

## Analysis report for the Demographic Behavioral

Almazan, Naomi Heart C., Tajon, Jaiven Kamyille

### Introduction

This dataset contains health and behavioral records of individuals from various backgrounds, aimed at exploring the relationships between demographics, lifestyle habits, and health indicators. It includes variables such as age, sex, height, weight, BMI, physical activity hours per week, smoking/drinking status, patient satisfaction, and health literacy.

### Methods used for analysis

The dataset was first cleaned by removing irrelevant columns and handling missing values using (na.omit). Variable names were standardized for clarity, and categorical variables were converted into factors with appropriate labels. Descriptive statistics were generated using the psych package to summarize numerical variables. Several visualizations were created using ggplot2, including histograms, boxplots, scatter plots, and bar charts to explore distributions and relationships within the data. A two-sample t-test was conducted to examine mean BMI differences between sexes. Additionally, a correlation matrix was computed for all numeric variables to identify potential linear relationships.

### Key results and figures

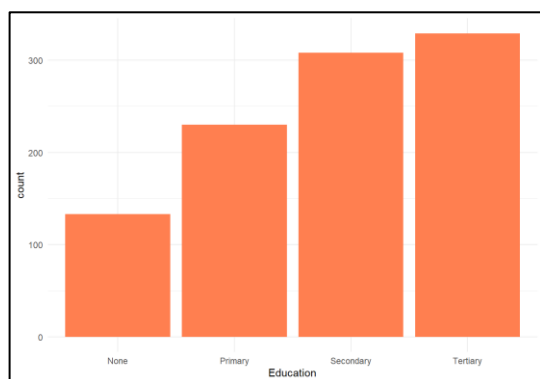


Figure 1: Bar Plot of Levels in Education

Figure 1. The barplot shows that secondary and tertiary education levels are the most common among the individuals in the dataset. This indicates that a majority of individuals have moderate to high educational attainment.

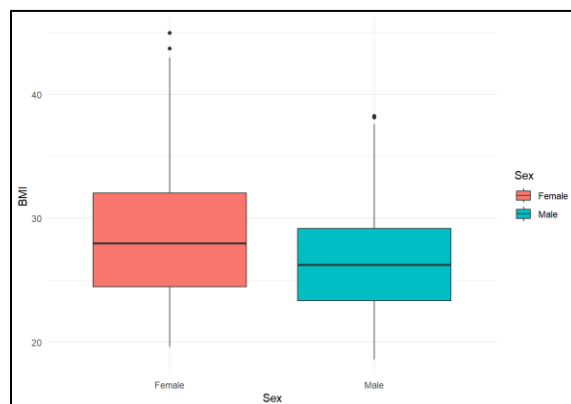


Figure 2: Boxplot of BMI and Sex

Figure 2. This boxplot reveals that Females tend to have a slightly higher median BMI compared to males. However, the overlap in interquartile ranges and similar spread in both sexes indicates **no substantial difference** in BMI distribution.

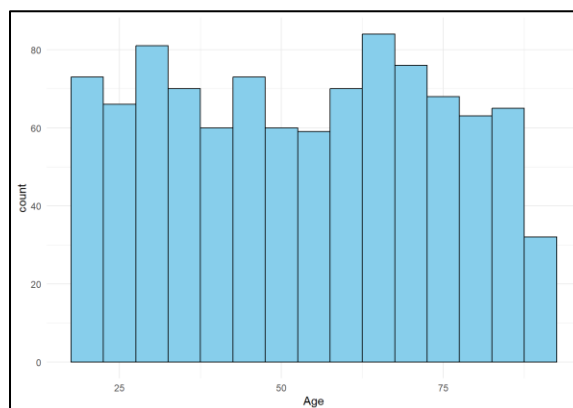
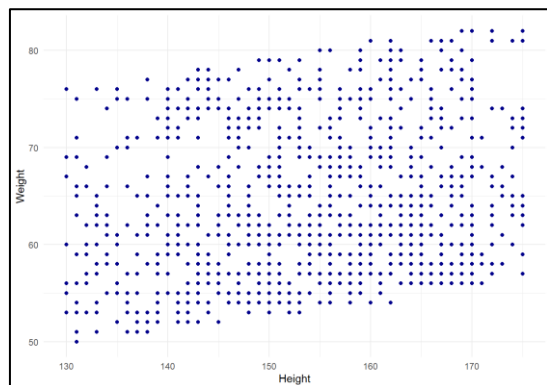


Figure 3: Histogram of Age

Figure 3. The majority of the patients fall between the ages of 20 to 50 years old. This suggests that the dataset mostly represents young to middle-aged adults, and a few old individuals. This Age profile influences health behaviour and risk factors present in the population(dataset).



**Figure 4. Scatterplot of Height vs Weight**

Figure 4. Both height and weight have a clear positive linear relationship; as height increases, weight also increases. This is biologically expected and confirms that body structure is consistent with proportionality. However, some outliers occurs, indicating this is underweight and overweight individuals for their height.

Statistic	Value
t-value	6.70
Degrees of Freedom (df)	956.37
p-value	$3.57 \times 10^{-11}$
95% Confidence Interval	[1.41, 2.57]

**Table 1. T-Test**

There is a statistically significant difference in mean BMI between females and males ( $p < 0.001$ ). Females have a higher average BMI than males by about 2 units. Since the confidence interval does not include 0, this difference is unlikely due to random chances.

	Weight	Height	BMI	PA Hours	Health Lit
Weight	1	0.19	0.51	-0.03	0.10
Height	0.19	1	-0.74	0.01	-0.01
BMI	0.51	-0.74	1	-0.04	0.08
PA Hours	-0.03	0.01	-0.04	1	0.01
Health Lit	0.10	-0.01	0.08	0.01	1

**Table 2. Correlation Matrix**

- BMI is positively correlated with Weight ( $r = 0.51$ ), as expected.
- BMI is strongly negatively correlated with Height ( $r = -0.74$ ), since taller individuals may have lower BMI for the same weight.
- Health Literacy is weakly positively correlated with BMI and Weight, but not strongly enough to imply causation.
- Physical Activity Hours has very weak or negligible correlation with most other variables.

### Interpretation

The dataset provided an overview of patients' demographic, behavioral, and health-related factors. It included variables such as age, sex, BMI, physical activity, education, and health literacy. From this analysis, the age distribution was skewed toward young to middle-aged adults, suggesting a generally healthy and active group. While BMI was distributed across a

broad range, the comparison between male and female BMI values showed no statistically significant difference, indicating that sex alone may not be a strong determinant of BMI within this population.

The result was the strong positive correlation between height and weight, showing internal consistency and biological validity in the dataset. However, scatter plots revealed outliers in individuals whose weight did not match their height indicating potential underweight or overweight cases. Furthermore, educational level appeared skewed toward secondary and tertiary levels, which reflects a population that may have greater access to health knowledge and services. Their Health literacy and patient satisfaction were positively correlated, indicating that individuals with a better understanding of health-related information tend to report higher satisfaction with their healthcare experience.

## Conclusion

The findings suggest that while biological factors like height and weight maintain expected relationships, behavioral and social variables (like education and satisfaction) offer insights into health management. The lack of strong associations between BMI and individual behaviors such as sex or age points toward the need for a holistic, personalized approach in health interventions. Improving health literacy, supporting education, and promoting balanced lifestyle habits appear to be effective strategies in enhancing overall patient satisfaction and well-being.