

Homework 2

Answer the following questions using: R.

Submitted homework solutions should be well written and should include all R code and R output. R Markdown is a useful tool for doing this nicely.

1. Consider the following data:

```
group1 <- c(2.9736,0.9448,1.6394,0.0389,1.2958)
group2 <- c(0.7681,0.8027,0.2156,0.0740,1.5076)
group3 <- c(4.8249,2.2516,1.5609,2.0452,1.0959)
```

- (a) Perform the usual ANOVA F -test
 - (b) Perform the permutation F -test
 - (c) Perform the Kruskal-Wallis test
 - (d) Discuss the results of these three tests. (i.e. What are the p-values? What is the conclusion? Are the conclusions different? How similar are the p-values? Are any assumptions violated? etc.)
2. Download this data set: <http://bit.ly/Zg51IA>
It contains 50 observations divided equally between 5 groups. First perform the Kruskal-Wallis test and report your results. Then, if necessary, perform multiple comparisons using the Wilcoxon rank-sum test using the Bonferroni correction to control the family-wise (or experiment-wise) error rate at $\alpha=0.05$. Report your results and any significant differences that you find.
 3. Page 212, Question 4: Perform the Kruskal-Wallis test on the data to determine if there is a difference between the median lengths for the YOY gizzard shad populations in the four Kokosing Lake sites. Make sure to state the null and alternative hypotheses, the p-value, and your conclusion.

```
#Here is the Data
site1<-c(46,28,46,37,32,41,42,45,38,44)
site2<-c(42,60,32,42,45,58,27,51,42,52)
site3<-c(38,33,26,25,28,28,26,27,27,27)
site4<-c(31,30,27,29,30,25,25,24,27,30)
```

4. (Grad students only) Suppose $k=4$, $n_1=n_2=n_3=1$ and $n_4=2$. Obtain the form of the exact null (H_0) distribution of H for the case of no tied data.

5. (Grad students only) The Bonferroni correction is known to be conservative. This means that we are guaranteed to control family-wise error rate at less than or equal to α , but often the true family-wise error rate is much less than this.

Perform a simulation study to compute the true family-wise error rate when performing multiple comparisons under the null hypothesis that the means of all groups are the same. Perform the simulation study using the following parameters:

- $\alpha = 0.2$
- $k=6$
- $n_j=20$ for $j=1,2,3,4,5,6$,
- Data in each group is generated from a normal distribution with variance 1
- Use the Wilcoxon rank-sum test for multiple comparisons.

table(detected)