Probability and stochasticity

Today's agenda:

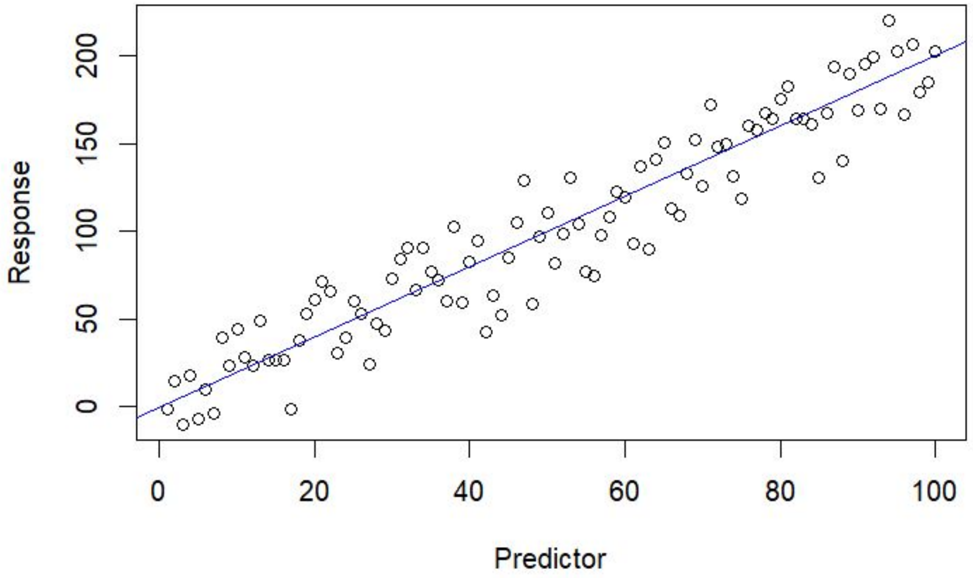
- Check-in
- Variability
- Probability theory
- Analyzing probability theory

Check-in

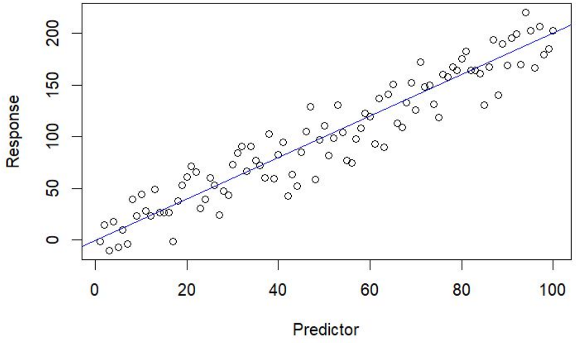
- Reminder: Assignment 1 was due last Friday
- Typo in lecture notes (notes said read chapter 4, syllabus said chapter 3)
  - If you’ve worked through chapter 3, great!
  - If you worked through chapter 4, work through chapter 3 when you get a chance (esp 3.4-3.6)

Variability

- Increasing recognition of importance biologically (e.g., intraspecific variation)
- Important component of statistical models (“signal” vs “noise”)



Variability

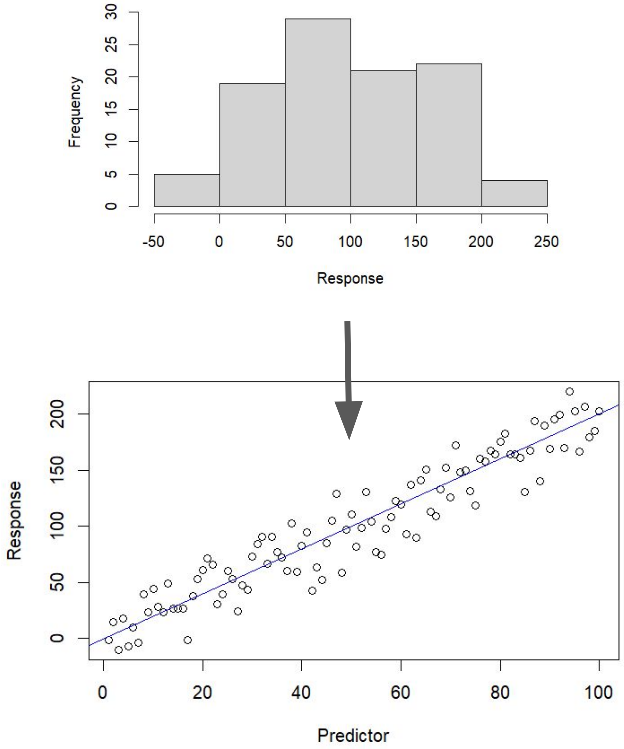


- Different sources of variability
  - Unmeasured variables (e.g., environmental stochasticity)
  - Measurement error
  - Inherent uncertainty
  - Demographic stochasticity (50% survival in a population of 3)
  - Nonlinear dynamics can increase/decrease variation

Variability

Different ways of handling variability

- Try to explain it vs. treat as error



Probability theory refresher

Probabilities sum to 1

Probability theory refresher

For mutually exclusive events: Probability A or B = P(A) + P(B)

P(Dead or alive) = P(Dead) + P(Alive) = 1

- If P(Dead) = 0.4, P(Alive) = 0.6

P(SIR) = P(S) + P(I) + P(R) = 1

- If P(S) = 0.1, and P(R) = 0.2, P(S or R) = 0.3
- Then P(I) = 0.7

Probability theory refresher

	Male	Female
Blue	20	30
Green	15	35

For non-mutually exclusive events,

Have to account for double counting:

P(Blue OR Male) =

P(Blue) + P(Male) - P(Blue AND Male) = 0.65

Corrects for double-counts

Probability theory refresher

Conditional probability: Probability of something, given something else

P(A|B) = P(A and B) / P(B)

P(Male|Blue) = P(Blue male) / P(Blue)

0.40 = 0.20/.50

	Male	Female
Blue	20	30
Green	15	35

Probability theory refresher

Independent variables: P(A and B) = P(A)*P(B)

If Sex and Color were independent, we'd expect:

P(Blue Male) = P(Blue) * P(Male)

0.175 = 0.5 * 0.35

0.20 != 0.175, measurement error? dependence?

	Male	Female
Blue	20	30
Green	15	35

Questions so far?

What to know about probability distributions

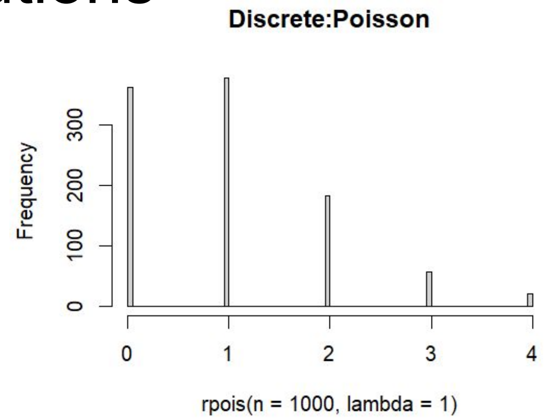
2 classes: Discrete vs continuous

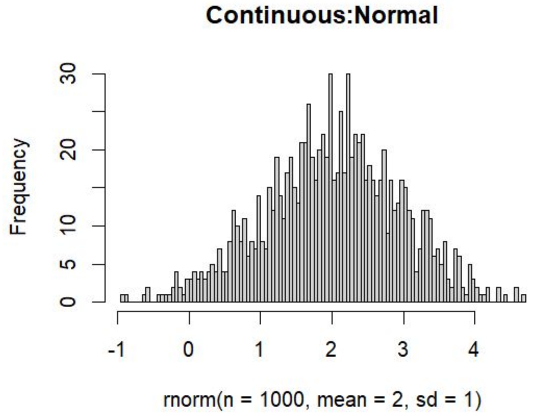
Discrete:

- **Integers**
- E.g., individuals, # events

Continuous:

- **Any real number** (or a subset, e.g. positives)
- E.g., body sizes, speeds, etc.
- P(any exact number) = 0
- Focus on P(some number or more extreme)
  - E.g. P(x >= 4)





Continuous or discrete?

Organisms

- Body Mass
- Age
- Leaf Area
- Population size
- Population density
- Species richness
- Functional diversity
- Individuals per transects
- Mean body mass

Environment

- Temperature
- Elevation
- Depth
- Hurricane severity

Evolution

- Base pair substitution
- Evolutionary rate
- Species per taxon

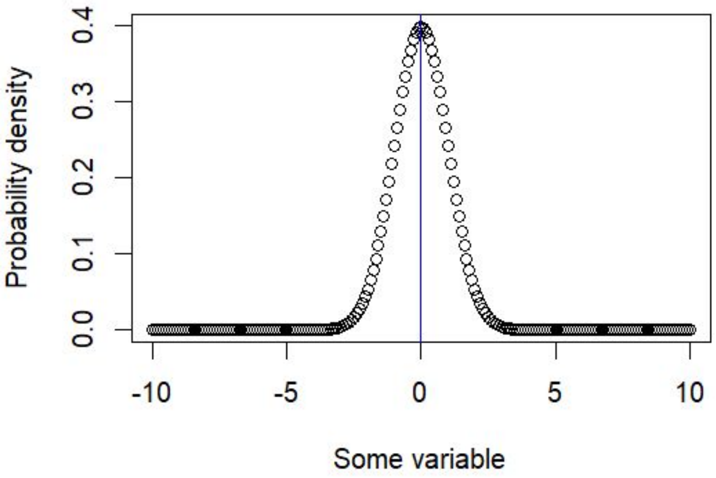
What to know about probability distributions

Central tendencies:

Mean

- "Average"
- Expected value (denoted *E*[*x*])
- Also denoted $\overline{x}$

Use function `mean()`

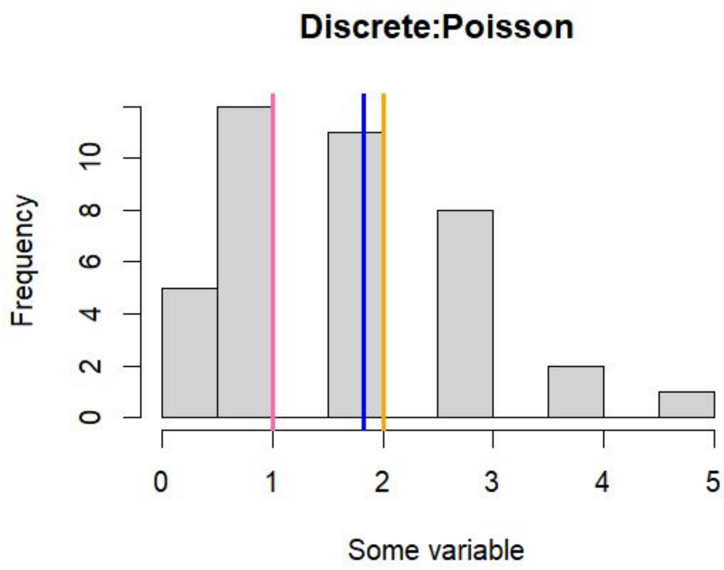


What to know about probability distributions

Central tendencies:

Median

- More robust than mean
- Less sensitive to outliers and skew
- Use function `median()`



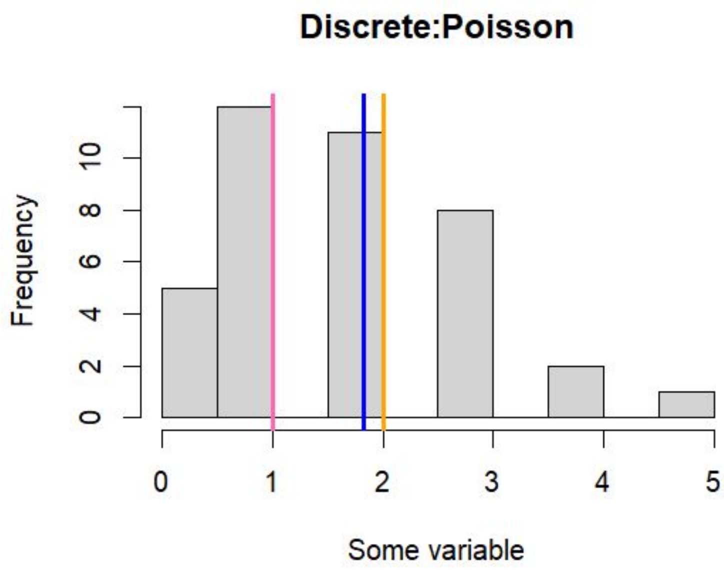
Blue = mean
Orange = median
Pink = mode

What to know about probability distributions

Central tendencies:

Mode

- Most likely or most common value
- No built-in function
- DescTools package has `Mode()`



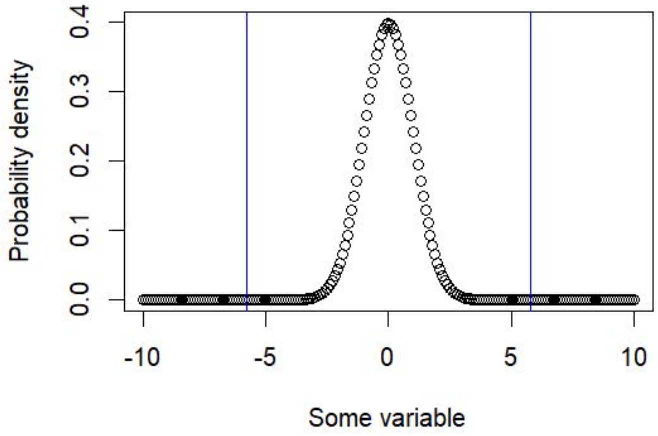
Blue = mean
Orange = median
Pink = mode

What to know about probability distributions

Spread:

Variance and standard deviation

- Variance:
  - SD^2
  - Additive
  - Units of mean squared
  - Function `var()`
- SD
  - √variance
  - Not additive
  - Same units as mean
  - Function `sd()`



Blue lines: Mean +/- SD

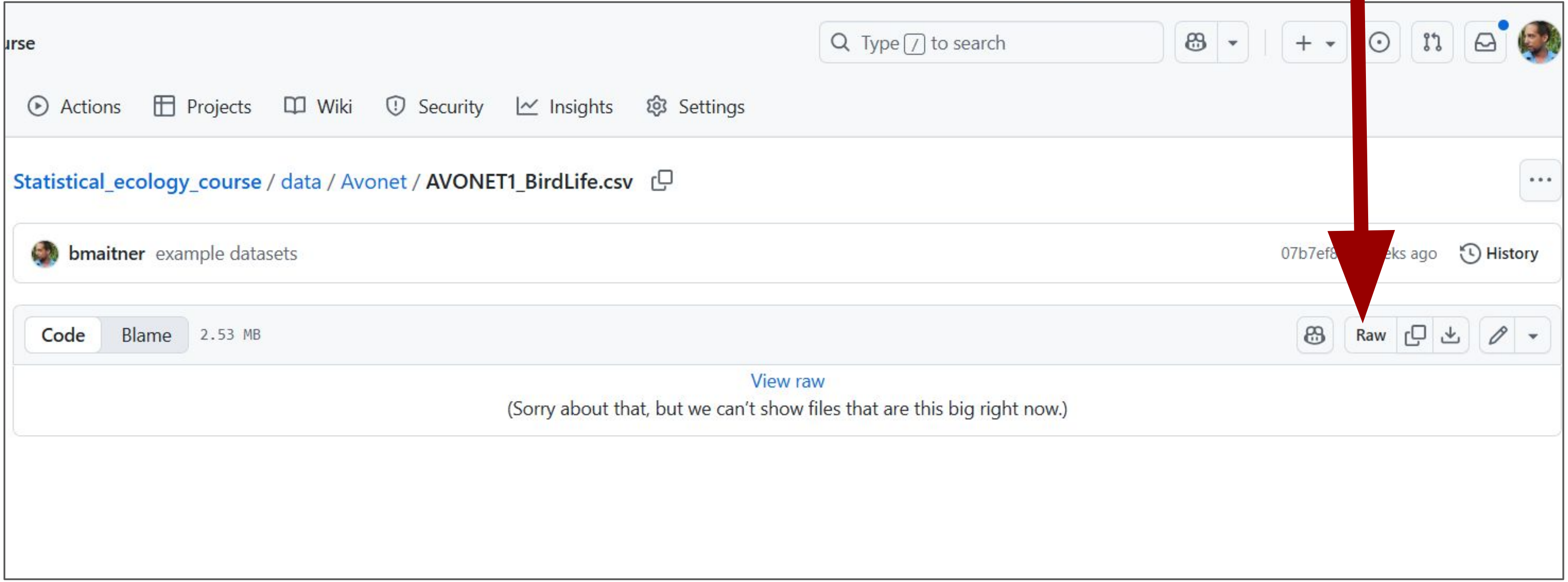
Activities

- Load data and calculate mean, variance, std dev, median, mode, etc.
    - Use data on github

Working with probability distributions

To get link:
Right click on “Raw”, copy the URL

Load the avonet data



Working with probability distributions

Load the avonet data

```
avonet <-
read.csv("https://github.com/bmaitner/Statistical_ecology_co
urse/raw/refs/heads/main/data/Avonet/AVONET1_BirdLife.csv")
```

Working with probability distributions

Using the avonet data:

1) Generate a histogram of Hand-Wing Index (high HWI = long-distance flyer)
    a) Use `hist()` function
2) Calculate mean and median Hand-Wing Index
    a) `mean()` and `median()`
3) Generate a histogram of Body Mass
    a) Also try generating a histogram of the log of body mass (use `log10()` or `log()`)
4) Calculate mean and median body mass
    a) How similar are they?
5) (optional) Try calculating the mode of body mass
    a) `Mode()` in the DescTools package

Working with probability distributions

Using the avonet data:

1) Calculate the variance and standard deviation of Mass
    a) `var()` and `sd()`
2) Check that variance equals SD^2
    a) Use "==" to test whether things are equal
    b) R uses "^" to denote raising something to a power, e.g., `x^2` mean x^2
    c) For square roots, use either `sqrt()` or `x^0.5`
3) Calculate the mean and variance of Range.Size

Remainder of class:

- Load in your focal data
- Use today's functions to better understand your key variables

Before next class:

- Look over 4.5 and 4.6