# Natural Language Processing System Performance Settings

Data Analysis | Significance Test | Effect Size | Post-test Power Analysis | Download Test Results

System File: ?
Select File uploaded: score

Evaluation Unit Size: ?
1

Target Statistic: ?
(•) Mean ( ) Median

Random Seed: ?
Shuffle

Run

## Summary of statistics

| Score | mean | median | std_dev | min | max |
|---|---|---|---|---|---|
| score1 | 0.28464 | 0.25552 | 0.15964 | 0.02206 | 1.00000 |
| score2 | 0.28051 | 0.24905 | 0.15694 | 0.01216 | 1.00000 |
| difference | 0.00413 | 0.00070 | 0.11228 | -0.83215 | 0.51108 |
| difference (partitioned) | 0.00413 | 0.00070 | 0.11228 | -0.83215 | 0.51108 |

## Test statistic recommendation

**Normality:**The data distribution does not pass the normality test.

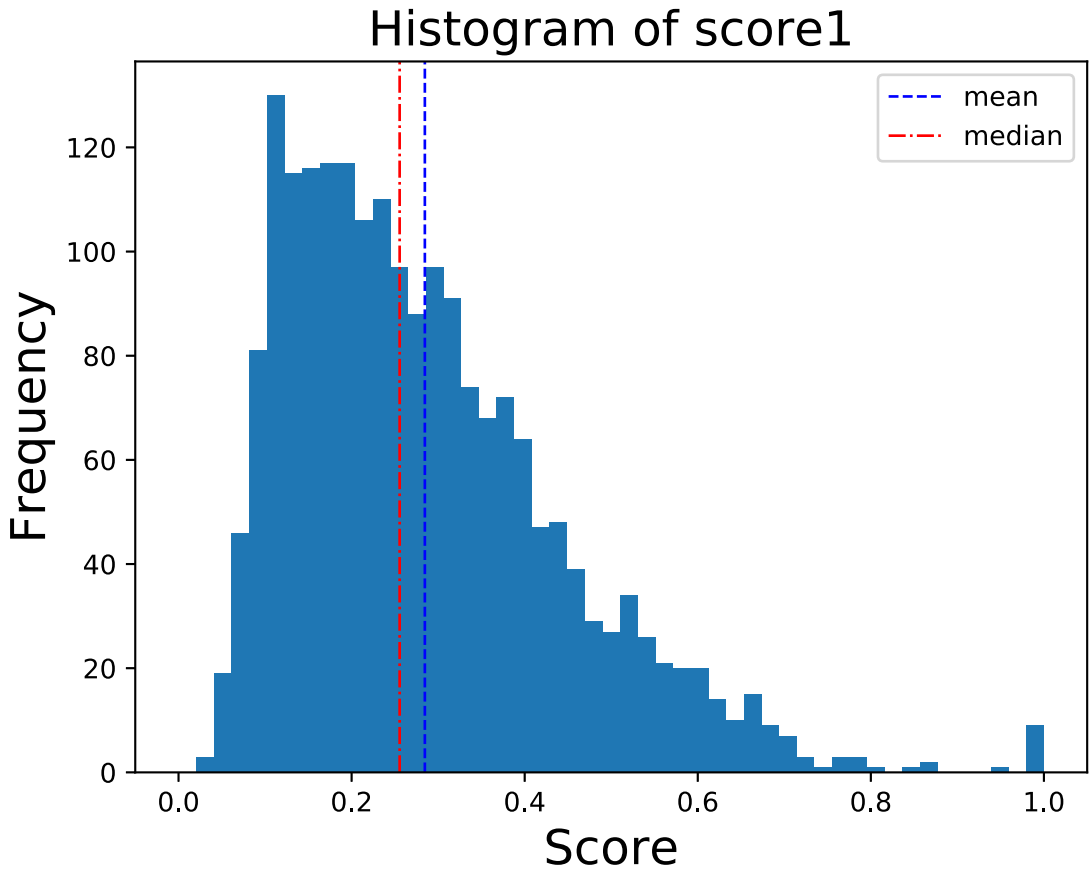
**Skewness:** The data distribution does not exceed the skew threshold.

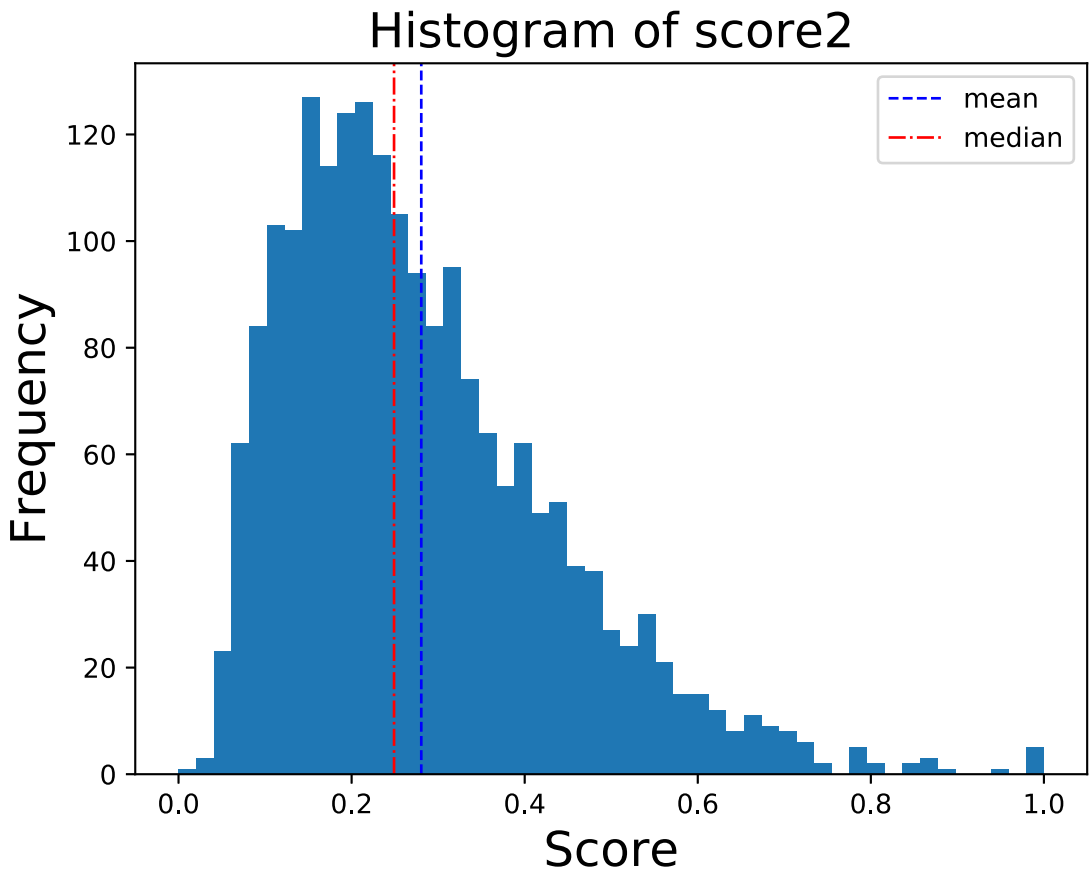
**Test statistic:** Based on the skewness, the recommended test statistic to use is: **mean**

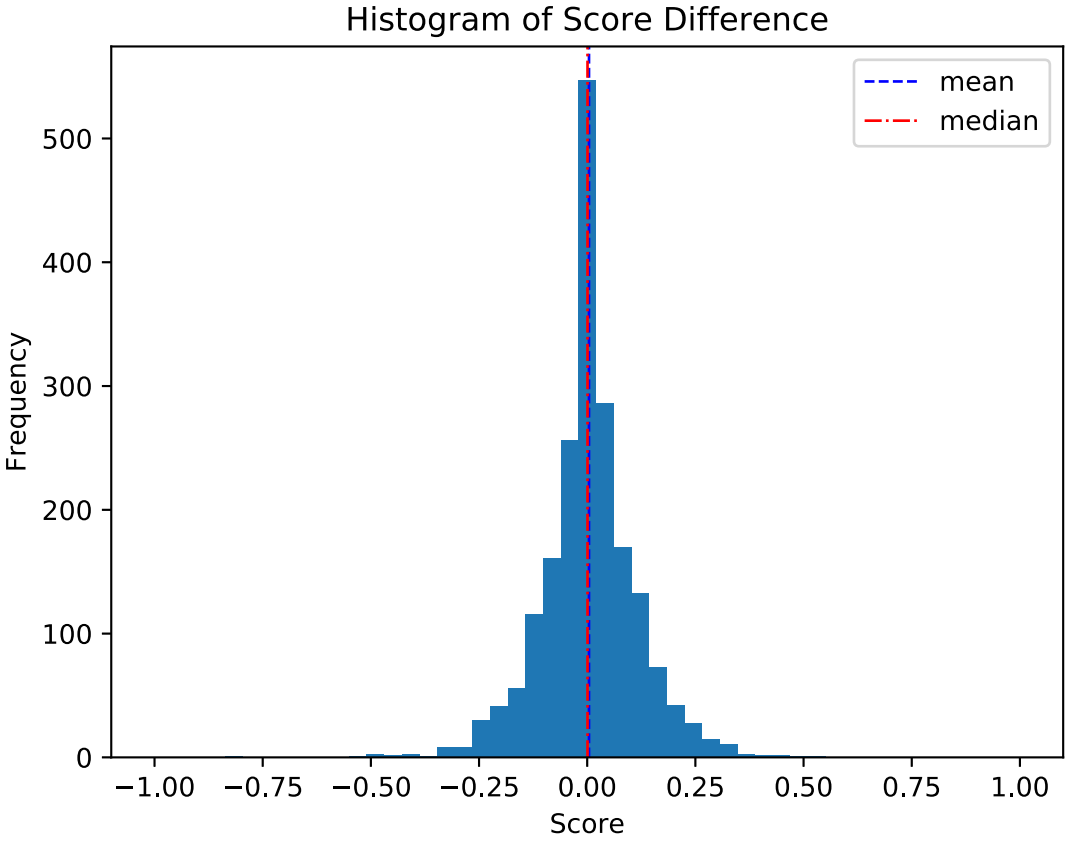
## The following table shows the recommended tests.

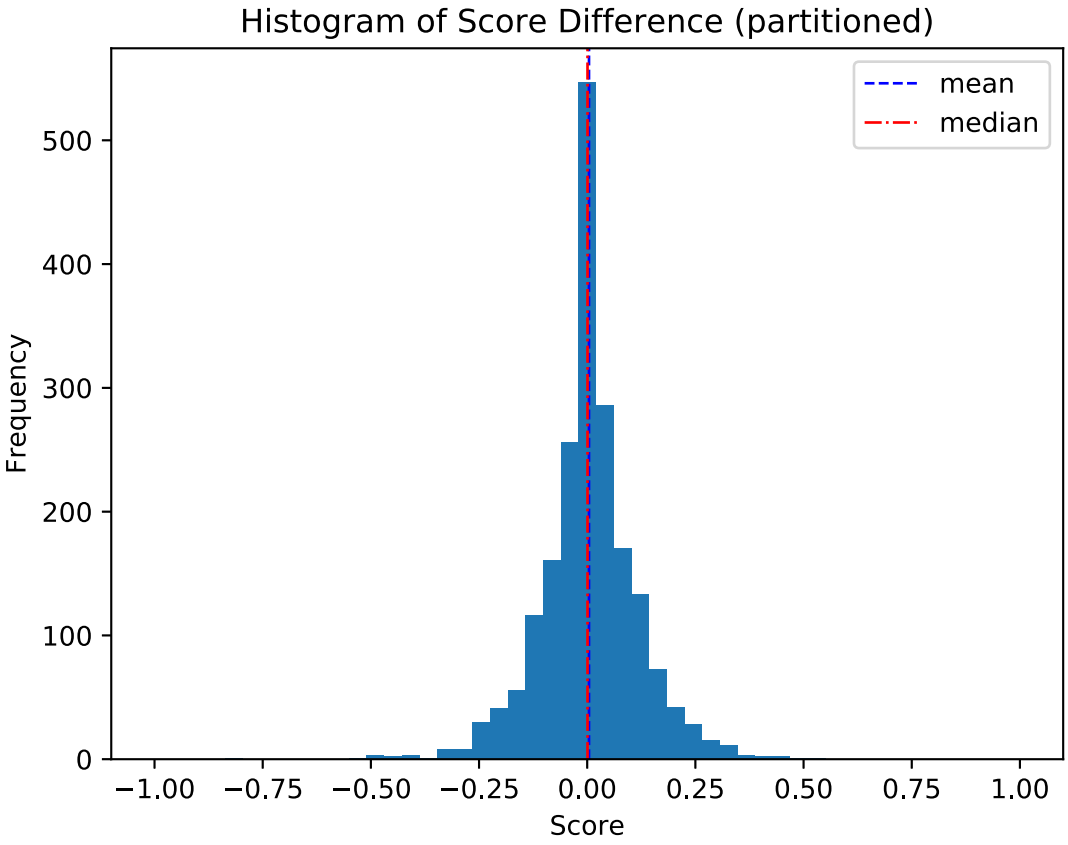
| Test | Reason |
|---|---|
| bootstrap | The bootstrap test based on t ratios does not assume normality, and thus is appropriate for testing for mean difference. |
| permutation | The sign test calibrated by permutation based on mean difference is nonparametric and does not assume normality. |
| wilcoxon | The Wilcoxon signed-rank test can be used for this case, but since it is a nonparametric test, it |

| | |
|---|---|
| | has relatively low statistical power. Also the null hypothesis is that the the pairwise difference has location 0. |
| sign | The (exact) sign test can be used for this case, but it has relatively low statistical power due to loss of information. Also, the null hypothesis is that the median is 0. |
| t | The student t test may be appropriate for non-normal data if the sample size is large enough, but the iid assumption must hold. |


Histogram of score1
mean
median
Frequency
Score
0
20
40
60
80
100
120
0.0
0.2
0.4
0.6
0.8
1.0


Histogram of score2
mean
median
Frequency
Score
0
20
40
60
80
100
120
0.0
0.2
0.4
0.6
0.8
1.0


Histogram of Score Difference
mean
median
Frequency
0
100
200
300
400
500
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Score


Histogram of Score Difference (partitioned)
mean
median
Frequency
0
100
200
300
400
500
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Score