

## Session 15: Missing Data — Pen-and-Paper Pair Exercise

PSY 410 | Data Science for Psychology

Name: \_\_\_\_\_ Date: \_\_\_\_\_

*No laptop today? No problem. This handout lets you practice the same skills on paper. Work with a partner who has a laptop and compare your work at the end.*

---

### The data: therapy\_survey

Survey data from a therapy study:

id	age	baseline_depression	followup_depression	satisfaction
1	25	22	12	4
2	30	25	23	3
3	NA	18	NA	NA
4	22	20	15	5
5	28	24	NA	4
6	NA	19	NA	NA
7	35	NA	NA	NA
8	26	21	16	5

---

### The task (same as the slide exercise)

1. How many participants are missing baseline data? Followup data?
2. How many participants have **complete data** (no NAs anywhere)?
3. Create a version that drops rows missing followup data
4. What percentage of participants completed the followup?

### Your pen-and-paper version

**Step 1: Count NAs per variable.** Go through each column and count the missing values:

Variable	Count of NAs
id	
age	
baseline_depression	
followup_depression	
satisfaction	

**Step 2: Find complete cases.** Mark each row as complete or not:

id	Any NAs in this row?	Complete?
1		
2		
3		
4		
5		
6		
7		
8		

**How many complete cases?** \_\_\_\_\_

**Step 3: Drop rows missing followup.** Cross out any row where `followup_depression` is NA. Which rows survive?

Surviving rows: \_\_\_\_\_

**How many rows remain?** \_\_\_\_\_

**Step 4: Calculate completion rate.**

- Total participants: \_\_\_\_\_
- Participants with followup data: \_\_\_\_\_
- Completion rate: \_\_\_\_\_ / \_\_\_\_\_ = \_\_\_\_\_ = \_\_\_\_\_%

**Step 5: Think about it.** Look at the rows you crossed out. Do participants missing followup data also tend to be missing other variables? What might this pattern suggest?

Your answer: \_\_\_\_\_

**Step 6: Write the code.** Fill in the blanks:

```
# 1. Count missing by variable
therapy_survey |>
  summarize(across(everything(), ~sum(_____(.x))))

# 2. Complete cases
therapy_survey |>
  drop_na() |>
  _____()

# 3. Drop if missing followup
therapy_survey |>
  drop_na(_____)

# 4. Completion rate
therapy_survey |>
  summarize(
    completion_rate = mean(!is.na(_____))
  )
```

---

### Check your work

Compare your counts, completion rate, and code with your partner's screen. Do your answers match?