Proposal

Sakura Hu, Ruikang Wang, Peizeng Yuan

2024-10-10

Contributions

Sakura Hu: Ruikang Wang: Peizeng Yuan:

Introduction

Data Description

Preliminary Results

Bibliography

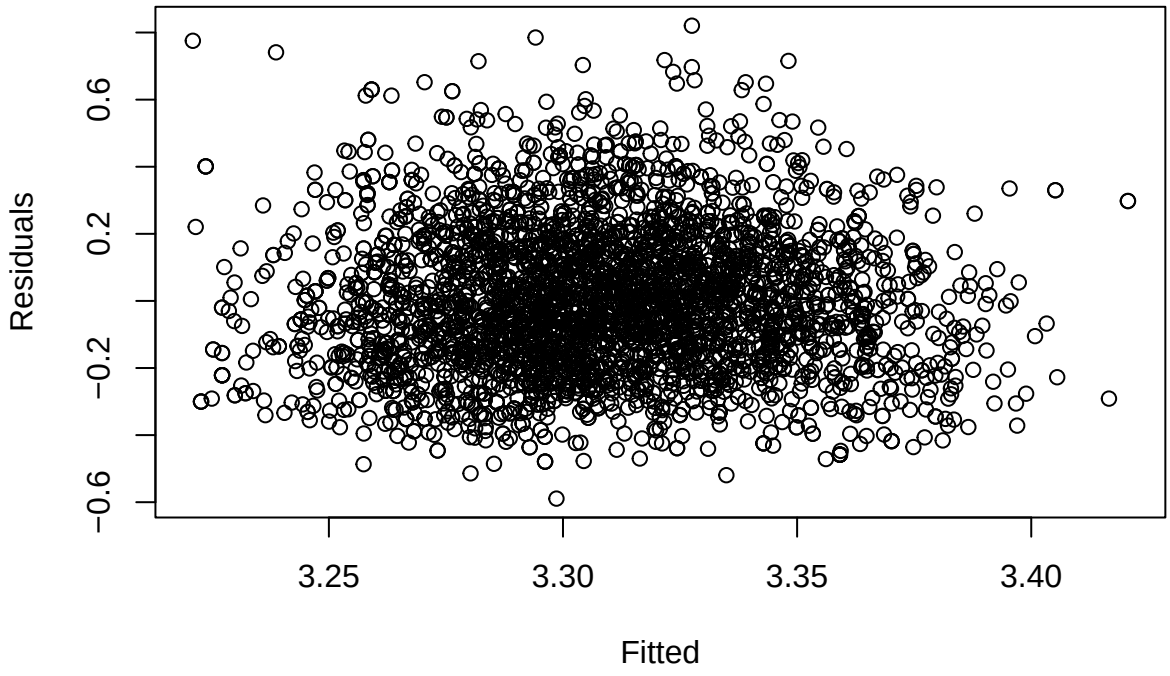
```
##
## The downloaded binary packages are in
##  /var/folders/km/sjr_hj0n7lj6hf7j8l7nrfpm0000gn/T//RtmpqiYkZZ/downloaded_packages
```

```
## # A tibble: 6 x 6
##     BMI Poverty PhysActiveDays   Age SleepHrsNight Gender
##   <dbl>   <dbl>          <int> <int>         <int> <fct>
## 1  27.2     5                5    45             8 female
## 2  27.2     5                5    45             8 female
## 3  27.2     5                5    45             8 female
## 4  23.7     2.2              7    66             7 male
## 5  23.7     5                5    58             5 male
## 6  26.0     2.2              1    54             4 male
```

```
##
## Call:
## lm(formula = log(BMI) ~ Poverty + PhysActiveDays + Age + SleepHrsNight +
##     Gender, data = nhanes_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58922 -0.14816 -0.01533  0.13461  0.82036
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.3596886  0.0229080 146.660  < 2e-16 ***
## Poverty        -0.0101187  0.0021355  -4.738 2.24e-06 ***
## PhysActiveDays -0.0043391  0.0019351  -2.242 0.025004 *
## Age             0.0012950  0.0002051   6.313 3.07e-10 ***
## SleepHrsNight  -0.0102776  0.0027065  -3.797 0.000149 ***
```

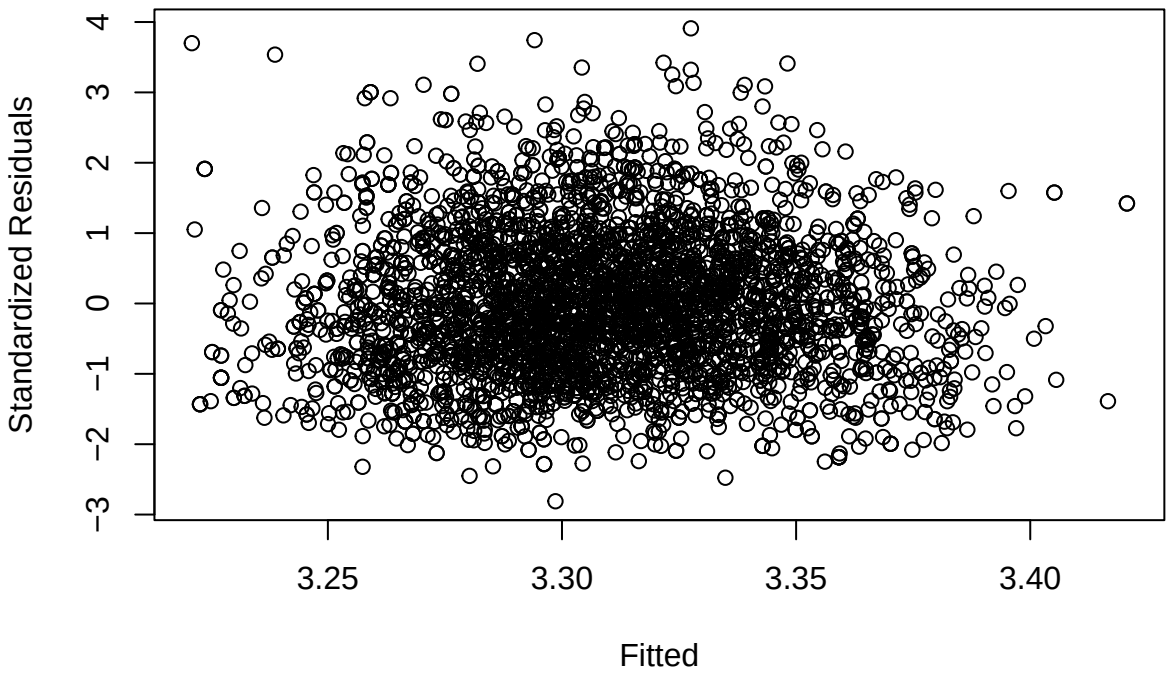
```
## Gendermale      0.0212251  0.0070538   3.009 0.002639 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2099 on 3567 degrees of freedom
## Multiple R-squared:  0.02335,    Adjusted R-squared:  0.02198
## F-statistic: 17.06 on 5 and 3567 DF,  p-value: < 2.2e-16
```

Figure1: fitted versus residual values



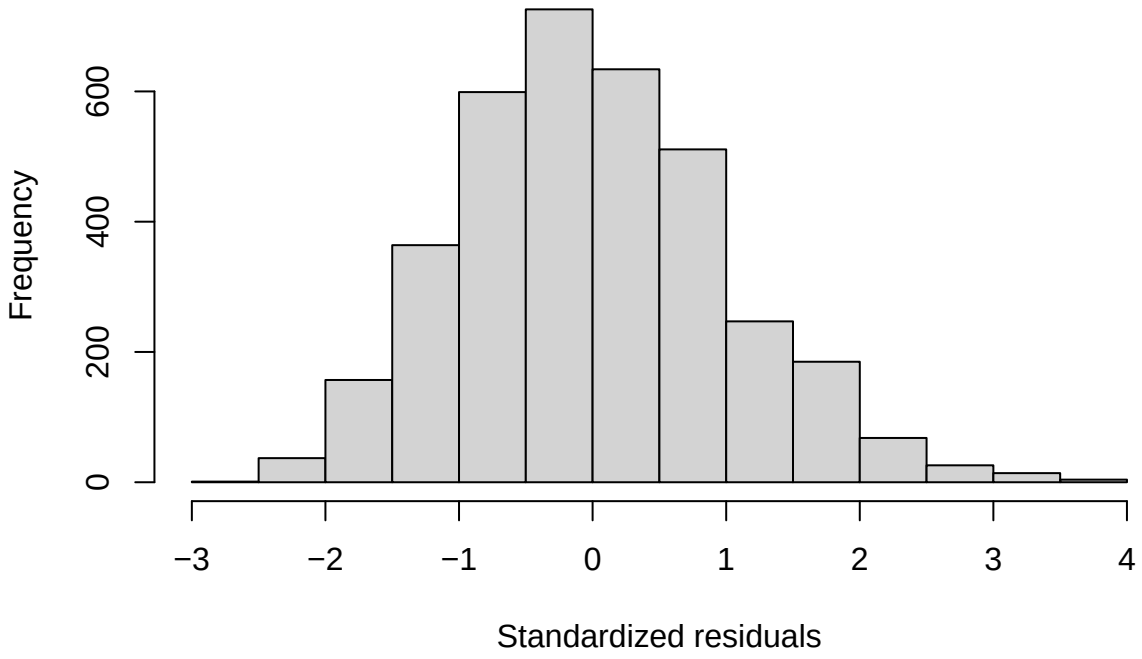

```r
fitted_values = fitted(lm_model)
sresidual_values = rstandard(lm_model)

plot(fitted_values, sresidual_values,
     main = "Figure2: fitted versus Standardized Residuals values",
     xlab = "Fitted",
     ylab = "Standardized Residuals")
```

Figure2: fitted versus Standardized Residuals values




```
hist(sresidual_values,
     main = "Figure3: Standardized residuals histogram",
     xlab = "Standardized residuals")
```

Figure3: Standardized residuals histogram




```r
# Set up the plotting area for a 2x3 grid (2 rows, 3 columns)
par(mfrow = c(2, 3))

# Plot residuals against Poverty
plot(nhanes_data$Poverty, residual_values,
     main = "Figure4: Poverty vs. Residuals",
     xlab = "Poverty",
     ylab = "Residuals")

# Plot residuals against Age
plot(nhanes_data$Age, residual_values,
     main = "Figure5: Age vs. Residuals",
     xlab = "Age",
     ylab = "Residuals")

# Plot residuals against Gender
plot(nhanes_data$Gender, residual_values,
     main = "Figure6: Gender vs. Residuals",
     xlab = "Gender",
     ylab = "Residuals")

# Plot residuals against PhysActiveDays
plot(nhanes_data$PhysActiveDays, residual_values,
     main = "NAHANES: Physical Active Days vs. Residuals",
     xlab = "Physical Active Days",
     ylab = "Residuals")
```

```r
# Plot residuals against SleepHrsNight
plot(nhanes_data$SleepHrsNight, residual_values,
     main = "NAHANES: Sleep Hours vs. Residuals",
     xlab = "Sleep Hours",
     ylab = "Residuals")

# Reset the plotting area to default
par(mfrow = c(1, 1))
```

Figure4: Poverty vs. Residual

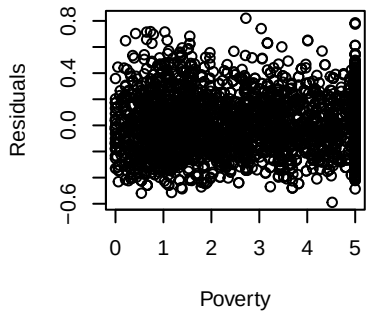


Figure5: Age vs. Residuals

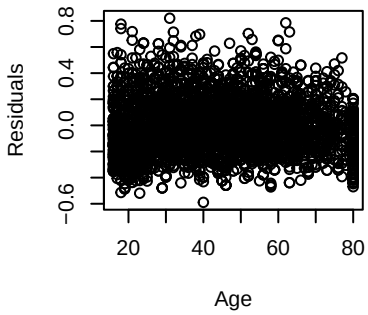
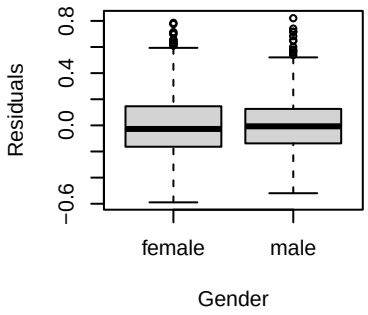
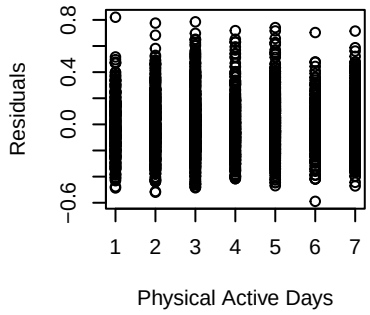


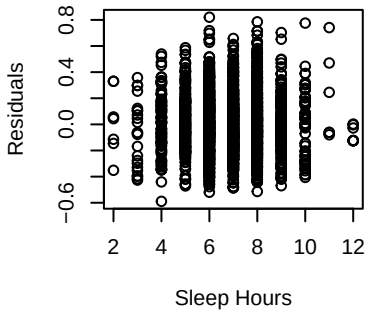
Figure6: Gender vs. Residual



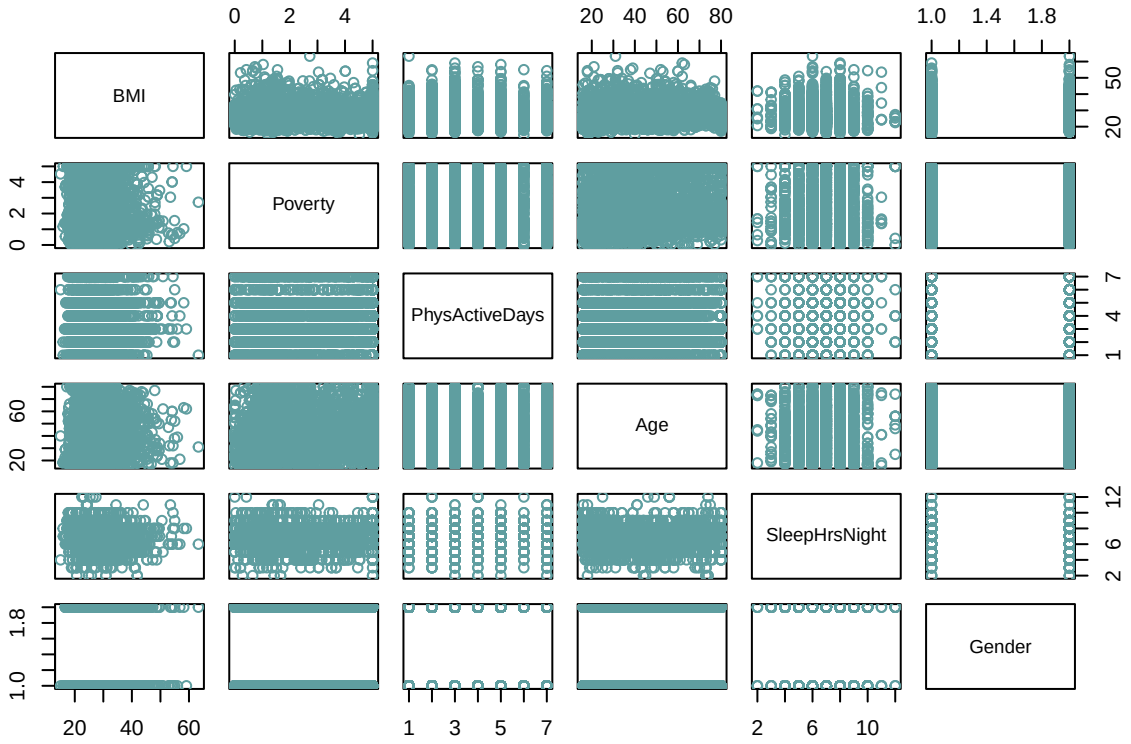
ANES: Physical Active Days vs. R



NAHANES: Sleep Hours vs. Resid

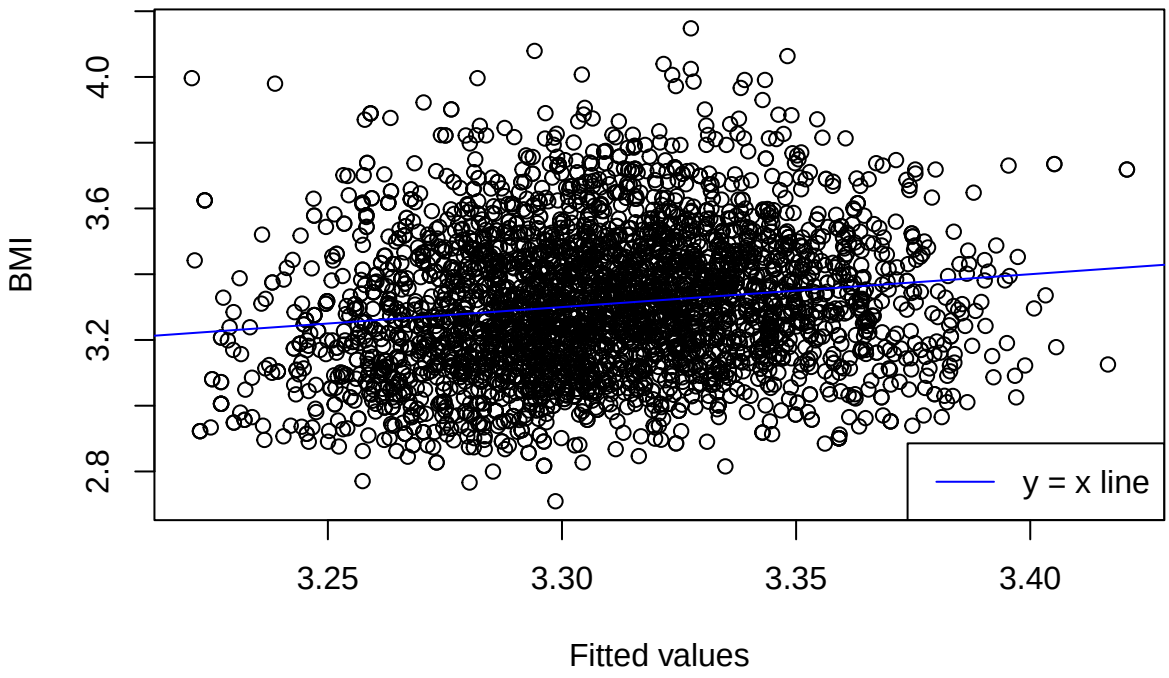



```r
plot(nhanes_data[, c(1, 2, 3, 4, 5, 6)], col="cadetblue")
```

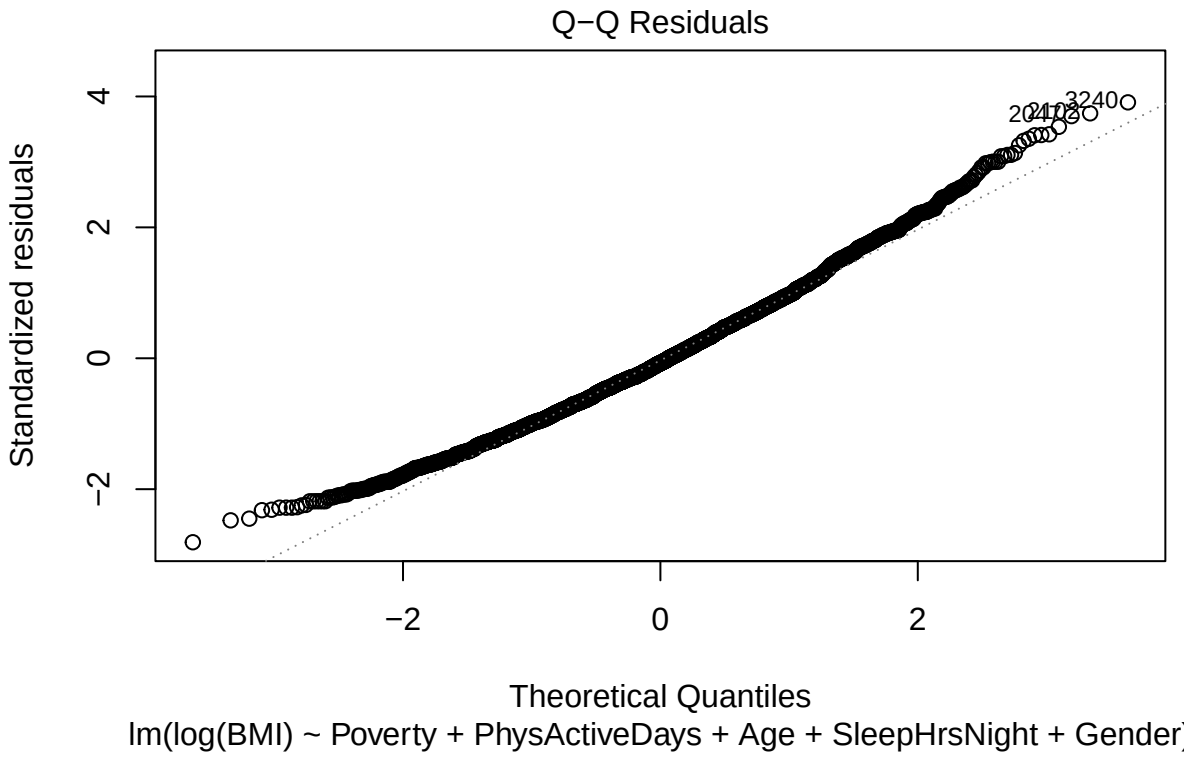



```
BMI_values = log(nhanes_data$BMI)
fitted_values = fitted(lm_model)

plot(fitted_values, BMI_values,
     main = "Figure7: BMI versus regression fit",
     xlab = "Fitted values", ylab = "BMI")
abline(0, 1,
       col=c("blue"), lty=1)
legend("bottomright", legend=c("y = x line"),
       col=c("blue"), lty=1)
```

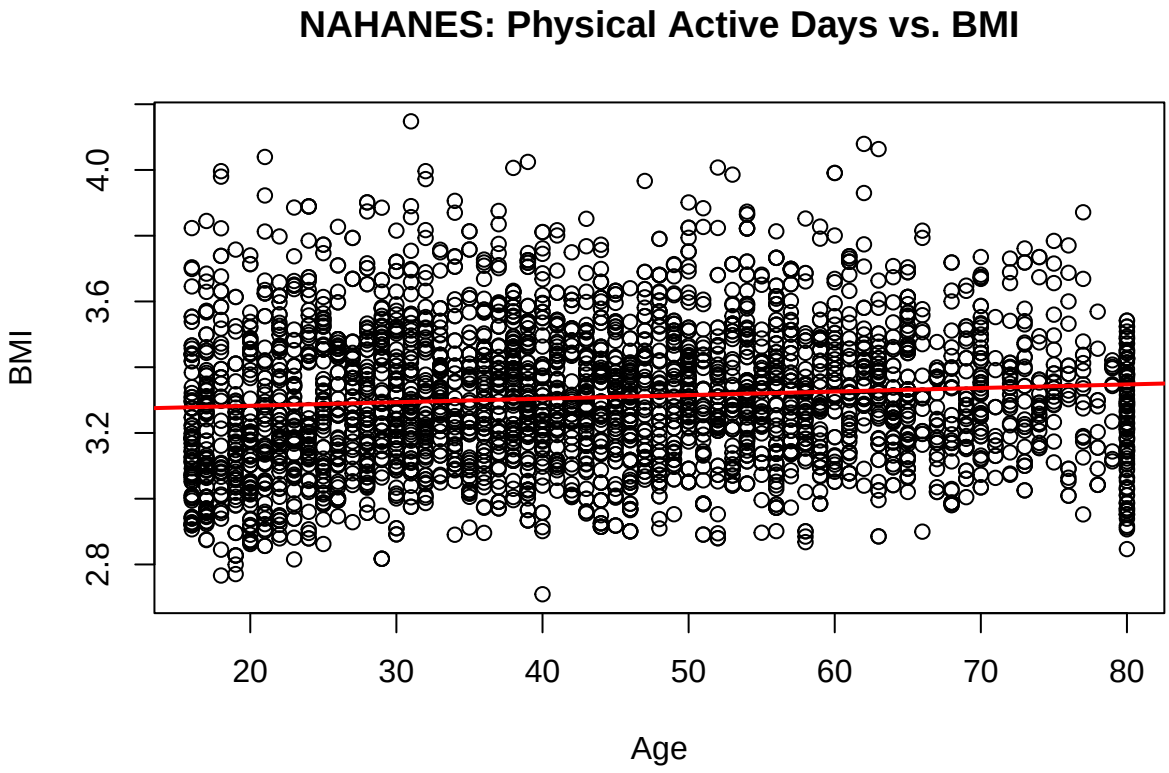
Figure7: BMI versus regression fit



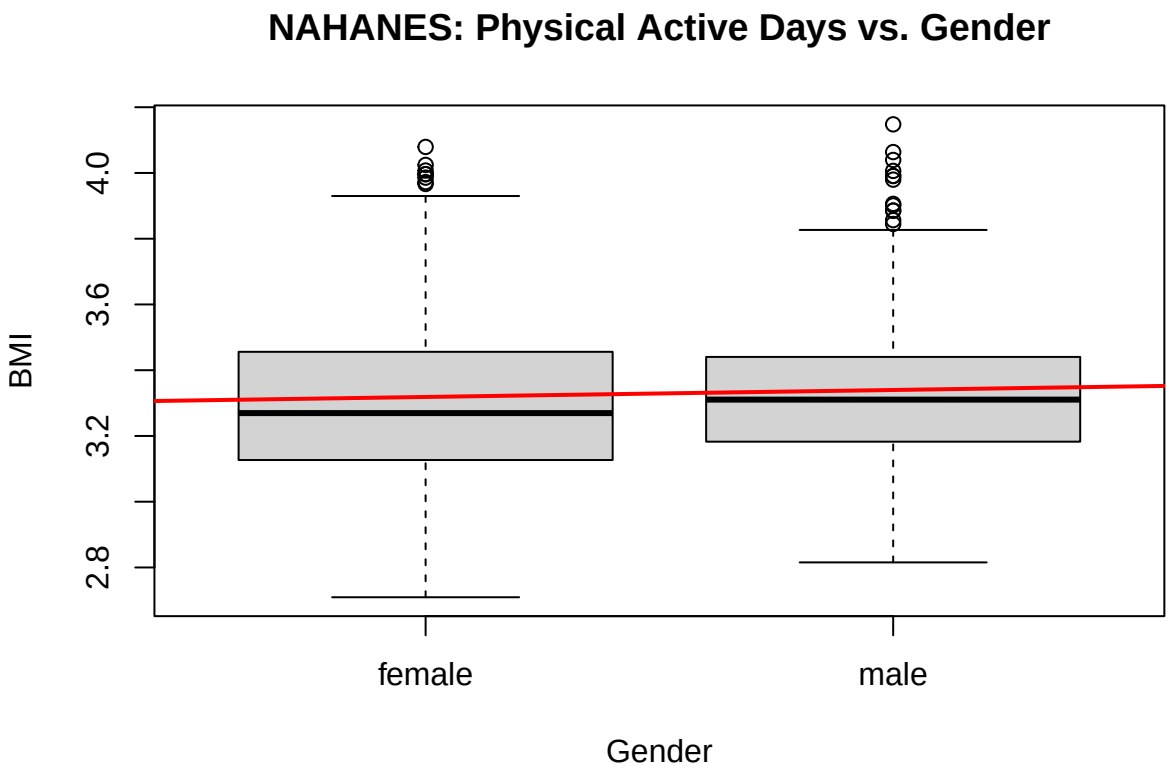

```
# Normal Q-Q plot
plot(lm_model, which = 2)
```



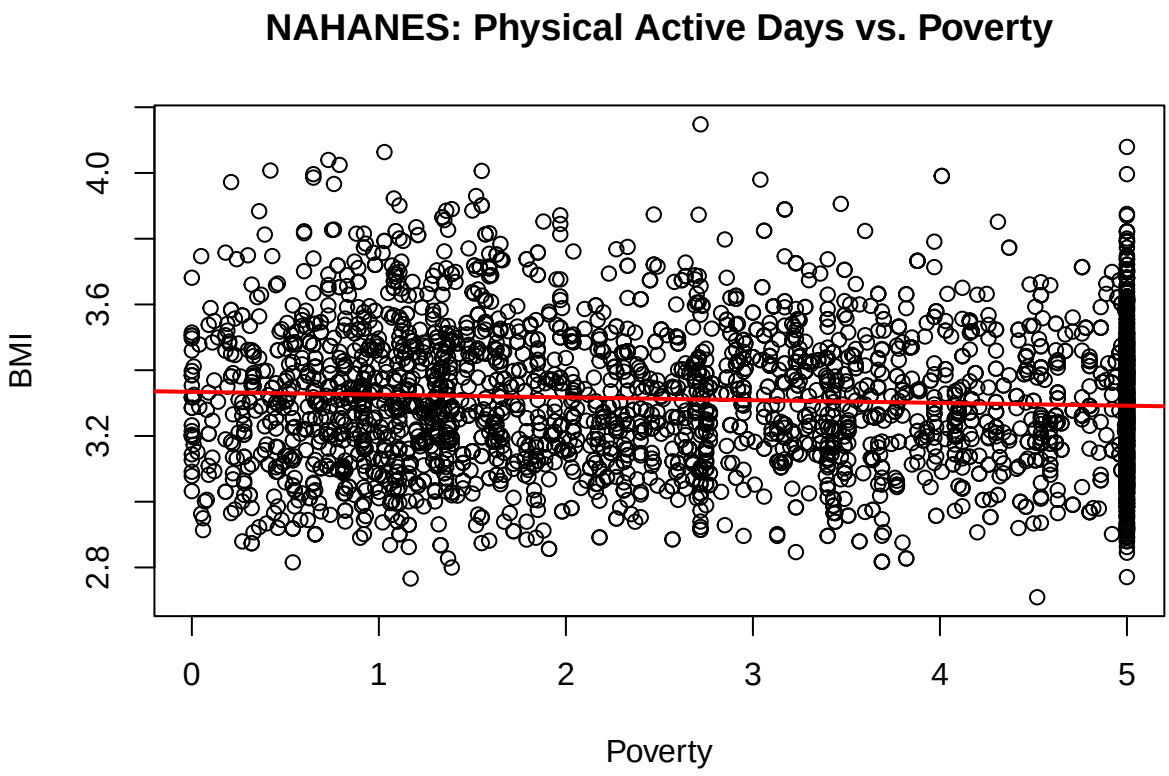

```r
# Scatter plot for continuous variables
plot(nhanes_data$Age, log(nhanes_data$BMI),
     main = "NAHANES: Physical Active Days vs. BMI",
     xlab = "Age",
     ylab = "BMI")


fit <- lm(log(nhanes_data$BMI) ~ nhanes_data$Age)

# Add the fitted regression line
abline(fit, col = "red", lwd = 2)
```



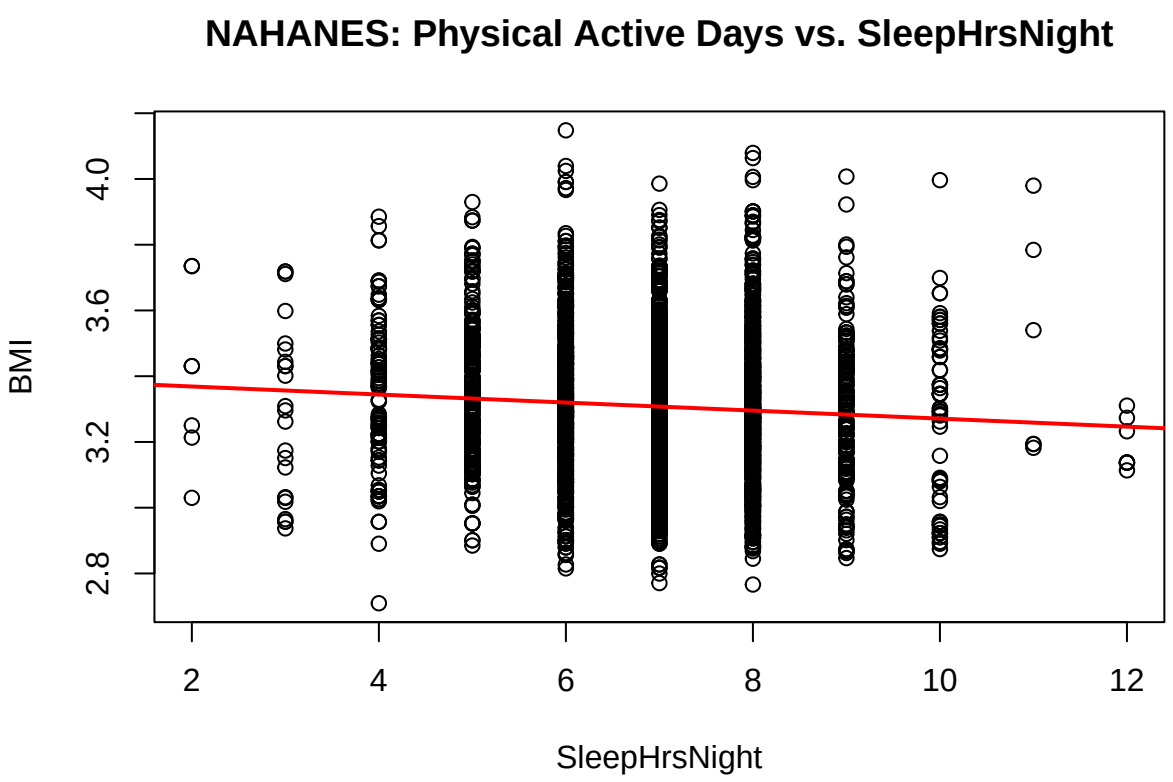

```r
# Scatter plot for continuous variables
plot(nhanes_data$Gender, log(nhanes_data$BMI),
     main = "NAHANES: Physical Active Days vs. Gender",
     xlab = "Gender",
     ylab = "BMI")

fit <- lm(log(nhanes_data$BMI) ~ nhanes_data$Gender)

# Add the fitted regression line
abline(fit, col = "red", lwd = 2)
```



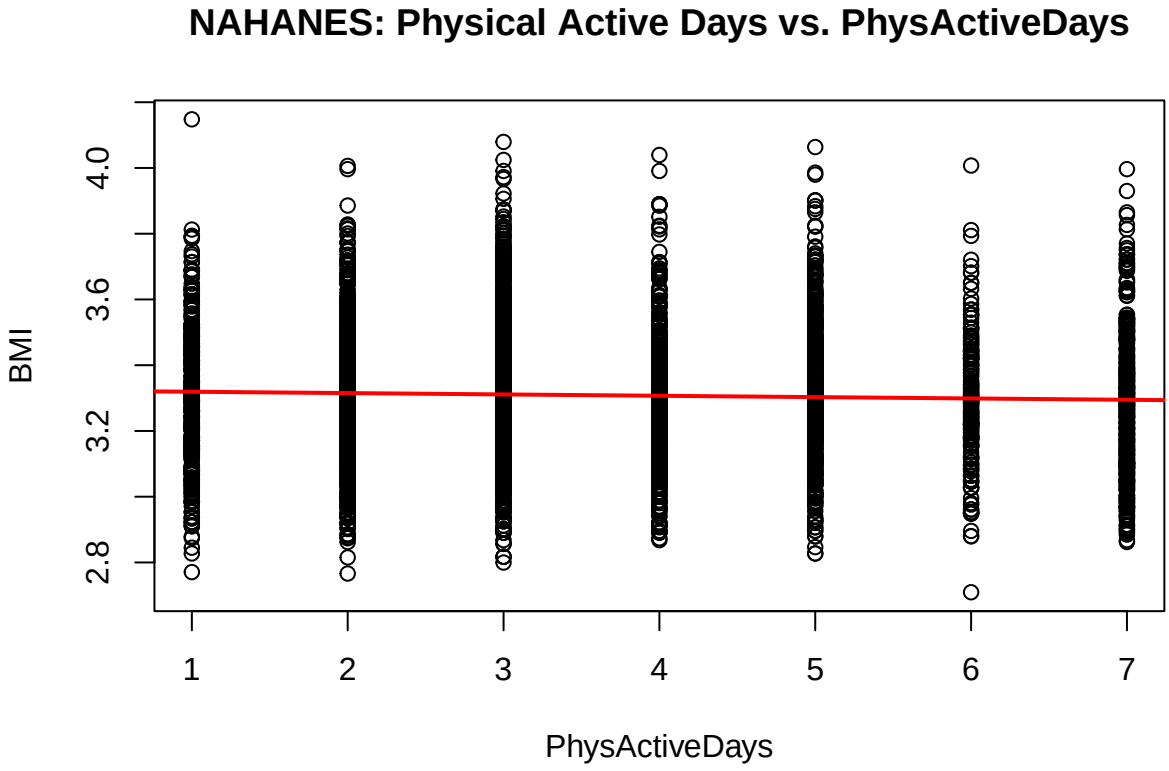

```r
plot(nhanes_data$Poverty, log(nhanes_data$BMI),
     main = "NAHANES: Physical Active Days vs. Poverty",
     xlab = "Poverty",
     ylab = "BMI")

fit <- lm(log(nhanes_data$BMI) ~ nhanes_data$Poverty)

# Add the fitted regression line
abline(fit, col = "red", lwd = 2)
```



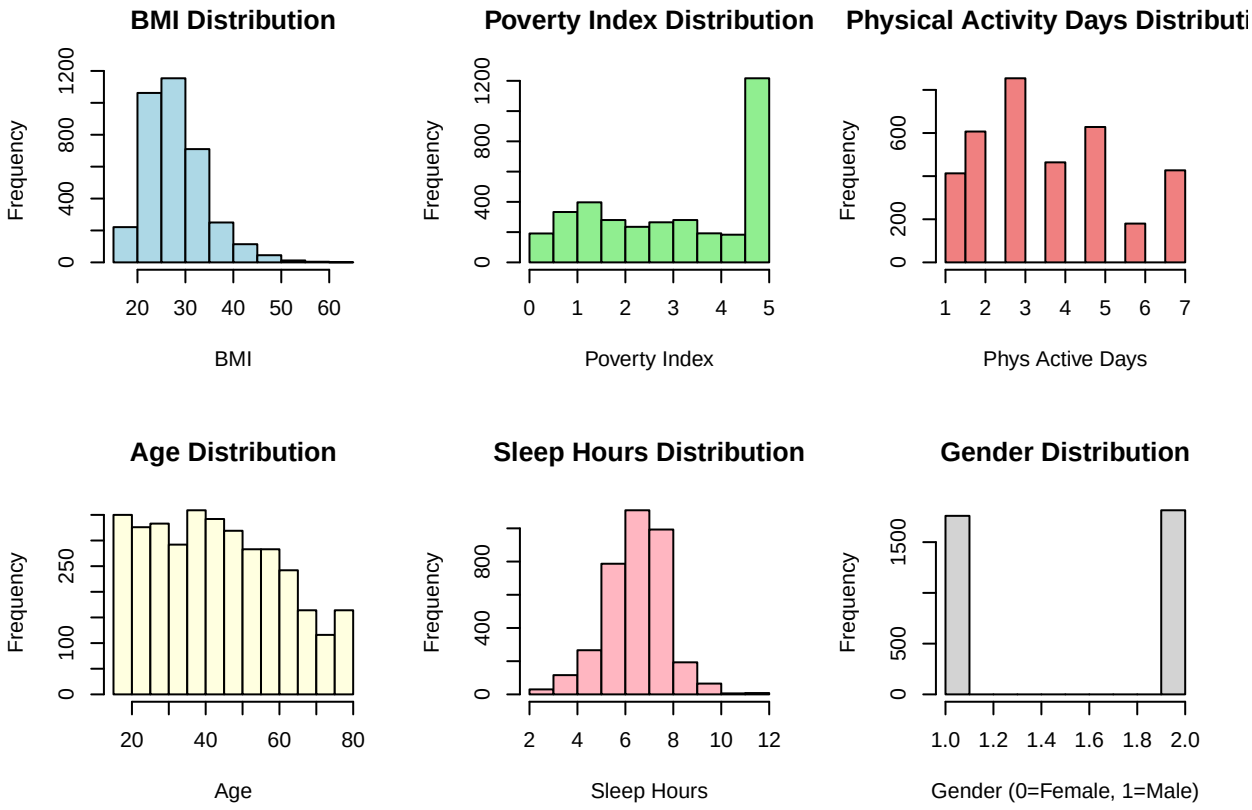

```r
plot(nhanes_data$SleepHrsNight, log(nhanes_data$BMI),
     main = "NAHANES: Physical Active Days vs. SleepHrsNight",
     xlab = "SleepHrsNight",
     ylab = "BMI")

fit <- lm(log(nhanes_data$BMI) ~ nhanes_data$SleepHrsNight)

# Add the fitted regression line
abline(fit, col = "red", lwd = 2)
```




```r
plot(nhanes_data$PhysActiveDays, log(nhanes_data$BMI),
     main = "NAHANES: Physical Active Days vs. PhysActiveDays",
     xlab = "PhysActiveDays",
     ylab = "BMI")
fit <- lm(log(nhanes_data$BMI) ~ nhanes_data$PhysActiveDays)

# Add the fitted regression line
abline(fit, col = "red", lwd = 2)
```




```r
# Plot histograms for each variable
par(mfrow = c(2, 3))  # Arrange plots in 2 rows and 3 columns
hist(nhanes_data$BMI, main="BMI Distribution", xlab="BMI", col="lightblue")
hist(nhanes_data$Poverty, main="Poverty Index Distribution", xlab="Poverty Index", col="lightgreen")
hist(nhanes_data$PhysActiveDays, main="Physical Activity Days Distribution", xlab="Phys Active Days", co
hist(nhanes_data$Age, main="Age Distribution", xlab="Age", col="lightyellow")
hist(nhanes_data$SleepHrsNight, main="Sleep Hours Distribution", xlab="Sleep Hours", col="lightpink")
hist(as.numeric(nhanes_data$Gender), main="Gender Distribution", xlab="Gender (0=Female, 1=Male)", col=
```




```r
par(mfrow = c(1, 1))  # Reset plot layout
```