Weighted Averages

Weighted averages come in two forms:

- Raw weighted averages
- Average of averages

Raw Weighted Averages – Frequency Distributions

We will begin by considering raw weighted averages, but first we must review the definition of an average. By definition, an average is always the ratio of two numbers. It is the sum of data points x_i divided by the number or count of data points, n. It is written as

$$\bar{x} = \frac{x_1 + x_2 +_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

When working with arbitrary numbers, we don't think about the units of the answer, but working with real data, there are always units, and the average is expressed as "the average of something per something." Let's consider the average scores of ten students on a test.

{100, 80, 80, 90, 80, 80, 90, 90, 90, 80}

The average of these scores is obviously the sum of the scores (860) divided by 10, or 86. We say the average score is 86, rarely mentioning the units, but more precisely, we should say, "the average score is 86 points per student." The units are "points per student." There is always a "per" to indicate what we are averaging over.

We will now turn this problem into a weighted average problem by grouping the scores into frequencies as illustrated below. This form of the data is called a frequency distribution.

Score	Students Frequency
100	1
90	4
80	5
Total	10

It is the same data but expressed in a slightly different form. Clearly, we can no longer just average the three scores and get (100+90+80)/3 = 90. We must weight the scores by the number of students who received each score. We must weight the scores by the item following the "per" in "points per student."

Weighted averages arise when multiple quantities are to be averaged, but the frequency, count, time, or activity associated with each quantity is unequal. To arrive at a correct average, each quantity must be multiplied by its corresponding weight. The weighted average of the scores can be written as

$$\frac{100(1)+90(4)+80(5)}{1+4+5}=86$$

Notice that if the frequency associated with each score were equal, then it would be unnecessary to weight the scores, because they would all be weighted equally. In general, the weighted average can be written as

$$\bar{x}=\frac{w_1x_1+w_2x_2+w_3x_3+\cdots+w_kx_k}{w_1+w_2+w_3+\cdots+w_k}=\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i}$$

This looks unwieldy, but hopefully, after a little study, it will start to make sense. The biggest potential for confusion is the use of the same letters and subscripts to represent slightly different things in both the average and weighted average formulas. In the weighted average formula, k is the number of unique values of the x_i's rather than the individual x_i's. Furthermore, in the weighted average formula, x_i itself is not all the individual x_i's but rather just the unique x_i's. In the example above, n=10 for the ten total scores, and k=3 for the three unique scores.

Let's consider another example of a raw weighted average that we are all familiar with: grade point average or gpa. If a person receives 40 credits at grade of C, 50 credits at a B, and 10 credits at an A. What is their gpa? This is a straight forward variation of the frequency distribution discussed above. Credits are really frequencies.

Grade	Credits Frequency
A = 4	10
B = 3	50
C = 2	40
Total	100

The letter grades must be weighted by the number of credits received at that grade.

$$\frac{40\times 3+50\times 3+10\times 4}{40+50+10}=\frac{310}{100}=3.1$$

In terms of units, gpa is grade per credit. That is, the person received an average grade of 3.1 for each credit hour. By stating the units in this manner, we see that we are once again weighting by what follows the "per" in the units. Rarely are the units actually stated, but they are always implied.

Another, not so obvious, form of weighted average occurs in averages of averages as discussed below.

Average of Averages

The straight forward average of averages occurs whenever multiple averages were computed separately and now need to be combined into a single average. For example, a student may have recorded separate gpa's for each year of college. In order to combine them into a composite gpa, they must be weighted by the number of credits taken each year. Consider the following data.

Year gpa	**Credits Frequency**
2.4	26
3.1	29
3.3	32
3.7	33

The overall gpa must be computed as a weighted average. It would be incorrect to just average the individual averages. The correct computation is

$$\frac{26(2.4)+29(31.)+32(3.3)+33(3.7)}{26+29+32+33}=3.16$$

Average of averages also arise whenever two ratios need to be averaged. Common examples include:

- Miles per hour (mph)
- Miles per gallon (mpg)
- Dollars per hour
- Dollars per pound
- Defects per one-hundred parts (percents)
- Units per hour

The thing that all these situations have in common is that, before they can be averaged, they must be **weighted by the units in the denominator** of the ratios. That is, they must be weighted by the units following the "per." Another thing they have in common is that they can be reconstituted as a new ratio of two totals. Of course, this is seen best with examples.

Speed - Miles per hour (mph)

In most cases, we don't even think of these ratios as averages, but they are in fact averages. Miles per hour (mph) is the average number of miles driven per hour. We are averaging "per hour." The average is computed by totaling the number of miles driven and dividing by the total number of hours required to drive those miles. When we are given two separate mph values to "average" we cannot simply add them and divide by two. We must weight them by what is in the denominator, in this case "hours."

Consider a 400 mile trip in which a car traveled half the distance at 40 miles per hour and the other half at 50 miles per hour. At first reading, since the distances traveled were equal, you might think that you could just add the two speeds and divide by 2. This answer would be quit close to the correct answer, but it would be incorrect. The quantities have to be **weighted by the quantity in the denominator, not the quantity in the numerator**. In this case, that is hours, not distance. Calculating the hours driven in each leg of the trip gives

$$\frac{200miles}{40mph}=5hours \text{ and } \frac{200miles}{50mph}=4hours$$

At this point, we could compute the correct weighted average as follows

$$\frac{40mph \times 5hours + 50mph \times 4hours}{5hours + 4hours} = \frac{400miles}{9hours}$$

Of course, once we have the total hours, we could do the calculation more directly as

$$\frac{total\_miles}{total\_hours} = \frac{400miles}{9hours} = 44.44mph$$

This problem could be even more misleading if the speeds were driven for a different number of miles. We would be drawn naturally to weighting them by the unequal number of miles driven. Consider, for example, if the car were drive 400 miles at 40 mph and 500 miles at 50 mph. Weighting by miles would give

$$\frac{400miles \times 40mph + 500miles \times 50mph}{900miles} = 45.55mph \quad \textbf{WRONG}$$

In this case the correct answer is (40+50)/2 since each leg of the trip was driven for 10 hours.

$$\frac{total\_miles\_driven}{total\_hours\_driven} = \frac{900miles}{20hours} = 45.0mph$$

In the examples, the numeric differences are not great, but none the less, one approach is wrong and the other is correct. One must always be on guard for it is easy to be misled.

Mileage - Miles per gallon (mpg)

This situation leads to the same trap as miles per hour. Assume a car drives 100 miles and gets 20 mpg, the drives 300 miles and gets 30 mpg. One's first reaction is to weight the two mileage numbers by the miles driven. This, however, would be incorrect. Weighting by miles gives

$$\frac{100 \times 20 + 300 \times 30}{400} = \frac{110}{4} = 27.5 \quad \textbf{WRONG}$$

which is incorrect. Just as in miles per hour, we need to look at total miles divided total gallons used. In effect, we need to weight by gallons, not miles.

$$\frac{total\_miles}{total\_gal} = \frac{400}{\frac{100}{20} + \frac{300}{30}} = \frac{400}{5 + 10} = 26.67$$

The proper formulation of the problem would become obvious it were stated as "a car drove 100 miles and used 5 gallons, and then drove 300 miles and used 10 gallons." We would automatically compute the mpg as

$$\frac{100+300}{5+10}=\frac{400}{15}=26.67$$

In the form of a weighted average, it would look like

$$\frac{20mpg\times 5gal+30mpg\times 10gal}{5gal+10gal}=\frac{400}{15}$$

As in mph, we needed to weight by the units in the denominator rather than the numerator.

Wages – Dollars per hour (dph)

By this point the pattern should be clear. Consider a person who earned $400 working for $5 per hour and $1000 working for $20 per hour. What was her average wage in dph? Now we are smart enough to now that we want to weight by hours not dollars. The person worked 80 hours for $5 dph and 50 hours for $20 dph. This can be computed straight forwardly as

$$\frac{total\_dollars}{total\_hours}=\frac{400+1000}{80+50}=\frac{1400}{130}$$

As a weighted average, this becomes

$$\frac{5dph\times 80hours+20dph\times 500hours}{80hours+50hours}=\frac{1400}{130}dph$$

Mixture - Dollars per pound (dpp)

At the risk of beating this to death, we will do one more example of "dollars per something," namely, "dollars per pound." How many pounds of cashew nuts worth $2 per pound must be mixed with pecans worth $3 per pound to create a mixture of 40 pounds of nuts worth $2.75 a pound? In terms of totals we have

$$\frac{total\_dollars}{total\_pounds}=dpp$$

This situation is made trickier by the fact that we are given the final average and the total pounds, from which we must find the weights. All of these examples can be cast in this manner, so it is good to see a problem from a slightly different angle. Let x equal the number of pounds of cashews at $2 and 40-x equal the number of pounds at $3. The resulting weighted average equation looks like

$$\frac{2x + 3(40 - x)}{40} = 2.75$$

Solving for x gives 119/5 = 23.8 pounds of cashews and 40 – 23.8 = 16.2 pounds of pecans.

Percents - Defects per 100 parts

Working with percents is more subtle than the previous examples. Consider two production runs of 120 parts and 80 parts. In the first run, 5% were defective, and in the second run, 10% were defective. What was the average percent of defective parts? The units are "defective parts per 100 parts." In terms of total ratios this can be written as

$$\frac{total\_defective\_parts}{total\_parts/100} = \frac{total\_defective\_parts}{total\_parts} \times 100 = \text{ \% defective}$$

Therefore, we have to weight by total number of parts, which results in the following weighted average ratio

$$\frac{120(.05) + 80(.10)}{120 + 80} = \frac{6 + 8}{200} = \frac{14}{200} = \frac{7}{100} = 7\%$$

Work – Units per hour

Most rate problems can be viewed as weighted average problems. Consider two machines X and Y. X produced 1000 bolts at the rate of 200 bolts per hour, and Y produced 1000 bolts at the rate of 250 bolts per hour. What was their average production rate in bolts per hour? Machine X worked for 4 hours, and machine Y worked for 5 hours, so, the average production rate was

$$\frac{total\_bolts}{total\_time} = \frac{2000}{4 + 5} = 222.22bph$$

$$\frac{250bolts \times 4bph + 200bolts \times 5bph}{4hours + 5hours}$$

Notice that, this is not (200+250)/2 = 225. However, if we were told that the two machines both worked for one hour (equal hours) at their respective rates, then the combined rate would in fact be 450 bph, or 225 bph on average. The wording is critical to properly interpret the problem.

Expected Value of X

Let's revisit the formula for weighted average

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_kx_k}{w_1 + w_2 + w_3 + \cdots + w_k} = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}$$

We can rewrite this formula in the form

$$\bar{x} = \frac{w_1}{W}x_1 + \frac{w_2}{W}x_2 + \frac{w_3}{W}x_3 + \cdots + \frac{w_k}{W}x_k = \frac{1}{W}\sum_{i=1}^{k} w_i x_i$$

where W is the total weights $\sum_{i=1}^{k} w_i$. The individual ratios $\frac{w_i}{W}$ are the decimal fractions or percents of total weights associated with each individual x. Because of the way in which these ratios are computed, they must sum to one and can be interpreted as percents or probabilities. If the ratios are written as p_i, then the equation for the average value of x can be written as

$$\bar{x} = \sum_{i=1}^{k} p_i x_i$$

When written in this form, the average value of x is called the expected value, and the numbers p_i are interpreted as the probabilities of x.

A simple, but useful, application of this formula is estimating the grade needed in the second half of a semester to achieve a certain final grade. The final grade GF is a weighted average of the current grade GC plus the remaining grade GR

$$GC \times p + GR \times (1 - p) = GF$$

where p is the percent of grade completed at some intermediate point during the semester. For example, a student who has a grade of 68 at the midterm (p=.5) wants to determine what grade is needed to reach a final grade of 80. Solving the equation

$$68(.5) + GR(.5) = 80$$

gives GR = 92.