MAT 3378 (Winter 2018)
Assignment 1
***Deadline:* Tuesday, January 30, 2018 (in class)**
There are a total of 5 questions.

1. A study published in the New England Journal of Medicine (august 2001) suggested that it is more dangerous to enter an hospital during the weekend. In a period of 10 years, researchers followed more than 4 million admissions to the emergency room in hospital in Ontario. Their results revealed that patients admitted on weekends have a much higher risk of death than those admitted to the emergency room from Monday to Friday.

   (a) Is it an observational study, an experimental study, or a mixed study ? (Why ?)

   (b) If you think you are very sick on a Saturday, should you wait until Monday to ask for medical help ? Explain.

   (c) Suggest possible explanations for these disturbing results.

2. Three treatments T_1, T_2, T_3 are to be studied in an experiment with a completely randomized design with $n_T = 9$ experimental units. Assuming a **balanced** design, use `R` to randomly allocate the experimental units to the 3 treatment groups.

3. Penicillin is made from Penicillium chrysogenum. A group responsible for its production look for the most favorable nutritive medium for the rapid development of their stock. They want to use a complete factorial design to study the influence of five factors (each with levels). They have 32 tests to realize. The objective of the study is to define the composition of the nutrient solution for the best performance in penicillin. The 5 factors, that is, the factors whose levels are monitored during the experiment, are the concentrations of the following products:

   - factor 1: corn liquor;
   - factor 2: lactose;
   - factor 3: precursor;
   - factor 4: sodium nitrate;
   - factor 5: glucose.

   The duration of the tests and the material constraints do not allow them to perform all the trials at one time. They must operate in two stages. They perform a first campaign of 16 tests, make a mandatory break of several weeks, and then carry out a second campaign of 16 tests.

   The problem is that between the two campaigns, many uncontrolled factors could change levels: the strain, the temperature, operators, etc. To attempt to control for

these different conditions, they include a factor called campaign with 2 levels (campaign 1, campaign 2).

(a) If we run a single campaign of 32 tests and assuming that randomization is used for the assignment units to the treatments, what type experimental design is used here?

(b) Suppose there are two campaigns of 16 tests. The campaign factor defines treatments or blocks? (Explain)

(c) Suppose there are two campaigns of 16 tests. What kind of experimental design is used here?

4. Exercise 15.12 from the textbook.

5. Consider a study where we investigate the effect of a type of golf ball on the distance. We have $n_T = 20$ golfers and 2 treatments (i.e. 2 types of golf balls). We are considering two type of designs:

   Design I: Each golfer is randomly assigned a type of golf ball. Half of the golfers are assigned to type 1 and the remaining golfers will use the type 2 golf balls. We ask the golfers to hit a few golf balls and we compute the average distance of these hits.
   Design II: Each golfer will hit each type of ball a few times and we will compute the average distance for each type of golf ball. We randomize to order of the balls within each golfer.

   (a) Give the name of the two experimental designs.

   (b) If you were to choose between these two designs, which would you prefer? (Why?)

   (c) Let us consider the design II. Denote Y_{ij} as the average distance for golfer i with golf ball type j, for $i = 1, \ldots, 20$ and $j = 1, 2$. Since the two observation Y_{i1} and Y_{i2} are from the same golfer, then we should consider them as correlated. We will assume the following:

   $$\rho\{Y_{i1}, Y_{i2}\} = \rho, \text{ for } i = 1, \ldots, 20,$$

   where ρ is some positive constant. It is reasonable to assume that the two measurements are going to be **positively** correlated, since we expect that a good golfer will have two long distances, and a weaker golfer would have two short distances.

   Assume that $\sigma^2\{Y_{ij}\} = \sigma^2$ for all $i = 1, \ldots, 20$ and $j = 1, 2$. To compare the effect of the type of golf ball on the distance, for golfer i, we will compute $Y_{i1} - Y_{i2}$. Show that

   $$\sigma^2\{Y_{i1} - Y_{i2}\} = 2\,\sigma^2\,(1 - \rho).$$

   Hint: Use properties of the covariance operator.

Remark: We see here that if the observations within the subject are highly correlated, the variance of the difference can be very small. This is the power of blocking at work.