Exploring Sleep Patterns

Yidan Chen

Abstract

In this project, we analyzed human sleep patterns using multiple linear regression models to identify significant predictors, including age, sex, race/ethnicity, physical activity, and alcohol consumption. Providing insights into the variability of sleep behaviors across populations.

Contents

1 Background Information

Sleep is a fundamental aspect of human health and plays a vital role in physical and mental health. Poor quality or duration of sleep may lead to a variety of health problems. However, sleep behavior is influenced by a variety of factors, such as lifestyle, genetics, work environment, etc., and thus the issue has long been an important area of public health research.

This analysis utilizes data from the 2017-2018 National Health and Nutrition Examination Survey (NHANES) to focus on the determinants of sleep duration, particularly on weekdays and weekends. By identifying key factors that influence sleep duration, this study aims to provide insights into potential human health and public health strategies to improve sleep hygiene and health in different populations.

2 Research Problem

In this project we focus on the following research questions:

- What are the main factors that influence sleep duration on weekdays and weekends (considering gender, age, and ethnicity)
- How does lifestyle (physical activity and alcohol consumption) affect sleep patterns?
- Are there significant differences in the factors that influence weekday sleep versus weekend sleep?

3 Analysis & Results

By producing the distribution of sleep time and outputting the summary data, we can intuitively get that the histogram of sleep time on weekdays shows an approximately normal distribution, with a peak value around 8 hours, indicating that the majority of people sleep about 8 hours on weekdays. Similar to weekdays, the histogram for

weekend sleep time is also bell-shaped, with a peak value close to 8 hours. However, it appears to be slightly wider, thus indicating that sleep duration is more variable on weekends.

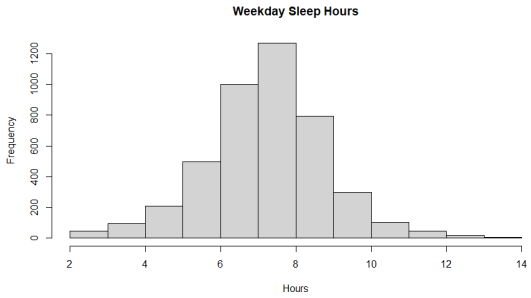


Figure 1: Weekday Sleep Hours

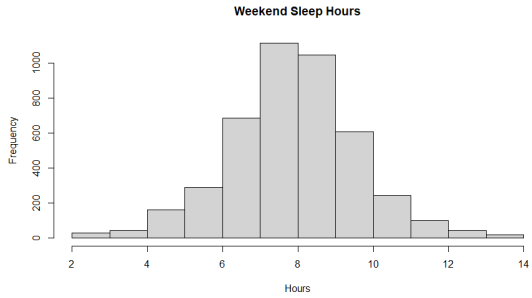


Figure 2: Weekend Sleep Hours

The analysis examined the factors influencing **weekday** and **weekend** sleep hours, highlighting both commonalities and differences.

Sex Differences:

- On weekdays, males sleep **0.15 hours less** than females on average ($p = 0.002$).
- On weekends, males sleep **0.21 hours less** than females ($p < 0.001$).

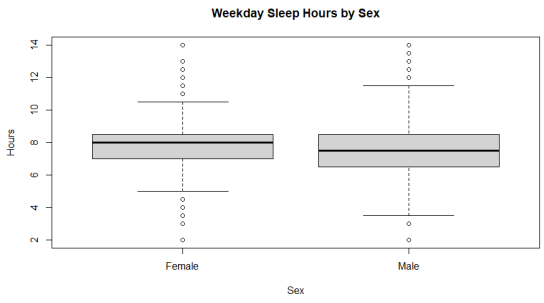


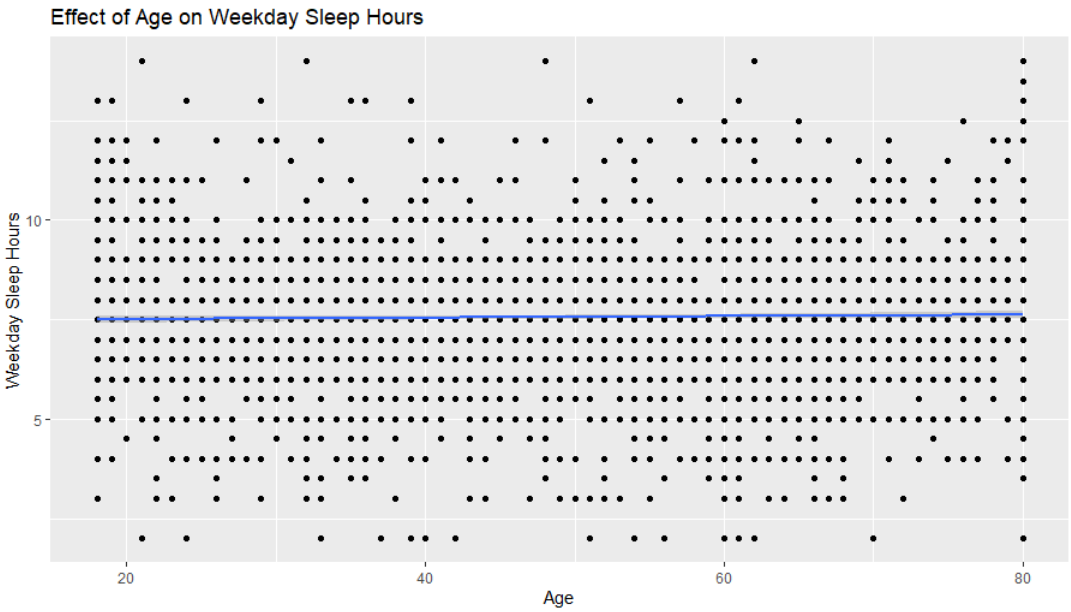
Figure 3: Weekday Sleep Hours

Figure 4: Weekend Sleep Hours

Age Effect:

- Age does not significantly influence weekday sleep hours ($p = 0.457$).

- On weekends, older individuals sleep **0.013 hours less** per year of age ($p < 0.001$).



Race/Ethnicity:

- On weekdays:
  - NH-Black individuals sleep **0.37 hours less** than NH-White individuals ($p < 0.001$).
  - NH-Other individuals sleep **0.16 hours less** than NH-White individuals ($p = 0.023$).
  - Hispanic individuals do not show significant differences ($p = 0.189$).
- On weekends:
  - Hispanic individuals sleep **0.36 hours more** than NH-White individuals ($p < 0.001$).
  - No significant differences are observed for NH-Black or NH-Other groups.

Physical Activity:

- Higher physical activity levels reduce sleep on both weekdays (**0.10-hour reduction**, $p < 0.001$) and weekends (**0.06-hour reduction**, $p < 0.001$).

Heavy Drinking:

- Heavy drinking does not significantly influence sleep hours on either weekdays ($p = 0.246$) or weekends ($p = 0.164$).

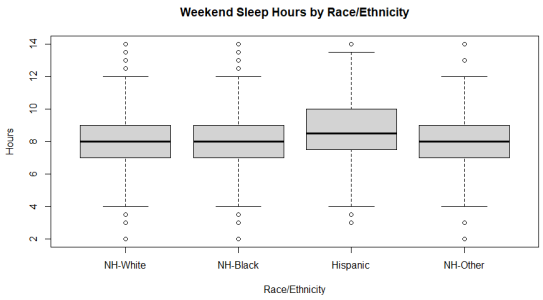


Figure 5: Weekday Sleep Hours

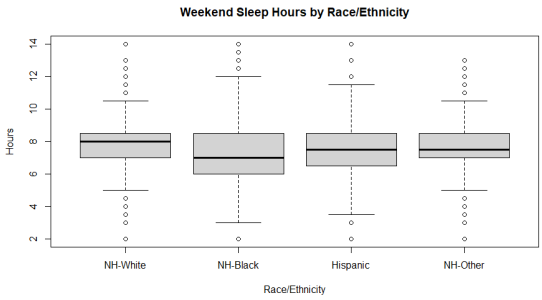


Figure 6: Weekend Sleep Hours

- **Weekday Sleep Hours:** Significant predictors include *sex*, *race/ethnicity*, and *physical activity*. The model explains **2.37%** of the variability in weekday sleep ($R^2 = 0.0237$).
- **Weekend Sleep Hours:** Significant predictors include *age*, *sex*, *race/ethnicity (Hispanic)*, and *physical activity*. The model explains **3.59%** of the variability in weekend sleep ($R^2 = 0.0359$).

```
Call:
lm(formula = slphr.wd ~ age + sex + race_eth + phy.act + hvy.drk,
    data = nhanes17)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8053 -0.8694  0.0895  0.9835  7.1059

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       7.960232   0.094493  84.241  < 2e-16 ***
age              -0.001043   0.001403  -0.744  0.45721
sexMale          -0.151219   0.049742  -3.040  0.00238 **
race_ethNH-Black -0.370720   0.065295  -5.678 1.45e-08 ***
race_ethHispanic -0.085184   0.064831  -1.314  0.18894
race_ethNH-Other -0.158835   0.069799  -2.276  0.02292 *
phy.act          -0.104879   0.014835  -7.070 1.80e-12 ***
hvy.drk          -0.048166   0.041512  -1.160  0.24599
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.599 on 4374 degrees of freedom
Multiple R-squared:  0.02373,   Adjusted R-squared:  0.02216
F-statistic: 15.19 on 7 and 4374 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = slphr.wn ~ age + sex + race_eth + phy.act + hvy.drk,
    data = nhanes17)

Residuals:
    Min      1Q  Median      3Q     Max
-6.2908 -1.0125  0.0003  1.0021  6.2660

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       9.083117   0.100800  90.111  < 2e-16 ***
age              -0.013542   0.001497  -9.046  < 2e-16 ***
sexMale          -0.210491   0.053062  -3.967 7.40e-05 ***
race_ethNH-Black -0.095848   0.069652  -1.376 0.168866
race_ethHispanic  0.359488   0.069158   5.198 2.11e-07 ***
race_ethNH-Other -0.047713   0.074457  -0.641 0.521677
phy.act          -0.059897   0.015825  -3.785 0.000156 ***
hvy.drk          -0.061611   0.044282  -1.391 0.164195
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.706 on 4374 degrees of freedom
Multiple R-squared:  0.03589,   Adjusted R-squared:  0.03435
F-statistic: 23.26 on 7 and 4374 DF,  p-value: < 2.2e-16
```

In this analysis, our goal was to identify important factors influencing weekday and weekend sleep duration, while ensuring the validity of the model through diagnostic checks. Our initial explorations focused on understanding the distribution of key variables, including weekday sleep duration (slphr.wd), weekend sleep duration (slphr.wn), and potential predictors such as age, gender, race/ethnicity, physical activity, and alcohol use. By aggregating statistics and visualizing (e.g., histograms, box plots, and scatter plots) these variables, we were able to further identify patterns, relationships, and potential outliers.

During the modeling phase, we constructed two separate multiple linear regression models for weekday sleep duration and weekend sleep duration, where predictors included age, gender, race/ethnicity, physical activity, and alcohol use. These variables were selected based on their theoretical relevance to sleep behavior and their availability in the dataset.

In model evaluation, we assessed the model using the following methods: 1) Statistical significance, using p-values to determine the strength of the relationship. 2) godness of the fit, which provides insight into the extent to which the model explains changes

in sleep duration through R and R^2 values. 3) Residual diagnostics,We checked residual linearity, normality, and homoskedasticity through r-constructed images to ensure that the model meets the key assumptions of linear regression. In summary, the experimental model was improved by diagnostic plots and visualizations that included support for interpretation and highlighted important trends, and only meaningful findings were presented to avoid overreporting

4 Conclusion and Discussion

This analysis identified key factors influencing sleep duration on weekdays and weekends, with age, gender and physical activity being the main influences on sleep, while alcohol use and were not significantly associated with sleep duration. In terms of gender, males slept less than females on both weekdays and weekends. Age was a significant factor influencing weekend sleep, with older adults sleeping for shorter periods of time. In terms of race, blacks sleep less on weekdays, while Hispanics sleep more on weekends.

In addition, physical activity shortens sleep duration. In addition, low R^2 may indicate underlying problems and limitations, as shown in the output of the previous section, our these models explained only 2.37% of weekday and 3.59% of weekend sleep variability, suggesting that other unmeasured factors (e.g., stress, diet, illness) may play an important role.

Future directions may include other predictors such as work hours, stress levels, and diet to better explain sleep variation

A Code

```r
#load("C:/Users/15104/Documents/Gwen/UV/U1/Linear Models/finalPorject/nhanes17.RData")

# Load necessary libraries
library(ggplot2)

# Check the structure
#str(nhanes17)
#head(nhanes17)
# Summarize sleep variables
summary(nhanes17$slphr.wd)
summary(nhanes17$slphr.wn)

# Plot histograms to visualize distributions
hist(nhanes17$slphr.wd, main = "Weekday Sleep Hours", xlab = "Hours")
hist(nhanes17$slphr.wn, main = "Weekend Sleep Hours", xlab = "Hours")
# Summarize key predictors
summary(nhanes17$age)
summary(nhanes17$sex)
summary(nhanes17$race_eth)
summary(nhanes17$phy.act)
summary(nhanes17$hvy.drk)

# Plot relationships with sleep hours
boxplot(slphr.wd ~ sex, data = nhanes17, main = "Weekday Sleep Hours by Sex", xlab = "Sex", ylab = "Hours")
boxplot(slphr.wn ~ sex, data = nhanes17, main = "Weekend Sleep Hours by Sex", xlab = "Sex", ylab = "Hours")
boxplot(slphr.wn ~ race_eth, data = nhanes17, main = "Weekend Sleep Hours by Race/Ethnicity", xlab = "Race/Ethnicity", ylab = "Hours")
boxplot(slphr.wd ~ race_eth, data = nhanes17, main = "Weekend Sleep Hours by Race/Ethnicity", xlab = "Race/Ethnicity", ylab = "Hours")

#Model1_for weekday
model_wd <- lm(slphr.wd ~ age + sex + race_eth + phy.act + hvy.drk, data = nhanes17)
summary(model_wd)
#model2_for weekend
model_wn <- lm(slphr.wn ~ age + sex + race_eth + phy.act + hvy.drk, data = nhanes17)
summary(model_wn)
```

```r

# Plot residuals for diagnostics
par(mfrow = c(2, 2))
plot(model_wd)
plot(model_wn)

# Visualize significant predictors
library(ggplot2)
ggplot(nhanes17, aes(x = age, y = slphr.wd)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Effect of Age on Weekday Sleep Hours", x =
     "Age", y = "Weekday Sleep Hours")
```

Listing 1: R Code for Sleep Analysis