# STAT 523 HW2

*Md Muhtasim Billah*

*1/29/2020*

## Problem 1.

You wish to compare four treatments for effectiveness in preventing flu: (1) a flu vaccine, (2) 1 gram of vitamin C per day, (3) a turmeric curcumin complex pill, and (4) a placebo taken daily. (A placebo is a dummy pill which contains no active ingredient and should have no physical effect.)
A. In clinical experiments involving people, it is strongly recommended that a placebo be included among the treatments. Explain why.
B. What is an appropriate response variable?
C. Suppose that 800 volunteer subjects are enrolled for the study. Draw a diagram showing a design of your experiment. Clearly indicate where randomization, replication and control are applied in the experiment.

## Answer:

### A.

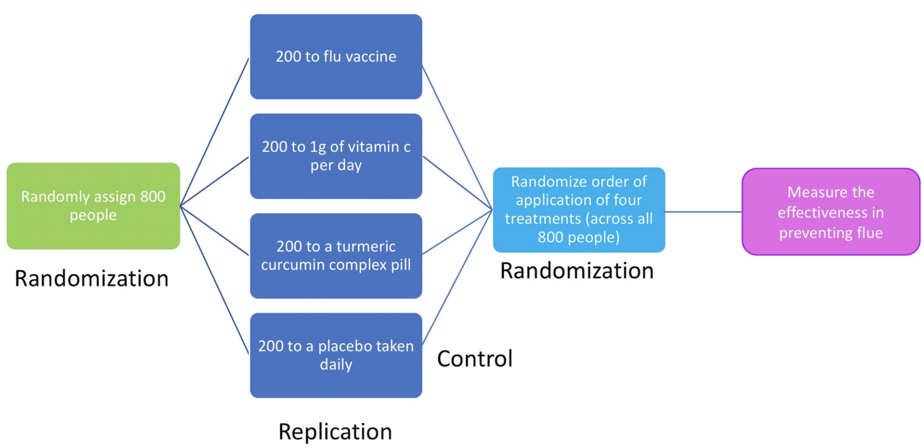
In clinical experiments involving people, it is strongly recommended that a placebo be included among the treatments. It is because, the people who are given placebo, work as the control group for such experiments. Having a control group is essential and one of the principles of effective experimentation. The experimental units (people in this case) who are given the placebos, they are actually not treated, which provides reference for comparisons among the other treatments. It also allows to control the effects of lurking variables.

### B.

An appropriate response variable for this experiment would be the effictiveness in preventing flue.

### C.

Diagram for designing the experiment is given below.



# Problem 2.

A chemical engineer is designing the production process for a new product. The chemical reaction that produces the product may have higher or lower yield, depending on the temperature and the stirring rate in the vessel in which the reaction takes place. The engineer decides to investigate the effects of combinations of two temperatures (55°C, 75°C) and four stirring rates (60 rpm, 80 rpm, 100 rpm, 120 rpm) on the yield of the process. She will process two batches of the product at each combination of temperature and stirring rate.
a. What are the experimental units and how many in total are required?
b. What are the treatments and how many are there?
c. Identify the control group, if there is one.
d. What is the response variable?
e. Identify 2 stages in the experiment where the principle of randomization should be applied. Describe how it is implemented in these stages, respectively.

# Answer:

## a.

The experimental units (EUs) are the batches of the product. There will be $4 \times 2 = 8$ combinations for the two temperatures and four stirring rates. Since two batches of the product will be processed for each combination, the total number of EUs will be $8 \times 2 = 16$.

## b.

The treatments are the combinations of the different temperatures and stirring rates. Since their are 8 combinations possible, the number of treatments will be 8. The 8 treatments are given in the following table.

| Treatments | Temperatures ($°C$) | Stirring rate (rpm) |
|---|---|---|
| Comb. 1 | 55 | 60 |
| Comb. 2 | 55 | 80 |
| Comb. 3 | 55 | 100 |
| Comb. 4 | 55 | 120 |
| Comb. 5 | 75 | 60 |
| Comb. 6 | 75 | 80 |
| Comb. 7 | 75 | 100 |
| Comb. 8 | 75 | 120 |

## c.

There is no control group for this experiment.

## d.

The response varible is the yeild of the process.

## e.

The randomization can be applied in two stages. In the first stage, the 16 batches (EUs) would be randomly assigned to eight combinations (treatments). Since, there are two batches (replicates) for each combination (treatment), these batches will be selected randomly and that would be the second stage of randomization.

# Problem 3

In a study, mosquitoes were divided into four groups of seven mosquitoes each: infected rhesus and sporozites present (Trt 1), infective rhesus and oocysts present (Trt 2), infective rhesus and no infection developed (Trt 3), and noninfective (Trt 4). Distances flown by the mosquitoes within 24 hours were recorded. The summary data are:

$$\bar{x}_1 = 3.7453, \bar{x}_2 = 4.7953, \bar{x}_3 = 5.5472, \bar{x}_4 = 4.8658, \sum\sum {x_{ij}}^2 = 683.3276$$

Use the ANOVA F test at level $\alpha = 0.10$ to decide whether there are any significant differences between true average flight times for the four treatments.
A. Show hand calculations for SST, SSTr, and SSE (Formulas 10B, C, D and that $x_i = \bar{x_i} * J$)
B. Follow the 4-step hypothesis test outline done in class (i.e. 1 - hypothesis/significance level, 2 - test statistic, 3 - rejection region, 4 - conclusion). You can also use the p-value approach in step 3 with R command 1 − pf(F, df1, df2) where F = test statistic, df1= treatment df, and df2 = MSE df.

# Answer:

## A.

Here,
Number of treatments, $I = 4$.
Number of replicates in each treatment group, $J = 7$.
Level of significance, $\alpha = 0.10$.

Now,
Total value for each treatment can be calculated with the given formula:

$$x_i = \bar{x_i} * J$$

So,
Total flight distance for the four treatments will be as below:

$$x_1 = \bar{x_{1.}} * J = 3.7453 * 7$$

$$x_2 = \bar{x_{2.}} * J = 4.7953 * 7$$

$$x_3 = \bar{x_{3.}} * J = 5.5472 * 7$$

$$x_4 = \bar{x_{4.}} * J = 4.8658 * 7$$

Calculating the values using R, we get the total flight distance for the four treatments as:

$$x_{1.} = 26.2171, x_{2.} = 33.5671, x_{3.} = 38.8304, x_{4.} = 34.0606$$

So, the grand total becomes,

$$x_{..} = x_{1.} + x_{2.} + x_{3.} + x_{4.}$$

$$x_{..} = 26.2171 + 33.5671 + 38.8304 + 34.0606$$

$$x_{..} = 132.6752$$

So,

$$x_{..}^2 = 17602.71$$

The correction factor (CF) is given by:

$$CF = \frac{x_{..}^2}{IJ}$$

$$CF = \frac{17602.71}{4 * 7}$$

$$CF = 628.6682$$

Now, using formula 10D from FP, we calculate SST, SSTr and SSE.

i) SST

$$SST = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}^2 - CF$$

$$SST = 683.3276 - 628.6682$$

$$SST = 54.6594$$

ii) SSTr

$$SSTr = \frac{\sum_{i=1}^{I} x_{i.}^2}{J} - CF$$

$$SSTr = \frac{x_{1.}^2 + x_{2.}^2 + x_{3.}^2 + x_{4.}^2}{J} - CF$$

$$SSTr = \frac{27.2171^2 + 33.5671^2 + 38.8304^2 + 34.0606^2}{7} - 628.6682$$

$$SSTr = 19.2525$$

iii) SSE

$$SSE = SST - SSTr$$

$$SSE = 54.6594 - 19.2525$$

$$SSE = 35.4069$$

## B.

Based on the SST, SSTr and SSE values, we can form the ANOVA Table as below.

| Source of Variation | Degree of freedom (df) | Sum of Squares (SS) | Mean Square (MS) | Test Statistic F |
|---|---|---|---|---|
| Treatment (Among) | $I - 1 = 3$ | $SSTr = 19.2525$ | $MSTr = SSTr/(I-1)$ $= 6.4175$ | $F = MSTr/MSE$ $= 4.35$ |
| Error (Within) | $I(J-1) = 24$ | $SSE = 35.4069$ | $MSE = SSE/I(J-1)$ $= 1.475$ | |
| Total | $IJ - 1 = 27$ | $SST = 54.6594$ | | |

Now, since we have the F-statistic, we can run our hypothesis test step by step.

**Step1:**

Null hypothesis, $H_0 : \mu_1 = \mu_2 = ... = \mu_I$.
Alternative hypothesis, $H_a : H_0$ is false.
Level of significane, $\alpha = 0.1$

**Step2:**

F-statistic,

$$F = \frac{MSTr}{MSE} = 4.35$$

**Step3:**

Using built in function in R, we find,

$$p - value = 0.01392537$$

**Step4:**

It is evident that,

$$0.01392537 < 0.1$$

$$p - value < \alpha$$

So, we can reject the null hypothesis and accept the alternate hypothesis. This indicates that there are significant differences among true average flight times for the four treatments.