Lab 2: Describing Generations' Sleep

Neeha Kotte, Tazrian Ahmed, Deva Empranthiri

Background: Why is this topic important?

- Sleep plays a vital role in maintaining healthy physical and mental well-being.
- Understanding the factors influencing individuals' sleep patterns is crucial to assessing overall health.
- Therefore, our overall research question is:

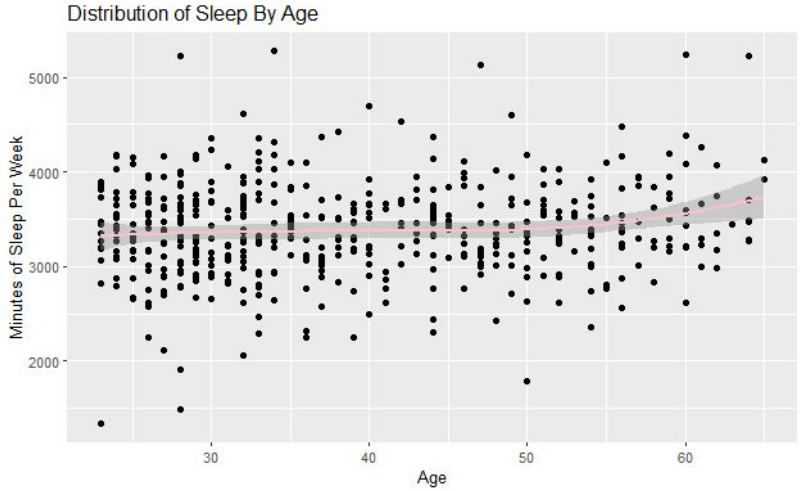
How does age correlate with an individual's sleep duration, and what role do sex and health status play in this association?

Data, Conceptualization & Operationalization

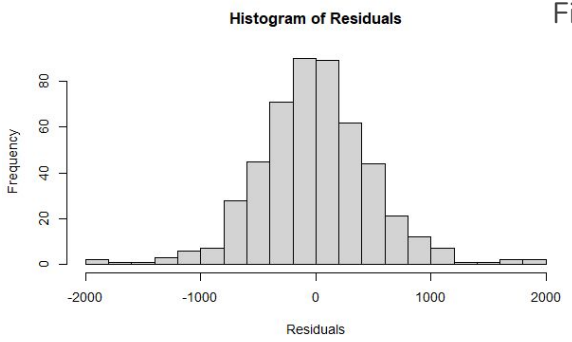
- Cross-sectional dataset, ‘sleep75’, sourced from *Introductory Econometrics: A Modern Approach, 5th Edition* (wooldridge package in R)

- **706** observations and **34** different variables:
  - **age** is a metric variable ranging from 23 years to 65 years old
  - **slpnaps** is a metric variable which represents the minutes of sleep at night (and during naps) an individual receives per week
  - **male** is an indicator variable (1 for males)
  - **gdhlth** is an indicator variable (1 if in good or excellent health)

Modeling Specification & Visualization

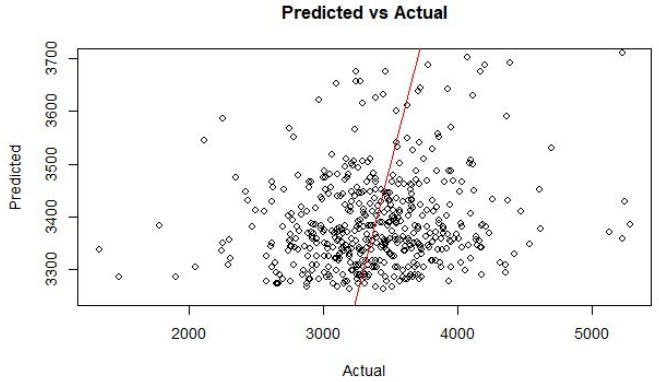


- Age vs. minutes of sleep (Including Naps) per week
- Age: Minimum of 23, maximum of 65
- Minutes of Sleep: Minimum of 1335, maximum of 5280
- No significant outliers
- Minutes of Sleep is normally distributed
- Age is uniformly distributed



Final Model: Sleep vs Age & Male & Good Health

- Male/Good Health: Binary Values (0, 1)
- Formula:
- Sleep ~ 4.446 * age - 191.446 * gdhlth - 73.072 * male + 3425
- R-squared: 0.02465



Results & Conclusion

```
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                slpnaps
## ------------------------------------------------
## age                             4.288*
##                                 (1.961)
##
## gdhlth                         -195.675**
##                                 (74.042)
##
## Constant                      3,394.931***
##                                 (108.045)
##
## ------------------------------------------------
## Observations                      495
## R2                               0.025
## Adjusted R2                      0.021
## Residual Std. Error       499.026 (df = 492)
## F Statistic              6.411** (df = 2; 492)
## ================================================
## Note:                *p<0.05; **p<0.01; ***p<0.001
```

Model 2

```
================================================
                          Dependent variable:
                      ---------------------------
                                slpnaps
------------------------------------------------
age                             4.446*
                                (1.960)

gdhlth                         -191.446**
                                (73.967)

male                            -73.072
                                (45.143)

Constant                      3,425.526***
                                (109.511)

------------------------------------------------
Observations                      495
R2                               0.031
Adjusted R2                      0.025
Residual Std. Error       498.206 (df = 491)
F Statistic              5.162** (df = 3; 491)
================================================
Note:                *p<0.05; **p<0.01; ***p<0.001
```

Model 3

- Three linear regression models were developed to predict `slpnaps` (minutes of sleep per week) based on different combinations of predictors: age, general health (`gdhlth`), and gender (`male`).
- Model 2 incorporated both age and good health as predictors. The coefficient for general health revealed that individuals with good or excellent health tend to have lower minutes of 'slpnaps', controlling for age, with a coefficient of -195.675. This model exhibited a slightly improved adjusted R-squared of 0.02144 than Model 1.
- The addition of gender (`male`) in Model 3 did not significantly enhance the model's predictive performance, as the coefficient for gender was not statistically significant.
- Among the models, Model 2 emerged as the most suitable for predicting `slpnaps` in this dataset.
- Further research could explore additional predictors or interactions to better capture the complexity of sleep behavior and enhance predictive accuracy.