

Problem Set 1

Edgar Aguilar

January 29

Answers to questions for PS1

Question #1

- I am using the data set from Congo-Brazzaville.
- The survey includes information from 1200 respondents.
- The responses were collected between 2023-06-20 and 2023-07-17.

Question #2

The ages were between 18 and 85, with a mean of 37.73 and median of 35. The other characteristics of the respondents are in the table below:

```
# Use kableExtra to show a final nicer table
library(kableExtra)

## Gender ----
# Show unique values in the GENDER column -- 1 = Man 2 = Woman. -- to check data looks good.

# Replace 1 with "Man" and 2 with "Woman", and calculate percentages. Generate tibble.
gender_tibble <- df1 %>%
  mutate(
    Q100 = case_when(
      Q100 == 1 ~ "Man",
      Q100 == 2 ~ "Woman",
    )
  ) %>%
  count(Q100, name = "Count") %>%
  mutate(Percentage = round((Count / sum(Count)) * 100, 2), Category = "Gender")
##category variable will help me put tibbles together in a single table after.

## Region ----
## replace numbers with region names bases on codebook and create tibble
region_tibble <- df1 %>%
  mutate(
    REGION = case_when(
      REGION == 1780 ~ "Brazzaville",
      REGION == 1781 ~ "Pointe-Noire",
      REGION == 1782 ~ "Kouilou",
      REGION == 1783 ~ "Niari",
      REGION == 1784 ~ "Bouenza",
      REGION == 1785 ~ "Lekoumou",
```

```

    REGION == 1786 ~ "Pool",
    REGION == 1787 ~ "Plateaux",
    REGION == 1788 ~ "Cuvette",
    REGION == 1789 ~ "Cuvette Ouest",
    REGION == 1790 ~ "Sangha",
    REGION == 1791 ~ "Likouala ",
  )
) %>%
count(REGION, name = "Count") %>%
mutate(Percentage = round((Count/sum(Count))*100, 2), Category = "Region")

## Location ----
# Replace 1 with "Urban" and 2 with "Rural", and calculate percentages
urbrur_tibble <- df1 %>%
  mutate(
    URBRUR = case_when(
      URBRUR == 1 ~ "Urban",
      URBRUR == 2 ~ "Rural",
    )
  ) %>%
count(URBRUR, name = "Count") %>%
mutate(Percentage = round((Count / sum(Count)) * 100, 2), Category = "Location")

## Language ----
# Replace and calculate
language_tibble <- df1 %>%
  mutate(
    Q2 = case_when(
      Q2 == 2 ~ "Français",
      Q2 == 1780 ~ "Lingala",
      Q2 == 1781 ~ "Kituba",
      Q2 == 1782 ~ "Lari",
      Q2 == 1783 ~ "Teke",
      Q2 == 9995 ~ "Other",
      Q2 == 9998 ~ "A refusé de répondre",
      Q2 == 9999 ~ "Ne sait pas",
      Q2 == -1 ~ "Manquant ",
    )
  ) %>%
count(Q2, name = "Count") %>%
mutate(Percentage = round((Count / sum(Count)) * 100, 2), Category = "Language")

## Summary Disruptive Statistics ----
# combine tibbles
summary_table <- bind_rows(
  gender_tibble,
  region_tibble,
  urbrur_tibble,
  language_tibble
)

```

```

# Merge Q100, REGION, URBRUR, and Q2 into one column called 'Variable'
updated_summary_table <- summary_table %>%
  mutate(
    Variable = coalesce(as.character(Q100), as.character(REGION), as.character(URBRUR), as.character(Q2))
  ) %>%
  select(Category, Variable, Count, Percentage) # Rearrange columns to the desired order

# Display the updated table
updated_summary_table %>%
  knitr::kable(caption = "Descriptive Statistics Summary Table") %>%
  kable_styling(full_width = FALSE)

```

Table 1: Descriptive Statistics Summary Table

Category	Variable	Count	Percentage
Gender	Man	599	49.92
Gender	Woman	601	50.08
Region	Bouenza	88	7.33
Region	Brazzaville	464	38.67
Region	Cuvette	48	4.00
Region	Cuvette Ouest	24	2.00
Region	Kouilou	32	2.67
Region	Lekoumou	32	2.67
Region	Likouala	40	3.33
Region	Niari	72	6.00
Region	Plateaux	56	4.67
Region	Pointe-Noire	240	20.00
Region	Pool	80	6.67
Region	Sangha	24	2.00
Location	Rural	360	30.00
Location	Urban	840	70.00
Language	Français	609	50.75
Language	Kituba	275	22.92
Language	Lari	15	1.25
Language	Lingala	248	20.67
Language	Other	25	2.08
Language	Teke	28	2.33

Question #3

- 21% of respondents perceive China's economic and political influence as negative while 68% see it as positive.
- 9% consider China's influence neither positive nor negative and around 2% are unsure.

```

# create frequency table
China_attitudes_frequency <- df1 %>%
  count(Q78A, name = "Count") %>%
  mutate(
    Percentage = round((Count / sum(Count)) * 100, 2),
    Response = case_when(
      Q78A == 1 ~ "Very negative",

```

```

Q78A == 2 ~ "Somewhat negative",
Q78A == 3 ~ "Neither positive nor negative",
Q78A == 4 ~ "Somewhat positive",
Q78A == 5 ~ "Very positive",
Q78A == 8 ~ "Refused",
Q78A == 9 ~ "Don't know",
Q78A == -1 ~ "Missing",
)
) %>%
select(Response, Count, Percentage)

# Display the table China
China_attitudes_frequency %>%
  knitr::kable(
    caption = "Attitudes toward China Table"
  ) %>%
  kable_styling(
    full_width = FALSE
  )

```

Table 2: Attitudes toward China Table

Response	Count	Percentage
Very negative	159	13.25
Somewhat negative	94	7.83
Neither positive nor negative	107	8.92
Somewhat positive	424	35.33
Very positive	397	33.08
Don't know	19	1.58

Question #4

- 21% of respondents perceive USA's economic and political influence as negative while 51% see it as positive.
- 25% of respondents consider it neither positive nor negative and 2% do not know.
- The perception of US influence as positive is 17% lower than China, but there is a greater percentage of people (16% more in comparison with China) that consider it neither positive or negative.

```

#create frequency table
USA_attitudes_frequency <- df1 %>%
  count(Q78B, name = "Count") %>%
  mutate(
    Percentage = round((Count / sum(Count)) * 100, 2),
    Response = case_when(
      Q78B == 1 ~ "Very negative",
      Q78B == 2 ~ "Somewhat negative",
      Q78B == 3 ~ "Neither positive nor negative",
      Q78B == 4 ~ "Somewhat positive",
      Q78B == 5 ~ "Very positive",
      Q78B == 8 ~ "Refused",
      Q78B == 9 ~ "Don't know",
      Q78B == -1 ~ "Missing",
    )
  )

```

```

) %>%
select(Response, Count, Percentage)

# Display the table USA
USA_attitudes_frequency %>%
knitr::kable(
  caption = "Attitudes Toward USA Table"
) %>%
kable_styling(
  full_width = FALSE
)

```

Table 3: Attitudes Toward USA Table

Response	Count	Percentage
Very negative	122	10.17
Somewhat negative	140	11.67
Neither positive nor negative	301	25.08
Somewhat positive	392	32.67
Very positive	218	18.17
Don't know	27	2.25

Question #5

- The results show that, on average, responses for the USA are 0.304 points higher than those for China. So, on average respondents perceive the US influence in economics and politics more favorably than China's. The high t-value (8.8) indicates that there is a significant difference between responses to China and the USA and the low p-value (below 0) suggests that the results are statistically significant. We reject null hypothesis that there is not difference in the means.

```

# Clean Q78A and Q78B using across
df_clean <- df1 %>%
  mutate(
    across(
      Q78A:Q78B,
      ~ if_else(.x %in% 1:5, .x, NA)
    )
  )

# Perform the paired t-test
t_test_result <- t.test(df_clean$Q78A, df_clean$Q78B, paired = TRUE)

# Display the t-test result
t_test_result

##
## Paired t-test
##
## data: df_clean$Q78A and df_clean$Q78B
## t = 8.8694, df = 1169, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## 0.2369654 0.3715816

```

```
## sample estimates:  
## mean difference  
##      0.3042735
```