Final Deliverable

Title: Which factors influence students' drinking habits?

Dataset: Stellenbosch University student survey (Google Form, Kaggle dataset)
https://www.kaggle.com/datasets/joshuanaude/effects-of-alcohol-on-student-performance

1. Introduction and Background:

College years are considered by many to be a time of personal growth, social connections, as well as new experiences, including alcohol consumption. In this project we will study the phenomenon of alcohol consumption among university students using data collected through a Google Forms survey from the students of Stellenbosch University in South Africa. More specifically, the aim of this project is to identify how socialising frequency and field of study influence drinking behaviour, while also assessing the stereotype that males consume more alcohol than females. These relationships will be explored using the methods of hypothesis testing and logistic regression, along with the method of descriptive statistics, which will provide basic information and visual representations on the concept of alcohol consumption.

2. Research Questions

- How does socializing frequency influence alcohol consumption?
- Is the field of study related to the amount of alcohol consumption per night?
- Are male students more likely to consume alcohol than female students?

3. Variables

Dependent Variable (Y): Alcohol consumption (drinks per night)
Independent Variables (X): Gender, Field of study, Socializing frequency

4. Descriptive Statistics

We performed Descriptive Statistics for alcohol consumption given socializing frequency to examine the way socializing frequency affects the number of drinks students consume per night.

Socializing frequency	Mean alcohol consumption	Median alcohol consumption	Standard Deviation
Never (0.0)	2.41	0.0	3.27
Once a week (1.0)	5.14	6.5	2.80
Only on weekends (1.5)	4.16	4.0	2.48
Twice a week (2.0)	5.44	6.5	2.66
Three times a week (3.0)	6.19	6.5	2.66
More than four times a week (5.0)	6.84	6.5	2.32

5. Hypothesis Testing

We ran one-way ANOVA test to compare the mean alcohol consumption per night across the eight faculties, and see if alcohol consumption is associated with the field of study.
We used $\alpha = 0.05$.

Null Hypothesis (H_0): The mean alcohol consumption is the same across all fields of study.

Alternative Hypothesis (H_1): At least one field of study has a different mean alcohol consumption.

ANOVA results: F-statistic: 1.678, p-value: 0.113

→Because $p = 0.113 \geq 0.05$, we fail to reject the Null Hypothesis(H_0).

6.Logistic regression

Our goal was to perform logistic regression in order to model alcohol consumption as a function of the genders and examine whether male students drink more than female ones.

Dependent variable: Y = 1 for consuming 4 more drinks/night, 0 for less

Independent variable: $X_i = 1$ if male, 0 if female

General model: $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 * X_i$ where $\beta_0 = 0.25$, $\beta_1 = 1.39$, $p_i = P(Y=1|(X_i))$

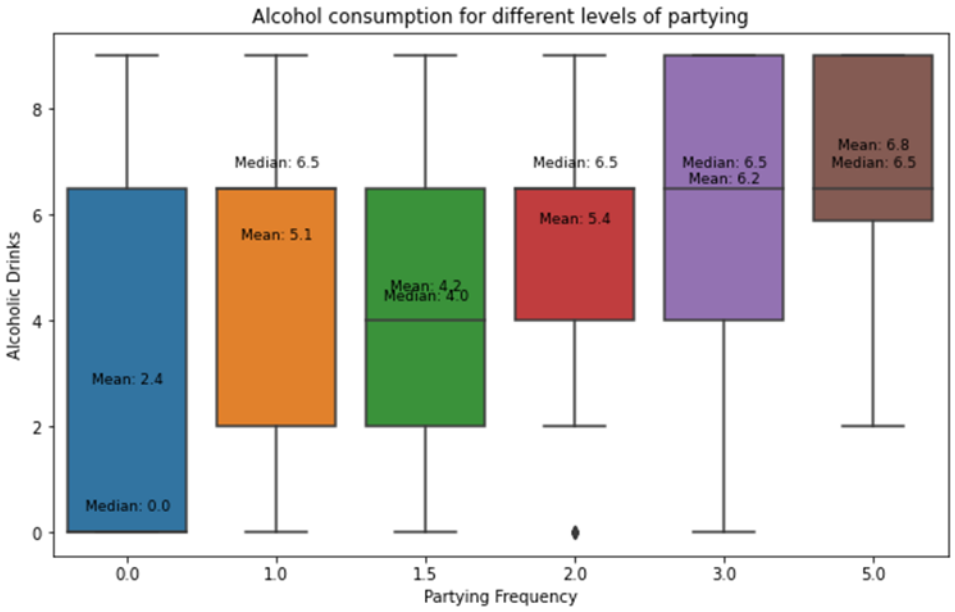
Predicted Probability(male): $p_1 = 0.838$

Predicted Probability(female): $p_0 = 0.563$

Confidence interval: (0.927,1.851)

Relative Risk: RR = 1.488

7.Plots



8. Conclusions

- Socializing frequency and alcohol consumption are positively correlated. Students who go out more often tend to consume more drinks on average per night.
- The One-Way ANOVA test showed that we cannot conclude that the field of study significantly influences the amount of alcohol consumed per night.
- The relative risk we conducted through logistic regression reveals that males are 48.8% more likely than females to be high alcohol consumers, which indicates that gender is a highly significant predictor of high alcohol consumption.