On The Average IV: Class Size

For my fourth data gymnastic, it's time to be practical. This is one that touches the experience of every college student: You are a college student, or a prospective college student. And one of the things that's important to you in evaluating a college is the average class size. I suppose you could always ask the college about its class size, or look it up in a published ranking of colleges, or read the number in the catalogue. But let's suppose you had to compute the number yourself — just to leave nothing to chance and to eliminate confusion. That's the way to understand it — do it yourself, at least on a hypothetical example.

O.K., let's start with another thought experiment — something too simple for reality, but clear enough for practice. Suppose my college has exactly two hundred students and one class — everybody takes it. With once class and no variation, that's easy, at least for the average: There is exactly one class, everyone has the same class, and the average class size is 200.

That's the simplest example. Now for something one step up in difficulty: My college just decided that some of the seniors need special treatment, perhaps some lab. work. Now there are two classes, one of size 190, one of size 10. What's the average class size? Obviously, the arithmetic is $(190 + 10)/2 = 100$. The average class size is 100.

Class 1	190
Class 2	10
Sum	200
Average	100

Let me push this a little further, trying to create something more interesting: It also turns out that we have two students with really special needs, and we have a faculty willing to accommodate. So, now there are four classes, one of size 190, one of size 8, and two of size 1. What's the average? Obviously, the arithmetic is (190 + 8 + 1 + 1)/4 = 50.

Class 1	190
Class 2	8
Class 3	1
Class 4	1
Sum	200
Average	50

Now, I ask you a question: Do you "believe" that answer, "50"? I'm not asking whether or not my arithmetic is correct. It is. I'm asking whether or not you would feel you had been cheated if this college had advertised: "Our average class size is 50." and you had chosen to attend? If you believe this answer, then let me give you an example that is more extreme: This time I'm going to give those ten seniors individual treatment, one on one with their professors. Again, what's the average? The arithmetic would seem to be (190 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1)/11 = 18.18, or 18 (approximately).

Class 1	190
Class 2	1
Class 3	1

Class 4	1
Class 5	1
Class 6	1
Class 7	1
Class 8	1
Class 9	1
Class 10	1
Class 11	1
Sum	200
Average	18.18

Still satisfied? If you find that number acceptable, then consider: Suppose you went to this college, with an advertised average class size of 18. And suppose that, somehow, now that you're there, the experience isn't exactly living up to your expectations. So one day at lunch you decide to check for yourself and you starting asking the other students: "My class has 190 students in it. How large is yours?" And sure enough, nearly all of the other students at the lunch table report that their class too has 190 students in it (in fact, you're all in the same class).

Suitably alarmed, but recognizing that your lunch table crowd may not represent the whole college, you decide to check further: You decide that it's time for a survey and so you send a questionnaire to every student: You want the facts. What's the size of your class?

And here's how the numbers come back:

Ten students report classes of size one. and one hundred and ninety students report classes of size 190. So, using your own data, what's the average? Well, you got two hundred responses to your questionnaire and so the arithmetic requires you to add up the numbers and divide by 200. That's

Student #1	1	10 students
Student #2	1	
Student #3	1	
...		
Student #10	1	
Student #11	190	190 students
Student #12	190	
Student #13	190	
...		
Student #200	190	
Sum	36,110	200 students
Average	180.5	

According to your survey, the average class size is 181 ! The college says the average class size is 18. But you asked the students and your data say that the average class size is 181 — ten times larger than advertised.

The Unit of Analysis (Again)

What's the problem? Don't say the problem is that, "Statistics lie," or that "The average (one of them) isn't the right number." Neither of these epithets is an intelligent response to a confusing situation — These are just different ways of re-stating the fact that there's a problem. But what *is* the problem? How can there be two radically different answers, both 18 and 181, as answers to the question "What's the average?"

Actually, both answers are correct. And both the college, that reports 18, and you, with your survey that reports 181, are using the same facts.

In this case, again, you straighten out the mess by asking yourself what is the *unit* you are interested in? What is the *unit of analysis*? If I were a prospective college student, looking out for my own best interest (defined as small classes), then I would take the student as the unit of analysis. I would look at the data that describe each unit, as you did in your survey, and then I would average them. And, just to be sure I got the thing right I would identify the units, as well as the numbers, when I wrote down the equation.

So, unit number one, that is *student* number one, reported 190 *classmates* in class. So, with respect to the student as the unit of analysis (look at the denominator), the first report is

$$190\ \frac{\text{classmates}}{\text{student}}$$

Read that as "190 classmates *per* student". And, to repeat, note the denominator and note the use of labels, "classmates" and "students"

Student number two came up with the same report, and now I have two instances of 190 classmates per student. And that would repeat for 190 of the 200 reports.

$$\left.\begin{array}{l} 190\ \dfrac{\text{classmates}}{\text{student}} \\ 190\ \dfrac{\text{classmates}}{\text{student}} \\ \text{....} \\ 190\ \dfrac{\text{classmates}}{\text{student}} \end{array}\right\}\ \text{190 reports}$$

And then for ten more students the report would be one classmate per student.

$$\left.\begin{array}{l} 1\ \dfrac{\text{classmate}}{\text{student}} \\ 1\ \dfrac{\text{classmate}}{\text{student}} \\ \text{....} \\ 1\ \dfrac{\text{classmate}}{\text{student}} \end{array}\right\}\ \text{10 reports}$$

So, altogether, with 200 reports I have 190 reports of "190 classmates per student" and 10 reports of "1 classmate per student." So adding up the 200 reports and dividing by 200: the average number of classmates, averaged over the 200 units, is 180.55 classmates *per student*.

And that *is* the correct answer — *for the student as the unit of analysis*. So where did the college's advertisement of 18 students per class come from? The answer is in the denominator: students *per class*? The college used a different unit: Trust me as a faculty member — students are not the only constituents of a college: There is an administration that sweeps the floors and there are faculty that teach the classes. And, like you, both of them ask: "How big are the classes? But they get a different answer. And there's the problem: We're not all asking the same question of these data. The English is sloppy enough to make it look the same. But you have to look very closely at the question: As a faculty member, or taking the faculty as the unit of analysis, there are exactly eleven units in these data, eleven classes.

Attaching the units to the numbers, and spelling it out in detail, I have one faculty member who reports 190 students. That is

$$190\ \frac{\text{students}}{\text{faculty}}$$

And I have ten happy faculty reporting one student per faculty member. So my data are

$$190\ \frac{\text{students}}{\text{faculty}} \qquad \text{1 report}$$

$$\left.\begin{array}{l} 1\ \dfrac{\text{student}}{\text{faculty}} \\ 1\ \dfrac{\text{student}}{\text{faculty}} \\ \cdots\cdot \\ 1\ \dfrac{\text{student}}{\text{faculty}} \end{array}\right\} \quad \text{10 reports}$$

And now, applying my arithmetic to these eleven numbers, one datum for each unit of analysis, I add up the numbers and divide by eleven, getting the answer eighteen, approximately, eighteen, not one hundred and eighty one.

And the moral is: Keep your eye on the unit of analysis that shows up, in this example, as the unit of the denominator — as something *per something*, where that last *something* is the unit of analysis. There are two answers here, referring to two different questions, and to two different units of analysis. Both answers are *numerically* correct, but that is not to say that both of the *answers* are correct: It all depends on the question you asked (or intended to ask — now that you are being more careful).

Writing it as a weighted mean, as

$$\text{mean} = \frac{\text{number} * \text{weight} + \text{number} * \text{weight} + \ldots + \text{number} * \text{weight}}{\text{sum of weights}}$$

in this example, using the class as the unit of analysis, each report had an implicit weight of 1, each class was given equal weight. But I can simplify the arithmetic by using the weight to represent the frequency of each value that was reported — using a weight that represents a number of the units of analysis. Thus, more efficiently

$$\text{mean} = \frac{190 \frac{\text{students}}{\text{class}} * 1\text{class} + 1 \frac{\text{student}}{\text{class}} * 10 \text{ classes}}{11\text{classes}}$$

giving me the 18.18 students per class.

Exercise: For practice, using weight for the average of classmates per student, …

Exercise: Consider the same problem, using medians

Exercise: It would be reasonable to assume that either one of these numbers will do: That which ever average you use, you will still get the same rank order of “best” colleges, that the best, measured in terms of classmates per student will still be the best in terms of students per class — even though the numbers will be different. Question: Is this assumption correct? When you compare two colleges, is it necessarily true that the college with the smallest number of students per class is also the college with the smallest number of classmates per student? Prove that this is not true by creating a counter example. Construct hypothetical data for two colleges such that one of the two has the smaller number of students per class while the other has the smaller number of classmates per student.

Homework:

I am enclosing three Excel files: These are enrollments during one quarter for all classes at Dartmouth. For your joy / curiosity / whatever I have included the entire data set, by class, as one file. These were the enrollments for Winter of 1991. For your homework, I have extracted the data for two departments. Practicing on something small — make up something even simpler until you’re sure of your self. A hypothetical department with only two classes. Then, and only then, work yourself up to data: Take a look at Physics and take a look at Government. In each case compute an average with the class as the unit of analysis — computing the average number of students per class — and compute an average with the student as the unit of analysis — computing the average number of classmates per student. Which department, Physics of Government, has the "best" average. If you've got the numbers right, then it will not be easy to answer that question. Explain, briefly, in English — pass the room mate test for clarity and succinctness. (And what did the college tell you about class size?) (Feel free to compute the averages for the college as a whole. You have the data, at least for one term.) (Hint: You can check your computations, or mine: For Physics I get an average of 18.84 students per class and an average of 86.29 classmates per student. Get that right, or correct me, and you are ready to tackle Government — and the whole.) Do you think that the rank order of departments by class size would change, depending upon the measure used to report "size"? Could the rank order of colleges change, depending upon the measure used to report "size"?

Class							
Department	Number	Section	Type	Time	Prof.	Limit	Enrollment
GOVT	5	1	LEC	11	Masters, Roger D	105	91
GOVT	6	1	LEC	9S	Arseneau, Robert B	105	89
GOVT	7	1	LEC	11	Becker, David G	40	45
GOVT	32	1	LEC	11	Winters, Richard F	50	42
GOVT	42	1	LEC	12	Kopstein, Jeffrey S	50	63
GOVT	48	1	LEC	2A	Lustick, Ian S	999	21
GOVT	57	1	LEC	10	Becker, David G	999	69
GOVT	62	1	LEC	11	Mather, Lynn M	60	60
GOVT	64	1	LEC	9L	Masters, Roger D	999	32
GOVT	70	1	LEC	10	Sullivan, Denis G	999	23
GOVT	80	1	LEC	ARR	Winters, Richard F	999	16
GOVT	84	1	LEC	3A	Mather, Lynn M	999	17
GOVT	86	1	LEC	2A	Sa'adah, M Anne	999	11
GOVT	99	1	LEC	3A	Becker, David G	999	10

(((Class)))							
Department	Number	Section	Type	Time	Prof.	Limit	Enrollment
PHYS	4	1	LEC	11	Thorstensen, John R	999	128
PHYS	13	1	LEC	10	Mook II, Delo E	999	113
PHYS	16	1	LEC	9L	Montgomery, David C	999	30
PHYS	24	1	LEC	11	Walsh, John E	999	12
PHYS	44	1	LEC	9L	La Belle, James W	999	14
PHYS	66	1	LEC	10	Denton, Richard E	999	10
PHYS	72	1	LEC	ARR	Lawrence III, Walter E	999	4
PHYS	82	1	LEC	ARR	Sturge, Michael	999	6
PHYS	85	1	LEC	ARR	Harris, Joseph D	999	2
PHYS	87	1	LEC	ARR	Harris, Joseph D	999	2
PHYS	102	1	LEC	10A	Yafet, Yako	999	6
PHYS	105	1	LEC	11	Hudson, Mary K	999	6
PHYS	110	1	LEC	ARR	Lotko, William	999	2
PHYS	121	1	LEC	ARR	Harris, Joseph D	999	0
PHYS	122	1	LEC	ARR	Yafet, Yako	999	0
PHYS	123	1	LEC	ARR	Harris, Joseph D	999	0
PHYS	127	1	LEC	ARR	Harris, Joseph D	999	0
PHYS	137	1	LEC	ARR	Harris, Joseph D	999	11
PHYS	157	1	LEC	ARR	Harris, Joseph D	999	12