ECO 602 Analysis of Environmental Data

FALL 2019 – UNIVERSITY OF MASSACHUSETTS

DR. MICHAEL NELSON

Inference Paradigms

Scatterbrain

I have a lingering feeling that I have forgotten something…

If you have sent me materials to look at and/or asked me a question via email and I haven’t responded, please send me a kind reminder!

Question set 2 timeline

Some recent, pertinent ECO events

Rebecca Hernandez UC Davis: The Sustainability and Ecology of Renewable Energy in an Unprecedented Energy Transition

- Rebecca demonstrated that as ecologists, we have very high statistical literacy, and we must use our skills to inform decision-making.
- Rebecca used the example of an engineer inferring properties of all solar farms from a sample of size 1.

Jason Carmignani: PhD dissertation defense

Jason's defense included some excellent uses of descriptive and inferential statistics including:

- Descriptive plots of continuous lake level data
- Inference using general linear mixed models

Today's Agenda

1. Bayesian thinking video
2. Quick review of distributions
3. Inference paradigms
4. Recap of stochastic and deterministic models

A Visual Guide to Bayesian Thinking
Juila Galef

https://www.youtube.com/watch?v=BrK7X_XlGB8

Check out her channel, it’s rad!

Distributions recap

1. What did we learn?

Probability theory and distributions: some essential terms*

Event, observation, realization, independence

Sample, sample space

Random variable

Probability mass, probability density

Cumulative probability

Empirical distribution

*Disclaimer: not an exhaustive list.

Inference paradigms

1. What is a paradigm?
2. Which inference paradigms have we heard of?

Inference paradigms

1. Parametric and non-parametric inference
2. We prefer to use parametric!
   1. Parametric inference depends on distributions.
   2. Parametric inference has lots of assumptions.
3. Sometimes we must use non-parametric.

Parametric inference paradigms

1. Frequentist
2. Bayesian
3. Likelihood is a shared concept used in both frequentist and Bayesian paradigms.

Frequentist paradigm

Frequentist inference is what we usually mean when we talk about statistics.

Frequentism was the dominant paradigm.

Frequentism has pros and cons.

Frequentist pros

- Makes no assumptions about prior knowledge of a system.
- Usually mathematically tractable.
- More widely known than Bayesian.
- Some describe the frequentist paradigm as less ‘subjective’

Frequentist cons

- Depends on the concept of repeated sampling.
- Assumes that true population parameters are unknown or unknowable.
- The repeated sampling concept and the assumption of the existence of a true but unknowable model lead to an unexpected interpretation of confidence.

Bayesian paradigm

Inspired by Bayes' rule about conditional probability.

Unconditional probability: What is the probability of event X?

Conditional probability: What is the probability of event X if we know that Y is true?

Bayesian paradigm: conditional probability

Bayes’ rule states *very* roughly that we must take into account our prior knowledge to calculate the accurate probability of an event.

Our inference about event X is **conditioned** by our knowledge of the outcome of event Y.

Let’s draw a Venn diagram:

Bayesian inference

Requires a **prior** probability distribution.

Calculates a **posterior** probability distribution from our prior distribution and our sample data.

The choice of prior distribution can be difficult or problematic.

Choosing a prior distribution is where some consider Bayseian inference to be more **subjective** than frequentism.

Bayesian pros

1. Does not assume the existence of a true model.
2. Does not depend on hypothetical repeated sampling.
3. Interpretation can seem more intuitive than frequentist interpretation
4. Prior information may be extremely relevant for inference!

Bayesian cons

1. The math can be much more complicated!
2. Less widely known.
3. Some say it is inappropriately subjective.
4. Implementation is generally more difficult and computationally intensive.

Likelihood

1. A common ground between Bayesian and frequentist approaches.
2. Maximum likelihood method attempts to find the population parameters that make the observed data the most likely.
3. The likelihood of an event is proportional to probability mass or density.

Maximum likelihood

1. Given a data set, find the distribution parameters that maximize the likelihood of the observations.
2. Example on board with a Normal distribution:
3. I really like the explanation of likelihood in the Bolker 2007 book, chapter 6: Ecological Models and Data in R

Frequentist confidence

1. Frequentist confidence comes from the repeated sampling concept.
2. We are confident that if we repeated our experiment many times, our estimates of the population parameters would be close to the true population values n% of the time.
3. Our confidence is focused on the process

Bayesian confidence

1. We are confident in our sample: our observed data are real and not just one possible outcome of our sampling procedure.
2. We can calculate a range of population parameter values that are reasonable based on our observed data.

Confidence/credibility

Frequentist: the true population parameter value either is or isn’t within an interval calculated from our data. We are confident that if we repeated the process many times, we would usually capture the true value in our interval.

Bayesian: given the evidence in our data, we think the true population parameter value lies within a concrete range.

This stuff is difficult and subtle!

We won’t worry about understanding the finer points upon first exposure.

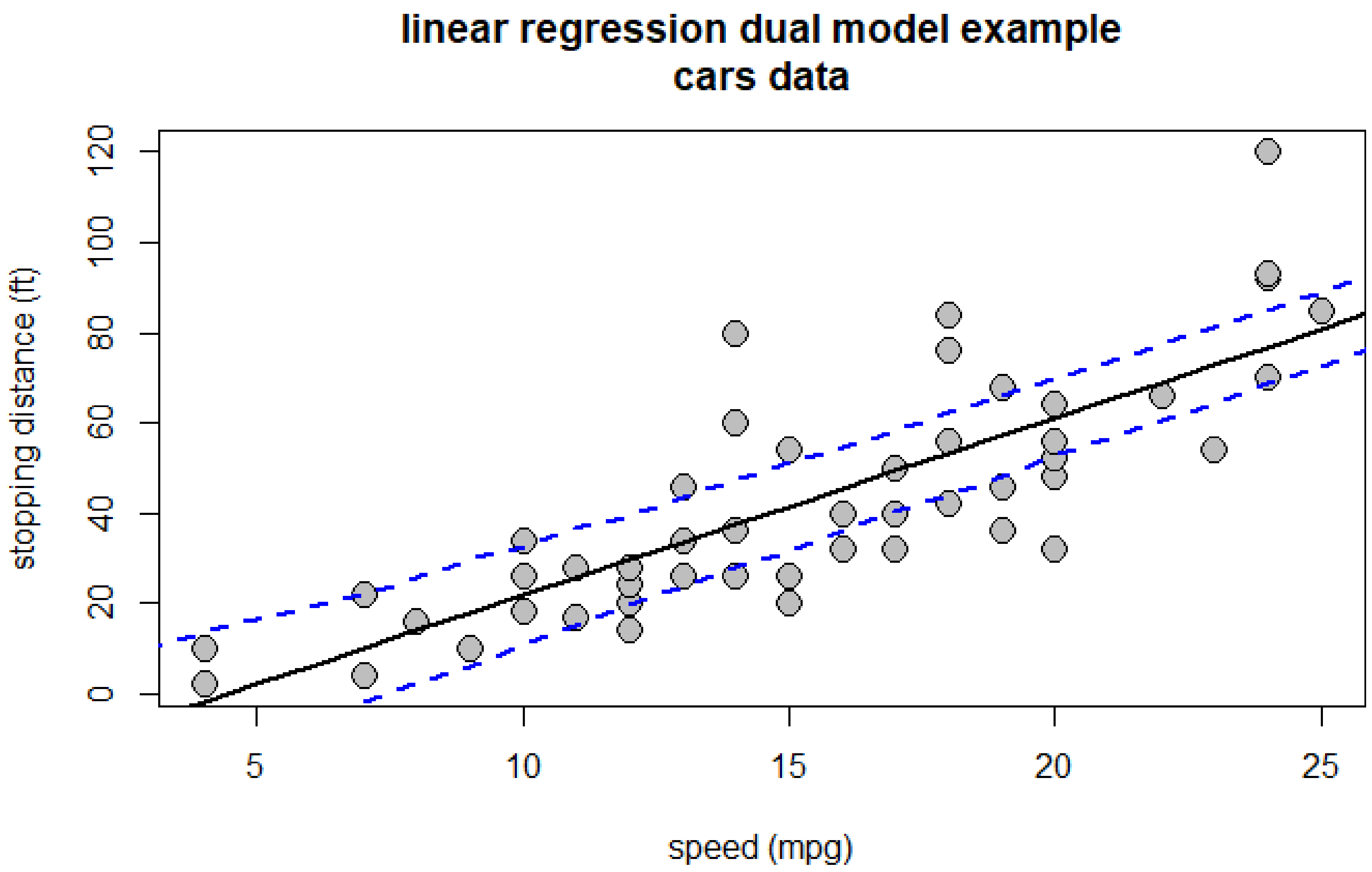
If your head is spinning, that’s ok! It means you have been paying attention.

In my experience, developing an intuition for these concepts occurs over time scales much longer than one semester.

Dual-model paradigm: stochasticity

1. Deterministic: model of means
2. Stochastic: model of variability
    1. Error, noise, variability in the sample and population

Dual-model: linear regression example



For next time:

Read the McGarigal chapter 5c: hypothesis testing

I haven't decided on the question set 3 questions yet.

I may merge sets 3 and 4.

I will update the distribution and due dates on the syllabus.

No class on Tuesday!