

Lab for Jan 27, 2009, **Introduction to Statistical Thinking**

A. Concepts:

- A way of making inferences in the face of uncertainty (epistemology).
- The statistical universe (ontology)
- Inductive vs. deductive reasoning
- Descriptive statistics vs. hypothesis testing
- What is a hypothesis?
- What is a null hypotheses?
- What is "error"? Type I & Type II error
- Precision vs. accuracy
- Sampling (and the concept of "randomness" and "bias")
- What is an experiment? How does it differ from an observation?
- Experimental Design - observations planned in advance
- Experimental and sample units
- Replication vs. pseudoreplication
- Control
- Causality

B. Data

- Population
- Sample
- Parameters
- Scales of measurement:
 - Nominal (only a name)
 - Ordinal (ranking of categories)
 - Interval or Ratio (continuous values)
- Central tendency:
 - mean
 - median

- Dispersion (why is this important?):

- range
- standard deviation and variance
- standard error
- quartiles

C. Likelihood, Bayes, and frequentist approaches

Likelihood = P(Data | Model & Parameters)

$$P(A \& B) = P(A|B)P(B) \quad P(A \& B) = P(B|A)P(A)$$

Maximum likelihood – pick the best model

$$\frac{P(M_1 | Data)}{P(M_2 | Data)} = \frac{P(Data | M_1)P(M_1)}{P(Data | M_2)P(M_2)}$$

Bayesian – calculate the posterior probability of each model

$$P(M_0 | Data) = \frac{P(Data | M_0)P(M_0)}{P(Data)} = \frac{P(Data | M_0)P(M_0)}{\sum_j P(Data | M_j)P(M_j)}$$

Estimate with MCMC

Nuisance parameters

ML: Joint estimation-Maximize likelihood for all parameters at once

$$\max[P(Data | \alpha, \beta)]$$

Bayes: Marginal estimation–Sum over all possible values of nuisance parameters

$$P(\alpha | Data) = \frac{p(\alpha) \int P(Data | \alpha, \beta) P(\beta) d\beta}{\iint P(Data | \alpha, \beta) P(\alpha) P(\beta) d\alpha d\beta}$$

Frequentist – Does the data support the given model over another?

Nested hypotheses

$$LRT = \ln \left(\frac{\max[P(Data | \alpha, \beta)]}{\max[P(Data | \alpha = \beta)]} \right)$$

Compare 2*LRT to C² to calculate p-value

Can also compare non-nested hypotheses (i.e. AIC)

Bayesian – How much does the data support this model relative to another?

$$BayesFactor = \frac{P(Data | M_1)}{P(Data | M_2)}$$

MCMC calculates uncertainty as a product of the search