

Problem Set 1

Jonathan Bachrach

2025-01-28

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

The data set includes the results of a survey conducted in Nigeria between March 5th and April 1st 2022. It includes 1600 responses to a range of questions regarding a wide range of topics relevant to Nigerians, including social and political issues, perception of foreign countries, government approval, etc.

This information was extracted using the code below:

```
afro <- read_csv("from_sav_data.csv")

select(afro,
  respondents = RESPNO,
  date = DATEINTR) %>%
summary(.) %>%
kable(.)
```

respondants	date
Length:1600	Min. :2022-03-05
Class :character	1st Qu.:2022-03-12
Mode :character	Median :2022-03-19
NA	Mean :2022-03-18
NA	3rd Qu.:2022-03-24
NA	Max. :2022-04-01

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

Gender data (variable Q100) was removed from the publicly-available data set, and was unavailable for this analysis.

Ages of respondents ranged from 18 to 97, with an average responsee age of 34.91 years, and a median age of 32. 75% of responsees were under 41.

```
afro <- read_csv("from_sav_data.csv")
view(afro)

select(afro,
  age = Q1) %>%
summary(.) %>%
kable(.)
```

age
Min. :18.00
1st Qu.:25.00
Median :32.00
Mean :34.91
3rd Qu.:41.00
Max. :97.00

Interviews were most frequently conducted in Hausa (524 interviews; 33%). Other frequently used languages include Yoruba (282 interviews), Igbo (186 interviews), and English (120 interviews). A smaller number of interviews were conducted in other languages such as Pidgin English, Efik, Epira, Fulani, and Isoko.

```
afro %>%
  group_by(Q2) %>%
  summarize(.,
            n()) %>%
  kable(.)
```

Q2	n()
1	120
620	524
621	186
622	282
623	39
624	20
625	2
626	30
627	1
628	25
629	1
630	5
631	15
632	19
633	7
634	15
635	23
636	1
637	2
638	2
639	14
640	7
642	12
643	10
644	17
645	13
646	12
9995	196

3) Describe attitudes about economic and political influence of China, Q78A in your data. Your answer should include a relative frequency table and a couple of quick sentences describing the data.

Questions 78A and 78B asked respondents to rank the influence of China and the United States on Nigeria on a scale from 1 (very negative) to 5 (very positive). Respondents could also refuse (8) or answer “I don’t know” (9).

Perceptions of both countries were similar, with an average response of 3.639 for China and 3.564 for the United States.

Interestingly, fewer individuals responded “I don’t know” when asked about China (31%) than about the United States (39%), indicating that China may have a slight edge in terms of public awareness.

4 and 5 were the most popular responses for both China and the USA, comprising 47% and 38% of each country’s total responses.

```
filter(afro, Q78A < 6, Q78B < 6) %>%
  select(.,
    china = Q78A,
    USA = Q78B) %>%
  summary(.) %>%
  kable(.)
```

china	USA
Min. :1.000	Min. :1.000
1st Qu.:3.000	1st Qu.:3.000
Median :4.000	Median :4.000
Mean :3.639	Mean :3.564
3rd Qu.:5.000	3rd Qu.:4.000
Max. :5.000	Max. :5.000

```
chinatab <- count(afro, Q78A) %>%
mutate(., proportion = prop.table(n))

kable(chinatab)
```

Q78A	n	proportion
1	93	0.058125
2	142	0.088750
3	105	0.065625
4	408	0.255000
5	340	0.212500
8	3	0.001875
9	509	0.318125

```
USAtab <- count(afro, Q78B) %>%
  mutate(., proportion = prop.table(n))

kable(USAtab)
```

Q78B	n	proportion
1	83	0.051875
2	125	0.078125
3	149	0.093125
4	385	0.240625
5	229	0.143125
8	5	0.003125
9	624	0.390000

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

See analysis of data above.

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

At the 5% significance threshold, a statistically significant difference exists between the average perception of China and the United States among Nigerians. The difference between the means of responses for each question in the sample data was 0.075; a p-value of 0.028 indicates that the observed difference in mean values is unlikely to be the result of chance.

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

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

##
## Paired t-test
##
## data: afro$Q78A and afro$Q78B
## t = 2.205, df = 948, p-value = 0.02769
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.008230023 0.141401168
## sample estimates:
## mean difference
##      0.0748156
```