Indices.xls

What is it?

It's an Excel file that enables you to calculate easily the values of 3 indices that are widely used in the analysis of elections:

Least squares index (measure of disproportionality, i.e. the disparity between vote shares and seat shares);
Effective number of elective parties (measure of vote fragmentation);
Effective number of legislative parties (measure of seat fragmentation).

How does it work?

Download the file Indices.xlsx from this site (www.tcd.ie/Political_Science/people/michael_gallagher/ElSystems/index.php)

It's a standard MS Excel file. For the election you want to analyse, enter the votes and seats won by each party. The file then tells you:

% votes and % seats for each party;
value of least squares index (LSq);
effective number of vote-winning (Eff no Pv) and seat-winning (Eff no Ps) parties.

Describe in detail what I need to do

The file contains 11 templates into which the details of an election can be entered. The first has already been filled in as an example, with the seat and vote values of the parties at the 1999 election in Finland (excluding the 1 seat from Åland).

The first blank template begins in row 17.

Enter the total number of votes in cell C18, and the total number of seats in cell C19. Then, for the first party, enter its vote total in cell D18 and its seat total in D19, and repeat across for all the other parties. It doesn't matter in what order the parties are entered (except that 'Others' should be the last one entered, for reasons explained below). But it is essential, obviously, that for each party, its votes and seats appear in the same column; the order in which the parties' vote totals are listed must be the same as the order in which their seat totals are listed.

Having entered the data, what do I look for?

First, examine cells B18 and B19 (under the 'Check / indices' heading). These are check cells. They tell you if there are any errors in the data. They should return a value of zero. If they don't, that means that the numbers of votes (or seats) you have entered for each party don't add up to the number in the Total cell. Either there is an error in your data or you have made an error when entering them. (However, very minor errors – for example, if you are entering percentages and due to rounding the total adds to 99.9 rather than to 100.0 – are not going to make any difference to the value of the indices that is worth being concerned about.)

If there are no errors, at the most basic level the file now tells you the percentage vote (and seat share) of each party. These figures are found in the fourth (and fifth) row of each template. These percentages should, of course, add to 100, so the cells at the left of these rows (B20 and B21 in the first template) should equal zero.

More usefully, the three indices will now appear in bold at the bottom left-hand corner of the template. In the first template, the figure in cell B22 is the value of the least squares index; B23 contains the effective number of elective parties; B24 contains the effective number of legislative parties.

How many parties can the file cope with?
It's designed to deal with elections with up to 500 parties or independent candidates. In the unlikely event that you want to analyse an election with more than 500 parties, a basic knowledge of Excel should enable you to figure out how to do this. While it is rare to come across an election result with results reported for 100 parties, that is partly because all too often the votes for smaller parties are aggregated into an 'Others' bloc (see below for how to deal with this). For example, Spain's 2004 election was contested by 104 parties (only 24 won 0.1 per cent of the votes or more) and its 2008 election by 99 parties (23 won 0.1 per cent of the votes or more). Ideally, every party, no matter how small, and each independent candidate should be treated as a separate unit, given that that is how they contested the election. In some countries, especially those using single-member constituency system electoral systems, it is common for independent candidates to stand at elections, and given that the votes for each independent candidate should be entered separately, it's possible that the number of cases might exceed 200. As explained below, this bunching of others often does not matter, and in any case there are ways of dealing with it.

Of course, when there are fewer than 500 parties, you only have to enter figures for the parties that did contest the election; you don't need to fill out the rows by entering zeros under the other columns.

How many elections can it deal with?
The file that you can download has 10 'blank' templates. Of course, once you have the file, you can extend the number of templates indefinitely simply by copying the template, or by 're-using' each template. But if you 'fill up' the programme that you download, you can always download it again.

I don't have raw vote totals, just percentages
No problem. Just enter the percentages in the 'votes' row. It doesn't affect the calculations at all.

In the data source I'm using, the votes of a lot of small parties and independents are bunched together in one bloc labelled 'Others'. Does this matter?
It might. Ideally, you should aim for as much disaggregation as possible. The vote (and seat) total of every small party, even every independent candidate, should ideally be entered separately. If your data source isn't as comprehensive as this, then there are various ways to proceed.

(i) Effective number of parties.

(a) The easiest approach is to treat the 'Others' as if they comprised just one party, in other words to enter the votes and seats for 'Others' as if for a regular party. This will give a figure that is somewhat smaller than the 'correct' value (the one we would get if we had votes on every single party, no matter how small), but unless the votes for 'Others' is quite large – over 10 per cent, say – the error will be very minor and this approach is acceptable albeit not optimal.

(b) A little more complicated, and likely to produce a value even closer to the correct value (though very slightly larger than it), is to omit the 'Others', which has the same effect as assuming that the category of 'Others' is made up of a large number of very small parties – an assumption that in most cases is valid. This is done by modifying the Excel template for that election so that the sum of the values

concerned does not include the 'Others' component. For example, if the 'Others' vote and seat totals have been entered in column J ('Others' should always be the last 'party' entered), modify the sum of the squares (cell C23 in the first template) so that it sums only the values in columns D to I. But, if you do this, be sure to enter the votes and seats for Others; simply omitting them would be a mistake. This will give a value that will almost certainly be very close to the 'correct' value (the one we would get if we had votes on every single party, no matter how small). Alternatively, simply do not enter the votes and seats for Others in the spreadsheet.

(c) Adopt what Rein Taagepera terms a 'least components' approach and assume that none of the parties lumped together is larger than the smallest party whose result is reported separately. (For example, if the 'Others' bloc has 10.5 per cent of the votes and the smallest party reported separately has 1.1 per cent, treat the Others as if they are 19 separate parties each with 0.55 per cent (half the size of the smallest party whose vote share is reported) and one party with 0.05 per cent.)

Conclusion: in most cases, approach (a) can safely be adopted, though (b) is usually preferable, (c) is likely to give an even more accurate estimate, and the average of (b) and (c) usually delivers the most accurate estimates of all.

(ii) Least squares index.

(a) Now, treating the lumped Others as if they make up a single party would be a serious mistake and is likely to produce a value significantly different from the correct one, especially if this group is not very proportionally represented, as is often the case. Instead, there are two possibilities.

(b) The simplest approach is to omit 'Others', making the assumption that all the parties in this category are very small. Once again, this is achieved by modifying the Excel template so that the sum of the squares does not include the 'Others' component. For example, if the 'Others' vote and seat totals have been entered in column J ('Others' should always be the last 'party' entered), modify the sum of the squares (cell C22 in the first template) so that it sums only the values in columns D to I. But, if you do this, be sure to enter the votes and seats for Others; simply omitting them would be a mistake. In most cases this approach will give a value reasonably close to the 'correct' value.

(c) As when calculating the effective number of parties, it will usually be an improvement to take the figure given by the least components approach, which is outlined above.

Conclusion: when the size of 'Others' is small (below 1 per cent) approach (b) is acceptable; otherwise, approach (c) is preferred, ad the average of (b) and (c) is likely to lead to the most accurate estimates.

For more detailed advice, see Appendix B of '*The Politics of Electoral Systems*', edited by Michael Gallagher and Paul Mitchell (Oxford: Oxford University Press, paperback edition 2008): http://www.oup.co.uk/isbn/0-19-925756-6.

An example might clarify this

Take the case of the Burkina Faso election of 22 November 2020. Altogether this was contested by 126 parties. Only 22 of these won 0.5% or more of the national vote total; the other 104 parties collectively won 322,935 votes (11.53% of the total). The official results published by the Burkinabe Conseil Constitutionnel helpfully supply full details of the votes of every party, so we know what the 'correct' values of the indices are and can investigate how close various methods would bring us to these if, as is the case for some countries, the results available to us simply listed the votes received by each of the strongest 22 parties and then stated that 'Others' won 322,935 votes and 11.53% of the total. Since none of these 104 tiny parties won a seat, the value of Ns is known and hence we do not include it in the illustration below, but the values of LSq and Nv would be uncertain if a result were reported with all small parties lumped together in this way. The four sets of values would be:

1. The correct values – with full data on each of the 126 parties, we know that these are LSq 8.16, Nv 6.45.
2. Estimate by simply omitting the Others (in effect assuming that each won such a tiny number of votes that its vote–seat disparity is effectively zero and hence can be ignored) – this gives LSq 8.09, Nv 6.46.
3. Estimate by the least components approach. The smallest party that won more than 0.5% of the national vote received 21,443 votes, so assume that the 322,935 votes received by 'Others' were received by 30 parties each receiving 10,722 votes (half the number won by the smallest party for which we have data) and by one party winning the remaining 1,275 votes. This gives LSq 8.22, Nv 6.44.
4. Take the bunched votes for Others, ie treat Others as if they constituted a party or a bloc that won 11.53% of the national vote but no seats; this gives LSq 11.48, Nv 5.95.

Conclusion: the approach of treating the lumped Others as if they constituted a single party is unambiguously a mistake; this approach should never be adopted. The least components approach may give slightly more accurate figures than simply ignoring the Others, but it obviously entails more work, and just omitting the Others will give figures very close to the correct ones. Most accurate of all is usually to average the estimates from method 2 and method 3, though the improvement over each of those two approaches is likely to be only marginal.

What are these indices anyway?

They have become the standard measures of fragmentation and disproportionality respectively. The 'effective number' measure was devised by Rein Taagepera and Markku Laakso; it tells us how fragmented a voting population or a parliament is. The idea of 'least squares' has a long history in the natural sciences and in the social sciences; this particular index tells us how far the distribution of seats across parties differs from the distribution of votes. You can read more about both measures in Appendix B of '*The Politics of Electoral Systems*', edited by Michael Gallagher and Paul Mitchell (Oxford: Oxford University Press, paperback edition 2008).