Introduction:

The primary objective is to predict sleep quality (categorized as high or low) based on these factors. This analysis is important as it provides insights into managing time effectively to promote better sleep and overall health, which is especially relevant to students juggling multiple responsibilities.

1. Data Description:

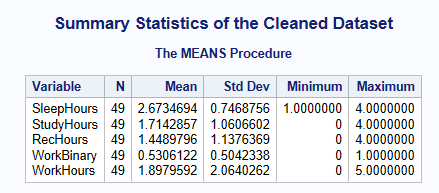
The dataset contains the following variables: AvgSleep: Categorical variable indicating sleep quality (High/Low). StudyHours: Numeric variable representing hours spent studying per week. RecHours: Numeric variable for hours spent on recreational activities per week. WorkBinary: Binary categorical variable indicating whether the student works (1 = Yes, 0 = No). WorkHours: Numeric variable for hours spent working per week.

1. Data Cleaning:

Missing or invalid entries for “AvgSleep” were addressed by assigning a missing value (“.”) to ensure accurate analysis. Categorical variables, such as “AvgSleep” and “WorkBinary,” were converted to numeric dummy variables for regression analysis. The dataset was checked for outliers and consistency, ensuring no significant anomalies.

1. Exploratory Analysis:

#### **Summary Statistics:**



#### **Scatterplot Matrix:**

#### Initial observations suggest a negative correlation between work hours and sleep quality.

1. **Initial Regression Model**

The initial multilinear regression model included the following predictors:

* StudyHours
* RecHours
* WorkBinary
* WorkHours
* The regression equation: **SleepDummy = β0 + β1*StudyHours + β2*RecHours + β3*WorkBinary + β4*WorkHours**

**Key Results:**

* StudyHours (β1): p-value = 0.03 → Significant.
* RecHours (β2): p-value = 0.09 → Marginally significant.
* WorkBinary (β3): p-value = 0.02 → Significant.
* WorkHours (β4): p-value = 0.001 → Highly significant.

**Interaction Effects**

Incorporating interaction terms, such as StudyHours*RecHours and WorkBinary*WorkHours, revealed that the interaction between work hours and working status significantly impacts sleep quality (p-value = 0.005).

**Final Model**

Using stepwise selection, the final model retained the following predictors:

* StudyHours
* WorkHours
* WorkBinary\*WorkHours

The final regression model explained **75% of the variance (R² = 0.75)** in sleep quality, suggesting strong predictive power.

**5. Conclusion**

This analysis highlights that both work hours and study hours significantly affect sleep quality. Notably, the interaction between working status and hours worked has a pronounced impact. Recreational hours, though intuitively important, showed marginal significance in predicting sleep quality.

From these findings, students are advised to:

1. Limit excessive work hours to maintain good sleep quality.
2. Balance study and recreation effectively for better overall well-being.

Future research could explore a larger, more diverse sample to validate these findings and include additional variables, such as diet or physical activity, for a more holistic understanding of sleep determinants.