

# Sleep Duration and Body Mass Index

Jordan Anderson (American University)

```
library(tidyverse) library(haven) library(janitor)
knitr::opts_chunk$set( message = FALSE, warning = FALSE )
```

## Load NHANES data

```
demo <- read_csv("demo.csv", show_col_types = FALSE) sleep <- read_csv("sleep.csv",
show_col_types = FALSE) bmx <- read_xpt("BMX_J.xpt")
demo <- clean_names(demo) sleep <- clean_names(sleep) bmx <- clean_names(bmx)
```

## Merge datasets

```
full <- demo %>% left_join(sleep, by = "seqn") %>% left_join(bmx, by = "seqn")
```

## Create analysis variables

```
df <- full %>% mutate( sleep_hours = na_if(sld012, 77), sleep_hours = na_if(sleep_hours,
99), sleep_hours = ifelse(sleep_hours < 1 | sleep_hours > 24, NA, sleep_hours),

bmi = bmx$bmi,

sex = factor(riagendr, levels = c(1, 2),
              labels = c("Male", "Female")),
age  = ridageyr,
race = factor(ridreth1),
pir  = indfmpir

) %>% drop_na(sleep_hours, bmi, age, sex, race, pir)
```

## Sleep groups for figures

```
df <- df %>% mutate( sleep_group = case_when( sleep_hours < 6 ~ "<6", sleep_hours < 7 ~ "6-7", sleep_hours < 8 ~ "7-8", TRUE ~ "8+" ), sleep_group = factor(sleep_group, levels = c("<6", "6-7", "7-8", "8+")) )
```

### Introduction

Sleep duration has been increasingly recognized as an important behavioral factor influencing metabolic and cardiovascular health. Prior research suggests that insufficient sleep is associated with higher body mass index (BMI), obesity, and related chronic conditions. However, the strength and shape of the relationship between sleep duration and BMI remains an open empirical question, particularly in nationally representative samples of U.S. adults.

In this study, we examine the association between self-reported sleep duration and measured body mass index using data from the National Health and Nutrition Examination Survey (NHANES). We begin by visually exploring the relationship between sleep duration and BMI, followed by grouped comparisons across sleep-duration categories. We then estimate multivariable linear regression models adjusting for demographic and socioeconomic covariates to assess the robustness of the observed relationship.

### Data and Methods Data

Data for this analysis were drawn from the National Health and Nutrition Examination Survey (NHANES), a nationally representative survey conducted by the Centers for Disease Control and Prevention. NHANES combines interview responses with physical examinations to collect detailed information on health behaviors, demographics, and measured health outcomes among U.S. adults.

Multiple NHANES datasets were merged for this study, including demographic characteristics, self-reported sleep duration, and measured body mass index. After excluding individuals with missing data on key variables, the final analytic sample consisted of 4,945 adult participants.

### Variables

The primary outcome variable is body mass index (BMI), measured in kilograms per square meter ( $\text{kg}/\text{m}^2$ ). The key explanatory variable is sleep duration, measured as the average number of hours slept per night.

Covariates included age, sex, race/ethnicity, and income-to-poverty ratio (PIR), which was used as an indicator of socioeconomic status.

### Analytical Approach

Descriptive statistics and visualizations were used to examine patterns in BMI across levels of sleep duration. Scatterplots and smoothed trend lines were used to assess the continuous relationship between sleep duration and BMI. Mean BMI values were also compared across discrete sleep duration groups.

Multivariable linear regression models were estimated to assess the association between sleep duration and BMI while adjusting for demographic and socioeconomic covariates. Statistical significance was evaluated using conventional p-value thresholds.

### Results Sleep Duration and Body Mass Index

Figure 1 presents a scatterplot of BMI against sleep duration, along with a fitted linear trend. The figure suggests a modest negative association between sleep duration and BMI, indicating that individuals who report longer sleep durations tend to have lower BMI on average.

```
ggplot(df, aes(x = sleep_hours, y = bmi)) + geom_jitter(alpha = 0.2, width = 0.05) +
  geom_smooth(method = "lm", se = FALSE) + annotate("text", x = mean(df$sleep_hours), y =
    max(df$bmi) + 1, label = "Figure 1", fontface = "bold" ) + labs( x = "Sleep Duration (hours)",
    y = "Body Mass Index (kg/m2)" ) + theme_minimal()
```

### Average BMI Across Sleep Duration Groups

To further explore this relationship, mean BMI was compared across sleep duration categories. Figure 2 displays average BMI values by sleep group.

```
fig2_df <- df %>% group_by(sleep_group) %>% summarise(mean_bmi = mean(bmi), .groups
  = "drop")
```

```
ggplot(fig2_df, aes(x = sleep_group, y = mean_bmi)) + geom_col() + annotate("text", x =
  2.5, y = max(fig2_df$mean_bmi) + 0.8, label = "Figure 2", fontface = "bold" ) + labs( x =
  "Sleep Duration Group (hours)", y = "Mean BMI (kg/m2)" ) + coord_cartesian(clip = "off")
+ theme_minimal()
```

### Modeled Relationship Between Sleep Duration and BMI

Figure 3 presents a smoothed visualization of BMI across the range of reported sleep durations. The fitted curve suggests a gradual decline in BMI as sleep duration increases, particularly between shorter and moderate sleep durations. At higher sleep durations, the relationship appears to level off, supporting the use of a linear modeling approach.

```
ggplot(df, aes(x = sleep_hours, y = bmi)) + geom_point(alpha = 0.15) + geom_smooth(se =
  TRUE) + annotate("text", x = mean(df$sleep_hours), y = max(df$bmi) + 1, label = "Figure
  3", fontface = "bold" ) + labs( x = "Sleep Duration (hours)", y = "Body Mass Index (kg/m2)"
  ) + theme_minimal()
```

### Multivariable Regression Results

A multivariable linear regression model adjusting for age, sex, race/ethnicity, and income-to-poverty ratio was estimated. Results indicate that sleep duration is significantly associated with BMI. Each additional hour of sleep is associated with an estimated 0.29-unit decrease in BMI ( $p < 0.001$ ), holding other factors constant.

```
m_bmi <- lm( bmi ~ sleep_hours + age + sex + race + pir, data = df )
summary(m_bmi)
```

## Discussion

This analysis examined the association between sleep duration and body mass index among U.S. adults using nationally representative NHANES data. Across multiple descriptive and regression-based approaches, shorter sleep duration was consistently associated with higher BMI.

Visualizations and grouped comparisons suggested that individuals reporting moderate sleep durations exhibited lower average BMI compared to those reporting shorter sleep durations. Regression results further confirmed a statistically significant inverse association between sleep duration and BMI after adjusting for demographic and socioeconomic covariates.

**While the analysis is observational and cannot establish causality, the findings are consistent with prior research linking insufficient sleep to adverse metabolic outcomes. Future research could extend this analysis by incorporating additional behavioral covariates or exploring potential non-linear patterns in greater detail.**