Assignment 2: Probability Distributions

# Overview

This assignment has two parts (full detail at the bottom). In the first part, I’ll give you a list of different types of data and you’ll have to decide what type of probability distribution they correspond to using the Bestiary of Distribution in your book (section 4.5). For the second part, you’ll use the same dataset from Assignment 1 and will decide which types of distributions your focal variables correspond to.

## Rationale

Understanding data distributions is important for data visualization and statistical analyses, as well as providing information on important biological processes. This assignment is designed to familiarize you with some of the more common types of statistical distributions in biology as well as getting you used to the process of trying to decide which is most appropriate for a given variable.

## Assignment Format

As before, this assignment can be submitted in one of two formats:

1. The output of an .Rmd file
2. See an example .Rmd file at: <https://github.com/bmaitner/Statistical_ecology_course/blob/main/example_R_markdown/Example.Rmd>
3. See example output at:

<https://github.com/bmaitner/Statistical_ecology_course/blob/main/example_R_markdown/Example.pdf>

<https://github.com/bmaitner/Statistical_ecology_course/raw/refs/heads/main/example_R_markdown/Example.docx>

<https://github.com/bmaitner/Statistical_ecology_course/raw/refs/heads/main/example_R_markdown/Example.html>

1. A fully-reproducible R script that can be run on any computer

a. See an example .R file at:

<https://github.com/bmaitner/Statistical_ecology_course/blob/main/example_R_commented/Example.R>

Assignments can be submitted in 2 ways:

1. Provide the file on canvas (e.g., .R or .html; undergrads only)
2. Upload the file to your own Github repository and then provide a link via Canvas (mandatory for graduate students, undergrads get extra credit).

## Grading

Students will be graded on both meeting the criteria outlined in the overview as well as the quality of their work.

*Grade Breakdown:*

* Part 1 answers 40%
* Part 2 accuracy and reasoning 40%
* Reproducibility 20%
* Uploaded to Github (Undergrads) +10%
* Not uploaded to Github (Grads) -10%

# Part 1: Hypothetical Distributions

For each of the following types of data, list which distribution type you think is most appropriate out of the options presented in Table 4.1 and Section 4.5 in your book. For each, briefly (one or two short sentences MAX) explain how you came to that conclusion. You can include these as either text in a .Rmd file output, or else just include them as comments in a .R file.

1. **Number of surviving individuals within a forest plot**
2. **Species abundance (counts of individuals in a plot)**
3. **Species richness (number of species in a sample)**
4. **Ages (in years) of surviving individuals within a reserve**
5. **Body size (e.g., body mass, length)**
6. **Population growth rate**
7. **Proportion of habitat covered by vegetation**
8. **Biomass per unit area**
9. **Seed counts per plant**
10. **Time to germination**
11. **Distance moved by an animal (step lengths in movement data)**
12. **Pollinator visitation rate (visits per flower per unit time)**
13. **Leaf area measurements**
14. **Environmental variables (temperature, rainfall, pH)**
15. **Prey consumed per day**
16. **Fish counts on reef transects**
17. **Proportion of coral cover on a reef**
18. **Larval settlement success (number settled out of larvae released)**
19. **Time to mortality under hypoxia experiments (fish, invertebrates)**
20. **Proportion of infected fish in a population (parasite prevalence)**

# Part 2: Focal Dataset Distributions

For your focal dataset (the one you used in Assignment 1), identify the numeric variables you’re interested in and explain which distribution you think is most appropriate. For this, you should provide evidence in the form of both histograms and brief, written arguments (written as comments if using .R files).