Week 2 Assignment

**Student Name**

**Department**

**Course Name**

**Lecturer Name**

**January 21, 2024**

## Introduction

I used the Student Performance dataset to create the solution for this assignment. The dataset includes information on student performance in various subjects. Every row depicts a different student, while the columns provide information regarding aspects of their performance and background. The dataset consist the following variables:

## Nominal Variables

1. Gender: Indicates the gender of the student (male/female).
2. Race/Ethnicity: Represents the racial or ethnic group to which the student belongs (e.g., Group A, Group B, etc.).
3. Lunch: Specifies the type of lunch the student receives (standard/free or reduced).
4. Test Preparation Course: Indicates whether the student completed a test preparation course (completed/none).

## Ordinal Variable

1. Parental Level of Education: Describes the highest level of education attained by the student’s parents.

## Interval/Ratio Variables

6. Math Score: The score obtained by the student in the math subject.  
7. Reading Score: The score obtained by the student in the reading subject.  
8. Writing Score: The score obtained by the student in the writing subject.

## Analysis

#load the dataset  
students\_data <- read.csv('/home/addis/Desktop/Projects/R/StudentsPerformance.csv')  
head(students\_data)

## gender race.ethnicity parental.level.of.education lunch  
## 1 female group B bachelor's degree standard  
## 2 female group C some college standard  
## 3 female group B master's degree standard  
## 4 male group A associate's degree free/reduced  
## 5 male group C some college standard  
## 6 female group B associate's degree standard  
## test.preparation.course math.score reading.score writing.score  
## 1 none 72 72 74  
## 2 completed 69 90 88  
## 3 none 90 95 93  
## 4 none 47 57 44  
## 5 none 76 78 75  
## 6 none 71 83 78

## Frequency Distribution for the Nominal and Ordinal Variables

# Nominal Variables  
gender\_freq <- table(students\_data$gender)  
race\_freq <- table(students\_data$race.ethnicity)  
lunch\_freq <- table(students\_data$lunch)  
test\_prep\_freq <- table(students\_data$test.preparation.course)  
  
# Display the results  
print("Nominal Variables:")

## [1] "Nominal Variables:"

print(gender\_freq)

##   
## female male   
## 518 482

print(race\_freq)

##   
## group A group B group C group D group E   
## 89 190 319 262 140

print(lunch\_freq)

##   
## free/reduced standard   
## 355 645

print(test\_prep\_freq)

##   
## completed none   
## 358 642

## Frequency Distribution for the Ordinal Variables

# Ordinal Variable  
parental\_education\_freq <- table(students\_data$parental.level.of.education)  
parental\_education\_mode <- as.character(names(which.max(parental\_education\_freq)))  
  
print("Ordinal Variable:")

## [1] "Ordinal Variable:"

print(parental\_education\_freq)

##   
## associate's degree bachelor's degree high school master's degree   
## 222 118 196 59   
## some college some high school   
## 226 179

print(paste("Mode:", parental\_education\_mode))

## [1] "Mode: some college"

## Cmmon Smmary Measures for the Interval or Ratio Variable

# Ratio Variables  
math\_summary <- summary(students\_data$math.score)  
reading\_summary <- summary(students\_data$reading.score)  
writing\_summary <- summary(students\_data$writing.score)  
  
print("Ratio Variables:")

## [1] "Ratio Variables:"

print(math\_summary)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 57.00 66.00 66.09 77.00 100.00

print(reading\_summary)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 17.00 59.00 70.00 69.17 79.00 100.00

print(writing\_summary)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10.00 57.75 69.00 68.05 79.00 100.00

## Interpretation

The frequency distribution illustrates a fairly even distribution of gender, with 518 female students and 482 male students in the dataset. The Race/ethnicity varies, with Group c being the most prevalent. From the data, we observe that most students have a standard lunch, and a significant portion completed test preparation course. Parental education shows diversity, with “some college the most common level of parental education.

The summary measures for ratio variables provide valuable insights into the distribution of student performance. The minimum math score is 0 while the maximum score of 100 indicating that some students scored the lowest possible, while others attained perfect scores. The median scores for math (66.00), reading (70.00), and writing (69.00) provide central tendencies, showcasing the middle points of the distributions. The mean scores, which are 66.09 for math, 69.17 for reading, and 68.05 for writing, indicate the average performance across students. The interquartile ranges (1st Qu. to 3rd Qu.) reflect the spread of scores, with higher variability observed in reading compared to math and writing.

## Reflection

Understanding data types and selecting appropriate descriptive statistics is crucial for data analysis. Different types of data, like nominal, ordinal, and interval/ratio, require different approaches. Nominal data uses frequency distributions, ordinal data uses median and quartiles, and interval/ratio data uses mean, standard deviation, and variance. Selecting the right descriptive statistics helps uncover patterns, trends, and relationships.