````
---
title: "Drinking Habits Analysis"
author: "Isobel Bodefeld"
date: "`r Sys.Date()`"
output: html_document
---

```{r setup, include=FALSE}
# Load required libraries
library(tidyverse)
```

# Introduction
This report analyzes drinking habits based on marital and parental status. We will examine the frequency of alcoholic beverage consumption and perform a statistical test to see if there is a significant difference between groups.

# Data Import
Let's start by loading the data.

```{r}
# Load the data
data <- read_csv("~/PSY 420M Psychological Methods and Statistics/Project/data.csv")

# Display the first few rows of the data
head(data)
```

# Frequency Plot
The frequency plot below shows the distribution of the number of days per week an alcoholic beverage is consumed, separated by group (married men with kids vs. single men without kids).

```{r}
# Frequency plot
ggplot(data, aes(x = value, color = group)) +
  geom_freqpoly(binwidth = 1) +
  labs(
    title = "Frequency Plot of Drinking Habits",
    subtitle = "Comparison between Married Men with Kids and Single Men without Kids",
    x = "# of Days per Week an Alcoholic Beverage is Consumed",
    y = "Count",
    color = "Group"
  ) +
  theme_minimal()
```

# Boxplot
The boxplot below illustrates the impact of marital and parental status on drinking habits.

```{r}
# Boxplot
ggplot(data, aes(x = group, y = value)) +
  geom_boxplot() +
  labs(
    title = "Impact of Marital and Parental Status on Drinking Habits",
    x = "Marital and Parental Status",
    y = "# of Days per Week an Alcoholic Beverage is Consumed"
  ) +
  theme_minimal()
```
````

Statistical Analysis
We perform an independent samples t-test to determine if there is a significant difference in drinking habits between the two groups.

```{r}
# Perform t-test
t_test_result <- t.test(value ~ group, data = data)

# Print the t-test result
t_test_result
```

Summary Statistics
Below are the summary statistics for each group, including the mean, standard deviation, standard error, and 95% confidence intervals.

```{r}
# Summary statistics by group
data_summary <- data %>%
  group_by(group) %>%
  summarize(
    count = n(),
    mean_value = mean(value),
    sd_value = sd(value),
    se_value = sd_value / sqrt(count),
    ci_95 = 1.96 * se_value
  )

# Display the summary statistics
data_summary
```

Conclusion
This analysis explored the drinking habits of different groups based on marital and parental status. The visualizations and statistical analysis provide insights into how these factors may influence drinking behavior.

Summary Statistics Interpretation

- **Groups**:
  - **Married with Kids**: This group consists of 100 men who are married and have children.
  - **Single without Kids**: This group consists of 100 men who are single and do not have children.

- **Mean Number of Days per Week Consuming Alcohol**:
  - **Married with Kids**: The average number of days per week that men in this group consume alcohol is approximately 1.98 days.
  - **Single without Kids**: The average number of days per week that men in this group consume alcohol is higher, at approximately 2.87 days.

- **Standard Deviation (SD)**:
  - **Married with Kids**: The standard deviation is 1.01, indicating that the number of days per week that alcohol is consumed by individuals in this group varies by about 1.01 days around the mean of 1.98 days.
  - **Single without Kids**: The standard deviation is 1.41, indicating more variability in the number of days per week that alcohol is consumed in this group, with a spread of about 1.41 days around the mean of 2.87 days.

- **Standard Error (SE)**:
  - **Married with Kids**: The standard error is 0.101, which reflects the precision of the sample mean. It suggests that if we were to take multiple samples of the same size, the mean number of days per week consuming alcohol would typically differ by about 0.101 days from the actual population mean.

  - **Single without Kids**: The standard error is 0.141, indicating slightly less precision in estimating the population mean compared to the "marriedwithkids" group.

- **95% Confidence Interval (CI)**:
  - **Married with Kids**: The 95% confidence interval for the mean is approximately 1.98 ± 0.199, or (1.78, 2.18). This means that we are 95% confident that the true mean number of days per week that married men with kids consume alcohol falls between 1.78 and 2.18 days.
  - **Single without Kids**: The 95% confidence interval for the mean is approximately 2.87 ± 0.277, or (2.59, 3.15). This means that we are 95% confident that the true mean number of days per week that single men without kids consume alcohol falls between 2.59 and 3.15 days.

Overall Interpretation
- **Comparison**: The data suggests that single men without kids tend to consume alcohol more frequently than married men with kids. On average, single men without kids consume alcohol about 0.89 days more per week than their married counterparts with children.

- **Variability**: The single men without kids group also shows greater variability in drinking habits, as indicated by a higher standard deviation.

- **Confidence**: The confidence intervals for the two groups do not overlap significantly, which might suggest a statistically significant difference in the mean number of drinking days per week between the two groups.