

# Quantitative Sociological Analysis

## Descriptive Statistics

### Exercise 3

February 11-13, 2025

# Summarizing descriptive statistics: Netflix

- Let's start to make sense of our Netflix survey data
  - and work toward making a descriptive statistics table
- First, we need to consider each variable's level of measurement
  - so we know how to appropriately summarize the data

## Exercise 3

Let's work in our groups. Review the survey and try to determine how responses to each respective question should be coded into a variable so we can make sense of these data.

Assign a scribe to take notes. Scratch paper will be fine, or maybe download this PPT and use that. You won't turn this part in, but we will share our thoughts as a class afterward.

# age

Date of birth

	Month	Day	Year
Please Select:	<input type="text" value="v"/>	<input type="text" value="v"/>	<input type="text" value="v"/>

- date of birth is a standard survey question because it is more reliable and valid than asking a respondent to report their age
  - age can be computed based on survey date – birth date
    - this also allows for age to be measured in days when precision is necessary
- age will be recorded in years
  - level of measurement: interval-ratio

# subjective age

Most days how old do you feel?

0 10 20 30 40 50 60 70 80 90 100

years old



- response options range from 0 to 100 years old
  - we don't need to convert this variable
- subjective age will be recorded in years
  - level of measurement: \_\_\_\_\_

Fill in the blank here, and other blanks  
as we move forward in this exercise

# sex

What was your biological sex assigned at birth?

☐ Male

☐ Female

- includes two mutually exclusive response options
  - example of a binary indicator variable
    - Why should this type of variable be coded 0 1, rather than 1 2?
- sex will be coded as 0 = male, 1 = female
  - Why will we name this variable female?
    - Hint: mean female = 0.73. Thus, what proportion are male?
  - level of measurement: \_\_\_\_\_

Answer the question, and other questions as we move forward in this exercise

# gender

Consider if masculinity and femininity existed on a polar spectrum. Most days how would you identify?

Strongly Masculine 0 1 2 3 Neither Masculine nor Feminine 4 5 6 7 8 Strongly Feminine 9 10

Most days I feel



- response options range from 0 to 10 on a spectrum
  - this provides several options for how to treat this measure...

# gender

## option (a)

Consider if masculinity and femininity existed on a polar spectrum. Most days how would you identify?

Strongly Masculine 0 1 2 3 Neither Masculine nor Feminine 4 5 6 7 8 Strongly Feminine 9 10

Most days I feel



- we will retain this measurement for our variable
  - Thus, higher scores reflect \_\_\_\_\_, whereas lower scores reflect \_\_\_\_\_, and score in the middle of this spectrum reflects \_\_\_\_\_.
- Level of measurement: response options can be rank ordered, so we could treat as ordinal
  - However, ordinal measures with many categories are often treated like interval-ratio
    - We will treat gender like an interval-ratio variable

# gender

## option (b)

Consider if masculinity and femininity existed on a polar spectrum. Most days how would you identify?

Strongly Masculine 0 1 2 3 Neither Masculine nor Feminine 4 5 6 7 8 Strongly Feminine 9 10

Most days I feel



- if we wanted to treat gender like an ordinal measure it could be helpful to collapse some response options so there were fewer categories
- What might that type of gender variable look like?
  - Note: many possible “correct” answers
    - Recall, a variable should best reflect its conceptualization rooted in perspective and theory



# region

Since living in the United States, in what region have you lived the longest?

☐ North

☐ East

☐ South

☐ West

- How should region be coded?
  - level of measurement: \_\_\_\_\_
    - Hint: the numbers don't really matter, so long as they are mutually exclusive
      - but ideal for coding to be somewhat intuitive

# rurality

Select the best option that describes your current place of residence

☐ urban

☐ suburban

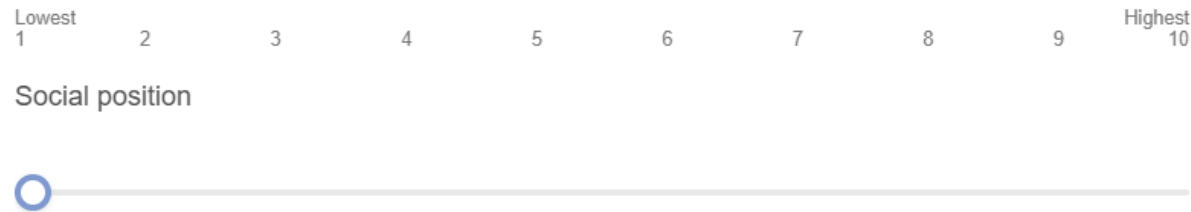
☐ rural

- will be coded as 1=urban, 2=suburban, 3=rural
  - level of measurement: \_\_\_\_\_
- Why would we want to name this variable urbanicity if it were coded as
  - 1=rural, 2=suburban, 3=urban?
    - Hint: like with the region variable, the numbers don't really matter but ideal to be intuitive
      - Why might naming this variable suburban as 1=rural, 2=urban, 3=suburban not be very intuitive?

# subjective socioeconomic status

Where would you place yourself on this sliding scale, relative to other people in the United States?

The scale has 10 points, with 1 being the lowest and 10 being the highest.  
Which point best represents your current social position?



- response options range from 1 (lowest) to 10 (highest), with equal intervals
  - because of so many response categories we will treat this variable as if it were...
    - level of measurement: \_\_\_\_\_
- What if we collapsed categories to create a ses variable like this...
  - 1-4 (1)=lower class, 5-7 (2)=middle class, 8-10 (3)=upper class
    - level of measurement: \_\_\_\_\_

# political party affiliation

Political party affiliation

☐ Strongly Republican

☐ Republican

☐ Independent/Other

☐ Democrat

☐ Strongly Democrat

- How would you code this variable, and what would you name it?
  - level of measurement: \_\_\_\_\_
    - Note: many possible “correct” answers

# composite variable

- a new variable created by combining other variables
  - often to reflect a broader construct
    - commonly referred to as a scale or an index
- sometimes based on well-established guidelines
  - like with many psychometric scales (e.g., anxiety, depression, attention)

# anxiety

How often have you experienced the following during the past two weeks?

	Not at all	Several days	More than half the days	Nearly every day
Feeling nervous, anxious, or on edge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not being able to stop or control worrying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Worrying too much about different things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouble relaxing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being so restless that it is hard to sit still	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Becoming easily annoyed or irritable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feeling afraid as if something awful might happen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- based on the [GAD-7 guidelines](#)...
- how should this variable be coded?
  - level of measurement: \_\_\_\_\_
- If we were to use the clinical threshold, which is  $\geq 10$ 
  - how would this binary indicator variable be coded?
    - level of measurement: \_\_\_\_\_

# depression

How often have you experienced the following during the past two weeks?

	Not at all	Several days	More than half the days	Nearly every day
Little interest or pleasure in doing things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feeling down, depressed, or hopeless	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouble falling or staying asleep, or sleeping too much	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feeling tired or having little energy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Poor appetite or overeating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feeling bad about yourself — or that you are a failure or have let yourself or your family down	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouble concentrating on things, such as reading the newspaper or watching television	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Thoughts that you would be better off dead, or of hurting yourself in some way	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- based on the [PHQ-9 guidelines](#)...
- how should this variable be coded?
  - too complex, just use box at bottom
  - level of measurement: \_\_\_\_\_
- For simplicity let's use a clinical threshold of  $\geq 10$  for a binary indicator

# attention

How often do you experience the following?

	Never	Rarely	Sometimes	Often	Very Often
make careless mistakes when you have to work on a boring or difficult project	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
have trouble keeping your attention focused when you are doing boring or repetitive work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
have difficulty concentrating on what people say to you, even when they are speaking to you directly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
have difficulty organizing tasks and activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
avoid, dislike, or are reluctant to engage in tasks that require sustained mental effort	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
forget to complete tasks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- based on the [ASRS guidelines...](#)
  - Part A only
- how should this variable be coded?
  - level of measurement: \_\_\_\_\_



# macro sociohistorical experience

Rank order the following events from most to least impactful on your current world view

COVID-19 pandemic
January 6 insurrection
Black Lives Matter movement
Roe v Wade reversal

- due to my poor operationalization, it's not worth wrapping our heads around how to construct this variable
  - unless you have any good ideas

# preference for comedy

Rank order the following genres from your most to least favorite to watch on a typical day

drama
action
horror
comedy
documentary

- we will construct this variable based on the rank ordered placement of comedy
  - 1=comedy is most favorite, 2=comedy is second favorite, 3=comedy is third favorite, 4=comedy is fourth favorite, 5=comedy is least favorite
    - level of measurement: \_\_\_\_\_

Note how this approach regrettably disregards potentially useful information – the rank order of the other four specific genres is likely useful for predicting Netflix consumer behavior. However, how complex might it be to include this information? Could this information be reflected in just one variable?

# genre rankings

Not concerned with this for now but may see something like this later.

- Permutations: all possible unique arrangements when order matters

$$P(n, n) = \frac{n!}{(n - n)!} = n!$$

- There are 5 genres
  - $5! = 120$

What did we all come up with?