Testing Differences Statistically with the Leiden Ranking

The Leiden Ranking 2011/2012 provides the *Proportion top-10% publications* ($PP_{top\ 10\%}$) as a new indicator. This indicator allows for testing performance differences between two universities for statistical significance.

On December 1, 2011, the Centre for Science and Technology Studies (CWTS) at Leiden University launched the Leiden Ranking 2011/2012 at http://www.leidenranking.com/ranking.aspx. The Leiden Ranking 2011/2012 measures the scientific performance of 500 major universities worldwide. The *Proportion top-10% publications* ($PP_{top\ 10\%}$) is added as a new indicator of impact. This indicator corresponds with the *Excellence Indicator* (*EI*) recently introduced in the SCImago Institutions Rankings (at http://www.scimagoir.com/pdf/sir_2011_world_report.pdf).

Whereas SCImago uses *Scopus* data, the Leiden Ranking is based on the *Web-of-Science* data of Thomson Reuters. In addition to the "stability intervals" provided by CWTS, values for both $PP_{top\ 10\%}$ and *EI* can be tested statistically for significant differences from expectation. Furthermore, the statistical significance of performance differences between universities can be tested by using the *z*-test for independent proportions (Bornmann *et al*., in press; Sheskin, 2011, pp. 656f.).

An Excel sheet can be downloaded from http://www.leydesdorff.net/leiden11/leiden11.xls into which the values for this indicator ($PP_{top\ 10\%}$) can be fed in order to obtain a *z*-value. The example in the download shows the results for Leiden University when compared with the University of Amsterdam (not significantly different; $p > 0.05$), and for Leiden University when compared with the expectation (the value is significantly above the expectation; $p < 0.001$). The values in the sheet can be replaced with values in the ranking for any university or any set of two universities.

The *z*-test

The *z*-test can be used to measure the extent to which an observed proportion differs significantly from expectation, and whether the proportions for two institutions are significantly different. In general, the test statistics can be formulated as follows:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

where: n_1 and n_2 are the numbers of all papers published by institutions 1 and 2 (under the column "*P*" in the Leiden Ranking); and p_1 and p_2 are the values of $PP_{top\ 10\%}$ of institutions 1 and 2. The pooled estimate for proportion (p) is defined as:

$$p = \frac{t_1 + t_2}{n_1 + n_2}$$

where: t_1 and t_2 are the numbers of top-10% papers of institutions 1 and 2. These numbers are calculated (in the sheet) on the basis of "*P*" and "$PP_{top\ 10\%}$" provided by the Leiden Ranking. When testing observed versus expected values for a single sample, $n_1 = n_2$. In that case, p_1 is the value of the $PP_{top\ 10\%}$, $p_2 = 0.1$, and $t_2 = 0.1 * n_2$ (that is, the expected number in the top-10%).

An absolute value of *z* larger than 1.96 indicates the statistical significance of the difference between two ratings at the five percent level ($p < 0.05$); the critical value for a test at the one percent level ($p < 0.01$) is 2.576. However, in a series of tests for many institutions, a significance level higher than five percent must be chosen because of the possibility of a family-wise accumulation of Type-I errors (the so-called Bonferroni correction; cf. Leydesdorff *et al*., 2011).

In summary, it seems fortunate to us that two major teams in our field (Granada and Leiden University) have agreed on using an indicator for the Scopus and WoS databases, respectively, that allows for testing of statistically significant differences of scientific performances. Of course, there remains the problem of interdisciplinarity/multidisciplinarity when institutional units such as universities are ranked. This could be counteracted by field-normalization and perhaps by fractionation of citations (1/NRef) in terms of the citing papers (Zhou & Leydesdorff, 2011).

Loet Leydesdorff [1] & Lutz Bornmann,[2]

[1] Amsterdam School of Communication Research, University of Amsterdam, Kloveniersburgwal 48, NL-1012 CX, Amsterdam, The Netherlands; loet@leydesdorff.net.
[2] Max Planck Society, Administrative Headquarters, Hofgartenstr. 8, 80539 Munich, Germany; Lutz.Bornmann@gv.mpg.de.

References

Bornmann, L., de Moya-Anegón, F., & Leydesdorff, L. (in press). The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011. *Journal of Informetrics;* preprint available at http://arxiv.org/abs/1110.2305.

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables in citation analysis one more time: Principles for comparing sets of documents *Journal of the American Society for Information Science and Technology, 62*(7), 1370-1381.

Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures* (5th edition). Boca Raton, FL: Chapman & Hall/CRC.

Zhou, P., & Leydesdorff, L. (2011). Fractional counting of citations in research evaluation: A cross- and interdisciplinary assessment of the Tsinghua University in Beijing. *Journal of Informetrics, 5*(3), 360-368.