Biostatistics

Module 2 2/20/17

Why do statistics?

* Data not by chance
* Objectivity, but actually kind of subjective; but once you hold yourself to a cutoff point, there is a level of objectivity; traditional frequentist statistics diverge here from maximum likelihood and Bayesian stats. Subjectivity, argues Jones, is removed when using Bayesian models
* Discriminate signal from noise
* Identify relationships or effects
* Attribute causation in experimental study
* Relating multiple variables

Reasons to use maximum likelihood, or maximum likelihood is good because:

* Flexible (frequentist holds itself to p-value .05): mechanistic and process-based questions using available data, whole-scale ecosystems, data are messy! Test multiple hypotheses at once
* Straight-forward
* Philosophy (more applicable to real world)

Review Intro Biostats

1. The normal distribution, bounded between -infiniti to +infiniti
   1. Why?
   2. Central limit theorem is basis of normal dist. Law of large numbers, averages
   3. Mathematically tractable
   4. Mean and variance are independent of each other
   5. Curve is called probability density function: F(x) = (1/sigma\*2pi)\*e^(x-u)^2/2sigma^2, relates x and the parameters and allows us to draw smooth curve and probability of drawing a particular value (height of trees, people, etc)
2. Linear models: NOT talking about y =mx +b; rather, additive models (ANOVA etc)
   1. Height = avg height of oak + Bi \*height of either species (encode oak and maple as {0,1}). Bi is the difference between oak and maple average heights
   2. Simple additive model, with parameters to explain differences in height between 2 populations. If bi is 0, then the heights are not statistically significantly different and there is no effect of that parameter

Maximum likelihood

Bayes Rule: probability of model given data

Log likelihood because you sum probability densities instead of multiplying (multiplying probabilities gives small numbers)

Continuous: decimals, gamma, normal, lognormal

Discrete: counts