3. Nonparametric methods

If the probability distributions of the statistical variables are unknown or are not as required (e.g. normality assumption violated), then we may still apply *nonparametric tests* in order to test whether two or more groups follow the same distribution. Such methods do not assume the data to follow any predefined distribution and are therefore also called *distribution-free methods*. In particular, they do not deal with any particular parameters of those distributions, hence the name *nonparametric methods*.

Nonparametric methods require nothing more than the variables to be measured at ordinal scale at least and they use only information from the ranks. This makes them on one hand more robust but on the other hand less efficient than their parametric counterparts, when the assumptions of the parametric tests are met.

3.1. The Sign Test

The sign test is used as a substitude for the paired-data t-test in the following situations:

- if the measurement scale is only ordinal,
- if the normality assumption is violated,
- as a quick substitude for the t-test.

It may also be used as a substitude for the single-sample t-test for location when the normality assumption is violated.

As a test for comparing two populations, the sign test is stated in terms of the probability that values of one population are greater than values of a second population that are paired two the first (e.g. responses X and Y of same consumers to two advertisements).

The null hypothesis of the two-sided test is that $p = P(X > Y) = 0.5$. The alternative hypothesis is that $p \neq 0.5$. (One sided alternatives would be $p > 0.5$ or $p < 0.5$.)

With n observation pairs (x_i, y_i), the test statistic is $\quad T =$ number of plus signs, $\quad$ where

$$x_i > y_i \rightarrow (+) \quad \text{and} \quad x_i < y_i \rightarrow (-)$$

and pairs with $x_i = y_i$ (ties) are discarded.

Under the null hypothesis: $T \sim \text{Bin}(n, \frac{1}{2})$ with $E(T) = \frac{n}{2}$ and $\text{Var}\,(T) = \frac{n}{4}$.

A large sample statistic is therefore:

$$Z = \frac{T - E(T)}{\sqrt{\text{Var}\,(T)}} = \frac{2T - n}{\sqrt{n}} \sim N(0,1) \text{ under } H_0.$$

<u>Example:</u> (Azcel, Example 14.1)
A management consultant wants to test if the same people rate their respective CEO differently on a scale from 1 (worst) to 5 (best) before and after receiving their MBA degree.

Example: (continued)

before	3	5	2	2	4	2	1	5	4	5	3	2	2	2	1	3	4
after	4	5	3	4	4	3	2	4	5	4	4	5	5	3	2	2	5
change	1	0	1	2	0	1	1	-1	1	-1	1	3	3	1	1	-1	1

In terms of plus and minus signs, the data is after removing the ties for CEO's 2 and 5:

$$+ + + + + - + - + + + + + - +,$$

such that: $\quad T =$ number of pluses $= 12$.

Now for $T \sim$Bin(15,0.5):

$$P(T \geq 12) = P(T \leq 3) = \sum_{k=0}^{3} \binom{15}{k} 0.5^{15}$$

$$= \left(1 + 15 + \frac{15 \cdot 14}{2} + \frac{15 \cdot 14 \cdot 13}{2 \cdot 3}\right) 0.5^{15}$$

$$\approx 0.0176,$$

yielding a two-sided p-value of 0.0352. There is thus evidence at 5% significance level that the attitude of employees towards their CEO changes after receiving an MBA. Since the rejection happened at the right tail, the attitude change is positive.

Applying the large sample approximation $Z = \frac{2T-n}{\sqrt{n}}$, we would have got applying a continuity correction:

$$P(T \geq 12) = P(T \leq 3) \approx P\left(Z \leq \frac{2 \cdot 3.5 - 15}{\sqrt{15}}\right) = \Phi(-2.065)$$

$$= 1 - \Phi(2.065) = 1 - 0.9805 = 0.0195,$$

yielding a two-sided p-value of 0.039.

The sign test can also be viewed as a test of the hypothesis that the *median* difference between two populations is zero. To see this, note that

$$H_0\colon\ P(X > Y) = 0.5 \ \Leftrightarrow\ P(X - Y > 0) = 0.5,$$

which is just the defining property of $\text{Median}_{(Y-X)} = 0$.

As such, the test my be adapted for testing whether the median of a single population is equal to any prespecified number a. The null and alternative hypotheses of the two-sided test are:

$$H_0\colon \text{Population median} = a,$$
$$H_1\colon \text{Population median} \neq a.$$

To conduct the test, we pair our data observations with a and perform the sign test. If the null hypothesis is true, then about one-half of the signs will be pluses and one-half minuses because, by definition of the median, one-half of the population values are above it and one-half below it.

Example: (Azcel)

Suppose we wish to test the null hypothesis that the median income in a certain region is \$24 000 per family per year. The randomly sampled data is (in thousands of dollars):

$$\begin{array}{c} 22,30,28,22,34,19,42,18,16,26 \\ 30,25,29,20,17,33,32,24,15,31 \end{array}$$

or in terms of $+$ and $-$ signs (wrt. 24):

$$\begin{array}{c} -++-+-+--+ \\ +++--++0-+, \end{array}$$

such that $T = 11$ and $n = 19$ (disregarding the single tie). Now for T ~Bin(19,0.5):

$$\begin{aligned} P(T \geq 11) &= P(T \geq 10) - P(T = 10) \\ &= 0.5 - \binom{19}{10} 0.5^{19} \approx 0.3238, \end{aligned}$$

such that the two-sided p-value is 0.6476. We can thus not reject the null hypothesis that the median equals 24.

You can automate the calculation of those p-values by using the SIGNTEST-function of the Real Statistics toolbox. The syntax is p=SIGNTEST(*range,median, tails*), where *range* contains the sample data, *median* contains the hypothesized median and *tails*$=$ 1 or 2 denotes a one- or two-sided test.

3.2. The Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is designed for the same purpose and may be applied in the same situations as the sign test, that is, for testing for (differences in) location in single populations or matched pairs, when the conditions for the corresponding t-tests are not met. The Wilcoxon signed-rank test is more efficient than the sign test as it uses the full information of the ranks rather than just their signs and is thus to be preferred whenever statistical software allows its use. Its only disadvantage is that its test statistic is somewhat harder to calculate by hand and that it is slightly less robust to outliers than the sign test.

As a paired-observations two-sample test, the hypotheses of the two-sided test are

H_0: median of pop. 1 = median of pop. 2,
H_1: median of pop. 1 $\neq$ median of pop. 2.

In order to perform the test, we list the pairs of observations (x_1, x_2) of the two populations and compute their difference $D = x_1 - x_2$. Then we rank the absolute values of the differences D discarding any ties with $D = 0$. In the next step, we form separate sums of the ranks of the positive and the negative differences.

The Wilcoxon T statistic is defined as the smaller of the two sums of ranks,

$$T = \min\left[\sum(+), \sum(-)\right],$$

where $\sum(+/-)$ is the sum of the ranks of the positive/negative differences.

Critical points of the test statistic T are tabulated. We reject the null hypothesis of the two-sided test if the computed value of the statistic is less than the $\frac{\alpha}{2}$ critical point from the table. We reject H_0 against the one-sided alternative H_1: median of pop. 1 $>$ ($<$) median of pop. 2 if $\sum(-)$ ($\sum(+)$) is less than the α critical point from the table.

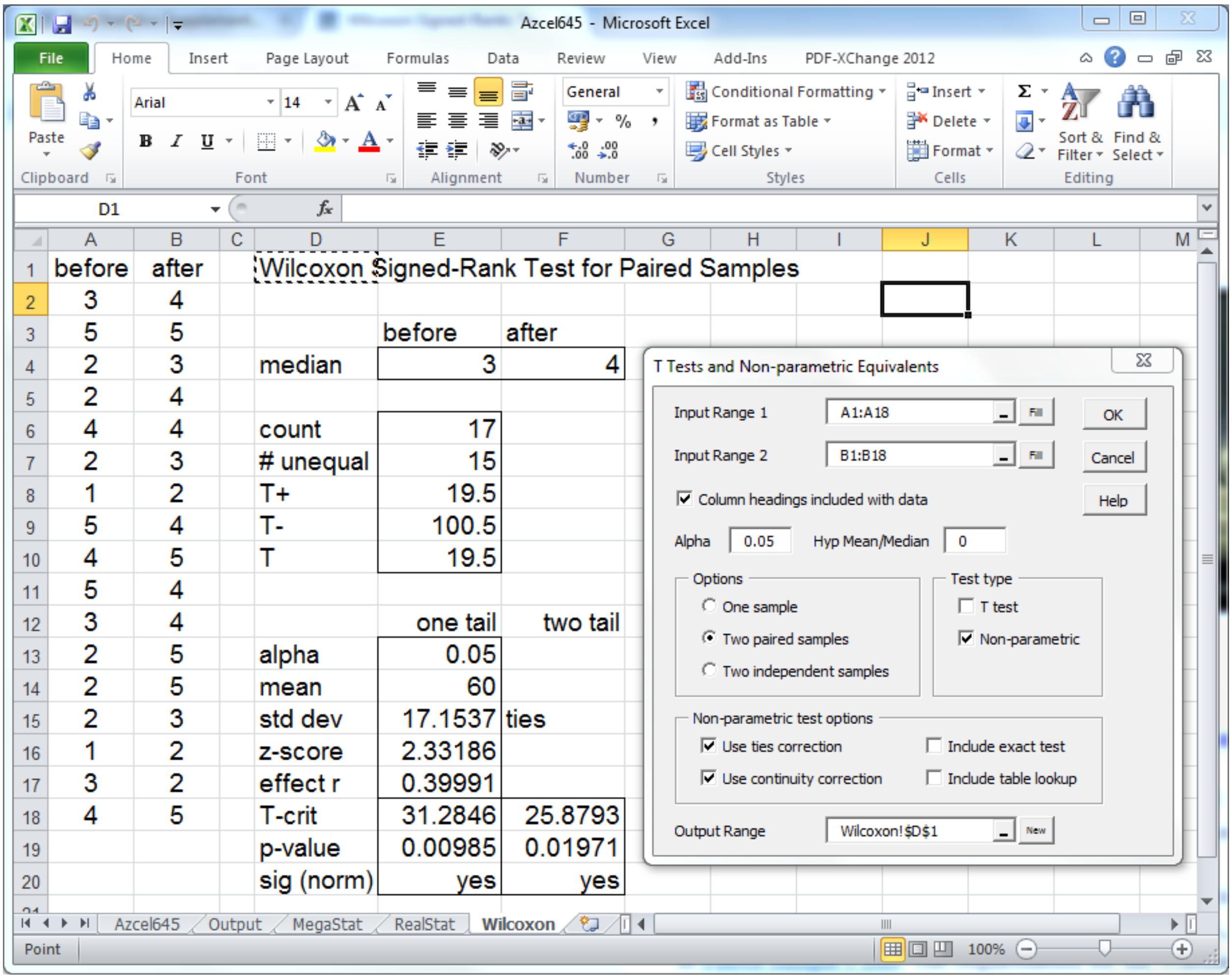
Example: (Satisfaction with CEO continued.)

CEO	Satisfaction before MBA	Satisfaction after MBA	D	$\|D\|$	Rank $\|D\|$	Rank pos.D	Rank neg.D
1	3	4	-1	1	6.5		6.5
2	5	5	0	0			
3	2	3	-1	1	6.5		6.5
4	2	4	-2	2	13		13
5	4	4	0	0			
6	2	3	-1	1	6.5		6.5
7	1	2	-1	1	6.5		6.5
8	5	4	1	1	6.5	6.5	
9	4	5	-1	1	6.5		6.5
10	5	4	1	1	6.5	6.5	
11	3	4	-1	1	6.5		6.5
12	2	5	-3	3	14.5		14.5
13	2	5	-3	3	14.5		14.5
14	2	3	-1	1	6.5		6.5
15	1	2	-1	1	6.5		6.5
16	3	2	1	1	6.5	6.5	
17	4	5	-1	1	6.5		6.5
					Sum:	19.5	100.5

After discarding the ties for CEO's 2 and 5, the effective sample size reduces to $n = 15$. Since the smaller sum is the one associated with the positive ranks, we define $T = \sum(+) = 19.5$. Looking up in a table we find that this is just below the 2% critical value of 20 for a two-sided test, which is at the same time the 1% critical value against the one sided alternative "median satisfaction before MBA $<$ median satisfaction after MBA". Note that we were not able to reject identical medians with such high significance with the sign test discussed earlier.

Wilcoxon Signed-Rank Test in Excel

You can get the Wilcoxon Signed-Rank Test from the 'T-Tests and Non-parametric Equations' tool of the Real Statistics toolbox. Choose 'Two paired samples' under Options and 'Non-parametric' under Test type.



The p-values from that tool are not exact. You can improve the approximation by checking 'Use ties correction' and 'Use continuity correction' under Non-parametic test options. You have also the option to ask for an exact test if $n \leq 25$, but the results are only correct if there are no ties.

3.3. Matched versus Independent Samples

Recall that both the sign test and Wilcoxon's signed rank test are tests for locations in matched samples, that is, the statistical units in both samples are assumed to be identical except for a certain characteristic, the impact of which we wish to test (often same objects in some before/after situation). The corresponding test for normally distributed observations or large samples ($n \geq 30$) is Student's t-test for matched samples.

The median test, on the contrary, tests for identical medians in independent samples and has already been discussed in a subsection of chapter 2 about the χ^2-tests for equality of proportions. We proceed with a more efficient test in independent samples, the Kruskal-Wallis test (2 samples: Mann-Whitney test).

3.4. The Kruskal-Wallis (W. rank sum) Test

Consider drawing k independent random samples, one for each out of k populations with observations on ordinal scale at least. The Kruskal-Wallis hypothesis test (also known as Wilcoxon rank sum test) is:

H_0: All k populations have the same distribution.

H_1: Not all k populations have the same distribution.

Although the test is stated in terms of distributions, it is most sensitive to differences in locations. Therefore, the procedure is actually used to test the ANOVA hypothesis of equality of k population means when the presumptions of ANOVA (normally distributed observations with equal variances in all populations) are not met.

In order to perform the test, we rank all data points in the entire set of all k populations from the smallest to the largest, without regard to which sample they come from. Then we sum all the ranks from each separate sample. Let n_i, $i = 1, \ldots, k$ denote the sample size from population i and let $n = n_1 + \cdots + n_k$. Defining R_i, $i = 1, \ldots, k$ as the sum of ranks from sample i, the Kruskal-Wallis test statistic is defined as:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1).$$

As long as each $n_i \geq 5$, H is approximately chi-square distributed with $k-1$ degrees of freedom under the null of equal distributions. Different distributions will lead to high values of H, so we reject the null hypothesis if the computed value of H exceeds $\chi_\alpha(k-1)$ for a given significance level α.

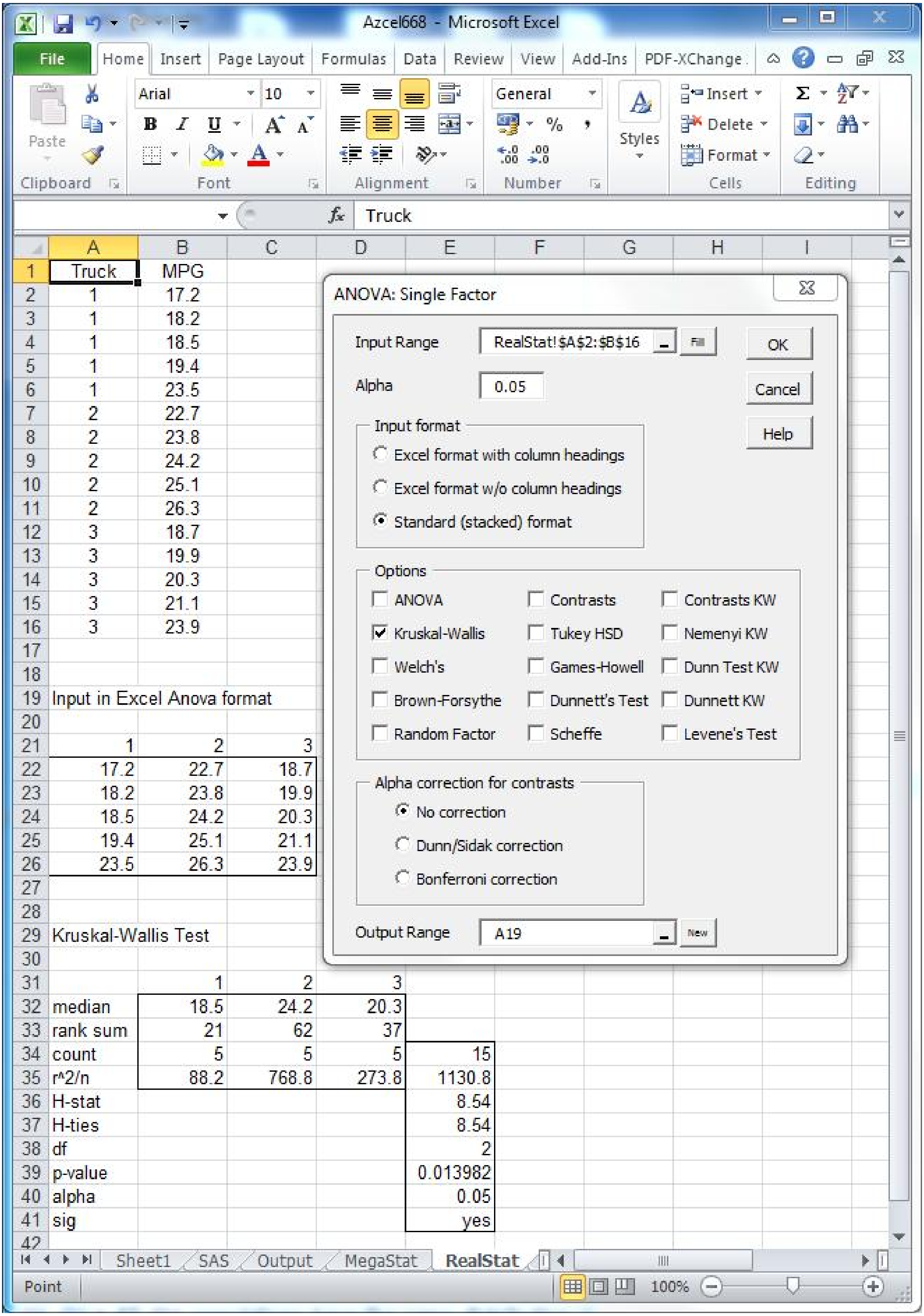
Example:

Three different trucks are tested for fuel efficiency as measured in miles per gallon (mpg).

Truck 1		Truck 2		Truck 3	
MPG	Rank	MPG	Rank	MPG	Rank
17.2	1	22.7	9	18.7	4
18.2	2	23.8	11	19.9	6
18.5	3	24.2	13	20.3	7
19.4	5	25.1	14	21.1	8
23.5	10	26.3	15	23.9	12
Sum:	21	Sum:	62	Sum:	37

$$\begin{aligned} H &= \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \\ &= \frac{12}{15 \cdot 16} \left(\frac{21^2}{5} + \frac{62^2}{5} + \frac{37^2}{5} \right) - 3 \cdot 16 \\ &= 8.54. \end{aligned}$$

The p-value of the test may be calculated in Excel as CHIDIST(8.54;2)=0.014, so the fuel efficiency of the different trucks is not identical.


Azcel668 - Microsoft Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-Ins
PDF-XChange
Paste
Clipboard
Arial
10
Font
Alignment
General
Number
Styles
Insert
Delete
Format
Cells
Editing
Truck
Truck MPG
1 17.2
1 18.2
1 18.5
1 19.4
1 23.5
2 22.7
2 23.8
2 24.2
2 25.1
2 26.3
3 18.7
3 19.9
3 20.3
3 21.1
3 23.9
Input in Excel Anova format
1 2 3
17.2 22.7 18.7
18.2 23.8 19.9
18.5 24.2 20.3
19.4 25.1 21.1
23.5 26.3 23.9
Kruskal-Wallis Test
1 2 3
median 18.5 24.2 20.3
rank sum 21 62 37
count 5 5 5 15
r^2/n 88.2 768.8 273.8 1130.8
H-stat 8.54
H-ties 8.54
df 2
p-value 0.013982
alpha 0.05
sig yes
ANOVA: Single Factor
Input Range
RealStat!$A$2:$B$16
Fill
OK
Alpha
0.05
Cancel
Help
Input format
Excel format with column headings
Excel format w/o column headings
Standard (stacked) format
Options
ANOVA
Contrasts
Contrasts KW
Kruskal-Wallis
Tukey HSD
Nemenyi KW
Welch's
Games-Howell
Dunn Test KW
Brown-Forsythe
Dunnett's Test
Dunnett KW
Random Factor
Scheffe
Levene's Test
Alpha correction for contrasts
No correction
Dunn/Sidak correction
Bonferroni correction
Output Range
A19
New
Sheet1
SAS
Output
MegaStat
RealStat
Point
100%

Remarks:

- You can get the Kruskal-Wallis test in excel from the 'ANOVA: Single Factor' tool, which you get within the Real Statistics toolbox from 'Analysis of Variance'.

- The less efficient median test was not applicable in this case because all expected frequencies were less than 5.

- In the case of only 2 samples, the test-statistic for the Kruskal-Wallis test may be written in alternative ways, known as the Mann-Whitney U test or (Wilcoxon) rank sum test. But the test results (p-values) remain unchanged.