Research Tools: The LL Calculator

The log-likelihood calculator in *The Prime Machine* has been designed to make it very easy for you to calculate and display statistics based on your own figures (you don't need to use tPM corpora for these; the calculator is not connected to the corpus or other online data.

You can find the **LL Calculator** on the **Research Tools** Tab. It does not matter whether you are online or working offline.

This button simply swaps over the figures you entered in the two blue columns. It makes it easy to reverse your test.

To use the LL Calculator, simply type in (or paste in) 4 numbers:

Study Corpus Frequency; Study Corpus Size; Reference Corpus Frequency; Reference Corpus Size. **Results will update automatically.**

You can use the other settings on this page to:

- Change the heading of the first column from "Words" to "Features" (this has no effect on the results).
- Increase the number of rows. Note: Log-likelihood values are based on independent tests for each row. Each row is treated as a separate test. It is possible using other tools to calculate log-likelihood for multiple rows, but this cannot be done in tPM.
- Change the rightmost column to display Bayes meaning (strong evidence, etc.), BIC Values (≥2, ≥6, ≥10), or a comment based on the p value (exceeds 15.13 threshold, etc.).
- Change columns 3 and 5 to show values as percentages rather than as per million.
- Show/Hide a column showing arrows to indicate whether the frequency is higher or lower in the Study corpus compared to the Reference corpus.
- Show/Hide negative log-likelihood values.

This calculator was inspired by the pre-2015 UCREL Log-likelihood online wizard http://ucrel.lancs.ac.uk/llwizard.html. Results for Log-likelihood and Bayes should be the same, except tPM's BIC values may be different for very small values of BIC when the total counts are very high. While it is generally true that negative BIC values will correspond to evidence in favour of the null hypothesis, the formula for BIC involves subtracting the natural log value of the combined corpus sizes, and in tPM this is not permitted to go below 0. Negative BIC scores are only obtained in tPM when the normalized frequency is lower in the Study Corpus than it is in the Reference Corpus.

Support

The Prime Machine is still undergoing development.

For further information see http://help.theprimemachine.com

Last Updated: Friday, April 13, 2018