**STATISTICS ASSIGNMENT**

**Instruction:**

*This is the Group assignment. You will create by yourself your own team of 4 to 5 people.* You are expected to write a formal report based on the guideline questions. However, you can explore yourself further findings based on the data given and techniques from the lectures.

Each of your teams will be assigned a team number from 1 up to 10 (Dzung will assign this number for each team. Your team will select 1000 observations randomly using “set.seed (n)” and “sample” command in R studio as your own data set and analyzing it. The number n in the set.seed command is your order number in the team list. For example, Team number 1 has the seed of 1.

**Objectives and Data description**

1. **Data description:**

This dataset provides a comprehensive overview of various factors affecting student performance in exams. It includes information on study habits, attendance, parental involvement, and other aspects influencing academic success.

Meaning of variables are as follows:

------------------------------------------------------------------------------------------------------------

|  |  |
| --- | --- |
| **Hours\_Studied** | Number of hours spent studying per week. |
| **Attendance** | Percentage of classes attended. |
| **Parental\_Involvement** | Level of parental involvement in the student's education (Low, Medium, High). |
| **Access\_to\_Resources** | Availability of educational resources (Low, Medium, High). |
| **Extracurricular\_Activities** | Participation in extracurricular activities (Yes, No). |
| **Sleep\_Hours** | Average number of hours of sleep per night. |
| **Previous\_Scores** | Scores from previous exams. |
| **Motivation\_Level** | Student's level of motivation (Low, Medium, High). |
| **Internet\_Access** | Availability of internet access (Yes, No). |
| **Tutoring\_Sessions** | Number of tutoring sessions attended per month. |
| **Family\_Income** | Family income level (Low, Medium, High). |
| **Teacher\_Quality** | Quality of the teachers (Low, Medium, High). |
| **School\_Type** | Type of school attended (Public, Private). |
| **Peer\_Influence** | Influence of peers on academic performance (Positive, Neutral, Negative). |
| **Physical\_Activity** | Average number of hours of physical activity per week. |
| **Learning\_Disabilities** | Presence of learning disabilities (Yes, No). |
| **Parental\_Education\_Level** | Highest education level of parents (High School, College, Postgraduate). |
| **Distance\_from\_Home** | Distance from home to school (Near, Moderate, Far). |
| **Gender** | Gender of the student (Male, Female). |
| **Exam\_Score** | Final exam score. |

------------------------------------------------------------------------------------------------------------

The main objective of this project is to explore the key variables, the relationship between them and the evaluation of the factors which affect student performance in the exam. The requirements below are the guidelines that you need to complete. You can do more analysis to explore the data if you find somethings meaningful.

*Some important points in the report and criteria evaluation for evaluation are:*

**Report guidelines:**

1. Introduction
2. Descriptive statistics

* Graphical method (plots, table)
* For one variable
* For the relationship between two variables (2 quanti, 2 quali, 1 quanti and 1 quali)
* Numerical method (for quantitative variables: Summary statistics)
* For one variable
* For the relationship

1. Do the appropriate tests for the Exam Score classifying by some key qualitative variables
2. Conclusion

*Keep in mind that the key variable is The* ***Exam\_Score.*** *This assignment is aimed at the evaluation of your skills in exploring a data set with given objectives.*

**Some criteria:**

* Clearly specifying objective of each analysis and the techniques used
* Clearly and correctly presenting the outputs given by R studio
* Clearly and correctly make the comments of the output
* If possible, make some implications from the output

**Some guideline questions:**

* Produce the plot for some key qualitative and quantitative variable
* Produce the cross tables and plot for the relationship between two qualitative variables (just select variables that you think their relationship is meaningful
* Produce the Box plot of Exam\_Score classifying by some ket qualitative variable
* Scatter plot of Exam\_Score with some key quantitative variables.
* Correlation analysis
* Do some basic tests (t test, Chi square test) for the relationship between Exam.Score and other variables.

**DEADLINE:**

**You are given 6 days to complete the task. The Deadline for submission is at 12pm on Saturday, 21st of December, 2004. Report of all class will be sent to representative (Dzung). Then Dzung sent 1 file to me by email (**[**thenm@neu.edu.vn**](mailto:thenm@neu.edu.vn)**).**

**Late submission will make the Penalty of 50% of your total score.**