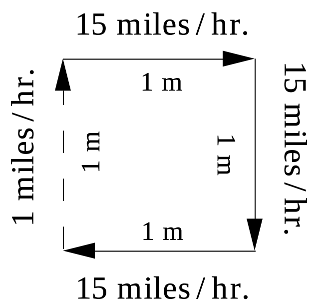
On the Average III: Racing

O.K., you have the principle: Watch the unit of analysis. You have the principle under control, but let me put your control to another test with Data Gymnastics exercise #2:

Consider a foot race on a 4 mile square track, 1 mile on an edge. I'm racing against myself, going for a personal record — and then some.



Here's my performance: I managed 15 miles per hour for the first mile (a four minute mile), I repeated at 15 miles per hour for the second mile, repeated again at 15 miles per hour for the third mile. And then I paid for the first three miles with heat stroke, plus a heart attack and sore feet — leaving me able to crawl the last mile in exactly one hour. What's was my average?

Well, it was

$(15 + 15 + 15 + 1) / 4 = 11.5$

If that looks reasonable to you, if 11.5 miles per hour seems like a reasonable estimate for the average, then explain this: I ran and crawled for a little more than an hour, for exactly one hour and twelve minutes. So how is it that I covered a distance of only four miles, which is considerably less than the 11.5 miles that I am supposed to have traveled in an average hour? That makes no sense. So, the answer, 11.5 miles per hour, must be wrong.

The way out of this anomaly is, again, to watch the unit of analysis. Does your average, 11.5, describe the typical mile: What I accomplished during the first mile averaged with what I accomplished in the second mile, and the third, and the fourth. Or does your average describe the typical hour or minute: Averaging what I did in the first four minutes with what I did in the next four minutes — with what I did in the next four minutes, and what I did in the next sixty minutes? What is the unit of analysis? Is it a mile, or is it a minute? It matters.

That is the nature of the puzzle for this bit of mental gymnastics, but let me tell you how I really figure out the solution to a problem like this: The rule for problem solving in this situation is another one of those strategies that you never admit to in a final report. The rule is work backward. Create a simple thought experiment: Figure out the answer for a hypothetical example, *then* figure out the method that works for the hypothetical example, and then go back to the data where you can solve the problem. What you do not want to do is just forge ahead, however bravely, with "2 unknowns": Trying to use an unknown method in search of an unknown answer. So you do a thought experiment. You think up a problem for which the answer is known. You work on this problem until the method becomes clear, and then you go back to the data.

So, working backward, what do I know about this problem? Actually, working backward, I can begin with the answer: I know that the total mileage was four miles and that the total time was 1 hour and 12 minutes, or 1.2 hours. So the answer is going to be 4 miles per 1.2

hours, which is 3.33 miles per hour. I have the answer. Now, what's the method? What type of average would have given me the right result?

Well, the method is not what I did above, not (15 + 15 + 15 + 1)/4 = 11.5, but let's take a look at it using the units and using the weighted mean. What went wrong?

$$\text{mean} = \frac{15\frac{\text{miles}}{\text{hour}} * 1\text{mile} \;+\; 15\frac{\text{miles}}{\text{hour}} * 1\text{ mile} \;+\; 15\frac{\text{miles}}{\text{hour}} * 1\text{mile} \;+\; 1\frac{\text{mile}}{\text{hour}} * 1\text{ mile}}{4\text{miles}}$$

If you look at that in detail, it makes no sense: the units used as the units of analysis are inconsistent (and that's what got me the wrong answer): When I wrote 15 miles *per hour*, I have a unit of time where I should expect the unit of analysis. But when I wrote 1 *mile* for the weight I said that I was organizing the analysis in miles. I can't do both (not at the same time). I have to pick one. In fact, I can organize the problem in terms of time, or I can organize it in terms of distance — but which ever I do I have to do it consistently.

Lets try it in terms of hours: I started out at 15 miles/hour. Hours are in the denominator so I'll give this speed a weight in hours. Thinking it through, I kept up this speed for four minutes in order to complete the first file. Four minutes is one-fifteenth of an hour — there's the weight. I did it again, same speed, for the next one-fifteenth of an hour, completing the second mile. I did it again, same speed, for the next one-fifteenth of an hour, completing the third mile. And then I completed my circuit at one mile per hour, continuing at that rate for an entire hour until I was finished. Using the hour as the unit, here is a consistent weighted average:

$$\text{mean} = \frac{15\frac{\text{miles}}{\text{hour}} * \frac{1}{15}\text{ hour} + 15\frac{\text{miles}}{\text{hour}} * \frac{1}{15}\text{ hour} + 15\frac{\text{miles}}{\text{hour}} * \frac{1}{15}\text{ hour} + 1\frac{\text{mile}}{\text{hour}} * 1\text{ hour}}{1\frac{3}{15}\text{ hours}}$$

which simplifies to

$$\text{mean} = \frac{1\text{ mile} + 1\text{mile} + 1\text{mile} + 1\text{ mile}}{1.2\text{ hours}}$$

and to

$$\text{mean} = \frac{4\text{ miles}}{1.2\text{hours}}$$

which gives

$$\text{mean} = 3.33\frac{\text{miles}}{\text{hour}}$$

which is correct.

There's my reward for thinking clearly. Now I've got the formula for the right answer. I've thought it through. I've been very careful to keep the unit of analysis simple and in the right place, and I got the right answer. The moral: Watch the unit of analysis. Check, with a simple problem — just to be sure. And understand that averages, in the real world, require patience and careful thinking, — even for "easy" problems.

Exercise: I picked one unit, the hour. But suppose I had picked the mile as the unit, re-expressing my first mile as 4 minutes *per mile* (and my fourth mile as 60 minutes *per mile*)? Now I am organizing the analysis in miles and I need a weight in miles. Will it work? (Partial answer: It can't give me exactly the same answer because the first answer was in miles per hour and the second answer will be in hours per mile (or minutes per mile). Keeping that in mind, keeping in mind that the correct answer in hours per mile is exactly the inverse of the correct answer in miles per hour — the answer will be *almost* the same and it will be correct. Do it and explain.)

Homework. Most of the hard work of data analysis is undertaken flat on your back, staring at the ceiling, with your eyes closed — at least for me. Whatever your equivalent posture, assume it: You won't need much computing for this one, except to write down your answers:

#1: I've seen versions of this in magazines for people who get a little nuts about precise measurements of aerobic performance and measures of improvement (or change) from day to day:

> Question: Dear Dr. C.: I understand the rules for measuring my aerobic capacity. But living in hill country, far from a track, how do I measure my performance when my route takes me up hills and down?
>
> Answer: Run a circular route. For every mile you have to run uphill you will be treated to a mile downhill and so the effort and time, on a flat route, are equivalent to average effort and average time on a circular route — no adjustments necessary.

Problem: Obviously Dr. C. lives in flat country, otherwise his heart would long ago have punished his brain for a nice simple logical idea — that is wrong. Explain to the poor man why his averaging doesn't work, lest his mind kill his body on their first trip to the mountains. Create a simple numerical example, using a weighted average to illustrate the problem.