Deck 5

* Frequentism (classical statistics)
  + What is it?
    - An inferential statistic paradigm
    - Easy to communicate because widely known and used
  + Key features
    - No assumptions using prior knowledge
    - Based on the idea of hypothetical repeated (infinite) sampling of an unknowable population
    - Less subjective than Bayesian statistics
  + Benefits
    - Widely known
    - Many software tools
    - Assumptions are often reasonable
  + Drawbacks
    - Repeated sampling assumption: non-intuitive interpretation of confidence and significance
    - Focus on the process rather than the results of a single experiment
    - Focus on hypothesis testing
  + Common methods
    - Linear regression, ANOVA, T-tests, generalized linear models
  + Key assumptions
    - Pop is large and unknowable
    - Set of true parameter values
    - Can never know the true parameter values
    - Inference is based on hypothetical repeated sampling
  + Confidence Intervals
    - Does not mean I am 95% certain that the mean length is between two values
    - It does mean that the length is captured 95% of the time it is recorded
* Sources of Error
  + Types of Error
    - Measurement
      * Imperfect instrumentation
      * Mismeasurements
      * Outside of our detection range
    - Process
      * The inherent randomness of a system
      * Demographic examples
        + Birth and death rate have outliers of people who have more or less kids and die younger than average
      * Environmental stochasticity
        + Weather or climate fluctuations
        + Extreme events
    - Model
      * Mis-specifying a model
      * Omission of important predictors
      * Including unimportant predictors
      * Wrong type of model or sampling method
    - Sampling
      * Errors that arise from non-representative samples
* Hypothesis Testing
  + Classical approach
    - Uses parametric distributions and their assumptions to do inference
  + Hypothesis testing
    - A way to quantify the strength of evidence against a null hypothesis
    - Null hypothesis captures what we expect to see if associations are random
    - Straw man hypothesis mean there is a slope of 0, there are no relationships in the data
    - A decision criterion (p-value) is used as a limit to fail or reject the null hypothesis in favor of alternative
  + Alternative hypotheses
    - 2-tailed
      * Non-directional
      * Increased light might be associated with greater or lower biomass
      * Covers both directions of the data, kinda ambiguous
    - 1-tailed hypothesis
      * Directional
      * Based on previous experience, can make an informed decision on what we think will happen
      * 1- tailed can give us better statistical power by accepting values on one side of the dataset