Problem Set 1

Byeolha Kim

Question 1: Briefly describe the dataset. What country did you choose? How many respondents are there in the survey, and when were the interviews conducted?

```r
# Get basic dataset information
country <- "Mozambique"
num_respondents <- nrow(df)
interview_year <- "2022"  # Based on provided source

# Print dataset description
cat("The dataset contains", num_respondents, "respondents from", country,
    ", and the interviews were conducted in", interview_year, ".")
```

```
## The dataset contains 1120 respondents from Mozambique , and the interviews were conducted in 2022 .
```

Question 2: Describe your respondents. Using appropriate descriptive statistics, tell me about their ages, distribution of male vs female respondents, language, etc.

```r
# Analyze demographic characteristics (age, gender)
respondent_summary <- df %>%
  summarise(
    total_respondents = n(),
    avg_age = mean(Q1, na.rm = TRUE),  # Replace Q1 with the correct column name for age
    male_count = sum(Q101 == 1, na.rm = TRUE),  # Replace Q101 with gender column
    female_count = sum(Q101 == 2, na.rm = TRUE)
  )

# Display respondent summary
kable(respondent_summary, caption = "Summary of Respondents' Demographics")
```

Table 1: Summary of Respondents' Demographics

total_respondents	avg_age	male_count	female_count
1120	34.29286	1112	1

```r
library(knitr)

# Create a summary table with relevant statistics
summary_data <- data.frame(
  Metric = c("Total Respondents", "Average Age", "Male Respondents", "Female Respondents"),
  Value = c(1120, 34.29286, 1112, 1)
)

# Display the formatted table
kable(summary_data, caption = "Summary of Respondents' Demographics")
```

Table 2: Summary of Respondents' Demographics

Metric	Value
Total Respondents	1120.00000
Average Age	34.29286
Male Respondents	1112.00000
Female Respondents	1.00000

Question 3: Describe attitudes about the economic and political influence of China (Q78A) in your data.

```r
# Frequency analysis for China's influence (Q78A)
china_influence <- df %>%
  count(Q78A) %>%
  mutate(percentage = n / sum(n) * 100)

# Display results
kable(china_influence, caption = "Perceptions of China's Influence")
```

Table 3: Perceptions of China's Influence

Q78A	n	percentage
1	72	6.4285714
2	168	15.0000000
3	91	8.1250000
4	313	27.9464286
5	310	27.6785714
8	2	0.1785714
9	164	14.6428571

Result: The survey results on perceptions of China's influence (Q78A) indicate that the largest proportion of respondents view China's influence positively, with 27.95% rating it as "4" and 27.68% rating it as "5" on the scale. A smaller portion of respondents, 6.43%, rated China's influence at the lowest level ("1"), while 15% rated it as "2". Notably, 14.64% of respondents provided a response of "9," which may indicate missing or uncertain responses. Overall, the data suggests that opinions on China's influence are generally favorable, with a significant portion of respondents perceiving it positively.

Question 4: Repeat this process for Q78B about the influence of the United States.

```
# Frequency analysis for U.S. influence (Q78B)
us_influence <- df %>%
  count(Q78B) %>%
  mutate(percentage = n / sum(n) * 100)

# Display results
kable(us_influence, caption = "Perceptions of U.S. Influence")
```

Table 4: Perceptions of U.S. Influence

Q78B	n	percentage
1	47	4.1964286
2	146	13.0357143
3	162	14.4642857
4	298	26.6071429
5	235	20.9821429
8	2	0.1785714
9	230	20.5357143

Result: The survey results on perceptions of U.S. influence (Q78B) show that the majority of respondents have a positive view, with 26.61% rating it as "4" and 20.98% rating it as "5". However, 20.54% of respondents provided a response of "9," which may indicate missing or uncertain responses. Comparatively, fewer respondents rated the U.S. influence negatively, with 4.20% giving it the lowest rating ("1"). Overall, the data suggests a generally favorable perception of U.S. influence, though uncertainty remains significant among respondents.

Question 5: Use a paired t-test to evaluate the difference between perceptions.

```
# Clean the data for valid responses (1-5 scale)
df_clean <- df %>%
  mutate(
    Q78A_clean = ifelse(Q78A %in% 1:5, Q78A, NA),
    Q78B_clean = ifelse(Q78B %in% 1:5, Q78B, NA)
  ) %>%
  drop_na(Q78A_clean, Q78B_clean)

# Perform paired t-test
t_test_result <- t.test(df_clean$Q78A_clean, df_clean$Q78B_clean, paired = TRUE)

# Display t-test results
t_test_result
```

```
##
##  Paired t-test
##
```

3

```
## data:  df_clean$Q78A_clean and df_clean$Q78B_clean
## t = 0.24582, df = 856, p-value = 0.8059
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.07334964  0.09435314
## sample estimates:
## mean difference
##      0.01050175
```

Findings: The paired t-test results indicate that there is no statistically significant difference in the perceptions of China's and the U.S.'s influence among respondents. The test yielded a *t-value of 0.24582* with *856 degrees of freedom*, and *a p-value of 0.8059*, which is well above the common significance threshold of 0.05. This suggests that the null hypothesis, which states that the mean difference between the two perceptions is zero, cannot be rejected.

Additionally, *the 95% confidence interval* for the mean difference ranges from *-0.0733 to 0.0944*, further supporting that any observed differences are likely due to random variation. The estimated mean difference of *0.0105* is very small, indicating that, on average, respondents rated China's and the U.S.'s influence similarly.

Conclusion: Overall, the results suggest that perceptions of China's and the U.S.'s influence in Mozambique do not differ significantly.