AMS 597 Spring 2021
Homework 2 Due 03/02/2021 10:00 AM

Instruction: Submit your homework via Blackboard. If you have difficulty submitting via Blackboard, email both TAs directly. No late homework will be accepted, based on the time the email is sent out. If you are submitting via email (not recommended), put this as the subject of your email
AMS597 2021: Homework/Exam #2. ID: XXXXXXXXX. Name: XXX XXX

1. (a) Use `sample` function to generate a random vector that follows a multinomial distribution with probability (0.1, 0.2, 0.4, 0.3).
   (b) Using only random uniform generator (DO NOT use `sample`), generate a random vector that follows a multinomial distribution with probability (0.1, 0.2, 0.4, 0.3).

2. Generate 100 exponentially distributed random variables with rate 2, and plot their empirical distribution function.

3. Use the following dataset to answer the questions.
   `http://www.ams.sunysb.edu/~pfkuan/Teaching/AMS597/Data/d_logret_6stocks.txt`
   (a) Perform a t-test for American Express with the null hypothesis that the mean of its log return is zero.
   (b) Perform a Wilcoxon signed-rank test for American Express with the null hypothesis that the mean of its log return is zero.
   (c) Perform a two-sample t-test to conclude if the mean log return of Pfizer and American Express are same or not.
   (d) Compare the variance of log returns for Pfizer and American Express.
   (e) Perform a two-sample Wilcoxon test to conclude if the mean log returns of Pfizer and American Express are same or not.

4. Write your own function `my.t.test` which can perform both one and two sample t-test. For two sample t-test, it can perform both equal and unequal variance version. Your `my.t.test` will take the following argument (1) the vector `x`, (2) optional vector `y` if it is two sample t-test, (3) type of alternative hypothesis `alternative`, (4) the mean or mean difference that you are testing `mu`. Your function `my.t.test` should contain a routine to check for equal variance assumption using the F test. If the p-value of the F test is ≤ 0.05, then it will perform two sample t-test with unequal variance assumption. Your function `my.t.test` should return the test statistic `stat`, degrees of freedom `df` and p-value `p.value`. You may use `var.test()` or write your own F test, but you should not use `t.test()`.

5. Write your own function Wilcoxon rank sum test `my.wilcox.test` which can perform both exact and normal approximation test for two-sided alternative hypothesis. Your function will compute the p-value using normal approximation if n_1 and $n_2 \geq 12$. Otherwise, it will

compute the p-value using exact test. Your function will return the test statistics `W1` and `W2`, p-value `p.value` and a message indicating the type of test used (`normal` or `exact`). You may assume there is no ties in the data.

6. A regression through the origin model may be used when specific knowledge about the problem at hand suggests that the response variable is zero if and only if the predictor variable is zero. For such problems, the model can be written as

$$Y_i = \beta X_i + \epsilon_i; i = 1; \ldots; n$$

where ϵ_i's are iid $N(0, \sigma^2)$ random noise.

   (a) Derive the least-squares estimate of β.

   (b) Run the following R code:

```
set.seed(123)
x <- rnorm(50)
y <- 2*x+rnorm(50)
```

   Use the formula you derived in part (a) above to estimate based on this data, and draw a scatterplot of the data with the fitted line overlaid.

   (c) Perform the regression through origin using the R function `lm` on the same data.

   (d) Write your own function `my.kendall` to compute Kendall's τ between two variables and apply your function to x and y generated above.