Rank-Based Tests for Clustered Data with R Package clusrank

Yujing Jiang
University of Connecticut

Xin He
University of Maryland

Mei-Ling Ting Lee
University of Maryland

Bernard Rosner
Harvard University

Jun Yan
University of Connecticut

Abstract

Nonparametric tests such as Wilcoxon rank sum test and Wilcoxon signed rank test are widely used in the situation where the underlying distribution of the population is far from normal or simply unknown. One necessary assumption for the appropriateness of the null distribution of the test statistic is that each observation is independent, however, this is also an assumption which is violated quite often in practice. For instance, in the study of human eyes, each person is the unit of randomization, whereas the data is collected from both eyes, therefore we sould expect correlation between the data collected from the two eyes from the same person. To account for the clustering effect, modifications of these two tests have been proposed by Rosner, Glynn, and Lee (2003) and Rosner, Glynn, and Lee (2006). The modified tests work for both balanced and unbalanced data, i.e., cluster size is idential or variable. In addition, the modified rank sum test can also deal with stratified data. No R package is available so far for nonparametric tests for clustered data. The package **clusrank** is a realization of the test procedures from the two papers mentioned above with both small and large sample tests.

Keywords: Wilcoxon rank sum test, Wilcoxon signed rank test.

1. Introduction

Clustered data often arise in biomedical studies, e.g., when the research objects are eyes, ears, teath, etc. In these cases, the observations can be classified into a number of distinct groups or "clusters", where the observations are more similar within each cluster than when they are from different clusters. For instance, when measurements are taken from both eyes of each patient, the measurments from the same person should be correlated whereas measurements taken from different patients should be independent. Quite often, the goal of study is to compare the measurements on the response variable from the control group and the treatment group or the measurements before and after a treatment to see if the treatment effect is present. The parametric statistical tests which explicitly account for clustering include an adjusted version of the standard two-sample t test which account for the intracluster correlation (i.e., correlation among obervations within the clusters) when the response variable follows a normal distribution and a adjusted χ^2 statistic when the response variable is binary (Donner and Banting 1988).

When the response variable is neither normal or binary, e.g., the observations are ordinal, the aforementioned tests are no longer suitable. An attractive alternative is non-parametric test. However, standard non-parametric tests requires the observations to be independent, which is violated when the data is clustered. If the intracluster correlation is positive, then the variance of the standard test statistic will be underestimated. Rosner *et al.* (2003) proposed a Wilcoxon rank sum test for clustered data to compare two groups assuming that the members within each cluster is exchangeble and objects from the same cluster belong to the same group.Their method is able to deal with unbalanced data (i.e., data with variable cluster sizes) and stratified data (e.g., data from a multi-center study which is stratified by centers). Another modified rank sum test for clustered data was proposed by Datta and Satten (2005), their method is valid even when the members of the same cluster came from different groups, or when the intraclulster correlation is determined by the group membership. To compare paired observations, an adjusted version of wilcoxon signed rank test was proposed by Larocque (2005) which invovles an estimate of variance based on certain sum of square over independent clusters, but this test procedure is not distribution free. Rosner *et al.* (2006) proposed another modified signed rank test assuming a common intercluster correlation and estimate it from absolute rank of observations. Datta and Satten (2008) proposed a signed rank test procedure based on the principle of general with-in cluster resampling and this test can handle the case when the cluster size is informative, i.e., cluster size depends on the group.

Despite the popularity of clustered data in a wide range of contexts such as clinical trials, longitudinal study, social science, etc, there is no available R package nor function publically available for non-parametric test for clustered data yet. Therefore a package pacted with functions for this purpose will help researchers and scientists in related areas as a ready tool for testing clustered data. In this paper we present the **clusrank** package, a realization of test procedures presented in Rosner *et al.* (2003) and Rosner *et al.* (2006). Large-sample inference based on asymptotic distribution of test statistics and small-sample inference based on permutation are provided for both rank sum test and signed rank test. The tests provided can also handel unbalanced data when there are clusters with different sizes, in addition, for rank sum test, the effect of stratification as an extra cofounding variable can also be accounted for.

The order of this paper is as following, section 2 introduces Wilcoxon rank sum test and Wilcoxon signed rank test for clustered data, when data is balanced or unbalanced. In addition, the modification of rank sum test when data is stratified is also discussed. Section 3 is a real data analysis for a eye study. The article is summarized in section 4.

2. Tests

2.1. Wilcoxan Rank Sum Test for Clustered Data

Hypothesis

Suppose there are two groups under different treatments, X and Y, the null hypothesis of the rank sum test is that the probability of an observation from a treatment X exceeding an observation from treatment Y is the same as an observation from treatment Y exceeding an observation from treatment X. Specifically, assume the data came in clusters and let X_{ij} denote the score for the jth subunit from the ith cluster in the first group, $i = 1, \ldots, m; j = 1, \ldots, g_i$ and Y_{kl} denote the score for the lth subunit from the kth cluster in the second group, $k = 1, \ldots, n; l = 1, \ldots, h_k$. The clustered Wilcoxon rank sum statistic $W_{c,obs}$ is defined as

$$W_{c,obs} = \sum_{i=1}^{m} \sum_{j=1}^{g_i} \text{Rank}(X_{ij}) \tag{1}$$

where ranks are determined based on the combined sample of all subunits over the X and Y clusters. Subunits for a given cluster are assumed to be exchangeable and each cluster only contains members from one treatment group.

Balanced data

Under balanced designs, the cluster sizes are the same. Since the score assigned to clusters under treatments X and Y are identically distributed under the null hypothesis, we can pool X and Y clusters together and refer to a combined set of Z clusters, where Z_{ij} =score for the jth subunit of the ith cluster, $j = 1, \ldots, g, i = 1, \ldots, m + n = N$. Under H_0, suppose that m of the N clusters are assigned at random to the X treatment and the remaining n clusters to the Y treatment. Let δ_i be an indicator function that equals 1 when the ith cluster is assigned to the X group and 0 when the ith cluster is assigned to the Y group. The distribution of the clustered rank sum statistic $W_{c,obs}$ is

$$W_c = \sum_{i=1}^{N} \delta_i R_{i+} \quad \text{where } R_{i+} = \sum_{j=1}^{g} R_{ij} \tag{2}$$

where R_{ij} =rank of the jth subunit in the ith cluster among all gN subunits over all Z clusters. It is shown that under H_0,

$$E(W_c) = gm(gN + 1)/2 \tag{3}$$

and

$$\mathrm{Var}(W_c) = [mn/\{N(N-1)\}] \sum_{i=1}^{N} \{R_{i+} - g(1 + gN)/2\}^2 \tag{4}$$

So a natural large sample test statistic based on (2), (3), and (4) is

$$Z_c = \{W_c - E(W_c)\}/\{\mathrm{Var}(W_c)\}^{1/2} \tag{5}$$

Z_c is asymptotically normal if both $m \to \infty$ and $n \to \infty$. When the sample size is small, permutation test can be performed.

Unbalanced data

For unbalanced designs, the null distribution of the sum of rank assigned to each cluster are not identically distributed across clusters with different sizes. Let (m_g, n_g) = number of clusters of size g assigned to the X and Y treatment respectively. Denote $N_g = m_g + n_g$ for $g = 1, \ldots, g_{max}$ as the total number of clusters for clusters with size g, and $N = \sum_{g=1}^{g_{max}} N_g$ as the total number of clusters in the sample. Let $R_{ij,g}$ = rank for the jth subunit in the ith cluster of size g, $g = 1, \ldots, g_{max}, i = 1, \ldots, N_g, j = 1, \ldots, g$, where $R_{ij,g}$s are computed based on the combined sample. The rank sum statistic can then be written as:

$$W_{c,obs} = \sum_{g=1}^{g_{max}} \sum_{i \in I_{g,obs}} R_{i+,g} \tag{6}$$

where $R_{i+,g}$ = sum of ranks of all subunits in the ith cluster of size g, $i = 1, \ldots, N_g$. The distribution corresponding to $W_{c,obs}$ is

$$W_c == \sum_{g=1}^{g_{max}} = \sum_{g=1}^{g_max} \sum_{i=1}^{N_g} \delta_{i,g} R_{i+,g} \tag{7}$$

where $\delta_{i,g} = 1$ if the ith cluster of size g is assigned to group X, and is 0 otherwise. The corresponding $E(W_c)$ and $\mathrm{Var}(W_c)$ are as following:

$$E(W_c) = \sum_{g=1}^{g_{max}} m_g (R_{++,g}/N_g) \tag{8}$$

$$\text{Var}(W_c) = \sum_{g=1}^{g_{max}} [m_g n_g / \{N_g(N_g - 1)\}] \sum_{i=1}^{N_g} (R_{i+,g} - R_{++,g}/N_g)^2 \tag{9}$$

A large-sample test statistics based on W_c is as following:

$$Z_c = \{W_c - E(W_c)\}/\{\text{Var}(W_c)\}^{1/2} \tag{10}$$

Again, when sample size is small, permutation test can be applied based on (7)

When applying the ranksum test for the clustered data, a formula is used as a interface where the response variable is on the righthand side and covariates on the lefthand side. Special functions `group`, `cluster` and `stratum` are used to indicate the function of each variable, `group` indicates the id of the treatment group, `cluster` indicates the membership of an observation in a specific cluster and `stratum` indicates the stratum an observation belongs to.

```
library(clusrank)
data(crd)
## Using large-sample test.
cluswilcox.test(z ~ group(group) + cluster(id), data = crd)

##
##  Wilcoxon rank sum test for clutered data
##
## data:  z from crd, cluster: id, group: group,
## Rank sum statistic = 19505
## Expected value of rank sum statistic = 20013
## Variance of rank sum statistic = 631030
## Test statistic = -0.63924, p-value = 0.5227
## difference in locations = 0
## The data is unbalanced
## alternative hypothesis: true difference in locations is not equal to 0

## For small sample, using the permutation test.
cluswilcox.test(z ~ group(group) + cluster(id), data = crd, permutation = TRUE)

##
##  Wilcoxon rank sum test for clutered data
##
## data:  z from crd, cluster: id, group: group,
## Rank sum statistic = 19505
## p-value = 0.556
## location = 0
## The data is unbalanced
## alternative hypothesis: true location is not equal to 0
```

Data with Stratification

Futher more, if the data is also stratified, the test statistic and its null distribution will need modification to control for stratification as an extra confounding variable. Suppose there are V strata, let $(m_{g,v}, n_{g,v})$ = number of clusters of size g in stratum v assigned to the$X$ and Y clusters of size g in stratum v respectively, $g = 1, \ldots, g_{max}, v = 1, \ldots, V$. Let $R_{i+,g,v}$ be the rank sum for the subunits in the ith cluster of size g in the vth stratum. The rank sum statistic is defined as

$$W_{c,obs} = \sum_{g=1}^{g_{max}} \sum_{v=1}^{V} \left(\sum_{i \in I_{g,v,obs}} R_{i+,g,v} \right) \tag{11}$$

where $I_{g,v,obs}$ is the observed subset of $m_{g,v}$ unique indices selected from $\{1, \ldots, N_{g,v}\}$, corresponding to cluster of size g in stratum v that are assigned to the X treatment. The corresponding expectation and variance of the null distribution of $W_{c,obs}$

$$\begin{aligned} E(W_c) &= \sum_{g=1}^{g_{max}} \sum_{v=1}^{V} m_{g,v} R_{++,g,v}/N_{g,v} \\ \mathrm{Var}(W_c) &= \sum_{g=1}^{g_{max}} \sum_{v=1}^{V} [m_{g,v} n_{g,v} / \{N_{g,v}(N_{g,v} - 1)\}] \\ &\quad \times \sum_{i=1}^{N_{g,v}} (R_{i+,g,v} - R_{++,g,v}/N_{g,v})^2 \end{aligned}$$

The large sample test statistic is then

$$Z_c = \{W_c - E(W_c)\}/\{\mathrm{Var}(W_c)\}^{1/2}. \tag{12}$$

Again, permutation test can be applied when sample size is small.

Following is an illustration of the rank sum test for the stratified data, the `crdStr` data is a test data set comes with the package:

```
data(crdStr)
cluswilcox.test(z ~ group(group) + cluster(id) + stratum(stratum), data = crdStr)

##
##  Wilcoxon rank sum test for clutered data
##
## data:  z from crdStr, stratum: stratum, cluster: id, group: group,
## Rank sum statistic = 222250
## Expected value of rank sum statistic = 227660
## Variance of rank sum statistic = 20338000
## Test statistic = -1.198, p-value = 0.2309
## difference in locations = 0
##
## alternative hypothesis: true difference in locations is not equal to 0
```

2.2. Wilcoxon signed rank test for clustered data

Hypothesis

Let $X_{ij}(Y_{ij})$ denotes the baseline (follow-up) score for the jth subunit in the ith cluster (subject) and define $Z_{ij} = Y_{ij} - X_{ij}, j = 1, \ldots, g_i; i = 1, \ldots, m$. Within each cluster, the difference scores are assumed to be independent and identically distributed. The signed rank test is used to find out if the population is shifted after the treatment. Formally, hypothesis being tested is

$$H_0 : \text{the difference score } Z \text{ is symmetric about } 0$$

vs

$$H_1 : Z \text{ is symmetric about } \gamma, \gamma \neq 0.$$

Rank $|Z_{ij}|$ over the total of $G = \sum_{i=1}^{m} g_i$ subunits from the m clusters and let $S_{ij} = R_{ij}V_{ij}$, where R_{ij} = rank of $|Z_{ij}|$ within the total data set of G subunits over m clusters, and $V_{ij} = \text{sign}(Z_{ij})$.

Balanced Data

If the data is balanced, the clustered Wilcoxon signed rank statistic is defined as following:

$$T_c^{(obs)} = \sum_{i=1}^{m} S_{i+} \equiv \sum_{i=1}^{m}\sum_{j=1}^{g} R_{ij}V_{ij}, \tag{13}$$

where $S_{i+} = \sum_{j=1}^{g} S_{ij}$ which is the sum of the rank within i^{th} cluster and only consider nonzero Z_{ij} in the computation of signed ranks. When considering the randomization distribution corresponding to T_c, the unit of randomization is the cluster. Let $\delta_1, \ldots, \delta_m$ be i.i.d. random variables each taking on the values +1 and -1 with probability 1/2, then the distribution of $T_c^{(obs)}$ is:

$$T_c = \sum_{i=1}^{m} \delta_i S_{i+}. \tag{14}$$

It is shown that under H_0, $E(T_c) = 0$ and $\text{Var}(T_c) = \sum_{i=1}^{m} S_{i+}^2$. Standarize $T_c^{(obs)}$ with $E(T_c)$ and $\text{Var}(T_c)$, the large sample test statistic is defined as

$$Z_c = T_c \Bigg/ \left(\sum_{i=1}^{n} S_{i+}^2\right)^{1/2} \sim N(0,1) \qquad \text{under } H_0. \tag{15}$$

When the sample size is small, a permutation test can be applied.

For the signed rank test, the input should be a numeric vector which contains the difference between the paired observations, or two vectors, where vector `x` contains observations before the treatment, and `y` contains observations after the treatment. This manner of input is restricted to the signed rank test.

```
## Large sample signed rank test for clustered data
data(crsd)
cluswilcox.test(z, cluster = id, data = crsd)

##
##  Wilcoxon signed rank test for clutered data
##
## data:  z, cluster: id from crsd
## rank statistic = -110
## Variance of  rank statistic = 30178
## test statistic = -0.63321, p-value = 0.5266
## total number of observations = 40, total number of clusters = 20
##
## alternative hypothesis: true location shift is not equal to 0
```

```
## Small sample test
data(crsd)
cluswilcox.test(z, cluster = id, data = crsd, permutation = TRUE)

##
##  Wilcoxon signed rank test for clutered data using permutation
##
## data:  z, cluster: id from crsd
## rank statistic = -110
## p-value = 0.444
## total number of observations = 40, total number of clusters = 20
##
## alternative hypothesis: true location shift is not equal to 0
```

Unbalanced Data

When the data is unbalanced the test statistic is defined as:

$$T_c^{(obs)} = \sum_{i=1}^{m} w_i \bar{S}_i \tag{16}$$

where $\bar{S}_i = S_{i+}/g_i, w_i = 1/\mathrm{Var}(\bar{S}_i)$ under H_0. The randomization distribution corresponding to $T_{c,s}^{obs} = \sum_{i=1}^{m} \delta_i w_i \bar{S}_i$. δ is defined as in (14). The test statistic is defined as

$$Z_{c,s} = T_{c,s} \Bigg/ \left(\sum_{i=1}^{m} \hat{w}_i^2 \bar{S}_i^2 \right)^{1/2} \sim N(0,1) \text{ under } H_0, \tag{17}$$

where $\hat{w}_i = g_i/[\widehat{Var}(S_{ij})\{1+(g_i-1)\hat{\rho}_{s,cor}\}]$, $\hat{\rho}_{s,cor} = \hat{\rho}_s \left(1 + \frac{1-\hat{\rho}_s^2}{m-5/2}\right)$,$\hat{\rho}_s = \max\,[\hat{\sigma}_A^2/(\hat{\sigma}_A^2+\hat{\sigma}^2), 0]$, $\hat{\sigma}^2 = \sum_{i=1}^{m}\sum_{j=1}^{g_i}(S_{ij} - \bar{S}_i)^2/(G-m), \hat{\sigma}_A^2 = \max\,[\{\sum_{i=1}^{m} g_i(\bar{S}_i - \bar{\bar{S}})^2/(m-1) - \hat{\sigma}^2\}/g_0, 0], g_0 = [\sum_{i=1}^{m} g_i - \sum_{i=1}^{m} g_i^2/\sum_{i=1}^{m} g_i]/(m-1)$ and $\widehat{Var}(S_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{g_i}(S_{ij} - \bar{\bar{S}})^2/(G-1)$.
An illustration of use of the test is as following:

```
data(crsdUnb)
cluswilcox.test(z, cluster = id, data = crsdUnb)

##
##  Wilcoxon signed rank test for clutered data
##
## data:  z, cluster: id from crsdUnb
## adjusted rank statistic = -0.015709
## Variance of  adjusted rank statistic = 0.00123
## test statistic = -0.44794, p-value = 0.6542
## total number of observations = 748, total number of clusters = 142
## The signed rank test statistics is adjusted since the data is unbalanced.
## alternative hypothesis: true location shift is not equal to 0
```

3. Real Data Analysis

3.1. Data

To illustrate the usage of the package, we are going to perform the clustered Wilcoxon rank sum test on a real data set in a study of eyes. Age-related macular degeneration (AMD) is a disease that blurs the sharp centeral vision by affecting macula, a oval yellow spot near the center of the retina of the human eye. The complement factor H R1210C is a large protein that circulates in human plasma. The variant of this protein confers the strongest genetic risk for AMD and earlier age at onset. The objective of the study is to characterize the observable traits of this variant. The study was carried out by the Seddon Lab (Seddon, Sharma, and Adelman 2006; Ferrara and Seddon 2015) with 143 patients (283 eyes) involved, including 62 patients with the rare variant. The degree of severity of AMD was graded based on the Clinical Age-Related Maculopathy Staging (CARMS) system for each enrolled eye. The CARMS system has a 5-step scale, where 1 to 3 represent no symptom, earlier and intermediate severity respectively, 4 and 5 represent two different symptoms of the advanced AMD, geographic atrophy and neovascular disease respectively. The CARMS grades were assessed separately for these two advanced stages, i.e., we are going to carry out the analysis on two subsets of the observations respectively: the subset of observations with CARMS grade 1, 2, 3, or 4, and the subset of observations with CARMS grade 1, 2, 3, or 5. Since high correlation between eyes of the same patient is expected while observations from different patients can be assumed as independent, the data is clustered in pairs and each subject is a cluster. The data also provided information on age and sex, and an extra variable combined age and sex, which could be used to stratify the data. In this analysis, the CARMS grade is the response variable and treatment refers to the presence of the complement factor H R1210C variant. The data set contains 7 variables: ID is the subject identifier which is the cluster id.

```
## Carry out clustered rank sum test for the subset
## with CARMS grade 1, 2, 3 and 4.
data(sedlab)
cluswilcox.test(CARMS ~ cluster(ID) + stratum(Agesex) + group(Variant),
                data = sedlab, subset = CARMS %in% c(1, 2, 3, 4))

##
##  Wilcoxon rank sum test for clutered data
##
## data:  CARMS from sedlab, stratum: Agesex, cluster: ID, group: Variant,
## Rank sum statistic = 8792
## Expected value of rank sum statistic = 10867
## Variance of rank sum statistic = 258650
## Test statistic = -4.0797, p-value = 4.509e-05
## difference in locations = 0
## The data is unbalanced
## alternative hypothesis: true difference in locations is not equal to 0
```

```
## Carry out clustered rank sum test for the subset
## with CARMS grade 1, 2, 3 and 5.
data(sedlab)
cluswilcox.test(CARMS ~ cluster(ID) + stratum(Agesex) + group(Variant),
                data = sedlab, subset = CARMS %in% c(1, 2, 3, 5))

##
##  Wilcoxon rank sum test for clutered data
```

```
##
## data:  CARMS from sedlab, stratum: Agesex, cluster: ID, group: Variant,
## Rank sum statistic = 14502
## Expected value of rank sum statistic = 15583
## Variance of rank sum statistic = 340860
## Test statistic = -1.8519, p-value = 0.06404
## difference in locations = 0
## The data is unbalanced
## alternative hypothesis: true difference in locations is not equal to 0
```

When controlled for the stratification covariate `Agesex`, the p-value for CARMS grades 1-4 is less than 0.001, which implies strong correlation between the presence of the complement factor H R1210C variant and the severity of AMD when treating geographic atrophy as the advanced stage. The p-value for CARMS grades 1,2,3 and 5 is 0.06, again the evidence is relatively strong the the presence of the variant does affect the severity of AMD when treating neovascular disease as the advanced stage.

4. Summary

In this artical, Wilcoxon rank sum test and Wilcoxon signed rank test adjusted for the clustering effect in the data are introduced. The usage of the R package **clusrank** carrying out the two tests is illustrated with examples. Both the tests are able to handel unbalanced data. In addition, the clustered Wilcoxon rank sum test also permits an extra covariate as stratification variable.

Acknowledgement

Data **sedlab** is provided by Seddon Lab (Seddon *et al.* 2006; Ferrara and Seddon 2015).

References

Datta S o, Satten GA (2005). "Rank-Sum Tests for Clustered Data." *Journal of the American Statistical Association*, **100**(471), 908–915. doi:10.1198/016214504000001583.

Datta S, Satten GA (2008). "A Signed-Rank Test for Clustered Data." *Biometrics*, **64**(2), 501–507. ISSN 1541-0420. doi:10.1111/j.1541-0420.2007.00923.x. URL http://dx.doi.org/10.1111/j.1541-0420.2007.00923.x.

Donner A, Banting D (1988). "Analysis of Site-specific Data in Dental Studies." *Journal of Dental Research*, **67**(11), 1392–1395. doi:10.1177/00220345880670110601.

Ferrara D, Seddon JM (2015). "Phenotypic Characterization of Complement Factor H R1210C Rare Genetic Variant in Age-Related Macular Degeneration." *JAMA Ophthalmology*, **133**(7), 785–791. doi:10.1001/jamaophthalmol.2015.0814. /data/Journals/OPHTH/934175/eoi150027.pdf, URL +http://dx.doi.org/10.1001/jamaophthalmol.2015.0814.

Larocque D (2005). "The Wilcoxon Signed-Rank Test for Cluster Correlated Data." In P Duchesne, B RÉMillard (eds.), *Statistical Modeling and Analysis for Complex Data Problems*, pp. 309–323. Springer US. ISBN 978-0-387-24554-6. doi:10.1007/0-387-24555-3_15. URL http://dx.doi.org/10.1007/0-387-24555-3_15.

Rosner B, Glynn RJ, Lee MLT (2003). "Incorporation of Clustering Effects for the Wilcoxon Rank Sum Test: A Large-Sample Approach." *Biometrics*, **59**(4), 1089–1098.

Rosner B, Glynn RJ, Lee MLT (2006). "The Wilcoxon Signed Rank Test for Paired Comparisons of Clustered Data." *Biometrics*, **62**(1), 185–192.

Seddon JM, Sharma S, Adelman RA (2006). "Evaluation of the Clinical Age-related Maculopathy Staging System." *Ophthalmology*, **113**(2), 260–266.

Affiliation:

Jun Yan
Department of Statistics
Professor of Statistics
University of Connecticut
215 Glenbrook Rd. Unit 4120, Storrs, CT 06269, USA.

Email:jun.yan@uconn.edu