Homework 3 STAT301 Name Charles Sandahl

1. Dr. Brown would like to compare his students' performance after he taught three classes with three different pedagogies. He used his students' final exam scores as the measurements. He took a sample of 12 students' final scores from pedagogy A, 12 from pedagogy B and 12 from C. Use the 0.05 significance level to test the claim that there is a significant difference between the mean final scores of the three pedagogies. The data is on D2L, named as Student final scores.
   a) Step 1: Formulate the null and alternative hypotheses in symbols

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: At\ least\ one\ \mu_i\ is\ different$$

   Where μ_1 Is pedagogy A, μ_2 is pedagogy B, and μ_3 is pedagogy C

   Step 2: Choose the one-way ANOVA test

   b) Step 3: Find the correct test statistic and p-value for testing the means from R output

$$F - value = 5.926$$
$$p - value = 0.00632$$

   c) Step 4: Make a decision based on the significance level of 0.05; and explain your decision in the context.
   The p-value is 0.006, which is significantly lower than our α-value of 0.05 we reject the null hypothesis which suggests that there is evidence there is a significant difference between the mean final scores of the three pedagogies.

   d) If there are significant differences, perform a pairwise comparison to determine where the differences lie.
   There are significant differences between pedagogies B-A and C-A.

2. The data CO2 on d2l recorded the CO2 emission from the burning of fossil fuels (metric tonnes of CO2 per person) (Source: https://cdiac.ess-dive.lbl.gov/). The data recorded the CO2 emission per person from three countries: USA, Russia and India. Use the 0.05 significance level to test the claim that the average CO2 emission per person are significantly different between the three counties. Note, the format of the data is wide format. You will need to use the `gather()` OR the `pivot_longer()` `function` in R to convert it to a long format with one independent variable and one dependent variable so the format is analysis ready.
   a) Step 1: Formulate the null and alternative hypotheses in symbols

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: At\ least\ one\ \mu_i\ is\ different$$

   Where μ_1 Is USA, μ_2 is Russia, and μ_3 is India

Step 2: Choose the one-way ANOVA test

b) Step 3: Find the correct test statistic and p-value for testing the means from R output

$$F - value = 1524$$
$$p - value = < 2\text{e} - 16$$

c) Step 4: Make a decision based on the significance level of 0.05; and explain your decision in the context.

The p-value is <2e-16, which is significantly lower than our α-value of 0.05 we reject the null hypothesis which suggests that there is evidence that the average CO2 emission per person are significantly different between the three counties.

d) If you discover the significant difference, Where the differences lie?

There are significant differences in all three of the countries with all three having very small p-values

3. A clinical research was conducted to compare three treatments (drug A, B and C) on treating depression. The researchers recruited 24 volunteers and randomly assigned 8 to take drug A, 8 to take B, and 8 to take C. The baseline characteristics of the 24 volunteers are the same. The researchers measured the depression scores after the volunteers took the medications. The data is listed in the table below. Suppose the data follow normal distribution. The lower depression score, the more effective the drug is. Use the 0.05 significance level to test the claim that the average depression scores after taking the drugs are significantly different.

Drug	score1	score2	score3	score4	score5	score6	score7	score8
Drug A	51	45	33	41	36	38	39	33
Drug B	23	31	23	20	19	26	28	17
Drug C	22	18	29	32	41	20	16	23

a) Enter the data in R with the correct format (one IV and one DV)

b) Step 1: Formulate the null and alternative hypotheses in symbols

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: At\ least\ one\ \mu_i\ is\ different$$

Where μ_1 Is drug A, μ_2 is drug B, and μ_3 is drug C

Step 2: Choose the one-way ANOVA test

c) Step 3: Find the correct test statistic and p-value for testing the means from R output

$$F - value = 14.44$$
$$p - value = 0.000113$$

d) Step 4: Make a decision based on the significance level of 0.05; and explain your decision in the context.

The p-value is <2e-16, which is significantly lower than our α-value of 0.05 we reject the null hypothesis which suggests that there is evidence that the average depression scores after taking the drugs are significantly different.

e) Which two drugs had significantly different average depression scores?
There seems to be a significant difference between Drug B-Drug A and Drug C-Drug A.

4. A representative from a seed-producing company has been on a seed-collecting trip in the hope of finding improved varieties of millet. She wants to discover whether the yields from different local seed sources are significantly different. An experiment is set up with five different seed sources (A, B, C, D and E). For each seed source, she planted seven replications of plots. She measured the yield from each plot at the end of the growing season. She is interested in testing the main effect of seed sources with the significance level of 0.05. The data is as following. Perform an appropriate test to answer the following questions:

A	B	C	D	E
1.4	1.5	0.3	1.5	1.1
1.2	1.8	0.5	1.9	1.5
1.1	1.6	0.8	1.2	0.7
1.0	1.3	0.2	1.6	1.4
0.9	1.2	0.6	1.8	1.2
1.8	1.9	0.5	1.5	1.3
1.2	1.7	0.3	0.6	1.7

a) Enter the data in R with the correct format.
b) Step 1: State the null and alternative hypotheses

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_a: At\ least\ one\ \mu_i\ is\ different$$

Where μ_1 Is seed A, μ_2 is seed B, μ_3 is seed C, μ_4 is seed D and μ_5 is seed E

Step 2: Choose the one-way ANOVA test
c) Step 3: Find the test statistic and p-value for testing the significant difference in mean

$$F - value = 13.49$$
$$p - value = 2.11\text{e} - 06$$

d) Step 4: What conclusion can you draw based on the p-value and the significance level of 0.05?
The p-value is 2.11e-06, which is significantly lower than our α-value of 0.05 we reject the null hypothesis which suggests that there is evidence that yields from different local seed sources are significantly different

e) Step 5: Which two seed sources are significantly different from each other based on the significance level of 0.05?
There are significant differences between seeds C-A, C-B, D-C, and E-C.

5. Suppose that we have a different data for the chest deceleration for three different size categories (small, midsize and large) of cars. In the sample, we have a total of 21 cars. We still want to use the significance level of 0.05 and test the claim that the different size categories have the different chest deceleration in the standard crash test. Suppose the equal variance assumption is satisfied. Part of the output from SPSS is listed below. Use the information from the SPSS output, calculate the MS_B, MS_W, df_B, df_W, and the F statistic.

Sources	Sum of squares	Df	Mean square	F statistic	P-value
Between groups	200.857	?	?	?	0.061
Within groups	549.714	?	?		

Show your calculations:

$df_B =$ $3-1=2$

$df_W =$ $21-3=18$

$MS_B =$ $\frac{200.857}{2} = 100.4285$

$MS_W =$ $\frac{549.714}{18} = 30.3174$

$F =$ $\frac{100.4285}{30.3174} = 3.31256$