Masters Project Milestone Report

Name of the student: Jingning Yao

Name of the supervisor(s): Dr. Emi Tanaka and Dr. Katharine Turner

Title of the project (may be tentative): Opimization of Row-Column Designs

Summary:

Introduction

When testing theories about the relationship between two or more variables, an experiment allows researchers to observe the relationship between the variables in a controlled environment that better allows them to infer validity of their theory. Not all experiments, however, yield valid results. A well-designed experiments is needed to obtain valid and efficient results.

Row-column design is a special type of experimental design where the units have a rectangular array structure. This unit structure is common in the biological field, e.g. a crop field trial where the plots are laid out in a regular grid structure. The main idea of this design method lies in optimal allocation of experimental conditions (i.e. treatments) across two dimensions (namely rows and columns).

In designing experiments, we want good capability of achieving our aim, e.g. detecting any differences between treatments where there is a real difference. We also want to ensure these differences arise due to treatments instead of other causes.

Background

Completely randomised design. Completely randomized design is the simplest model for row-column design, which does not have any restriction on allocation, so treatments are randomly assigned to any position in experiment units. In this case, unexpected adjacency may occur, which may be undesirable and cause bias in the result (Piepho, Michel, and Williams 2018).

Randomised complete block design. We need stratum to build the designs. A stratum is a restriction imposed on the layout of design. Stratum helps to adjust and accommodate for field variation, and model the effect in a precise and theoretical way. Observations are usually modeled by linear model. A basic design widely discussed is randomized complete block design (RCB). The assumed model is followed.

$$Y_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij}$$

Here $i = 1, 2, 3, ..., r$ and r is the number of replicate. And $j = 1, 2, 3, ..., \nu$ where ν is the number of treatment. Y_{ij} is the observation of ith replicate, jth treatment;ρ are parameters for r replicates; τ_j are parameters for the ν treatments;μ is the overall mean and ϵ_{ij} is the the

error of the model.(Matheson, Harwood, and Williams 2024) RCB blocks every replicates and each blocks contain every kinds of treatment.

Placing all treatments in one replicate is possible when the number of treatment is small. When it come to larger number, larger number of plots in a replicate is required, in which case resource limitation should be considered and physical differences cannot be ignored. Incomplete block designs should be considered, which only some treatments are placed in one block.

Incomplete block design. In incomplete block design, for all treatment having same status, maximize the number of different pairwise comparisons in one block is the optimal design we wish to obtain. Incomplete block designs have one more block effect than RCB because blocks are different form each other. So the model is assumed as followed.

$$Y_{igh} = \mu + \rho_i + \beta_{ig} + \tau_j + \epsilon_{igh}$$

Here $Y_{igh}(i = 1, 2, \dots, r; g = 1, 2, \dots, s; h = 1, 2, \dots, k)$ refer to the observations come from r replicates with s incomplete blocks containing k plots.ϵ_{igh} is the the error of the model If we consider the design for every blocks, we can rewrite the incomplete block design model in a matrices-vectors form as:

$$Y = \mathbf{G}\mu + \mathbf{R}\rho + \mathbf{Z}\beta + \mathbf{X}\tau + \epsilon$$

Here R , Z and X are design matrices for replicates blocks and treatments. G is a ones vector. Y is the vector of all Y_{igh} likewise μ, ρ, β, τ and ϵ are vectors of parameters (Matheson, Harwood, and Williams 2024). Through this model we are trying to find a optimal design for a experiment.

A-optimality. A-optimality is a statistical criterion used in experimental design to evaluate and select the best experimental designs option. The A-optimality criterion focuses on the variances of the parameter estimates, mainly basing on variance-covariance matrix.(Jones, Allen-Moyer, and Goos 2021)

General Plan

- Conduct a method or algrithm for finding the optimal or near-optimal row-column design for a experiment through statistical criterion and mathematical derivation.
- Explore the area of optimizing row-column design based on A-optimality criterion with restriction of achieving the neighborhood balance and evenness of distribution. Specifically, I will review Piepho, Michel, and Williams (2018) Matheson, Harwood, and Williams (2024), Jones, Allen-Moyer, and Goos (2021) and other more related materials, and translate into a more mathematically rigorous framework.
- I will stimulate a experiment a through R code in Rstudio, take the algorithm into application, and verify conclusions I get.

References

Jones, Bradley, Katherine Allen-Moyer, and Peter Goos. 2021. "A-Optimal Versus d-Optimal Design of Screening Experiments." *Journal of Quality Technology* 53 (4): 369–82.

Matheson, AC, CE Harwood, and ER Williams. 2024. *Experimental Design and Analysis for Tree Improvement*. CSIRO PUBLISHING.

Piepho, Hans-Peter, Volker Michel, and Emlyn Williams. 2018. "Neighbor Balance and Evenness of Distribution of Treatment Replications in Row-Column Designs." *Biometrical Journal* 60 (6): 1172–89.

Student's signature **Supervisor's signature**

Date: 2024-03-08