Simulation study to ensure the objective function

The aim of this section is to perform a simulation study to verify the objective function that we derived is able us to obtain the optimal design effectively. This simulation study composed of first attempt to generate all essentially different Phase 2 designs of a given set of design parameters. For this case, we uses experiments with six runs, four tags, two trays, 12 plants and three experiments as an examples of this study. Once we have a list of designs, we will take a best design, the worst design and one design with an efficiency that in the middle. We will then compare three designs of their variances of the treatment comparison from 1000 sets of simulation datasets.

Phase 1 and Phase 2 experiments

Before we decide how to generate these essentially different designs, we need to understand the layout of the experimental structure of the Phase 1 and Phase 2 experiments. The Phase 1 experiment consists of three treatments assigned to 12 plants in total. Six of 12 plants are allocated to one of two trays. Each sample is then harvested from each of 12 plants and then further subdivided into two subsample. Thus, there are total of 24 subsamples from the Phase 1 experiment to allocated to the Phase 2 experiment. In the Phase 2 experiment, we uses 6 runs and 4-plex labelling system to measure the protein abundances of each subsample in the Phase 1 experiment.

Generate all essential designs

To allocate 24 subsamples of the Phase 1 experiment to the Phase 2 experiment, there can be 6.2 times 10^23 ways, which computationally unfeasible to complete. Thus, we need to come up with a more efficient way to generate all these different assignments.

Instead of determining how to allocate each subsample of Phase 1 experiment to the Phase 2 experiment, we can reduce down the number of essential allocations from the relationship between the experiment structures within each Phase of experiment. For example, each tray needs to contain 12 Plants, where four of these 12 Plants must be assigned to one of three treatment groups.

The assignment can be partitioned into three steps: first step decide how to assign two trays to runs and tags, second step is to decide the assignment of three treatment groups to 12 units of a single tray, and last step is to decide the allocation of Plant to each treatment.

In the first step, there are three ways to assign trays to runs and tags. The first way is to assign the trays where Tray effects are intentionally confounded with the Tag effects. The second way is to assign trays where Tray effects are intentionally confounded with the Runs effects. The third way is to assign trays where Trays effects are orthogonal to both Runs and Tray effects.

Once we have confirmed the allocation of trays to runs and tags, we can then decide how to allocate the treatments groups to 12 units of a single tray. Since each treatment group is replicated four times in a tray, there are 34650 unique assignments. However, there are total of two trays, so there are 600 million ways to allocate pick any two random assignments, out of 34650, for two different trays.

Furthermore, there is still one more step to complete the assignment, which is allocate two plants to four experimental units of each treatment groups. There can only be three ways, which are (A, A, D, D), (A, D, A, D) and (A, D, D, A), where ‘A’ and ‘D’ denotes Plant IDs in one of the treatment group.

For this study, we focuses on generating two million designs for each of three ways to assign trays to runs and tags: Tray effects are intentionally confounded with the Tag effects, Tray effects are intentionally confounded with the Runs effects, and Trays effects are orthogonal to both Runs and Tray effects. We will then compare the Phase 1 Plants average efficiency factors, treatment degrees of freedom, residual degrees of freedom and treatment average efficiency factors between different designs. Note that it takes about 2 hours to generate three million designs, however, we only store up to three million designs at a time in my machine.

From the two million designs generated, when trays are assigned to that Tray effects are intentionally confounded with the Tag effects, there are 292 designs with Phase 1 Plants average efficiency factor equals to one. Over of these 292 designs, 277 designs has treatment DF equals to 2 in the Between Plants Within Trays Within Runs stratum. From these 277 designs, there are 3 designs with residual DF in the Between Plants Within Trays Within Runs stratum equals to 6 as the other 274 designs has residual DF less than 6. These 3 designs has treatment average efficiency factors of 0.9375, 0.8571 and 0.8571. The design with treatment average efficiency factors of 0.9375 has the same property of the optimal design generated from the objective function.

From the two million designs generated, when trays are assigned to that Tray effects are intentionally confounded with the Run effects, there are 3304 designs with Phase 1 Plants average efficiency factor equals to one. There are 3179 out of these 3304 designs that has treatment DF of 2 in the Between Plants Within Trays Within Runs stratum. From these 3179 designs, there are 5 designs has residual DF in the Between Plants Within Trays Within Runs stratum equals to 5, which is the highest DF from these 3179 designs. These 5 designs has treatment average efficiency factors ranging from 0.8705 to 0.5455. In summary, the best design generated from this set has the residual DF of 5 and treatment average efficiency factor of 0.8705, which is not as good as the best design generated from the previous set, when trays are assigned to that Tray effects are intentionally confounded with the Tag effects.

From the two million designs generated, when trays are assigned to that Tray effects are intentionally confounded with the Run effects, there are 352 designs with Phase 1 Plants average efficiency factor equals to one. There are 313 out of these 352 designs that has treatment DF of 2 in the Between Plants Within Trays Within Runs stratum. From these 313 designs, there are 4 designs has highest residual DF of 5 in the Between Plants Within Trays Within Runs stratum. These 4 designs has treatment average efficiency factors ranging from 0.75 to 0.5237. Thus, the best design generated from this set has the residual DF of 5 and treatment average efficiency factor of 0.75, which is worse than best design generated from the previous two sets, where when trays are assigned to that Tray effects are intentionally confounded with the Tag effects and Run affects.

Using the best design generated from these three different sets, the next step is then to compare their variances of the treatment comparison from 1000 sets of simulation datasets.