UM–SJTU Joint Institute

Applied Regression Analysis (STAT4130)

Term Project Review Report

Sleep Efficiency Prediction Using Linear Regression

Instructed by

Dr. Ailin Zhang

July 26, 2023

Name	**ID**
Cao Yanzhuo	520370910021

The report explores the relationships between sleep efficiency (SE), different sleep stages, and individuals' health habits. The report uses linear regression and various techniques to achieve its objectives, including constructing three models, evaluating their performance, and drawing conclusions.

Strengths: 1. The report takes a comprehensive approach to analyze various factors influencing sleep efficiency, including sleep stages, health habits, and lifestyle variables. This allows for a more holistic understanding of the topic. 2. The data set used in this study contains a reasonable number of cases and includes personal information, sleep-related factors, and health habits, which allows for a comprehensive analysis of the factors that influence sleep efficiency. 3. The report highlights the significance of sleep efficiency by linking it to various health conditions, such as snoring, asthma, and cognitive impairment. This emphasizes the importance of quality sleep for physical and mental health. 4. The models constructed in this study are evaluated using root mean squared error, R-squared, F-statistics, and residual vs. fitted values plot, which provides a thorough evaluation of the model performance.

Weaknesses: 1. The residual plots are very weird since the residuals are on the two different edges of the plot in Model 3. Also, in Model 2 the residuals are in two parallel lines, which shows that the residuals are always negative when the fitted value is small and is always positive when the fitted value is large. 2. The number of variables the model regresses is too small, resulting in an adjusted R square that is too small to be meaningful (in Model 2 it is just 0.25). 3. In 2.2.3, I don't understand why the 4 plots can conduct that we should use F-test to determine if we can use the reduced model. I think that we can just do the regression and see if we can delete some variables, and use F-test to check the result. 4. Lack of External Validation: The report does not mention using an external dataset for model validation, which could raise concerns about overfitting or the model's ability to predict beyond the current dataset. 5. Lack of explanation: The study does not provide a clear explanation of the methods used to select the predictor variables or the rationale behind the model construction. This could make it difficult for other researchers to replicate the study or understand the results.

Overall, the report demonstrates commendable effort in exploring the relationships between sleep efficiency, sleep stages, and health habits through linear regression analysis. By addressing the mentioned weaknesses, particularly by incorporating external validation, providing clearer explanations, and ensuring appropriate model specification, the report's quality and impact could be further enhanced. As it stands, the report is scored at 75/100.