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

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.

Solution
Experimental design is the structure at which to go about identifying causal relatioships between variables. It is the set of procedures to test a hypothesis.
An incorrect experimental design leads to type 1 and type 2 errors (false positives and negatives), inappropriate conclusions of causal relationships between variables, and ethical issues. Without an accurate and appropriate experimental design, hypothesis tests will result in false conclusions or even become impossible to carry out. For example, one important issue regarding experimental design is randomization, which is the assignment of treatments to experimental units. When this is not done correctly, experimental unit/samples and the overall nature of the experiment will become biased and thus produce faulty results. Moreover, identifying and deciding on the appropriate control, treatment, and experimental groups/units are a major aspect of creating a good experimental design. Deriving correlations between variables will become confusing and difficult upon failure to do so.

1. (b)

In your own words, describe the difference between an experimental unit and a treatment unit. Why

does this distinction matter?

Solution
An experimental unit is the unit which treatments are applied to. Treatment unit is often referred to as factors, which is an entity that is applied to control and experimental units in order to achieve effects necessary for hypothesis testing. It is important to distinguish between the two because one is being applied to the other, and confusion over these terms will result in a poor/faulty experimental design. Another important distiction is the difference between an experimental unit and a sampling unit. Using the example taken in one of the lecture videos, the different types of social media platforms receiving the ads is the experimental unit, whereas the users on the platforms are the sampling units. In other words, sampling units are where the response is measured from.

Problem 2: Proving the Intuition

Show that, for the randomized complete block design: $SS_{total} = SS_{treat} + SS_{block} + SS_R$

Solution
The total sum of squares SS_{total} is derived in the steps below:

$$SS_{total} = a\sum i=1 b\sum j=1 (y_{i,j} - \bar{y}_{..})^2$$
$$= a\sum i=1 b\sum j=1 (\bar{y}_{i.} - \bar{y}_{..})^2 + b\sum j=1 a\sum i=1 (\bar{y}_{.j} - \bar{y}_{..})^2 + a\sum i=1 b\sum j=1 (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$
$$= ba\sum i=1 (\bar{y}_{i.} - \bar{y}_{..})^2 + ab\sum j=1 (\bar{y}_{.j} - \bar{y}_{..})^2 + a\sum i=1 b\sum j=1 (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$
$$= SS_{Trt} + SS_{Block} + SS_R = SS_{Total}$$

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.

Solution
The RCBD can only be used under the fact that there is one observation for each block-treatment combination. Testing interactions between the treatment and blocks would result in a situation similar to one of the assignment problems, where there are 0 degrees of freedom for the error and not enough data points to determine the other parameters.

Problem 4: 99 Designs for 99 Problems

For each of the following design patterns, give an example (that wasn't given in the videos) 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

Solution

1. CRD: Laboratory experiments. CRD is best used for experiments with homogeneous experimental units. A laboratory setting makes controlling environmental effects fairly easy. Other designs do not offer such advantage.
2. RCBD: Agricultural experiments. Naturally, agricultural experiments have similar experimental units grouped into replicates/blocks. Other designs do not have this property, making it difficult to use for agricultural experiments.
3. Factorial Design: Social experiments. Factorial design, as the name suggests, has the advantage of being able to involve multiple independent variables, or factors, in the experiment. A social experiment, such as standardized test scores based on gender/weather/location, involves multiple factors that may contribute to the response measurement. The CRD and CRBD is not suitable for such experiments involving multiple factors.

In [ ]: