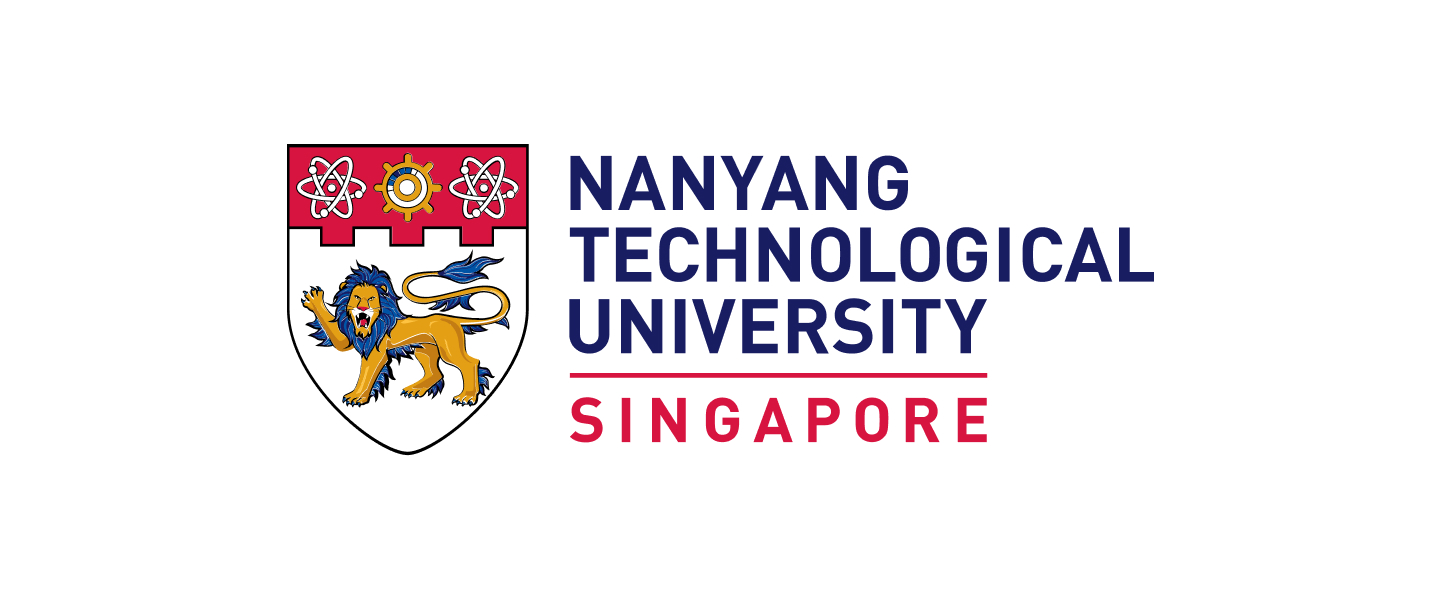
****

**MH3511 Data Analysis with Computer**

**Group Project**

A Study on the Factors that influence Sleep Efficiency

***Abstract:***

***Sleep efficiency is a vital determinant of sleep quality and is associated with various detrimental health outcomes such as obesity and impaired cognitive function. Furthermore, the study depicts a statistical assessment that endeavours to investigate the correlation between sleep efficiency and multiple variables including age, gender, duration of sleep, lifestyle habits, and exercise frequency. The analysis aims to elucidate the links between sleep efficiency and various factors and aims to offer valuable insights for enhancing sleep quality and overall health. The conclusions from this analysis can be employed to induce targeted interventions focused on improving sleep efficiency and promoting superior health outcomes.***

# Content Page

1. Introduction ……………………………………………………………………………… 3
2. Data Description ………………………………………………………………………… 3
3. Description and Cleaning of Dataset …………………………………………………. 4

3.1 Summary statistics for the main variable of interest, sleep efficiency …… 4

3.2 Summary statistics for other variables ………………………………………. 5

3.2.1 Age ………………………………………………………………………… 5

3.2.2 Rem Sleep Percent …..…………………………………………………. 5

3.2.3 Deep Sleep Percent …..…………….…………………………………. 5

3.2.4 Light Sleep Percent …...…………………………………………….…. 5

3.2.5 Gender …………….………………………………………………………. 6

3.2.6 Bedtime …………….……………………………………………………… 6

3.2.7 Sleep duration ………….…………………………………………………. 6

3.2.8 Caffeine Consumption …….……………………………………………… 7

3.2.9 Alcohol Consumption ……….…………………………………………… 7

3.2.10 Smoking Status …………….……………………………………………. 7

3.2.11 Exercise Frequency ………….………………………………………….. 7

1. Statistical Analysis ……………………………………………………………………….. 8

4.1 Relation between sleep efficiency and age …………………………………... 8

4.2 Relation between sleep efficiency and bedtime ……………………………… 10

4.3 Relation between sleep efficiency and sleep duration ………………………. 11

4.4 How does smoking affect sleep efficiency in each gender? ………………… 11

4.5 How does caffeine affect sleep efficiency in general? ………………………..13

4.6 How does exercise frequency affect sleep efficiency in each gender? ……..13

4.7 How does alcohol consumption affect sleep efficiency in each gender? …...15

4.8 Is there a correlation between sleep efficiency and three sleep stages? …..16

1. Conclusion and Discussion ……………………………………………………………….18

## Introduction

Sleep efficiency serves as a crucial indicator of the quality of an individual's sleep, and as a result, it is a vital predictor of overall health and well-being. According to studies, poor sleep efficiency is linked to a plethora of negative outcomes, including an elevated risk of chronic illnesses like obesity, diabetes, and heart disease, as well as impaired cognitive function and mood disorders.

The objective of this statistical analysis is to investigate the relationship between sleep efficiency and a variety of factors, including age, gender, sleep duration, caffeine and alcohol consumption, smoking status, and exercise frequency, to name a few. By utilising this dataset, our goal is to address the following queries:

1. Is there a distinction in sleep efficiency across different age groups and genders?
2. Is there any correlation between sleep efficiency and the three sleep phases?
3. Does the bedtime affect sleep efficiency?
4. Does the sleep duration impact sleep efficiency?
5. Do individuals with different lifestyles, such as smoking status, alcohol and caffeine consumption and exercise frequency, exhibit distinct sleep efficiency?

Through analysing the data and identifying significant correlations between these variables and sleep efficiency, this report aims to offer insights into the factors that contribute to healthy sleep and determine potential areas for intervention to enhance sleep quality. This report can benefit individuals, healthcare providers, and researchers by providing an improved comprehension of the factors that affect sleep efficiency and offering potential strategies for enhancing sleep quality and overall health. The outcomes of this analysis can be utilised to inform personalised interventions, such as behavioural therapies or lifestyle improvements, to improve sleep efficiency and promote better health outcomes.

## Data Description

The dataset being used is obtained from [Kaggle](https://www.kaggle.com/code/hexenmeiser/sleep-efficiency-dataset-eda-and-scoring/input). The dataset consists of 452 observations with 15 variables. Before conducting our exploratory analysis, we first performed a preliminary data cleaning to ensure the data is ready to be analysed.

* We made sure that the total percentage of the three sleep stages is equal to 100 for each record. If it is not equal to 100 then we will drop the specific row.
* We dropped irrelevant columns:
  + ID
  + Wakeup time - sleep duration and bedtime variables already exists in data
  + Awakening - it is irrelevant to the analysis
* We converted the datetime in bedtime into “%H.%M” format as we are only concerned with the time each individual sleeps. After converting to the desired format, we categorised the wakeup time into three categories, which are early, late and very late respectively.

After dropping irrelevant columns, we have a total of 452 observations with 13 variables:

1. Age - age of the individual
2. REM sleep percentage - the percentage of total sleep time spent in REM sleep
3. Deep sleep percentage - the percentage of total sleep time spent in deep sleep
4. Light sleep percentage - the percentage of total sleep time spent in light sleep
5. Gender - male or female
6. Bedtime - the time individual goes to bed each night (in HH.MM)
7. Sleep duration - the total amount of time the individual slept (in hours)
8. Sleep efficiency - a measure of the proportion of time in bed spent asleep
9. Caffeine consumption - the amount of caffeine consumed in the 24 hours prior to bedtime (in mg)
10. Alcohol consumption - the amount of alcohol consumed in the 24 hours prior to bedtime (in oz)
11. Smoking status - whether or not the individual smokes
12. Exercise frequency - the number of times the individual exercises each week

## Description and Cleaning of Dataset

We looked into each of the variables closely to check if there are any outliers to be removed, null values to be filled and any highly skewed data to be transformed.

**3.1 Summary statistics for the main variable of interest, sleep efficiency.**

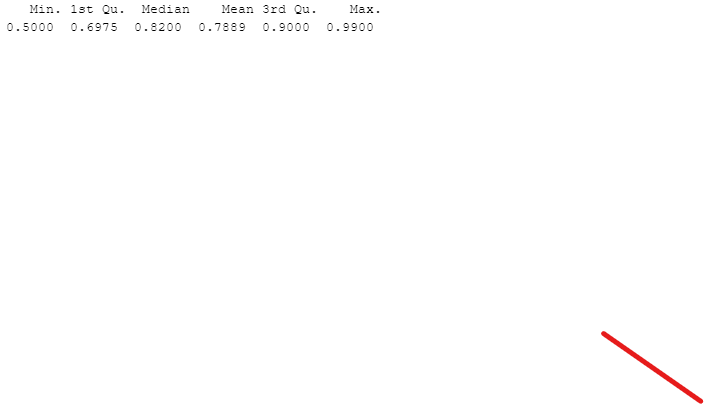
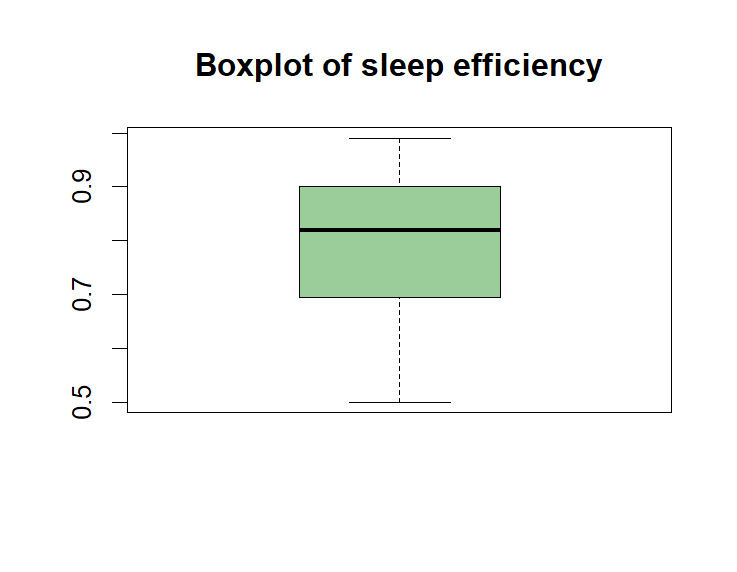
|  |  |  |
| --- | --- | --- |
|  |  |  |

The sleep efficiency data seems to be left-skewed. Thus, we performed log and square root transformations to attempt to reduce skewness. However, the results seem to make no difference and thus we decided to stick with the original data.

We then plotted QQ-plot and performed a shapiro-wilk test to check if sleep efficiency is normally distributed and the results showed that it is not. Thus, for the following analysis, we will use non-parametric tests instead.

|  |  |
| --- | --- |
|  | Shapiro-Wilk normality test  data: sleep$Sleep.efficiency  W = 0.91702, p-value = 4.633e-15 |

We then plotted the boxplot to check if there are any outliers. No outliers identified.



**3.2 Summary statistics for other variables**

3.2.1 Age

|  |  |  |
| --- | --- | --- |
|  |  | * No outlier is removed. * Age ranges from 9 to 69. |

3.2.2 REM sleep percentage

|  |  |  |
| --- | --- | --- |
|  |  | * There are no outliers in the data. |

3.2.3 Deep sleep percentage

|  |  |  |
| --- | --- | --- |
|  |  | * We notice that there is one outlier in the data. |

3.2.4 Light sleep percentage

|  |  |  |
| --- | --- | --- |
|  |  | * There are no outliers in the data. |

3.2.5 Gender

|  |  |
| --- | --- |
|  | * No outlier is removed. |

3.2.6 Bedtime

|  |  |
| --- | --- |
|  | * No outlier is removed. |

3.2.7 Sleep duration

|  |  |
| --- | --- |
|  | * The number of records for sleep duration of 5 and 10 are 8. * The number of records for sleep duration of 5.5 is 3. * The number of records are relatively low so we treat them as outliers and remove them. |

3.2.8 Caffeine consumption

|  |  |  |
| --- | --- | --- |
| Before filling NA | After filling NA | * No outlier is removed * Caffeine consumption contains NULL values which was filled with the median * After filling the NA records with median 25 (as data is skewed), the number of records for the different levels of caffeine consumption is as follows. |

3.2.9 Alcohol consumption

|  |  |  |
| --- | --- | --- |
| Before filling NA | After filling NA | * No outlier is removed * Alcohol consumption contains NULL values which was filled with the median |

3.2.10 Smoking status

|  |  |
| --- | --- |
|  | * There are no outliers in the data * There are no NA values in the data |

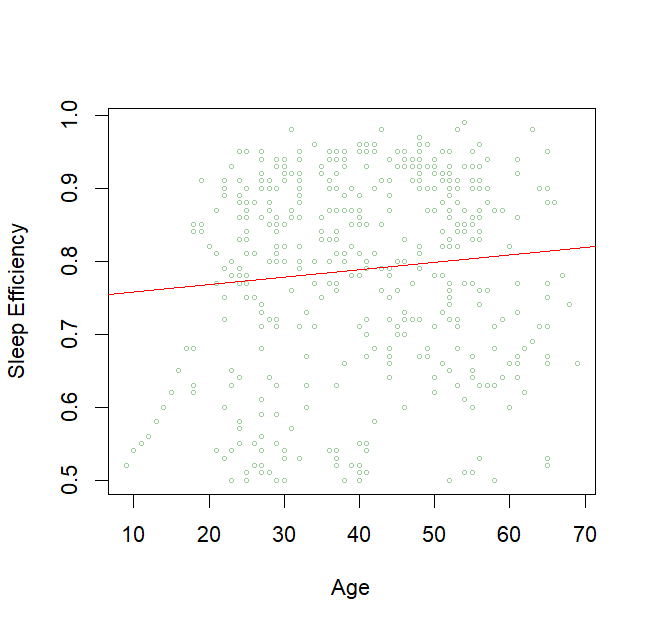
3.2.11 Exercise frequency

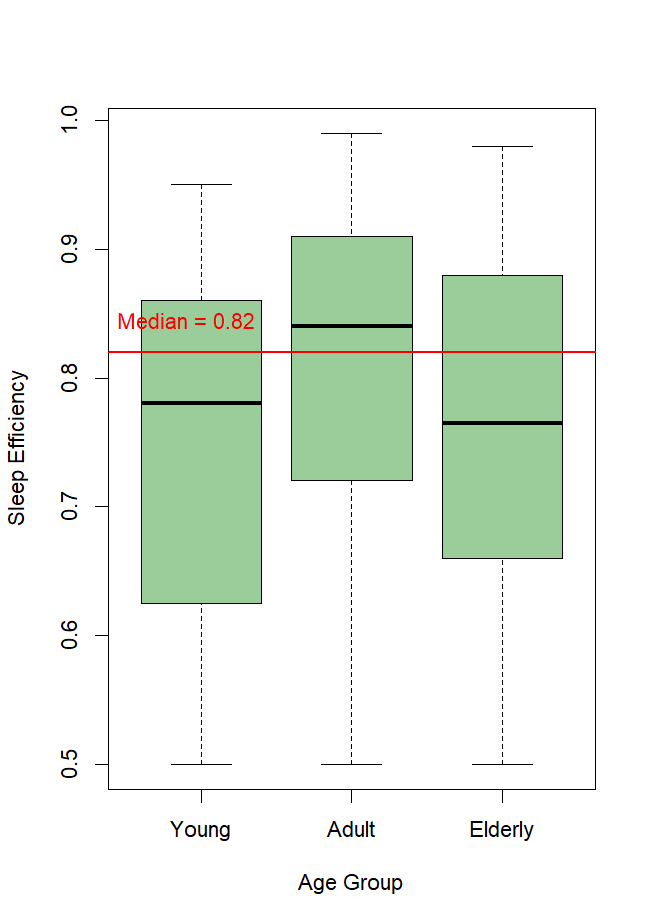
|  |  |  |
| --- | --- | --- |
| Before filling NA | After filling NA | * No outlier is removed * Exercise frequency contains NULL values which was filled with the median |

## Statistical Analysis

**4.1 Relation between Sleep Efficiency and Age**

In this section, we will discuss the relationship between age and sleep efficiency.



We commonly associate older age with lower sleep efficiency, however the scatterplot suggests that there is in fact no distinct linear relationship between age and sleep efficiency. Age and sleep efficiency also has a correlation coefficient of 0.0983 which shows weak correlation. 

On further analysis, we compared the sleep efficiency across different age groups.

* Young (age<=25)
* Adult (25<age<55)
* Elderly(age>=55)

The box plot shows that the Elderly has the lowest median sleep efficiency whereas the Adult has the highest median sleep efficiency. The sleep efficiency of the Young comes in second place. Median is used as a basis of comparison as Sleep Efficiency shows an asymmetric distribution.

The observations can be attributed to…

1. The natural process of ageing. Through secondary research, we found that sleep efficiency tend to decrease with ageing (Li et al., 2018).
2. The research mentioned that sleep efficiency decreases significantly with age in adulthood. The threshold set to differentiate the Adult and Elderly age group could have been when the transition took place which would explain the drastic fall in sleep efficiency when comparing the 2 age groups.

We investigated whether the data from different age groups are identically distributed.

|  |  |
| --- | --- |
| *H0*: All the three age groups gives the same distribution of sleep efficiency.  *H1*: At least one age group gives a different distribution of sleep efficiency. | Kruskal-Wallis rank sum test  data: AgeSedf$Sleep.efficiency and AgeSedf$AgeGroup  Kruskal-Wallis chi-squared = 14.923, df = 2, p-value = 0.0005749 |

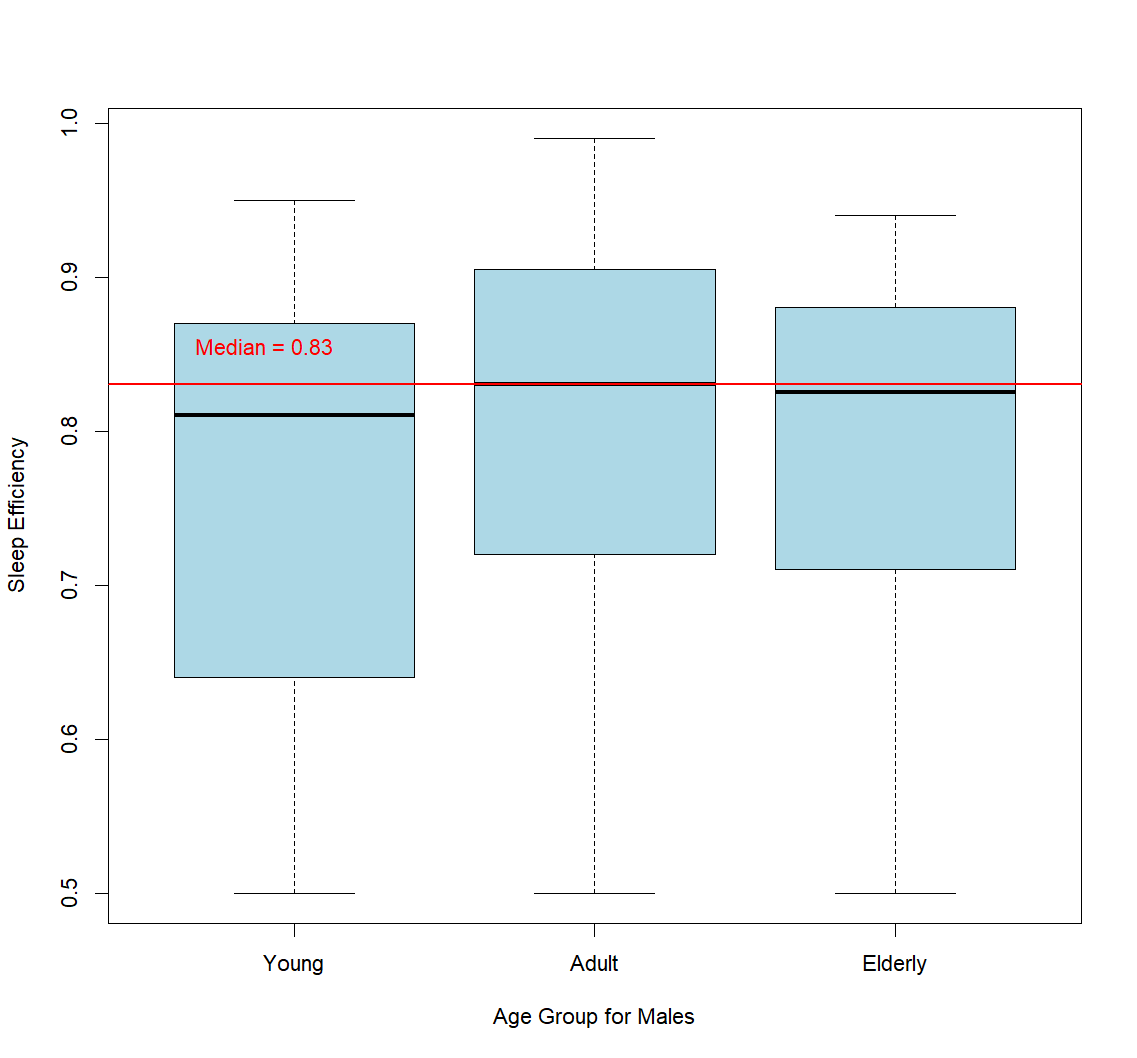
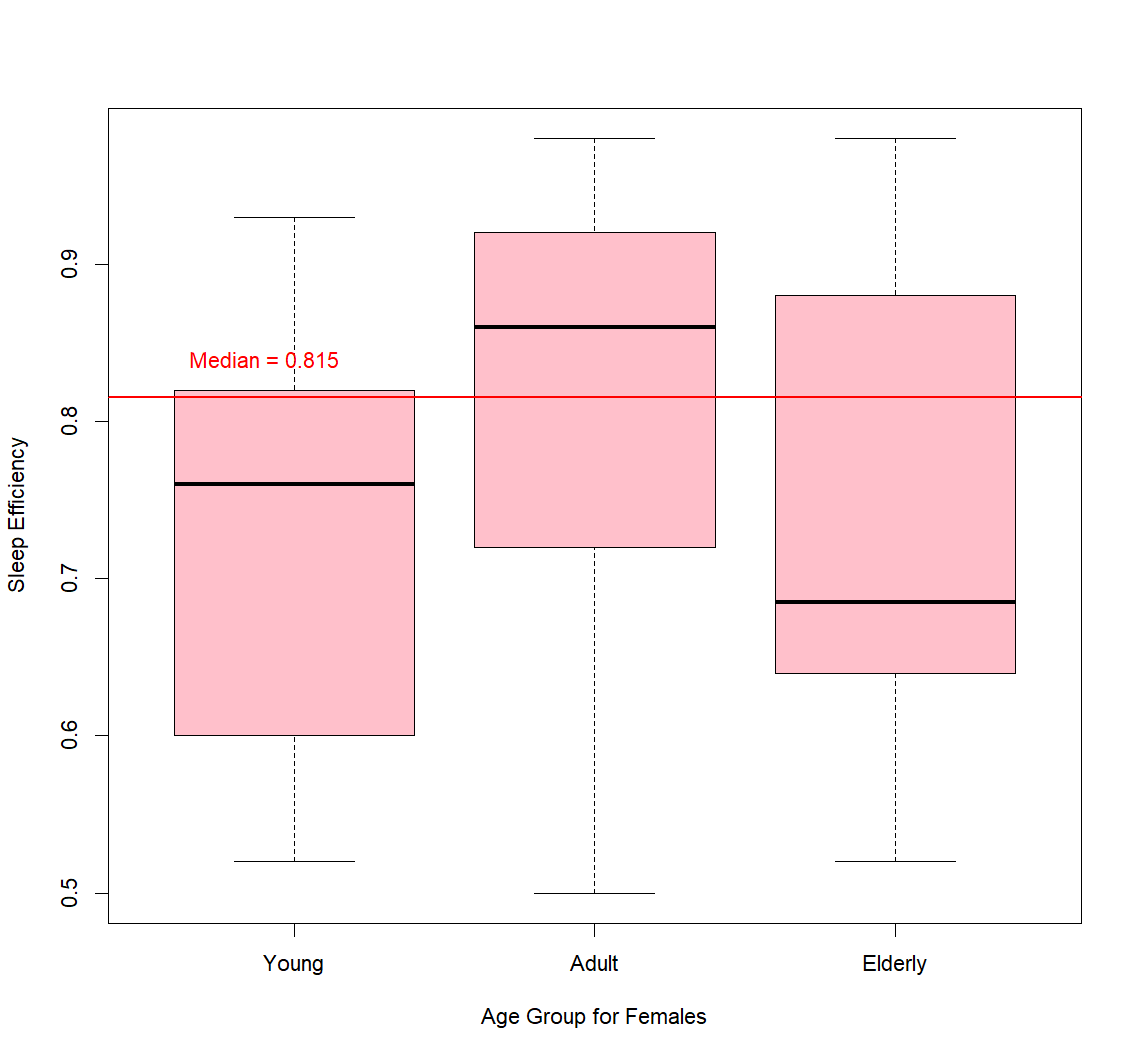
At a significance level of 0.05, we rejected *H0*. Thus, we concluded that age groups do not affect sleep efficiency.

We further conducted a pairwise wilcoxon test to investigate which age groups give different distributions.

|  |
| --- |
| Pairwise comparisons using Wilcoxon rank sum test with continuity correction  data: AgeSedf$Sleep.efficiency and AgeSedf$AgeGroup  Adult Elderly  Elderly **0.02204** -  Young **0.00051** 0.27564  P value adjustment method: none |

At significance level of 0.05, there is sufficient evidence to reject the hypothesis that the Elderly and Adult, Young and Adult have identical distributions. We concluded that Adulthood affects sleep efficiency extensively which may be possibly due to it being the phase of life when people transitioned from Young and transition to Elderly.

Will age affect the sleep efficiency of different genders differently?



For females, we observed that it has a similar ranking from the overall data. Adult has the best median sleep efficiency followed by the Young and lastly the Elderly. Since the distributions for all 3 age groups are relatively different, we investigate only the pairwise distributions.

|  |
| --- |
| Pairwise comparisons using Wilcoxon rank sum test with continuity correction  data: femaleAge$Sleep.efficiency and femaleAge$AgeGroup  Adult Elderly  Elderly 0.0540 -  Young **0.0013** 0.7611  P value adjustment method: none |

At significance level of 0.05, there is sufficient evidence to reject the hypothesis that the Young and Adult have identical distributions. Hence, we conclude that age affects females’ sleep efficiency. This could also be attributed to transitioning to adulthood whereby females may find stability in career, childcare, home etc.

For males, however, while adults continue to have the best sleep efficiency, it is observed that the Elderly have a higher median sleep efficiency than the Young. Furthermore, the medians and the distributions are similar, hence we conducted Kruskal-Wallis test to check if they are indeed identically distributed.

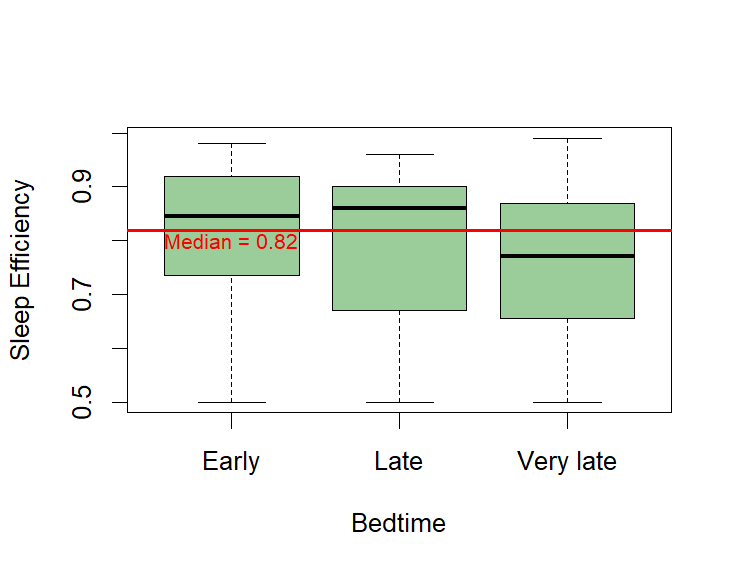
|  |  |
| --- | --- |
| *H0*: All the three age groups give the same distribution of sleep efficiency for males.  *H1*: At least one age group gives a different distribution of sleep efficiency for males. | Kruskal-Wallis rank sum test  data: maleAge$Sleep.efficiency and maleAge$AgeGroup  Kruskal-Wallis chi-squared = 3.1638, df = 2, p-value = 0.2056 |

At significance level of 0.05, there is insufficient evidence to reject the hypothesis that the distribution of sleep efficiency across the age groups for males are identically distributed. Hence, age may not be a strong factor affecting sleep efficiency in males.

**4.2 Relation between sleep efficiency and bedtime**

In this section, we wanted to investigate whether sleep efficiency depends on what time the individual goes to bed.

The following boxplot illustrated the distribution of sleep efficiency across different bedtimes.



The boxplot shows that the distributions look similar. Thus, to determine that, Kruskal-Wallis test is conducted to test the similarity of distribution among the three bedtime.

|  |  |
| --- | --- |
| *H0*: All the three bedtime gives the same distribution of sleep efficiency;  *H1*: At least one bedtime gives a different distribution of sleep efficiency. | Kruskal-Wallis rank sum test  data: sleep$Sleep.efficiency and sleep$BedtimePeriod  Kruskal-Wallis chi-squared = 14.454, df = 2, p-value = 0.0007267 |

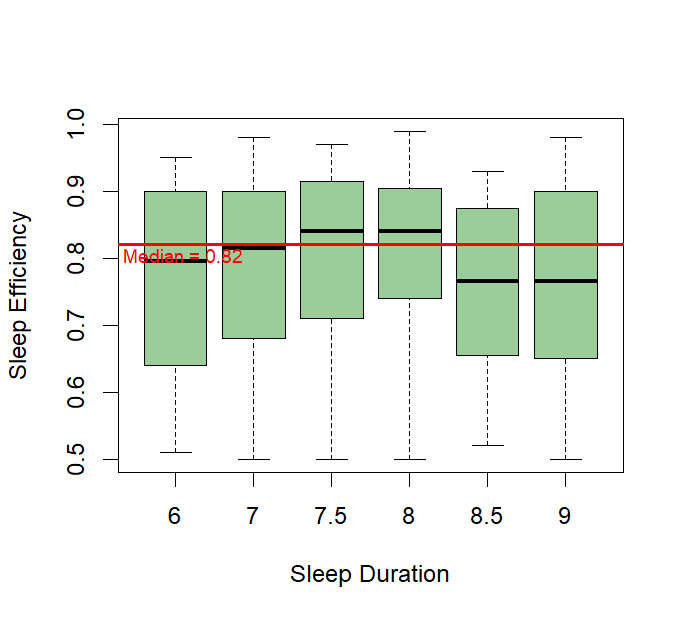
The Kruskal-Wallis test returns a p-value of 0.0007 , indicating that we need to reject the null hypothesis at a significance level of 0.05, where the distributions of three populations are similar. We further conducted pairwise comparisons to investigate which population has a different distribution.

|  |
| --- |
| Pairwise comparisons using Wilcoxon rank sum test with continuity correction  data: sleep$Sleep.efficiency and sleep$BedtimePeriod  Late Very late  Very late **0.04540** -  Early 0.16829 **0.00016**  P value adjustment method: none |

We can observe that the very late bedtime period has a different distribution with both early and late bedtime periods whereas the distribution between early and late is identical. Therefore, we concluded that sleep efficiency varies during the bedtime period. Based on the boxplot, individuals who sleep very late have a higher tendency to experience lower sleep efficiency.

**4.3 Relation between sleep efficiency and sleep duration**

In this section, we tried to answer the question “Is sleep efficiency affected by sleep duration?” The following boxplot illustrated the distribution of sleep efficiency across different sleep duration.



We then conducted Kruskal-Wallis test to determine if the distribution of sleep efficiency is identical for each sleep duration.

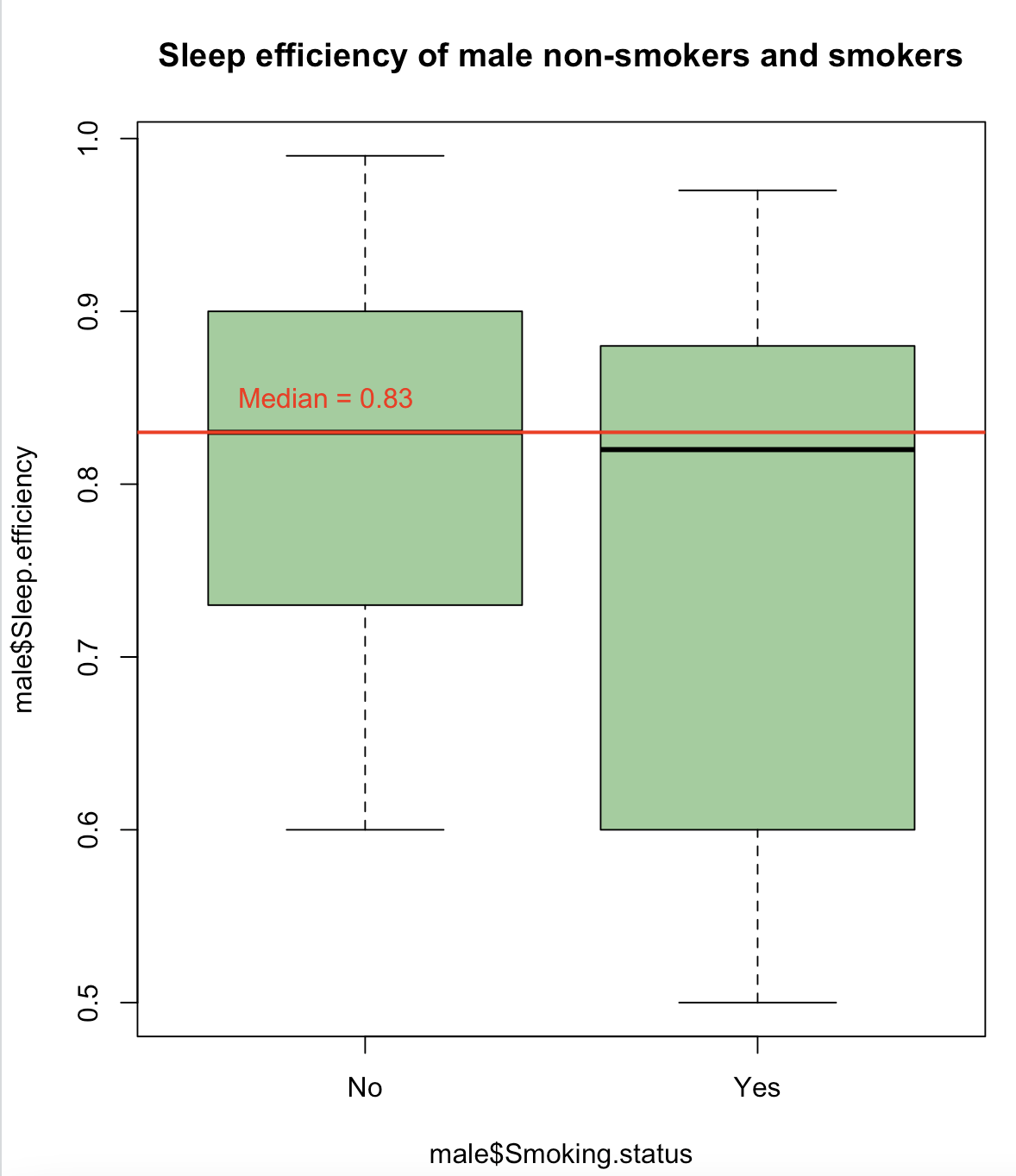
|  |  |
| --- | --- |
| *H0*: All sleep duration gives the same distribution of sleep efficiency;  *H1*: At least one sleep duration gives a different distribution of sleep efficiency. | Kruskal-Wallis rank sum test  data: trimmed.sleep$Sleep.efficiency and trimmed.sleep$Sleep.duration  Kruskal-Wallis chi-squared = 6.7531, df = 5, p-value = 0.2397 |

The Kruskal-Wallis test returns a p-value of 0.2397, indicating that the distributions are not significantly different at a significance level of 0.05. Therefore, we concluded that sleep efficiency is independent of individual sleep duration.

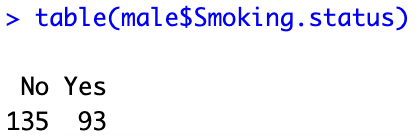
**4.4 Relation between sleep efficiency and smoking status**

We wanted to investigate if smoking affects sleep efficiency in male and females differently.

The following is an illustration of the distribution of male smokers and non-smokers using the

box plot.

The mean for both male smokers and non-smokers are below the median.



Using CLT, we used the t test since the sample size is more than 30 for each group.

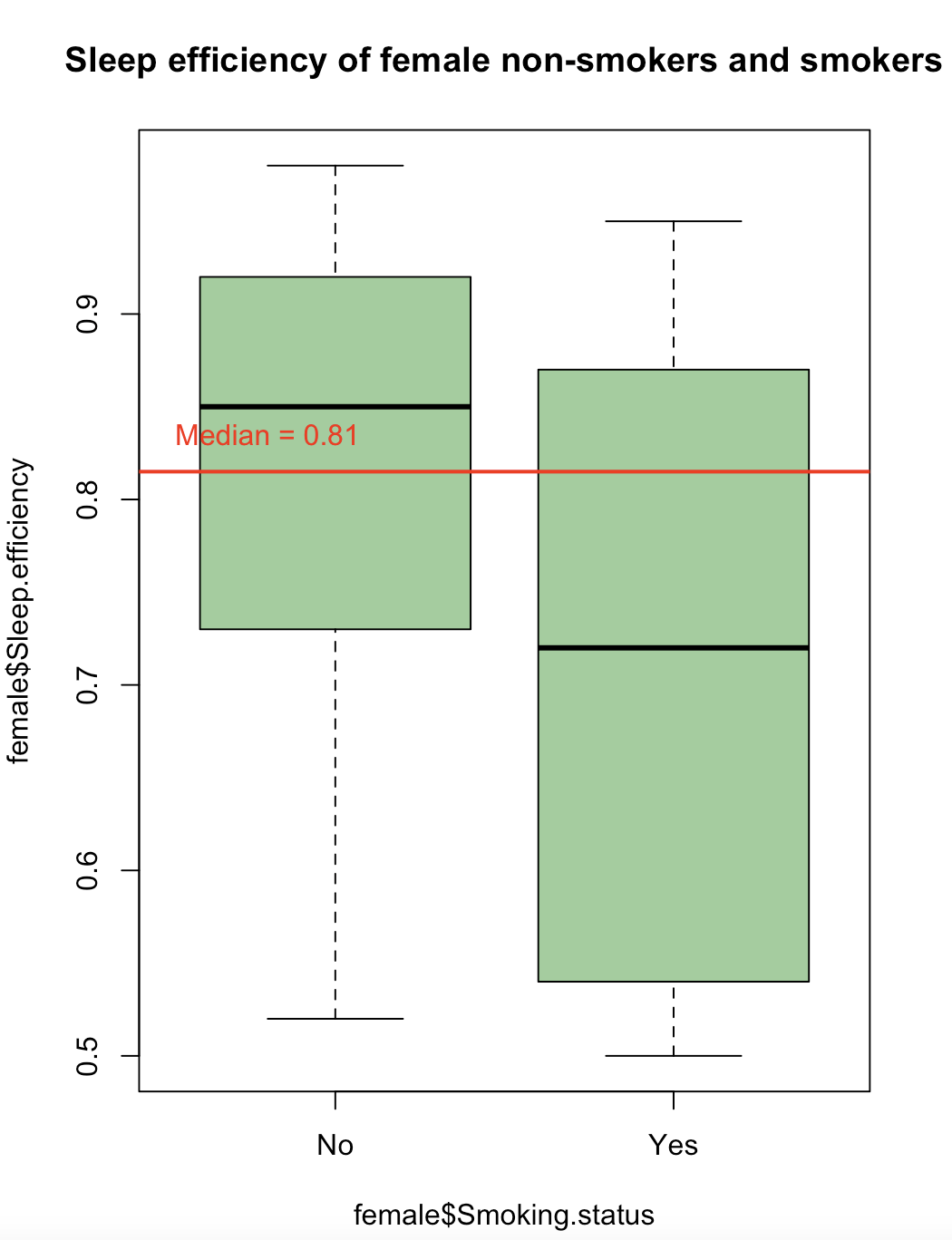
*H0*: Mean of sleep efficiency of male smokers = the mean of sleep efficiency of male non smokers

H1 : Mean of sleep efficiency of male smokers ≠ the mean of sleep efficiency of male non smokers

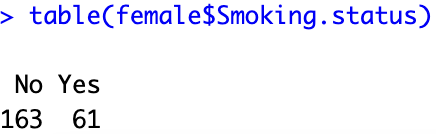
We conducted a variance test to determine if the two populations are having the same variance and with the results obtained, we used the t-test to compare the mean of these two independent samples.

|  |  |
| --- | --- |
| F test to compare two variances  data: maleSmoke$Sleep.efficiency and maleNoSmoke$Sleep.efficiency  F = 2.4985, num df = 92, denom df = 134, p-value = 1.318e-06  alternative hypothesis: true ratio of variances is not equal to 1  95 percent confidence interval:  1.724727 3.668163  sample estimates:  ratio of variances  2.49849 | Welch Two Sample t-test  data: maleSmoke$Sleep.efficiency and maleNoSmoke$Sleep.efficiency  t = -3.2794, df = 142.3, p-value = 0.001308  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  -0.09618198 -0.02383714  sample estimates:  mean of x mean of y  0.7547312 0.8147407 |

We do not reject the null hypothesis that the mean of sleep efficiency of male smokers is lesser than or equal to the mean of sleep efficiency of male non-smokers.



The mean for female non-smokers is lower than its median and the mean for female smokers is higher than its median.



Using CLT, we used the t test since the sample size is more than 30 for each group.

*H0*: Mean of sleep efficiency of female smokers = the mean of sleep efficiency of female non smokers

*H1*: Mean of sleep efficiency of female smokers ≠ the mean of sleep efficiency of female non smokers

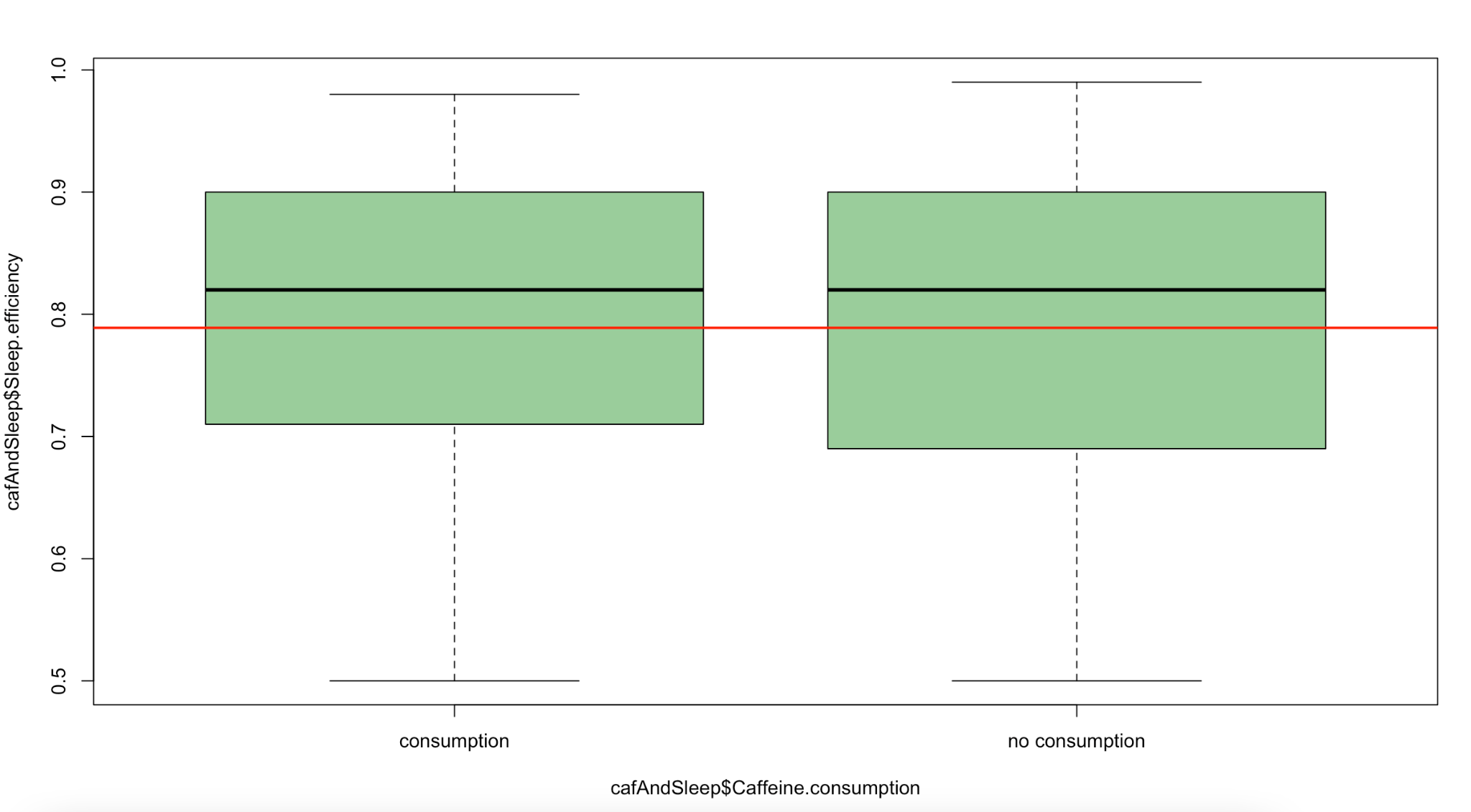
We used the variance test to determine whether the variances are equal and the t-test to compare the mean of these two independent samples.

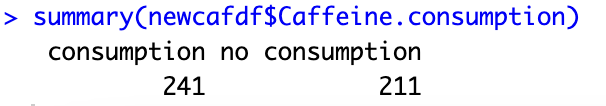
|  |  |
| --- | --- |
| F test to compare two variances  data: femaleyesSmoke$Sleep.efficiency and femalenoSmoke$Sleep.efficiency  F = 1.9962, num df = 60, denom df = 162, p-value = 0.0006527  alternative hypothesis: true ratio of variances is not equal to 1  95 percent confidence interval:  1.336242 3.107989  sample estimates:  ratio of variances  1.996225 | Welch Two Sample t-test  data: femaleyesSmoke$Sleep.efficiency and femalenoSmoke$Sleep.efficiency  t = -4.9826, df = 83.518, p-value = 3.343e-06  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  -0.16170740 -0.06944417  sample estimates:  mean of x mean of y  0.7034426 0.8190184 |

We do not reject the null hypothesis that the mean of sleep efficiency of female smokers is lesser than or equal to the mean of sleep efficiency of female non-smokers.

**4.5 Relation between sleep efficiency and caffeine consumption**

Since the number of records for 75,100 and 200 are less we re-categorised the data into consumption and no consumption.





The box plot for consumption and no consumption.

We performed the non-parametric wilcoxon test on the dataset to test if the distributions for sleep efficiency for consumption and no consumption of caffeine is the same.

|  |  |
| --- | --- |
| *H0*: Distribution of sleep efficiency with caffeine consumption = Distribution of sleep efficiency with no caffeine consumption  *H1*: Distribution of sleep efficiency with caffeine consumption is not the same as the distribution of sleep efficiency with no caffeine consumption | Wilcoxon rank sum test with continuity correction  data: nocons$Sleep.efficiency and cons$Sleep.efficiency  W = 25082, p-value = 0.8041  alternative hypothesis: true location shift is not equal to 0 |

Since the p-value is greater than 0.05, we will not reject the null hypothesis that the distributions are the same. Hence, there is no difference between caffeine consumption and no caffeine consumption on sleep efficiency.

**4.6 Relation between sleep efficiency and exercise frequency**

Then we perform Kruskal-Wallis test with the cleaned dataset for male and female groups.

*H0*: All groups with different exercise frequency give the same distribution of sleep efficiency.

*H1*: At least one group gives a different distribution of sleep efficiency

|  |  |
| --- | --- |
| Exercise Frequency for Males | Exercise Frequency for Females |
| Kruskal-Wallis rank sum test  data: male$Sleep.efficiency and male$Exercise.frequency  Kruskal-Wallis chi-squared = 21.086, df = 4, p-value = 0.0003045 | Kruskal-Wallis rank sum test  data: female$Sleep.efficiency and female$Exercise.frequency  Kruskal-Wallis chi-squared = 32.304, df = 5, p-value = 5.172e-06 |

Given the p-value is extremely low, we reject the null hypothesis that different exercise frequency levels give the same distribution of sleep efficiency.

Afterwards, we make a plot with Exercise.frequency against Sleep.efficiency for males and females. We can see that for both, there is a positive correlation between Exercise.frequency and Sleep.efficiency.

|  |  |
| --- | --- |
| Scatterplot for Male Group | Scatterplot for Female Group |
|  |  |

From the boxplots below, we can clearly see that for both genders, people who exercise tend to have higher sleep efficiency than people whose exercise frequency is 0.

|  |
| --- |
| Exercise Frequency against Sleep Efficiency (Male Group) |
|  |
| Exercise Frequency against Sleep Efficiency (Female Group) |
|  |

**4.7 Relation between sleep efficiency and alcohol consumption**

We perform Kruskal-Wallis rank sum test with the cleaned dataset for male and female groups. Given the p-value is extremely low, we reject the null hypothesis that different Alcohol Consumption levels give identical results on sleep efficiency.

|  |  |
| --- | --- |
| Alcohol Consumption for Male | Alcohol Consumption for Female |
|  |  |

Afterwards, we make a plot with Alcohol.consumption against Sleep.efficiency for males and females. We can see that for both genders, there is a negative correlation between Alcohol.consumption and Sleep.efficiency.

|  |  |
| --- | --- |
| Scatterplot for Male Group | Scatterplot for Female Group |
|  |  |

We made the boxplots of Sleep Efficiency against Alcohol Consumption by categorising data into three groups–No Alcohol Consumption, Low Alcohol Consumption and High Alcohol Consumption. From the boxplots, we can see that most people who do not consume alcohol experience higher sleep efficiency than the other two groups.

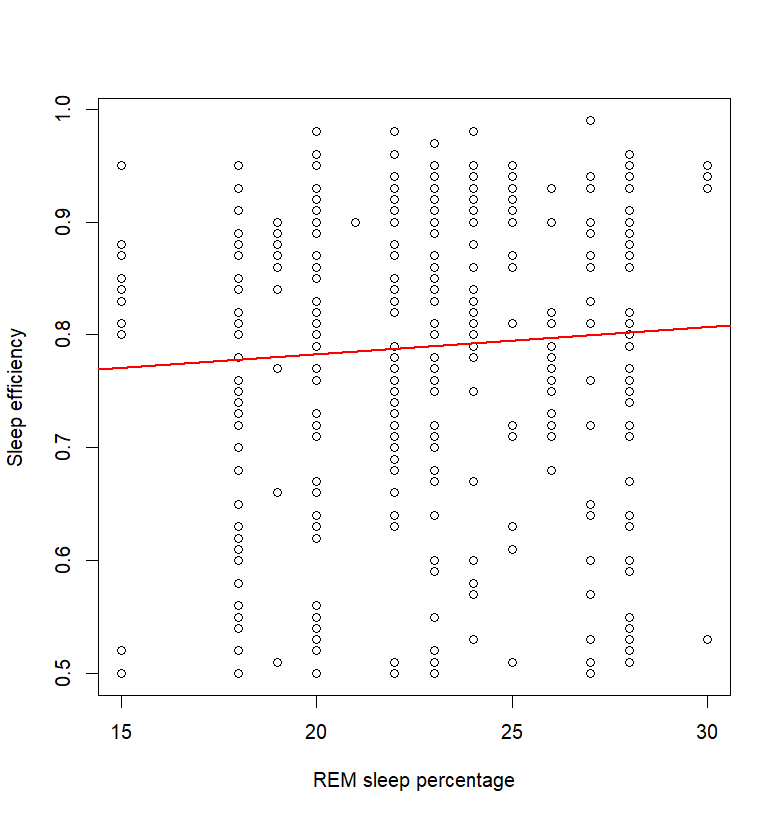
|  |
| --- |
| Alcohol Consumption against Sleep Efficiency (Male Group) |
|  |
| Alcohol Consumption against Sleep Efficiency (Female Group) |
|  |

**4.8 Relation between sleep efficiency and three sleep stages(Light, Deep, REM)**

There are three types of sleep:

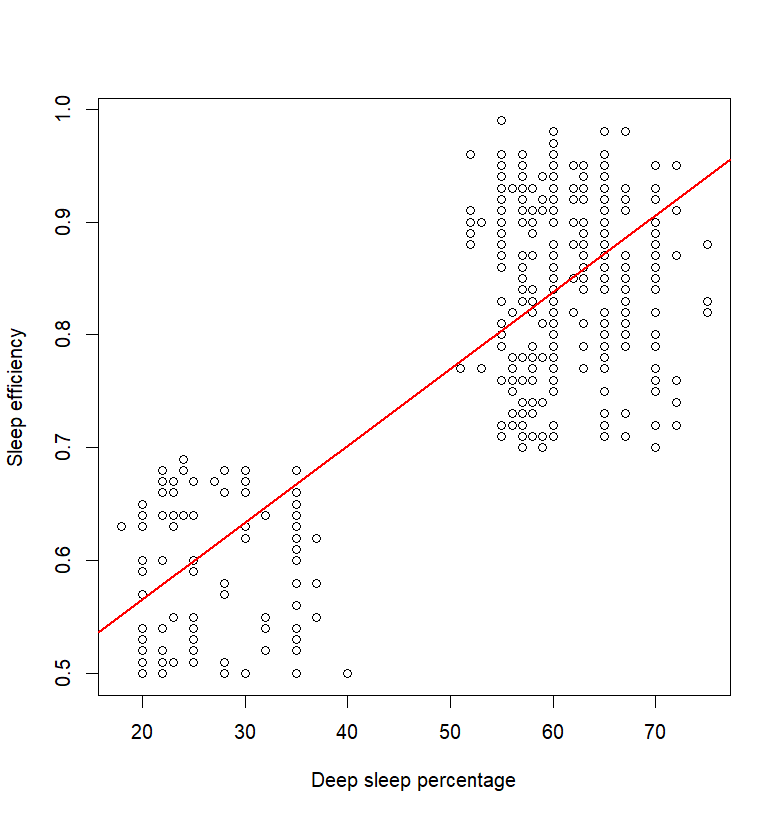
1. REM Sleep: Rapid eye movements, high brain activity, and vivid dreams.
2. Deep Sleep: Slow brain waves, deep relaxation, and minimal muscle activity.
3. Light Sleep: Easily disturbed, brain activity slows down, eye movements stop.

4.8.1 Scatter Plots with Linear Regression



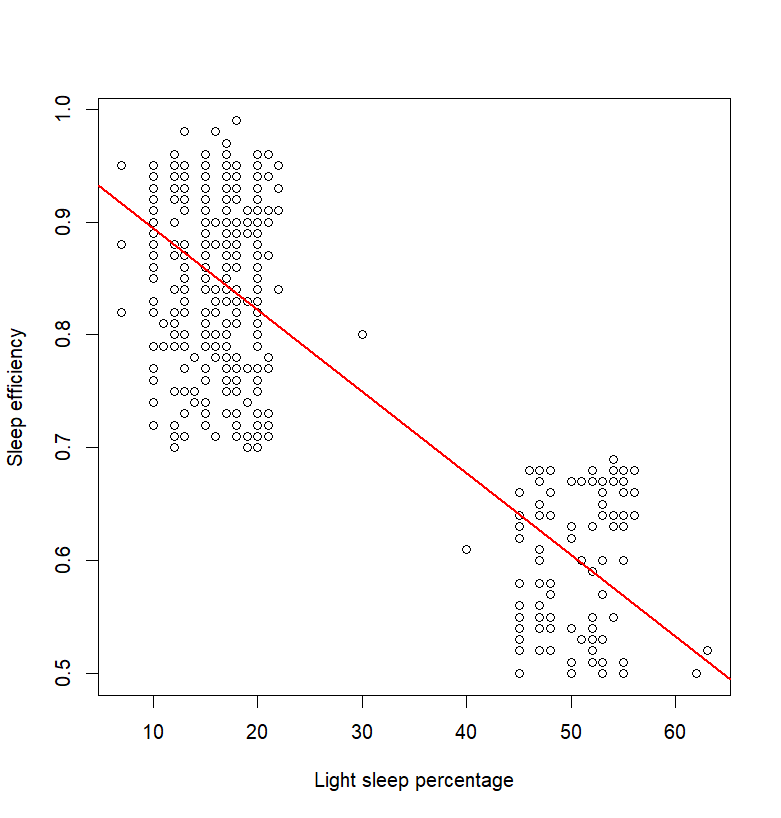
**REM Sleep:**

Here we see a scatter plot with a nearly flat linear regression line between “REM sleep percentage" and "sleep efficiency", this suggests that there is a weak or no correlation between these two variables. In other words, changes in the percentage of REM sleep do not correspond to changes in sleep efficiency.



**Deep Sleep**

Here we see a scatter plot with a linear regression line between "deep sleep percentage" and "sleep efficiency", and the line has a positive slope, this indicates a strong positive correlation between the two variables. In other words, as the percentage of deep sleep increases, the sleep efficiency also increases. This suggests that deep sleep plays an important role in promoting better sleep efficiency.



**Light Sleep**

Here we see a scatter plot with a linear regression line between "light sleep percentage" and "sleep efficiency", and the line has a negative slope, this indicates an inverse relationship between the two variables. In other words, as the percentage of light sleep increases, the sleep efficiency decreases. This suggests that spending a higher percentage of time in light sleep may negatively impact sleep efficiency

4.8.2 Correlation using the Pearson coefficient

To test the correlation between sleep efficiency and the three sleep stages, we used the Pearson correlation coefficient.

|  |
| --- |
| > cor(sleep\_data$Sleep.efficiency, sleep\_data$REM.sleep.percentage, method = "pearson")  [1] 0.06236245  > cor(sleep\_data$Sleep.efficiency, sleep\_data$Deep.sleep.percentage, method = "pearson")  [1] 0.7873351  > cor(sleep\_data$Sleep.efficiency, sleep\_data$Light.sleep.percentage, method = "pearson")  [1] -0.819204 |

Based on the Pearson correlation coefficients, there is a weak positive correlation between "Sleep efficiency" and "REM sleep percentage" with a coefficient of 0.062. This suggests that sleep efficiency may increase slightly as the percentage of REM sleep increases, but the relationship is not particularly strong. However, there is a significant tendency for sleep efficiency to increase as the percentage of deep sleep increases, with a strong positive correlation coefficient of 0.787. Conversely, there is a strong negative correlation coefficient of -0.819 between sleep efficiency and light sleep percentage, which indicates a substantial tendency for sleep efficiency to decrease as the percentage of light sleep increases. Although correlation does not imply causation, these correlations provide valuable insight into the factors that affect sleep efficiency. Other factors may also influence the relationship between these variables; nevertheless, these correlations offer a valuable understanding of the relationship between these sleep-related statistics.

## Conclusion and Discussion

We can summarise the factors influencing sleep efficiency based on age, gender, lifestyle, and habits. Elderly individuals tend to have the lowest sleep efficiency followed by children, while adults have the highest. Gender also plays a role, affecting sleep efficiency across different age groups. Late sleepers are more likely to have lower sleep efficiency compared to individuals who sleep earlier or on time. Interestingly, sleep duration doesn't seem to have much of an effect on sleep efficiency. However, smoking can result in lower sleep efficiency for both men and women, and caffeine consumption appears to have no impact. On the other hand, exercise has a positive effect on sleep efficiency, with frequent exercise associated with higher sleep quality. Additionally, consuming lower levels of alcohol can lead to higher sleep efficiency. Lastly, deep sleep is considered the most important stage of sleep for achieving higher sleep efficiency.

However, since our datasets are not large enough, the results may not fully represent the sleep profile of the global population. Additionally, people might have different sleep routines so our findings might not be applicable on certain individuals. Thus, this report is just a general guideline on what factors could affect sleep efficiency and individuals who are trying to fix their sleep schedules may refer to it.

## References

Li, J., Vitiello, M. V., & Gooneratne, N. S. (2018, March). *Sleep in normal aging*. Sleep medicine clinics. Retrieved April 18, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5841578/