

Iterative Proportional Fitting For A Two-Dimensional Table

By

Eddie Hunsinger
Demographer

Alaska Department of Labor and Workforce Development

May 2008

What Is IPF?

IPF stands for 'Iterative Proportional Fitting', and is sometimes referred to as 'Raking'.

IPF is a procedure for adjusting a table of data cells such that they add up to selected totals for both the columns and rows (in the two-dimensional case) of the table. The unadjusted data cells may be referred to as the 'seed' cells, and the selected totals may be referred to as the 'marginal' totals.

The next slide is meant to show what exactly the above text means, and what before and after IPF tables look like...

What Is IPF?

Before After

Selected
Marginal (Column) Totals

		11	9	8
5	1	2	1	
15	3	5	5	
8	6	2	2	

Selected
Marginal (Row) Totals

Seed Cells

IPF →

		11	9	8
5	1.51	2.31	1.18	
15	4.20	5.35	5.45	
8	5.28	1.34	1.37	

Fitted Cells

How To Do IPF?

So how were those fitted cells found?

You'll see that it's very simple. No math beyond arithmetic is required to do it.

For that two-dimensional (row and column marginals) table, the next slide describes exactly how it was done...

How To Do IPF?

Step 1: Each row of seed cells was proportionally adjusted to equal the marginal row totals (specifically, each cell was divided by the actual sum of the row of cells, then multiplied by the marginal row total).

Step 2: Each column of (already row-adjusted) cells was proportionally adjusted to equal the marginal column totals.
This is the end of the first 'Iteration'.

Steps 3 and higher: The above steps were repeated until the selected level of convergence was reached.

Make Sense? If Not:

Here is a link to two-dimensional IPF done step-by-step in a simple Excel spreadsheet (just follow the 'Iteration' numbers).

IPF was proposed in part of a 1940 research paper, by Deming and Stephan, as a way to make an un-biased guess at individual cells of data, based on knowledge from samples and estimated ('known') totals. Here is a link to that paper, with description of the procedure (see pp 439-442).

I've learned that there was an earlier (1937) documentation of the technique by a researcher named Kruithoff, but I can't find a link to that paper.

Four Important Points

1. The marginal column totals and the marginal row totals must sum to the same value for IPF to work.
 2. IPF will work with any non-zero seed values, but a more (proportionally) accurate seed will probably provide a more accurate result.
 3. IPF yields values with decimals. If you need to get rid of decimals and have the cells still add up, a separate rounding procedure will usually be necessary.
 4. IPF adjustments converge to fit marginal totals within any selected level of closeness, but in most cases will never exactly sum to all of the pre-determined marginal totals.

Two More Important Points

5. IPF cannot work if any of the marginal cell values are equal to zero. A workaround proposed by Nels Tomlinson is to make the value for such cells very small compared to all of the other cells in the marginal (e.g. .001 if all of the other marginal cells are whole numbers).

6. No adjustment will be made to seed cells with a value of zero. If you would like to allow adjustment for such a cell, a proposed workaround is to make its value very small compared to all of the other seed cells (e.g. .001 if all of the other seed cells are whole numbers).

Two References

Deming, W. and Stephan, F. (1940) 'On least square adjustment of sampled frequency tables when the expected marginal totals are known.' *Ann. Math. Statist.*, 6, 427–444.

Norman, P. (1999) 'Putting iterative proportional fitting on the researcher's desk.' Working Paper 99/3. School of Geography, University of Leeds, Leeds.

The End

So that's all I have to tell about IPF for a two-dimensional table, and two-dimensional tables is as far as most IPF users go. BUT, you can also perform IPF on more-than-two-dimensional tables.

[Here is a link to my web page with more IPF information.](#)