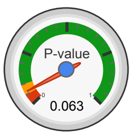
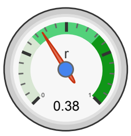
Wilcoxon Signed Rank test calculator

Reporting results in APA style

Results of the Wilcoxon Signed-Rank test indicated that there is a non-significant medium difference between Before (*Mdn* = 4 ,*n* = 56) and After (*Mdn* = 4 ,*n* = 56), *Z* = 1.9, *p* = .063, r = 0.4.

Parameter	Value
P-value	0.06295
Surprisal (S-value)	3.9897
Effect Size (r)	0.3796
Z	1.8596
W, (W-, W+)	89, (89, 211)
Number of pairs (N)	56
Non-zero difference pairs (n)	24
Ties Correction	166.5
S.E	32.5346
Average of differences ($\bar{x}_d$)	0.1786
SD of differences (S_d)	0.8966
Normality p-value	6.912e-8
Skewness	-1.3081
Skewness Shape	**Asymmetrical**, left/negative (pval=0)
Excess kurtosis	4.3977
Kurtosis Shape	**Leptokurtic**, long heavy tails (pval=0)
Outliers	-3, -3





Wilcoxon Signed-Rank-test, using Z distribution (two-tailed)

[Validation]

Because the data contains ties, equal differences (C_{ties} = 166.5), we use the **normal approximation**.

1. H_0 hypothesis

Since the p-value > α, H_0 can not be rejected.
The population's change is considered to be equal to the expected change (0).
In other words, the difference between the sample change and the expected change is not big enough to be statistically significant.
A non-significance result can not prove that H_0 is correct, only that the null assumption can not be rejected.

2. P-value

The p-value equals **0.06295**, ($P(x \le 1.8596) = 0.9685$). It means that the chance of type I error, rejecting a correct H_0, is too high: 0.06295 (6.29%). The larger the p-value the more it supports H_0.

3. Test statistic

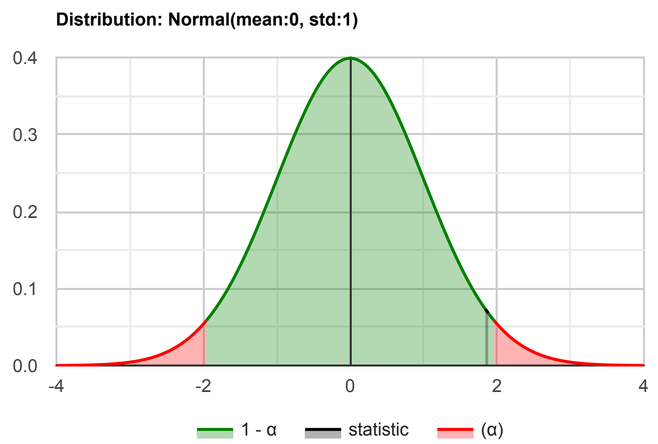
W can get values in the following range:[0, 300]
W+ = 211, W+ ~ N(150, 32.5346²).
The test statistic **Z** equals **1.8596**, which is in the 95% region of acceptance: [-1.96, 1.96].

4. Effect size

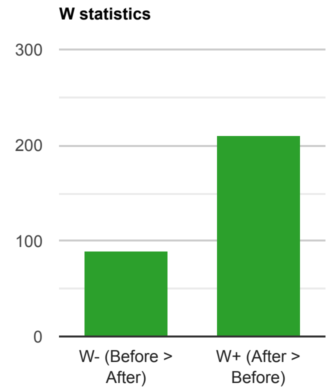
The observed effect size **r** is **medium**, **0.38**. This indicates that the magnitude of the difference between the mean ranks is medium.

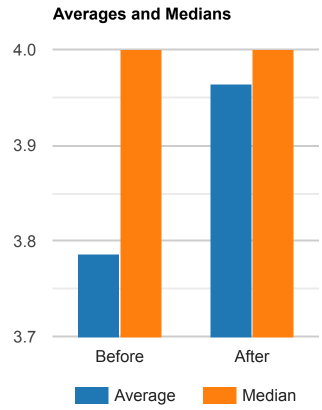
The observed **common language effect size** is **0.3**, this is the probability that a random value from **Before** is greater than a random value from **After**.

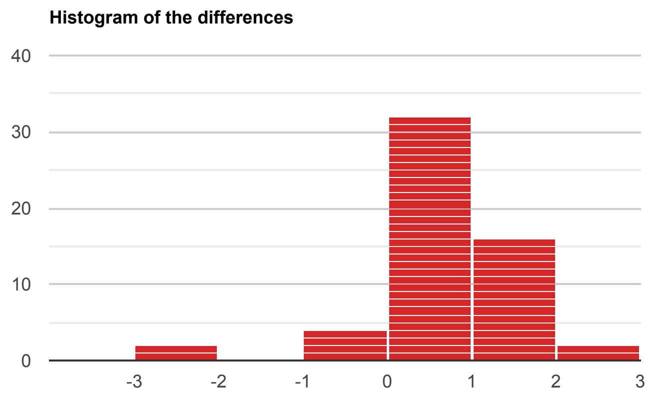
If you like the page, please share or like. Questions, comments and suggestions are appreciated. (**statskingdom@gmail.com**)



Significance level (α)


W statistics
300
200
100
0
W- (Before > After)
W+ (After > Before)


Averages and Medians
4.0
3.9
3.8
3.7
Before
After
Average
Median


Histogram of the differences
40
30
20
10
0
-3
-2
-1
0
1
2
3

Validation

The data doesn't meet all the Wilcoxon signed-rank assumptions.

Sample size

The number of pairs with a non-zero difference is **24**.
For the normal approximation you usually need at least sample size 16.

Outliers

Outliers' detection method: Tukey Fence, k=1.5.
The difference column contains 2 potential outliers, which is 3.57% of the observations. (-3, -3).
The **Wilcoxon signed-rank**-test is robust to the presence of outliers.
Outliers have **minimal impact** on non-parametric tests and can be optimal for outlier-rich data, yet identifying the cause of outliers is important as they may suggest a normal data distribution.

Normality

Normality is **not** an assumption for the Wilcoxon Signed-Rank test! We only check for normality to determine if a better test could be used. Normality was assessed using the Shapiro-Wilk Test with a significance level of α=0.05
When this test was run on the differences, the resulting p-value was **6.912e-8**. It is assumed that the data distribution is not normal. In other words, the difference between the data sample and the normal distribution is big enough to be statistically significant.
Since the number of pairs is bigger than 29 (which is a general guideline), you can use the **paired t** test, which is considered a more powerful test than the Wilcoxon Signed-Rank test.