**Homework #8**

**Total number of questions (available points) for this homework is 35.**

1. The Kruskal-Wallis test is appropriate when you wish to compare\_\_\_\_\_?
   1. Three or more independent samples
   2. Three of more related samples
   3. A sample mean to a known population parameter
   4. One independent sample and one related sample
2. The test statistic that is calculated for a Kruskal-Wallis test is which of the following?
   1. P-value
   2. Z-statistic
   3. H-statistic
   4. r –value
3. The test statistic that is calculated for a Kruskal-Wallis has a distribution from the family of \_\_\_\_\_?
   1. Normal distributions
   2. Chi-square distributions
   3. Z-distributions
   4. t-distributions
4. The degrees of freedom for the Kruskal-Wallis test statistic are determined by\_\_\_\_\_\_\_\_?
   1. The number of subjects minus 1
   2. The sum of number of subjects in each group minus the number of groups
   3. The number of groups being compared
   4. The number of groups being compared minus 1
5. When a significant Kruskal-Wallis test is found, then you must proceed with pairwise comparisons to determine *which* group mean differences are significant. The test for significance of differences for these pairwise comparisons is performed using which of the following tests?
   1. Independent t-tests
   2. Related samples t-tests
   3. Additional Kruskall-Wallis tests
   4. Mann-Whitney tests
6. In SPSS, when you find a significant Kruskal-Wallis test, you may ask SPSS to conduct pairwise comparisons in one of two ways. Which of the following options will result in the largest inflation of familywise error rate?
   1. All pairwise
   2. Stepwise step down

Let’s try an example using the Kruskal-Wallis test. You wish to assess the effects of 3 different modalities to reduce pain so that you can proceed with therapeutic exercises. You have a total sample size of 14 (N=14) with 5 receiving ice (n=5), 5 receiving hot packs (n=5), and 4 receiving ultrasound (n=4). Pain is measured on a visual analog scale from 0 mm (no pain) to 100 mm (excruciating pain). Scores in the dataset are measures of *change* in pain following the modality. Please open dataset in your folder titled **Kruskal-Wallis dataset**.

Start by splitting the dataset by groups and look at histograms, kurtosis, skewness and K-S tests within each group as you have done in the past. You will see that the US group in particular is problematic and we could have predicted this because there are only 4 subjects and we cannot even assess normality. (In fact this is probably too small of a sample set to do anything with but let’s just go with it!)

So, although our data is interval/ratio scaled, we have very small sample sizes in our groups AND we have unequal sample sizes so we have decided to use the non-parametric Kruskal-Wallis test to analyze differences. Since we already had decided to do a non-parametric test, we really did not need to assess for assumptions such as normality but it is just good habit to inspect/explore your data before analyzing.

If you were calculating by hand, you would start by putting all the data together and ranking it from lowest to highest, and summing these ranks for each group (just as you did last week) for the Mann Whitney test. The summed ranks (Ri) for each group, would then be used in the formula on page 230 to calculate H and we would then compare this test statistic to a critical value in the **Chi Square distribution** to determine significance. I have had you do this type of calculation in the past and I am confident that you can do it correctly, so let’s let SPSS do that calculation for us. (Oh thank goodness, you say!)

Run the Kruskal-Wallis test as described by Dr. Field in the text and video. When you select Kruskal-Wallis, you will need to tell SPSS what to do if it finds a significant overall test. If you select “None” from the dropdown menu of multiple comparisons, SPSS will only run the Kruskal-Wallis overall test of significance (sometimes referred to as omnibus test). This tests the overall hypothesis that the groups (modalities) come from different populations.

1. What is the calculated H-statistic? 7.340
2. What is your p-value for the Kruskal-Wallis test? 0.025

Since this test is significant, you know that the groups come from different populations but we don’t know which groups differ on measures of pain change. So, we need to run additional tests. On your Kruskal-Wallis test, if you selected “All pairwise” for the multiple comparisons, SPSS would have conducted comparisons on all possible pairs.

By the way, the total number of pairwise comparisons is defined as **k(k-1)/2** where k is the number of groups. For 3 groups (k=3), there would be a total of 3 comparisons.

NOTE: When you run the test now, you will again see results of the Kruskal-Wallis test with Box plots, test statistic and p-value. To see the results of the multiple comparisons that you requested, click on “View” on the bottom of your screen. This will show a dropdown menu that you can select “pairwise comparisons” to view those results.

1. Cut and paste the resulting figure and table here. To do this, go to the “Edit” function in the Model viewer and select “Copy Auxiliary View”.

A picture containing chart

Description automatically generated

1. What does the Test Statistic reported in the table for each of the three tests represent?

The test statistic reported in the table for each of the three tests represent whether or not each modality reduced pain or not, in order to continue therapeutic exercises.

1. What does the Std. Test statistic tell you? (SPSS calculates this value by dividing the Test Statistic Value by the Standard Error). Should look familiar to you…

The standard test statistic tells us how these modalities would be measured if the groups were not being compared against each other, and instead just run against a normal curve (i.e., converted into z-scores).

1. Notice that there are two Significance columns (Sig. and Adj. Sig.). What does the Adj. significance tell you? (ie. what is it adjusting?)

It is adjusting the significance level to the smallest familywise in order to declare particular comparisons statistically significant.

1. What do you conclude from your multiple comparison tests?

I conclude that we should reject the null hypothesis, since there was found to be a statistically significant p-value of 0.025 when looking at the distribution of change in pain, measured in mm, is the same across the modality categories.

1. What is the effect size (*r*) for each of the three comparisons?

*r* for ice-hot packs = -0.152

*r* for ice-ultrasound = -0.273

*r* for hot packs-ultrasound = -0.257

1. Please write a statement of your results like you would if you were reporting your results in a manuscript.

Pain was significantly lessened when using the modalities of ultrasounds, ice packs, and hot packs before therapeutic exercises, H = 7.340, p = 0.025. Pairwise comparisons with adjusted p-values showed that there were no significant differences between using ice packs to combat pain before therapeutic exercises compared to using hot packs (p = 1.00, *r* = -0.152). There were also no significant differences between using hot packs to combat pain before therapeutic exercises compared to using ultrasounds (p = 0.062, *r* = -0.257). There was a significant difference in pain between those using an ice pack and those using an ultrasound before therapeutic exercises (p = 0.042, *r* = -0.273).

Notice that for the multiple comparisons, you could also have selected “Stepwise step-down” rather than all pairwise. This would result in fewer comparisons than all pairwise. With only 3 groups to compare, it really doesn’t make sense to select this option since you are only doing 3 pairwise tests. SPSS makes a Bonferroni adjustment for the multiple comparisons to control for inflation of Type I error.

1. If you had 5 groups to compare, how many pairwise comparisons would be possible? 10

So you can see that with more than 3 groups, it would make more sense to select the stepwise option for multiple comparisons because it would cut down on the number of tests that would be conducted. Go ahead and select that option with your dataset just to give it a try.

1. Paste the output table for stepwise comparisons herTable

   Description automatically generated

Diagram, box and whisker chart

Description automatically generated

This is a bit more difficult to interpret but basically what it is telling you is that the US group is significantly different from the other groups (but does not compare with each group), and that Ice and Hot packs are not significantly different from each other (p=.750).

NOTE: It tells us that the US group is significantly different from the other groups but does NOT provide a p-value. We just know that the p-value is <.05.

1. Please write a statement of your results of the Kruskal-Wallis and multiple comparisons using the step-down analysis. (be sure to refer to Field p. 241 for examples)

Pain was significantly affected by using different modalities (ice packs, hot packs, and ultrasounds), H = 7.340, p = 0.025. Step-down follow up analysis showed that if ultrasounds are used before therapeutic exercises, it significantly reduces pain compared to the other modalities; however, using ice and hot packs before therapeutic exercises had no significant effect on pain, p = 750.

The remaining questions relate to the parametric ANOVA.

1. SST is the total sum of squares or total amount of variation in our dataset. Which of the following best describes total sum of squares when comparing independent groups?
   1. The sum of squared differences between each individual group mean
   2. The sum of squared differences between each individual data point and the grand mean
   3. The sum of squared differences between each group mean and the grand mean
   4. The sum of squared differences between each individual score and the group mean to which that person belongs to
2. How are degrees of freedom determined for the total sum of squares (SST)?
   1. Number of people in each group minus 1 (n-1)
   2. Total number of people in the sample minus 1 (N-1)
   3. Number of groups minus 1 (k-1)
   4. Number of people in each group (n) minus number of groups (k)
3. Model sum of squares (SSM) is the amount of variation explained by our model. Which of the following best describes the model sum of squares when comparing independent groups?
   1. The sum of squared differences between each individual group mean
   2. The sum of squared differences between each individual data point and the grand mean
   3. The sum of squared differences between each group mean and the grand mean
   4. The sum of squared differences between each individual score and the group mean to which that person belongs to
4. If you are running an ANOVA to test for differences between 4 independent groups, how many degrees of freedom will the model sum of squares have?
   1. 4
   2. 3
   3. 2
   4. 1
5. The residual sum of squares (SSR) is the variation that cannot be explained by our model. Which of the following best describes SSR when comparing independent groups?
   1. The sum of squared differences between each individual group mean
   2. The sum of squared differences between each individual data point and the grand mean
   3. The sum of squared differences between each group mean and the grand mean
   4. The sum of squared differences between each individual score and the group mean to which that person belongs to
6. Recalling the important relationship that we studied early in the semester: **Test stastic= variance explained by the model / variance not explained by the model,** which of the following would you predict would define our F- statistic?
   1. MST / MSR
   2. MSM/ MSR
   3. None of the above
7. When conducting an ANOVA, it is important to test for homogeneity of variance (HOV) between the groups being compared. If your Levene’s test shows that the HOV is violated, which of the following typically makes the best adjustment to the F-ratio in terms of preserving statistical power?
   1. Welch test
   2. Brown Forsythe test
   3. Games- Howell test
   4. None of these adequately adjust for violated HOV

Please answer Questions 26-31 using the following SPSS output.



1. How many groups were compared in this analysis? 3
2. What is the total number of subjects in this dataset (ie. N)? 49
3. In this output, which of the following terms represents SSM?
   1. Between Groups SS
   2. Within Groups SS
   3. Total Groups SS
   4. SSM is not represented in this output
4. In this output, which of the following terms represents SSR?
   1. Between Groups SS
   2. Within Groups SS
   3. Total Groups SS
   4. SSM is not represented in this output
5. Calculate Mean Square(Within Groups) 47.611
6. Calculate the F-ratio. 16.62
7. What does the term familywise error rate mean?
   1. The error rate across statistical tests conducted on several different datasets
   2. The error rate across statistical tests conducted on one dataset to answer the same research question
   3. The error rate across statistical tests conducted on one dataset to answer any number of research questions
8. When testing assumptions to compare means for several independent groups using an ANOVA, it is important to test normality \_\_\_\_\_
   1. Across the entire sample
   2. Separately for each group
   3. For differences between the groups
9. Since the F-ratio is a ratio of systematic variation to unsystematic variation, a calculated F-statistic of less than 1 automatically signifies:
   1. Non-significant effect
   2. Significant effect
   3. Deviation from normality
   4. Cannot tell from the information presented
10. To assess the meaning of a calculated F-ratio, the value is compared to a critical value from which of the following distributions?
    1. Normal distribution
    2. t-distribution
    3. F-distribution
    4. Chi square distribution