```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

C2M4: Peer Reviewed Assignment

Outline:

The objectives for this assignment:

1. Get a better understanding of Experimental design patterns.
2. Prove some of the background intuition in blocking and interblock interactions.
3. Understand how and when to apply different model strucutres for different experimental designs.

General tips:

1. Read the questions carefully to understand what is being asked.
2. This work will be reviewed by another human, so make sure that you are clear and concise in what your explanations and answers.

Problem 1: Experimental Design

The goal of this problem is to get you thinking about how experiments are designed and how data is collected, because those influence what models we end up using.

1. (a)

In your own words, define experimental design. Describe some negative effects of making an incorrect experimental design decision.

An *experimental design* is a study carefully designed so that researchers can maintain control over a "treatment", variable, e.g., the administering of a medication, and observe changes in a response variable, e.g., some measure of wellness. We might contrast this with an observational study, in which researchers cannot control the administration of any treatments on units in a sample. Instead, researchers collect and analyze data without changing existing conditions.

An incorrect experimental design can result in incorrect conclusions about the relationships between treatment and response. This is especially important because the goal of experimental designs are used to understand causal relationships.

1. (b)

In your own words, describe the difference between an experimental unit and a sample unit. Why does this distinction matter?

An *experimental unit* is an entity to which a treatment is applied. If we independently assign the treatments to several experimental units, we have performed a *replication*.

Importantly, an *experimental unit* may not be the same as a *sampling unit*. The latter is an entity on which the response is measured.

To sharpen this distinction, consider the following example.

Company X performs an experiment to compare effects of four different advertising campaigns for improving click rates. Four social media platforms are chosen to host the ad campaigns, each of which will show the add to 1000 internet users. Experimenters randomly assign each campaign to a different social media platforms.

In this example, the treatment is the advertising campaign. But what are the experimental units? In this case, they are not users, but social media platforms, since social media platforms are the entities to which treatments are applied. Users are then the sampling units.

This distinction might seem trivial, but we will learn that it is not. For example, if you thought that the experimental units were users, then you would falsely believe that you had many replications of each ad campaign, when in fact, you do not have replication, because there are four campaigns, and for platforms. In a future lesson, we will learn how blocking might belp us with replication.

Problem 2: Proving the Intuition

Show that, for the randomized complete block design:

$$SS_{total} = SS_{treat} + SS_{block} + SS_R$$

$$
\begin{aligned}
TSS &= \sum_{i=1}^{t}\sum_{j=1}^{r}\left(Y_{i,j}-\bar{Y}\right)^2 \\
&= \sum_{i=1}^{t}\sum_{j=1}^{r}\left(Y_{i,j}+\bar{Y}_{i\cdot}-\bar{Y}_{i\cdot}+\bar{Y}_{\cdot j}-\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot}+\bar{Y}_{\cdot\cdot}-\bar{Y}_{\cdot\cdot}\right)^2 \\
&= \sum_{i=1}^{t}\sum_{j=1}^{r}\left[\left(Y_{i,j}-\bar{Y}_{i\cdot}-\bar{Y}_{\cdot j}+\bar{Y}_{\cdot\cdot}\right)+\left(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot}\right)+\left(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}\right)\right]^2 \\
&= \sum_{i=1}^{t}\sum_{j=1}^{r}\left(Y_{i,j}-\bar{Y}_{i\cdot}-\bar{Y}_{\cdot j}+\bar{Y}_{\cdot\cdot}\right)^2+\sum_{j=1}^{r}t\left(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot}\right)^2+\sum_{i=1}^{t}r\left(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}\right)^2 \\
&= \ldots = SS_R + SS_{block} + SS_{treat}
\end{aligned}
$$

the "**...**" term means showing that the "cross terms" are all equal to zero. For example:

$$
\sum_{i=1}^{t}\sum_{j=1}^{r}\left(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot}\right)\left(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}\right)=\sum_{i=1}^{t}\left(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}\right)\sum_{j=1}^{r}\left(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot}\right)=
$$

$$
\sum_{i=1}^{t}\left(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}\right)\left[n\bar{Y}_{\cdot\cdot}-n\bar{Y}_{\cdot\cdot}\right]=0
$$

Problem 3: Interblock Interactions

Describe why, in a randomized complete block design (RCBD), it is not possible to test whether interactions exist between the treatment and blocks.

The RCBD model is

$$
Y_{i,j}=\mu+\tau_i+\beta_j+\varepsilon_{i,j},
$$

for $i=1,\ldots,t$ and $j=1,\ldots,r$. This model has $1+t-1+r-1=t+r-1$ parameters. If we added an interaction term, we would need to estimate a parameter for every interaction between the ith treatment and the jth block. This would require $n=tr$ parameters, which is the number of data points. So, we wouldn't have the data to estimate standard errors or conduct hypothesis tests.

Problem 4: 99 Designs for 99 Problems

For each of the following design patterns, give an example (that wasn't given in our video lessons) for an experiment that would best lend itself to the specified design pattern. Make sure to explain why the specified design is more applicable for your experiment than the other design patterns.

1. Complete Randomized Design (CRD)
2. Complete Randomized Block Design (CRBD)
3. Factorial Design

Various answers given in lecture videos.

```
In [ ]:
```