Session 14: Joins — 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: Three tables from a therapy study

`baseline`

id	age	baseline_depression
1	25	22
2	30	25
3	22	18
4	35	30
5	28	20

`treatment`

id	condition
1	CBT
2	Control
3	CBT
4	Control

(Note: participant 5 is missing from this table)

`followup`

id	followup_depression
1	12
2	23
3	10
5	18

(Note: participant 4 is missing, but participant 5 is here)

The task (same as the slide exercise)

1. Create a dataset with baseline info and treatment condition (keep all participants)
2. Add followup data (keep all from step 1)
3. How many participants are missing followup data?

Your pen-and-paper version

Step 1: Join `baseline` and `treatment`. Draw lines connecting matching `id` values between the two tables. Then fill in the joined result using `left_join()` (keep all rows from `baseline`):

id	age	baseline_depression	condition
1			
2			
3			
4			
5			

What value does `condition` get for participant 5? _____

Why? _____

Step 2: Now join the result with `followup`. Again using `left_join()`, fill in:

id	age	baseline_depression	condition	followup_depression
1				
2				
3				
4				
5				

Step 3: Count the missing followup data.

How many participants have `NA` for `followup_depression`? _____

Which participant(s)? _____

Step 4: Think about why. What might explain why participant 4 is missing followup data but participant 5 is not?

Your answer: _____

Step 5: Write the code. Fill in the blanks:

```
# Step 1
baseline_treatment <- baseline |>
  ________join(treatment, by = "____")

# Step 2
complete_data <- baseline_treatment |>
  ________join(followup, by = "____")

# Step 3
sum(is.na(complete_data$____________________))
```

Check your work

Compare your joined tables and code with your partner's screen. Do your answers match?