

GH 811

Is Sleep Duration Associated with Obesity in the United States? In-class Exercise

Background

In this exercise, you will use linear regression to determine the association between BMI and average hours of sleep. You will also explore data graphically, assess model fit and make model predictions. The analyses will be carried out using the constructed dataset “bmi_data.csv”.

Objectives

1. Use `ggplot2` to generate one or more graphs expressing the relationships in your data. Stratify on “smoke” and use linear smoothing with confidence intervals. Label your plots with a title and name the axes appropriately.
2. You are concerned that BMI levels may not be normally distributed. You will need to explore the data to determine which transformation of BMI to use in the regression. Generate a histogram of BMI (make sure you display the density rather than frequency). Also, generate density histograms for all variants of BMI you may consider using in your regression. Which transformation or variant of BMI looks best? (Use this BMI in regression)
3. First, explore correlation between your chosen variant of BMI and sleep. Report the correlation coefficient. Next, run a linear regression with BMI as your dependent variable and sleep as your independent variable. Interpret the β coefficients and the R^2 .
4. Assess the fit of the model by plotting the residuals or using R's `plot()` function. Does the model seem like a good fit?
5. You would like to operationalize sleep into clinically relevant categories. To do this you break sleep into quartiles and create a categorical variable capturing these cut points. You then run a linear regression with BMI as the dependent variable and your newly created sleep variable as the independent variable. (You will need to run `as.factor`) Interpret the β coefficients and the R^2 .
6. Discuss limitations and next steps if you were to continue this analysis.

Presentation

You are not limited to the steps above. Feel free to be creative in your analyses! You will work in teams of 2 or 3 for this exercise and have 1 hour to work within your assigned group. At the end of the allotted time, one team will present their findings to the class.