**Statistical Analyses**

***Data Visualization:***

**1.a. Comparative Boxplot of Alcoholic Drinks per Week by Class Standing**

A screenshot of a computer code

AI-generated content may be incorrect.

A chart with different colored squares

AI-generated content may be incorrect.

The comparative boxplot suggests:

* Sophomores have the highest median, followed by juniors and seniors, then freshman. This suggests that sophomores drink more *on average* than other classes.
* Sophomores show the widest spread (highest IQR), suggesting high variability in drinking habits. Seniors have the narrowest IQR, suggesting more consistent drinking habits.
* Upperclassmen (sophomores to seniors) have extreme outliers.
* All in all, **all upperclassmen** (sophomores to seniors) **drink more than freshman**.

**1.b. Comparative Boxplot of Overall GPA by Gender**

A computer code with text

AI-generated content may be incorrect.

A graph of a person and person

AI-generated content may be incorrect.

The comparative boxplot suggests:

* Female students have a slightly higher median GPA than male students.
* Male students show slightly more variability in performance, but females cluster more towards the end.
* One female student has a very low GPA (~2.0), which is unusual given the otherwise strong performance of the group.
* All in all, **females outperform males** in **GPA** at **every quartile**.

***Hypothesis Testing:***

**2.a. Stress Level by Gender**

1) Hypothesis

*μF* = average stress level for female students

*μM* = average stress level for male students

***H₀: μF = μM*** *(no difference in stress)*  
***H₁: μF ≠ μM*** *(difference in stress)*

2) Preparation

* Alpha = 5%
* n > 30
* 𝝈-unknown
* Independent samples (female vs. male students)
* Data Provided

So, a two-tailed, two-sample t-test is performed.

3) Computation & Comparison

A computer screen shot of text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

4) Interpretation

* p-value (0.003) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis**.
* We have **enough statistical evidence** to conclude that the stress level for female and male students are **different**.

**2.b. Weekday Sleep by Gender**

1) Hypothesis

*μF* = average sleep on weekdays for female students

*μM* = average sleep on weekdays for male students

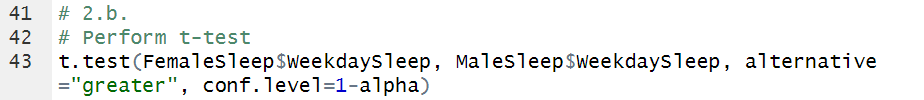
***H₀: μF ≤ μM*** *(females sleep less or equal than males)*  
***H₁: μF > μM*** *(females sleep more than males)*

2) Preparation

* Alpha = 5%
* n > 30
* 𝝈-unknown
* Independent samples (female vs. male students)
* Data Provided

So, a right-tailed, two-sample t-test is performed.

3) Computation & Comparison



A screenshot of a computer

AI-generated content may be incorrect.

4) Interpretation

* p-value (0.154) > alpha (0.05)
* At **α=0.05**, we **fail to reject** the **null hypothesis**.
* We can conclude that there is **no significant evidence** that female students sleep more than male students on weekdays.

***Simple Linear Regression:***

**3. Stress Score Predicted by Anxiety Score**

1) Hypothesis

***H₀:*** *There is* ***no relationship*** *between anxiety score and stress score.*   
***H₁:*** *There is* ***a relationship*** *between anxiety score and stress score.*

2) Preparation

* Alpha = 5%
* Both StressScore (dependent) and AnxietyScore (independent) are numeric.

A computer screen shot of a program

AI-generated content may be incorrect.

A graph with a red line and a line between it

AI-generated content may be incorrect.

The scatterplot suggests **slightly strong positive** **linear** relationship between Stress Score and Anxiety Score.

3) Computation & Comparison

A screenshot of a computer code

AI-generated content may be incorrect.

4) Interpretation

* p-value (<2e-16) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis**.
* We can conclude thatanxiety is a **significant predictor** of stress.

**4. Stress Score Predicted by Depression Score**

1) Hypothesis

***H₀:*** *There is* ***no relationship*** *between depression score and stress score.*   
***H₁:*** *There is* ***a relationship*** *between despression score and stress score.*

2) Preparation

* Alpha = 5%
* Both StressScore (dependent) and DepressionScore (independent) are numeric.

A computer screen shot of a code

AI-generated content may be incorrect.

A graph of a stress score

AI-generated content may be incorrect.

The scatterplot suggests **moderate positive** **linear** relationship between Stress Score and Depression Score.

3) Computation & Comparison



A screenshot of a computer

AI-generated content may be incorrect.

4) Interpretation

* p-value (<2e-16) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis**.
* We can conclude thatdepression is a **significant predictor** of stress.

***Multiple Linear Regression:***

**5. Stress Score Predicted by Happiness, Weekday Sleep and Weekend Sleep**

1) Hypothesis

***H₀: βHappiness = βWeekdaySleep = βWeekendSleep = 0*** *(None of the predictors significantly predict StressScore.)*   
***H₁: At least one β ≠ 0*** *(At least one predictor significantly predicts StressScore.)*

2) Preparation

* Alpha = 5%
* F-test statistic

3) Computation & Comparison

A white background with black text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

4) Interpretation

* Overall p-value (2.665e-08) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis** and there is **at least one predictor** that is **statistically significant**.
* **Happiness**: Significant => p-value (3.84e-09) < alpha
* **WeekdaySleep**: Not significant => p-value (0.213) > alpha
* **WeekendSleep**: Not significant => p-value (0.117) > alpha
* **Happiness significantly predicts stress** while sleep variables do not.

**6. Stress Score Predicted by Anxiety and Depression**

1) Hypothesis

***H₀: βAnxietyScore = βDepressionScore = 0*** *(Neither AnxietyScore nor DepressionScore significantly predict StressScore.)*   
***H₁: At least one β ≠ 0*** *(At least one predictor significantly predicts StressScore.)*

2) Preparation

* Alpha = 5%
* F-test statistic

3) Computation & Comparison

A close-up of a white background

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

4) Interpretation

* Overall p-value (2.2e-16) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis** and there is **at least one predictor** that is **statistically significant**.
* **AnxietyScore**: Significant => p-value (<2e-16) < alpha
* **DepressionScore**: Significant => p-value (4.48e-14) > alpha
* **Both anxiety** and **depression significantly predicts stress** in the model, with anxiety showing a stronger effect.

***Chi Square for Categorical Variables:***

**7. Association between Gender and Alcohol Use**

1) Hypothesis

***H₀:*** *Gender and alcohol use are* ***not associated****.*  
***H₁:*** *Gender and alcohol use are* ***associated****.*

2) Preparation

* Alpha = 5%
* Both Gender and AlcoholUse are categorical.
* Check expected counts ≥ 5 in all cells (for Chi-square validity)

3) Computation & Comparison

A screenshot of a computer program

AI-generated content may be incorrect.

The two-way table suggests:

* Alcohol use patterns vary significantly by gender, with females more likely in the "Light" and "Moderate" groups, while males are slightly more likely to be in the "Heavy" group.

4) Interpretation

* Test Statistic: χ² = 11.96
* p-value (0.007517) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis** and can conclude that there is a **significant association** between gender and alcohol use.

**8. Association between Gender and All-nighters**

1) Hypothesis

***H₀:*** *Gender and all-nighters are* ***not associated****.*  
***H₁:*** *Gender and all-nighters are* ***associated****.*

2) Preparation

* Alpha = 5%
* Both Gender and All-nighters are categorical.
* Check expected counts ≥ 5 in all cells (for Chi-square validity)

3) Computation & Comparison

**A computer screen shot of a computer code

AI-generated content may be incorrect.**

The two-way table suggests:

* There is a **difference** in all-nighter behavior by gender. **Males** show a **higher** frequency of pulling all-nighters than females.

4) Interpretation

* Test Statistic: χ² = 8.57
* p-value (0.0034) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis** and can conclude that there is a **significant association** between gender and all-nighters.

***ANOVA:***

**9. GPA Differences by Class Standing**

1) Hypothesis

***H₀:******μFreshman = μSophomore = μJunior = μSenior*** *(There is* ***no significant difference*** *in mean GPA among class standings.)****H₁: At least one*** *class standing has a* ***significantly******different*** *mean GPA from the others.*

2) Preparation

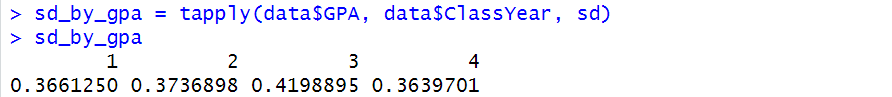
* Variables
  + Independent: ClassYear (categorical, 4 levels)
  + Dependent: GPA (continuous)
* ANOVA Conditions
  + **Normality**: n > 30

A graph of a graph

AI-generated content may be incorrect.

The **histogram** of GPA suggests a reasonably normal distribution. Since our sample size is 40 (**n > 30**), we assume normality by the **Central Limit Theorem (CLT).**

* + **Independence:** Assume random sampling
  + **Homogeneity/ Equality of Variances**



A computer screen shot of a code

AI-generated content may be incorrect.

To check the assumption of homogeneity of variances, we verified that no group’s standard deviation exceeded **twice** that of another group. Since this condition was met (FALSE result from the test), we assume **equal variances** and proceed with ANOVA.

3) ANOVA Test

A **one-way ANOVA** is conducted to compare the mean GPA among the class years.

A screenshot of a computer

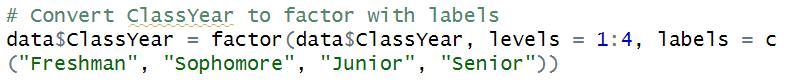
AI-generated content may be incorrect.

ANOVA Test Interpretation

* p-value (2.91e-07) < alpha (0.05)
* At **α=0.05**, we **reject** the **null hypothesis** and can conclude that there is **at least one** class standing has a **significantly different** mean GPA from the others.

4) Post Hoc Analysis (Tukey’s HSD Test)

Since the ANOVA test showed significance, we perform **Tukey’s HSD** to determine **which groups differ**.



A screenshot of a computer code

AI-generated content may be incorrect.

Tukey’s HSD Test Interpretation

* 3 out of 6 groups - **Sophomore vs. Freshman** (p = 0.0000001), **Junior vs. Freshman** (p = 0.00028), **Senior vs. Freshman** (p = 0.00058) - show a significant difference.

5) Conclusion

* **Freshmen** GPAs **significantly lower** than Sophomores (−0.40), Juniors (−0.31), and Seniors (−0.30)
* **Upperclassmen** (Sophomore–Senior) show **no significant differences**.

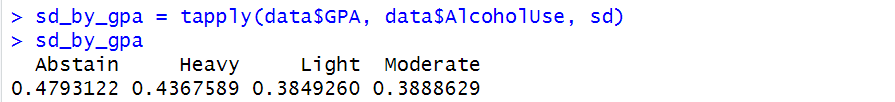
**10. GPA Differences by Alcohol Use**

1) Hypothesis

***H₀:******μAbstain = μLight = μModerate = μHeavy*** *(There is* ***no significant difference*** *in mean GPA across alcohol use groups.)****H₁: At least one*** *alcohol use group has a* ***significantly******different*** *mean GPA from the others.*

2) Preparation

* Variables
  + Independent: AlcoholUse (categorical, 4 levels)
  + Dependent: GPA (continuous)
* ANOVA Conditions
  + **Normality**: n > 30 (Checked normality for GPA in question 9)
  + **Independence:** Assume random sampling
  + **Homogeneity/ Equality of Variances**



A screenshot of a computer code

AI-generated content may be incorrect.

To check the assumption of homogeneity of variances, we verified that no group’s standard deviation exceeded **twice** that of another group. Since this condition was met (FALSE result from the test), we assume **equal variances** and proceed with ANOVA.

3) ANOVA Test

A **one-way ANOVA** is conducted to compare the mean GPA among the alcohol use groups.

A close-up of numbers

AI-generated content may be incorrect.

ANOVA Test Interpretation

* p-value (0.299) > alpha (0.05)
* At **α=0.05**, we **fail to reject** the **null hypothesis** and can conclude that there is **no significant difference** in mean GPA across alcohol use groups.