Two-Way ANOVA Test

Sarah Katherine Stansell

Math 301-03

11-11-19

Exercise plays an important role in people’s lives due to its health benefits. Because of the importance of exercise, I chose to analyze the interaction between one’s sex and classification (freshmen, sophomore, junior, or senior) and their effect on the number of hours of exercise people get in a typical week. People might be interested in this data because it will give information on who might be more likely to exercise based on their sex or classification. I received my data from a simple random sample from the Fall 2019 Statistics Survey response and this data is accurate because it was given to me from my professor as a simple random sample.

Because I am analyzing how two factors can affect the number of hours of exercise people get per week I will use a Two-Way ANOVA test. A two-way ANOVA test is an efficient way to study two factors simultaneously. In this test, I have three comparisons: means for the first factor, means for the second factor, and means for the interaction of the factors. To start the test, we must define our factors(independent variables) and response variables (dependent variables). Factor 1 is sex which includes either male or female. Factor 2 is class which includes freshmen, sophomores, juniors, and seniors. My response variable is the mean number of hours of exercise people get in a typical week(in minutes).

To begin, I must define the null and alternative hypothesis. For factor 1, the null hypothesis is H0: μM=μF. This means that the mean number of hours of exercise all men get in a typical week in minutes equal the mean number of hours of exercise all females get in a typical week in minutes. The alternative hypothesis, contrary to the null hypothesis, is Ha: μM≠μF. μM equals the mean number of hours of exercise all men get in a typical week in minutes. μF equals the mean number of hours of exercise all females get in a typical week in minutes.

For factor 2 the null hypothesis is H0: μA= μB= μC= μD. The alternative hypothesis is Ha: μA≠μB or μA≠μC or μA≠μD or μB≠μC or μB≠μD or μc≠μd.. μA equals the mean number of hours of exercise all freshmen get in a typical week in minutes. μB equals the mean number of hours of exercise all sophomores get in a typical week in minutes. μC equals the mean number of hours of exercise all juniors get in a typical week in minutes. μD equals the mean number of exercise all seniors get in a typical week in minutes.

I am also testing the interaction between the two variables. The null hypothesis is H0: there is no interaction. The alternative hypothesis is Ha: there is an interaction.

To continue with the Two-Way ANOVA test we must see if our data meets the test conditions. Data must be from a simple random sample and each sample of my data is from a simple random sample. Additionally, data must be from independent samples. I assume my data is from independent samples but I do not know for sure so this could be a problem. I will now create boxplots to check for skew and outliers. Ideally there should be no outliers and no strong skew of my data. See figure 1 below to see how I inserted data into SPSS.

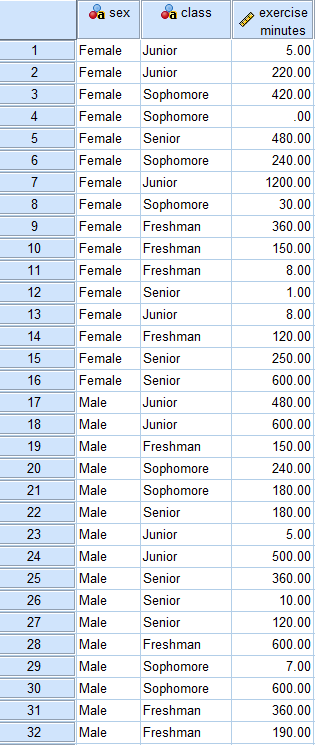


Figure 1

I then created boxplots for factor 1(Sex) and found an outlier for females and a slight skew for males. Refer to figure 2 below. The outliers mean that there could be a problem with my data. The boxplots for factor 2 (class) shows that there is a strong skew for freshmen and a slight skew for juniors, but no outliers. Refer to figure 3 below. The strong skew could be a problem.

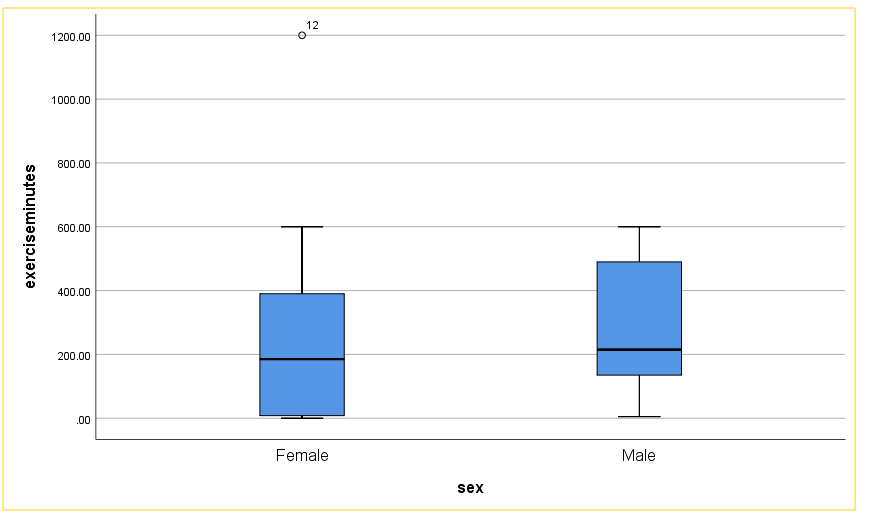


Figure 2 Figure 3

Next, I need to check the standard deviations of both factors to see if the ratio is less than 2. Refer to figure 4 below to see that I took the standard deviation of females divided by the standard deviations of males and got 315.84015 / 216.48768 = 1.4589. The ratio is less than 2, which means it’s therefore safe to assume that σF=σM.

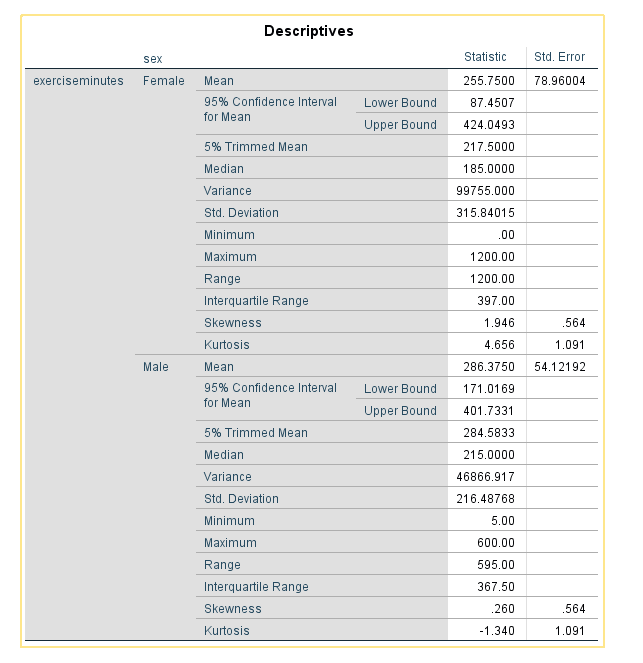


Figure 4

For factor 2, refer to figure 5 and 6 to see that I divided the largest standard deviation(juniors) by the smallest standard deviation(freshmen) to get 412.37526 / 187.19108 = 2.20296. Because the ratio is greater than 2, it is not safe to assume σA= σB= σC= σD.

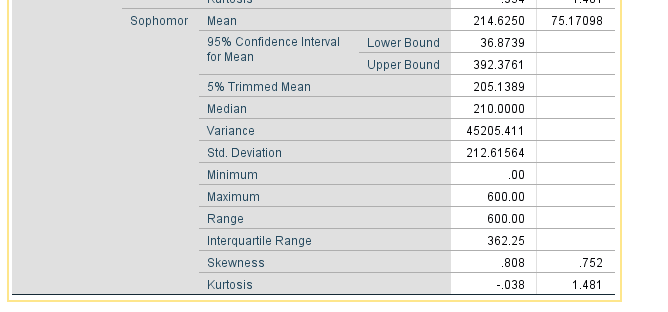
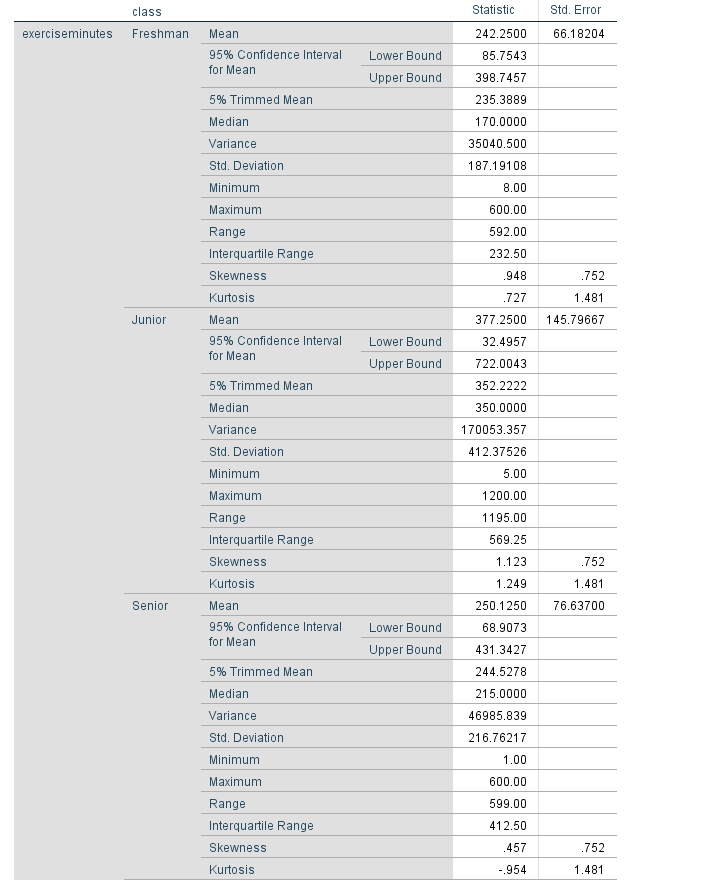


Figure 5 Figure 6

Next I created an interaction graph to determine if there is a possibility of interaction between sex and class. This interaction graph creates line plots of the sample means. When referring to figure 7 it is evident that there is a possible interaction because the lines are not parallel and they intersect.

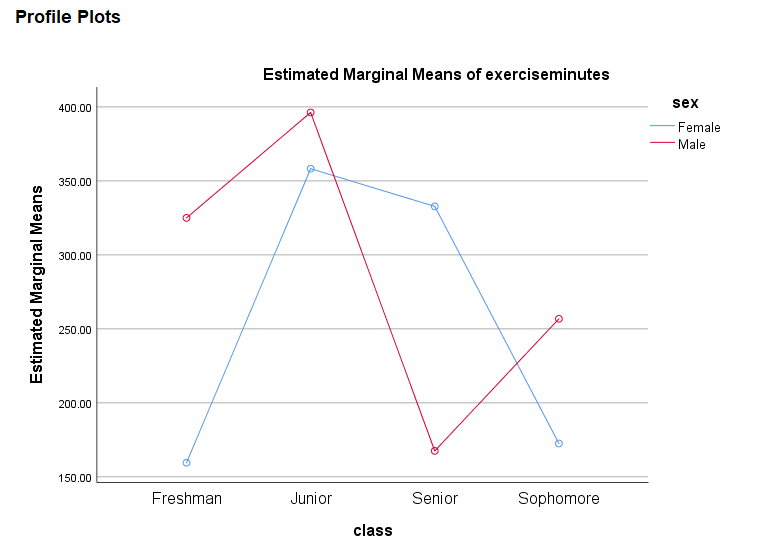


Figure 7

To continue with the Two -Way ANOVA test I chose a significance level of .01. I chose .01 because some data has outliers and is skewed and there is not a ton of data. To find if there is significant evidence that the difference in the mean number of hours of exercise per week varies based on sex and class I need descriptive statistics for both factors. By looking at figure 8 I see that for factor 1 (sex) the f value= .092 and the p-value= .764. Because of the p-value being greater than our significance level of .01 there is not significant evidence that there is a difference in the mean number of hours of exercise per week based on sex. For factor 2(class), the f value = .515 and the p-value =.676. Because of the p-value being greater than our significance level of .01 there is not significant evidence that there is a difference in the mean number of hours of exercise per week based on class. To test the interaction between the two factors I find that f=.487 and the p-value= .695. Once again, because of the p-value being greater than our significance level of .01 there is not significant evidence that there is an interaction between the two factors.

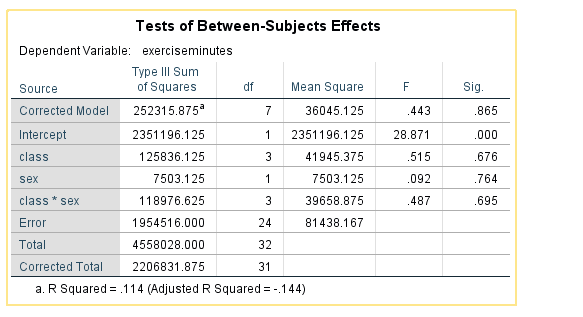


Figure 8

Technically my data shows that there is no evidence for an interaction between the two factors but if there were to be evidence I would need to compare them further using the Bonferroni method. I will continue and explain the process of using the Bonferroni method even though my data doesn’t require it. I would need to go under data and split the file for sex and use Bonferroni for class. When looking at figure 9, it is evident that there is no significant evidence that there is a difference in the mean hours of exercise in a typical week based on sex and class. This is because the significance levels are all above the significance level of .01.

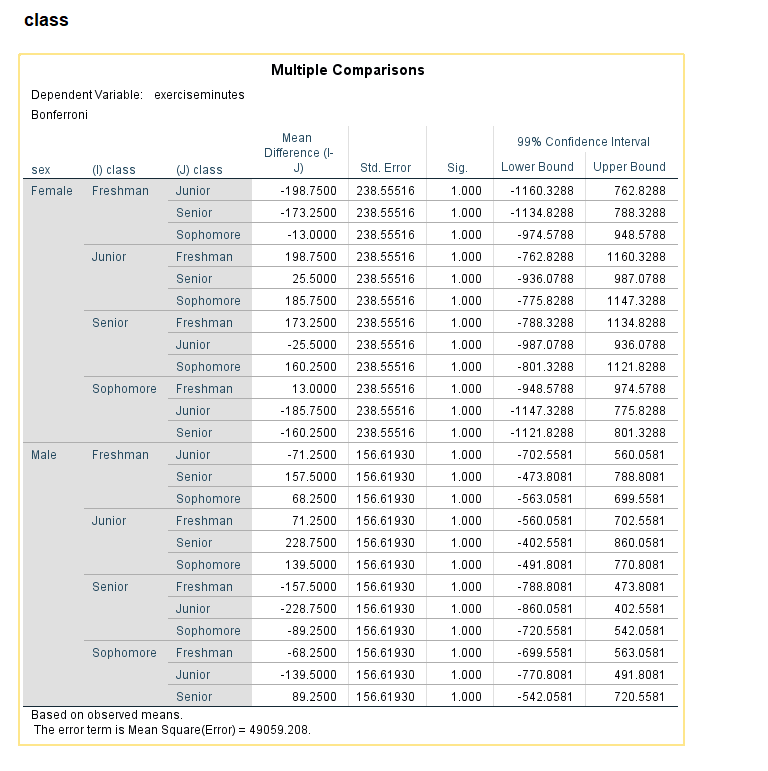


Figure 9

Overall the p-values for all factors were greater than my significance level of .01. That means for factor 1 (sex) that there was no significant evidence that the mean number of hours of exercise all females and males get in a typical week are different. For factor 2 (class) that means there is no significant evidence that the mean number of hours of exercise all freshmen, sophomores, juniors, and seniors get in a typical week are different. For the interaction between the two factors, the p-value was greater than the significance value of .01 so there is no significant evidence that there is an interaction. Because there was no significant evidence technically the Bonferroni method was not necessary but I conducted it anyway and it supported my findings that that there is no significant evidence that either sex or class affects the mean number of hours of exercise people get in a typical week (in minutes).