

# onewaytests: An R Package for One-Way Tests in Independent Groups Designs

by Osman Dag, Anil Dolgun and Naime Meric Konar

**Abstract** One-way tests in independent groups designs are the most commonly utilized statistical methods with applications on the experiments in medical sciences, pharmaceutical research, agriculture, biology, engineering, social sciences and so on. In this paper, we present the **onewaytests** package to investigate treatment effects on the dependent variable. The package offers the one-way tests in independent groups designs, which include ANOVA, Welch's heteroscedastic  $F$  test, Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances, Brown-Forsythe test, Alexander-Govern test, James second order test and Kruskal-Wallis test. The package also provides pairwise comparisons, graphical approaches, and assesses variance homogeneity and normality of data in each group via tests and plots. A simulation study is also conducted to give recommendations for applied researchers on the selection of appropriate one-way tests under assumption violations. Furthermore, especially for non-R users, a user-friendly web application of the package is provided. This application is available at <http://www.softmed.hacettepe.edu.tr/onewaytests>.

## Introduction

There are many statistical procedures for one-way independent groups designs. The one-way fixed effects analysis of variance (ANOVA) is the most common one being used for comparing  $k$  independent group means. Independence of the observations in each group and between groups, normality (i.e.,  $k$  samples come at random from normal distribution) and variance homogeneity (i.e.,  $k$  populations have equal variances) are three basic assumptions of ANOVA. ANOVA is a valid and powerful test for identifying group differences provided that these assumptions are held. However, most of the data sets in practice are not normally distributed and group variances are heterogeneous, and it is difficult to find a data set that satisfies both assumptions. When the assumptions underlying ANOVA are violated, the drawn inferences are also invalid. There have been remarkable efforts to find an appropriate and robust test statistic under non-normality and variance heterogeneity. Several procedures alternative to ANOVA were proposed, including Welch's heteroscedastic  $F$  test (Welch, 1951), Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances (Welch, 1951), Alexander-Govern test (Alexander and Govern, 1994), James second order test (James, 1951, 1954), Brown-Forsythe test (Brown and Forsythe, 1974a,b) and Kruskal-Wallis test (Kruskal and Wallis, 1952).

In the literature, there are many studies in which the comparison of ANOVA and its alternative tests was made with respect to type I error rate and power when the assumptions of ANOVA are not satisfied. The effects of sample size (balanced/unbalanced case), non-normality, unequal variances, and combined effects of non-normality and unequal variances were investigated in details (Wilcox, 1988; Cribbie et al., 2012; Lantz, 2013; Gamage and Weerahandi, 1998; Parra-Frutos, 2013; Cribbie et al., 2007). In the light of these studies, we conduct a Monte Carlo simulation study in an attempt to give recommendations for applied researchers on the selection of appropriate one-way tests.

ANOVA, Welch's heteroscedastic  $F$  test, Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances, Kruskal-Wallis test, and Brown-Forsythe test are available under some packages (given in Table 1) on the Comprehensive R Archive Network (CRAN). Alexander-Govern test and James second order test are also found to be robust to assumption violations, but have been overlooked as their calculation and implementation were not easily available elsewhere.

In this paper, we introduce an R package, **onewaytests** (Dag et al., 2017) which implements Alexander-Govern test and James second order test, in addition to the ANOVA, Welch's heteroscedastic  $F$  test, Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances, Kruskal-Wallis test, and Brown-Forsythe test. Besides the omnibus tests, the package provides pairwise comparisons to investigate which groups create the difference providing that a statistically significant difference is obtained by the omnibus test. Moreover, the package offers several graphic types such as grouped box-and-whisker plot and error bar graph. Also, it provides statistical tests and plots (i.e. Q-Q plot and histogram) to assess variance homogeneity and normality, in the package version 1.5 and later. The **onewaytests** package is publicly available on the CRAN.

The organization of this paper is presented as follows. First, we give the theoretical background on the one-way tests in independent groups designs. Second, we introduce the **onewaytests** package and demonstrate the applicability of the package using two real-life datasets. Third, the web interface of the **onewaytests** package is introduced. A Monte Carlo simulation study is also conducted to give recommendations for applied researchers on the selection of appropriate tests included in the package.

Results of this simulation study and the general conclusions on the effects of assumption violations are mentioned.

### One-way tests in independent groups designs

In this section, the statistical tests that are used to test the equality of several populations in the one-way independent groups designs are explained. Among these tests, ANOVA, Welch’s heteroscedastic  $F$  test (Welch, 1951), Alexander-Govern test (Alexander and Govern, 1994), James second order test (James, 1951, 1954) and Brown-Forsythe test (Brown and Forsythe, 1974a,b) test the null hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$  versus alternative  $H_1 : \text{at least one } \mu_j (j = 1, 2, \dots, k) \text{ is different}$ . Welch’s heteroscedastic  $F$  test with trimmed means and Winsorized variances (Welch, 1951) tests the equality of population trimmed means,  $H_0 : \mu_{t1} = \mu_{t2} = \dots = \mu_{tk}$  versus alternative  $H_1 : \text{at least one } \mu_{tj} \text{ is different}$ , where  $\mu_{tj}$  represents the trimmed mean of the  $j$ th population ( $j = 1, 2, \dots, k$ ). Kruskal-Wallis test (Kruskal and Wallis, 1952) tests the null hypothesis  $H_0 : \theta_1 = \theta_2 = \dots = \theta_k$  versus alternative  $H_1 : \text{at least one } \theta_j (j = 1, 2, \dots, k) \text{ is different}$ . For the Kruskal-Wallis test,  $\theta_j$  represents sum of the ranks of the  $j$ th population.

### ANOVA

The one-way fixed effects analysis of variance (ANOVA) is the most common statistical procedure to test the equality of  $k$  independent population means. Underlying assumptions associated with ANOVA include homogeneity of group variances, normality of distributions and statistical independence of errors. Under these conditions, the ANOVA test statistic,

$$F = \frac{\sum_j n_j (\bar{X}_j - \bar{X}_{..})^2 / (k - 1)}{\sum_i \sum_j (X_{ij} - \bar{X}_j)^2 / (N - k)} \tag{1}$$

follows an  $F$  distribution with  $k - 1$  degrees of freedom for the numerator and  $N - k$  degrees of freedom for the denominator. In Equation (1),  $k$  is the number of groups,  $N$  is the total number of observations, and  $n_j$  is the number of observations in the  $j$ th group.  $X_{ij}$  is the  $i$ th observation ( $i = 1, 2, \dots, n_j$ ) in the  $j$ th group ( $j = 1, 2, \dots, k$ ),  $\sum_j n_j = N$ ,  $\bar{X}_{..}$  is the overall mean, where  $\bar{X}_{..} = \sum_j n_j (\bar{X}_j) / N$  and  $\bar{X}_j$  is sample mean for the  $j$ th group, where  $\bar{X}_j = \sum_i X_{ij} / n_j$ . ANOVA is more powerful if the assumptions of normality and variance homogeneity hold true. Non-normality has minimal effect on the type I error when the variances are equal (Lantz, 2013) but when the variances are not equal, ANOVA provides poor control over the type I and type II error rates (Bishop, 1976).

### Welch’s heteroscedastic $F$ test

Welch (1951) proposed a heteroscedastic alternative to ANOVA that is robust to the violation of variance homogeneity assumption. Under the null hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ , the test statistic  $F_w$ ,

$$F_w = \frac{\sum_j w_j (\bar{X}_j - X'_{..})^2 / (k - 1)}{\left[ 1 + \frac{2}{3} ((k - 2)\nu) \right]} \tag{2}$$

follows an  $F$  distribution with degrees of freedom  $k - 1$  for the numerator and  $1/\nu$  degrees of freedom for the denominator. In Equation (2),  $w_j = n_j / S_j^2$ ,  $S_j^2 = \sum_i (X_{ij} - \bar{X}_j)^2 / (n_j - 1)$ ,

$$X'_{..} = \frac{\sum_j w_j \bar{X}_j}{\sum_j w_j},$$

and

$$\nu = \frac{3 \sum_j \left[ \left( 1 - \frac{w_j}{\sum_j w_j} \right)^2 / (n_j - 1) \right]}{k^2 - 1}.$$

Welch’s heteroscedastic  $F$  test is robust and less sensitive to heteroscedasticity when compared to the ANOVA as within groups variance is based on the relationship between the different sample sizes in the different groups instead of a simple pooled variance estimate (Lantz, 2013).

**Welch’s heteroscedastic  $F$  test with trimmed means and Winsorized variances**

Welch’s heteroscedastic  $F$  test with trimmed means and Winsorized variances (Welch, 1951) is a robust procedure that tests the equality of means by substituting trimmed means and Winsorized variances for the usual means and variances. Therefore, this test statistic is relatively insensitive to the combined effects of non-normality and variance heterogeneity (Keselman et al., 2008). Let  $X_{(1)j} \leq X_{(2)j} \leq \dots \leq X_{(n_j)j}$  be the ordered observations in the  $j$ th group and  $g_j = \lfloor \epsilon n_j \rfloor$ ,  $\epsilon$  is the proportion to be trimmed in each tail of the distribution. After trimming, the effective sample size for the  $j$ th group becomes  $h_j = n_j - 2g_j$ . Then  $j$ th sample trimmed mean is

$$\bar{X}_{tj} = \frac{1}{h_j} \sum_{i=g_j+1}^{n_j-g_j} X_{(i)j},$$

and  $j$ th sample Winsorized mean is

$$\bar{X}_{wj} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij},$$

where

$$Y_{ij} = \begin{cases} X_{(g_j+1)j} & \text{if } X_{ij} \leq X_{(g_j+1)j} \\ X_{ij} & \text{if } X_{(g_j+1)j} < X_{ij} < X_{(n_j-g_j)j} \\ X_{(n_j-g_j)j} & \text{if } X_{ij} \geq X_{(n_j-g_j)j}. \end{cases}$$

The sample Winsorized variance is

$$s_{wj}^2 = \frac{1}{(n_j - 1)} \sum_{i=1}^{n_j} (Y_{ij} - \bar{X}_{wj})^2.$$

Let

$$q_j = \frac{(n_j - 1)s_{wj}^2}{h_j(h_j - 1)},$$

$$w_j = \frac{1}{q_j},$$

$$U = \sum_j w_j,$$

$$\bar{X} = \frac{1}{U} \sum_j w_j \bar{X}_{tj},$$

$$A = \frac{1}{k - 1} \sum_j w_j (\bar{X}_{tj} - \bar{X})^2,$$

$$B = \frac{2(k - 2)}{k^2 - 1} \sum_j \frac{(1 - w_j/U)^2}{h_j - 1}.$$

Under  $H_0 : \mu_{t1} = \mu_{t2} = \dots = \mu_{tk}$ , Welch’s heteroscedastic  $F$  test with trimmed means and Winsorized variances  $F_{wt}$ ,

$$F_{wt} = \frac{A}{B + 1} \tag{3}$$

follows an approximately  $F$  distribution with  $k - 1$  and  $\nu'$  degrees of freedom, where  $\nu'$  is

$$\nu' = \left( \frac{3}{k^2 - 1} \sum_j \frac{(1 - w_j/U)^2}{h_j - 1} \right)^{-1}.$$

$F_{wt}$  is not only less sensitive to heteroscedasticity and non-normality but also robust to the negative effects of outliers as it utilizes the trimmed means and Winsorised variances (Keselman et al., 2008).

**Brown-Forsythe test**

Brown and Forsythe (1974a,b) proposed the following test statistic:

$$F_{BF} = \frac{\sum_j n_j (\bar{X}_j - \bar{X}_{..})^2}{\sum_j (1 - n_j/N) S_j^2} \tag{4}$$

Under the null hypothesis,  $F_{BF}$  statistic has an approximately  $F$  distribution with  $k - 1$  and  $f$  degrees of freedom, where  $f$  is obtained with

$$f = \left( \sum_j c_j^2 / (n_j - 1) \right)^{-1},$$

and

$$c_j = \frac{(1 - n_j/N) S_j^2}{\left[ \sum_j (1 - n_j/N) S_j^2 \right]}.$$

$F_{BF}$  is a modification of ANOVA which has the same numerators as the ANOVA but an altered denominator. This test is more powerful when some variances appear unusually low (Brown and Forsythe, 1974a).

**Alexander-Govern test**

Alexander and Govern (1994) presented another robust test that is alternative to ANOVA, and it is used when group variances are not homogeneous. The test statistic for Alexander-Govern test is

$$\chi_{AG}^2 = \sum_{j=1}^k z_j^2. \tag{5}$$

Under the null hypothesis,  $\chi_{AG}^2$  is distributed as  $\chi^2$  distribution with  $k - 1$  degrees of freedom. In Equation (5),

$$z_j = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855c)}{(10b^2 + 8bc^4 + 1000b)},$$

where  $c = [\alpha \times \ln(1 + t_j^2/v_j)]^{1/2}$ ,  $b = 48\alpha^2$ ,  $\alpha = v_j - 0.5$  and  $v_j = n_j - 1$ . The  $t$  statistic for each group is

$$t_j = \frac{\bar{X}_j - X^+}{S'_j}. \tag{6}$$

The variance-weighted mean is  $X^+ = \sum_{j=1}^k w_j \bar{X}_j$  and the standard error for the  $j$ th group is

$$S'_j = \left[ \frac{\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2}{n_j(n_j - 1)} \right]^{1/2}.$$

Also, weights for each group are defined as

$$w_j = \frac{1/S_j'^2}{\sum_{j=1}^k 1/S_j'^2}. \tag{7}$$

$\chi_{AG}^2$  is another modification of ANOVA under heterogeneity of variance using the normalizing transformation of the one-sample  $t$  statistic. This test provides a good control of type I and II error rates for normally distributed data, but it is not robust to non-normality (Myers, 1998).

**James second order test**

An alternative test to ANOVA was proposed by James (1951). This test statistic ( $J$ ) is

$$J = \sum_j t_j^2, \tag{8}$$

where  $t_j$  is given in Equation (6). The test statistic,  $J$ , is compared to a critical value,  $h(\alpha)$ , where

$$\begin{aligned}
 h(\alpha) = & r + \frac{1}{2}(3\chi_4 + \chi_2)T + \frac{1}{16}(3\chi_4 + \chi_2)^2 \left(1 - \frac{k-3}{r}\right) T^2 \\
 & + \frac{1}{2}(3\chi_4 + \chi_2)(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) \\
 & + (2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) \\
 & + \frac{1}{4}(-R_{12}^2 + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2)(3\chi_4 - 2\chi_2 - 1) \\
 & + (R_{23} - 3R_{22} + 3R_{21} - R_{20})(5\chi_6 + 2\chi_4 + \chi_2) \\
 & + \frac{3}{16}(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) \\
 & + \frac{1}{16}(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2)(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) \\
 & + \frac{1}{4}(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2) + \frac{1}{4}(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2).
 \end{aligned}$$

For any integers  $s$  and  $t$ ,  $R_{st} = \sum_j (n_j - 1)^{-s} w_j^t$ ,  $\chi_{2s} = r^s / [(k - 1)(k + 1) \dots (k + 2s - 3)]$  where  $r$  is the  $(1 - \alpha)$  centile of a  $\chi^2$  distribution with  $k - 1$  degrees of freedom,  $T = \sum_j (1 - w_j)^2 / (n_j - 1)$  and  $w_j$  is defined in Equation (7). If the test statistic,  $J$ , exceeds  $h(\alpha)$ , then the null hypothesis is rejected.

The James second order test has been acknowledged as the best option for both normal heteroscedastic data (Alexander and Govern, 1994) and non-normal symmetric heteroscedastic data (Oshima and Algina, 1992). The disadvantage of this method is the complexity of the computation of critical values.

### Kruskal-Wallis test

Kruskal and Wallis (1952) proposed the nonparametric alternative to ANOVA. Let  $r_{ij}$  denote the rank of  $X_{ij}$  when  $N = n_1 + \dots + n_k$  observations are ranked from smallest to largest.  $R_j = \sum_{i=1}^{n_j} r_{ij}$  is the sum of ranks assigned to the observations in the  $j$ th group and  $\bar{R}_j = R_j / n_j$  is the average rank for these observations. Under these definitions, the Kruskal-Wallis test statistic is given by

$$\chi_{KW}^2 = \frac{1}{S^2} \left( \sum_{j=1}^k \frac{R_j^2}{n_j} - \frac{N(N+1)^2}{4} \right), \tag{9}$$

where

$$S^2 = \frac{1}{N-1} \left( \sum_{j=1}^k \sum_{i=1}^{n_j} r_{ij}^2 - \frac{N(N+1)^2}{4} \right).$$

When  $n_j \rightarrow \infty$ ,  $\chi_{KW}^2$  has an asymptotic  $\chi^2$  distribution under the null hypothesis  $H_0 : \theta_1 = \theta_2 = \dots = \theta_k$ . We reject  $H_0$  if  $\chi_{KW}^2 \geq \chi_{k-1, \alpha}^2$  where  $\chi_{k-1, \alpha}^2$  is the upper  $\alpha$  percentile for the  $\chi^2$  distribution with  $k - 1$  degrees of freedom. Note that, when there are no ties,  $S^2$  simplifies to  $N(N + 1) / 12$ .

The Kruskal-Wallis test is robust to non-normality as it utilizes ranks instead of actual values. However, it assumes that the observations in each group come from populations with the same shape of distribution. Therefore, if different groups have different shapes (e.g., some are skewed to the right and another is skewed to the left), the Kruskal-Wallis test may give inaccurate results (Fagerland and Sandvik, 2009).

There are other R packages including one-way tests in independent groups designs; namely, **stats** (R Core Team, 2017), **lawstat** (Gastwirth et al., 2017), **coin** (Hothorn et al., 2006), **car** (Fox and Weisberg, 2011), **WRS2** (Mair et al., 2017), **welchADF** (Villacorta, 2017). The **aov.test** (**onewaytests**) can be seen as a simplified version of anova (**stats**) and Anova (**car**), but the latter two are more flexible and comprehensive in a way that they can handle two or more variables (e.g., two-way ANOVA, multivariate ANOVA, repeated measures ANOVA, etc.). Amongst the alternatives, the **onewaytests** is the first package including seven different one-way independent design tests that are covered in one package. A brief comparison between these packages and **onewaytests** is given in Table 1. In Table 1, the numbers in parentheses indicate the rankings of the functions in terms of computational speed and the checks indicate the availability of the specific one-way design test in the packages. The

**onewaytests** package has the fastest running times except for Welch's heteroscedastic  $F$  test, where the **stats**'s function elapse time is 0.001 seconds faster. Note that the computational time is system-specific and may vary according to the number of iterations and the inclusion of additional arguments in the functions.

**Table 1:** Comparison of R packages including one-way independent design tests

Test	stats	car	lawstat	welchADF	WRS2	coin	onewaytests
$F$	(2)✓	(3)✓					(1)✓
$F_w$	(1)✓			(4)✓	(3)✓		(2)✓
$F_{wt}$				(3)✓	(2)✓		(1)✓
$F_{BF}^2$			(2)✓				(1)✓
$\chi_{AG}^2$							✓
$J$							✓
$\chi_{KW}^2$	(2)✓					(3)✓	(1)✓

## Demonstration of the onewaytests package

The **onewaytests** package includes several functions specially designed for one-way independent groups designs along with functions for assessing fundamental assumptions via relevant tests and plots. In this section, we demonstrate the usage of **onewaytests** package by using two different data sets from different fields.

### Iris data

In this part, we work with iris data set, collected by [Anderson \(1935\)](#), available in R. [Fisher \(1936\)](#) introduced this data set as an application of the linear discriminant analysis. At present, the popularity of the data set still continues within a variety of studies; examples include the studies related to data mining ([Hu, 2005](#)), multivariate normality assessment ([Korkmaz et al., 2014](#)), and so on.

This data set includes iris flowers' length and width measurements of sepal and petal characteristics in centimeters along with a grouping variable indicating the type of iris flowers (setosa, virginica and versicolor). For illustrating the implementation of our package, we use sepal length measurements as a response variable and iris types as a grouping variable. This data set has a total of 150 observations (50 observations in each type of iris flowers).

After installing and loading **onewaytests** package, the functions designed for one-way independent groups designs are available to be used. One-way tests in this package generally give test statistics and p-values to decide on the hypothesis of the statistical process, except for James second order test. In this test, test statistic and critical value are given as an output since the asymptotic distribution of the test statistic is not available.

The **onewaytests** package is also able to give some basic descriptive statistics of the given groups.

```
# obtain some basic descriptive statistics
R> describe(Sepal.Length ~ Species, data = iris)

           n Mean  Std.Dev Median Min Max 25th 75th Skewness Kurtosis NA
setosa    50 5.006 0.3524897   5.0 4.3 5.8 4.800  5.2 0.1164539 2.654235  0
versicolor 50 5.936 0.5161711   5.9 4.9 7.0 5.600  6.3 0.1021896 2.401173  0
virginica  50 6.588 0.6358796   6.5 4.9 7.9 6.225  6.9 0.1144447 2.912058  0
```

For all functions in the **onewaytests** package, the argument must be specified as a formula in which left and right-hand sides of the formula give sample values and its corresponding groups, respectively. The left and right-hand sides of the formula must have one variable. The variable on the left must be a numeric while the variable on the right must be a factor. Otherwise, each function returns an error message.

The functions for one-way design tests are coded from scratch. The function for pairwise comparison is coded for pairwise comparison of one-way independent groups designs by using the p-value adjustment methods available in the `p.adjust` function under **stats** package. For checking assumptions, the tests are used from different packages (**stats**, **car** and **nortest** ([Gross and Ligges, 2015](#))) and adapted for one-way independent groups designs. The function for the graphics is coded from scratch using the **ggplot2** package ([Wickham, 2009](#)).

## One-way tests in independent groups designs

The `onewaytests` package includes seven one-way tests in independent groups designs. In this part, the implementations of these tests are presented.

### ANOVA: `aov.test(...)`

The `aov.test` is utilized to test the equality of  $k$  independent population means.

```
R> aov.test(Sepal.Length ~ Species, data = iris, alpha = 0.05, na.rm = TRUE,
  verbose = TRUE)
```

```
One-Way Analysis of Variance
-----
data : Sepal.Length and Species

statistic : 119.2645
num df    : 2
denom df  : 147
p.value   : 1.669669e-31

Result    : Difference is statistically significant.
-----
```

Here, the statistic is the ANOVA test statistic distributed as  $F$  with the degrees of freedom for the numerator (num df) and denominator (denom df). Also, `p.value` is the significance value of the test statistic. Since the `p.value`, derived from `aov.test`, is lower than 0.05, it can be concluded that there is statistically significant difference between the iris species ( $F = 119.2645$ ,  $p\text{-value} = 1.669669 \times 10^{-31}$ ).

`alpha` is the level of significance to assess the statistical difference. Default is set to `alpha = 0.05`. `na.rm` is a logical value indicating whether NA values should be stripped before the computation proceeds. Default is `na.rm = TRUE`. `verbose` is a logical for printing output to R console. Default is set to `verbose = TRUE`. These arguments are available in the functions for one-way tests and checking assumptions. The users who would like to use the statistics in the output in their programs can use the following codes.

```
R> result <- aov.test(Sepal.Length ~ Species, data = iris, alpha = 0.05, na.rm = TRUE,
  verbose = FALSE)

# the ANOVA test statistic
R> result$statistic
[1] 119.2645

# the degrees of freedom for numerator and denominator
R> result$parameter
[1] 2 147

# the p-value of the test
R> result$p.value
[1] 1.669669e-31
```

Here, the codes for how to obtain the statistics from the ANOVA output are given. Since all one-way design tests return similar outputs, similar codes are not repeated in the other tests. For all tests, the level of significance is taken as 0.05.

### Welch's heteroscedastic $F$ test: `welch.test(...)`

One may use the `welch.test` function in the `onewaytests` package to perform Welch's heteroscedastic  $F$  test.

```
R> welch.test(Sepal.Length ~ Species, data = iris)
```

```
Welch's Heteroscedastic F Test
-----
data : Sepal.Length and Species
```

```

statistic : 138.9083
num df    : 2
denom df  : 92.21115
p.value   : 1.505059e-28

```

```
Result    : Difference is statistically significant.
```

In the output, similar to `aov.test`, the statistic is the Welch's test statistic distributed as F with the degrees of freedom for the numerator (`num df`) and denominator (`denom df`). One can conclude that the difference between the species is statistically significant ( $F_w = 138.9083$ ,  $p\text{-value} = 1.505059 \times 10^{-28}$ ).

### Welch's heteroscedastic F test with trimmed means and Winsorized variances: `welch.test(...)`

The `welch.test` function in the `onewaytests` package is also used to perform Welch's heteroscedastic F test with trimmed means and Winsorized variances.

```
R> welch.test(Sepal.Length ~ Species, data = iris, rate = 0.1)
```

```
Welch's Heteroscedastic F Test with Trimmed Means and Winsorized Variances
```

```
-----
data : Sepal.Length and Species
```

```

statistic : 123.6698
num df    : 2
denom df  : 71.64145
p.value   : 5.84327e-24

```

```
Result    : Difference is statistically significant.
```

Here, the statistic is the test statistic distributed as F with the degrees of freedom for the numerator (`num df`) and denominator (`denom df`). Moreover, the `rate` is the rate of observations trimmed and Winsorized from each tail of the distribution. If `rate = 0`, it performs Welch's heteroscedastic F test. Otherwise, one can perform Welch's heteroscedastic F test with trimmed means and Winsorized variances. Default is set to `rate = 0`. One can conclude that there is a statistically significant difference between the species ( $F_{wt} = 123.6698$ ,  $p\text{-value} = 5.84327 \times 10^{-24}$ ).

### Brown-Forsythe test: `bf.test(...)`

One may use the `bf.test` function in the `onewaytests` package to perform Brown-Forsythe test to compare more than two groups.

```
R> bf.test(Sepal.Length ~ Species, data = iris)
```

```
Brown-Forsythe Test
```

```
-----
data : Sepal.Length and Species
```

```

statistic : 119.2645
num df    : 2
denom df  : 123.9255
p.value   : 1.317059e-29

```

```
Result    : Difference is statistically significant.
```

In the output, the statistic is the Brown-Forsythe test statistic distributed as F with the degrees of freedom for numerator (`num df`) and denominator (`denom df`). It can be concluded that there is a statistically significant difference between the iris species ( $F_{BF} = 119.2645$ ,  $p\text{-value} = 1.317059 \times 10^{-29}$ ).

### Alexander-Govern test: `ag.test(...)`

The `ag.test` function in the `onewaytests` package is used to perform Alexander-Govern test.



```
R> ag.test(Sepal.Length ~ Species, data = iris)
```

```
Alexander-Govern Test
```

```
-----
data : Sepal.Length and Species
```

```
statistic : 146.3573
parameter : 2
p.value   : 1.655451e-32
```

```
Result    : Difference is statistically significant.
-----
```

Here, statistic is the Alexander-Govern test statistic distributed as  $\chi^2$  with the degrees of freedom (parameter). One can conclude that the difference between the species is statistically significant ( $\chi^2_{AG} = 146.3573$ , p-value =  $1.655451 \times 10^{-32}$ ).

### James second order test: `james.test(...)`

One may use the `james.test` function in the **onewaytests** package to perform James second order test.

```
R> james.test(Sepal.Length ~ Species, data = iris, alpha = 0.05)
```

```
James Second Order Test
```

```
-----
data : Sepal.Length and Species
```

```
statistic      : 279.8251
criticalValue  : 6.233185
```

```
Result        : Difference is statistically significant.
-----
```

Here, alpha is the significance level, statistic is the James second order test statistic,  $J$ , `criticalValue` is the critical value,  $h(\alpha)$ , corresponding to the significance level,  $\alpha$ . If  $J$  exceeds  $h(\alpha)$ , then the null hypothesis is rejected. Since  $J = 279.8251$ , obtained from `james.test`, is higher than  $h(\alpha) = 6.233185$ , it can be concluded that there is a statistically significant difference between the iris species.

### Kruskal-Wallis test: `kw.test(...)`

The `kw.test` function in the **onewaytests** package is utilized to perform Kruskal-Wallis test.

```
R> kw.test(Sepal.Length ~ Species, data = iris)
```

```
Kruskal-Wallis Test
```

```
-----
data : Sepal.Length and Species
```

```
statistic : 96.93744
parameter : 2
p.value   : 8.918734e-22
```

```
Result    : Difference is statistically significant.
-----
```

In the output, statistic is the Kruskal-Wallis test statistic distributed as  $\chi^2$  with the degrees of freedom (parameter). One can conclude that the difference between the species is statistically significant ( $\chi^2_{KW} = 96.93744$ , p-value =  $8.918734 \times 10^{-22}$ ).

### Pairwise comparisons

In this part, we present the pairwise comparisons to investigate which groups create the difference. We utilized the `p.adjust` function under the **stats** package (R Core Team, 2017) for our `paircomp`

function. The `paircomp` function has also the same p-value adjustment methods as `p.adjust`, including `bonferroni`, `holm` (Holm, 1979), `hochberg` (Hochberg, 1988), `hommel` (Hommel, 1988), `BH` (Benjamini and Hochberg, 1995), `BY` (Benjamini and Yekutieli, 2001), and `none`. The default is set to "bonferroni". Pairwise comparisons are made by adjusting p-values according to the specified method, except when the method is James second order test, which requires adjusting the significance level instead of the p-value.

One-way tests return a list with class "owt" except for James second order test. The reason for returning a list with class "owt" is to introduce the output to the `paircomp` function for pairwise comparison adjusting p-values according to the specified method. Besides, James second order test returns a list with class "jt" introducing the output to the `paircomp` function for pairwise comparison by adjusting the significance level instead of the p-value.

In the iris example, there is a statistically significant difference between iris species in terms of sepal length measurements. All pairwise comparisons of groups can be conducted using `paircomp` function. For simplicity, the Bonferroni correction is applied to show the usage of the pairwise comparisons following the significant result obtained by Alexander-Govern test and James second order test.

```
# Alexander-Govern test
R> out <- ag.test(Sepal.Length ~ Species, data = iris, verbose = FALSE)
R> paircomp(out, adjust.method = "bonferroni")

Bonferroni Correction (alpha = 0.05)
-----
  Level (a) Level (b)      p.value  No difference
1   setosa versicolor 8.187007e-17      Reject
2   setosa  virginica 1.105024e-25      Reject
3 versicolor  virginica 5.913702e-07      Reject
-----

# James second order test
R> out <- james.test(Sepal.Length ~ Species, data = iris, verbose = FALSE)
R> paircomp(out, adjust.method = "bonferroni")

Bonferroni Correction (alpha = 0.0166666666666667)
-----
  Level (a) Level (b) statistic criticalValue  No difference
1   setosa versicolor  110.6912      5.959328      Reject
2   setosa  virginica  236.7350      5.992759      Reject
3 versicolor  virginica   31.6875      5.938643      Reject
-----
```

In both Alexander-Govern and James second order tests with Bonferroni correction, statistical differences between all types of iris flowers are significant in terms of sepal length measurements.

### Checking assumptions via tests and plots

Two main assumptions, normality and variance homogeneity, can be checked through the `onewaytests` package. One can assess the variance homogeneity via `homog.test`, which has options including Levene (Levene's test), Bartlett (Bartlett's test) and Fligner (Fligner-Killeen test). Also, the normality of data in each group can be checked through `nor.test`, which has options including SW (Shapiro-Wilk test), SF (Shapiro-Franca test), LT (Lilliefors test known as Kolmogorov-Smirnov test), AD (Anderson-Darling test), CVM (Cramer-von Mises test), PT (Pearson Chi-square test). Moreover, the `nor.test` has options to assess the normality of data in each group through plots (Q-Q plots and Histograms with normal curves).

```
# Bartlett's homogeneity test
R> homog.test(Sepal.Length ~ Species, data = iris, method = "Bartlett")

Bartlett's Homogeneity Test
-----
data : Sepal.Length and Species

statistic : 16.0057
parameter : 2
p.value   : 0.0003345076
```

Result : Variances are not homogeneous.

Bartlett's homogeneity test results reveal that the variances between iris species are not homogeneous ( $\chi^2 = 16.0057$ , p-value = 0.0003345076).

# Shapiro-Wilk normality test

```
R> nor.test(Sepal.Length ~ Species, data = iris, method = "SW", plot = "qqplot-histogram")
```

Shapiro-Wilk Normality Test

data : Sepal.Length and Species

	Level	Statistic	p.value	Normality
1	setosa	0.9776985	0.4595132	Not reject
2	versicolor	0.9778357	0.4647370	Not reject
3	virginica	0.9711794	0.2583147	Not reject

Shapiro-Wilk normality test results state that there is not enough evidence to reject the normality of sepal length values in each iris species since all p-values are greater than 0.05. Also, the normality of data in each group can be assessed visually by Q-Q plots and histograms with normal curves (Figure 1).

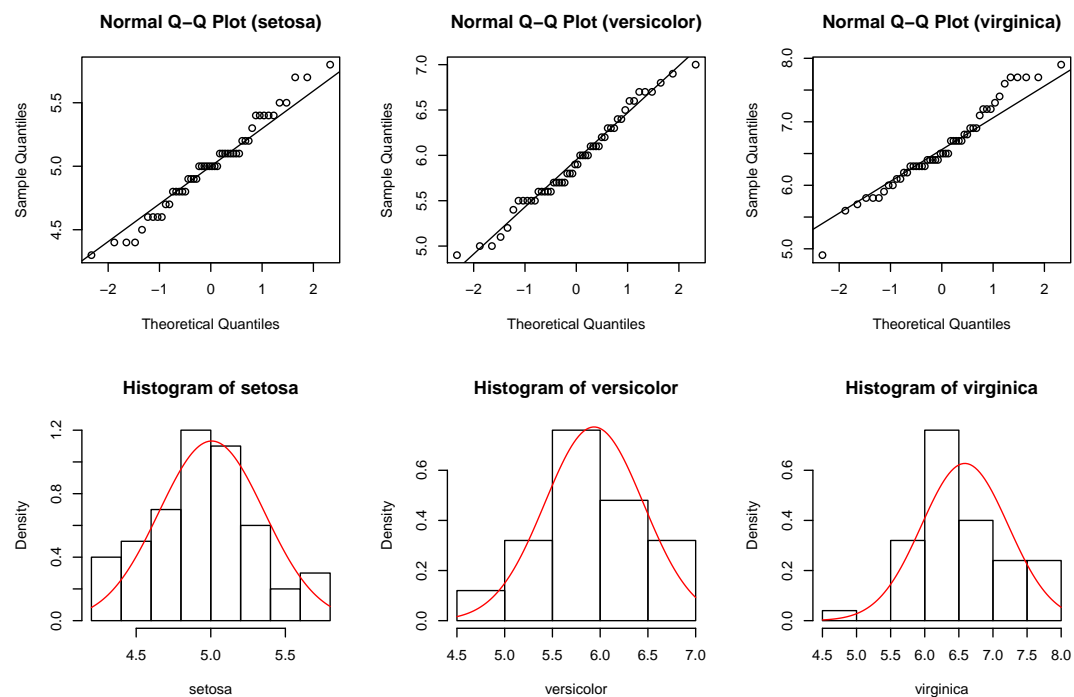


Figure 1: Q-Q plots and histograms with normal curves

### Graphical approaches

The users can obtain several graphic types of given groups via the `gplot`, which has options involving box-and-whisker plot with violin line - a rotated density line on each side - (Figure 2a), box-and-whisker plot (Figure 2b), mean  $\pm$  standard deviation graph (Figure 2c) and mean  $\pm$  standard error graph (Figure 2d). These graphics can be obtained via the following codes:

```
# Box-and-whisker plot with violin line
```

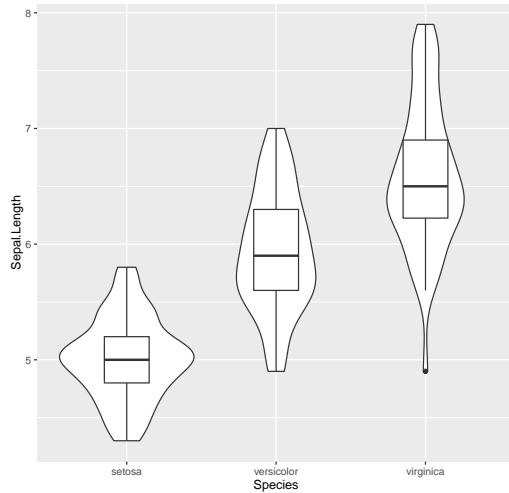
```
R> gplot(Sepal.Length ~ Species, data = iris, type = "boxplot")
```

```
# Box-and-whisker plot
```

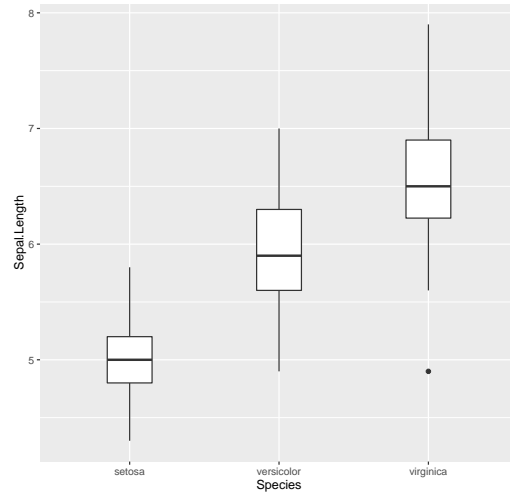
```
R> gplot(Sepal.Length ~ Species, data = iris, type = "boxplot", violin = FALSE)
```

```
# Mean +- standard deviation graph
R> gplot(Sepal.Length ~ Species, data = iris, type = "errorbar", option = "sd")

# Mean +- standard error graph
R> gplot(Sepal.Length ~ Species, data = iris, type = "errorbar", option = "se")
```



(a) Box-and-whisker plot with violin line



(b) Box-and-whisker plot

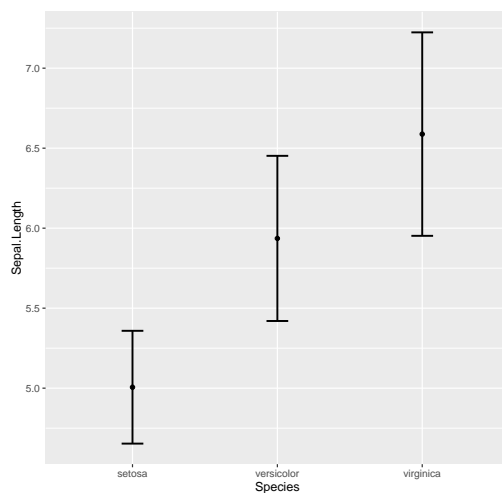
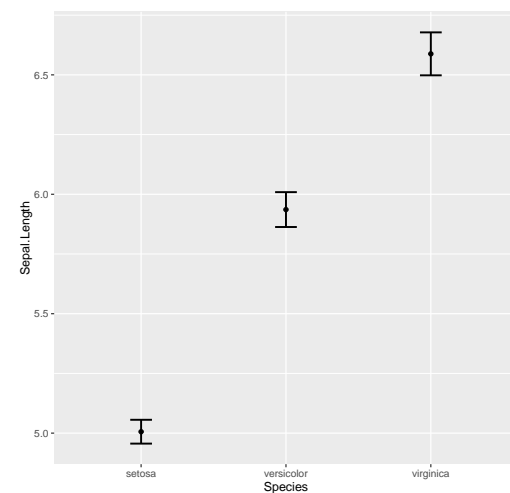
(c) Mean  $\pm$  standard deviation graph(d) Mean  $\pm$  standard error graph

Figure 2: Graphics of given groups

### German breast cancer data

In this part, we utilize German breast cancer study group (GBSG) data set, available in the **mfp** package (Ambler and Benner, 2015) in R. This data set was collected from a cohort study and measurements from patients with primary node positive breast cancer were taken between July 1984 and December 1989. Seven risk factors, including menopausal status, tumor size, tumor grade, age, number of positive lymph nodes, progesterone and oestrogen receptor concentrations were examined in the work done by Sauerbrei et al. (1999). For illustrative purposes, we take the dependent variable as recurrence-free survival times of patients, of whom an event for recurrence-free survival occurs, and the factor as tumor grade. This factor is a three-level categorical variable: 1 = grade I, 2 = grade II, and 3 = grade III. A total of 299 observations (18, 202, 79 observations in each group, respectively) are available.

The objective of adding GBSG dataset is showing the usage of the package in practice rather than demonstrating the usage of all functions in the **onewaytests** package. After checking the normality and variance homogeneity assumptions, an appropriate one-way test is decided to compare groups.

Also, pairwise comparisons are applied when it is necessary to determine which groups create the difference.

After installing and loading the **mfp** package, GBSG dataset can be reached by using the following R code.

```
# load GBSG data
R> data("GBSG")

# select the patients of whom an event for recurrence-free survival occurs
R> GBSG_subset <- GBSG[GBSG$cens == 1,]

# obtain the descriptive statistics of the tumor grades in terms of recurrence-free
# survival times
R> describe(rfst ~ tumgrad, data = GBSG_subset)
```

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skewness	Kurtosis	NA
1	18	1052.1111	444.5332	969.0	476	1990	729	1290.25	0.8712495	2.938206	0
2	202	845.9505	511.2683	729.5	72	2456	487	1160.75	0.9484978	3.253876	0
3	79	616.6076	432.2091	476.0	98	2034	312	758.00	1.4487566	4.698735	0

The assumptions of variance homogeneity and normality are assessed by using `homog.test` and `nor.test`, respectively.

```
# Bartlett's homogeneity test
R> homog.test(rfst ~ tumgrad, data = GBSG_subset, method = "Bartlett")
```

Bartlett's Homogeneity Test

-----  
data : rfst and tumgrad

statistic : 3.262419  
parameter : 2  
p.value : 0.1956927

Result : Variances are homogeneous.  
-----

Bartlett's homogeneity test results reveal that there is no enough evidence to reject the variance homogeneity ( $\chi^2 = 3.262419$ ,  $p\text{-value} = 0.1956927$ ) since  $p\text{-value}$  is larger than 0.05.

```
# Shapiro-Wilk normality test
R> nor.test(rfst ~ tumgrad, data = GBSG_subset, method = "SW")
```

Shapiro-Wilk Normality Test

-----  
data : rfst and tumgrad

	Level	Statistic	p.value	Normality
1	1	0.9097324	8.510408e-02	Not reject
2	2	0.9195909	4.749653e-09	Reject
3	3	0.8489033	1.708621e-07	Reject

-----

Shapiro-Wilk normality test results state that there is not enough evidence to reject the normality of recurrence - free survival times of the patients with tumor grade I since  $p\text{-value}$  is greater than 0.05. The normality of recurrence - free survival times of the patients with tumor grade II and III is not met since the  $p\text{-value}$  is smaller than 0.05. Our simulation study results suggest that ANOVA is the appropriate one-way test in such case.

```
R> aov.test(rfst ~ tumgrad,data = GBSG_subset)
```

One-Way Analysis of Variance

-----  
data : rfst and tumgrad

```

statistic : 8.875494
num df    : 2
denom df  : 296
p.value   : 0.000180542

```

```
Result    : Difference is statistically significant.
```

Since the p-value, derived from aov. test, is lower than 0.05, it can be concluded that there is a statistically significant difference between the tumor grades with respect to recurrence - free survival times ( $F = 8.875494$ , p-value = 0.000180542).

On the other hand, we observe that recurrence - free survival times of the patients with tumor grade II and III have few outliers and positively-skewed distributed. Liao et al. (2016) suggests that Kruskal-Wallis test should be used in such cases. Therefore, this test is also utilized along with ANOVA to compare groups.

```
R> kw.test(rfst ~ tumgrad, data = GBSG_subset)
```

```
Kruskal-Wallis Test
```

```
data : rfst and tumgrad
```

```

statistic : 23.42841
parameter : 2
p.value   : 8.176855e-06

```

```
Result    : Difference is statistically significant.
```

One can conclude that the difference between the tumor grades with respect to recurrence - free survival times is statistically significant ( $\chi^2_{KW} = 23.42841$ , p-value =  $8.176855 \times 10^{-06}$ ) since p-value is smaller than 0.05.

Both ANOVA and Kruskal-Wallis test results reveal that there is a statistically significant difference between tumor grade groups in terms of recurrence-free survival times. In the next step, we need to determine which groups create the difference.

```
# ANOVA
```

```
R> out <- aov.test(rfst ~ tumgrad, data = GBSG_subset, verbose = FALSE)
R> paircomp(out, adjust.method = "bonferroni")
```

```
Bonferroni Correction (alpha = 0.05)
```

```

-----
Level (a) Level (b)      p.value  No difference
1         1         2 0.2980174599    Not reject
2         1         3 0.0006698433      Reject
3         2         3 0.0014901832      Reject
-----

```

```
# Kruskal-Wallis test
```

```
R> out<- kw.test(rfst ~ tumgrad, data = GBSG_subset, verbose = FALSE)
R> paircomp(out, adjust.method = "bonferroni")
```

```
Bonferroni Correction (alpha = 0.05)
```

```

-----
Level (a) Level (b)      p.value  No difference
1         1         2 0.0949942647    Not reject
2         1         3 0.0001333143      Reject
3         2         3 0.0002457434      Reject
-----

```

Pairwise comparisons with Bonferroni correction after both ANOVA and Kruskal-Wallis test indicate that there is no statistically significant difference between the patients with tumor grade I and

those with tumor grade II in terms of recurrence-free survival times. On the other hand, there exists a statistically significant difference between the patients with tumor grade I/II and those with tumor grade III.

### Web interface of onewaytests package

The objective of this package is to provide the users with one-way tests in independent groups designs, pairwise comparisons, graphical approaches, and assess variance homogeneity and normality of data in each group via tests and plots. At times, it is difficult for new R users or applied researchers to deal with R codes. Thus, we have developed a web interface of **onewaytests** package by using **shiny** (Chang et al., 2017). The web interface is available at <http://www.softmed.hacettepe.edu.tr/onewaytests>.

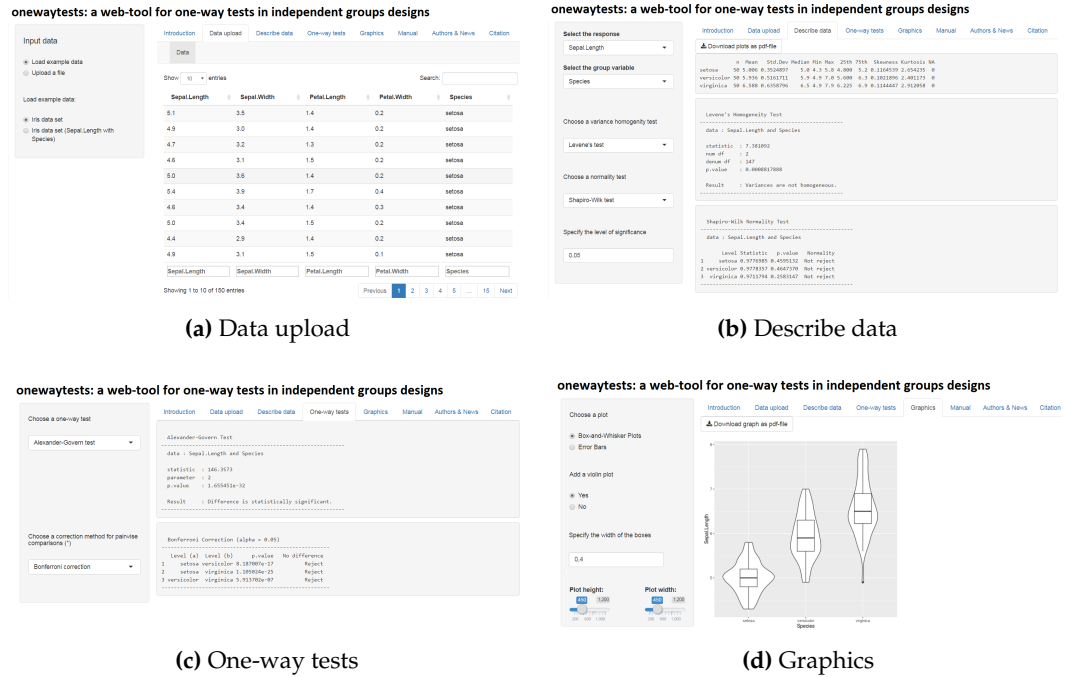


Figure 3: Web interface of onewaytests package

Users can upload their data to the tool via *Data upload* tab (Figure 3a). There exist two demo datasets on this tab for the users to test the tool. Basic descriptive statistics can be obtained through *Describe data* tab (Figure 3b). Moreover, assumptions of variance homogeneity and normality can be checked via this tab to decide on which one-way test is appropriate to test the statistical difference. Variance homogeneity is checked by variance homogeneity tests (Levene’s test, Bartlett’s test, Fligner-Killeen test). The normality of data in each group is assessed by normality tests (Shapiro-Wilk, Cramer-von Mises, Lilliefors (Kolmogorov-Smirnov), Shapiro-Francia, Anderson-Darling, Pearson Chi-Square tests) and plots (Q-Q plot and Histogram with a normal curve). After describing the data, users can test statistical difference between groups through *One-way tests* tab (Figure 3c). In this tab, there exist one-way tests (ANOVA, Welch’s heteroscedastic *F*, Welch’s heteroscedastic *F* test with trimmed means and Winsorized variances, Brown-Forsythe, Alexander-Govern, James second order test, Kruskal-Wallis tests) and pairwise comparison methods (Bonferroni, Holm, Hochberg, Hommel, Benjamini-Hochberg, Benjamini-Yekutieli, no corrections). Moreover, users can obtain the graphics of given groups (Figure 2) through *Graphics* tab (Figure 3d).

### Simulation study

In this section, it is aimed to compare the performances of seven tests in terms of type I error and adjusted power (Lloyd, 2005), and give some general suggestions on which test(s) should be used or avoided under the violations of assumptions. The adjusted power,  $R(\alpha)$ , is

$$R(\alpha) = \Phi(\delta + \Phi^{-1}(\alpha)), \tag{10}$$

with

$$\hat{\delta} = \Phi^{-1}(1 - \hat{\beta}) - \Phi^{-1}(\hat{\alpha}).$$

In Equation (10),  $\hat{\alpha}$  is the estimated type I error,  $\hat{\beta}$  is the estimated type II error,  $\alpha$  is the nominal size for type I error,  $\Phi$  and  $\Phi^{-1}$  are the cumulative and inverse cumulative distribution functions of the standard normal distribution, respectively.

### Simulation design

A Monte Carlo simulation is implemented to illustrate the performances of these tests for the scenarios in which the assumptions of normality and/or variance homogeneity are held or not. The algorithm of the simulation study can be depicted as follows:

- i) Generate three random samples from normal or skew normal distribution with means  $\mu_1, \mu_2$  and  $\mu_3$  and standard deviations  $\sigma_1, \sigma_2, \sigma_3$ . Group standard deviations are taken as  $\sigma_1 = \sigma_2 = \sigma_3 = 1$  for homogeneous case, and  $\sigma_1 = 1, \sigma_2 = \sqrt{2}, \sigma_3 = 2$  for heterogeneous case. Skewness  $\gamma$  is set to  $\gamma = 0$  for normal distribution,  $\gamma = 0.5$  for positive skew normal distribution and  $\gamma = -0.5$  for negative skew normal distribution with different sample size combinations (balanced and unbalanced). Set  $\mu_1 = \mu_2 = \mu_3 = 0$  for gathering type I error;  $\mu_1 = 0, \mu_2 = 0.25, \mu_3 = 0.5$  or  $\mu_1 = 0, \mu_2 = 0.5, \mu_3 = 1$  or  $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$  for power.
- ii) Check whether groups are different by the corresponding one-way test in independent groups designs at the level of significance  $\alpha$  ( $\alpha = 0.05$ ).
- iii) Repeat steps i) - ii) for 10,000 times and calculate the probability of rejecting the null hypothesis when the null hypothesis is true (type I error) or false (power).
- iv) Calculate adjusted power (given in Lloyd (2005)) by using the estimated type I error and power found in iii).

### Results

In this section, the performances of one-way tests in independent groups designs are investigated through type I error and adjusted power. All results are not given here to protect the content integrity, but attached as supplementary tables 2–4.

#### Type I error rates

In this part, we compare seven one-way tests in terms of their type I errors. We observed that the type I errors get closer to nominal level as sample size increases under the data generated from normal distribution ( $\gamma = 0$ ) with equal variances ( $\sigma_1 = \sigma_2 = \sigma_3 = 1$ ). Kruskal-Wallis test tends to be more conservative compared to the rest of them in the unbalanced small sample sizes ( $n_1 = 6, n_2 = 9, n_3 = 15$ ). When variance homogeneity is not held under normality, unbalance of sample sizes causes a decline in type I error rates of ANOVA and Kruskal-Wallis test regardless of sample size.

When the data in each group come from right skew normal distribution ( $\gamma = 0.5$ ) with unbalanced small sample size setting, type I error rate of ANOVA is estimated to be 0.052 under variance homogeneity ( $\sigma_1 = \sigma_2 = \sigma_3 = 1$ ) whereas it declines to 0.026 under variance heterogeneity ( $\sigma_1 = 1, \sigma_2 = \sqrt{2}, \sigma_3 = 2$ ). The type I error rate of ANOVA is negatively affected by the combined effect of heteroscedasticity and unbalanced sample size. Especially, when the smaller variances are associated with groups having smaller sample size, the empirical type I error rate of ANOVA is halved compared to the homoscedastic case. Kruskal-Wallis test is dramatically affected by variance heterogeneity. It tends to be conservative in the unbalanced case whereas it becomes liberal in the balanced case.

James second order test, Alexander-Govern test, Welch's heteroscedastic  $F$  test and Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances control the type I error rate at the nominal size for all simulation scenarios. Brown-Forsythe test is not able to control type I error rate; especially when the variances are heterogeneous, it becomes liberal.

#### Adjusted powers

Adjusted power is important to compare tests having different type I errors since it adjusts power with respect to type I error. The results are illustrated in Figures 4–5 to see the clear difference among the tests.

Within the tests discussed in this study, ANOVA is the best one with respect to adjusted power when the sample sizes are not equal under normality ( $\gamma = 0$ ) and variance homogeneity ( $\sigma_1 = \sigma_2 =$



$\sigma_3 = 1$ ). Under the same condition, ANOVA and Brown-Forsythe test are superior to the rest of them in terms of adjusted power when the design of sample size is balanced. As expected, ANOVA performs poor under normality when variances are unstable ( $\sigma_1 = 1, \sigma_2 = \sqrt{2}, \sigma_3 = 2$ ); however, Alexander-Govern, James and Welch's heteroscedastic  $F$  tests have higher adjusted powers compared to other tests. Brown-Forsythe and Kruskal-Wallis tests perform poorly compared to the others under the same condition.

When the data are generated from positive skew normal distribution ( $\gamma = 0.5$ ) with equal variances, ANOVA and Kruskal-Wallis test perform better than other tests. ANOVA is slightly better than Kruskal-Wallis test for small sample sizes and vice versa is true for the medium sample sizes. James, Alexander-Govern and Welch's heteroscedastic  $F$  tests have the highest adjusted powers among the whole tests for all scenario combinations under the data generated from a positive skew normal distribution with heterogeneous variances.

ANOVA, Brown-Forsythe test and Kruskal-Wallis test perform better than the other tests when the data are generated from negative skew normal distribution ( $\gamma = -0.5$ ) with homogeneous variances and equal sample sizes; however, Kruskal-Wallis test is slightly superior to ANOVA and Brown-Forsythe test. The adjusted power of ANOVA seems not to be affected by the skewness and it performs best under the same condition when the design of sample size is unbalanced. When the data are generated from the negative skew normal distribution with heterogeneous variances, Kruskal-Wallis test performs best for low ( $\mu_1 = 0, \mu_2 = 0.25, \mu_3 = 0.5$ ) and medium ( $\mu_1 = 0, \mu_2 = 0.5, \mu_3 = 1$ ) effect sizes while Welch's heteroscedastic  $F$  test with trimmed mean and Winsorized variance has higher performance compared to other tests for high effect size ( $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$ ).

It is noted that the Kruskal-Wallis test is affected by the skewness of the distribution especially under heterogeneity of variance ( $\sigma_1 = 1, \sigma_2 = \sqrt{2}, \sigma_3 = 2$ ). Kruskal-Wallis test has the highest adjusted power when the data are generated from a negatively skewed distribution ( $\gamma = -0.5$ ). On the other hand, it has the lowest adjusted power when the data are generated from a positively skewed distribution ( $\gamma = 0.5$ ).

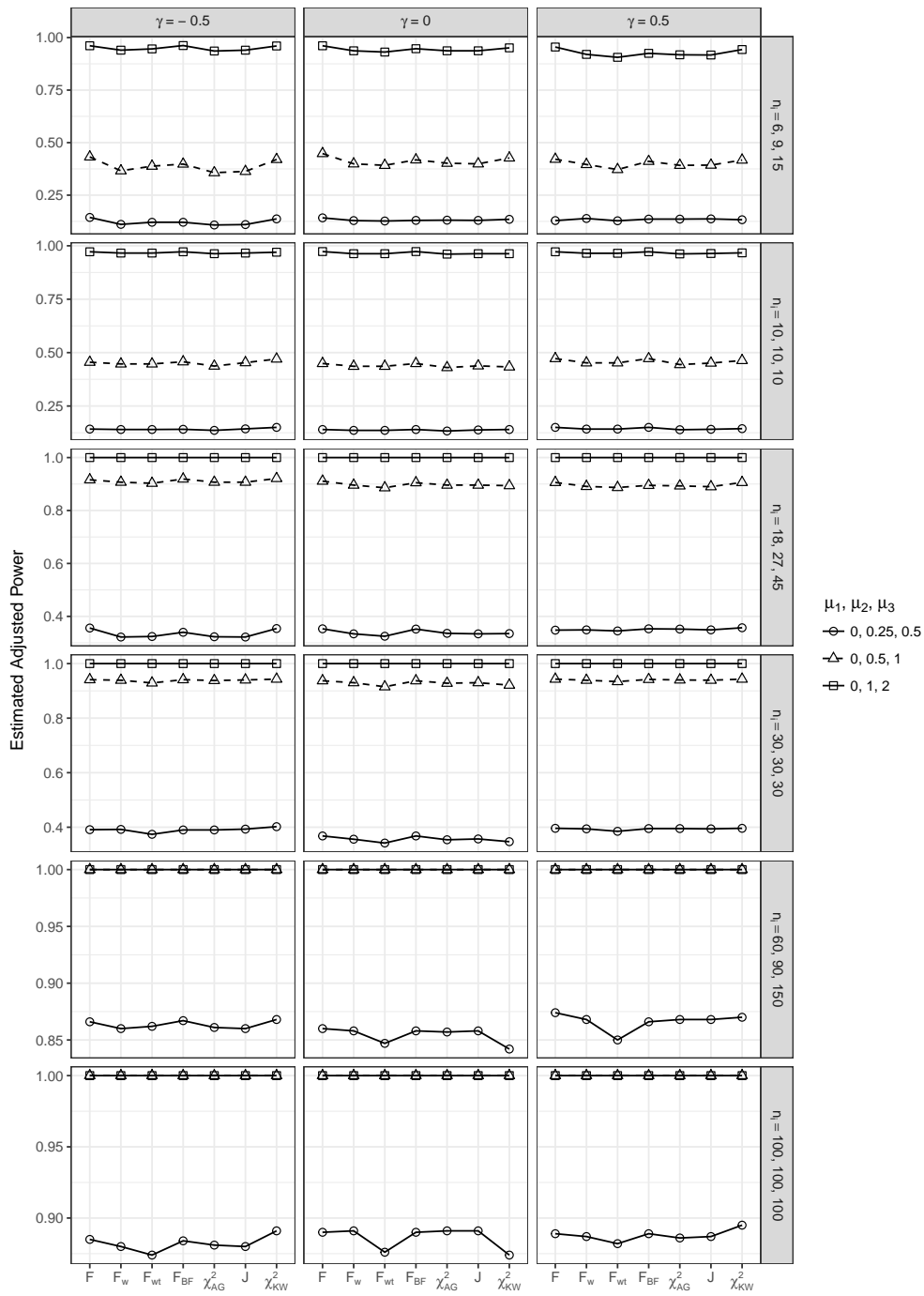


Figure 4: Estimated adjusted power results under homogeneous-variance ( $\sigma_1 = \sigma_2 = \sigma_3 = 1$ ) scenarios

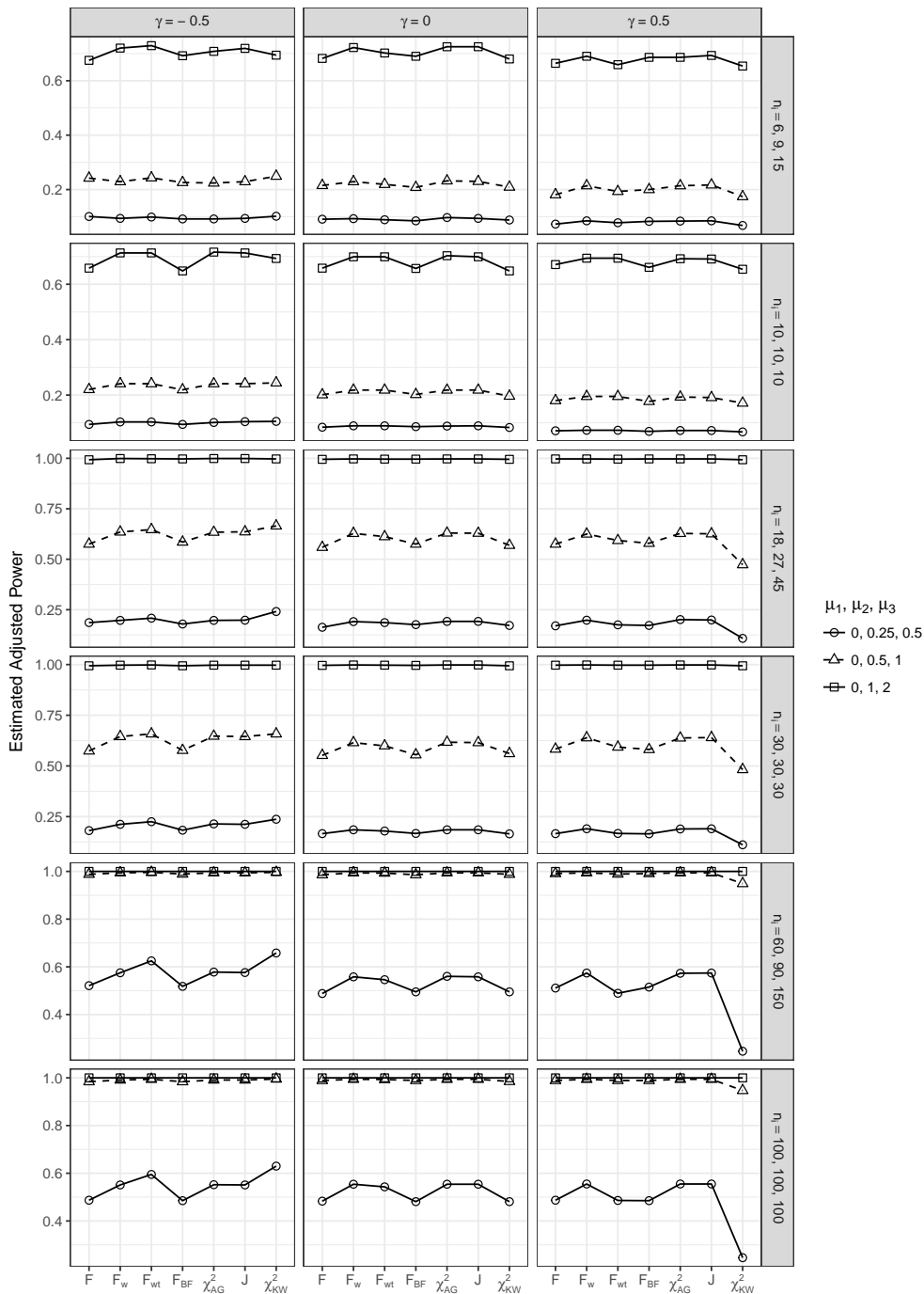


Figure 5: Estimated adjusted power results under heterogeneous-variance ( $\sigma_1 = 1, \sigma_2 = \sqrt{2}, \sigma_3 = 2$ ) scenarios

### Summary and further research

One-way tests in independent groups designs are the most commonly utilized statistical methods with applications on the experiments in medical sciences, pharmaceutical research, agriculture, biology, chemistry, engineering, and social sciences. In this paper, we present the **onewaytests** package for researchers to investigate treatment effects on the dependent variable.

The **onewaytests** package includes well-known one-way tests in independent groups designs including one-way analysis of variance, Kruskal-Wallis test, Welch heteroscedastic *F* test, Brown-Forsythe test. In addition to these well-known tests, Alexander-Govern test, James second order test

and Welch heteroscedastic  $F$  test with trimmed means and Winsorized variances, not available in most statistical software, are able to be reached via this package. Also, pairwise comparisons can be applied if the statistically significant difference is obtained.

Normality and variance homogeneity are the vital assumptions for the selection of the appropriate test. Therefore, this package also enables the users to check these assumptions via variance homogeneity and normality tests. It also enables the users some basic descriptive statistics and graphical approaches.

ANOVA is the most commonly used method for one-way independent groups designs. However, this method has two certain assumptions, homogeneity of variances and normality, to be satisfied. When these assumptions are not met, there are some other alternatives for one-way independent groups designs, which include Welch's heteroscedastic  $F$  test, Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances, Alexander-Govern test, James second order test, Brown-Forsythe test and Kruskal-Wallis test.

In this paper, we also compared seven one-way tests in **onewaytests** package with respect to the type I error rate and adjusted power. In the light of Monte Carlo simulation, it is noted that the normality assumption does not have a negative effect on the adjusted power of the tests as severe as variance homogeneity assumption. ANOVA is suggested to be used if the variance homogeneity assumption is held. Otherwise, Alexander-Govern, James second order and Welch's heteroscedastic  $F$  tests are recommended to be utilized. It is pointed out that under the negatively skew normal distribution with heterogeneous variances, Kruskal-Wallis test performs best with small and medium effect sizes while the Welch's heteroscedastic  $F$  test with trimmed means and Winsorized variances has the highest adjusted power with large effect size.

At present, the **onewaytests** package offers the one-way tests in independent groups designs, pairwise comparisons, graphical approaches, and assessment of variance homogeneity and normality via tests and plots. The package and its web-interface will be updated at regular intervals.

## Acknowledgment

We thank the anonymous reviewers for their constructive comments and suggestions which helped us to improve the quality of our paper.

## Bibliography

- R. A. Alexander and D. M. Govern. A New and Simpler Approximation for ANOVA under Variance Homogeneity. *Journal of Educational and Behavioral Statistics*, 19(2):91–101, 1994. URL <https://doi.org/10.2307/1165140>. [p175, 176, 178, 179]
- G. Ambler and A. Benner. *Mfp: Multivariable Fractional Polynomials*, 2015. R package version 1.5.2. [p186]
- E. Anderson. The Irises of the Gaspe Peninsula. *Bulletin of the American Iris society*, 59:2–5, 1935. [p180]
- Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1): 289–300, 1995. [p184]
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001. URL <https://doi.org/10.1214/aos/1013699998>. [p184]
- T. A. Bishop. *Heteroscedastic ANOVA, MANOVA, and Multiple-Comparisons*. PhD thesis, The Ohio State University, 1976. [p176]
- M. B. Brown and A. B. Forsythe. The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means. *Technometrics*, 16(1):129–132, 1974a. URL <https://doi.org/10.2307/1267501>. [p175, 176, 178]
- M. B. Brown and A. B. Forsythe. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974b. [p175, 176, 178]
- W. Chang, J. Cheng, J. Allaire, Y. Xie, and J. McPherson. *shiny: Web Application Framework for R*, 2017. R package version 1.0.1. [p189]

- R. A. Cribbie, R. R. Wilcox, C. Bewell, and H. Keselman. Tests for Treatment Group Equality When Data Are Nonnormal and Heteroscedastic. *Journal of Modern Applied Statistical Methods*, 6(1):117–132, 2007. URL <https://doi.org/10.22237/jmasm/1177992660>. [p175]
- R. A. Cribbie, L. Fiksenbaum, H. Keselman, and R. R. Wilcox. Effect of Non-Normality on Test Statistics for One-Way Independent Groups Designs. *British Journal of Mathematical and Statistical Psychology*, 65(1):56–73, 2012. URL <https://doi.org/10.1111/j.2044-8317.2011.02014.x>. [p175]
- O. Dag, A. Dolgun, and N. M. Konar. *Onewaytests: One-Way Tests in Independent Groups Designs*, 2017. R package version 1.5. [p175]
- M. W. Fagerland and L. Sandvik. The Wilcoxon–Mann–Whitney Test under Scrutiny. *Statistics in Medicine*, 28(10):1487–1497, 2009. URL <https://doi.org/10.1002/sim.3561>. [p179]
- R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2): 179–188, 1936. URL <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>. [p180]
- J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, 2nd edition, 2011. [p179]
- J. Gamage and S. Weerahandi. Size Performance of Some Tests in One-Way ANOVA. *Communications in Statistics-Simulation and Computation*, 27(3):625–640, 1998. URL <https://doi.org/10.1080/03610919808813500>. [p175]
- J. L. Gastwirth, Y. R. Gel, W. L. W. Hui, V. Lyubchich, W. Miao, and K. Noguchi. *lawstat: Tools for Biostatistics, Public Policy, and Law*, 2017. R package version 3.1. [p179]
- J. Gross and U. Ligges. *Nortest: Tests for Normality*, 2015. R package version 1.0-4. [p180]
- Y. Hochberg. A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 75(4): 800–802, 1988. [p184]
- S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2): 65–70, 1979. [p184]
- G. Hommel. A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. *Biometrika*, 75(2):383–386, 1988. [p184]
- T. Hothorn, K. Hornik, M. A. Van De Wiel, and A. Zeileis. A Lego System for Conditional Inference. *The American Statistician*, 60(3):257–263, 2006. URL <https://doi.org/10.1198/000313006X118430>. [p179]
- Y.-C. Hu. A new fuzzy-data mining method for pattern classification by principal component analysis. *Cybernetics and Systems: An International Journal*, 36(5):527–547, 2005. URL <https://doi.org/10.1080/01969720590944294>. [p180]
- G. S. James. The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38(3/4):324–329, 1951. URL <https://doi.org/10.2307/2332578>. [p175, 176, 178]
- G. S. James. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, 41(1-2):19–43, 1954. URL <https://doi.org/10.2307/2333003>. [p175, 176]
- H. Keselman, J. Algina, L. M. Lix, R. R. Wilcox, and K. N. Deering. A Generally Robust Approach for Testing Hypotheses and Setting Confidence Intervals for Effect Sizes. *Psychological Methods*, 13(2): 110–129, 2008. URL <https://doi.org/10.1037/1082-989X.13.2.110>. [p177]
- S. Korkmaz, D. Goksuluk, and G. Zararsiz. MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, 6(2):151–162, 2014. [p180]
- W. H. Kruskal and W. A. Wallis. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952. [p175, 176, 179]
- B. Lantz. The Impact of Sample Non-Normality on ANOVA and Alternative Methods. *British Journal of Mathematical and Statistical Psychology*, 66(2):224–244, 2013. URL <https://doi.org/10.1111/j.2044-8317.2012.02047.x>. [p175, 176]
- H. Liao, Y. Li, and G. Brooks. Outlier Impact and Accommodation Methods: Multiple Comparisons of Type I Error Rates. *Journal of Modern Applied Statistical Methods*, 15(1):23, 2016. URL <https://doi.org/10.22237/jmasm/1462076520>. [p188]

- C. J. Lloyd. Estimating Test Power Adjusted for Size. *Journal of Statistical Computation and Simulation*, 75(11):921–933, 2005. URL <https://doi.org/10.1080/00949650412331321160>. [p189, 190]
- P. Mair, F. Schoenbrodt, and R. Wilcox. *WRS2: Wilcox Robust Estimation and Testing*, 2017. 0.9-2. [p179]
- L. Myers. Comparability of the James' Second-Order Approximation Test and the Alexander and Govern A Statistic for Non-Normal Heteroscedastic Data. *Journal of Statistical Computation and Simulation*, 60(3):207–222, 1998. URL <https://doi.org/10.1080/00949659808811888>. [p178]
- T. Oshima and J. Algina. Type I Error Rates for James's Second-Order Test and Wilcox's Hm Test under Heteroscedasticity and Non-Normality. *British Journal of Mathematical and Statistical Psychology*, 45(2):255–263, 1992. URL <https://doi.org/10.1111/j.2044-8317.1992.tb00991.x>. [p179]
- I. Parra-Frutos. Testing Homogeneity of Variances with Unequal Sample Sizes. *Computational Statistics*, 28(3):1269–1297, 2013. URL <https://doi.org/10.1007/s00180-012-0353-x>. [p175]
- R Core Team. *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>. [p179, 183]
- W. Sauerbrei, P. Royston, H. Bojar, C. Schmoor, M. Schumacher, G. B. C. S. Group, and others. Modelling the effects of standard prognostic factors in node-positive breast cancer. *British Journal of Cancer*, 79: 1752–1760, 1999. URL <https://doi.org/10.1038/sj.bjc.6690279>. [p186]
- P. J. Villacorta. The welchADF Package for Robust Hypothesis Testing in Unbalanced Multivariate Mixed Models with Heteroscedastic and Non-Normal Data. *The R Journal*, 2017. [p179]
- B. L. Welch. On the comparison of several mean values - an alternative approach. *Biometrika*, 38(3/4): 330–336, 1951. URL <https://doi.org/10.2307/2332579>. [p175, 176, 177]
- H. Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, 2009. [p180]
- R. R. Wilcox. A New Alternative to the ANOVA F and New Results on James's Second-Order Method. *British Journal of Mathematical and Statistical Psychology*, 41(1):109–117, 1988. URL <https://doi.org/10.1111/j.2044-8317.1988.tb00890.x>. [p175]

Osman Dag  
Department of Biostatistics  
Hacettepe University Faculty of Medicine, Sıhhiye, Ankara  
Turkey  
[osman.dag@hacettepe.edu.tr](mailto:osman.dag@hacettepe.edu.tr)

Anil Dolgun  
School of Science  
RMIT University, Melbourne  
Australia  
[anil.dolgun@rmit.edu.au](mailto:anil.dolgun@rmit.edu.au)

Naime Meric Konar  
Department of Biostatistics  
Hacettepe University Faculty of Medicine, Sıhhiye, Ankara  
Turkey  
[nmeric.konar@hacettepe.edu.tr](mailto:nmeric.konar@hacettepe.edu.tr)

## Supplementary Material

**Table 2:** Estimated type I error rates

Distribution	$\sigma_1, \sigma_2, \sigma_3$	$n_1, n_2, n_3$	$F$	$F_w$	$F_{wt}$	$F_{BF}$	$\chi^2_{AG}$	$J$	$\chi^2_{KW}$	
Normal ( $\gamma = 0$ )	1, 1, 1	6, 9, 15	0.046	0.047	0.047	0.046	0.046	0.047	0.041	
		10, 10, 10	0.052	0.050	0.050	0.050	0.049	0.051	0.047	
		18, 27, 45	0.049	0.051	0.051	0.048	0.051	0.051	0.051	0.048
		30, 30, 30	0.051	0.052	0.053	0.051	0.052	0.052	0.052	0.051
		60, 90, 150	0.050	0.049	0.050	0.051	0.049	0.049	0.049	0.050
		100, 100, 100	0.051	0.050	0.052	0.051	0.050	0.050	0.050	0.050
	1, $\sqrt{2}$ , 2	6, 9, 15	0.023	0.045	0.047	0.051	0.043	0.046	0.029	
		10, 10, 10	0.058	0.051	0.051	0.053	0.050	0.052	0.054	
		18, 27, 45	0.025	0.050	0.050	0.053	0.049	0.050	0.032	
		30, 30, 30	0.058	0.052	0.051	0.056	0.051	0.052	0.055	
		60, 90, 150	0.025	0.051	0.051	0.055	0.051	0.051	0.034	
		100, 100, 100	0.059	0.052	0.052	0.059	0.052	0.052	0.059	
Skew Normal ( $\gamma = 0.5$ )	1, 1, 1	6, 9, 15	0.052	0.055	0.053	0.051	0.056	0.055	0.047	
		10, 10, 10	0.048	0.048	0.048	0.046	0.048	0.050	0.045	
		18, 27, 45	0.053	0.054	0.052	0.052	0.053	0.054	0.050	
		30, 30, 30	0.047	0.046	0.045	0.047	0.045	0.046	0.046	
		60, 90, 150	0.046	0.046	0.050	0.046	0.046	0.046	0.049	
		100, 100, 100	0.052	0.052	0.052	0.052	0.052	0.052	0.050	
	1, $\sqrt{2}$ , 2	6, 9, 15	0.026	0.052	0.051	0.053	0.052	0.053	0.032	
		10, 10, 10	0.059	0.051	0.051	0.054	0.051	0.053	0.052	
		18, 27, 45	0.023	0.051	0.053	0.055	0.050	0.051	0.038	
		30, 30, 30	0.053	0.046	0.046	0.052	0.046	0.046	0.057	
		60, 90, 150	0.022	0.048	0.052	0.052	0.048	0.048	0.052	
		100, 100, 100	0.058	0.050	0.053	0.058	0.050	0.050	0.082	
Skew Normal ( $\gamma = -0.5$ )	1, 1, 1	6, 9, 15	0.048	0.050	0.051	0.048	0.053	0.051	0.045	
		10, 10, 10	0.051	0.048	0.048	0.048	0.048	0.048	0.043	
		18, 27, 45	0.048	0.051	0.049	0.049	0.050	0.051	0.047	
		30, 30, 30	0.048	0.047	0.048	0.048	0.047	0.047	0.045	
		60, 90, 150	0.049	0.051	0.050	0.050	0.050	0.051	0.049	
		100, 100, 100	0.053	0.054	0.054	0.053	0.053	0.054	0.052	
	1, $\sqrt{2}$ , 2	6, 9, 15	0.024	0.047	0.049	0.048	0.047	0.049	0.031	
		10, 10, 10	0.058	0.051	0.051	0.054	0.050	0.052	0.053	
		18, 27, 45	0.024	0.049	0.049	0.053	0.048	0.049	0.037	
		30, 30, 30	0.054	0.048	0.048	0.052	0.047	0.048	0.059	
		60, 90, 150	0.022	0.050	0.050	0.054	0.049	0.050	0.054	
		100, 100, 100	0.058	0.054	0.055	0.058	0.054	0.054	0.085	

**Table 3:** Adjusted power results

Distribution	$\sigma_1, \sigma_2, \sigma_3$	$n_1, n_2, n_3$	$\mu_1, \mu_2, \mu_3$	$F$	$F_w$	$F_{wt}$	$F_{BF}$	$\chi^2_{AG}$	$J$	$\chi^2_{KW}$
Normal ( $\gamma = 0$ )	1, 1, 1	6, 9, 15	0, 0.25, 0.5	<b>0.141</b>	0.129	0.127	0.130	0.132	0.130	0.136
			0, 0.5, 1	<b>0.447</b>	0.400	0.392	0.417	0.404	0.399	0.428
			0, 1, 2	<b>0.961</b>	0.937	0.931	0.948	0.937	0.937	0.951
		10, 10, 10	0, 0.25, 0.5	0.140	0.136	0.136	<b>0.141</b>	0.133	0.138	0.139
			0, 0.5, 1	0.449	0.436	0.436	<b>0.451</b>	0.430	0.437	0.433
			0, 1, 2	<b>0.973</b>	0.963	0.963	<b>0.973</b>	0.961	0.964	0.963
		18, 27, 45	0, 0.25, 0.5	<b>0.353</b>	0.335	0.324	0.351	0.334	0.334	0.335
			0, 0.5, 1	<b>0.911</b>	0.897	0.886	0.905	0.895	0.896	0.894
			0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
		30, 30, 30	0, 0.25, 0.5	0.367	0.356	0.343	<b>0.368</b>	0.356	0.356	0.347
			0, 0.5, 1	<b>0.937</b>	0.930	0.915	<b>0.937</b>	0.929	0.930	0.920
			0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
	60, 90, 150	0, 0.25, 0.5	<b>0.861</b>	0.858	0.848	0.859	0.858	0.858	0.842	
		0, 0.5, 1	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	
	100, 100, 100	0, 0.25, 0.5	<b>0.891</b>	0.890	0.877	<b>0.891</b>	<b>0.891</b>	0.890	0.874	
		0, 0.5, 1	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	
	1, $\sqrt{2}$ , 2	6, 9, 15	0, 0.25, 0.5	0.091	0.093	0.089	0.086	<b>0.096</b>	0.093	0.087
			0, 0.5, 1	0.214	0.228	0.220	0.209	<b>0.232</b>	0.229	0.207
			0, 1, 2	0.682	0.721	0.703	0.691	<b>0.724</b>	0.723	0.677
		10, 10, 10	0, 0.25, 0.5	0.084	<b>0.089</b>	<b>0.089</b>	0.086	0.087	<b>0.089</b>	0.082
			0, 0.5, 1	0.200	0.217	0.217	0.200	0.217	<b>0.219</b>	0.195
			0, 1, 2	0.657	0.697	0.697	0.655	<b>0.702</b>	0.700	0.647
18, 27, 45		0, 0.25, 0.5	0.164	0.192	0.186	0.175	0.192	<b>0.193</b>	0.170	
		0, 0.5, 1	0.562	0.629	0.612	0.574	<b>0.630</b>	<b>0.630</b>	0.568	
		0, 1, 2	0.995	<b>0.997</b>	0.996	0.996	<b>0.997</b>	<b>0.997</b>	0.995	
30, 30, 30		0, 0.25, 0.5	0.167	0.185	0.180	0.166	<b>0.187</b>	0.185	0.165	
		0, 0.5, 1	0.553	0.615	0.600	0.554	<b>0.619</b>	0.615	0.561	
		0, 1, 2	0.996	0.997	0.997	0.996	<b>0.998</b>	0.997	0.994	
60, 90, 150	0, 0.25, 0.5	0.491	0.559	0.545	0.496	0.558	<b>0.560</b>	0.496		
	0, 0.5, 1	0.987	<b>0.994</b>	0.992	0.987	<b>0.994</b>	<b>0.994</b>	0.988		
	0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
100, 100, 100	0, 0.25, 0.5	0.483	0.553	0.545	0.482	<b>0.554</b>	0.553	0.483		
	0, 0.5, 1	0.988	<b>0.994</b>	0.992	0.988	<b>0.994</b>	<b>0.994</b>	0.985		
	0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
Skew Normal ( $\gamma = 0.5$ )	1, 1, 1	6, 9, 15	0, 0.25, 0.5	0.128	<b>0.137</b>	0.127	0.135	0.136	<b>0.137</b>	0.134
			0, 0.5, 1	<b>0.419</b>	0.395	0.372	0.410	0.392	0.393	0.418
			0, 1, 2	<b>0.955</b>	0.919	0.906	0.924	0.918	0.917	0.944
		10, 10, 10	0, 0.25, 0.5	<b>0.150</b>	0.141	0.141	0.149	0.138	0.141	0.144
			0, 0.5, 1	<b>0.472</b>	0.450	0.450	0.470	0.443	0.450	0.463
			0, 1, 2	<b>0.972</b>	0.964	0.964	0.971	0.961	0.964	0.967
		18, 27, 45	0, 0.25, 0.5	0.348	0.350	0.345	0.352	0.351	0.350	<b>0.359</b>
			0, 0.5, 1	<b>0.907</b>	0.891	0.887	0.895	0.892	0.891	<b>0.907</b>
			0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
		30, 30, 30	0, 0.25, 0.5	0.397	0.395	0.386	0.397	0.396	0.395	<b>0.398</b>
			0, 0.5, 1	0.942	0.939	0.934	0.942	0.940	0.939	<b>0.943</b>
			0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
60, 90, 150	0, 0.25, 0.5	<b>0.875</b>	0.868	0.851	0.867	0.869	0.868	0.870		
	0, 0.5, 1	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
	0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
100, 100, 100	0, 0.25, 0.5	0.890	0.887	0.882	0.890	0.887	0.887	<b>0.894</b>		
	0, 0.5, 1	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
	0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		



**Table 4:** Adjusted power results continued

Distribution	$\sigma_1, \sigma_2, \sigma_3$	$n_1, n_2, n_3$	$\mu_1, \mu_2, \mu_3$	$F$	$F_w$	$F_{wt}$	$F_{BF}$	$\chi^2_{AG}$	$J$	$\chi^2_{KW}$	
Skew Normal ( $\gamma = 0.5$ )	$1, \sqrt{2}, 2$	6, 9, 15	0, 0.25, 0.5	0.072	<b>0.085</b>	0.077	0.083	<b>0.085</b>	<b>0.085</b>	0.068	
			0, 0.5, 1	0.181	<b>0.216</b>	0.193	0.199	0.215	<b>0.216</b>	0.172	
			0, 1, 2	0.664	0.691	0.659	0.684	0.687	<b>0.693</b>	0.652	
		10, 10, 10	0, 0.25, 0.5	0.072	0.072	0.072	0.069	0.072	<b>0.073</b>	0.067	
			0, 0.5, 1	0.180	<b>0.194</b>	<b>0.194</b>	0.176	<b>0.194</b>	0.192	0.170	
			0, 1, 2	0.672	<b>0.693</b>	<b>0.693</b>	0.661	<b>0.693</b>	0.692	0.653	
		18, 27, 45	0, 0.25, 0.5	0.169	0.199	0.174	0.171	<b>0.200</b>	0.198	0.109	
			0, 0.5, 1	0.573	0.625	0.591	0.579	<b>0.627</b>	0.625	0.475	
			0, 1, 2	<b>0.997</b>	<b>0.997</b>	0.995	0.996	<b>0.997</b>	<b>0.997</b>	0.993	
	30, 30, 30	0, 0.25, 0.5	0.167	<b>0.189</b>	0.167	0.166	0.188	<b>0.189</b>	0.111		
		0, 0.5, 1	0.584	<b>0.639</b>	0.593	0.582	0.638	<b>0.639</b>	0.483		
		0, 1, 2	0.997	<b>0.998</b>	0.997	0.997	<b>0.998</b>	<b>0.998</b>	0.994		
	60, 90, 150	0, 0.25, 0.5	0.514	<b>0.573</b>	0.489	0.516	<b>0.573</b>	<b>0.573</b>	0.246		
		0, 0.5, 1	0.991	<b>0.994</b>	0.989	0.991	<b>0.994</b>	<b>0.994</b>	0.949		
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
	100, 100, 100	0, 0.25, 0.5	0.487	<b>0.554</b>	0.486	0.487	<b>0.554</b>	<b>0.554</b>	0.245		
		0, 0.5, 1	0.990	<b>0.994</b>	0.990	0.990	<b>0.994</b>	<b>0.994</b>	0.946		
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
	Skew Normal ( $\gamma = -0.5$ )	$1, 1, 1$	6, 9, 15	0, 0.25, 0.5	<b>0.144</b>	0.111	0.120	0.120	0.107	0.111	0.135
				0, 0.5, 1	<b>0.432</b>	0.364	0.386	0.396	0.355	0.365	0.417
				0, 1, 2	0.961	0.940	0.946	<b>0.962</b>	0.935	0.940	0.959
			10, 10, 10	0, 0.25, 0.5	0.142	0.142	0.142	0.142	0.138	0.142	<b>0.150</b>
				0, 0.5, 1	0.455	0.449	0.449	0.449	0.439	0.451	<b>0.470</b>
				0, 1, 2	<b>0.972</b>	0.966	0.966	<b>0.972</b>	0.963	0.966	0.970
18, 27, 45			0, 0.25, 0.5	<b>0.358</b>	0.322	0.325	0.340	0.322	0.322	0.356	
			0, 0.5, 1	0.917	0.908	0.903	0.920	0.907	0.908	<b>0.921</b>	
			0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	
30, 30, 30		0, 0.25, 0.5	0.391	0.392	0.376	0.390	0.390	0.393	<b>0.401</b>		
		0, 0.5, 1	0.941	0.939	0.929	0.941	0.937	0.939	<b>0.943</b>		
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
60, 90, 150		0, 0.25, 0.5	0.867	0.860	0.862	<b>0.868</b>	0.861	0.860	0.867		
		0, 0.5, 1	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
100, 100, 100		0, 0.25, 0.5	0.884	0.880	0.874	0.884	0.880	0.880	<b>0.892</b>		
		0, 0.5, 1	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
		0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>		
$1, \sqrt{2}, 2$		6, 9, 15	0, 0.25, 0.5	0.101	0.094	0.098	0.092	0.091	0.094	<b>0.102</b>	
			0, 0.5, 1	0.240	0.229	0.243	0.225	0.223	0.230	<b>0.249</b>	
			0, 1, 2	0.672	0.720	<b>0.728</b>	0.691	0.707	0.720	0.694	
		10, 10, 10	0, 0.25, 0.5	0.094	0.103	0.103	0.094	0.102	<b>0.104</b>	<b>0.104</b>	
			0, 0.5, 1	0.220	0.240	0.240	0.220	0.242	0.242	<b>0.243</b>	
			0, 1, 2	0.658	0.713	0.713	0.649	<b>0.717</b>	0.713	0.692	
	18, 27, 45	0, 0.25, 0.5	0.188	0.197	0.207	0.178	0.197	0.197	<b>0.242</b>		
		0, 0.5, 1	0.579	0.635	0.645	0.582	0.633	0.636	<b>0.666</b>		
		0, 1, 2	0.993	<b>0.999</b>	<b>0.999</b>	0.996	<b>0.999</b>	<b>0.999</b>	0.998		
30, 30, 30	0, 0.25, 0.5	0.180	0.213	0.225	0.183	0.213	0.212	<b>0.238</b>			
	0, 0.5, 1	0.574	0.646	0.658	0.576	0.647	0.646	<b>0.659</b>			
	0, 1, 2	0.994	0.997	<b>0.998</b>	0.994	0.997	0.997	0.997			
60, 90, 150	0, 0.25, 0.5	0.518	0.576	0.624	0.517	0.577	0.576	<b>0.659</b>			
	0, 0.5, 1	0.988	0.994	<b>0.996</b>	0.988	0.994	0.994	<b>0.996</b>			
	0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>			
100, 100, 100	0, 0.25, 0.5	0.487	0.552	0.595	0.486	0.552	0.552	<b>0.630</b>			
	0, 0.5, 1	0.984	0.992	0.994	0.984	0.992	0.991	<b>0.995</b>			
	0, 1, 2	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>			