Wilcoxon Rank Sum Test Project Report

Emmanuel Asante

12/1/2018

Introduction

The Wilcoxon Rank Sum Test is a a non-parametric Test that is used to test if samples come from the same distribution by testing if they have the same centers (usually median). The hypothesis of the tests are:

$ H_0 $: The two populations have the same distribution.
$ H_1 $: The two populations from different distributions.

Assumptions:

1. It Assumes that the observations are independent.
2. The two populations have equal spread or Variance.

Motivation

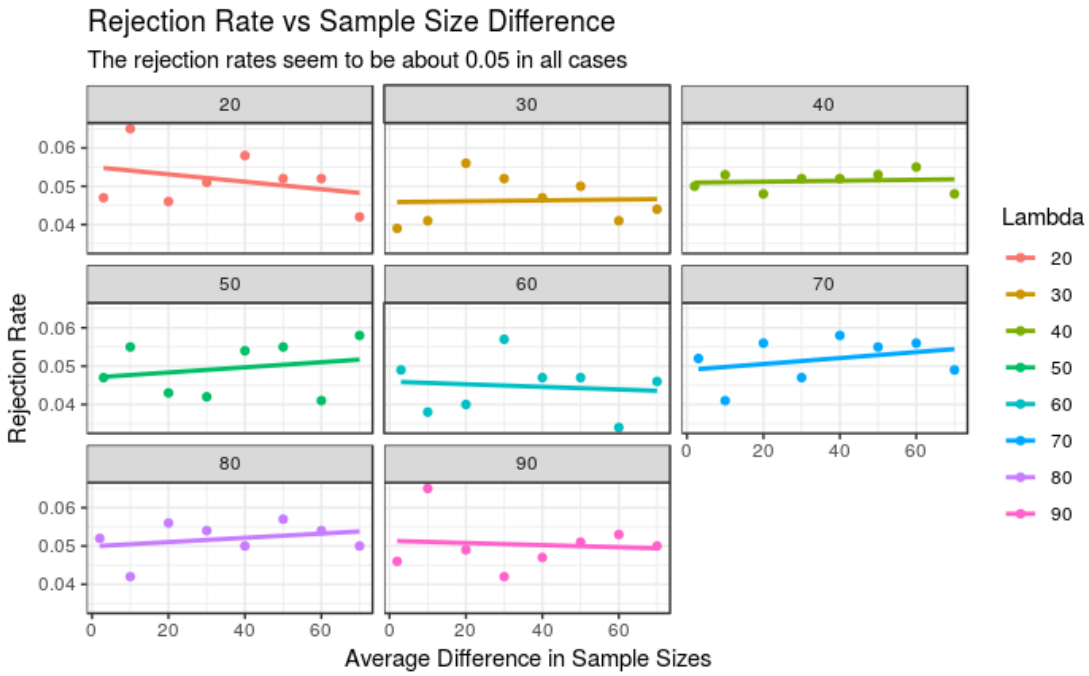
My interest was to investigate the relationship between rejection rates and the power of the Wilcoxon Rank Sum Test for different sample sizes. I did this by dividing my study into two parts. First, I investigated this relationship for cases in which the population distributions are symmetric. Then I investigated the case where the population distributions non-symmetric. In both cases, a fundamental distribution that I used was the Poisson distribution. This was to ensure that the assumption of equal spread was satistfied. The questions that I attempted to answer are as follows:

1. How does symmetry of the population distribution relate to the rejection rates of the test?
2. How does the difference in sample size affect the rejection rate?
3. For non-symmetric distribution, how does the difference in sample sizes affect the power of the test.

Results

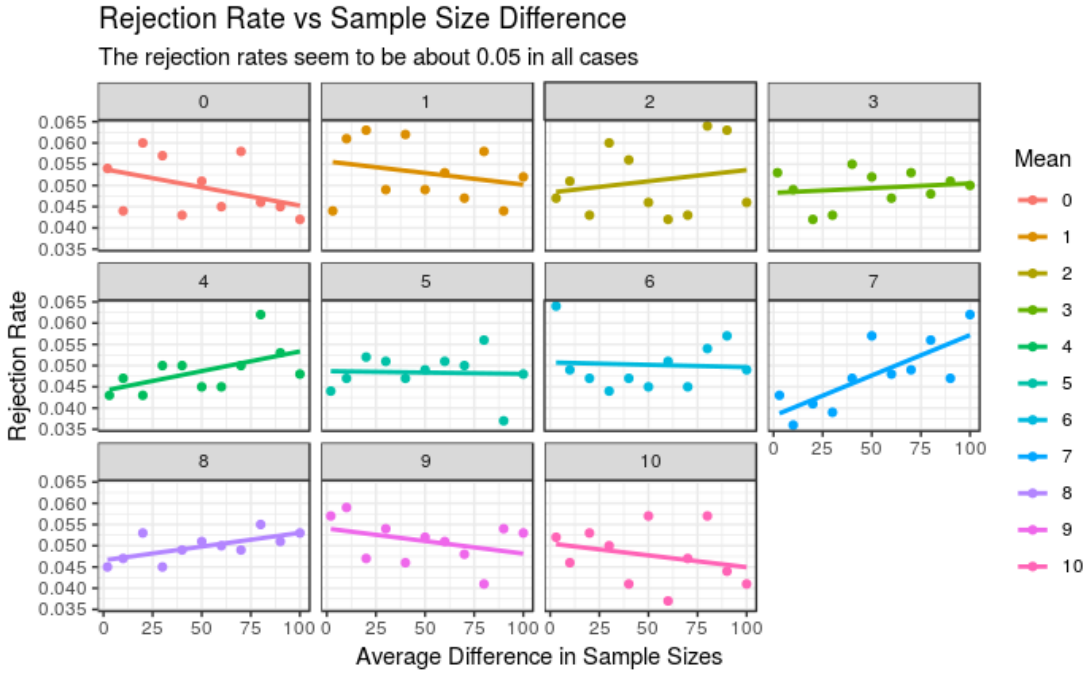
Poisson Population Case

The rejection rates seem to be approximately uniform for the different sizes of lambda for most of the various values of differences in sample sizes. This may show that the difference in sample sizes may not(subject to further investiigation) a significant effect on the rejection rate of the wilcoxon Rank sum test.This can be said of samples that are taken from the same symmetric($\lambda \geq 20$) population. In order naively infer this observation to other symmetric distributions, I will use another symmetric distribution to investigate this observation.



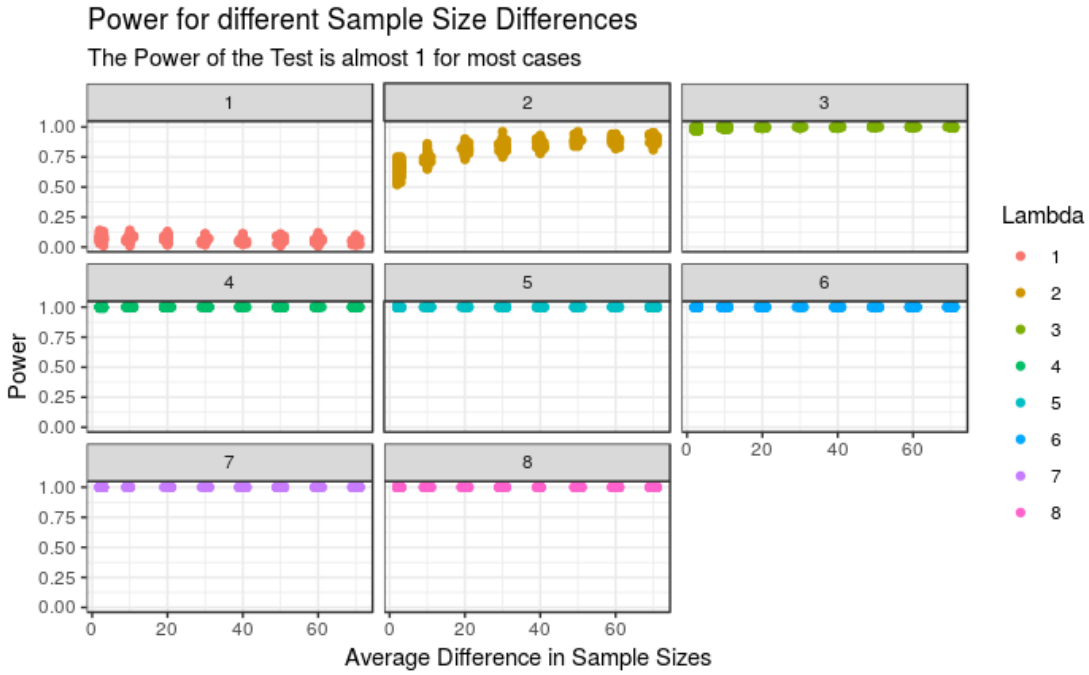
Normal Population Case

The rejection rates seem to fluctuate about 0.050 irrespective of the combination of the sample size difference and the mean combination. The rejection rates seem to deecrease for increasing sample size differences in cases where the population mean is 0, 1, 4, 9, 10. It is increasing for the other cases although the slopes are small except for when the mean = 7.



Different Skewed populations with same variance

For different Skewed populations with same variance, I observed that the power of the test for the case where the $\lambda = 2$ seem to increase with increasing sample size differences. Curiously, for $\lambda = 1$, I obtain the rejection rate to be almost 0.



Future Work

Future work should further investigate the findings for
1. several other population distributions.
2. Also the assumptions should be tested for different values of parameters.

References

1. University of Virginia Library :link
2. Dr. Charlotte Wickham's Notes for ST 541 Fall 2018.
3. R4DS by Dr. Hadley Wickham