Introductions
Most environmental data violates the assumptions of these tests.
What is statistical modeling?
What is statistical modeling?

“Confronting models with data”

- Model fitting / parameter estimation
- Model comparison
- Estimation, partitioning, and propagation of uncertainties
What is statistical modeling?

“Confronting models with data”

Design the statistical analysis to fit the data rather than the data to fit the test
What is a model?
What is a model?

A conceptual, graphical, or mathematical representation / abstraction of some empirical process(es).

A mathematical function that formalizes our conceptual model / theory

\[ f(x) = a \]
What is a model?

Models are HYPOTHESES
Syllabus
Course Materials

- Reading assignments, lecture slides, project details, etc. are all posted on the lab website http://people.bu.edu/dietze/Bayes2020/EE509.htm

- Primary Text:

- Software:
  - R / RStudio
  - OpenBUGS / JAGS
  - Git / GitHub
Grading

Grading will be based on lab reports, a semester-long project, and four exams.

Lab reports/problem sets (10 points each) = 150

Semester project = 95
  project proposal 2/14 (10)
  model description 3/6 (15)
  preliminary analysis 4/10 (20)
  Final report before exam 4 (50)

Exams (30, 25, 30, 30 points ) [non-cumulative] = 115

Total = 360
Labs

- LAB IS MANDATORY

- Labs will be posted in git repository https://github.com/mdietze/EE509

- Due FOLLOWING WEEK by the start of lab

- Must be turned in individually

- Can work together
Semester Project

• Final product: “Journal article” on a data analysis
  • You choose topic
  • ENCOURAGED to use your own data
  • Analysis must be new, use concepts from class
  • “Methods” heavy
• Four milestones
• One lab is peer critique
Lecture & Exams

• Four sections
  • Probability theory and Maximum Likelihood
  • Bayesian methods
  • Hierarchical/mixed models
    - Linear regression \(\rightarrow\) nonlinear, non-gaussian
  • Advanced topics
    - Time series
    - Spatial
Exams

- Multiple Choice
- Matching
- Fill in the blank
- Short Answer / Derivation
- ~15 questions
Expectations

- You have seen basic calculus at some point
  - Primarily need to follow derivations
- Basic familiarity with statistical concepts
  - e.g. experimental design, randomization, mean, median, variance
- Open mind
- You will work hard
- You won't 'get' Bayes the first time they see it (but will need to by the 2\textsuperscript{nd} exam)
Objectives

- Literacy
  - Read and evaluate advanced stats used in papers
- Proficiency
  - underlying statistical concepts
  - Software: R, JAGS
- Exposure to advanced topics
- Paradigm shift
A bit more on motivation....

Data are usually complex

Violate the assumptions of classical tests

This complexity can be addressed with modern techniques
Example:
How much light is a tree getting?
Example:
How much light is a tree getting?

- Dominant
- Intermediate
- Suppressed
Linear models

Logistic

Multinomial

Non-zero ECA observations

\[ \lambda^{(e)} > 0 \]

Light availability

Linear scale

R.S.

Status observations

\[ \lambda^{(s)} \]

Probability

Linear scale

Zero ECA observations

\[ \lambda^{(e)} > 0? \]

Log scale

Model light estimate

\[ \lambda^{(m)} \]

Field

Model

\[ \beta_0 \beta_1 \]

\[ c_0 \]

\[ \nu_l \]

\[ c_1 \nu_e \]

\[ a_0 a_1 \nu_m \]

Posterior light estimate
Problem Characteristics

- Multiple data constraints
- Non-linear relationships
- Non-Normal residuals
- Non-constant variance
- Latent variables (response variable not being observed directly)
- Distinction between observation error and process variability
- Missing data
Statistical Paradigms

- Classical (e.g. sum of squares)
- Maximum Likelihood
- Bayesian
## Statistical Paradigms

<table>
<thead>
<tr>
<th></th>
<th>Statistical Estimator</th>
<th>Method of Estimation</th>
<th>Output</th>
<th>Data Complexity</th>
<th>Prior Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical</td>
<td>Cost Function</td>
<td>Analytical Solution</td>
<td>Point Estimate</td>
<td>Simple</td>
<td>No</td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>Probability Theory</td>
<td>Numerical Optimization</td>
<td>Point Estimate</td>
<td>Intermediate</td>
<td>No</td>
</tr>
<tr>
<td>Bayesian</td>
<td>Probability Theory</td>
<td>Sampling</td>
<td>Probability Distribution</td>
<td>Complex</td>
<td>Yes</td>
</tr>
</tbody>
</table>
# Statistical Paradigms

<table>
<thead>
<tr>
<th></th>
<th>Statistical Estimator</th>
<th>Method of Estimation</th>
<th>Output</th>
<th>Data Complexity</th>
<th>Prior Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical</td>
<td>Cost Function</td>
<td>Analytical Solution</td>
<td>Point Estimate</td>
<td>Simple</td>
<td>No</td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>Probability Theory</td>
<td>Numerical Optimization</td>
<td>Point Estimate</td>
<td>Intermediate</td>
<td>No</td>
</tr>
<tr>
<td>Bayesian</td>
<td>Probability Theory</td>
<td>Sampling</td>
<td>Probability Distribution</td>
<td>Complex</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The unifying principal for this course is statistical estimation based on **probability**
Next lecture

- Will cover basics of probability theory
- Read
  - Clark 2007 - Chapter 1
  - **Hilborn and Mangel p39-62** (course website)
- Optional
  - Clark 2007 – Appendix D (Probability)
  - Otto and Day – Appendix 1 (Math) and 2 (Calculus) (course website)