2:05-2:30pm: Intro to Statistical Modeling *(White Board Exercise)*

Goal: *Have students understand a statistical model conceptually, including its component parts. Have them ask a precise statistical question related to their research question and define its parts.*

People: Would be helpful to have one facilitator and one scribe.

2:05-2:10: ASK: *What is an example of a statistical model?*
- List on white board. I'll bet you get a lot! Keep going until the list includes a linear regression
- Say we're going to focus on this simplest example to start: linear regression.

2:10-2:15: ASK: *What does a linear regression attempt to demonstrate?*
- Show a *relationship* between two variables
- Discuss *correlation* vs. *causation*
- Statistical models describe **patterns** and highlight **correlations**
- The **model** is the equation: $y = mx + b$

2:15-2:25: ASK: *What are the components of a statistical model?*
- Response variable = y
- Predictor variable(s) = x
- Distribution = depends on type of data
  - When data is not normally distributed, this is called a "generalized linear model"
  - Draw some distributions and describe the associated data:
    - Gaussian (normal)
    - Binomial (0-1)
    - Poisson (count)
    - Negative Binomial (count)
- The "link" function allows you to view non-linear x-y relationships in the distributions above in a linear way.
  - For instance, the "log" link is used in a Poisson model to re-project x-y onto a simple, linear plane. (Draw this re-projection)
  - Link functions by distribution:
    - Gaussian = identity (it's already linear!)
    - Binomial = logit
    - Poisson = log
    - Negative Binomial = log

- **2:25-2:30: ASK:** *What statistical question can we ask about our research topic? What will the components of the associated statistical model look like?*
  - Make sure to define response + predictor variables + distribution + link