Week 2 Assignment

Student Name

2024-01-29

# Overview of the Dataset

#load the data  
student\_data<-read.csv("/home/addis/Desktop/Projects/R/Data Analysis/StudentsPerformance.csv")  
head(student\_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

# Variable Selection:

## Qualitative Variables:

### 1. Gender:

This variable represents the gender of the students.

### 2. Test Preparation Course:

This variable indicates whether students completed a test preparation course or not.

## Quantitative Variables:

###1. Math Score: A numerical score representing students’ performance in math.

### 2. Reading Score:

A numerical score representing students’ performance in reading.

# Analysis

## Qualitative Variables - Contingency Table and Association Calculation:

### Contingency Table

# Create a contingency table for Gender and Test Preparation Course  
cont\_table\_qual <- table(student\_data$gender, student\_data$test.preparation.course)  
cont\_table\_qual

##   
## completed none  
## female 184 334  
## male 174 308

### Association Calculation:

# Perform the chi-squared test  
chi\_squared\_test <- chisq.test(cont\_table\_qual)  
chi\_squared\_test

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: cont\_table\_qual  
## X-squared = 0.015529, df = 1, p-value = 0.9008

## Quantitative Variables - Correlation and Scatter Plot:

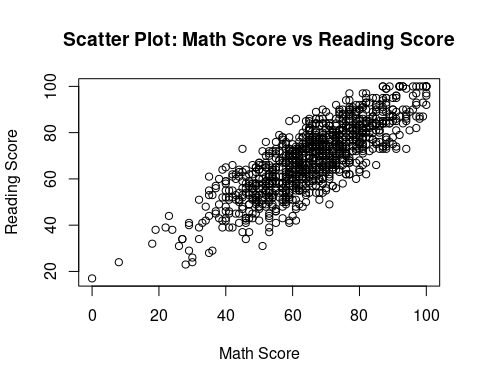
### Calculate the correlation between Math Score and Reading Score

correlation\_coefficient\_quant <- cor(student\_data$math.score, student\_data$reading.score)  
correlation\_coefficient\_quant

## [1] 0.8175797

### Scatter plot for Math Score and Reading Score

# Create a scatter plot for Math Score and Reading Score  
plot(student\_data$math.score, student\_data$reading.score, main = "Scatter Plot: Math Score vs Reading Score", xlab = "Math Score", ylab = "Reading Score")



# Interpretation

## Qualitative Variables

### Contingency table

The contingency table indicates that the dataset had more females (518 in total) compared to males (482). The test preparation course was attended by more females than the males (184 vs. From 174, there are relatively more female learners who did not complete the test preparation course than males (334 vs. 308).

### Chi-squared Statistic (X-squared)

Chi-squared statistic is a measure of difference between the observed and expected frequencies in contingency table. In this case, our calculated value is very low (0.015529).

### Degrees of Freedom (df):

The degrees of freedom simply refer to the number values in a statistic’s final computation that is free or open for speculation. In this case, df = 1 implying a binary outcome (done or none).

### P-value:

The p-value is important in determining whether the differences observed are significant. At p-value of 0.90, the evidence is not strong enough to reject this null hypothesis. Alternatively, there is no statistically significant relationship between gender and course completion.

## Interpretation- Quantitative Variables

The correlation coefficient value is 0.8176 showing a strong relationship between math scores and reading scores. A correlation coefficient that is close to 1 describes a strong positive linear relationship. In this connection, the tendency for reading scores to rise is very high when math results increase.

# Reflection

Establishing the link between gender and test preparation course completion is essential for detecting possible differences. In case the completion rates dramatically vary depending upon gender this fact may indicate areas for special interventions. Moreover, insights into the high positive correlation of math and reading performances are valuable in an educational setting. It means that a good math student is likely to be a better reader while the opposite scenario might also apply. This insight could influence teaching approaches, curriculum design or intervention to help students in both subjects.