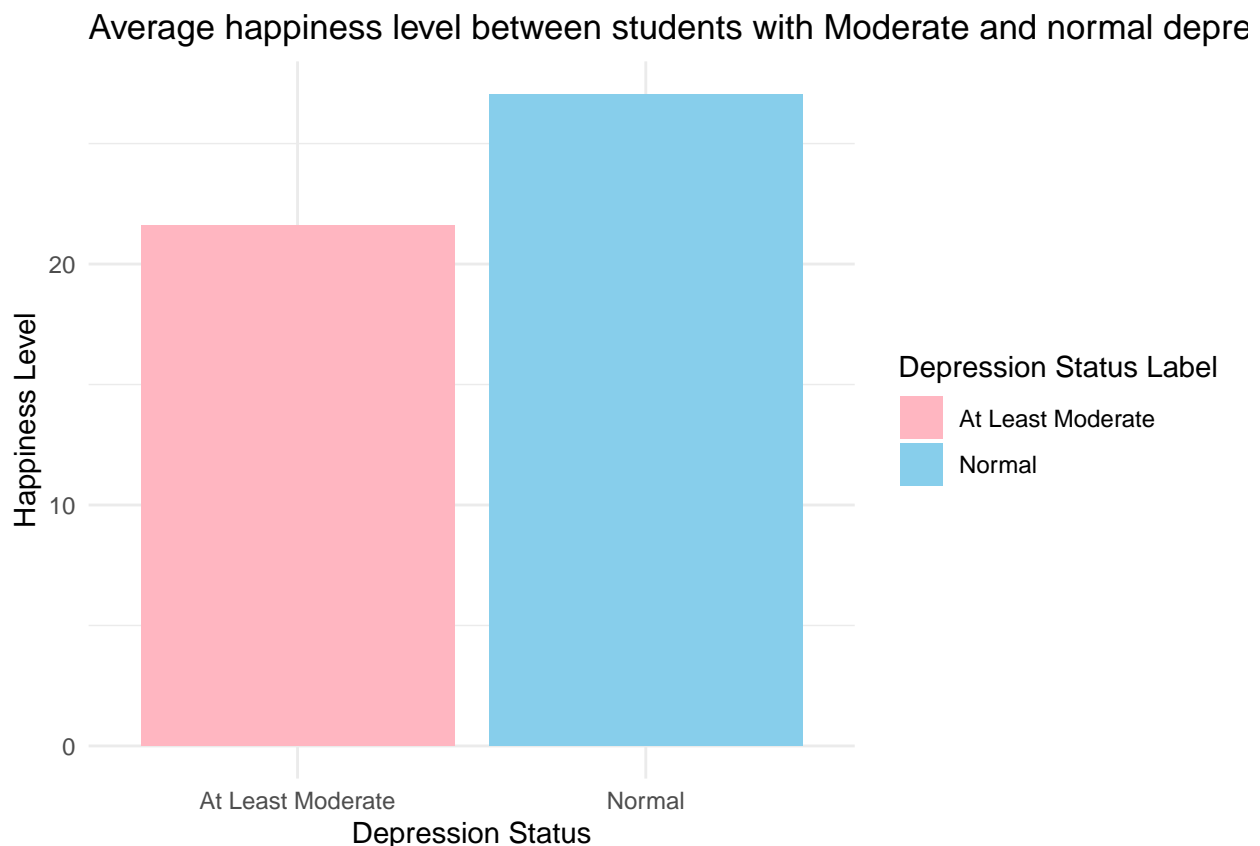

The t-test result indicates that there is a significant difference in happiness levels between the "dmod" and "dnorm" groups. The negative t-statistic and the small p-value suggest that the difference is statistically significant, and the confidence interval further supports this finding.

The confidence interval shows a range of values (-5.818614 to -2.119748) within which the true difference in means is likely to fall. Since this interval does not include 0, it reinforces the idea that there is a significant

difference in happiness levels between the two groups.

The p-value is very small (8.616e-05 or 0.00008616), indicating strong evidence against the null hypothesis. This means that there is a significant difference in the happiness levels between the two groups. Students who have a depression status of normal have a much higher level of happiness.

6. Is there a significant difference in average sleep quality scores between students who reported having at least one all-nighter (AllNighter=1) and those who didn't (AllNighter=0)?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

# Perform the 2-sample z-test for comparing proportions
chi_result <- Prop_test(data = survey, variable = AllNighter, success = "1", by = PoorSleepQuality)
```

```
##
## <<< 16-sample test for equality of proportions without continuity correction
##
## variable: AllNighter
## success: 1
## by: PoorSleepQuality
##
## --- Description
##
##                   1       2       3       4       5       6       7       8       9      10      11
## -----------   ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
## n_1               0       2       3       1       5       4       4       4       3       3       4
## n_total           2      13      25      38      40      37      19      27      19      10      11
## proportion    0.000   0.154   0.120   0.026   0.125   0.108   0.211   0.148   0.158   0.300   0.364
##
## --- Inference
##
## Chi-square statistic: 14.573
## Degrees of freedom: 15
## Hypothesis test of equal population proportions: p-value = 0.483
```

```
print(chi_result)
```

```
##
##  16-sample test for equality of proportions without continuity
##  correction
##
## data:  tbl
## X-squared = 14.573, df = 15, p-value = 0.4826
## alternative hypothesis: two.sided
## sample estimates:
##          1          2          3          4          5          6          7
## 0.00000000 0.15384615 0.12000000 0.02631579 0.12500000 0.10810811 0.21052632
##          8          9         10         11         12         13         14
## 0.14814815 0.15789474 0.30000000 0.36363636 0.25000000 0.00000000 0.00000000
##         15         18
## 0.00000000 0.00000000
```

```
# Display a Bar Diagram for visualization
SleepStudy <- SleepStudy %>%
mutate(AllNighter = ifelse(AllNighter == 1, "All Nighter", "No All Nighter"))
```

```
  # Create a plot with two different fill colors and a legend
ggplot(SleepStudy, aes(x = AllNighter, y = PoorSleepQuality, fill = AllNighter)) +
  stat_summary(fun = "mean", geom = "bar", position = "dodge") +
  labs(title = "Stress Score by All-Nighter",
  x = "All Nighter",
  y = "Poor Sleep Quality") +
    theme_minimal() +
scale_fill_manual(values = c("All Nighter" = "lightpink", "No All Nighter" = "skyblue"), name = "All Nig
```

Stress Score by All–Nighter
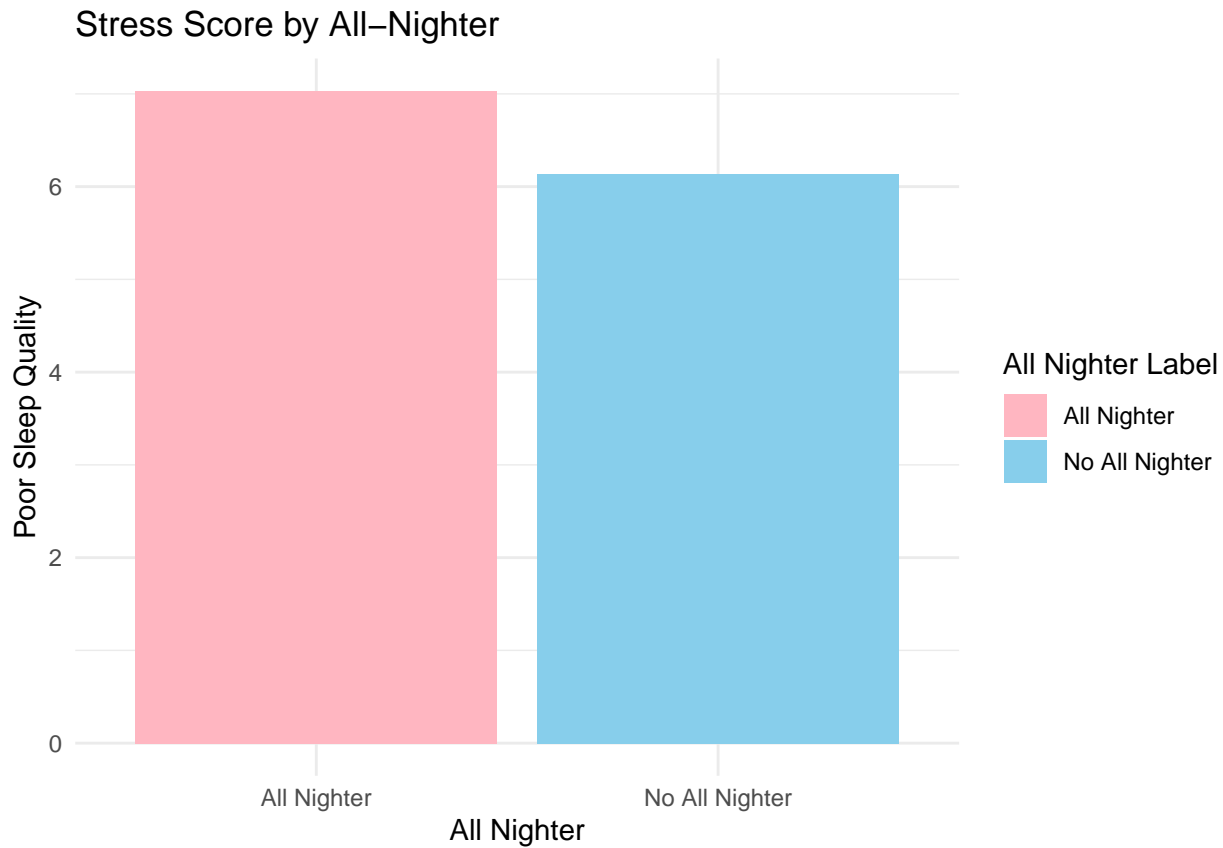
The chi-square statistic is 14.573 with 15 degrees of freedom, resulting in a p-value of 0.4826. The null
hypothesis, which assumes equal population proportions, is tested against the alternative hypothesis that
there is a difference in proportions. The p-value of 0.4826 is larger than the conventional significance level of
0.05, suggesting a lack of evidence to reject the null hypothesis. Therefore, there is no strong indication that
the distribution of "AllNighter" varies significantly across different levels of "PoorSleepQuality." However, by
looking at the diagram above, we can see that students who does not spend an all nighter have a much better
sleep quality in average.

**7. Do students who abstain from alcohol use have significantly better stress scores than those
who report heavy alcohol use?**

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

# Subset the data for abstainers and heavy drinkers
abstainers_data <- SleepStudy %>% filter(AlcoholUse == "Abstain")
heavy_drinkers_data <- SleepStudy %>% filter(AlcoholUse == "Heavy")

# Perform a t-test2
```

```
t_test_result <- t.test(abstainers_data$StressScore, heavy_drinkers_data$StressScore)

# Print the t-test results
print(t_test_result)

##
##  Welch Two Sample t-test
##
## data:  abstainers_data$StressScore and heavy_drinkers_data$StressScore
## t = -0.62604, df = 28.733, p-value = 0.5362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.261170  3.327346
## sample estimates:
## mean of x mean of y
##  8.970588 10.437500
```

```
SleepStudy <- SleepStudy %>% filter(AlcoholUse %in% c("Heavy", "Abstain"))

  SleepStudy <- SleepStudy %>%
  mutate(AlcoholUse = ifelse(AlcoholUse == "Abstain", "AbstainUse", "HeavyUse"))

  # Create a plot with two different fill colors and a legend
ggplot(SleepStudy, aes(x = AlcoholUse, y = StressScore, fill = AlcoholUse)) +
  stat_summary(fun = "mean", geom = "bar") +
  labs(title = "Stress Score by Alcohol Use",
  x = "Use of Alcohol",
  y = "Measure of amount of stress") +
    theme_minimal() +
    scale_fill_manual(values = c("AlcoholUse", "AbstainUse" = "skyblue", "HeavyUse" = "lightpink"), name
```

Stress Score by Alcohol Use
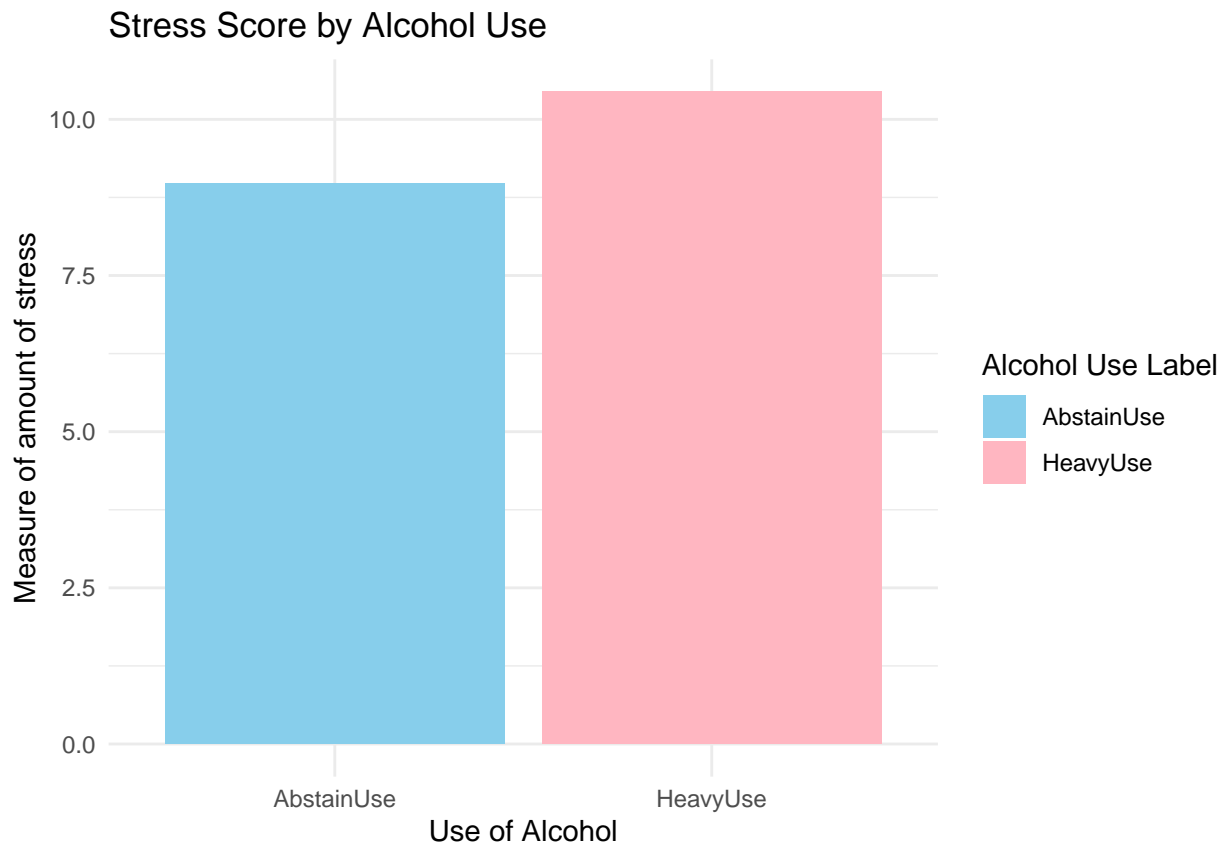
The confidence interval provides a range of values within which you can be 95% confident that the true difference in means lies. In this case, the confidence interval ranges from -6.261170 to 3.327346. This interval includes 0, suggesting that there is a possibility that the true difference in means could be zero.

It can be shown through the plot diagram that students with heavy use of Alcohol have a much significantly higher stress score than students who abstain from alcohol.

8. Is there a significant difference in the average number of drinks per week between students of different genders?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

prop_result <- Prop_test(data = survey, variable = Gender, success = "1", by = Drinks)
```

```
##
## <<< 18-sample test for equality of proportions without continuity correction
##
## variable: Gender
## success: 1
## by: Drinks
##
## --- Description
##
##                    0       1       2       3       4       5       6       7       8       9      10
## -----------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
## n_1              14       2       3       5       2       8       4      10       7       8      20
## n_total          33       9      16      30      18      31      23      22      14      11      26
```

17

```
## proportion      0.424   0.222   0.188   0.167   0.111   0.258   0.174   0.455   0.500   0.727   0.769
##
## --- Inference
##
## Chi-square statistic: 70.646
## Degrees of freedom: 17
## Hypothesis test of equal population proportions: p-value = 0.000
```

```r
print(prop_result)
```

```
##
##  18-sample test for equality of proportions without continuity
##  correction
##
## data:  tbl
## X-squared = 70.646, df = 17, p-value = 0.00000001669
## alternative hypothesis: two.sided
## sample estimates:
##         0         1         2         3         4         5         6         7
## 0.4242424 0.2222222 0.1875000 0.1666667 0.1111111 0.2580645 0.1739130 0.4545455
##         8         9        10        12        13        14        15        18
## 0.5000000 0.7272727 0.7692308 0.8888889 1.0000000 1.0000000 1.0000000 1.0000000
##        20        24
## 1.0000000 1.0000000
```

```r
SleepStudy <- SleepStudy %>%
  mutate(Gender = ifelse(Gender == 1, "Male", "Female"))

ggplot(SleepStudy, aes(x = Gender, y = Drinks, fill = Gender)) +
 stat_summary(fun = "mean", geom = "bar") +
  labs(title = "Average Number of Drinks / Week by Gender",
       x = "Gender",
       y = "Average Num. of Drinks/Week") +
  theme_minimal() +
  scale_fill_manual(values = c("Gender", "Male" = "skyblue", "Female" = "lightpink"), name = "Gender")
```

Average Number of Drinks / Week by Gender

The chi-square statistic is 70.646 with 17 degrees of freedom, resulting in a very low p-value of 0.00000001669. This small p-value leads to rejecting the null hypothesis, suggesting that there is a significant association between gender and the reported number of drinks. The alternative hypothesis indicates a two-sided test, emphasizing that the proportions differ across gender categories in terms of reported drinks. As also shown in the bar diagram, we can see that male students spend on average more number of drinks per week than female students.

9. Is there a significant difference in the average weekday bedtime between students with high and low stress (Stress=High vs. Stress=Normal)?

```
High_stress <- data.frame(SleepStudy %>% filter(grepl('high', SleepStudy$Stress)))

Normal_stress <- data.frame(SleepStudy %>% filter(grepl('normal', SleepStudy$Stress)))

  t_test_result <- t.test(High_stress$WeekdayBed,Normal_stress$WeekdayBed)
  print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  High_stress$WeekdayBed and Normal_stress$WeekdayBed
## t = -1.0746, df = 87.048, p-value = 0.2855
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4856597  0.1447968
## sample estimates:
## mean of x mean of y
```

```
##  24.71500  24.88543
```
```
# Perform the 2-sample t-test for comparing means
mean_result <- ttest(WeekdayBed~Stress, data=survey, alternative="two_sided")
```
```
##
## Compare WeekdayBed across Stress with levels normal and high
## Response Variable:  WeekdayBed, WeekdayBed
## Grouping Variable:  Stress,
##
##
## ------ Describe ------
##
## WeekdayBed for Stress normal:  n.miss = 0,  n = 197,  mean = 24.885,  sd = 1.028
## WeekdayBed for Stress high:  n.miss = 0,  n = 56,  mean = 24.715,  sd = 1.053
##
## Mean Difference of WeekdayBed:  0.170
##
## Weighted Average Standard Deviation:   1.033
##
##
## ------ Assumptions ------
##
## Note: These hypothesis tests can perform poorly, and the
##       t-test is typically robust to violations of assumptions.
##       Use as heuristic guides instead of interpreting literally.
##
## Null hypothesis, for each group, is a normal distribution of WeekdayBed.
## Group normal: Sample mean assumed normal because n > 30, so no test needed.
## Group high: Sample mean assumed normal because n > 30, so no test needed.
##
## Null hypothesis is equal variances of WeekdayBed, homogeneous.
## Variance Ratio test:  F = 1.108/1.056 = 1.049,  df = 55;196,  p-value = 0.792
## Levene's test, Brown-Forsythe:  t = -0.054,  df = 251,  p-value = 0.957
##
##
## ------ Infer ------
##
## --- Assume equal population variances of WeekdayBed for each Stress
##
## t-cutoff for 95% range of variation: tcut =  1.969
## Standard Error of Mean Difference: SE =  0.156
##
## Hypothesis Test of 0 Mean Diff:  t-value = 1.089,  df = 251,  p-value = 0.277
##
## Margin of Error for 95% Confidence Level:  0.308
## 95% Confidence Interval for Mean Difference:  -0.138 to 0.479
##
##
## --- Do not assume equal population variances of WeekdayBed for each Stress
##
## t-cutoff: tcut =  1.988
## Standard Error of Mean Difference: SE =  0.159
##
## Hypothesis Test of 0 Mean Diff:  t = 1.075,  df = 87.048, p-value = 0.286
```

```
## 
## Margin of Error for 95% Confidence Level:  0.315
## 95% Confidence Interval for Mean Difference:  -0.145 to 0.486
## 
## 
## ------ Effect Size ------
## 
## --- Assume equal population variances of WeekdayBed for each Stress
## 
## Standardized Mean Difference of WeekdayBed, Cohen's d:  0.165
## 
## 
## ------ Practical Importance ------
## 
## Minimum Mean Difference of practical importance: mmd
## Minimum Standardized Mean Difference of practical importance: msmd
## Neither value specified, so no analysis
## 
## 
## ------ Graphics Smoothing Parameter ------
## 
## Density bandwidth for Stress normal: 0.407
## Density bandwidth for Stress high: 0.536
```

Two–Group Plot with Means

Compare WeekdayBed for Stress normal and high

Classic t–test of 0 Mean Diff: t = 1.089, df = 251,   p–value = 0.277
95% Confidence Interval for Mean Difference: −0.138 to 0.479
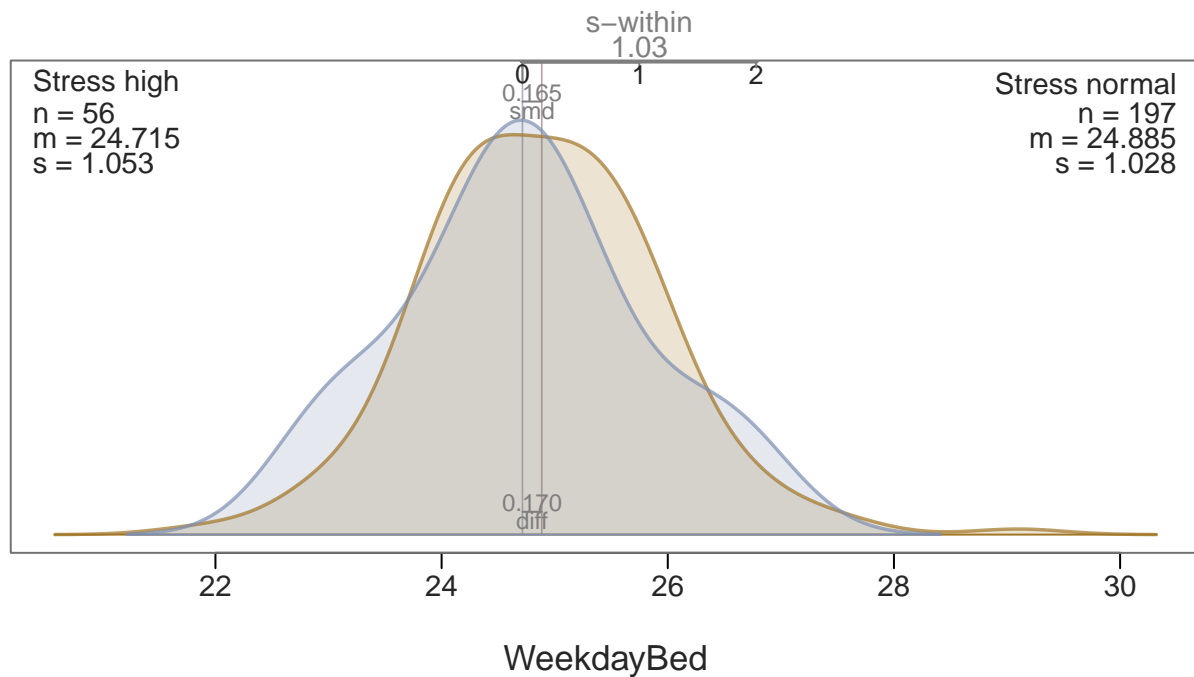


```
print(mean_result)
```

```
## $value1
```

```
## [1] "normal"
##
## $group1
##    [1] 25.75 25.70 27.44 23.50 25.90 25.35 23.90 25.00 25.30 25.60 22.90 24.00
##   [13] 25.77 25.25 25.00 24.63 24.60 24.60 24.40 23.90 25.00 22.90 23.80 25.00
##   [25] 23.80 25.80 24.60 22.83 23.90 24.40 25.40 23.30 24.90 26.60 26.80 25.57
##   [37] 24.12 25.20 23.20 25.30 27.00 25.30 26.13 24.40 23.00 24.55 25.60 24.10
##   [49] 23.10 24.33 24.20 23.80 24.45 25.75 24.40 24.20 23.70 25.60 24.10 24.15
##   [61] 25.00 27.67 25.50 24.40 21.80 24.10 24.05 25.20 25.10 23.90 25.30 24.40
##   [73] 25.10 24.20 26.10 25.40 24.25 24.00 24.60 23.55 23.83 25.20 25.17 26.40
##   [85] 24.70 25.65 25.20 25.80 24.80 25.85 22.60 24.30 25.25 23.75 26.30 24.75
##   [97] 23.80 25.10 26.50 25.90 25.60 25.34 24.63 25.77 25.70 26.60 23.07 25.46
##  [109] 25.80 26.00 26.13 23.90 24.67 24.35 24.75 24.10 24.70 24.80 24.10 24.50
##  [121] 24.90 23.55 25.10 27.30 25.40 24.40 24.90 26.90 26.13 24.70 24.50 26.00
##  [133] 24.30 26.00 24.00 25.95 24.50 24.50 24.30 24.40 25.50 25.75 24.75 25.35
##  [145] 23.70 25.70 25.15 26.70 23.65 24.30 25.00 24.87 25.85 26.50 24.90 25.50
##  [157] 25.05 25.13 25.13 24.20 24.20 25.70 23.73 24.25 24.20 24.09 25.70 26.15
##  [169] 25.25 24.90 25.50 26.10 25.36 29.10 25.90 22.30 24.40 26.30 25.15 24.80
##  [181] 24.80 25.50 24.90 23.70 24.15 24.20 24.90 25.20 24.20 25.70 25.35 25.35
##  [193] 24.00 23.60 24.45 24.38 23.35
##
## $value2
## [1] "high"
##
## $group2
##  [1] 23.80 24.40 26.20 25.92 22.90 24.00 24.50 24.70 24.40 24.05 26.63 25.17
## [13] 24.95 23.65 25.00 25.20 24.50 26.45 23.00 24.40 24.30 24.65 24.25 22.87
## [25] 25.00 25.20 25.27 25.05 25.30 22.83 24.75 24.60 26.50 24.90 23.30 24.10
## [37] 25.45 24.75 23.00 24.27 24.40 26.10 24.75 24.70 23.20 24.95 26.80 25.80
## [49] 25.20 24.10 26.58 23.35 25.40 23.20 24.60 26.75
```

```r
# Bar Diagram
SleepStudy <- SleepStudy %>%
  mutate(Stress = ifelse(Stress == "high", "high", "normal"))

ggplot(SleepStudy, aes(x = Stress, y = WeekdayBed, fill = Stress)) +
 stat_summary(fun = "mean", geom = "bar", position = "dodge") +
  labs(title = "Average Weekday Bedtime by Stress",
       x = "Stress",
       y = "Average Weekday Bedtime") +
  theme_minimal() +
  scale_fill_manual(values = c("Stress", "high" = "skyblue", "normal" = "lightpink"), name = "Stress")
```

Average Weekday Bedtime by Stress
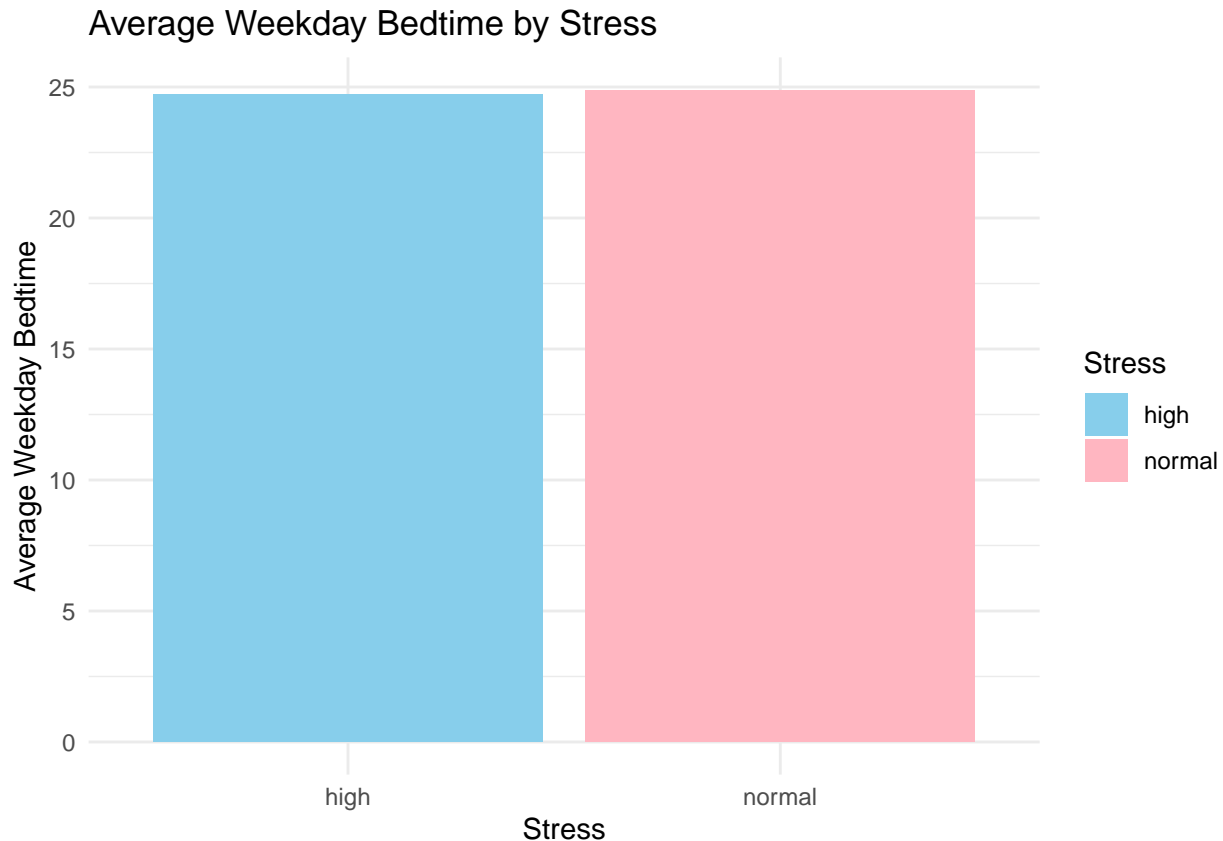
The t-test for the mean difference has a t-value of 1.089 with 251 degrees of freedom and a p-value of 0.277. The 95% confidence interval for the mean difference is -0.138 to 0.479.

There is no strong evidence of a significant difference in weekday bedtime between students with normal and high stress levels. The analysis considers various statistical aspects, assumptions, and provides descriptive and inferential results for a comprehensive understanding of the comparison.

10. Is there a significant difference in the average hours of sleep on weekends between first two year students and other students?

```
SleepStudy = read.csv("https://www.lock5stat.com/datasets3e/SleepStudy.csv")

first_second_year <- data.frame(SleepStudy %>% filter(grepl("^[1-2]", SleepStudy$ClassYear)))
third_fourth_year <- data.frame(SleepStudy %>% filter(grepl("^[3-4]", SleepStudy$ClassYear)))

t_test_result <- t.test(first_second_year$AverageSleep, third_fourth_year$AverageSleep)

print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  first_second_year$AverageSleep and third_fourth_year$AverageSleep
## t = 0.59977, df = 244.21, p-value = 0.5492
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1660165  0.3113800
## sample estimates:
```

```
## mean of x mean of y
##   7.997817   7.925135
```

```
ggplot(SleepStudy, aes(x = ClassYear, y = AverageSleep)) +
  geom_bar(stat = "summary", fun = "mean", fill = "skyblue", color = "black") +
  labs(title = "Average Sleep on Weekends for First two class Year and other Class",
       x = "ClassYear",
       y = "Average Sleep on Weekends") +
  theme_minimal()
```

Average Sleep on Weekends for First two class Year and other Class


```
SleepStudy <- SleepStudy %>%
mutate(ClassYear = ifelse(ClassYear %in% c("1", "2"), "first_second_year", "other_year"))

# Create a plot with two different fill colors and a legend
ggplot(SleepStudy, aes(x = ClassYear, y = AverageSleep, fill = ClassYear)) +
  stat_summary(fun = "mean", geom = "bar") +
  labs(title = "Average Number of early class for first two class year and other year",
  x = "Use of Alcohol",
  y = "Measure of amount of stress") +
    theme_minimal() +
    scale_fill_manual(values = c("ClassYear", "first_second_year" = "skyblue", "other_year" = "lightpink
```

Average Number of early class for first two class year and other year
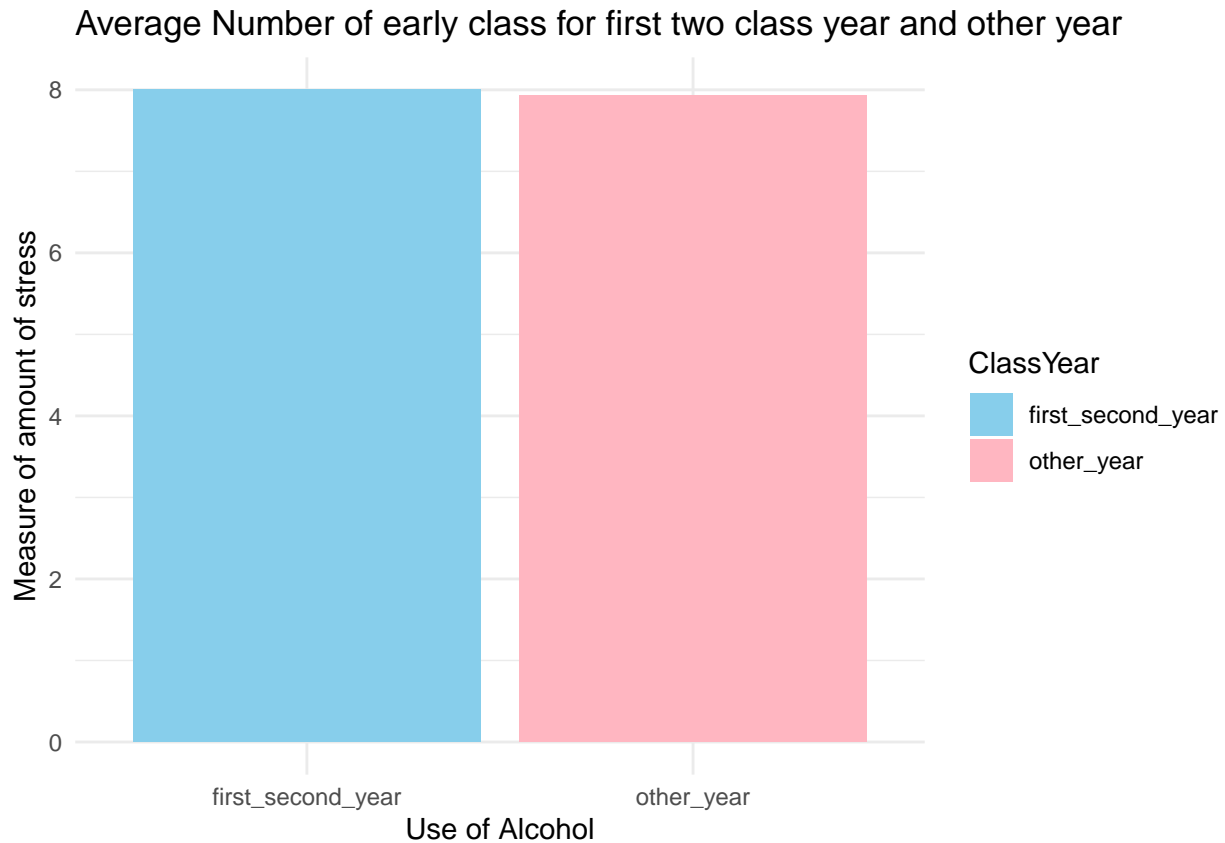
The p-value is 0.5492, which is not very small. A higher p-value suggests weaker evidence against the null hypothesis, indicating that there may not be a significant difference between the two groups.

The confidence interval ranges from -0.1660165 to 0.3113800. This interval includes 0, suggesting that there is a possibility that the true difference in means could be zero.

The t-statistic is relatively small, the p-value is not very small, and the 95% confidence interval includes zero, all of which indicate that the observed difference in means may not be statistically significant.

Summary

The analysis of various factors related to college students' sleep patterns reveals several insights. While gender differences in GPA were not statistically significant, there is a slight variation favoring female students. Early classes significantly impact class attendance, indicating the importance of scheduling considerations. Sleep preferences, as classified by "larks" or "owls," did not show a strong correlation with cognitive skills. Mental health, particularly depression status, significantly influences happiness levels, emphasizing the interplay between mental well-being and emotional states. Sleep quality, as influenced by all-nighters, showed no statistically significant impact, but students without all-nighters tended to have better sleep quality. Heavy alcohol use correlated with higher stress scores, highlighting the potential impact of lifestyle choices on stress levels.

In the study, there wasn't a big difference in grades between males and female students, but having early morning classes made students miss more classes. The way people prefer to sleep didn't seem to affect how smart they are. If someone was moderately depressed, they tended to be less happy, and students who didn't drink alcohol had lower stress. The study gives us a peek into how sleep, mental health, and lifestyle choices connect for college students and affect how they feel overall.

Reference

- The data link: https://www.lock5stat.com/datapage3e.html
- ChatGPT: https://openai.com/
- STAT 353 PPT Notes: https://rpubs.com/scsustat/Stat353PPT