

Problem Set 1

Ikhyun Lee

2025-01-24

1.

I chose Uganda. There are 2,400 respondents in the survey. The interviews were conducted from 01/07/22 to 01/25/22.

2.

Basic descriptions of respondents are provided in a table below.

```
# Gender Summary
gender <- df %>%
  count(Q100, name = "N") %>%
  mutate(Per = round((N / sum(N)) * 100, 1),
         Gender = recode(as.character(Q100), "1" = "Male", "2" = "Female")
  )

# Gender description for table
gender_summary <- data.frame(
  Variable = "Gender",
  Description = paste0(
    paste(
      gender$Gender, ": n=", gender$N, ", ", gender$Per, "%",
      collapse = "; "
    )
  )
)

# Age Summary
age <- df %>%
  summarize(
    `Lowest Age` = min(Q1, na.rm = TRUE),
    `Highest Age` = max(Q1, na.rm = TRUE),
    `Average Age` = round(mean(Q1, na.rm = TRUE), 1)
  )

# Age description for table
age_summary <- data.frame(
  Variable = "Age",
  Description = paste0(
    "Lowest Age: ", age$`Lowest Age`, "; ",
    "Highest Age: ", age$`Highest Age`, "; "
  )
)
```

```

    "Average Age: ", age$`Average Age`
  )
)

# Language Summary
language <- df %>%
  count(Q2, name = "N") %>%
  mutate(Per = round((N / sum(N)) * 100, 1),
         Q2 = as.character(as_factor(Q2)))
  ) %>%
  arrange(desc(N))

# Select top 2 languages and combine the rest
top_languages <- language %>% slice(1:2)
others <- language %>%
  slice(3:n()) %>%
  summarize(
    Q2 = "Others",
    N = sum(N),
    Per = round(sum(Per), 1)
  )
language_combined <- bind_rows(top_languages, others)

# Language description for table
language_summary <- data.frame(
  Variable = "Language",
  Description = paste0(
    paste(language_combined$Q2, ": n=", language_combined$N, ", ", language_combined$Per, "%", collapse = " ")
  )
)

# Final Table
bind_rows(
  gender_summary,
  age_summary,
  language_summary
) %>%
kable(col.names = c("Variable", "Description"), align = "lc",
      caption = "Description of Respondents",
      digits= 1L,
      Format = 'latex',
      booktabs = TRUE) %>%
kable_styling(latex_options = c("hold_position"))

```

Table 1: Description of Respondents

Variable	Description
Gender	Male : n= 1202 , 50.1 %; Female : n= 1198 , 49.9 %
Age	Lowest Age: 18; Highest Age: 85; Average Age: 36.4
Language	Luganda : n= 589 , 24.5 %; Lusoga : n= 292 , 12.2 %; Others : n= 1519 , 63.2 %

3.

Almost half (46.8%) of the respondents are positive about China's influence on Uganda (These respondents answered 4 and 5). This proportion is approximately twice that of the negative public opinion (23.7%) labeled as 1 and 2.

```
#frequency table
df %>%
  count(Q78A, name = 'N') %>%
  mutate(Per = N / sum(N)*100) %>%
  kable(
    caption= 'Attitudes about Influence of China',
    digits= 1L,
    Format = 'latex',
    booktabs = TRUE
  ) %>%
  kable_styling(latex_options = c("hold_position"))
```

Table 2: Attitudes about Influence of China

Q78A	N	Per
1	280	11.7
2	288	12.0
3	96	4.0
4	647	27.0
5	475	19.8
8	2	0.1
9	612	25.5

4.

44.9% of the respondents (who answered 4 and 5) are positive about US's influence on Uganda. This proportion is greater than three times that of the negative public opinion (1 and 2).

```
#frequency table
df %>%
  count(Q78B, name = 'N') %>%
  mutate(Per = N / sum(N)*100) %>%
  kable(
    caption = 'Attitudes about Influence of US',
    digits = 1L,
    Format = 'latex',
    booktabs = TRUE
  ) %>%
  kable_styling(latex_options = c("hold_position"))
```

5.

Result of the paired t -test below indicates that the mean of Q78A is lower than Q78B as t -statistic is -5.258. P-value is 1.661e-07 which is lower than 0.05. Thus, we accept the alternative hypothesis that true mean

Table 3: Attitudes about Influence of US

Q78B	N	Per
1	122	5.1
2	233	9.7
3	147	6.1
4	668	27.8
5	411	17.1
8	2	0.1
9	817	34.0

difference between Q78A and Q78B is not equal to 0. In plain language, we can say that respondents' attitudes about the influence of China is slightly lower than that of the US on average.

```
df =
  df %>%
  mutate(
    across(
      Q78A:Q78B,
      ~if_else(.x %in% 1:5, .x, NA)
    )
  )

t.test(df$Q78A, df$Q78B, paired = TRUE)
```

```
##
## Paired t-test
##
## data: df$Q78A and df$Q78B
## t = -5.258, df = 1537, p-value = 1.661e-07
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -0.2785383 -0.1271834
## sample estimates:
## mean difference
## -0.2028609
```