ECO 602
Analysis of Environmental Data

FALL 2019 – UNIVERSITY OF MASSACHUSETTS

DR. MICHAEL NELSON

Welcome!

Today's agenda

1. A little about me.
2. Course overview and philosophy.
3. A little about you.
4. Description and inference
5. Models and types of data

My background: education and research interests

Plant biologist by training

Spatial analysis

Ecological modeling

Agent-based modeling

Invasive species

My background: Teaching

Biostats

Spatial agent-based modeling

Spatial analysis

Disclaimers

I don’t know everything about statistics. Not even close.

I am not a math whiz or a statistician by training.

I am intimidated by unfamiliar statistical techniques.

I am terrified by masters unfamiliar statistical techniques.

Disclaimers

I’m still afraid of making mistakes.

I’ve applied many analysis techniques covered in this course but…

There are others that I don’t have first-hand experience with.

But…

Anti-Disclaimers

I’ve learned to use, appreciate, and even understand, many statistical analyses.

I’ve managed to teach statistics to biology and geography students.

I love to deepen my understanding of statistical analyses for my own research and understanding.

I love teaching analyses from a non-statistician point of view (since I’m not a statistician).

My experience learning data analysis and statistics

I had minimal formal training as an undergrad and afterward as a lab technician.

I thought statistics would be easy!

I had a lot of misconceptions.

It turns out, it's not an easy subject after all.

My experience learning data analysis and statistics

My first course covered graduate level stats and programming. I succeeded at neither.

I was very self-conscious of my mathematical abilities.

I couldn’t see the bigger picture through the details.

But then…

My experience learning data analysis and statistics

I started talking about analyses with my graduate student peers.

I tried hard to develop some intuition and literacy.

I started to feel comfortable with R programming.

I became fascinated by randomness and the idea of comparing simulated data with real observations.

My experience learning data analysis and statistics

Two core principles that I use when I start to get lost in the details:

1. What would we expect to observe in our study system if it were completely random?
2. Much of statistics comes down to estimating the center and spread of data.

What this course covers

Statistical literacy: a bird's eye view

Critical thinking about data

Analytical methods you will encounter

Paradigms and philosophies of statistics

Survey of distributions and probability basics

What this course does not cover

Theoretical and mathematical basis of statistics

Implementation of data analyses

Probability theory

Designing experiments

Proving theorems

My teaching approach

This course is for you. You are here because you want to be.

I appreciate feedback. I want to know what is working and what isn't.

It is my goal for this course to be challenging, but not overwhelming. I need your help to gauge this.

My teaching approach

Lecturing at you for 75 minutes is not ideal. Our meetings will consist of lecture combined with other, active components.

This approach won't succeed without your active participation.

My teaching approach

Your time (and mine) is valuable.

Let’s be respectful and efficient.

I want you to ask questions, in class, via email, in office hours.

Course Format

Class meetings are 75 minutes.

I don’t like to lecture at you for 75 minutes.

Meetings will be made up of lecture, question and answer periods, and small group or whole class discussions.

Your participation is essential.

Assignments and Grading

There are no in-class exams

Assignments and Grading

There are four components to your grade:

1. Short in-class quizzes, 15% of total
2. 4 group take-home problem sets, 50% of total
3. Critical review of a paper, 15% of total
4. Final take-home problem set, 20% of total

Brief overview of course content

Weeks 1 – 3: overview, characterizing and exploring data

Weeks 4 – 6: models and inference

Weeks 7 – 9: inference paradigms

Weeks 10 – 11: bestiary of analysis types

Weeks 12 – 13: recap and summary

I want to hear from you!

What do you remember from statistics class?

Variance

Central tendency

Binomial distribution

Bayes rule

Mean

Central limit theorem

Z-values

Normal distribution

Standard error

Sum of squares

Regression

T-tests

Probability density function

Confidence intervals

What do you remember from statistics class?

Don’t panic!

You don’t need to remember the terms on the previous slide.

Statistics is inherently difficult.

We will focus on broad concepts, literacy, and intuition.

What do you hope to learn in this course?

Let’s dive in to some content!

Why do we need data and analyses?

Describe systems

Discover and explain patterns

Explore relationships

Support or refute claims

Explore hypotheses

Why do we need data and analyses?

Communication, with scientists and non-scientists

Build models

Understand interactions

Formulate and test hypotheses

Quantify uncertainty

Description and Inference

A useful, and often underappreciated distinction:

Statistic: A number that summarizes the properties of a collection of specific observations.

Parameter: A number that estimates properties of an entire population from a collection of observations.

These terms are used loosely and you must know the context

Descriptive Statistics

Characterize the properties of a set of observations

Common approaches:

1. Numerical summaries
2. Graphical summaries

Descriptive Statistics

Measures of central tendency

- Mean
- Median
- Mode

Descriptive Statistics

Measures of variability

- Data range
- Interquartile range
- Variance and standard deviation
- Distributions
- Data symmetry

Inferential Statistics

Estimate properties of a larger population from a sample of observations.

Understand and/or quantify relationships among sample attributes and the environment.

Estimate parameters for a model of the population.

Compare different models of the population

Models

We use models every day.

They are simplified representation of reality.

We need mental models to get through the day.

This may seem obvious and simplistic, but it is profound.

Some models (that we might not realize are models)

Sunrise/sunset times

Seasons

Tides and moon phases

Weather forecasts

Our mental, and graphical maps of campus

Interpersonal interactions

Model Thinking: How to think like a modeler

What are the important parts of a system?

How could I represent attributes in a model?

What data would I need?

How might system components interact?

Model Thinking: How to think like a modeler

What components are vital, and which may be unimportant?

What are the possible sources of randomness and uncertainty?

How do I expect the system to behave?

How could I use data to support or refute my model?

Models in science

Simplified representation of how we think a system behaves.

George Box: “All models are wrong, but some are useful”

Phenomenological

Mechanistic

Models: Deterministic Components

Describes a process, usually a mathematical function of one or more variables, that always produces the same output from the same input. Explore relationships

The behavior of deterministic processes are often well characterized and facilitate predictions of future behavior.

But…

Models: Deterministic Components

Some deterministic functions do not produce predictable output!

They have reproduceable behavior, but long-term predictions are impossible.

The fields of dynamical systems and chaos: butterfly effect, sensitive dependence on initial conditions.

We won't focus on these, but it is important to know they exist.

Logistic population growth, strange attractors, etc.

Models: Stochastic Components

A stochastic process produces different output for repeated trials using the same input.

Stochastic models describe random processes.

Stochastic models can capture the variability, or noise, that is not explained by deterministic model components.

Uncertainty

Models: Exploring Relationships

Predictor variables

Response variables

Types of Data

A short list of data classifications:

Numeric	**Continuous**	**Discrete**
Categorical	**Ordinal**	**Binary**

Can you think of others?

For next class:

Critically read the four scenarios in the McGarigal reading.

While you read, consider:

- Types of data
- Uncertainty
- How do the results support (or fail to support) the claims?