MATH 204 - EXERCISES 1

These exercises are not for assessment

The following questions relate to a **completely randomized design** (CRD) with k treatment groups. We use the following notation:

- n_i is the number of experimental units in the ith treatment group.
- $n = n_1 + \cdots + n_k = \sum_{i=1}^{k} n_i$ is the total sample size.
- x_{ij} is the measured response for the jth unit in the ith group.
- sample mean for treatment i
$$\overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$
- sample variance for treatment i
$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2$$
- overall sample mean
$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}$$
- Sum of Squares for Treatments (SST)
$$SST = \sum_{i=1}^{k} n_i (\overline{x}_i - \overline{x})^2$$
- Sum of Squares for Error (SSE)
$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2$$
- Overall Sum of Squares (SS)
$$SS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x})^2$$

1. By writing
$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i + \overline{x}_i - \overline{x})^2$$
and noting that
$$(x_{ij} - \overline{x}_i + \overline{x}_i - \overline{x})^2 = (x_{ij} - \overline{x}_i)^2 + 2(x_{ij} - \overline{x}_i)(\overline{x}_i - \overline{x}) + (\overline{x}_i - \overline{x})^2$$
show that
$$SS = SST + SSE$$

2. The ANOVA F-test for comparing the treatment means, $\mu_1, \ldots, \mu_k$, in k treatment groups in a CRD uses the two statistics

$$\begin{aligned} MST &= \frac{SST}{k-1} = \frac{1}{k-1}\sum_{i=1}^{k} n_i(\overline{x}_i - \overline{x})^2 \\ MSE &= \frac{SSE}{n-k} = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2. \end{aligned}$$

For testing

$$\begin{aligned} H_0 &: \mu_1 = \cdots = \mu_k \\ H_a &: \text{At least two treatment means different} \end{aligned}$$

the test statistic is

$$F = \frac{MST}{MSE}$$

and if H_0 is **true**

$$F \sim \text{Fisher-F}(k-1, n-k).$$

The rejection region for the test with significance level α is

$$F > F_\alpha(k-1, n-k)$$

where $F_\alpha(\nu_1, \nu_2)$ is the $1-\alpha$ percentage point of the Fisher-F distribution with ν_1 and ν_2 degrees of freedom (*see pages 899-905 of McClave and Sincich 10th Ed*)

Use the ANOVA F-test to test for a difference in means between the $k = 3$ treatments in the following example. Use $\alpha = 0.05$.

Data: The example that follows is based on a study by Darley and Latané (1968), designed to discover whether the presence of other people has an influence on whether a person will help someone in distress. The experimenter (a female graduate student) had the subject being tested wait in a room with either 0, 2, or 4 other people. The experimenter announces that the study will begin shortly and walks into an adjacent room. In a few moments the person(s) in the waiting room hear her fall and complain of ankle pain. The research question to be answered is whether the number of people in the waiting room influences the response time of the subject.

The **response variable** is the number of seconds it takes the subject to help the experimenter. The single **factor** is the number of other people in the waiting room, and there are three **factor levels** (0, 2 and 4).

The data observed in the study were as follows: measurements are in seconds.

Treatment		
0	2	4
25	30	32
30	33	39
20	29	35
32	40	41
	36	44

Reference: Darley, J.M., and Latané, B. (1968). Bystander intervention in emergencies: Diffusion of resposiblity. *Journal of Personality and Social Psychology*, 8(4), 377-383.