Simulation Study For The Comparison of T-Test and the Wilcoxon-Test

Vacide Avsar   Guangming Liu

December 8, 2008

Abstract

This paper tries to figure out the differences between the results of the t-test and the Wilcoxon test in a simulation study.

1 Introduction

In this project, we are trying to compare the two statistical hypothesis tests, Student's t-test and the Wilcoxon test by a simulation study.

Student's t-Test is any statistical hypothesis test in which the test statistic has a Student's t distribution if the null hypothesis is true.

Different hypothesis tests make different assumptions about the distribution of the random sample in the data. One of the assumptions for the t-test is that the data are independently sampled from a normally distributed population. This assumption about the population distribution makes the t-test be a parametric statistical test. In some cases, the data within two correlated samples may fail to meet this assumption. When this happens, an appropriate non-parametric alternative test can be found. One of these non-parametric alternative tests is called the Wilcoxon Signed-Rank Test.

Like the t-test, Wilcoxon test involves comparison of the differences between measurements. On the other hand, it does not require assumptions about the form of the distribution of the measurements. It should therefore be used whenever the distributional assumptions that underlie t-test cannot be satisfied.

In our simulation study, we'll compare the size and the power of these two tests in two different cases. For the first case, we'll take the samples from a normally distributed population and for the other case, we'll have samples from uniformly distributed population. Uniform distribution is the one which is not normal but symmetric.

To see the effects clearly, we'll keep sample size, **n** small and the number of simulated datasets, **S** large enough:

$$n = 10$$

$$S = 10000$$

For each case, we'll test whether the means of the two samples are equal or not. We expect that the results for the size and power of tests will alter for these two different cases. Comparison of the results will show us which test is more relevant and safer.

Let's start with the first case.

2 Tests for Normally Distributed Population

We generate two sets of S datasets both from normal distribution. The significance level is 0.05 .

2.1 Means of the samples are equal (Case 1)

First, we evaluate the size of the two tests. For this, we take means equal. So,

$$\mu_1 = \mu_2 = 1$$

The variances of two samples are different:

$$(\sigma_1)^2 = \frac{1}{3}$$

$$(\sigma_2)^2 = \frac{1}{12}$$

Our null hypothesis is (Ho: $\mu 1 = \mu 2$)

We use R to generate two samples. Then, under our null hypothesis, we calculate the proportion of rejection of the test by determining the proportion of the p-value that are < 0.05 since we use $\alpha = 0.05$. If the proportion does achieve the given significance level, we get a relevant size of the test. After calculation, we get the size of the t-test as **0.0521** and the size of the Wilcox-test as **0.0507**. (For the results and codes, please see Appendix part). Since they're very close to significance level, they can be said as relevant tests. Then, we can calculate the power of the tests.

2.2 Means are not equal (Case 2)

Now, we will approximate the power of the two tests under an alternative hypothesis which says that μ_1 and μ_3 are not equal.

We reuse the sample 1 with $\mu_1 = 1$ from Case 1 and then generate another data set sample 3 with $\mu_2 = 2$. The variances are not equal:

$$(\sigma_1)^2 = \frac{1}{3}$$

$$(\sigma_2)^2 = \frac{3}{4}$$

To get the power of the t-test and wilcoxon test, again we will calculate the proportion of the p-value that were < 0.05.

Calculations show that the power of the t-test is **0.8094** and the power of the Wilcox-test is **0.7612**. (See the appendix for the detailed R codes and calculations) .

3 Tests for Uniformly Distributed Populations

Now, we generate two sets of S datasets both from uniform distribution. The significance level is 0.05 again.

3.1 Means are equal(Case 3)

First, we evaluate the size of the two tests. For this case, we take means equal. So,

$$\mu_1 = \mu_2 = 1$$

We let the first population follow from uniform distribution, ($min = 0$,$max = 2$) and the second population follow uniform distribution, ($min = 0.5$, $max = 1.5$). We can calculate the variance for both populations according to the definition of variance. So for this case, we have the variances as:

$$(\sigma_1)^2 = \frac{1}{3}$$

$$(\sigma_2)^2 = \frac{1}{12}$$

Our null hypothesis is (Ho: $\mu 1 = \mu 2$)

Firstly, we use R to generate two samples from the uniform distribution with($min = 0$, $max = 2$) and the uniform distribution with ($min = 0.5$, $max = 1.5$). Then we evaluate the size of the t-test and the Wilcoxon test under the null hypothesis (Ho: $\mu 1 = \mu 2$). We calculate the proportion of rejections of Ho by determining the proportion of the p-value that were < 0.05. As a result of calculation, we get the size of the t-test as **0.0561** and the size of the Wilcox-test as **0.0554**. (For the results and codes, please see Appendix part). The size of the tests are very close to significance level, and thus tests are relevant. Now, we can calculate the power of the test.

3.2 Means are not equal (Case 4)

Now, we approximate the power of the two tests under an alternative hypothesis which says that μ_1 and μ_2 are not equal.

We reuse the first sample from uniform distribution of $(min = 0, max = 2)$ from Case 3 and let the second sample taken from uniform distribution of $(min = 0.5,\ max = 3.5)$. Means are different. And also, the variances of the two populations are not equal. For this case, we have:

$$\mu_1 = 1$$

$$\mu_3 = 2$$

$$(\sigma_1)^2 = \frac{1}{3}$$

$$(\sigma_2)^2 = \frac{3}{4}$$

To get the power of the t-test and Wilcox-test, we calculate the proportion of the p-value that were < 0.05 again. The results show that the power of the t-test is **0.8206** and the power of the Wilcox-test is **0.7179**. (See the appendix for the detailed R codes and calculations) .

4 Results

Table 1 shows the results from the cases of normally distributed populations. And Table 2 shows the results from the cases of uniformly distributed populations.

	normal distribution		t-test	Wilcoxon test
Mean equal(Case 1)	μ1=1, μ2=1, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_1)^2 = \frac{1}{12}$	size	0.0521	0.0507
Mean not equal(Case 2)	μ=1, μ12=2, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{3}{4}$	power	0.8094	0.7612

Table 1: **Size and Power of t-test and Wilcoxon test for uniform distribution populations**

	Uniform distribution		t-test	Wilcoxon test
Mean equal(Case 3)	μ1=1, μ2=1, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_1)^2 = \frac{1}{12}$	size	0.0561	0.0554
Mean not equal(case 4)	μ=1, μ2=2, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{3}{4}$	power	0.8206	0.7179

Table 2: **Size and Power of t-test and Wilcoxon test for uniform distribution populations**

From the results of both tests, it's clear that both sizes are very close to 0.05. This means that both of the sizes of t-test and wilcoxon test achieve the given significance level $\alpha = 0.05$.

Then we can compare the power of t-test and Wilcoxon test. We know that the power means the probability of rejecting the null hypothesis given the alternative hypothesis is true. For normally distributed population samples, the power of t-test and wilcoxon test is 0.8094 and 0.7612 respectively. So it can be concluded that the t-test is more powerful and safer than the wilcox test here. For uniformly distributed population samples, the power of the t-test is 0.8206 whereas the power of the Wilcox-test is 0.7179. Thus, t-test is also more powerful in this case.

5 Conclusion and analysis

			t-test	Wilcoxon test
normal distr.: (case 2)	μ=1, μ2=2, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{3}{4}$	power	0.8094	0.7612
Uniform distr.: (case 4)	μ=1, μ2=2, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{3}{4}$	power	0.8206	0.7179

Table 3: **Power comparison for normal distribution and uniform distribution t-test Wilcoxon test**

Table 3 shows the comparison of the power from normal distribution and uniform distribution. The results indicate that:

The power of the t-test is slightly larger than that of the wilcoxon test in both cases. This means that the t-test is slightly powerful than the wilcoxon test regardless of the nonmormality of data.

Both of the t-test and wilcoxon test give slightly smaller rejection proportion (power) when the data are uniformly distributed compared with the normal distribution cases.

In this study, we choose a small sample size n = 10. We may add the sample size as a new factor and consider the cases with large sample size to see how the size and power will differ for the t-test and wilcoxon test.

We may also consider the cases when one sample from normal distribution and the other one from uniform distribution. And in the future we may also study the power of t-test and Wilcoxon test under more alternative hypotheses.

6 Appendix: R code and the output

Here is the R code to generate the datasets and caculate the size and power of t-test and wilcoxon test.

6.1 Case 1

samples from normally distributed populations with equal mean but unequal variance. $\mu_1 = \mu_2 = 1$, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{1}{12}$

```
> S <- 10000
> n <- 10
> mu1 <- 1
> mu2 <- 1
> sigma1 <- sqrt(1/3)
> sigma2 <- sqrt(1/12)
> sampnorm1 <- matrix(rnorm(n * S, mu1, sigma1), byrow = S, ncol = n)
> sampnorm2 <- matrix(rnorm(n * S, mu2, sigma2), byrow = S, ncol = n)
> tp1 <- rep(0, S)
> for (i in 1:S) {
+     tp1[i] <- t.test(sampnorm1[i, ], sampnorm2[i, ], alternative = "two.sided"
+ }
> tsize1 <- sum(tp1 < 0.05)/S
> tsize1
```

```
[1] 0.0474
```

```
> wp1 <- rep(0, S)
> for (i in 1:S) {
+     wp1[i] <- wilcox.test(sampnorm1[i, ], sampnorm2[i, ], alternative = "two.s
+ }
> wsize1 <- sum(wp1 < 0.05)/S
> wsize1
```

```
[1] 0.0499
```

6.2 Case 2

samples from normally distributed populations with unequal mean and unequal variance. $\mu_1 = 1$, $\mu_2 = 2$, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{3}{4}$

```
> S <- 10000
> n <- 10
> mu1 <- 1
> mu2 <- 2
> sigma1 <- sqrt(1/3)
> sigma3 <- sqrt(3/4)
> sampnorm1 <- matrix(rnorm(n * S, mu1, sigma1), byrow = S, ncol = n)
> sampnorm3 <- matrix(rnorm(n * S, mu2, sigma3), byrow = S, ncol = n)
> tp2 <- rep(0, S)
> for (i in 1:S) {
+     tp2[i] <- t.test(sampnorm1[i, ], sampnorm3[i, ], alternative = "two.sided"
+ }
> tpower2 <- sum(tp2 < 0.
> tpower2 <- sum(tp2 < 0.05)/S
> tpower2
```

```
[1] 0.8004
```

```
> wp2 <- rep(0, S)
> for (i in 1:S) {
+     wp2[i] <- wilcox.test(sampnorm1[i, ], sampnorm3[i, ], alternative = "two.s
+ }
> wpower2 <- sum(wp2 < 0.05)/S
> wpower2
```

```
[1] 0.7649
```

6.3 Case 3

samples from uniformly distributed populations with equal mean and unequal variance. $\mu_1 = \mu_2 = 1$, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{1}{12}$

```
> S <- 10000
> n <- 10
> sampunif1 <- matrix(runif(n * S, min = 0, max = 2), nrow = S,
+     ncol = n)
> sampunif2 <- matrix(runif(n * S, min = 0.5, max = 1.5), nrow = S,
+     ncol = n)
> tp3 <- rep(0, S)
> for (i in 1:S) {
+     tp3[i] <- t.test(sampunif1[i, ], sampunif2[i, ], alternative = "two.sided"
+ }
> tsize3 <- sum(tp3 < 0.05)/S
> tsize3
```

```
[1] 0.0555
```

```
> wp3 <- rep(0, S)
> for (i in 1:S) {
+     wp3[i] <- wilcox.test(sampunif1[i, ], sampunif2[i, ], alternative = "two.s
+ }
> wsize3 <- sum(wp3 < 0.05)/S
> wsize3
```

```
[1] 0.0572
```

6.4 Case 4

samples from uniformly distributed populations with unequal mean and unequal variance. $\mu_1 = 1$, $\mu_2 = 2$, $(\sigma_1)^2 = \frac{1}{3}$, $(\sigma_2)^2 = \frac{3}{4}$

```
> S <- 10000
> n <- 10
> sampunif1 <- matrix(runif(n * S, min = 0, max = 2), nrow = S,
+     ncol = n)
> sampunif3 <- matrix(runif(n * S, min = 0.5, max = 3.5), nrow = S,
```

```
+     ncol = n)
> tp4 <- rep(0, S)
> for (i in 1:S) {
+     tp4[i] <- t.test(sampunif1[i, ], sampunif3[i, ], alternative = "two.sided"
+ }
> tpower4 <- sum(tp4 < 0.05)/S
> tpower4

[1] 0.8142

> wp4 <- rep(0, S)
> for (i in 1:S) {
+     wp4[i] <- wilcox.test(sampunif1[i, ], sampunif3[i, ], alternative = "two.s
+ }
> wpower4 <- sum(wp4 < 0.05)/S
> wpower4

[1] 0.7221
```