Measurements of Central Tendency

The major measurements of central tendency are mean, median, mode. Other measures such as standard deviation, variance, and range are addressed elsewhere in this *Student Laboratory Handbook*.

Mean: The mean is the most commonly used measure of central tendency. There are different types of mean, viz. arithmetic (pronounced *air-ith-mét-tic*) mean, weighted mean, geometric mean, and harmonic mean. Each has its own significance and use in various areas of physics and teaching.

The **arithmetic mean** is taken as an alternative for the commonly-used word average. This type of mean is calculated by dividing the sum of the values in a set by their number. That is,

$$\bar{X} = \frac{1}{n}\sum_{1}^{n} x_n$$

For instance, what is the mean value of the following test scores? $56, 35, 96, 48, 79,$ and 95? The sum of the individual terms is 409 and the number of scores equals 6. Thus, the arithmetic mean is $409/6 = 68.2$ approximately.

Repeated samples drawn from the same population tend to have similar means. The mean is therefore the measure of central tendency that best resists the fluctuation between different samples. Its use, however, does have some drawbacks. The main disadvantage of mean is that it is sensitive to extreme values or outliers, especially when the sample size is small. Therefore, it is not an appropriate measure of central tendency for skewed (lopsided, not normally distributed) distribution.

Consider the following test scores: 5%, 8%, 7%, and 95%. The mean score is 43.75%. This score does not accurately reflect the general performance of the student.

Even more importantly, consider how a score of 0% (missed assignment) figures into the determination of grade when the other scores are 70%, 85%, and 77%. The first three scores suggest a 77.3% performance on a grade scale of 70% to 80% equals a letter grade of C. But when the missed assignment is figured in, the average is only 58% which is a letter grade of F on a typical 90-80-70-60 grade scale. <u>The lesson not to be missed here is that missed assignments can have a highly deleterious effects when it comes to a course grade! All students in a class should be made aware of this fact.</u>

The **weighted mean** takes into account the relative importance of a number in a group of numbers. Consider test and quiz scores in a grade book recorded as percentages. If a test is worth 3 times as much as each of 3 quizzes, then a weighted mean is calculated to properly reflect students' performance in a course. An arithmetic mean would not express the importance of the test score among along with three quiz scores:

Quiz 1	Quiz 2	Quiz 3	Test
58.6%	84.5%	89.5%	55.9%

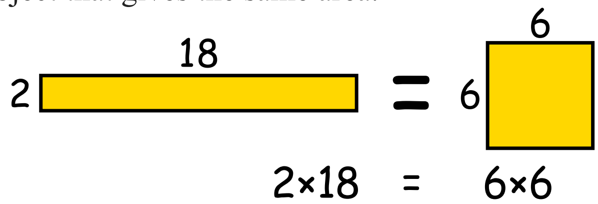
The arithmetic mean of the above scores is 72.1%. The weighted mean is considerably different:

$$1 * Quiz\ 1\ +\ 1 * Quiz\ 2\ +\ 1 * Quiz\ 3\ +\ 3 * Test\ =\ 400.3\%/6\ =\ 66.7\%$$

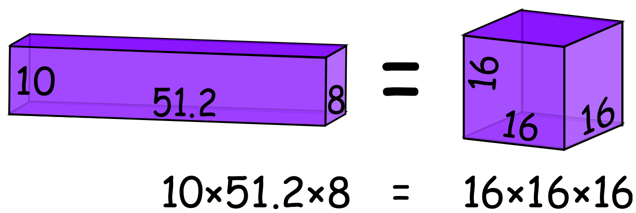
Because the test is weighted so much more in the determination of the course grade, it's value should be represented accurately in the calculation of a central value. The difference between 72.1% and 66.7% is substantial – it could make the difference in a course letter grade in this case.

The **geometric mean** is a way of taking widely varying dimensions into account. It is defined for **n** numbers as follows: multiply them all numbers together and then take the n^{th} root. That is, the geometric mean of n numbers a_1 to a_n is $\sqrt[n]{a_1 * a_2 * \dots * a_n}$.

For instance, consider the geometric mean of 2 and 18. Geometric mean = $\sqrt[2]{2 * 18} = \sqrt[2]{36}$ = 6. Six is the mean of a two-dimensional object that gives the same area.



Consider the geometric mean of 8, 10, and 51.2. Geometric mean = $\sqrt[3]{10 * 51.2 * 8} = \sqrt[3]{4096} = 16$. Sixteen is the mean of a three-dimensional object that gives the same volume.



The **harmonic mean** is another way to apply weighting, but it is substantially different from the weighted mean because it uses a "shortcut" to eliminate one unknown.

The arithmetic mean does not provide the correct information in situations that require weighting, for instance in averages involving <u>constant</u> rates. For example, if a truck goes on a round trip following the same route there and back totaling 100 miles with a constant speed of 50 miles per hour on the outbound leg and 25 miles per hour on the return leg, the average speed is NOT 37.5 miles per hour [(50*mph*+25*mph*)/2]. This is so because the TIMES of travel will be different for each leg of the trip. On the outbound portion of the trip, the time required to go 50 miles is only one hour. On the return portion of the trip, the travel time required to go 50 miles is two hours due to traveling at half the original speed. The "true average speed" is one in which the total travel time is the same as if the truck had traveled the whole distance at that average speed. <u>A simple arithmetic mean of the two speeds does not give the true average speed because it does not take into account the times spent at those individual speeds</u>. The 25*mph* portion of the trip should be double weighted due to the two hours in comparison to the 50*mph* portion of the trip at only one hour.

When weighted on the basis of time spent, then one can determine the true average speed as follows:

$$\text{rate}_{\text{average}}*(\text{total time}) = \text{rate}_1*\text{time}_1 + \text{rate}_2*\text{time}_2$$
$$v_{\text{averge}}*3hr = 25mph*2hr + 50mph*1hr$$
$$v_{\text{average}} = (50mi + 50mi)/3hr$$
$$v_{\text{average}} = 100mi/3hr$$
$$v_{\text{average}} = 33^1/_3mph$$

In the example above the total distance is 100*mi.* The total time is 3*hr.* Thus, the true average speed will be 100*mi*/3*hr* or $33^1/_3$*mph.* This is significantly different from the *arithmetic* mean of 37.5*mph.*

Believe it or not, there is yet another way to get the true average that utilizes weighted means. It's called the technique of the *harmonic* mean. In situations involving rates, the harmonic mean is another way to determine the correct, time-weighted average rate. The harmonic mean is defined as follows:

$$H = \frac{n}{\left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \ldots\right)}$$

where r_n represents the rates of the various components, and n the number of components. In a round trip, rate equals the speed and $n = 2$. For instance, if a vehicle travels a certain distance at a constant speed of 6*m/s* and then the same distance again at a constant speed of 4*m/s,* then its average speed is the *harmonic* mean that has a value of 4.8*m/s*. The calculation is as follows:

$$H = \frac{n}{\left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \ldots\right)} = \frac{2}{\left(\frac{1}{6m/s} + \frac{1}{4m/s}\right)} = 4.8m/s$$

This same approach applies to more than two segments given a series of sub-trips at different constant speeds if each sub-trip covers the same *distance.* Then the true average speed is the *harmonic* mean of all the sub-trip speeds.

Now, let's look at another example. An airplane flies three legs of an equilateral triangle (all legs the same length). If the plane flies a constant ground speed of 150*mph* on the first leg, a constant ground speed of 200*mph* on the second leg, and constant ground speed of 250*mph* on the third leg, what is the average ground speed over the course of the flight? (Note that you are not given any distances or times.)

$$H = \frac{3}{\left(\frac{1}{150mph} + \frac{1}{200mph} + \frac{1}{250mph}\right)} = 191.5mph$$

This is significantly different from an arithmetic mean of 200*mph.* The process also helps you solve a problem that otherwise is not soluble without a knowledge of individual distances and times for each leg of the trip.

If and only if each sub-trip takes the same amount of *time*, then the average speed will be the *arithmetic* mean of all the sub-trip speeds. This is rarely the case in physics.

Median: The median of a set of data is a value or quantity lying at the midpoint of a frequency distribution of observed values or quantities, such that there is an equal probability of falling above or below it. The median of 4, 6, 12, 13, 15 is 12 because it is the middle number in a rank ordered set of odd data points. If even number of data points such as 4, 6, 12, 13, 15, and 18, the median is ½ way between the two numbers either side of the middle position which in this case is 12.5. In both cases, half the numbers are less than the median and half of the numbers is more than the median. Consider the difference between average and median income of the following values: $10,000, $15,000, $25,000, and $220,000. The mean income is $67,500. The median income is $20,000. Which do you think more accurately represents what the typical income is among this group?

Mode: The mode of a set of data is the value that occurs most frequently in that set. That is for the set of numbers 2, 2, 3, 5, 7, and 8 the mode is 2.