

Natural Language Processing System Performance

Settings

Data Analysis

Significance Test

Effect Size

Post-test Power Analysis

Download Test Results

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.

Choose the significance test to run

☐ bootstrap

☐ permutation

☐ wilcoxon

☐ sign

☐ t

Significance level α : ?

0.00

Run