PSYC 5301 – Course introduction

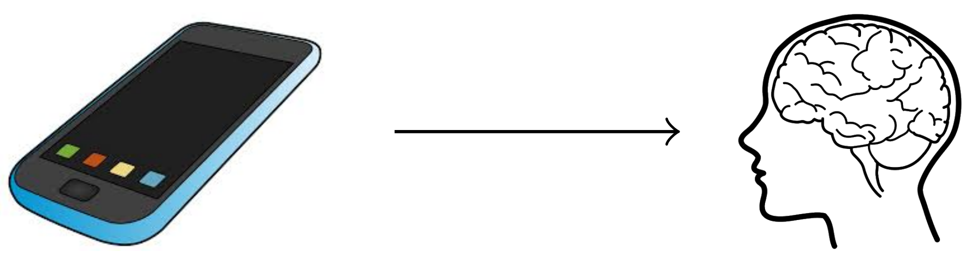
Thomas J. Faulkenberry, Ph.D.

January 19, 2021

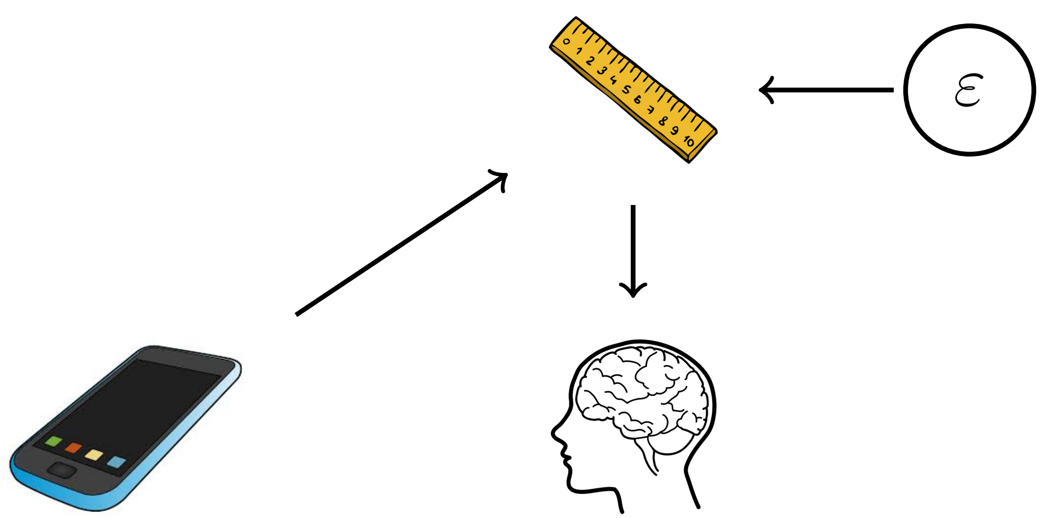
A simple example

I want to know if increased social media use leads to increased depression. How do I find out?

What I *think* I'm dealing with:

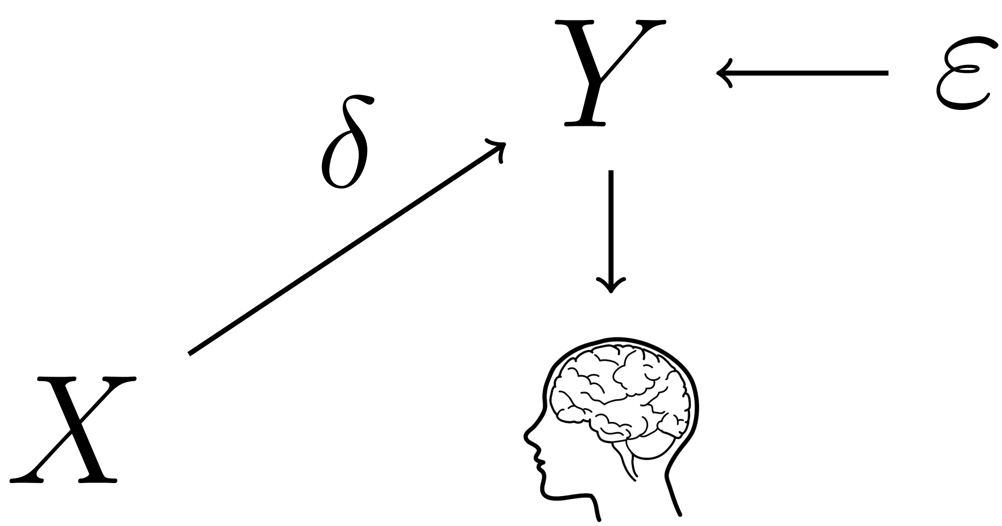


What I'm *really* dealing with:



Our question becomes a *mathematical* one – does social media use lead to an increase in *measurements* of depression?

Our **model** turns into this:



where

- X = social media use (0 = infrequent, 1 = frequent)
- Y = score on depression scale
- δ = "effect" of social media use
- ε = measurement error

The model is then specified as a *linear* equation:

$$Y = \delta \cdot X + \varepsilon$$

We then translate our research questions into mathematical ones:

1. does social media increase depression?
   - if "no", then $\delta = 0$
   - if "yes", then $\delta > 0$
   - which is it?

To answer this, we use *hypothesis testing* (also called /model comparison)

- Define two models:
  - $\mathcal{H}_0 : \delta = 0$
  - $\mathcal{H}_1 : \delta > 0$
- See which model best fits the observed data
  - frequentist method: compute p-value, i.e., the probability of observing the data if the $\mathcal{H}_0$ is true). If p is small, then data are *rare* under $\mathcal{H}_0$, so we *reject* $\mathcal{H}_0$ in favor of $\mathcal{H}_1$.
  - Bayesian method: compute Bayes factor, i.e., the relative likelihood of observing data under the two models. This will tell us which model gets support from the data (and how much evidence)

If $\mathcal{H}_1 : \delta > 0$ is best model (i.e., the effect of social media use on depression is *positive*), then we might want to know *how large the effect is*.

To answer the "how large" question, we use *parameter estimation*.

- frequentist methods: construct *95% confidence interval*
  - if we repeated the experiment an infinite number of times, then 95% of the constructed intervals would contain the true *population* parameter δ
  - note: this says nothing about the *probability* that the true δ is in our constructed interval – it just says the process works "in the long run"
- Bayesian methods: construct *95% credible interval*
  - after observing data, we construct a (posterior) distribution of values for δ. There is a 95% probability that this interval contains the true δ

Note: it makes no sense to perform parameter estimation if $\mathcal{H}_0$ is the better model? Why – because the model explicitly sets $\delta = 0$!

So, that's the basic game this semester. We will learn all about the classic experimental designs (i.e., ways to handle specific types of research questions). Inherent in each design is:

1. how to write the model equations
2. how to perform model comparison (frequentist and Bayesian)
3. how to perform parameter estimation

What are the classic designs?

- single factor designs (i.e., ANOVA)
- repeated measures designs
- multiple factor designs
- covariate designs
- regression models