Students Survey

Statistics Assignment

Instructions

- Analyze the provided dataset using R.
- Answer all questions, supporting your findings with charts, tables, and explanations.
- Submit a well-structured report (PDF format), including all code and outputs.

Part 1: Descriptive Statistics

1. **Basic Summary**:
    - For the following columns: Age, Mid-exam score, and GPA:
        - Calculate the mean, median, standard deviation, range, minimum, and maximum.
        - Identify the most frequent (mode) value.
2. **Frequency Distributions**:
    - Create frequency tables for:
        - Gender
        - Year of study
        - Distance from home to the university
3. **Percentile Analysis**:
    - Calculate the 25th, 50th, and 75th percentiles for GPA.

Part 2: Exploratory Data Analysis (EDA)

4. **Age Distribution**:
    - Plot a histogram or density plot for the Age column.
    - Answer:
        - What is the most common age?
        - Are there any outliers?
5. **Satisfaction Levels**:
    - Compare the satisfaction levels (university experience and BI major) across genders using boxplots.
    - Identify which group has the highest and lowest satisfaction ratings.
6. **Home Distance vs. Recommendation**:
    - Investigate the relationship between home distance from the university and the likelihood of recommending the university using:
        - A scatterplot or boxplot
        - Statistical summary (mean and median likelihood by distance group)
7. **Correlation Analysis**:
    - Calculate and visualize correlations between:
        - Mid-exam score
        - GPA

        - University satisfaction
        - Recommendation likelihood
    - Highlight the strongest relationships.

Part 3: Inferential Statistics

8. **Hypothesis Testing (T-Test)**:
    - Test if there is a significant difference in GPA between:
        - Male and female students
        - Students who work versus those who don't
9. **ANOVA**:
    - Analyze if the year of study has a significant effect on:
        - GPA
        - Satisfaction with the university experience
10. **Chi-Square Test**:
    - Test if there is an association between:
        - High school education type and GPA categories (e.g., low, medium, high)
        - Working status and satisfaction with the university experience

Part 4: Regression Analysis

11. **Simple Linear Regression**:
    - Build a model predicting GPA using the mid-exam score.
    - Assess:
        - R-squared value
        - Significance of the predictor
12. **Multiple Linear Regression**:
    - Extend the model to include:
        - Satisfaction with the university
        - Satisfaction with the BI major
        - Type of work (encoded appropriately if categorical)
    - Interpret:
        - Coefficients for each predictor
        - Overall model fit
13. **Logistic Regression**:
    - Build a logistic regression model to predict whether students would recommend the university (Yes/No) based on:
        - Satisfaction levels
        - Type of work
        - Year of study

Part 5: Specific Questions

14. **Key Insights**:

- Answer the following:
  - What is the average GPA for students in each year of study?
  - Which group has the highest average satisfaction with the university?
  - Are students who live closer (<5 KM) more likely to recommend the university compared to those farther away?

15. **Outliers**:
    - Identify potential outliers in:
      - GPA
      - Mid-exam scores
    - Use boxplots or IQR methods to justify your findings.

Part 6: Bonus Tasks

16. **Clustering**:
    - Perform clustering (e.g., K-means) on:
      - GPA
      - Mid-exam score
      - Satisfaction levels
    - Analyze the clusters:
      - How many clusters are optimal?
      - What characterizes each cluster?
17. **Interactive Visualization**:
    - Create an interactive dashboard (using R's Shiny or Python's Plotly/Dash) to visualize:
      - GPA distributions
      - Satisfaction levels by group
      - Relationships between predictors and outcomes