Measurement Objectives, the Median and the Mean

David Rogers ♦ www.RogersTeam.co.uk

One often wants to calculate a single number as being representative of a whole collection of numbers. For example, an Applications Development (AD) department that does many projects and measures the productivity of many of them, always wants a single number as the answer to the question 'what is the department's productivity', and the obvious way to do it is to produce some kind of average of the productivities of the individual projects. [1]

Mean and Median

Everyone knows that to calculate the average of a set of values, one adds up all the values and divides the answer by the number of values. The **mean** is exactly the same thing as the average: the two words are synonyms.

The median is entirely different. To find the median, arrange the set into a list in value sequence (either ascending or descending): then the **median** value is the value half-way through the list.

The mean of a set of numbers is affected by the values of all the numbers in the set, including any 'outlier' values, which the median is not. For example in Table 1 the median is 65,000: it is unaffected by the fact that the first salary is very different from the rest. But the first salary has a considerable impact on the mean (which is 97,727).

Employee Number	Salary	
1	450,000	
2	85,000	
3	80,000	
4	75,000	
5	70,000	
6	65,000	Median
7	60,000	
8	55,000	
9	50,000	
10	45,000	
11	40,000	

Table 1

[1] In fact, because of advancing technology it is seriously doubtful whether any single 'AD departmental productivity' can still be useful for commercial purposes (see 'Proposal to the international metrics and benchmarking community for splitting the measurement of Applications Development', which was presented to the UK Software Measurement Association (UKSMA) in September 2003). However this brief explanation of the meanings of terms ignores that problem.

Weighted Average

There is a simple trap waiting for the unwary. The productivity of the Applications Development department should not be calculated as the average of the individual productivities 10, 10, 20, 21 etc. This is really an average of averages, and only in special cases is it a meaningful number.

When calculating the productivity of an AD department from a number of projects, the *weighted* average is used, because it reflects the greater importance (in terms of cost) of the big projects. To calculate the **weighted average** or **weighted mean**: add up all the function points, add up all the project staff months, and divide the one by the other.

Project Number	Function Points	Staff-Months	Productivity FP/SM	Notice that the rows are ordered by Productivity
1	3600	360	10	
2	100	10	10	
3	400	20	20	
4	210	10	21	
5	220	10	22	
6	460	20	23	Median
7	240	10	24	
8	375	15	25	
9	200	5	40	
10	120	2	60	
11	100	1	100	
Totals:	6025	463		

Table 2

Weighted average productivity is 6025/463 = 13 FP/SM. Median is 23 FP/SM.

Average of Averages

In practise, the occasions on which on would use an average of averages are rare; but as an example of when an average of averages would be useful, consider the following (rather artificial) illustration. Suppose that the set of projects above is itself only a subset of some larger set of projects, and for some reason the question arises: is the subset typically representative of the larger set? There will be many criteria for 'typically representative', such as cost, duration, technology etc. - and one important criterion is project productivity: that is, one should consider whether the subset contains a disproportionate number of high- or low-productivity projects. For the purpose of answering that very narrow question, *project scale is irrelevant:* it does not matter that project number 1 cost 360 times as much as project number 11, and therefore it would be correct to calculate the average of the individual productivities :

$$\frac{(10+10+20+21+22+23+24+25+40+60+100)}{11} = 32 \text{ FP/SM.}$$

One could then compare this with a similarly calculated average productivity of the larger set of projects.

Objectives should determine methods of measurement and calculation

When confronted by any choices in measurement, it is essential to refer back to the *business objective*.

For example, suppose for some reason one wanted to establish what level of salary would be representative of the eleven employees listed in Table 1. If it were for some accounting purpose, such as forecasting future salary bills, the mean would probably be appropriate: the outlier (i.e. untypically high) salary should obviously be taken into account and influence the result. But a lawyer in a court case involving someone's loss of earnings might well argue that the median is more appropriate, because the single high salary has nothing to do with the unfortunate person's loss.

The objectives of a measurement programme should determine what is measured, and how, and how results are treated. That may seem to state the obvious: but it is often overlooked. Usually the objective of an ADM measurement programme really is to determine the overall efficiency of the development organization: then, only the weighted average of the productivities of all the development projects will do. But even here the measurer must probe carefully to check that the real question is being answered.

For instance consider the projects listed in Table 2. Perhaps management knows that project 1 – which because of its size has a big impact on the weighted average - had all kinds of difficulties that are already thoroughly understood, were unique at the time and unrepeatable in the future. Remember the objective of the calculation: to come up with a single number which will be as representative as possible of the departmental productivity. In this case the *most useful information* (as opposed to the most arithmetically accurate number) might be obtained by omitting project 1 and taking the weighted average productivity of the other ten projects:

$$\frac{(100+400+210+220+460+240+375+200+120+100)}{(10+20+10+10+20+10+15+5+2+1)} = 24 \text{ FP/SM.}$$

It is still best – i.e. most useful - to use the weighted average for projects 2 to 11 because apart from project 1, cost probably *does* matter to the management, so project 3 *should* have a much bigger impact on the result than project 10.

David Rogers is an independent consultant specialising in the quality control and correction of Applications Development & Maintenance (ADM) benchmarks.