Agricultural Yield 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 (yield) and the independent variables (precipitation, amount of irrigation, and type of fertilizer). 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 yield and the independent variables based on the fitted model.

2 Description of Variables

1. **Farm ID (`farm_id`):**

   - This variable represents the unique identifier for each farm. Each farm is assigned a specific ID number.

2. **Yield (`yield`):**

   - Yield refers to the quantity of agricultural produce (e.g., crop yield) obtained from each farm. It is the dependent variable in this dataset. Yield may be affected by other independent variables.

3. **Precipitation (`precipitation`):**

   - Precipitation represents the amount of rainfall received by each farm during the growing season, measured in millimeters (mm). It is an independent variable in the dataset. Yield may be positively or negatively affected by precipitation, depending on various factors such as crop type and stage of growth.

4. **Amount of Irrigation (`irrigation`):**

   - Amount of irrigation denotes the volume of water applied to each farm through artificial irrigation methods during the growing season, measured in liters per hectare (L/ha). It is an independent variable in the dataset. Yield may be positively influenced by irrigation, especially in regions with insufficient rainfall or during dry periods.

5. **Type of Fertilizer (`fertilizer`):**

   - Type of fertilizer indicates the type of nutrients supplied to each farm's soil to enhance crop growth and yield. It is a categorical independent variable in the dataset, with two levels: organic and chemical. Yield may be affected differently depending on the type of fertilizer used.