MEANS OF RATIOS OR RATIOS OF MEANS?

Marc Wuyts
September 2000

An example

A household consisting of 2 persons with income of $200, will have an income per capita of $100. Conversely, if you know that average income of the HH is $100 and that it consists of 2 persons, you can deduce that total income equals $200. This is simple mathematics.

The same is not true, however, when working with group averages. If a group of households have an average income of $200 and average HH size of 2, it does not follow that the average of HH per capita incomes is $100. Nor can you multiply this average of HH per capita incomes with average HH size to compute the average household income. Yet these are very common mistakes to make.

To see this, consider three households with the following incomes and HH sizes:

<table>
<thead>
<tr>
<th>HH Income</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH size</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

But, averaging income per head across HHs yields,

| HH income/head | 50  | 200 | 100  | 350/3 = 116.67 |

A different result altogether.

More formally,

The mean of the ratios of two variables is not equal to the ratio of their means!

That is,

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{Y_i} \neq \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i}; \quad i = 1, \ldots, n
\]

In fact, if the purpose is to calculate the average income per head across households, the correct way to go about this is to calculate:
• the ratio of the means (or sums) of both variables (the right hand side in the equation above),
• and, hence, not the mean of their ratios (its left hand side).

In the simple example above, this yielded $200/2 = 100$.

**An elaboration: using logarithms and geometric means**

Note, furthermore, that,

$$\log\left(\frac{X_i}{Y_i}\right) = \log(X_i) - \log(Y_i)$$

and, hence,

$$\frac{1}{n} \sum_i \log\left(\frac{X_i}{Y_i}\right) = \frac{1}{n} \sum_i \log(X_i) - \frac{1}{n} \sum_i \log(Y_i)$$

or,

$$\log\left[\prod_i \frac{X_i}{Y_i}\right]^{1/n} = \log\left[\prod_i X_i\right]^{1/n} - \log\left[\prod_i Y_i\right]^{1/n}$$

$$= \log \left[\frac{\prod_i X_i}{\prod_i Y_i}\right]^{1/n}$$

from which it follows that,

$$\left[\prod_i \frac{X_i}{Y_i}\right]^{1/n} = \left[\frac{\prod_i X_i}{\prod_i Y_i}\right]^{1/n}$$

Hence, the geometric mean of a ratio of two variables equals the ratio of their geometric means.

**Working with transformed data**

This latter result is particularly useful when both X and Y are approximately lognormally distributed such that the logarithmic transformation transforms both variables towards normality.

In this case, for each untransformed variable (X and Y), the geometric mean and the median will be approximately equal, but different from its arithmetic mean. Why this is so, is explained in Mukherjee, C., H. White and M. Wuyts (1998) *Econometrics and Data Analysis for Developing Countries*, London: Routledge, chapter 3, in particular, pp. 97-105.