The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)

Thorsten Pohlert

latest revision: 2015-11-19

 Please cite this package as:

T. Pohlert (2014). *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR).* R package. http://CRAN.R-project.org/package=PMCMR.

See also `citation("PMCMR")`.

Contents

1 Introduction

For one-factorial designs with samples that do not meet the assumptions for one-way-ANOVA (i.e., i) errors are normally distributed, ii) equal variances among the groups, and, iii) uncorrelated errors) and subsequent post-hoc tests, the Kruskal-Wallis test (`kruskal.test`) can be employed that is also referred to as the Kruskal–Wallis one-way analysis of variance by ranks. Provided that significant differences were detected by the Kruskal-Wallis-Test, one may be interested in applying post-hoc tests for pairwise multiple comparisons of the ranked data (Nemenyi's test, Dunn's test, Conover's test). Similarly, one-way ANOVA with repeated measures that is also referred to as ANOVA with unreplicated block design can also be conducted via the Friedman test (`friedman.test`). The consequent post-hoc pairwise multiple comparison tests according to Nemenyi and Conover are also provided in this package.

2 Comparison of multiple independent samples (One-factorial design)

2.1 Kruskal and Wallis test

The linear model of a one-way layout can be written as:

$$y_i = \mu + \alpha_i + \epsilon_i, \tag{1}$$

with y the response vector, μ the global mean of the data, α_i the difference to the mean of the i-th group and ϵ the residual error. The non-parametric alternative is the Kruskal and Wallis test. It tests the null hypothesis, that each of the k samples belong to the same population ($H_0 : \bar{R}_{i.} = (n+1)/2$). First, the response vector y is transformed into ranks with increasing order. In the presence of sequences with equal values (i.e. ties), mean ranks are designated to the corresponding realizations. Then, the test statistic can be calculated according to Eq. 2:

$$\hat{H} = \left[\frac{12}{n\,(n+1)}\right] \left[\sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3\,(n+1) \tag{2}$$

with $n = \sum_i^k n_i$ the total sample size, n_i the number of data of the i-th group and R_i^2 the squared rank sum of the i-th group. In the presence of many ties, the test statistics $\hat{H}$ can be corrected using Eqs. 3 and 4

$$C = 1 - \frac{\sum_{i=1}^{i=r} \left(t_i^3 - t_i\right)}{n^3 - n}, \tag{3}$$

with t_i the number of ties of the i-th group of ties.

$$\hat{H}^* = \hat{H}/C \tag{4}$$

The Kruskal and Wallis test can be employed as a global test. As the test statistic $\bar{H}$ is approximately χ^2-distributed, the null hypothesis is withdrawn, if $\hat{H} > \chi^2_{k-1;\alpha}$. It should be noted, that the tie correction has only a small impact on the calculated statistic and its consequent estimation of levels of significance.

2.2 Kruskal-Wallis – post-hoc tests after Nemenyi

Provided, that the globally conducted Kruskal-Wallis test indicates significance (i.e. H_0 is rejected, and H_A : 'at least on of the k samples does not belong to the same population' is accepted), one may be interested in identifying which group or groups are significantly different. The number of pairwise contrasts or subsequent tests that need to be conducted is $m = k\,(k-1)\,/2$ to detect the differences between each group. Nemenyi proposed a test that originally based on rank sums and the application of the *family-wise error* method to control Type I error inflation, if multiple comparisons are done. The Tukey and Kramer approach uses mean rank sums and can be employed for equally as well as unequally sized samples without ties (Sachs, 1997, p. 397). The null hypothesis $H_0 : \bar{R}_i = \bar{R}_j$ is rejected, if a critical absolute difference of mean rank sums is exceeded.

$$|\bar{R}_i - \bar{R}_j| > \frac{q_{\infty;k;\alpha}}{\sqrt{2}} \sqrt{\left[\frac{n\,(n+1)}{12}\right]\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}, \tag{5}$$

where $q_{\infty;k;\alpha}$ denotes the upper quantile of the studentized range distribution. Although these quantiles can not be computed analytically, as *df* $= \infty$, a good approximation is to set *df* very large: such as $q_{1000000;k;\alpha} \sim q_{\infty;k;\alpha}$. This inequality (5) leads to the same critical differences of rank sums ($|R_i - R_j|$) when multiplied with n for $\alpha = [0.1, 0.5, 0.01]$, as reported in the tables of Wilcoxon and Wilcox (1964, pp. 29–31). In the presence of ties the approach presented by Sachs (1997, p. 395) can be employed (6), provided that $(n_i, n_j, \ldots, n_k \geq 6)$ and $k \geq 4$.:

$$|\bar{R}_i - \bar{R}_j| > \sqrt{C\chi^2_{k-1;\alpha}\left[\frac{n\,(n+1)}{12}\right]\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}, \tag{6}$$

where C is given by Eq. 3. The function `posthoc.kruskal.nemenyi.test` does not evaluate the critical differences as given by Eqs. 5 and 6, but calculates the corresponding level of significance for the estimated statistics q and χ^2, respectively.

In the special case, that several treatments shall only be tested against one control experiment, the number of tests reduces to $m = k - 1$. This case is given in section 2.8.

2.3 Examples using `posthoc.kruskal.nemenyi.test`

The function `kruskal.test` is provided with the library `stats` (R Core Team, 2013). The data-set `InsectSprays` was derived from a one factorial experimental design and can be used for demonstration purposes. Prior to the test, a visualization of the data (Fig 1) might be helpful:

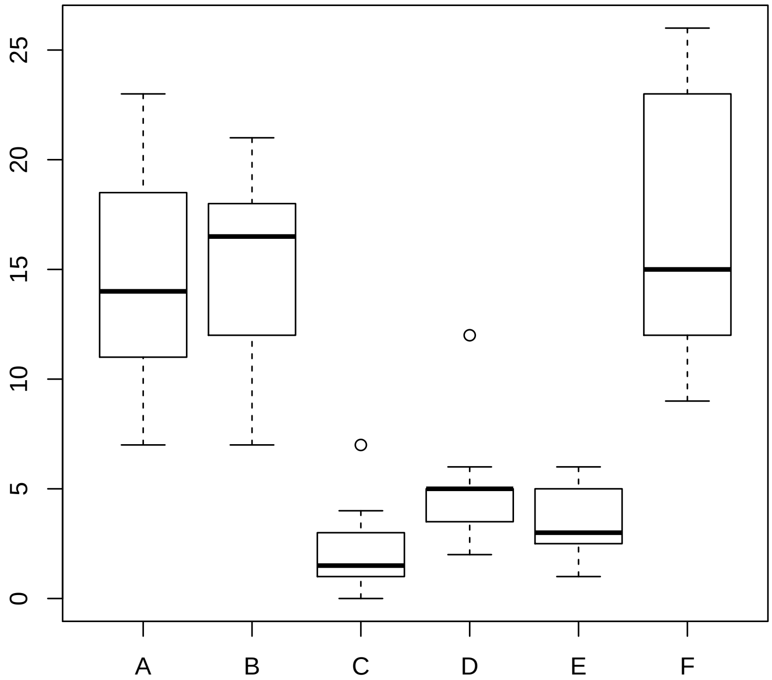

Figure 1: Boxplot of the `InsectSprays` data set.

Based on a visual inspection, one can assume that the insecticides *A, B, F* differ from *C, D, E*. The global test can be conducted in this way:

```
> kruskal.test(count ~ spray, data=InsectSprays)

        Kruskal-Wallis rank sum test

data:  count by spray
Kruskal-Wallis chi-squared = 54.6913, df = 5, p-value = 1.511e-10
```

As the Kruskal-Wallis Test statistics is highly significant ($\chi^2(5) = 54.69, p < 0.01$), the null hypothesis is rejected. Thus, it is meaningful to apply post-hoc tests with the function `posthoc.kruskal.nemenyi.test`.

```
> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> posthoc.kruskal.nemenyi.test(x=count, g=spray, dist="Tukey")

        Pairwise comparisons using Tukey and Kramer (Nemenyi) test
                   with Tukey-Dist approximation for independent samples

data:  count and spray

  A       B       C       D       E
B 0.99961 -       -       -       -
C 2.8e-05 5.7e-06 -       -       -
D 0.02293 0.00813 0.56300 -       -
E 0.00169 0.00047 0.94109 0.97809 -
F 0.99861 1.00000 3.5e-06 0.00585 0.00031

P value adjustment method: none
```

The test returns the lower triangle of the matrix that contains the p-values of the pairwise comparisons. Thus $|\bar{R}_A - \bar{R}_B| : n.s.$, but $|\bar{R}_A - \bar{R}_C| : p < 0.01$. Since PMCMR-1.1 there is a formula method included. Thus the test can also be conducted in the following way:

```
> posthoc.kruskal.nemenyi.test(count ~ spray, data=InsectSprays, dist="Tukey")

        Pairwise comparisons using Tukey and Kramer (Nemenyi) test
                   with Tukey-Dist approximation for independent samples

data:  count by spray

  A       B       C       D       E
B 0.99961 -       -       -       -
C 2.8e-05 5.7e-06 -       -       -
D 0.02293 0.00813 0.56300 -       -
E 0.00169 0.00047 0.94109 0.97809 -
F 0.99861 1.00000 3.5e-06 0.00585 0.00031

P value adjustment method: none
```

As there are ties present in the data, one may also conduct the Chi-square approach:

```
> (out <- posthoc.kruskal.nemenyi.test(x=count, g=spray, dist="Chisquare"))

        Pairwise comparisons using Nemenyi-test with Chi-squared
                       approximation for independent samples

data:  count and spray

  A       B       C       D       E
B 0.99985 -       -       -       -
C 0.00037 9.4e-05 -       -       -
D 0.08359 0.03812 0.73938 -       -
E 0.01113 0.00391 0.97354 0.99070 -
F 0.99945 1.00000 6.2e-05 0.02955 0.00281

P value adjustment method: none
```

which leads to different levels of significance, but to the same test decision. The arguments of the returned object of class `pairwise.h.test` can be further explored. The statistics, in this case the χ^2 estimations, can be taken in this way:

```
> print(out$statistic)

            A            B          C          D        E
B  0.09741248           NA         NA         NA       NA
C 22.70093702 25.772474315         NA         NA       NA
D  9.68046043 11.720034247  2.7330908         NA       NA
E 14.76750381 17.263698630  0.8495291  0.5351027       NA
F  0.16383657  0.008585426 26.7218417 12.3630375 18.04226
```

The test results can be aligned into a summary table as it is common in scientific articles. However, there is not yet a function included in the package PMCMR. Therefore, Table 1 was manually created.

2.4 Kruskal-Wallis – post-hoc test after Dunn

Dunn (1964) has proposed a test for multiple comparisons of rank sums based on the z-statistics of the standard normal distribution. The null hypothesis ($H0$), the probability of observing a randomly selected value from the first group that is larger than a randomly selected value from the second group equals one half, is rejected, if a critical absolute difference of mean rank sums is exceeded:

$$|\bar{R}_i - \bar{R}_j| > z_{1-\alpha/2*} \sqrt{\left[\frac{n\,(n+1)}{12} - B\right]\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}, \tag{7}$$

with $z_{1-\alpha/2*}$ the value of the standard normal distribution for a given adjusted $\alpha/2*$ level depending on the number of tests conducted and B the correction term for ties, which was taken from Glantz (2012) and is given by Eq. 8:

$$B = \frac{\sum_{i=1}^{i=r} \left(t_i^3 - t_i\right)}{12\,(n-1)} \tag{8}$$

The function `posthoc.kruskal.dunn.test` does not evaluate the critical differences as given by Eqs. 7, but calculates the corresponding level of significance for the estimated statistics z, as adjusted by any method implemented in `p.adjust` to account for Type I error inflation. It should be noted that Dunn (1964) originally used a Bonferroni adjustment of p-values. For this specific case, the test ist sometimes referred as to the Bonferroni-Dunn test.

2.5 Example using `posthoc.kruskal.dunn.test`

We can go back to the example with InsectSprays.

```
> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> posthoc.kruskal.dunn.test(x=count, g=spray, p.adjust.method="none")

        Pairwise comparisons using Dunn's-test for multiple
                         comparisons of independent samples

data:  count and spray

  A       B       C       D       E
B 0.75448 -       -       -       -
C 1.8e-06 3.6e-07 -       -       -
D 0.00182 0.00060 0.09762 -       -
```

Table 1: Mean rank sums of insect counts ($\bar{R}_i$) after the application of insecticides (Group). Different letters indicate significant differences ($p < 0.05$) according to the Tukey-Kramer-Nemenyi post-hoc test. The global test according to Kruskal and Wallis indicated significance ($\chi^2(5) = 54.69, p < 0.01$).

Group	$\bar{R}_i$	
C	11.46	a
E	19.33	a
D	25.58	a
A	52.17	b
B	54.83	b
F	55.62	b

```
E 0.00012 3.1e-05 0.35572 0.46358 -
F 0.68505 0.92603 2.2e-07 0.00043 2.1e-05

P value adjustment method: none
```

The test returns the lower triangle of the matrix that contains the p-values of the pairwise comparisons. As `p.adjust.method="none"`, the p-values are not adjusted. Hence, there is a Type I error inflation that leads to a false postive discovery rate. This can be solved by applying e.g. a Bonferroni-type adjustment of p-values.

```
> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> posthoc.kruskal.dunn.test(x=count, g=spray, p.adjust.method="bonferroni")

        Pairwise comparisons using Dunn's-test for multiple
                         comparisons of independent samples

data:  count and spray

  A       B       C       D       E
B 1.00000 -       -       -       -
C 2.7e-05 5.5e-06 -       -       -
D 0.02735 0.00904 1.00000 -       -
E 0.00177 0.00047 1.00000 1.00000 -
F 1.00000 1.00000 3.3e-06 0.00640 0.00031

P value adjustment method: bonferroni
```

2.6 Kruskal-Wallis – post-hoc test after Conover

Conover and Iman (1979) proposed a test that aims at having a higher test power than the tests given by inequalities 5 and 6:

$$\left|\bar{R}_i - \bar{R}_j\right| > t_{1-\alpha/2;n-k} \sqrt{s^2 \left[\frac{n-1-\hat{H}^*}{n-k}\right] \left[\frac{1}{n_i} + \frac{1}{n_j}\right]}, \tag{9}$$

with $\hat{H}^*$ the tie corrected Kruskal-Wallis statistic according to Eq. 4 and $t_{1-\alpha/2;n-k}$ the t-value of the student-t-distribution. The variance s^2 is given in case of ties by:

$$s^2 = \frac{1}{n-1} \left[\sum R_i^2 - n\frac{(n+1)^2}{4}\right] \tag{10}$$

The variance s^2 simplifies to $n\,(n+1)\,/12$, if there are no ties present. Although Conover and Iman (1979) did not propose an adjustment of p-values, the function `posthoc.kruskal.conover.test` has implemented methods for p-adjustment from the function `p.adjust`.

2.7 Example using `posthoc.kruskal.conover.test`

```
> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> posthoc.kruskal.conover.test(x=count, g=spray, p.adjust.method="none")

        Pairwise comparisons using Conover's-test for multiple
                         comparisons of independent samples

data:  count and spray

  A       B       C       D       E
B 0.5314  -       -       -       -
C 3.7e-14 3.0e-15 -       -       -
D 3.1e-08 2.4e-09 0.0014  -       -
E 7.5e-11 5.6e-12 0.0676  0.1451  -
F 0.4175  0.8524  1.4e-15 1.1e-09 2.6e-12

P value adjustment method: none

> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> posthoc.kruskal.conover.test(x=count, g=spray, p.adjust.method="bonferroni")

        Pairwise comparisons using Conover's-test for multiple
                         comparisons of independent samples

data:  count and spray

  A       B       C       D       E
B 1.000   -       -       -       -
C 5.6e-13 4.5e-14 -       -       -
D 4.7e-07 3.6e-08 0.021   -       -
E 1.1e-09 8.5e-11 1.000   1.000   -
F 1.000   1.000   2.1e-14 1.7e-08 3.9e-11

P value adjustment method: bonferroni
```

2.8 Dunn's multiple comparison test with one control

Dunn's test (see section 2.4), can also be applied for multiple comparisons with one control (Siegel and Castellan Jr., 1988):

$$|\bar{R}_0 - \bar{R}_j| > z_{1-\alpha/2*} \sqrt{\left[\frac{n\,(n+1)}{12} - B\right]\left[\frac{1}{n_0} + \frac{1}{n_j}\right]}, \tag{11}$$

where $\bar{R}_0$ denotes the mean rank sum of the control experiment. In this case the number of tests is reduced to $m = k - 1$, which changes the p-adjustment according to Bonferroni (or others). The function `dunn.test.control` employs this test, but **the user need to be sure that the control is given as the first level in the group vector**.

2.9 Example using `dunn.test.control`

We can use the `PlantGrowth` dataset, that comprises data with dry matter weight of yields with one control experiment (i.e. no treatment) and to different treatments. In this case we are only interested, whether the treatments differ significantly from the control experiment.

```
> require(stats)
> data(PlantGrowth)
> attach(PlantGrowth)
> kruskal.test(weight, group)

        Kruskal-Wallis rank sum test

data:  weight and group
Kruskal-Wallis chi-squared = 7.9882, df = 2, p-value = 0.01842

> dunn.test.control(x=weight,g=group, p.adjust="bonferroni")

        Pairwise comparisons using Dunn's-test for multiple
                         comparisons with one control

data:  weight and group

     ctrl
trt1 0.53
trt2 0.18

P value adjustment method: bonferroni
```

According to the global Kruskal-Wallis test, there are significant differences between the groups, $\chi^2(2) = 7.99, p < 0.05$. However, the Dunn-test with Bonferroni adjustment of p-values shows, that there are no significant effects.

If one may cross-check the findings with ANOVA and multiple comparison with one control using the LSD-test, he/she can do the following:

```
> summary.lm(aov(weight ~ group))

Call:
aov(formula = weight ~ group)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.0710 -0.4180 -0.0060  0.2627  1.3690 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.0320     0.1971  25.527   <2e-16 ***
grouptrt1    -0.3710     0.2788  -1.331   0.1944    
grouptrt2     0.4940     0.2788   1.772   0.0877 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6234 on 27 degrees of freedom
Multiple R-squared:  0.2641,       Adjusted R-squared:  0.2096 
F-statistic: 4.846 on 2 and 27 DF,  p-value: 0.01591
```

The last line provides the statistics for the global test, i.e. there is a significant treatment effect according to one-way ANOVA, $F(2, 27) = 4.85, p < 0.05, \eta^2 = 0.264$. The row that starts with `Intercept` gives the group mean of the control, its standard error, the t-value for testing $H0 : \mu = 0$ and the corresponding level of significance. The following lines provide the difference between the averages of the treatment groups with the control, where $H0 : \mu_0 - \mu_j = 0$. Thus the `trt1` does not differ significantly from the `ctr`, $t = -1.331, p = 0.194$. There is a significant difference between `trt2` and `ctr` as indicated by $t = 1.772, p < 0.1$.

2.10 van der Waerden test

The van der Waerden test can be used as an alternative to the Kruskal-Wallis test, if the data to not meet the requirements for ANOVA (Conover and Iman, 1979). Let the Kruskal-Wallis ranked data denote $R_{i,j}$, then the normal scores $A_{i,j}$ are derived from the standard normal distribution according to Eq. 12.

$$A_{i,j} = \phi^{-1}\left(\frac{R_{i,j}}{n+1}\right) \tag{12}$$

Let the sum of the i-th score denote A_j. The variance S^2 is calculated as given in Eq. 13.

$$S^2 = \frac{1}{n-1} \sum A_{i,j}^2 \tag{13}$$

Finally the test statistic is given by Eq. 14.

$$T = \frac{1}{S^2} \sum_{j=1}^{k} \frac{A_j^2}{n_j} \tag{14}$$

The test statistic T is approximately χ^2-distributed and tested for significance on a level of $1 - \alpha$ with $dg = k - 1$.

2.11 Example using `vanWaerden.test`

```
> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> vanWaerden.test(x=count, g=spray)

        Van der Waerden normal scores test

data:  count and spray
Van der Waerden chi-squared = 50.302, df = 5, p-value = 1.202e-09
```

2.12 post-hoc test after van der Waerden for multiple pairwise comparisons

Provided that the global test according to van der Waerden indicates significance, multiple comparisons can be done according to the inequality 15.

$$\| \frac{A_i}{n_i} - \frac{A_j}{n_j} \| > t_{1-\alpha/2;n-k} \sqrt{S^2 \;\; \frac{n-1-T}{n-k} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \tag{15}$$

The test given in Conover and Iman (1979) does not adjust p-values. However, the function has included the methods for p-value adjustment as given by `p.adjust`.

2.13 Example using `posthoc.vanWaerden.test`

```
> require(PMCMR)
> data(InsectSprays)
> attach(InsectSprays)
> posthoc.vanWaerden.test(x=count, g=spray, p.adjust.method="none")

        Pairwise comparisons using van der Waerden normal scores test for
               multiple comparisons of independent samples
```

```
data:  count and spray

  A       B       C       D       E
B 0.6366  -       -       -       -
C 6.9e-12 9.8e-13 -       -       -
D 9.0e-06 1.5e-06 0.0008  -       -
E 5.5e-08 8.1e-09 0.0316  0.1919  -
F 0.2323  0.4675  5.0e-14 8.6e-08 4.1e-10

P value adjustment method: none
```

3 Comparison of multiple joint samples (Two-factorial unreplicated complete block design)

3.1 Friedman test

The linear model of a two factorial unreplicated complete block design can be written as:

$$y_{i,j} = \mu + \alpha_i + \pi_j + \epsilon_{i,j} \tag{16}$$

with π_j the j-th level of the block (e.g. the specific response of the j-th test person). The Friedman test is the non-parametric alternative for this type of k dependent treatment groups with equal sample sizes. The null hypothesis, $H0 : F(1) = F(2) = \ldots = F(k)$ is tested against the alternative hypothesis: at least one group does not belong to the same population. The response vector y has to be ranked in ascending order separately for each block $\pi_j : j = 1, \ldots m$. After that, the statistics of the Friedman test is calculated according to Eq. 17:

$$\hat{\chi}_R^2 = \left[\frac{12}{nk\,(k+1)} \sum_{i=1}^{k} R_i\right] - 3n\,(k+1) \tag{17}$$

The Friedman statistic is approximately χ^2-distributed and the null hypothesis is rejected, if $\hat{\chi}_R > \chi^2_{k-1;\alpha}$.

3.2 Friedman – post-hoc test after Nemenyi

Provided that the Friedman test indicates significance, the post-hoc test according to Nemenyi (1963) can be employed (Sachs, 1997, p. 668). This test requires a balanced design ($n_1 = n_2 = \ldots = n_k = n$) for each group k and a Friedman-type ranking of the data. The inequality 18 was taken from Demsar (2006, p. 11), where the critical difference refer to mean rank sums ($|\bar{R}_i - \bar{R}_j|$):

$$|\bar{R}_i - \bar{R}_j| > \frac{q_{\infty;k;\alpha}}{\sqrt{2}} \sqrt{\frac{k\,(k+1)}{6n}} \tag{18}$$

This inequality leads to the same critical differences of rank sums ($|R_i - R_j|$) when multiplied with n for $\alpha = [0.1, 0.5, 0.01]$, as reported in the tables of Wilcoxon and Wilcox (1964, pp. 36–38). Likewise to the `posthoc.kruskal.nemenyi.test` the function `posthoc.friedman.nemenyi.test` calculates the corresponding levels of significance and the generic function `print` depicts the lower triangle of the matrix that contains these p-values. The test according to Nemenyi (1963) was developed to account for a family-wise error and is already a conservative test. This is the reason, why there is no p-adjustment included in the function.

3.3 Example using `posthoc.friedman.nemenyi.test`

This example is taken from Sachs (1997, p. 675) and is also included in the help page of the function `posthoc.friedman.nemenyi.test`. In this experiment, six persons (block) subsequently received six different diuretics (groups) that are denoted A to F. The responses are the concentration of Na in urine measured two hours after each treatment.

```
> require(PMCMR)
> y <- matrix(c(
+ 3.88, 5.64, 5.76, 4.25, 5.91, 4.33, 30.58, 30.14, 16.92,
+ 23.19, 26.74, 10.91, 25.24, 33.52, 25.45, 18.85, 20.45,
+ 26.67, 4.44, 7.94, 4.04, 4.4, 4.23, 4.36, 29.41, 30.72,
+ 32.92, 28.23, 23.35, 12, 38.87, 33.12, 39.15, 28.06, 38.23,
+ 26.65),nrow=6, ncol=6,
+ dimnames=list(1:6,c("A","B","C","D","E","F")))
> print(y)
```

```
     A     B     C    D     E     F
1 3.88 30.58 25.24 4.44 29.41 38.87
2 5.64 30.14 33.52 7.94 30.72 33.12
3 5.76 16.92 25.45 4.04 32.92 39.15
4 4.25 23.19 18.85 4.40 28.23 28.06
5 5.91 26.74 20.45 4.23 23.35 38.23
6 4.33 10.91 26.67 4.36 12.00 26.65
```

Based on a visual inspection (Fig. 2), one may assume different responses of Na-concentration in urine as related to the applied diuretics.

The global test is the Friedman test, that is already implemented in the package `stats` (R Core Team, 2013):

```
> friedman.test(y)
```

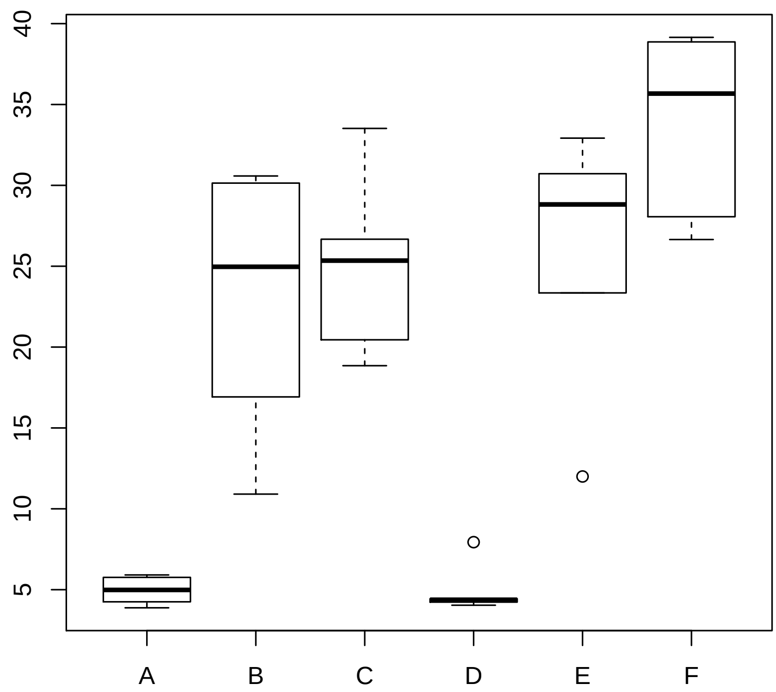

Figure 2: Na-concentration (mval) in urine of six test persons after treatment with six different diuretics.

```
        Friedman rank sum test

data:  y
Friedman chi-squared = 23.3333, df = 5, p-value = 0.0002915
```

As the Friedman test indicates significance ($\chi^2(5) = 23.3, p < 0.01$), it is meaningful to conduct multiple comparisons in order to identify differences between the diuretics.

```
> posthoc.friedman.nemenyi.test(y)

        Pairwise comparisons using Nemenyi multiple comparison test
             with q approximation for unreplicated blocked data

data:  y
```

```
  A      B      C      D      E
B 0.1880 -      -      -      -
C 0.0917 0.9996 -      -      -
D 0.9996 0.3388 0.1880 -      -
E 0.0395 0.9898 0.9996 0.0917 -
F 0.0016 0.6363 0.8200 0.0052 0.9400

P value adjustment method: none
```

According to the Nemenyi post-hoc test for multiple joint samples, the treatment F based on the Na diuresis differs highly significant ($p < 0.01$) to A and D, and E differs significantly ($p < 0.05$) to A. Other contrasts are not significant ($p > 0.05$). This is the same test decision as given by (Sachs, 1997, p. 675).

3.4 Friedman – post-hoc test after Conover

Conover (1999) proposed a post-hoc test for pairwise comparisons, if Friedman-Test indicated significance. The absolute difference between two group rank sums are signifcantly different, if the following inequality is satisfied:

$$|R_i - R_j| > t_{1-\alpha/2;(n-1)(k-1)} \sqrt{\frac{2k\left(1 - \frac{\hat{\chi}^2_R}{n(k-1)}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{k} R_{i,j}^2 - \frac{nk(k+1)^2}{4}\right)}{(k-1)\,(n-1)}} \tag{19}$$

Although Conover (1999) originally did not include a p-adjustment, the function has included the methods as given by `p.adjust`, because it is a very liberal test. So it is up to the user, to apply a p-adjustment or not, when using this function.

3.5 Example using `posthoc.friedman.conover.test`

```
> require(PMCMR)
> y <- matrix(c(
+ 3.88, 5.64, 5.76, 4.25, 5.91, 4.33, 30.58, 30.14, 16.92,
+ 23.19, 26.74, 10.91, 25.24, 33.52, 25.45, 18.85, 20.45,
+ 26.67, 4.44, 7.94, 4.04, 4.4, 4.23, 4.36, 29.41, 30.72,
+ 32.92, 28.23, 23.35, 12, 38.87, 33.12, 39.15, 28.06, 38.23,
+ 26.65),nrow=6, ncol=6,
+ dimnames=list(1:6,c("A","B","C","D","E","F")))
> friedman.test(y)

        Friedman rank sum test

data:  y
Friedman chi-squared = 23.3333, df = 5, p-value = 0.0002915
```

```
> posthoc.friedman.conover.test(y=y, p.adjust="none")

        Pairwise comparisons using Conover's multiple comparison test
                     for unreplicated blocked data

data:  y

  A       B       C       D       E
B 0.00014 -       -       -       -
C 3.0e-05 0.55547 -       -       -
D 0.55547 0.00067 0.00014 -       -
E 6.5e-06 0.24321 0.55547 3.0e-05 -
F 8.0e-08 0.00621 0.02468 3.3e-07 0.08511

P value adjustment method: none
```

References

Conover WJ (1999). *Practical Nonparametric Statistics.* 3rd edition. Wiley.

Conover WJ, Iman RL (1979). "On multiple-comparison procedures." *Technical report*, Los Alamos Scientific Laboratory.

Demsar J (2006). "Statistical comparisons of classifiers over multiple data sets." *Journal of Machine Learning Research*, **7**, 1–30.

Dunn OJ (1964). "Multiple comparisons using rank sums." *Technometrics*, **6**, 241–252.

Glantz SA (2012). *Primer of biostatistics.* 7 edition. McGraw Hill, New York.

Nemenyi P (1963). *Distribution-free Multiple Comparisons.* Ph.D. thesis, Princeton University.

R Core Team (2013). *R: A language and environement for statistical computing.* Vienna, Austria. URL `http://www.R-project.org/`.

Sachs L (1997). *Angewandte Statistik.* 8 edition. Springer, Berlin.

Siegel S, Castellan Jr NJ (1988). *Nonparametric Statistics for The Behavioral Sciences.* 2nd edition. McGraw-Hill, New York.

Wilcoxon F, Wilcox RA (1964). *Some rapid approximate statistical procedures.* Lederle Laboratories, Pearl River.