

Homework 1

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. The data in the code below are simulated exam scores. Suppose the exam was given in the semester after the course content was revised, and the previous median exam score was 70. We would like to know whether or not the median score has increased.

```
#test scores
testScores <- c(79,74,88,80,80,66,65,86,84,80,78,72,71,74,86,96,77,81,76,80,
               76,75,78,87,87,74,85,84,76,77,76,74,85,74,76,77,76,74,81,76)
```

- (a) Summarize the data. Is there any initial evidence that the hypothesis is true?
 - (b) State the null and alternative hypothesis that we wish to test here. Also state the level α that you will be using for this test.
 - (c) Test the hypothesis by using the binomial test.
 - (d) State your conclusion and p-value.
2. A certain data set has eight distinct observations, four from each treatment, and all of the observations from treatment 1 are bigger than the observations from treatment 2. What is the one-sided p -value associated with the permutation test? Show how you arrived at this p -value.
 3. Students in the introductory statistics class were asked how many brothers and sisters they have and whether their hometown is urban or rural. The data is below.

```
siblings<-data.frame(hometown=c(rep("rural",24),rep("urban",17)),
                     siblings=c(3,2,1,1,2,1,3,2,2,2,2,5,1,4,1,1,1,1,6,2,2,
                                2,1,1,1,0,1,1,0,0,1,1,1,8,1,1,1,0,1,1,2))
```

- (a) Test for a significant difference between rural and urban areas using the Wilcoxon rank-sum test. (Make sure to state your hypothesis, p-value, and conclusion!)
 - (b) Test for a significant difference using a permutation test. (Hint: Is checking all possible permutations feasible here?)
4. Create a fictitious data set where the Wilcoxon rank-sum test and the two-sample t-test lead to different conclusions at the 5% level of significance.

5. (Graduate Students Only) The simulated data below are from two normal distributions with equal means and possibly unequal variances. Test for differences in the scale parameters using the Ansari-Bradley test. Make sure to state the hypotheses, your conclusion, and your p-value.

```
#Treatment 1 data
trt1 <- c(21.9,20.2,19.4,20.3,19.6,20.4,18.4,20.1,22.0,18.9)
#Treatment 2 data
trt2 <- c(20.2,13.8,21.8,19.2,19.6,25.5,17.0,17.6,19.5,22.2)
```