Hey Marta –

Just to make sure I’ve got this right, I’ll share with you my ‘written’ interpretations of the various measures. So, I’ll repeat the template definition of the Foster-Greer-Thorbecke (FGT) class of poverty indices, and define Pα as:

where n is the sample size, i subscripts the family or individual, y is the relevant measure of welfare, z is the poverty line, and I is an indicator function which takes the value of one if the statement is true and zero otherwise. Technically, since I didn’t include probability weights, these are all sample estimates, but let’s just assume it’s self-weighted and I’ll refer to these as estimates of the population.

If we refer to the income gap as the difference between the income level of the poor and the poverty line, then P1 gives us the mean income gap expressed as a proportion of the poverty line, where the mean is formed over the entire population and counts the nonpoor as having zero income gap. The key element here is n in the denominator which means we take the average over the entire population. Just for using below, I’ll introduce np as the population of poor people (the cumulative sum of the indicator function). Since P0 tells us the proportion of the population that is poor, if we were to divide either P1 (or P2) by P0, the interpretation flips from being over the entire population to being over the population of poor people.

OK, much too wordy. So now I’ll paste in your email and provide my reactions.

ren poverty\_gap poverty\_gap\_index  /\*Here I’m assuming poverty\_gap\_index is P1; this is standard so pretty sure that must be correct \*/

gen total\_shortfall= poverty\_gap\_index \*poverty\_line\_215 \* population

So, if I follow correctly, this is P1\*z\*n, which I agree, should give

where the indicator function ensures that the sum is taken only for the poor. Total\_shortfall seems like a reasonably apt name.

gen avg\_shortfall=total\_shortfall/ headcount\_215

So, I’m pretty sure headcount\_215 is the total count of people who are poor (at IPL), or in my shorthand, np. If this is wrong, and headcount\_215 is P0 at the IPL, then what I write below is wrong. Anyway, if it’s np, then avg\_shortfall seems to be: P1\*z\*n / np. Or, more usefully,

Here, I’d probably refer to this as average shortfall of the poor, but avg\_shortfall seems descriptive. So, from this we’d learn something like the average poor person falls one dollar short of the poverty line.

gen income\_gap\_ratio=(total\_shortfall/headcount\_215) / poverty\_line\_215

Wow. This reads like really awkward code to me, but if I follow correctly, then I think this is:

*So, following the example above, the shortfall of a dollar would now be expressed as a ratio of the poverty line. So, if z=2.15, then this would tell us that the dollar shortfall is 0.465. I don’t recall exactly where I read Ravallion’s discussion on this, but calling this the income gap ratio seems reasonable.*

*Maybe the most useful thing to note, is that this should be P1/P0. Instead of taking the poverty gap averaged over the population, P1/P0 takes the average over the population of poor people. I tend to find P1/P0 as one of the more useful stats to convey to the general public. It also helps people to understand what P1 is NOT.*

 //I think this is Ravallion’s income gap ratio – or at least that is what their code says

Does that look correct to you? As I said I might be interpreting their code wrong, but if you spotted anything off I could try to figure out what the issue starting from there.

So, everything you’ve written seems right to me. I’d argue from a coding standpoint, we might want to do this differently. Similarly, we’ll want to use more uniform naming conventions (the ones they are unfortunately treat shortfall as synonymous with income gap and this is really confusing. So, income\_gap\_ratio should be shortfall\_ratio; or alternatively, avg\_shortfall should be avg\_income\_gap. But switching between these two terms to designate the difference between y and z is bad form.

To your question though, how you’ve interpreted it, matches with my sense of this.

**//Definitions**

* Average shortfall: average shortfall from the poverty line (averaged across the population in poverty). Yep, seems right.
* Total shortfall:  total amount of money that would be theoretically needed to lift the incomes or expenditure of all people in poverty up to the poverty line (w/out considering behavioral response). Assuming this is: , then this description seems right to me.
* Income gap ratio: average shortfall from the poverty line (averaged across the population in poverty) expressed as a share of the poverty line. Yep, seems right.
* Poverty gap index: mean shortfall of the total population from the poverty line counting the non-poor as having zero shortfall and expressed as a percentage of the poverty line. For those below the poverty line, the shortfall corresponds to the amount of money required in order to reach the poverty line. For those at or above the poverty line, the shortfall is counted as zero. The average shortfall is then calculated across the total population – both poor and non-poor – and then expressed as a share of the poverty line. Yep, seems right.