# Philosophical approaches to science & statistical inference

ZOL851
September 2nd, 2014

# What is a p-value?

- On the board….

# Goals for today

- Provide overview of scientific approaches.
- Overview of statistical approaches
- Recognize that the model we fit, and the way we fit it are two different issues.

# Readings for today

- Quinn & Dunham
- Chamberlin
- Platt
- Stephens

# R screencasts now available

- https://github.com/idworkin/ZOL851/blob/master/LinksToScreencasts.md

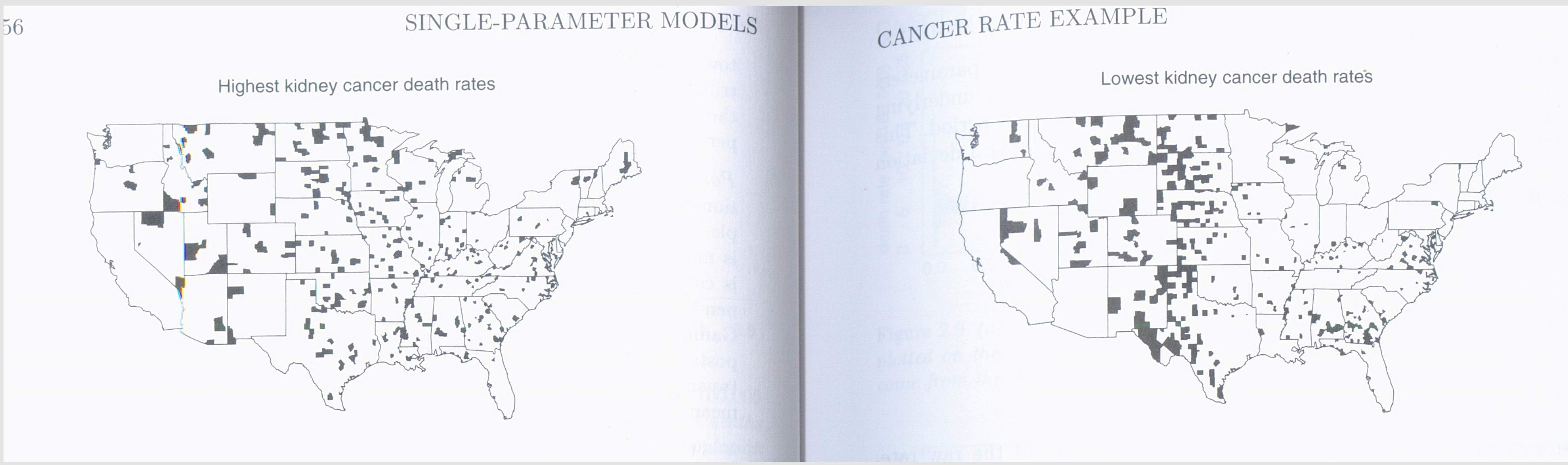
# Readings for Thursday (measurement theory)

- There are two readings on Measurement theory for Thursday. A short online article, and a long review article from last year.
- We will have a class discussion on Thursday about measurement theory. Come armed with a list of how it applies to your study system (you can use the check list in Houle).

# Readings for Thursday

- You will also want to read chapters 1 and 2 from Bolker by the weekend for the R session.
- Optionally you can also read Dalgaard chapters 1 & 2 ( more classic “stats” intro to R).

# Cancer incidence maps




Cancer incidence rates are based on sampling; in this case, they're sampling per 100,000 in each county. Thus, in their random sampling, the difference of a few individuals can account major differences in incidence rates.

Where do statistical methods fit into a scientific hypothesis driven research program?

# Motivating examples

- Pitchers et al (2014).
- Trying to infer the relative roles of selection and genetic variance (as well as trait types) on evolutionary rates.

# The Big picture

- Statistical inference should really be about providing a quantitative & mathematical formalism to both the ideas and the approach you take to science.
- Without an understanding of the approach we take to science, are the hypotheses we generate and test statistically useful?

Chalkboard

# Approaches to Science and Statistics

- Continua in Science
  - Description vs Inference
  - Hypothetico-deductive vs competition among multiple hypotheses
  - Experimentation vs Observation
- Approaches to statistics
  - Frequentist / Likelihood / Bayesian/ Nonparametric (randomization)

# Do we need hypotheses and predictions for science?

- “good science”?
- “productive science”?
- “fundable science”?

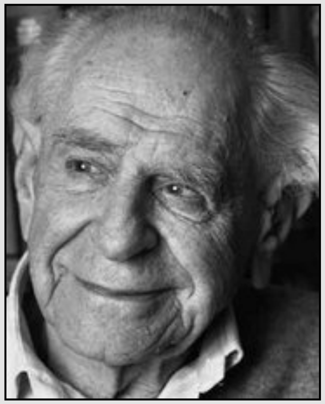
# Three Continua in Science

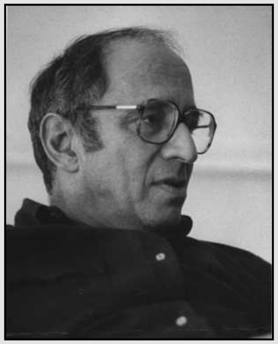
Description — Inference

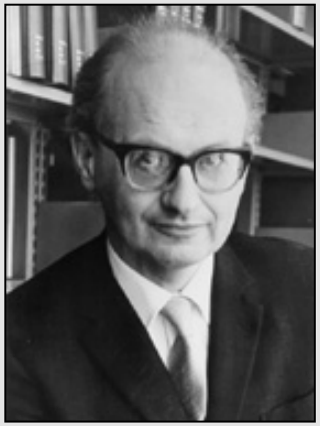
Hypothetico-deductive — Lakatosian Multiple hypotheses

Observation — Experiment

# Philosophers of Science

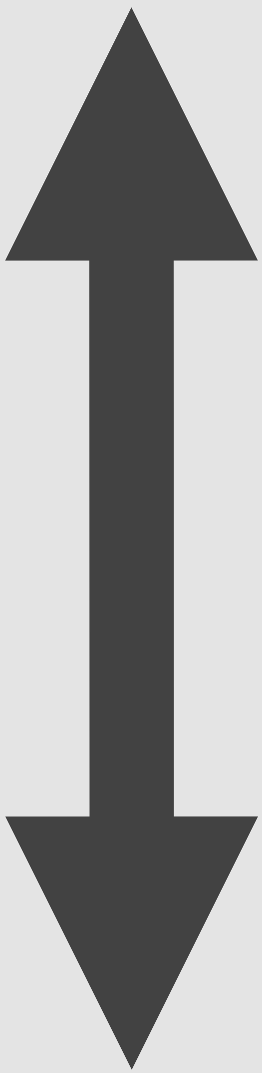






- Karl Popper
- Thomas Kuhn
- Imre Lakotos

# Continuum #1



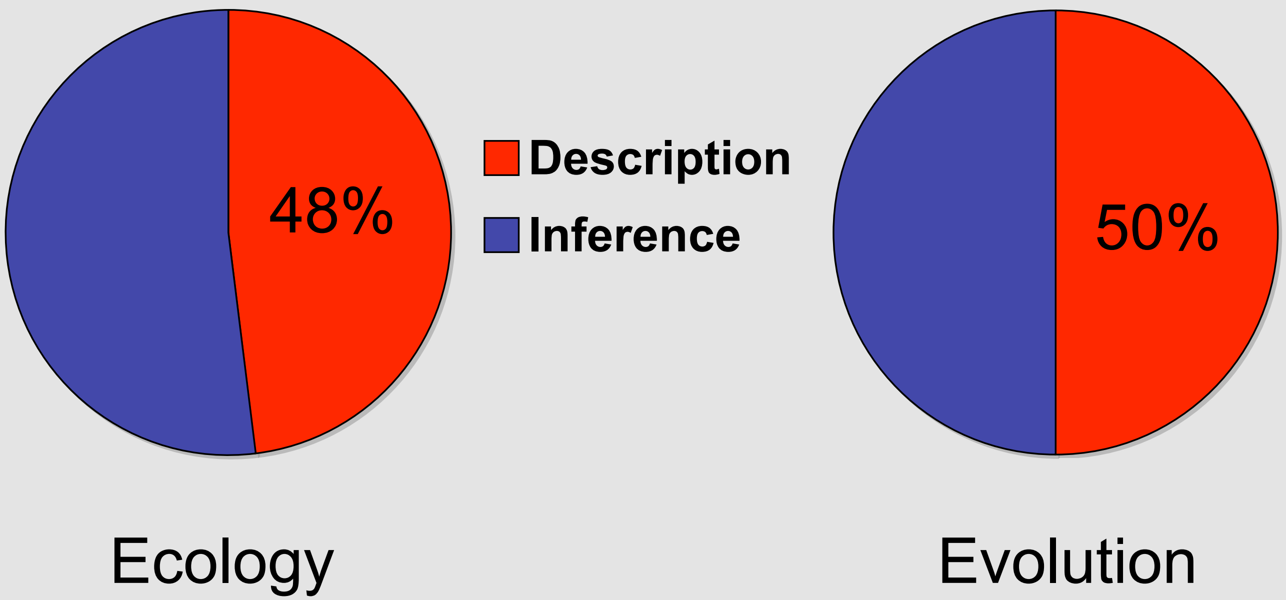
**Descriptive**
Hypothesis generating

side note: early phylogenetic trees were generated from discrete characters: these characters define this clade, and this clade is defined by these characters. Logic was circular, and the same data was later used to generate different trees.

Robert Sokal (early statistician) suggested (against the absolute truth/frequentist approach to stats) that we incorporate uncertainty to phylogenetic trees. Phylogenetics are NOT INFERENTIAL, but are rather hypothesis generating.

**Inferential**
Hypothesis testing

# Frequency of Observation and Hypothesis Testing




Data from the July 2005 issues

# Hypotheses and Predictions

- Hypotheses are suppositions about the way a system works. A proposed explanation. Includes some indication of process/mechanism.
  - e.g. Productivity in freshwater lakes is limited by the availability of Phosphorous

- Predictions state the specific expected outcome given that the hypothesis is true. Refers only to expected pattern.
  - e.g. Lakes with more P will have higher productivity.

- "If the intervening desert lowlands are impermeable barriers to dispersal, then there should be no relationship between genetic distance and geographic distance among mountaintop populations…"

- Could rephrase as:
  - Hypothesize that desert lowlands are barriers to dispersal.
  - Predict no relationship between genetic and geographic distance.
- Hypothesis is often implicit

# Some Things to Think About

- Where do hypotheses come from?
  - Need some observation (data) in order to generate hypotheses
- What makes a good hypothesis? (work in groups)

# Continuum #2

**Hypothetico-deductive**

Deduction of alternative hypotheses

Falsification (Platt/Popper?)

Major problem: one can reject ANY null hypothesis with large enough sample sizes/power
this is not biologically useful!

Many hypotheses can be true to varying degrees

**Lakatosian**

Competition among multiple hypotheses

Values based on successful predictions

# Strong Inference
# Platt (1964)

- Steps:
  1. Devise alternative hypotheses
  2. Devise a crucial experiment with alternative outcomes, each of which will exclude one or more of the hypotheses
  3. Carry out experiment so as to get a clean result
  4. Recycle the procedure making sub-hypotheses to refine the possibilities that remain
- This is the "...method of most rapid progress…"
- From Chamberlin's (1890) multiple working hypotheses.
  – No single favorite

# Quinn & Dunham, 1983

- The rejection of hypotheses is not as simple as it may first seem
- Often faced with varying degrees of support for various hypotheses
- Mutual causes do not lend themselves to univariate hypothesis tests
- Often difficult to formulate an appropriate null hypothesis to test against (false dichotomies)

- Generally Bayesian approaches are philosophically aligned with this as well, as we make statements about the posterior probabilities of the best approximating models.

# Work in groups to generate a list of +/- for each.

- Is there a clear cut answer for your system of study?

# Continuum #3

**Experimentation**

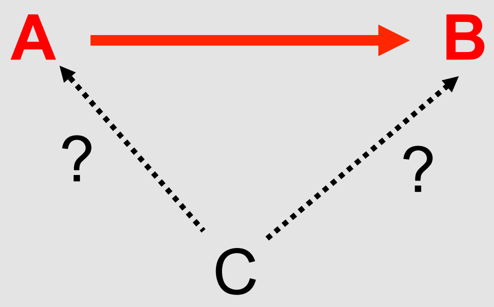
Highly Controlled as possible (i.e. manipulating just on or a few variables at a time).

**Observation**

Realism

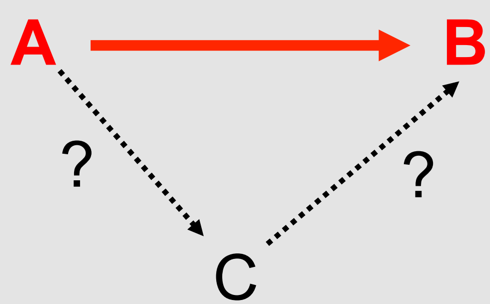
# Mechanism: more than just a match between hypothesis and data

- Correlation is NOT causation
- Associations can be completely unrelated
  - e.g. shoe size correlated with test scores
- Due to some third causal unmeasured variable



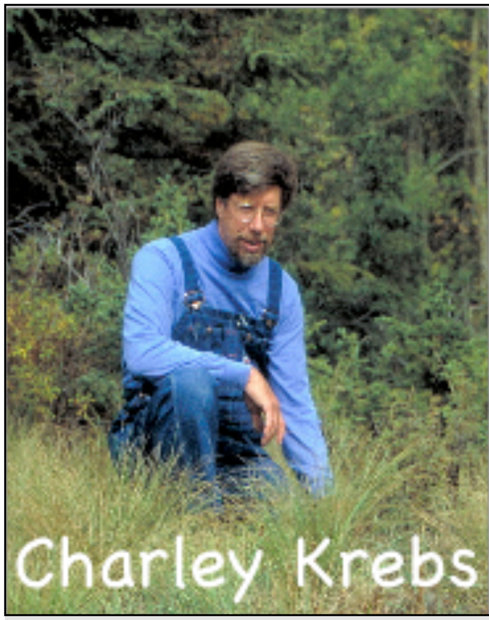
# Experiments as Mechanism

- Controlled, replicated, manipulative experiments
  - Hold everything constant
  - Change one thing
  - If areas without manipulation (control) differ from area with manipulations, then manipulations must be the cause
- But several challenges to this
- Must truly hold everything else constant
- Causality still slippery



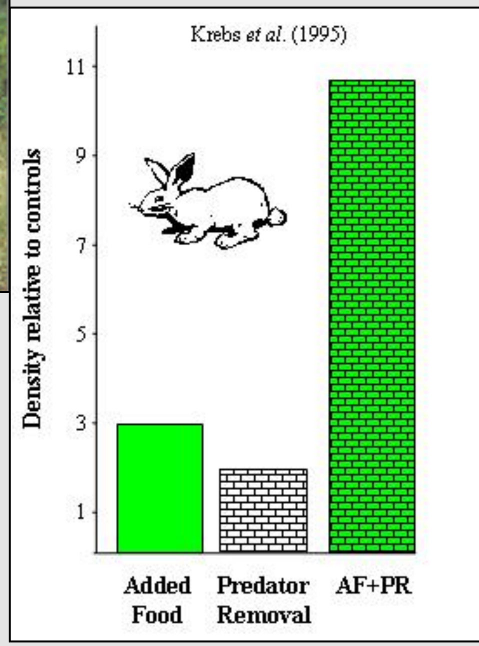
# Should we care about mechanism?

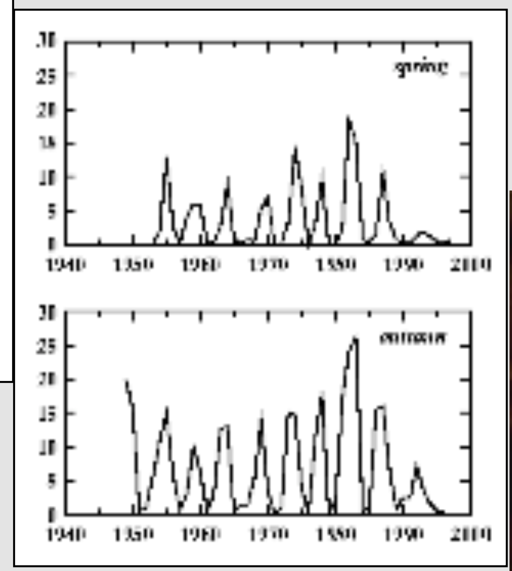
- Many would say yes, including NSF!
- Rob Peters: It’s hopeless to try and get at mechanisms in complex systems and we should just worry about prediction
- Some things are phenomenological and have no mechanism (e.g. gravity)




# Mechanism

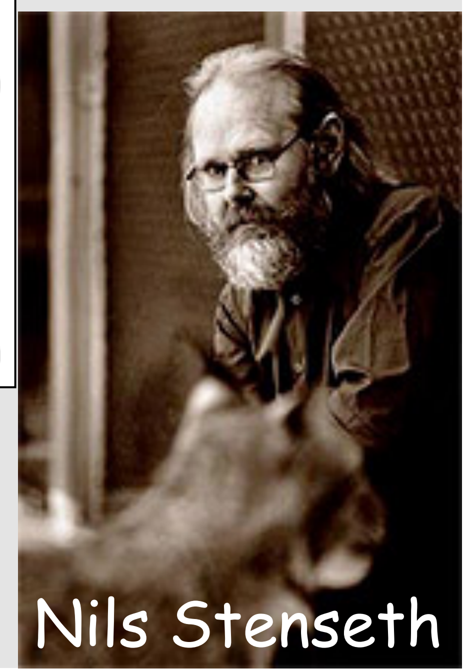
*Can never understand the cause of a biological phenomenon by studying pattern.*



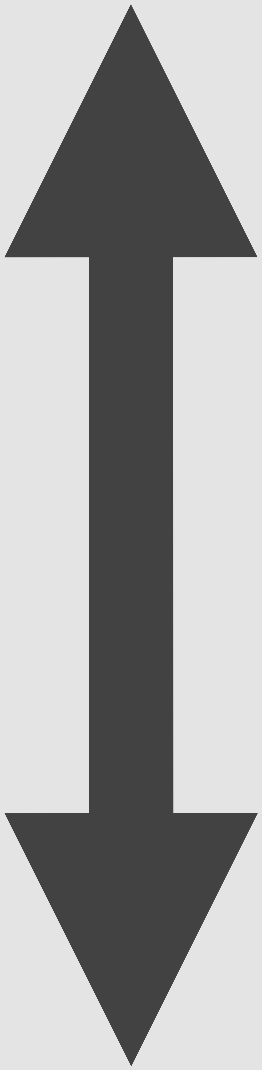




# Pattern

*The relevance of a mechanism can only be understood by observing natural patterns.*



# Continuum #3

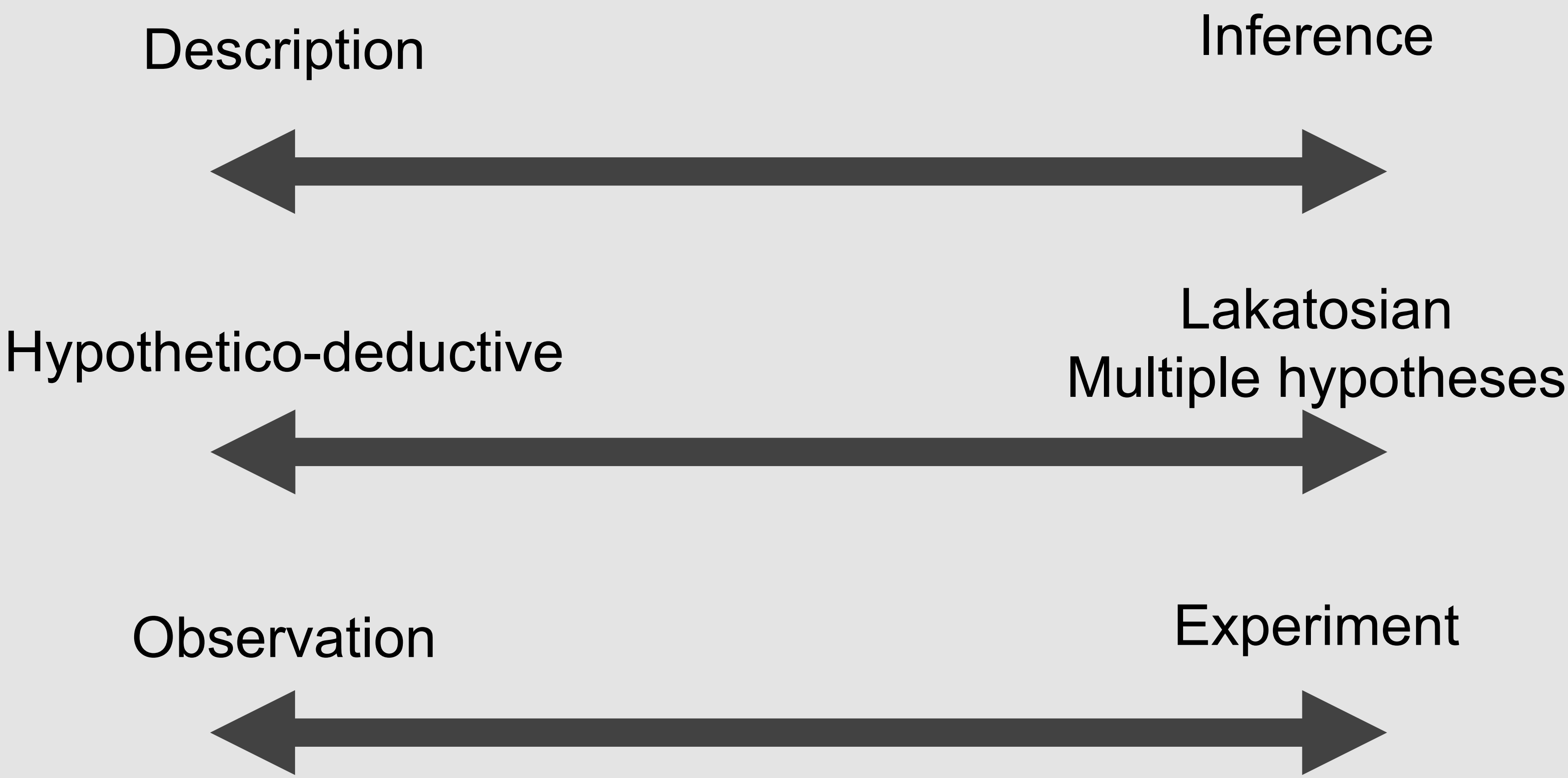


**Experimentation**
Control

**Observation**
Realism

- Lab Experiment
- Microcosm
- Parallel controlled field experiment
- Before/After field experiment
- Natural Experiment
- Observation
- Natural History

# Three Continua in Science



# Statistics and the scientific method

- How do we incorporate statistical thinking into our science?
- Fundamentally statistics are statements of probability.
- I.e. a p value is a statement about the probability of $P(\text{Data} \mid H_0)$.

# Statistics and the scientific method

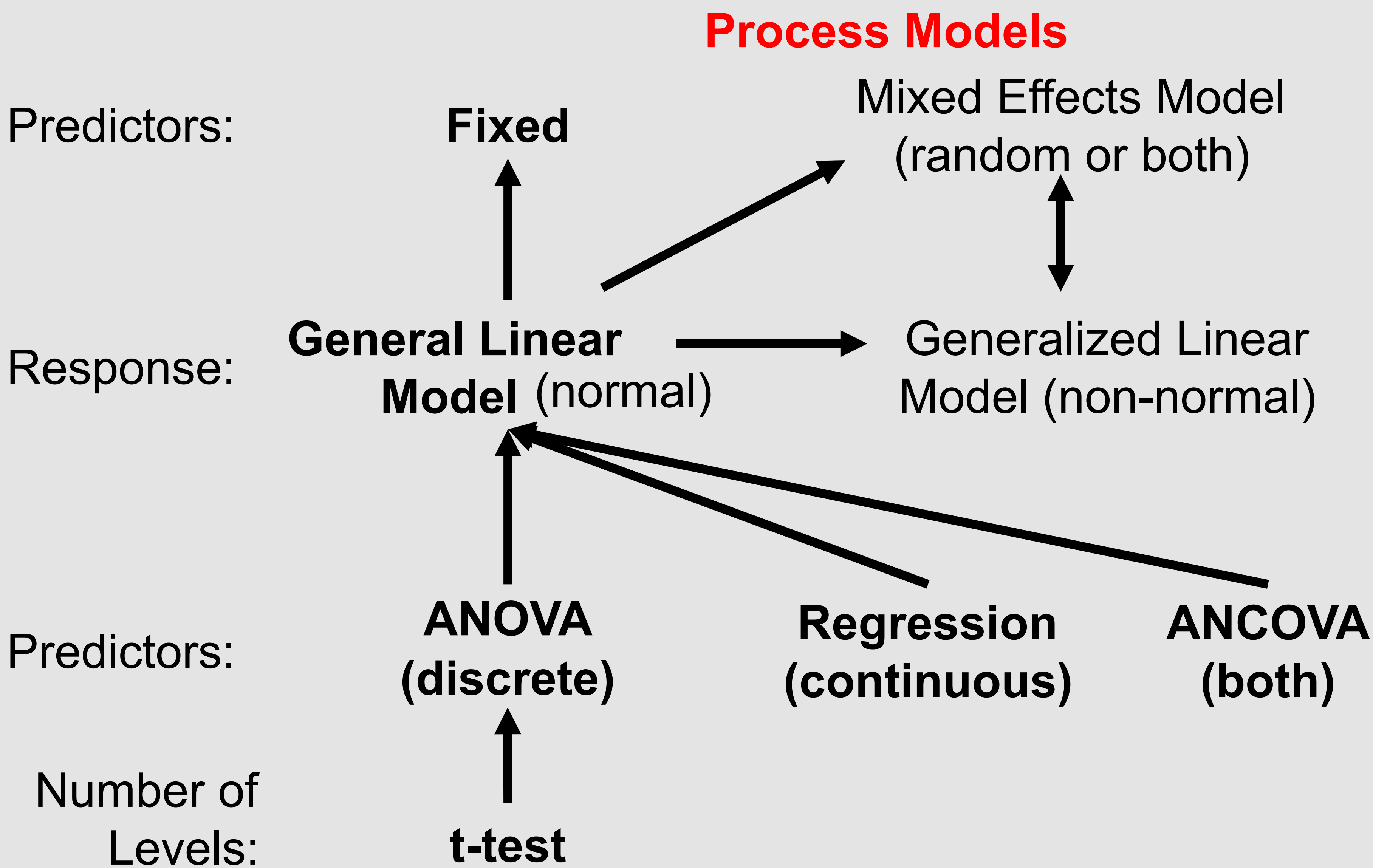
- I.e. a p value is a statement about the probability of $H_0$.
- When specifying a level for alpha (usually 0.05) this is arbitrary.
- Do you really feel more confident in rejecting a null hypothesis for a p=0.1, 0.05? 0.01?

# Approaches to Statistical inference

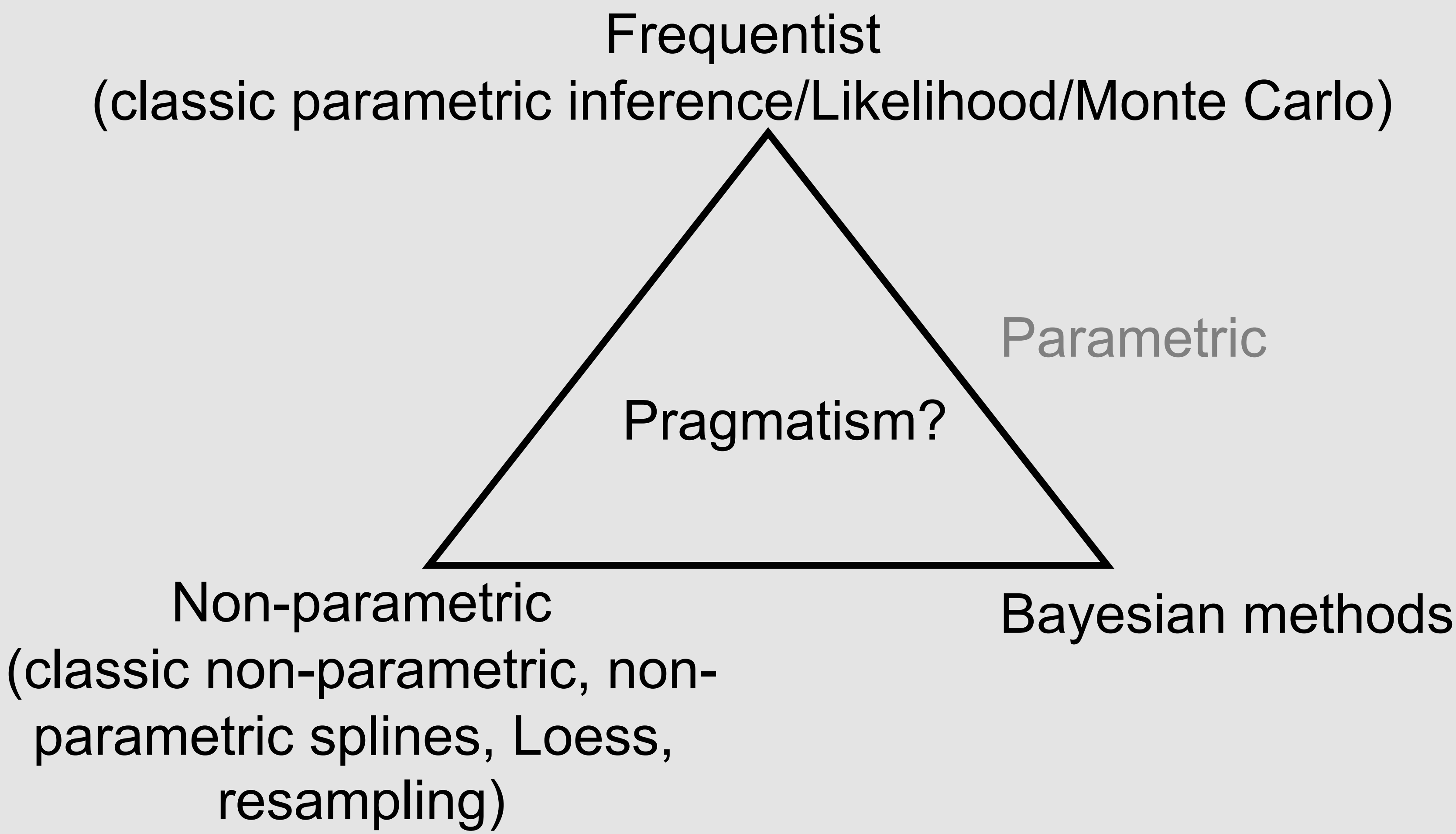
1. Classical/Frequentist
2. Likelihood
3. Bayesian
4. Nonparametric

# Continuity of Statistical models

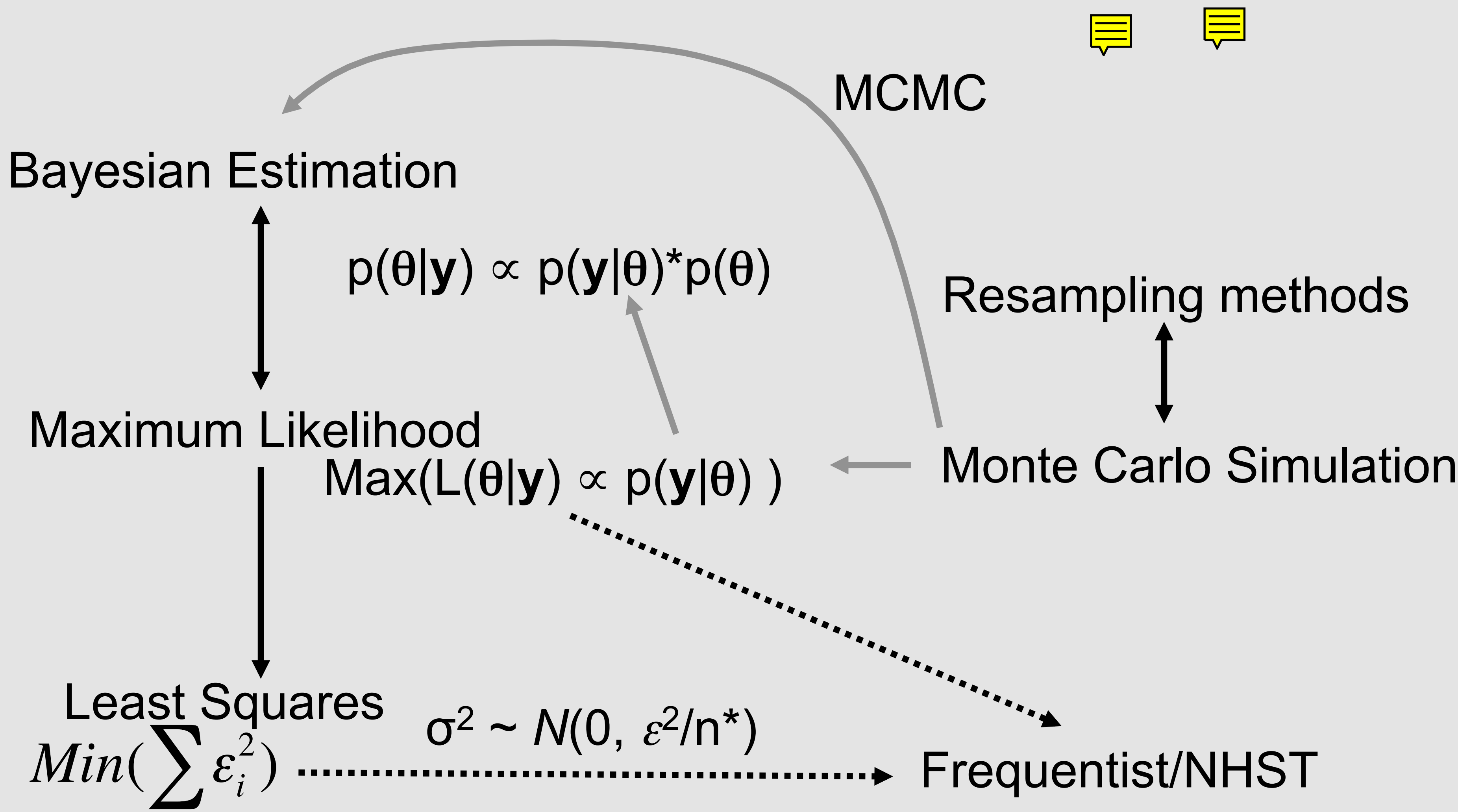
Frequentists/likelihood/Bayesian styles can be used with any of these models



# The philosophical axes of inference



# Relationship between Estimation/inference methods



# Better ways to re-organize the previous slide

- By the end of the class, who ever can get me the best organized slide displaying the relationships between the methods wins a prize.

# Approaches to Statistics

1. **Frequentist/NHST**
2. **Likelihood**
3. **Bayesian**
4. Nonparametric

NHST, Likelihood & Bayesian are all “parametric” in the sense that they depend on *apriori* decisions about the probability distributions used to model the variation in the data.

# Approaches to Statistics: NHST

- More properly called “hypothesis-testing” approach or

  Null hypothesis significance testing (NHST)
  - Frequentist is name given by the Bayesians (because of assumptions derived from sampling theory)
- Central idea is:
  - Test null hypothesis ($H_0$) vs. alternative ($H_a$)
  - Calculate a p-value (probability of rejecting the null hypothesis when it is in fact true)
  - Reject null, accept alternative if p “small”
- Hypothesis is predetermined, data are collected
- Probability of Data given the NULL hypothesis
    - $P(data|H_0)$

# Approaches to Statistics: NHST

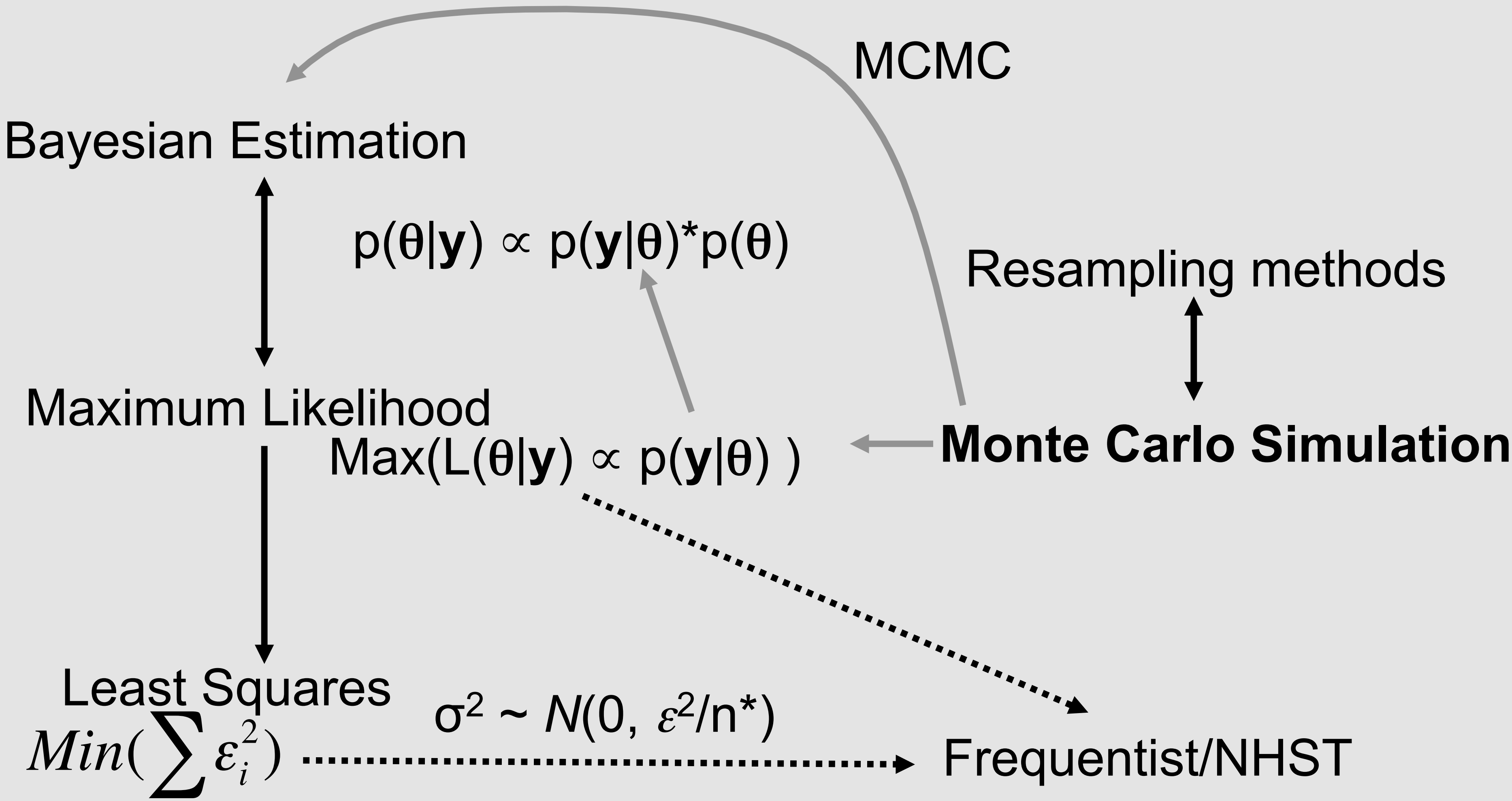
- Confidence intervals are calculated based on a combination of observed variation of the sample (from the standard error), and assumed properties of the distribution of the larger population (sampling theory).
- It is not only approaches based on Least Squared Errors (LSE) that utilize NHST. There are many examples of this for Likelihood, Monte Carlo simulation, resampling, and even some folks getting pseudo p-values from Posterior probabilities from a Bayesian approach.
- Using LSE does not mean that you are necessarily using NHST.

# NHST: critiques

- So often mis-used and mis-interpreted, especially because of “canned” methods.
- Inferences are not based on the observed properties of the data, but on comparisons of observed to unobserved samples representing unlikely draws of the population (i.e. testing $H_1$ vs $H_0$ using the tail of a distribution like the t or F distribution).
- P values are a crutch, and are so often incorrectly interpreted.
- Lots of parametric assumptions ( normality, iid, homogeneity of residual variation).
- Confidence intervals do not represent what people think. They represent how often the point estimate would lie within a given interval given repeated (and hypothetical) resampling.

What situations might a p-value have value? When does it lead us astray?

# Monte Carlo Methods



# Monte Carlo Simulations

- A parametric approach that shares a great deal philosophically with a frequentist framework.
- Based on parametric assumptions and parameter estimates, model the expected distributions for the parameters.
- Can be very useful for dealing with small sample sizes.
- Very dependent on the assumptions of the model framework.

# Likelihood based estimation & inference

MCMC

Bayesian Estimation

$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)^{*}p(\theta)$

Resampling methods

**Maximum Likelihood**

$\mathrm{Max}(L(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta))$

Monte Carlo Simulation

Least Squares

$Min(\sum \varepsilon_i^2)$

$\sigma^2 \sim N(0, \varepsilon^2/n^*)$

Frequentist/NHST

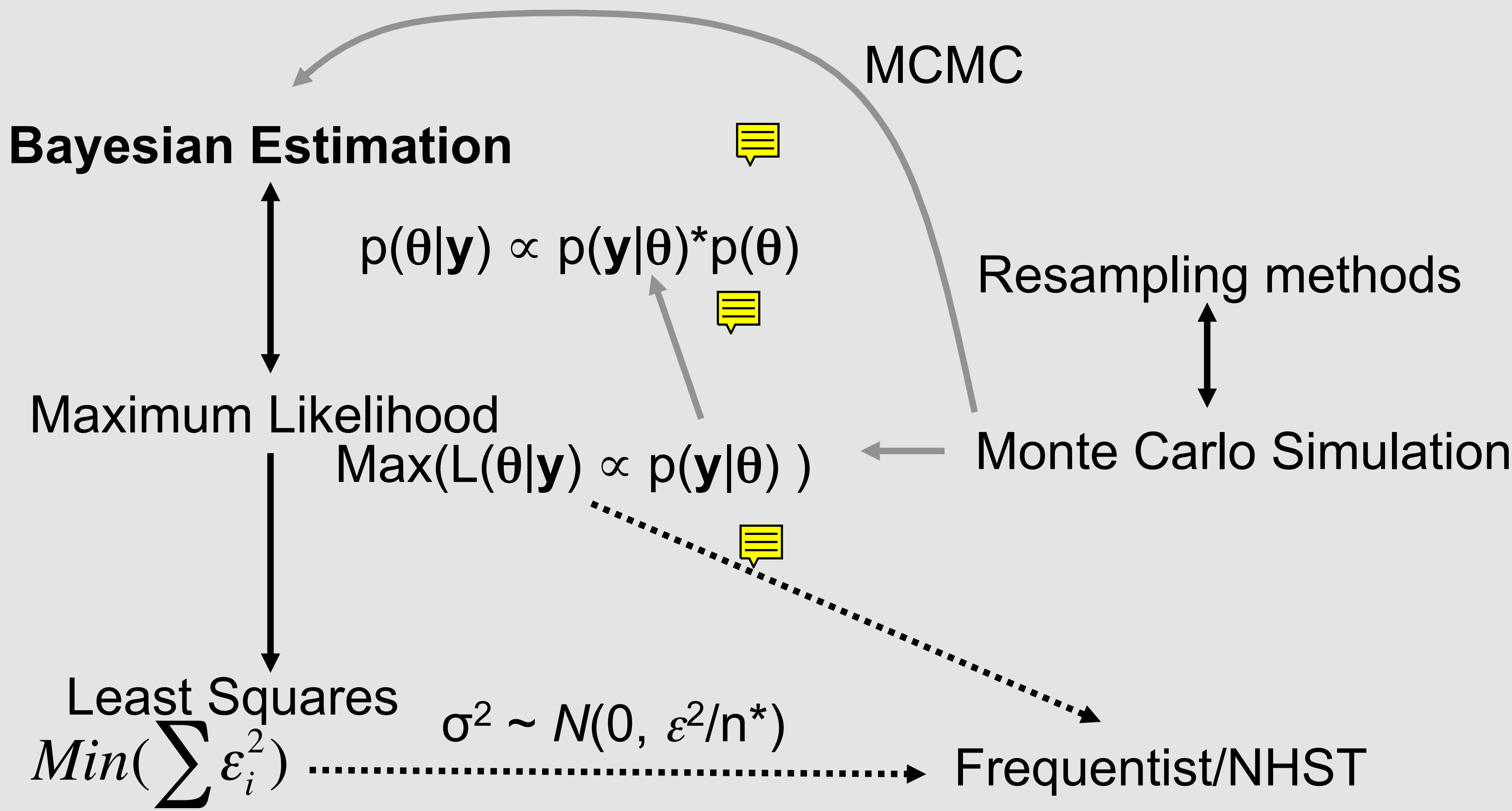
# Approaches to Statistics: Likelihood

- Estimation is based on ***assumed*** parametric distributions
- Data are taken as fixed and the relative probability of hypotheses are judged.
- Probability/Likelihood of hypotheses given the data
  - P(Data| Hypothesis) proportional to L(*Hypothesis*$_i$|Data)
- Likelihood methods can (and often are framed ) in a NHST framework, rather than as statements of relative probabilities.
- Forms the foundation of Bayesian statistics
- Likelihood allows us to to connect probability to information theory and the development of Information Theoretic approaches.

# Likelihood:critiques

- If used in an NHST framework, most of the same concerns apply as to the frequentist approach.
- Complex problems rarely have analytic solutions, so we must use approximations (but this is true for most complex problems).
- Estimates of variance (for small samples) can be biased. In practice for even moderate data sets, this is not a real problem. However for mixed models (models with random effects), this problem can become important (in particular for multivariate mixed models where you are estimating covariance matrices).

# Bayesian Estimation & Inference



# Bayesian estimation

$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)*p(\theta)$

- Bayesian's argue we are not interested in the relative probability of the data given the model/parameters.

- We know the data, we are interested in the probability of the model given the data!!!!!

- Need to assign a statement about **prior** belief in the model, $p(\theta)$ . Combine this with the likelihood to get the **posterior probability of the model**.

- Bayesian's are not interested in NHST methods, and instead treat the model parameters as random variables, the **posterior distribution**.

# Bayesian estimation

$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)*p(\theta)$

- Bayesian's are not interested in NHST methods, and instead treat all model parameters as random variables.

- Effectively Bayesian approaches focus on the interval estimates (as opposed to point estimates) analogous to the confidence intervals (**credible intervals**).

- A 95% credible interval (highest posterior density) is said to have a 95% probability of containing the true parameter estimate.

# Bayesian estimation: critiques

$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)*p(\theta)$

- Prior information: Subjective… (some say cheating). A strong enough prior can give you the answer you want regardless of the data

- Can be mathematically difficult. More generally computationally intensive (answers are approximate), and so requires some skills.

- There have been some arguments about whether a Bayesian approach could ever be “scientific” in a Popperian sense, as you do not reject hypotheses, you just assign relative probabilities to each.

# Approaches to Statistics: Nonparametric/randomization

- NOT based on predefined distribution.
- No underlying distribution assumed.
- Inference based on ranks or empirically derived distribution (resampling)
- Ranks
  - Traditional nonparametric techniques
    - e.g. Spearman rank correlation, Mann-Whitney U test
- Resampling
  - Computer intensive procedures
    - Jackknife/bootstrap/Monte Carlo methods

# Approaches to Statistics: Nonparametric/randomization

- Resampling methods can be used for a broad range of problems even without any knowledge of parametric form of the variation.
- While computationally intensive, they are easy to generate even by hand with simple spreadsheets.
- Can be used for “make your own statistics” sorts of problems.
- Based on the data, fewer assumptions (for instance bootstrapping does not require identical distributions of the data between treatments).

# Approaches to Statistics: Nonparametric/randomization

- Computationally intensive, for complex problems it can be difficult how to figure out how to perform the resampling
- Arguably limited in value as there is no formal logic as to how to use inductive approaches (go from your data set to the population at large).
- More generally they are a tool to help in making inferences or estimating uncertainty, and are not useful in thinking about the overall model.

# Class R style guide

http://www.msu.edu/~idworkin/ZOL851_style_guide.html