**1’Week09 Assignment due Oct28 before class: ANOVAs (7 points)**

This assignment is mostly a tutorial to help you better understand how to perform an ANOVA. Questions you need to answer to get points for this assignment are highlighted in yellow.

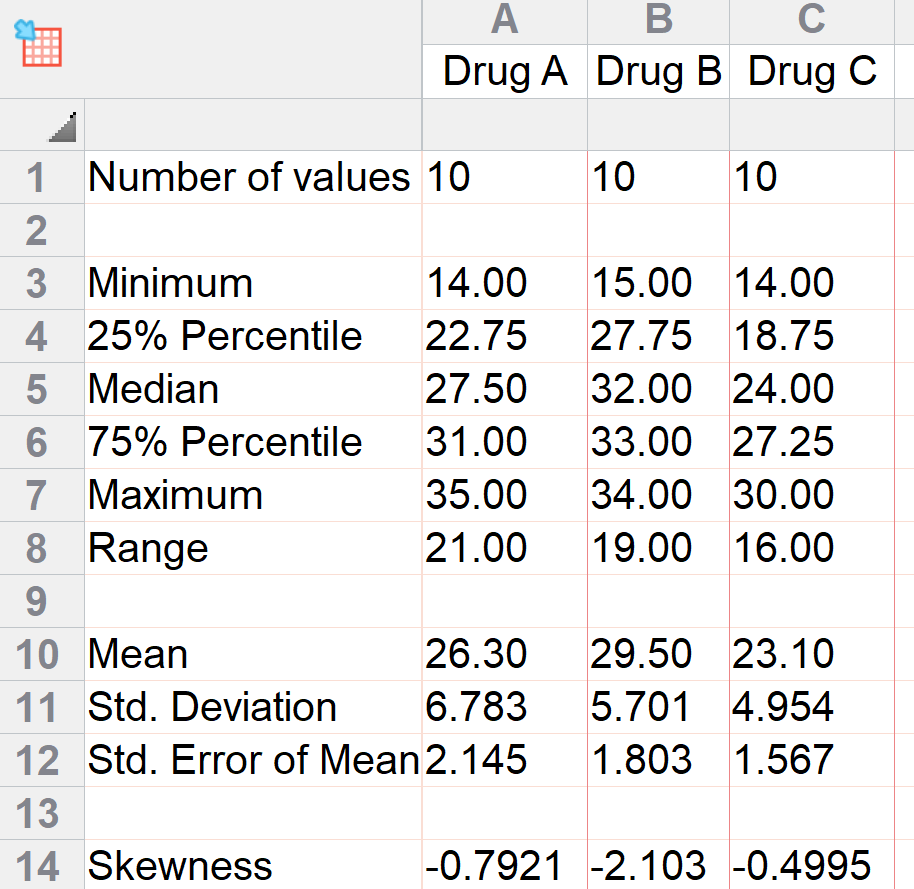
A team of researchers want to know which of three drugs, if any, is better at increasing lymphocyte percentage in people with lymphocytopenia, a condition that occurs when the lymphocyte count in the bloodstream is lower than normal. They recruited 30 people with borderline lymphocytopenia (lymphocyte% of 19%; normal is 20-40%). The subjects were randomized to one of three drugs. Each person took the drug and their lymphocyte percentage was measured 24 hours later. The data are in Week09\_lymphocytes.xlsx.

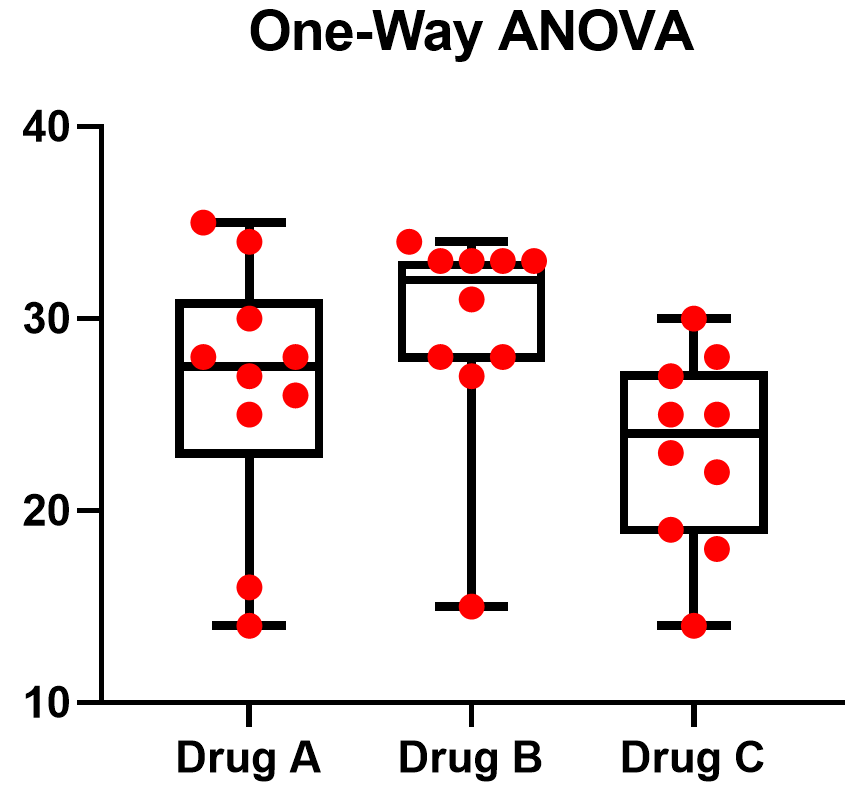
|  |  |  |
| --- | --- | --- |
| Drug A | Drug B | Drug C |
| 14 | 15 | 14 |
| 34 | 33 | 22 |
| 26 | 28 | 25 |
| 30 | 33 | 18 |
| 35 | 33 | 30 |
| 16 | 31 | 23 |
| 27 | 28 | 28 |
| 28 | 33 | 19 |
| 28 | 34 | 27 |
| 25 | 27 | 25 |

We have a continuous variable, lymphocyte%, and we want to know if it differs between three different drug groups. The data are meant for a one-way ANOVA if the data meet the test assumptions.

First, **the assumption of independence**. The groups are in different people and each value in each group is from a different person. Independence met.

Second, **the assumption of normality**. normality in each group and homoscedasticity between groups.





There appears to be a problem with normality. Drug A and Drug B appear to be fairly symmetrical with some skew to the left. Means and medians are similar. Drug B has much more skew and one “outlier” and the mean and median are farther apart than for the other groups. But the design is balanced (n=10 for each group), and the skew is in the same direction in each group, maybe this deviation from **normality can be tolerated by the ANOVA test**.

The last **the assumption of equal variance**, there appears to be homoscedasticity as the SD are fairly similar. The ratio of the largest to the smallest (6.78/4.95) is <2 (**You can also check the Brown-Forsythe test and Bartlett’s test in Graphpad Prism).**

What about the effect of the sample size? At a minimum, you need at least two samples in each group to calculate the group means and the SD. Our two-way ANOVA example about bluebooks and typing skills in our lecture had 3 in each group. However, it’s hard to check test assumptions with such small sample sizes. Remember the rule of thumb that if you have a sample size of 30, then you have enough samples size for a test to be robust, that means some deviation from normality and homoscedasticity may not bias the test result much. Although there are some differences of opinion, a rule of thumb is there should be at least 10 observations per independent group. In our ANOVA example, there are three independent groups, Drugs 1-3. Then there needs to be a minimum sample size of 30, which we have. Therefore, we may assume that the deviations from normality in our sample may not bias the test, as long as it is not too severe.

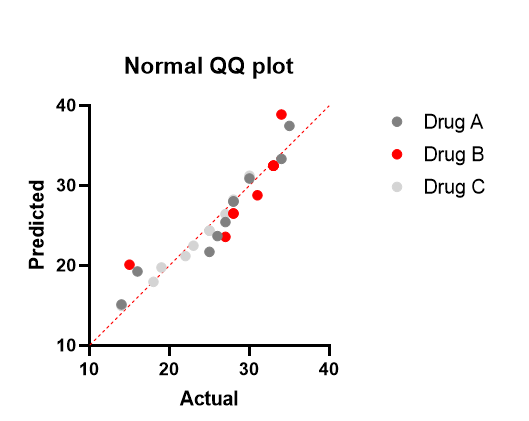
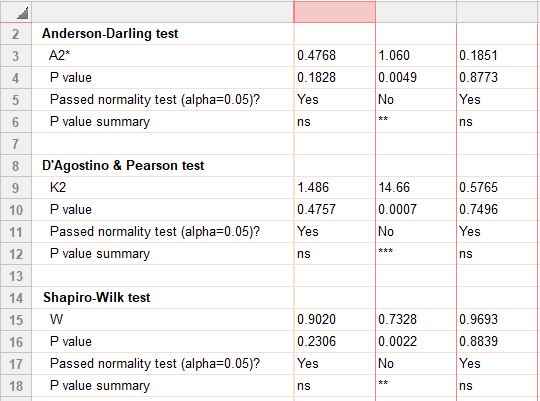
We will proceed with the one-way ANOVA assuming that that our test will be robust. Notice from the data summary above, Drug B has the highest mean lymphocyte% and Drug C the lowest. Just from this it seems that the drugs may have different effects and that Drug B increases lymphocyte% the most. Let’s see if our statistical test helps us determine if this is true.

Q1. What is the F statistic (with degrees of freedom) and the p-value? (1 pt)

**F(2,27)=2.981, p=0.0676.**

If you did the ANOVA correctly, you should have found **a p>0.05 so you fail to reject the null hypothesis** and conclude that lymphocyte% means are the same in the three different drug groups. Because the ANOVA p>0.05, you have **no reason to run a post-hoc test** to look at differences between groups.

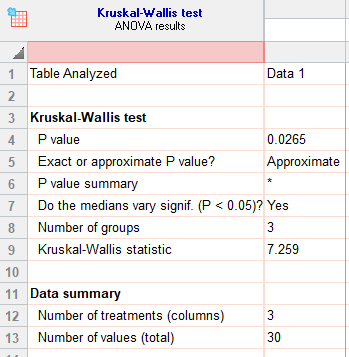
If you look at the data the means are 26.3, 29.5 and 23.1 for the three drugs. You may believe that these differences are **biologically significant**. Since the ANOVA found no differences, you know that **the within group variation was larger than the between group variation leading to a smaller F statistic**. Perhaps the data in **Drug B were too skewed** leading to a biased test result because you violated the assumption of normality.



I am now going to suggest doing something that could cause you to be accused of p-hacking (i.e., doing different analyses until you get the result you want). **P-hacking is something to be avoided** (we will discuss in a future lecture). But for the purposes of this example, we will run **the non-parametric Kruskal-Wallis**. We could have chosen to do this test first if **we decided data violated the normality assumption and this deviation from normally was too large** even with a sample of 30. Another option would be to **transform the data to create a more symmetrical distribution**. In this case, I tried transforming the data, but I was not successful in meeting the assumptions of normality. The non-parametric test is the best option.

Q2. What is the p-value for the Kruskal-Wallis test? (1 pt)

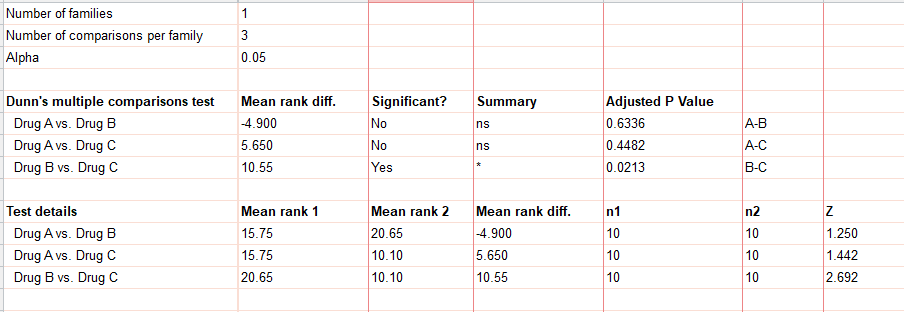
P=0.03 (0.0265)

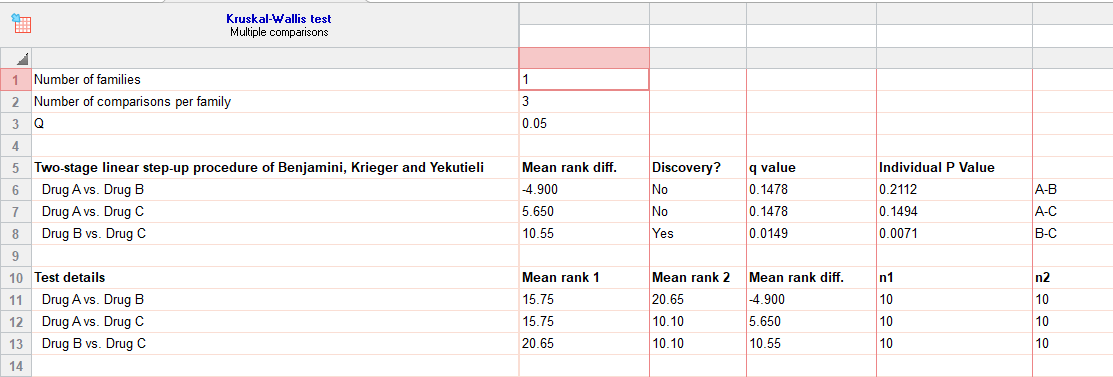


If you did the test correctly, you **rejected the null hypothesis** and concluded that **at least one group is different from the others**. Run a post-hoc test that compares all means to each other.

Q3. What group(s) are different? Include a measure of effect. (1 pt)

Drug B (median=32) is higher than Drug C (24)





OK. We reached **different conclusions** between **the parametric** and **non-parametric** tests on the same dataset. The question is, **which test is the least biased** (i.e., which is the correct application of a statistical test on the data). Assessing if data meet the assumptions of a test can often be a matter of opinion and arguments can many times be made for either choice of a parametric or non-parametric test. In this example, we may be **making a Type II error (false negative)** if we go with the results of the parametric test if **the assumption of normality is violated**.

What would I have done in reality (outside the confines of this teaching example)?. First, I would have **written a statistical plan that would explain in which situations** I would choose a particular test (we will cover this in more detail in a future lecture). I would **not have run a one-way ANOVA** on the raw data given **the large skew** in one of the groups. I would have tried to transform the data or done the Kruskal-Wallis test based on the strong skew and outlier in the Drug B data.

Now, let me throw in a big wrench into the works. The data are actually **repeated measures** of the effect of each drug in each of 10 individuals who were given the drugs serially. Drug A was given on Monday and lymphocyte% was measured on Tuesday. Drug B was given a week later (which was plenty of time for the effects of Drug A to wash out of their system). Lymphocyte% was measured the next day. Drug C was given one week after that. **The assumption of independence between drug groups is now violated**, **although there is still independence between the sets of repeated measures in each individual.**

Let’s check the other assumptions of a repeated measures ANOVA.

The **dependent variable**, lymphocyte%, is ***continuous***

The **categorical variable** has *>2 levels* (Drug A, Drug B, Drug C)

Data collection occurred over *equal time intervals* (0, 1, 2 weeks)

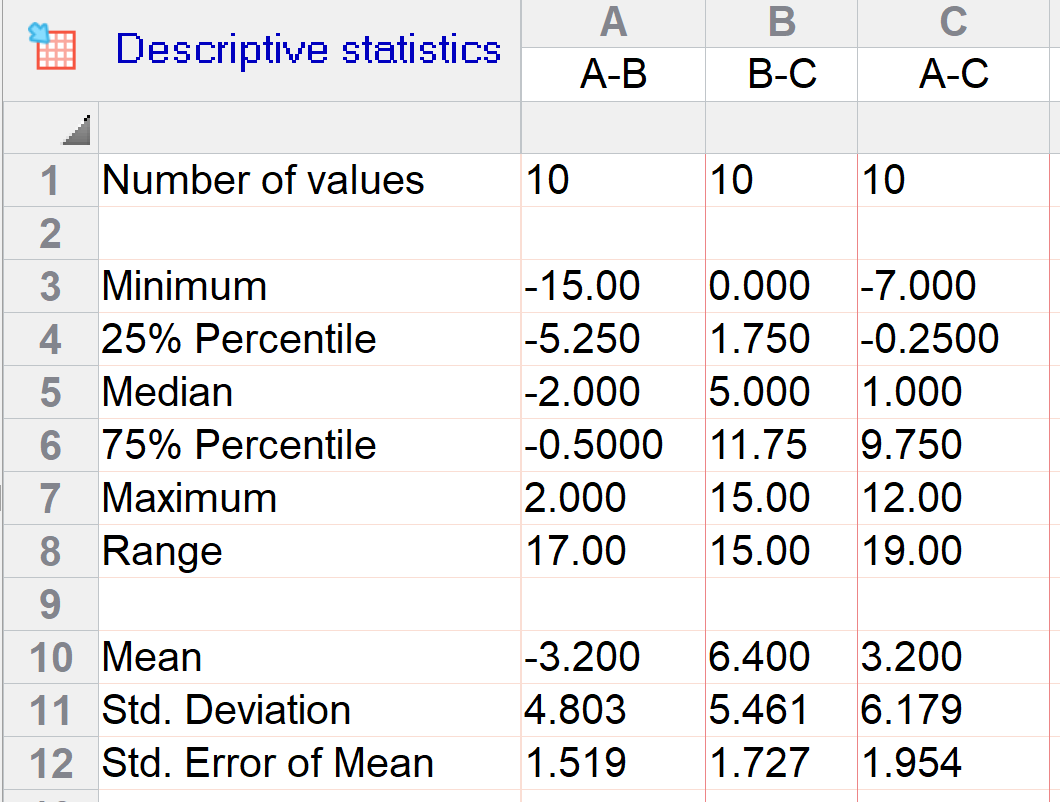
***Homoscedasticity*** between groups – **yes** we saw that

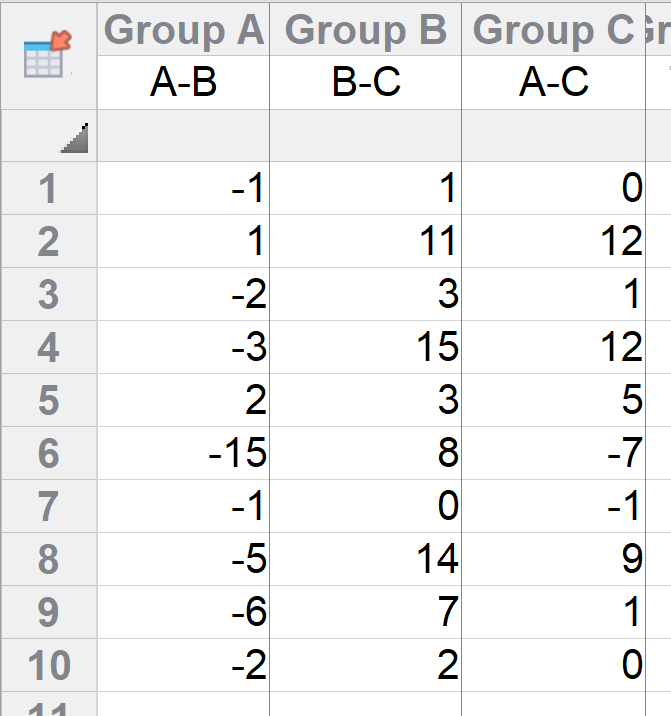
***No missing data*** – dataset is complete

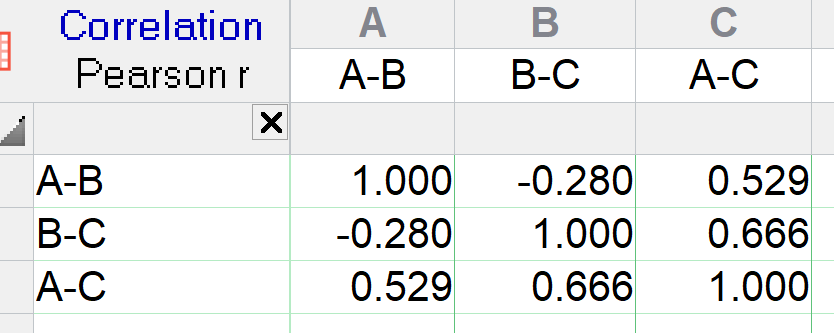
**Normality** within groups – problem; transformation did not help; **do mixed model or Friedman**

**Sphericity** – let’s assess for practice

Sphericity. Even though the problem of normality of the data precludes using the parametric test, we will still assess if the data would **meet the assumption of sphericity** for this tutorial. **This assumption is based on the difference between pairs of different groups**

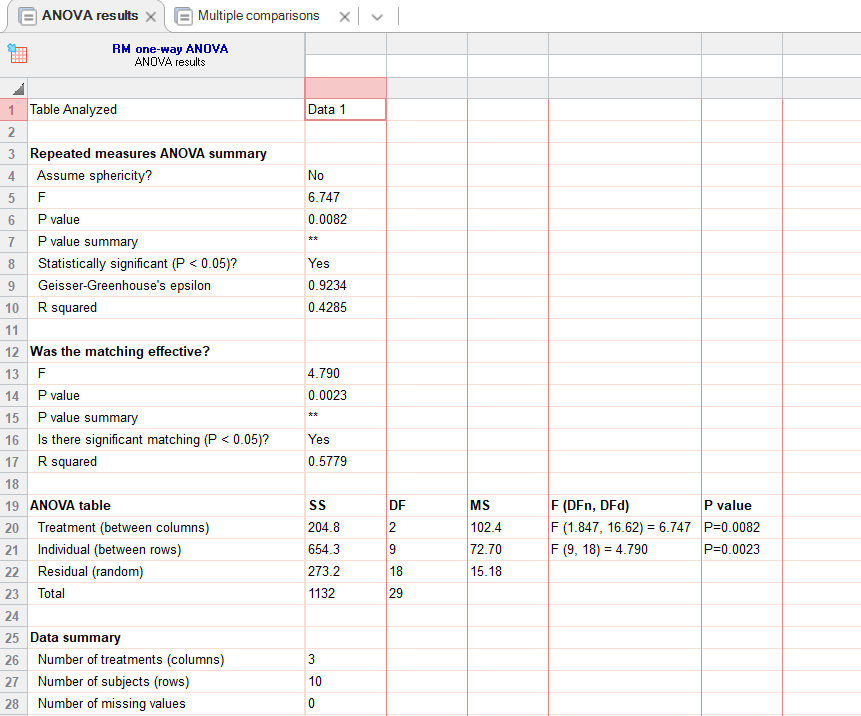






Looking at the descriptive data and correlation matrix above, it does **not appear that sphericity holds for the data**. Although the SD are similar, the **correlations between the differences is not the same** with both positive and negative correlation coefficients demonstrated with the data.

Even though we know the test results will be biased due to non-normality, we will practice preforming a RM ANOVA on the data.

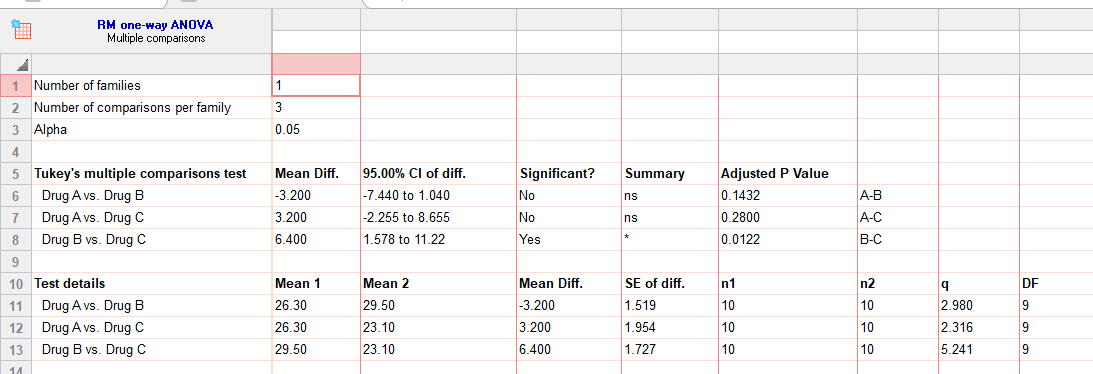




Q4. What is the F statistic, degrees of freedom and p-value for the difference between drug groups? (1 pt)

For the model with correction F(1,847,16,62)=6.747; p=0.008

If they report the results for the model without correction F(2,18)=6.747, p=0.006 or 0.007 – give 0.5 pt

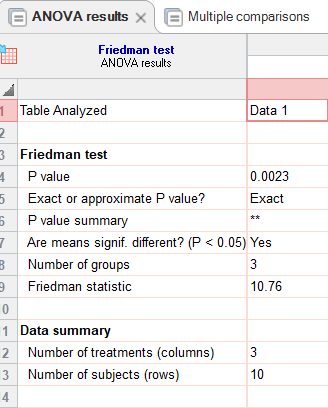




Q5. What drug groups are different from each other, if any? Include a measure of effect. (1 pt)

Drug B (mean=29.5) is higher than Drug C (23.1).

Now do the non-parametric Friedman test. Given that data failed the normality assumption, this is the test would should have done.





Q6. What is p-value for the difference between drug groups? (1 pt)

P=0.002

Q7. What drug groups are different from each other, if any? Include a measure of effect. (1 pt)

Drug B (median=32) is higher than Drug C (median=24).

