STAT4116: Project Proposal

Jiazhe Lin (u7487349)

1 Introduction

The selective breeding of crops is an ancient practice that dates back to the emergence of human civilisation. Following the industrial revolution and the rising demand for agronomic production, modern plant breeding began to leverage statistical tools to analyse cultivars for desired traits. The abundance of data in agricultural statistics in turn gave rise to numerous fundamental statistical theories, such as the analysis of variance (ANOVA) (reference).

A key aspect of the plant breeding cycle involves comparing varieties across different trials, which makes it possible to assess plant performance under varying soil, water, and other environmental conditions. The goal of this aggregated analysis is, first, to distinguish true genotypic performance from environmental influences, and second, to evaluate the stability of genotype performance across conditions.

With the objective of multi-environment trials relatively straightforward, the key challenge lies in modeling the interaction between genotype and environment effects, a phenomenon referred to as the G $\times$ E effect. One of the modern approach is to model some trait response as a linear mixed model with fixed environmental effect and random genotype, design and interaction, as well as auto-correlated residual errors to account for spatial correlation.

```
## [1]   323 17317
```

```
##       Me                Mum                Dad                fgen
##  Length:9333        Length:9333        Length:9333        Min.   : 0.000
##  Class :character   Class :character   Class :character   1st Qu.: 1.000
##  Mode  :character   Mode  :character   Mode  :character   Median : 3.000
##                                                           Mean   : 2.851
##                                                           3rd Qu.: 4.000
##                                                           Max.   :10.000
```

```
##       plot           col            row            gen            env
##  1      :  36   1      : 792   1      : 432   G0008  : 126   E01    : 288
##  2      :  36   10     : 792   10     : 432   G0010  : 126   E02    : 288
##  3      :  36   11     : 792   11     : 432   G0324  : 126   E04    : 288
##  4      :  36   12     : 792   12     : 432   G0013  :  72   E05    : 288
##  5      :  36   2      : 792   13     : 432   G0009  :  63   E07    : 288
##  6      :  36   3      : 792   14     : 432   G0002  :  55   E08    : 288
##  (Other):9288   (Other):4752   (Other):6912   (Other):8936   (Other):7776
##      yield
##  Min.   :0.3134
##  1st Qu.:2.3147
##  Median :3.1950
##  Mean   :3.3570
##  3rd Qu.:4.3300
##  Max.   :7.6599
##  NA's   :31
```

```
## # A tibble: 36 x 2
##    env   n_plots
##    <fct>   <int>
##  1 E01       288
##  2 E02       288
##  3 E03       192
##  4 E04       288
##  5 E05       288
##  6 E06       192
##  7 E07       288
##  8 E08       288
##  9 E09       192
## 10 E10       288
## # i 26 more rows
```

2 Research Question

3 Proposed Methodology

In this assignment, we aim to provide comprehensive comparison between the estimation accuracy between the classical linear mixed model, which can be written as a Bayesian posterior model and the Bayesian Lasso method.

4 Reference