

Student name: Quynh Bui

Student number: 7731733

Assignment 1

Question 1:

a.

This is an observational study because the researchers did not assign patients to the emergency room, but they followed the emergency room for 10 years and recorded the admissions.

b.

Even though through the study it seems that patients admits on the weekends have a much higher risk of death than those who admits to the emergency room from Monday to Friday, you should not wait until Monday if you are very sick. The chances of having higher risk of death on weekends might be because of many other reasons like:

- since it's the weekend people party and drink more than during the week.
- more people going out on the weekend might causes more accidents than during weekday. etc

Therefore if you wait until Monday, the chance of your sickness getting worse might be very high. c.

This result is called Simpson's Paradox. The explanation is from part b.

Question 2:

```
experimental_unit <- 1:9
treatment <- rep(c("T1", "T2", "T3"), 3)
data <- data.frame(sample(experimental_unit), treatment)
indexing<- function(tr)
{
  indices<-numeric(length(tr))
  unique.values<-unique(tr)
  for (k in unique.values)
  {
    count<-1
    for (i in which(tr==k))
    {
      indices[i]<-count
      count<-count+1
    }
  }
  return(indices)
}
```

```

indices <- indexing(treatment)
data <- data.frame(sample(experimental_unit), treatment, indices)
data <- reshape(data, idvar="treatment", timevar="indices", direction="wide")
data

```

	treatment	sample.experimental_unit..1	sample.experimental_unit..2	sample.experimental_unit..3
1	T1	2	8	9
2	T2	1	6	3
3	T3	4	5	7

Question 3:

- a.
If we run a campaign of 32 tests and assume that randomization is used for the assignment units to the treatments, then it's a Complete Randomized Design.
- b.
The 2 campaigns are blocking because since the experimental units are not homogeneous and we are dividing the experimental into 2 groups with each of 16 tests.
- c.
Suppose there are 2 campaign of 16 tests, this is called a Complete Block Design.

Question 4:

- a.
This is a mixed study design. We assign the version of the photograph to each officer. However we can not assign the gender to each officer.
- b.
Factor 1: Eye contact, 2 levels (Yes, No) Factor 2: Gender, 2 levels (Male, Female) There are 4 factorial level combinations.
- c.
This is an incomplete blocking design.
- d.
Basic unit is the experimental unit is the officer.

Question 5:

- a.
Design 1: This is an Incomplete Blocking Design
Design 2: This is a Complete Randomized Design
- b.
I would pick design 2 because it's more accurate to calculate the average the distance of the golf balls in the sense that each golfer can try both type of golf ball.
- c.
We know that $\rho[Y_{i1} - Y_{i2}] = \frac{\sigma^2[Y_{i1} - Y_{i2}]}{\sigma^2}$, hence $\sigma[Y_{i1} - Y_{i2}] = \rho[Y_{i1} - Y_{i2}]\sigma^2$
We can use properties of variance to show that:

$$\begin{aligned} & \sigma^2[Y_{i1} - Y_{i2}] \\ &= \sigma[Y_{i1} - Y_{i2}, Y_{i1} - Y_{i2}] \\ &= \sigma^2[Y_{i1}] - \sigma[Y_{i1}, Y_{i2}] - \sigma[Y_{i2} - Y_{i1}] + \sigma^2[Y_{i2}] \end{aligned}$$

$$\begin{aligned} &= 2\sigma^2 - 2\sigma^2[Y_{i1} - Y_{i2}] \\ &= 2\sigma^2 - 2\rho\sigma^2 \\ &= 2\sigma^2(1 - \rho) \end{aligned}$$