

Kruskal Wallis Test Calculator

Followed by post-hoc Dunn's test

Kruskal Wallis calculator with multiple comparisons, effect size, test power, outliers, and R syntax.

Significance level (α):

0.05

Outliers:

Included

Effect size (offsets):

0.3

Correction Method:

Bonferroni

Multiple comparisons method

Dunn's

Digits:

4

☐ Step by step

- ☒ Enter raw data directly
- ☐ Enter raw data from excel

components	Achieve quality attributes	Implement technologies	factors	Compare and evaluate	multiple technologies
4	4	3	3	3	3
3	4	3	2	3	4
3	4	4	3	3	3
4	5	4	2	2	4
3	4	4	4	5	4
2	4	3	4	4	4
3	4	5	2	3	4
3	2	5	3	4	4
4	4	5	3	4	5
4	4	3	4	3	3
3	4	4	4	2	4
4	3	3	3	2	4
4	3	4	3	3	
3	3	4	3	2	
4	4	2	2	2	

Header: you may rename 'Group1', 'Group2', etc.
Data: use or (comma) or as delimiters.
The tool ignores empty cells, non-numeric cells, or empty columns.

Calculate

Insert column

Delete column

Clear

Load last run

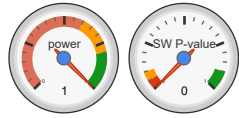
Load example

Reporting results in APA style

The Kruskal-Wallis H test indicated that there is a significant difference in the *dependent variable* between the different *groups*, $\chi^2(5) = 47.43$, $p < .001$, with a mean rank score of 215.18 for Design using patterns and components, 260.57 for Achieve quality attributes using components design", 222 for Implement technologies and components design, 162.4 for Analyze decision factors, 188.94 for Compare and evaluate technologies, 258.21 for Design using multiple technologies.

The Post-Hoc Dunn's test using a Bonferroni corrected alpha of 0.0033 indicated that the mean ranks of the following pairs are significantly different: **x₂-x₄**
x₂-x₅

[How to do with R?](#)



Groups:	Design using patterns and components	Achieve quality attributes using components design"	Implement technologies and components design	Analyze decision factors	Compare and evaluate technologies	Design using multiple technologies
Skewness:	-0.3033	-0.3771	0.1013	0.2847	0.3444	-0.06298
Skewness Shape:	<div>▲ Potentially</div>	<div>▲</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>
Excess kurtosis:	-0.3755	-0.2029	-0.7983	-0.7085	-0.628	0.6545
Tails Shape:	<div>▲ Potentially</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>	<div>▲ Potentially</div>
Normality	0.00001038	9.882e-13	0.01739	2.924e-7	1.898e-7	0.002909
Outliers:						5
Median:	4	4	3.5	3	3	4
Sample size (n):	50	189	20	83	87	12
Rank sum (R):	10759	49247	4440	13479	16437.5	3098.5
R ² /n:	2315121.62	12832100.58	985680	2188957.12	3105648.348	800058.5208

- Kruskal Wallis
- Calculators
- [One-way ANOVA](#)
- [Mann Whitney U](#)
- [Levene's test](#)
- [Two Way ANOVA](#)
- [Two Sample T-Test](#)
- [Two Sample Z-Test](#)

Kruskal-Wallis-test, using Chi-Square(df:5) distribution (right-tailed)[\[Validation\]](#)

1. H₀ hypothesis

Since the p-value < α, H₀ is rejected.
Some of the groups' mean ranks consider to be not equal.

In other words, the difference between the mean ranks of some groups is big enough to be statistically significant.
When selecting a value from each of the groups, there are some groups with a higher probability of containing the highest value than others.

2. P-value

The p-value equals **4.632e-9**, (P(x≤47.4347) = 1). It means that the chance of type I error (rejecting a correct H₀) is small: 4.632e-9 (4.6e-7%). The smaller the p-value the more it supports H₁.

3. Test statistic

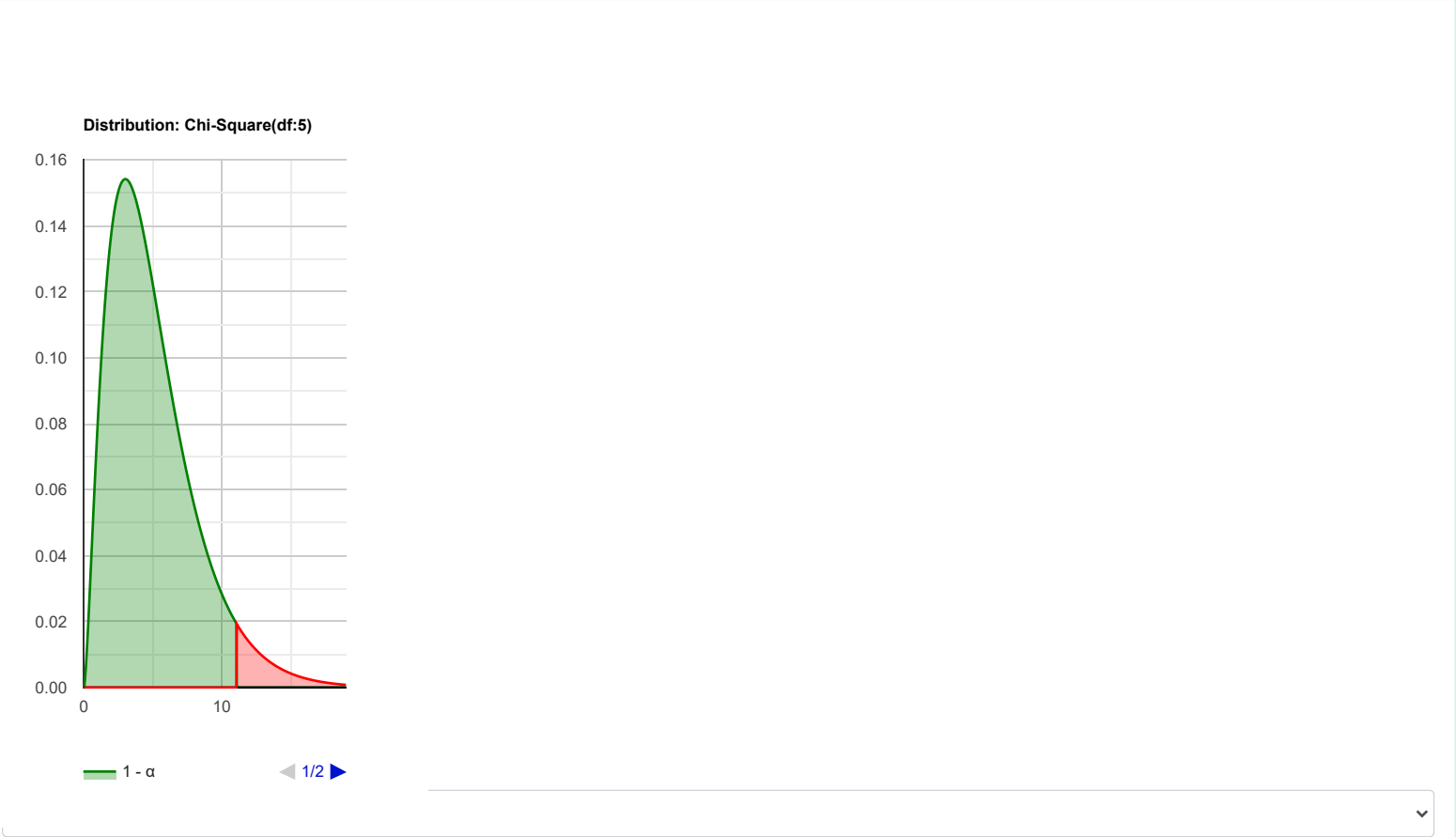
The test statistic **H** equals **47.4347**, which is not in the 95% region of acceptance: [0, 11.0705].

4. Effect size

The observed effect size η² is **medium, 0.098**. This indicates that the magnitude of the difference between the average is medium.

5. Multiple comparisons

The mean ranks of the following pairs are significantly different: **x₂-x₄ x₂-x₅**



Validation

● Test power

The test priori power is strong **1**

● Normality

The normality is **not** an assumption for the Kruskal-Wallis Test! We only check the normality to know if you could use a better test.

The normality was checked based on the [Shapiro-Wilk Test](#). (α=0.05)

When running the SW test on the residuals, the p-value is 0.000004744.

The ANOVA test is more powerful than the KW test and considered robust for moderate violation of the normality assumption. When checking the grrroups with a small sample size, less than 30, **2** groups don't have normally distributed data (the smaller p-value is 0.0000104).

The KW test is probably the correct test.

Multiple comparisons

Compares any pair of groups using the Kruskal Wallis test. In this case, the test is identical to the Mann-Whitney U test with normal approximation.
If you won't correct the significance level (α) and all the comparisons are independent, then the [type I error](#) may get to 0.5367. 1 - (1 - 0.05)¹⁵ = 0.5367.
This is the worse case, since usually some aspects of the multiple tests are in common, and the type I error will be lower.
In this case the Bonferroni correction would be an over correction and would reduce the test power.

<https://www.statskingdom.com/kruskal-wallis-calculator.html>

Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Compare and evaluate technologies	3	127
Design using multiple technologies	3	127
Design using multiple technologies	3	127
Design using multiple technologies	3	127
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Design using patterns and components	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5
Achieve quality attributes using components design"	4	288.5

$R_3 = 27.5 + 27.5 + 127 + \dots + 409.5 + 409.5 + 409.5 = 4440.$

$R_4 = 27.5 + 27.5 + 27.5 + \dots + 409.5 + 409.5 + 409.5 = 13479.$

$R_5 = 27.5 + 27.5 + 27.5 + \dots + 409.5 + 409.5 + 409.5 = 16437.5.$

$R_6 = 127 + 127 + 127 + \dots + 288.5 + 288.5 + 409.5 = 3098.5.$

$n = n_1 + n_2 + \dots + n_k = 441$

$$H' = \frac{12}{n(n+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \dots + \frac{R_k^2}{n_k} \right) - 3(n+1)$$

$$H' = \frac{12}{441(441+1)} \left(\frac{10759^2}{50} + \frac{49247^2}{189} + \frac{4440^2}{20} + \frac{13479^2}{83} + \frac{16437.5^2}{87} + \frac{3098.5^2}{12} \right) - 3(441+1) = 42.3976$$

$$H = \frac{H'}{1 - 0.1062} = 47.4347$$

$\text{MeanRank}_1 = 10759 / 50 = 215.18.$

$\text{MeanRank}_2 = 49247 / 189 = 260.5661.$

$\text{MeanRank}_3 = 4440 / 20 = 222.$

$\text{MeanRank}_4 = 13479 / 83 = 162.3976.$

$\text{MeanRank}_5 = 16437.5 / 87 = 188.9368.$

$\text{MeanRank}_6 = 3098.5 / 12 = 258.2083.$

Kruskal Wallis Test

The Kruskal-Wallis test also called one-way ANOVA on ranks is a non-parametric test.

Use the Kruskal-Wallis test calculator when your data doesn't meet the assumptions of the [one-way ANOVA test calculator](#).

The kruskal wallis test online checks the null assumption that when selecting a value from each of 'n' groups, each of these groups will have an equal probability of containing the highest value.

Target: To check if the difference between the ranks of two or more groups is significant, using a sample data

When the groups have a similar distribution shape, the null assumption is stronger and states that the medians of the groups are equal. When performing the Kruskal Wallis test, we try to determine, if the difference between the ranks reflects a significant difference between the groups, or is due to the random noise inside each group. The Chi-square statistic is an approximation for the exact calculation.

Right-tailed the Kruskal Wallis test can use only the right tail. [Why?](#)

Hypotheses

$H_0: MR_1 = .. = MR_k$

$H_1: \text{not}(MR_1 = .. = MR_k)$

MR - Mean rank.

Test statistic

$$H' = \frac{12}{n(n+1)} \sum \left(\frac{R_j^2}{n_j} \right) - 3(n+1)$$

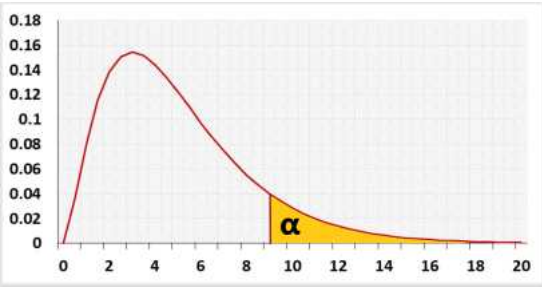
$$H = \frac{H'}{1 - \text{correction}}$$

R_j - the rank sum of group j.

n_j - the sample size of group j.

n - the total sample size across all groups, $n = n_1 + \dots + n_j$.

χ^2 distribution



Assumptions

- Independent samples from independent groups. One subject can't be in more than one group.
- The dependent variable is ordinal variable or continuous variable
- Two or more groups (the independent variable is categorical variable with two or more values)

Required Sample Data

- Sample data from all compared groups.

Multiple comparisons

Even if we know that not all the ranks are equal, we don't know which groups are not equal, hence we run a [Multiple comparisons](#) test to compare all the pairs.

We support two methods for the Multiple comparisons:

- **Dunn's test calculator** - takes into consideration the total number of groups (k) even when comparing only two groups
- **Mann Whitney U test** - the calculator uses the normal approximation of the Mann Whitney U test and supports the same results as the Kruskal Wallis test with two groups.

R Code

The following R code should produce the same results

```
if(!"MultNonParam" %in% installed.packages()){install.packages("MultNonParam")}
library(MultNonParam)
if(!"dunn.test" %in% installed.packages()){install.packages("dunn.test")}
library(dunn.test)
x1 <- c(3,2,3,4,3,4,4,4,3,3,4,4,4,5,5,4,4,4,4,3,5,4,4,3,4,2,2,5,4,4,3,3,2,2,4,3,3,4,3,2,3,3,4,4,3,4,4,3,4)
```

The dunn.test results show half of the p-values, that you should compare to half alpha.
This calculator results show full p-values.
When using corrections, the dunn.test corrects the p-value while this calculator corrects the alpha.