III. Geocentric Models

* Linear Regression
  + Geocentric: Describes associations, makes predictions, but mechanistically wrong
  + Gaussian: Abstracts from generative error model, replaces with normal distribution, mechanistically silent
  + Useful when handled with care
  + Many special cases: ANOVA, ANCOVA, t-test, others
* Why Normal Distribution?
  + Two arguments:
    - Generative: Summed fluctuations tend towards normal distribution
    - Inferential: For estimating mean and variance, normal distribution is the least informative distribution (Contains no other information than mean and variance)
  + Variable does not have to be normally distributed for normal model to be useful. It’s a machine for estimating mean/variance.
* Linear Regression Drawing the Owl
  + 1. Question / Goal / Estimand
  + 2. Scientific Model
* Generative Models
  + **Dynamic Models:** Incremental growth of the organism; both mass and height (length) derive from growth pattern; Gaussian variation result of summed fluctuations
  + **Static Models:** Changes in height result in changes in weight, but no mechanism; Gaussian variation result of growth history
    - **W = bH + U**
      * Ui ~ Normal(0, theta)
      * Hi ~ Uniform(130,170)
    - Describing Models
      * List the variables
      * Define each variable as a deterministic or distributional function of the other variables
  + 3. Statistical Model
    - Estimator: We want to estimate how the average weight change with height.
      * E(Wi|Hi) = a + bHi

A diagram of a function

AI-generated content may be incorrect.

* + - Prior Predictive Distribution
      * Priors should express scientific knowledge, but softly
        + When H=0, W=0
        + Weight increases (on avg) with height
        + Weight (kg) is less than height (cm)
        + Sigma must be positive
      * Understand the implications of priors through simulation
        + What do the observable variables look like with these priors?
    - Sermon on priors
      * There are no correct priors, only scientifically justifiable priors
      * Justify with information outside the data – like rest of model
      * Priors not so important in simple models
      * Very important/useful in complex models
      * Need to practice now: simulate, understand
  + 4. Validate Model
    - Simulation Based Validation
      * Bare minimum: Test statistical model with simulated observations from scientific model
      * Golem might be broken. Even working golems might not deliver what you hoped.
      * Strong Test: Simulation-Based Calibration
  + 5. Analyze Data
    - First Law of Statistical Interpretation
      * The parameters are not independent of one another and cannot always be independently interpreted.
      * Instead push out posterior predictions and describe/interpret those.