

# Biodiversity Dataset Description

Perry Williams, PhD

## 1 Guidelines for Evaluating Student Work

### 1. Develop Alternative Hypotheses:

- Students are required to develop at least 8 alternative hypotheses describing the relationship between the dependent variable (biodiversity) and the independent variables (distance to town, deforestation, soil pH, elevation, and rainfall). These hypotheses should explore various possible relationships, including linear, nonlinear, and interactive effects.

### 2. Fit Models to Data:

- For each hypothesis, students must fit a Bayesian statistical model to the dataset using either JAGS software or by implementing a custom Markov Chain Monte Carlo (MCMC) algorithm in R. The models should incorporate appropriate prior distributions and specify the likelihood function based on the chosen hypotheses.

### 3. Conduct Model Checking:

- After fitting the models, students should conduct model checking by assessing the goodness-of-fit using Bayesian p-values on the posterior predictive distribution. This step ensures that the fitted models adequately capture the variability in the observed data and provide reliable predictions.

### 4. Conduct Model Selection:

- Students are required to conduct model selection using the Deviance Information Criterion (DIC). DIC compares the goodness-of-fit and complexity of different models and helps identify the most parsimonious model that balances model fit and complexity.

### 5. Make Inference Using Top Model:

- Finally, students should make inference using their top-performing model. This involves interpreting the estimated parameters, assessing the uncertainty around parameter estimates, and drawing conclusions about the relationships between biodiversity and the independent variables based on the fitted model.

## 2 Description of Variables

### 1. Study Site ID (`site_id`):

- This variable represents the unique identifier for each study site within the rainforest. Each study site is assigned a specific ID number.

### 2. Biodiversity (`biodiversity`):

- Biodiversity refers to the number of different species within a given sample location. It is the dependent variable in this dataset. Biodiversity might be affected by other independent variables.

### 3. Distance to Town (`site_distance_to_town`):

- Distance to town represents the distance of each sample from the nearest town or human settlement in km. It is an independent variable in the dataset. Biodiversity is hypothesized to increase as distance to town increases.

### 4. Anthropogenic Effects (Deforestation) (`site_deforestation`):

- This variable represents the extent of anthropogenic effects, specifically deforestation, at each study site. It is as an independent variable in the dataset. It is an index between 0 and 1. 0 represents no deforestation. 1 represents total deforestation. Biodiversity is hypothesized to decrease with increasing deforestation.

### 5. Soil pH (`site_soil_ph`):

- Soil pH refers to the acidity or alkalinity of the soil at each study site. It is an independent variable in the dataset. Soil pH can influence the composition of plant species and consequently affect biodiversity.

### 6. Elevation (`site_elevation`):

- Elevation represents the height above sea level of each study site. It is as an independent variable in the dataset. Elevation can influence temperature, precipitation patterns, and habitat types, all of which can affect biodiversity.

### 7. Rainfall (`site_rainfall`):

- Rainfall represents the amount of precipitation received at each study site in millimeters (mm). It is an independent variable in the dataset. Rainfall is a critical factor influencing plant growth, habitat availability, and overall ecosystem dynamics, thereby affecting biodiversity.

### 3 Scientific Questions of Interest

#### 1. Impact of Anthropogenic Activities on Biodiversity:

- One key question of interest is to investigate how anthropogenic activities, such as deforestation, affect biodiversity within the rainforest. By analyzing the relationship between deforestation levels and biodiversity, we aim to understand the extent to which human activities contribute to changes in species richness and abundance.

#### 2. Influence of Environmental Factors on Biodiversity Patterns:

- Another important scientific question is to explore the role of environmental factors, including distance to town, soil pH, elevation, and rainfall, in shaping biodiversity patterns within the rainforest. By examining the relationships between these environmental variables and biodiversity, we seek to identify the primary drivers of species diversity and composition.

#### 3. Site-Level Variation in Biodiversity:

- Additionally, we are interested in examining the variation in biodiversity across different study sites within the rainforest. Samples within study sites may be more similar to each other because they are in close proximity. We want to account for this if necessary. By incorporating random effects for each study site, we aim to account for site-specific variability and explore how biodiversity varies spatially across the landscape.

#### 4. Interactions Between Environmental Factors and Anthropogenic Effects:

- Furthermore, we seek to investigate potential interactions between environmental factors and anthropogenic effects on biodiversity. For example, we may examine whether the influence of deforestation on biodiversity varies depending on the environmental context, such as soil pH or elevation.

#### 5. Conservation Implications:

- Finally, understanding the relationships between biodiversity and environmental factors, as well as the impact of human activities, has important conservation implications. By identifying key drivers of biodiversity loss or maintenance, we can inform conservation efforts and management strategies aimed at preserving the ecological integrity of the rainforest.