

Session 17: Correlation & Simple Regression — 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

Here is a small dataset of 8 students with hours of social media use per day and their GAD-7 anxiety score:

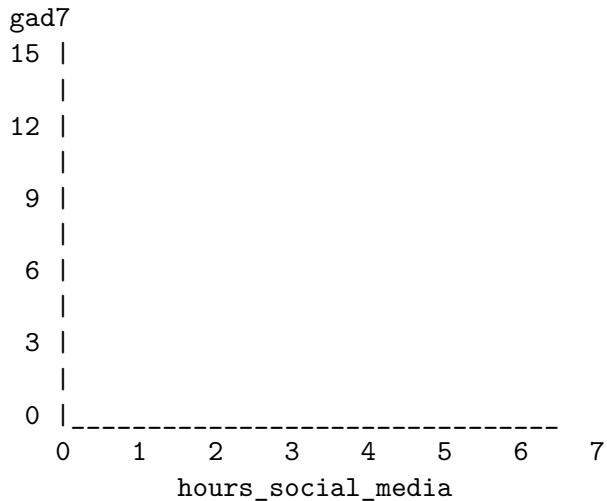
student	hours_social_media	gad7_score
1	2.1	5
2	5.3	12
3	1.0	3
4	3.8	9
5	6.2	14
6	4.5	8
7	0.5	4
8	3.0	7

The task

1. Plot the relationship between social media use and anxiety
2. Compute the correlation
3. Run a simple linear regression with `lm()`
4. Interpret the output

Your pen-and-paper version

Step 1: Sketch a scatterplot. Plot the 8 data points. Label axes clearly.



Does it look like there's a relationship? Positive or negative? Strong or weak?

Your answer: _____

Step 2: Predict the correlation. Based on your scatterplot, estimate what r will be. Circle your guess:

- Close to 0 (no relationship)
- Around 0.3 (weak positive)
- Around 0.6 (moderate positive)
- Around 0.9 (strong positive)
- Negative (inverse relationship)

Step 3: Interpret regression output. Your partner will run this code:

```
model <- lm(gad7_score ~ hours_social_media, data = student_data)
summary(model)
```

When they get the output, record these values:

Component	Value	What it means
Intercept (b0)		Predicted anxiety when social media hours = _____

Component	Value	What it means
Slope (b1)		For each additional hour, anxiety changes by _____
R-squared		_____ % of variation in anxiety is explained by social media use
p-value (slope)		Is the relationship statistically significant? (p < .05?)

Step 4: Make a prediction. Using the intercept and slope your partner found, predict the GAD-7 score for a student who uses social media 4 hours per day:

Predicted score = intercept + slope x 4 = _____ + _____ x 4 = _____

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

```
# Correlation
cor(student_data$_____, student_data$_____)

# Regression
model <- lm(_____ ~ _____, data = student_data)
summary(model)

# Visualize
ggplot(student_data, aes(x = _____, y = _____)) +
  geom_point() +
  geom_smooth(method = "____") +
  labs(
    title = "_____",
    x = "_____",
    y = "_____"
  ) +
  theme_minimal()
```

Step 6: Correlation vs. causation. Does this analysis prove that social media *causes* anxiety? Why or why not?

Your answer: _____

Check your work

Compare your scatterplot, correlation estimate, and code with your partner's screen. Do your answers match?