ECO 602
Analysis of Environmental Data

FALL 2019 – UNIVERSITY OF MASSACHUSETTS

DR. MICHAEL NELSON

Inference Paradigms, Hypothesis Testing

Today's Agenda

1. Inference paradigms recap
2. Recap of stochastic and deterministic models
3. Hypothesis Testing

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 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.

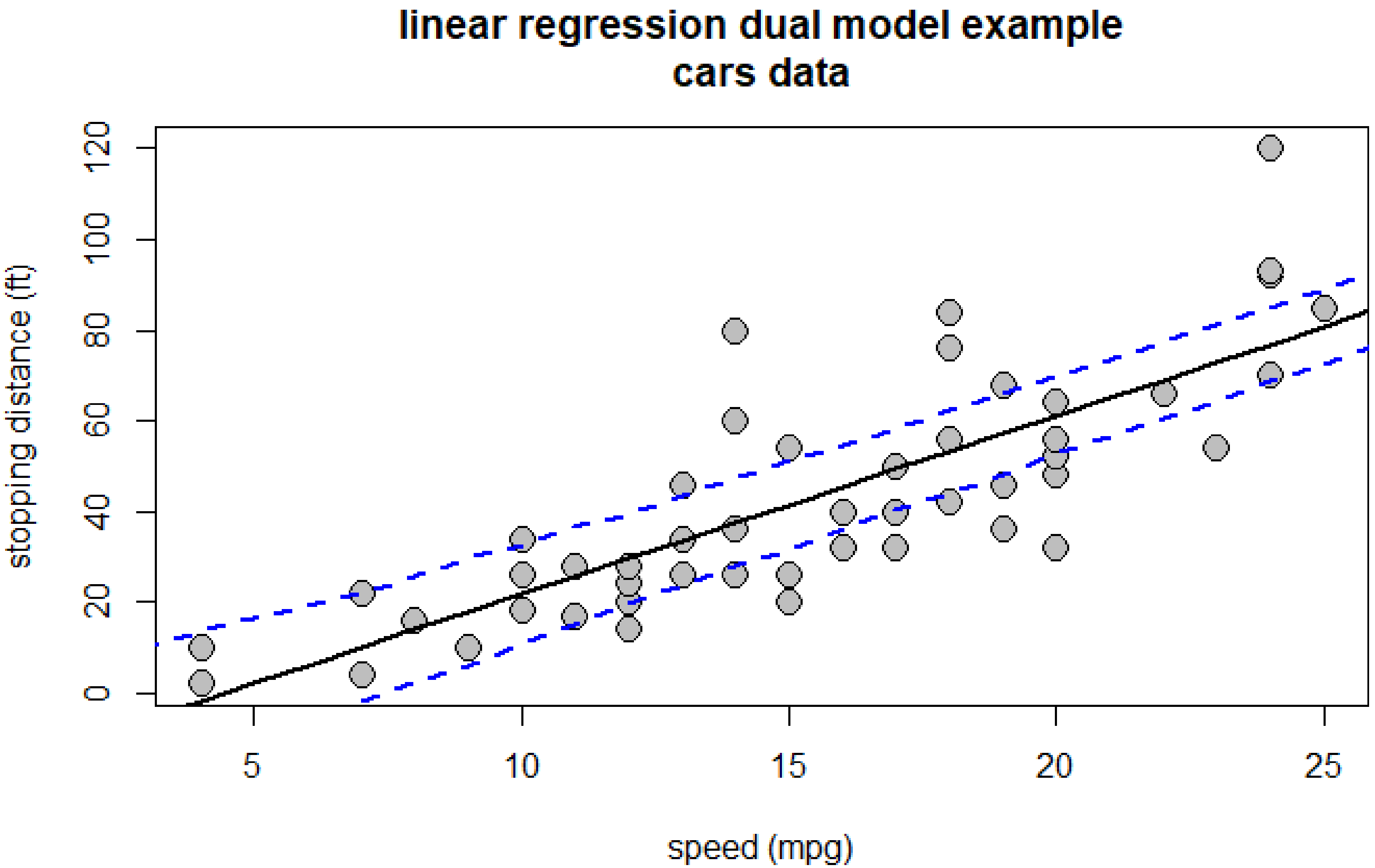
If your head is spinning, that’s ok! It means you have been paying attention.

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

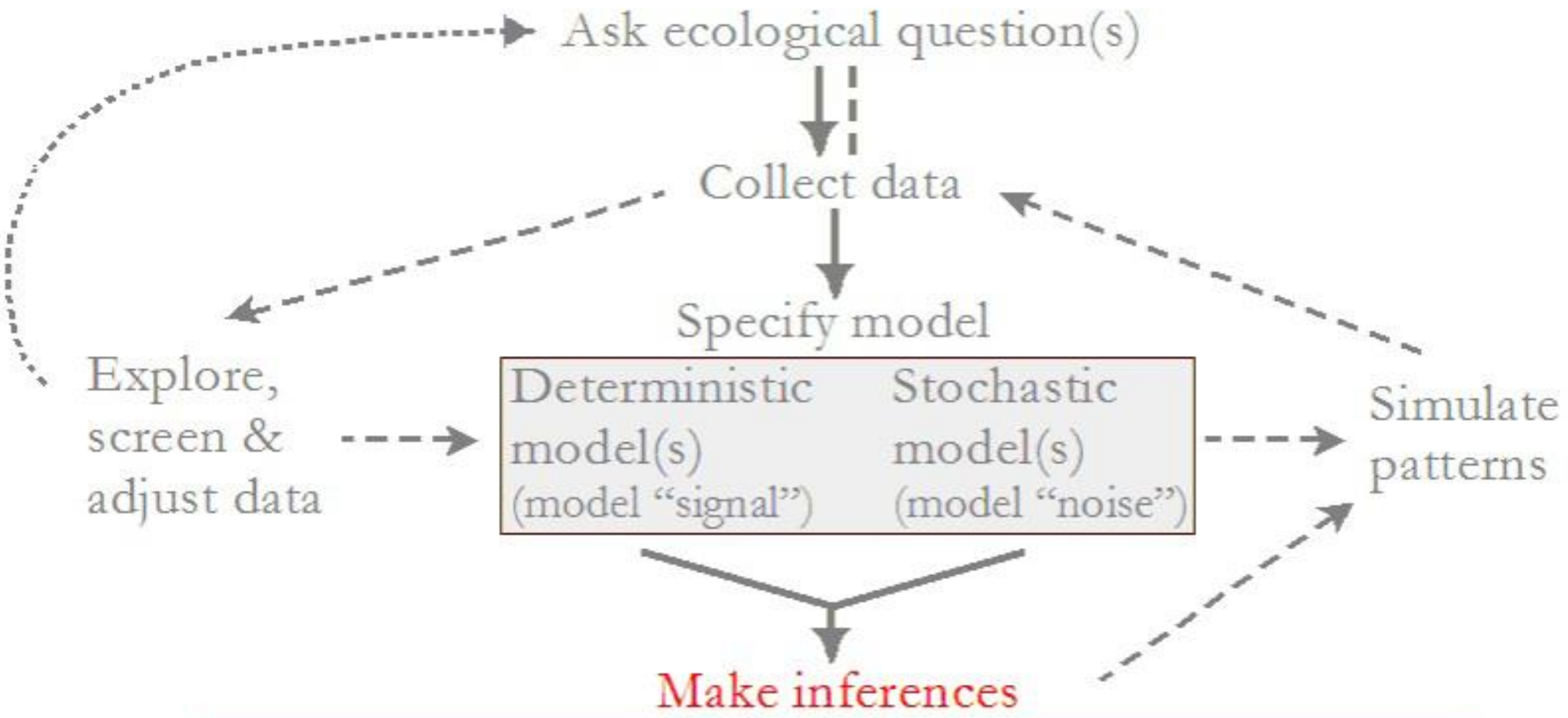
Dual-model paradigm: stochasticity

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

Dual-model: linear regression example



Hypothesis Testing Concepts



- Classical (frequentist) statistical inference generally involves testing null hypotheses, computing *p*-values, and often making decisions to reject the null hypothesis or not and assessing the associated decision errors.

Hypothesis Testing

We will examine hypothesis testing from a frequentist perspective.

A way to quantify the strength of evidence for or against a null hypothesis.

Null Hypotheses

We've considered these before in more conceptual model terms.

We need statistical models to conduct hypothesis testing.

Null Hypotheses: Parametric testing

Fit a deterministic function of the means and a **parametric distribution** to the stochastic component of our data.

Distribution parameters are based on the null hypothesis.

Null Hypotheses: Parametric testing

Population parameters are unknown/unknowable.

Sample parameters are statistics.

We have to use distributions that describe samples (not populations).

- These are usually modified versions of the familiar population distributions, for example the normal.

Null Hypotheses: Parametric testing

Parameters are estimated from sample, so they are **statistics.**

Estimated parameters have uncertainty.

P-Values: Frequentist paradigm

P-values have the same interpretation difficulties as other frequentist ideas, that stem from the repeated-sampling paradigm.

P-values can be controversial!

P-values are meant to quantify the strength of evidence against the null hypothesis.

P-Values

P-values are the estimated Type I error:

- False positive rate
- Probability of falsely rejecting a true null hypothesis

Related to confidence intervals (which we'll talk about later)

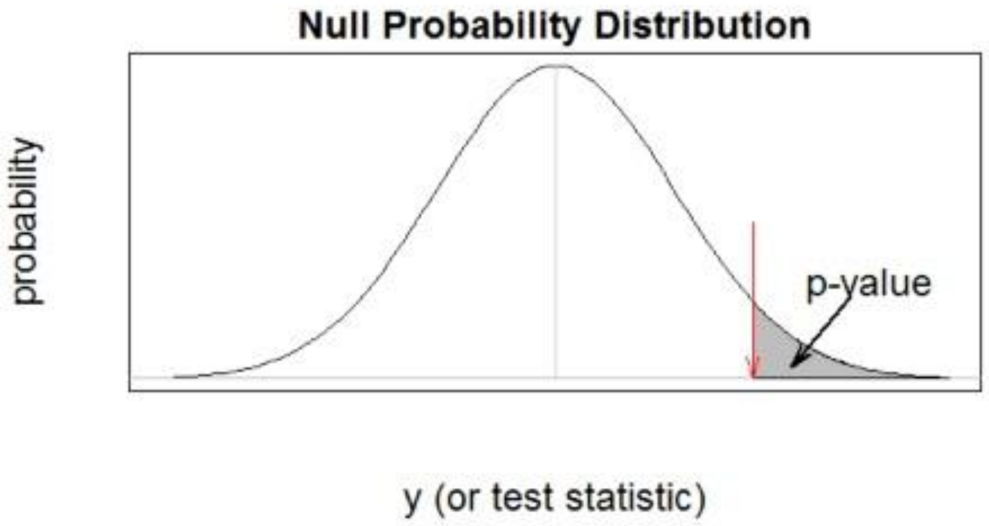
The proportion of repeated-sampling events for which we would falsely reject the null hypothesis.

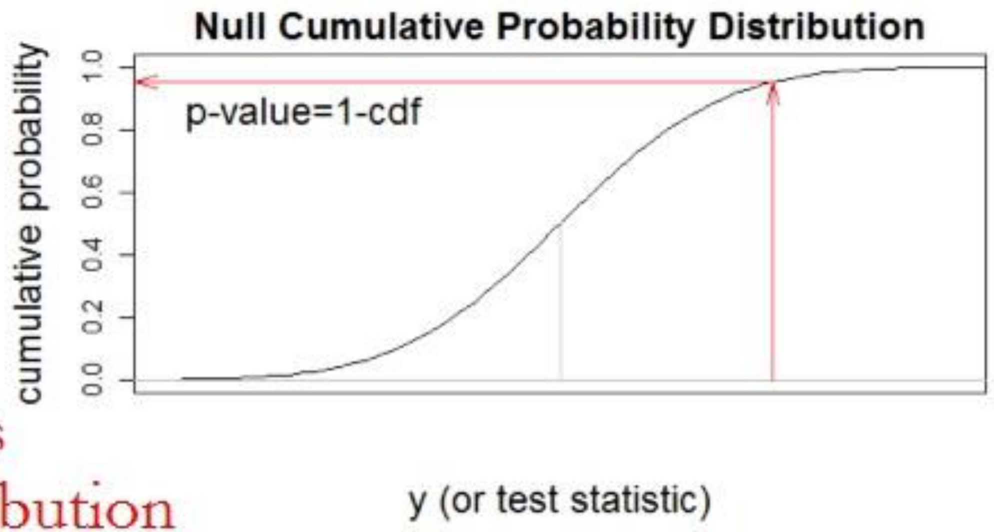
What does $p = 0.05$ mean?

Hypothesis Testing Concepts

P-values

- *Probability* of observing data (Y, or a statistic derived from it, e.g., slope, mean) as large or larger (one-sided evaluation) if the null hypothesis is true (i.e., data was derived from the null probability distribution)
- Strength of evidence against the null hypothesis



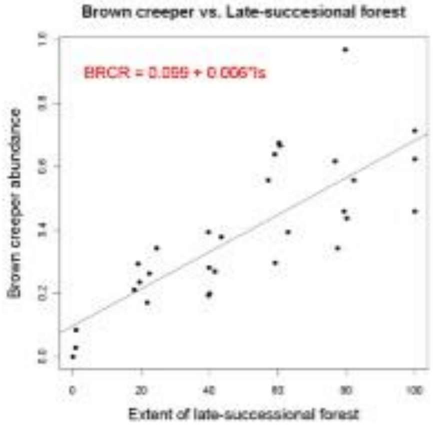


Remember, p-values are always calculated under the Null distribution

Hypothesis Testing Concepts

P-values

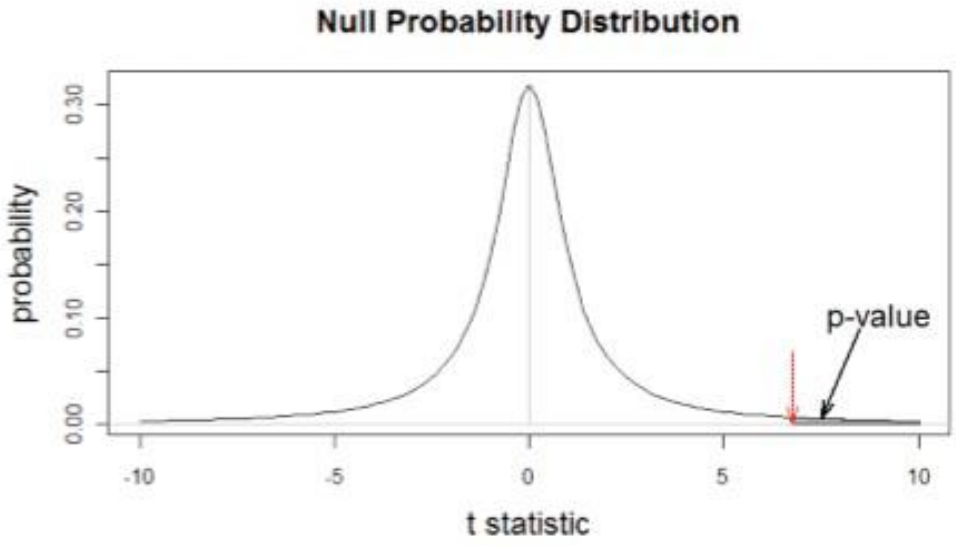
- *Parameters...*
  Probability of observing the value of φ (parameter estimate) under the null hypothesis (typically $\varphi = 0$), for any parameter with a sampling distribution.

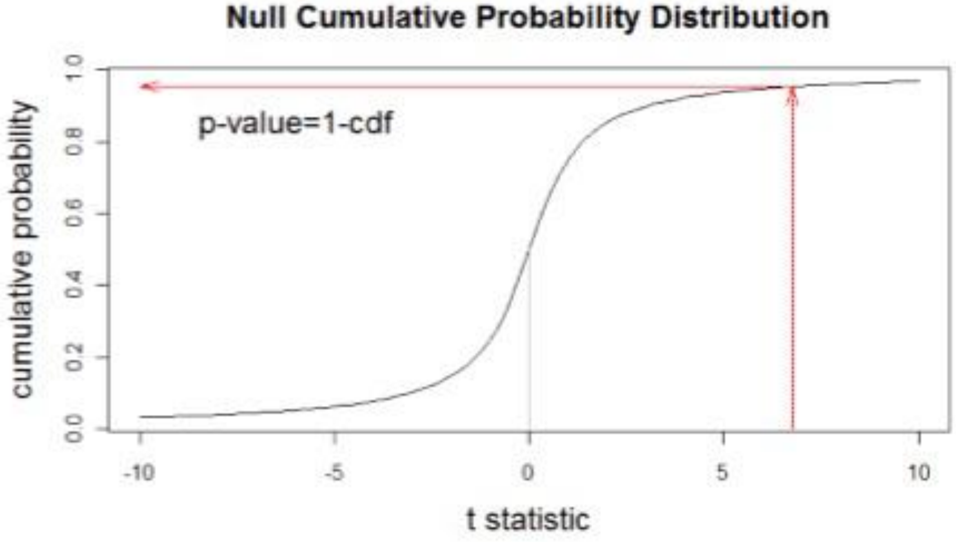


H_A: $\Pr(\varphi \neq 0)$

$$t_{\varphi} = \frac{\varphi_{obs} - \varphi_{null}}{SE_{\varphi}}$$

$$t_{\varphi} \sim t(\text{df}=1)$$



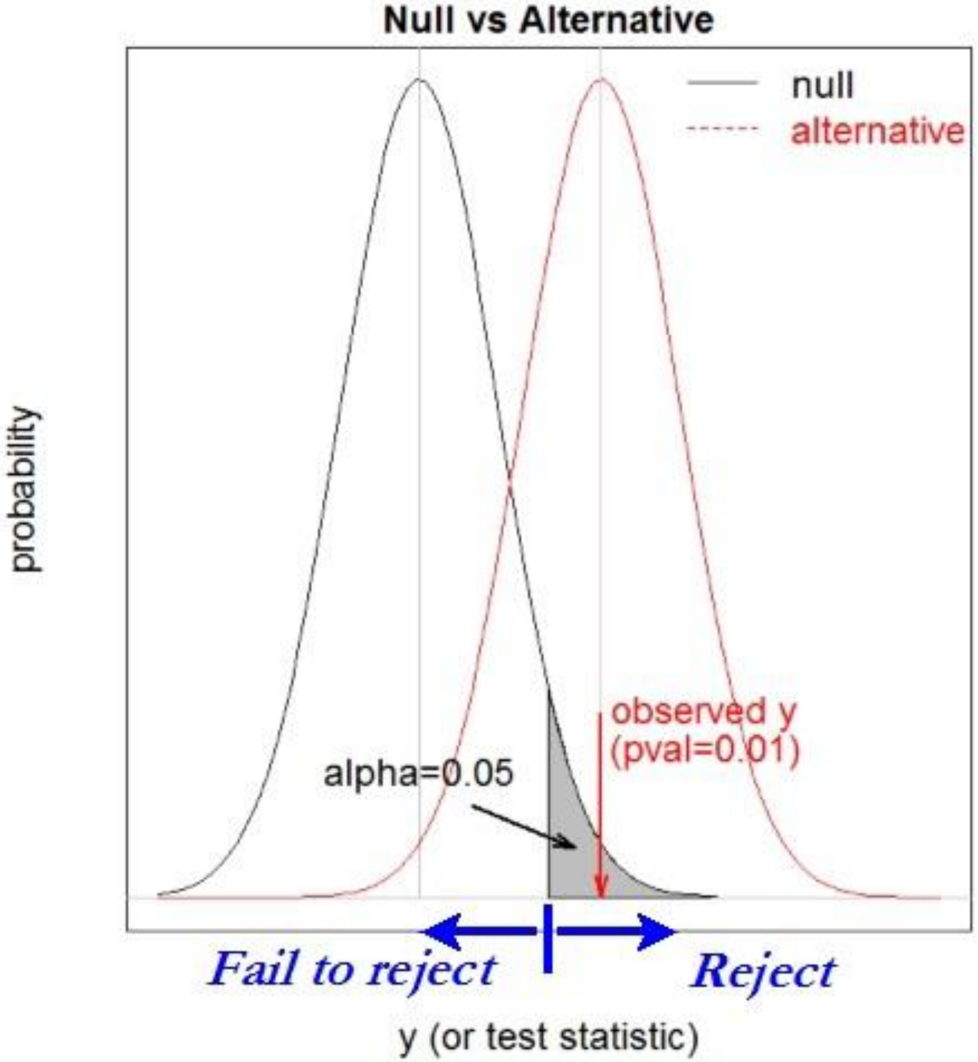


Hypothesis Testing Concepts

Neyman-Pearson decision framework

- *Reject* the null hypothesis if the *p*-value is less than a critical value (*alpha*), by convention usually ≤ 0.05
- *Fail to reject* the null hypothesis if the *p*-value is greater than *alpha* (i.e., there is insufficient evidence to disprove the null)

Remember, this applies to any probability distribution



False negatives: beta

Beta is the type II error rate: failing to reject a false null hypothesis.

We select a p-value cutoff ahead of time: alpha

The false negative rate depends on our choice of alpha and the data.

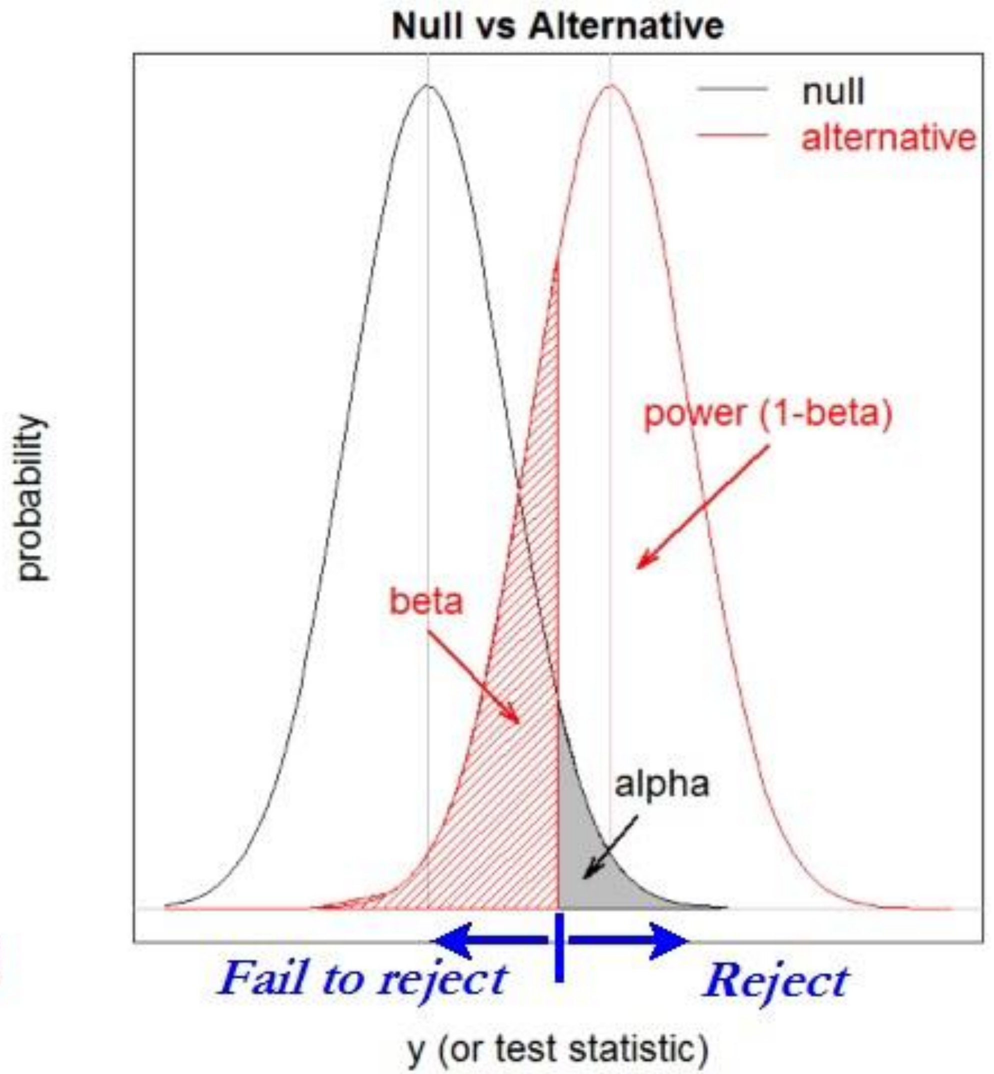
We cannot know beta until after we have collected data ☹

Hypothesis Testing Concepts

Neyman-Pearson decision framework

- *alpha* = probability of wrongly rejecting the null hypothesis (Type I error)
- *beta* = probability of wrongly accepting the null hypothesis (Type II error)
- *power* = probability of correctly rejecting the null hypothesis

alpha is under the null; *beta* and *power* are under the alternative



Power Analysis

Statistical Power: the probability that we **correctly reject** a **false null** hypothesis.

Statistical power is 1 – beta

We can’t know our statistical power until after we collect data…

Factors that influence statistical power

Sampling error, sample size

Population variability

Effect size

Our choice of alpha

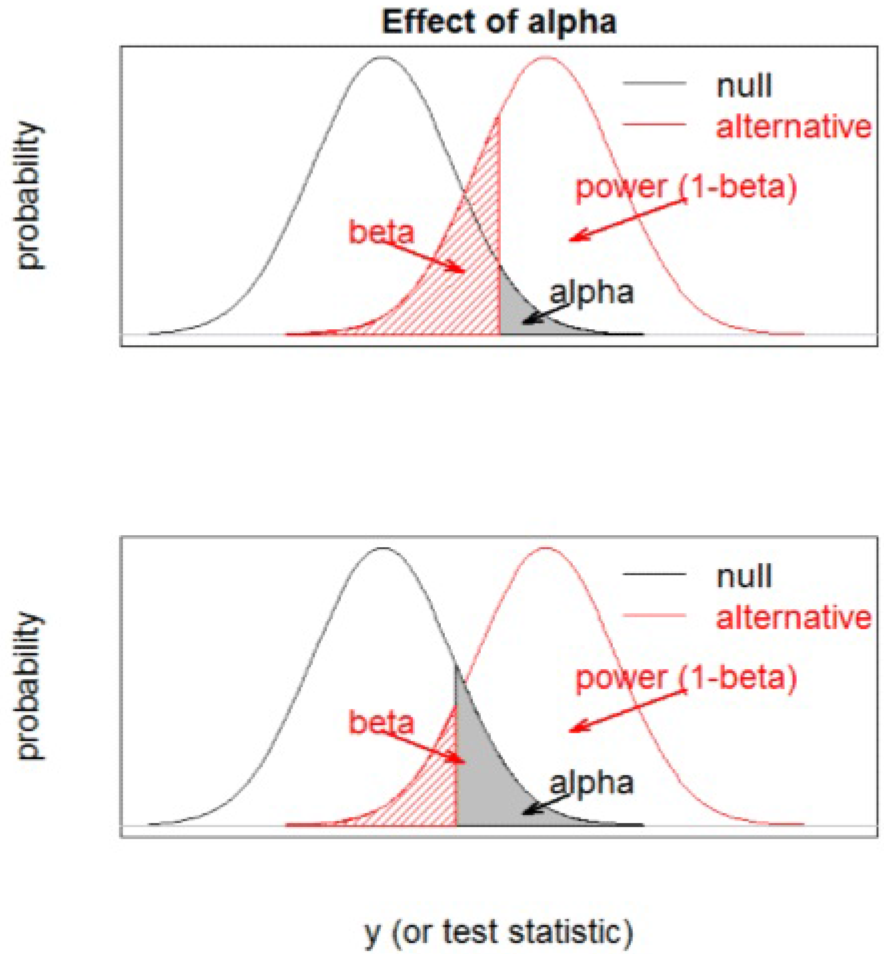
You cannot simultaneously decrease the false positive rate and increase statistical power!

Hypothesis Testing Concepts

Neyman-Pearson decision framework

Effect of alpha?

- Increasing alpha, increases power, all other things being equal

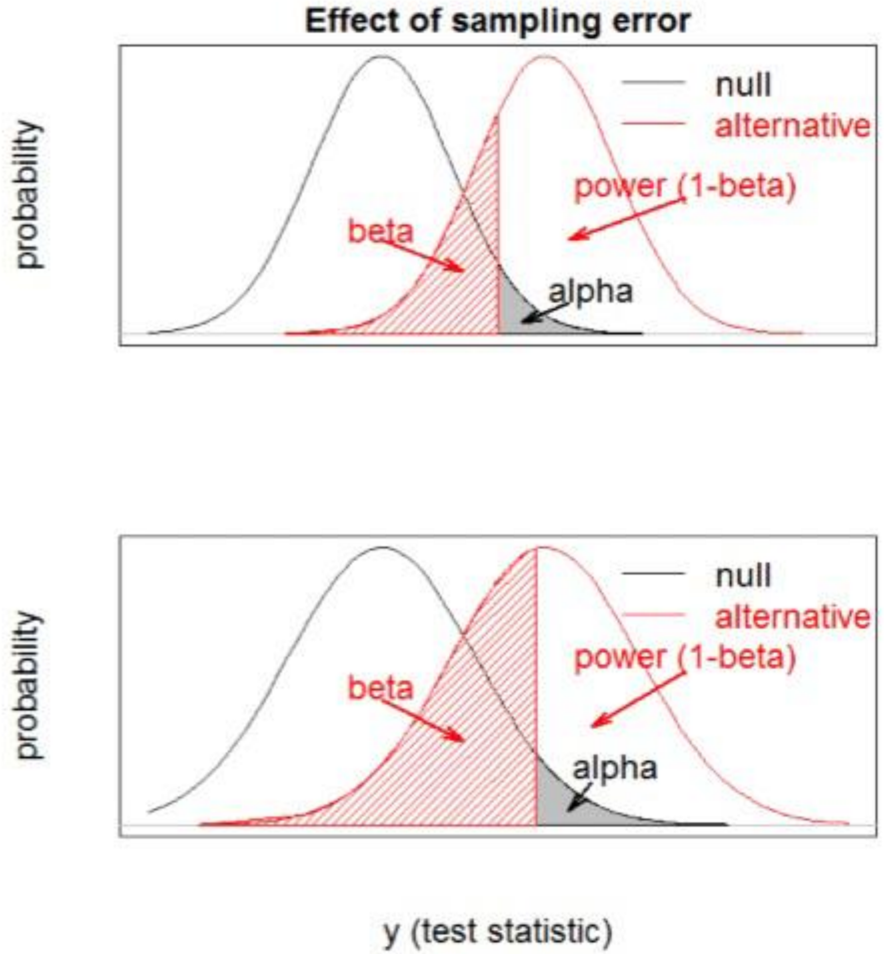


Hypothesis Testing Concepts

Neyman-Pearson decision framework

Effect of sampling variability (standard error)?

- Increasing sampling variability, either by increasing the variance in the underlying distribution or decreasing sample size (both effect sampling precision), decreases power, all other things being equal

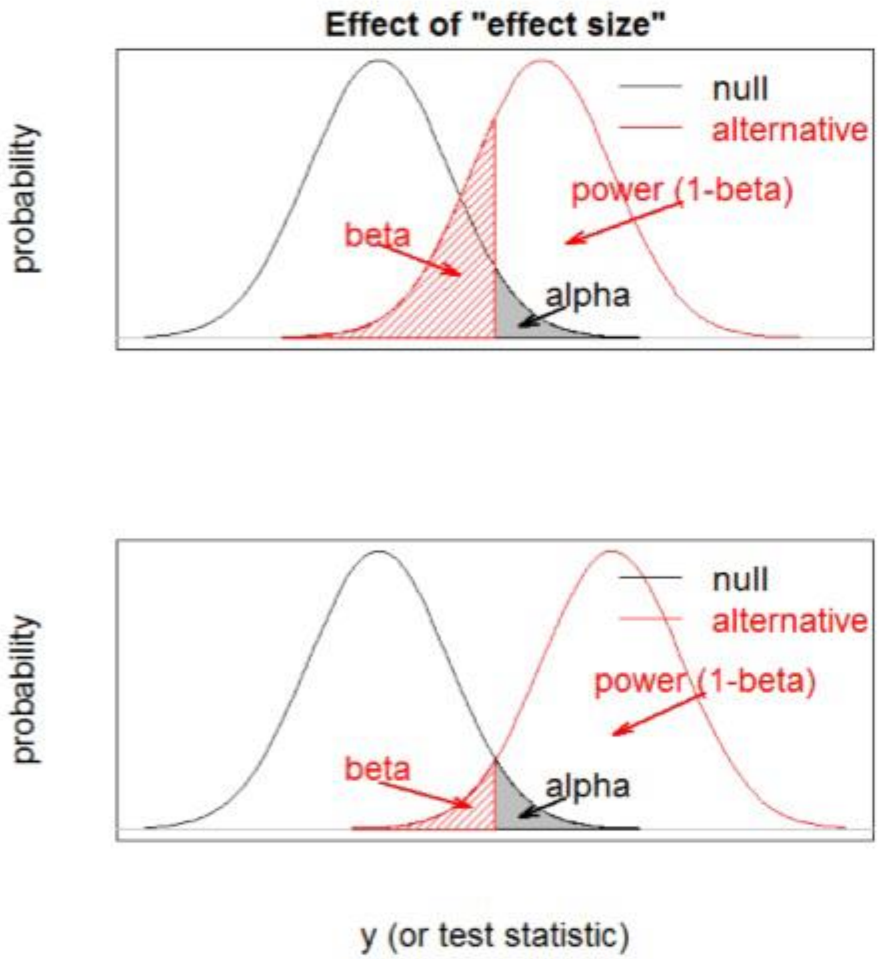


Hypothesis Testing Concepts

Neyman-Pearson decision framework

Effect of effect size?

- Increasing the effect size, increases power, all other things being equal



For next time

Finish hypothesis testing

Confidence intervals: McGarigal 6b