Arithmetic Mean, Harmonic Mean and Geometric Mean

Henri P. Gavin, Civil & Environmental Engineering, Duke University

January 26, 2005

Means

The way one determines a mean value of several quantities depends on the application or question one is trying to answer. The appropriate mean answers the question, "If all the quantities had the same value, what would that value have to be in order to achieve the same result?"

Arithmetic Mean

The *arithmetic mean* is the sum of the values divided by the total number of values. The arithmetic mean is relevant any time several quantities add together to produce a total.

Geometric Mean

The *geometric mean* of n values is the n-th root of the product of the values, $\sqrt[n]{\Pi x_i}$ In a formula: the geometric mean of $x_1, x_2, ..., x_n$ is $(x_1 \cdot x_2 \cdots x_n)^{1/n}$, which is $\sqrt[n]{x_1 \cdot x_2 \cdots x_n}$.

Any time you have a number of factors multiplying together to form a product, and you want to find the "average" factor, the answer is the geometric mean. The geometric mean of a data set is always less than or equal to the set's arithmetic mean (the two means are equal if and only if all members of the data set are equal). A geometric mean, unlike an arithmetic mean, tends to dampen the effect of extremely high or low values, which might bias the mean if a straight average (arithmetic mean) were calculated. (Show this using numerical examples.) In calculating the geometric mean, the numbers must all be positive.

Determining the average growth rate over a period of many years is a widely-used application of the geometric mean.

Suppose that an investment has these five annual returns: 10%, -20%, 0%, -10%, and 20%. What is the average rate of return? The arithmetic mean of these five returns is exactly 0%. However, the average return is not the arithmetic mean, because what these numbers represent is that on the first year your investment was multiplied (not added to) by 1.10, on the second year it was multiplied by 0.80, the third year it was multiplied by 1.00, and so on. The relevant quantity is the geometric mean of these five numbers. The question about finding the average rate of return can be rephrased as: "By what constant factor would your investment need to be multiplied by each year in order to achieve the same effect as multiplying by 1.10 one year, 0.80 the next, and so on?" The answer is the geometric mean.

The value of the investment is initially V. After the first year it is $(1.10)V$, after the second year it is $(0.80)(1.10)V$, after the third year it is $(1.00)(0.80)(1.10)V$, after the fourth year it is $(0.90)(1.00)(0.80)(1.10)V$, and after the fifth year it is $(1.20)(0.90)(1.00)(0.80)(0.90)V = (0.7776)V$ If the investment had had the same annual returns, r, for each of the five years, the value after five years would be $(r)(r)(r)(r)(r)V = r^5 V$. Therefore, the average rate of return for this example can be found from $r^5 V = 0.7776V$, or $r = \sqrt[5]{0.7776} = 0.951$, that is, on average over the five year period, the investment lost about 5%. Note that the order does not matter as long as the investment has no contributions or withdrawals during the five years.

If one considers the dollar amount by which the investment grew, then the arithmetic mean would be the correct mean to use. Suppose the investment grew by $100 in the first year, lost $220 in the second year, had no loss or gain in the third year, lost $88 in the fourth year, and gained $158.40 in the fifth year, (corresponding to an initial value of $1000), It would be perfectly correct to say that it has grown *by* an average of (100-220+0-88+158.4)/5 = -$29.60 yearly, over the five year period.

Another "geometric" example is the "average" dimension of a box that would have the same volume as length × width × height. For example, suppose a box has dimensions 50 × 80 × 100 cm, the "average" dimension of the box is

$\sqrt[3]{50 \times 80 \times 100} = 73.7$ cm

Any rectangular box with an average dimension of 73.7 cm would enclose the same 400,000 cm^3 or 0.4 m^3 volume.

Harmonic Mean

The *harmonic mean* of n values is the inverse of the sum of the inverses of the n values. In a formula, the harmonic mean of n values, $x_1, x_2, \ldots, x_n$ is $n/(1/x_1 + 1/x_2 + \cdots + 1/x_n)$, or $n/\sum 1/x_i$. Typically, the harmonic mean is appropriate for situations when the average of rates is desired. The harmonic mean answers the question "If the rate was constant over all distance intervals, what would the rate be for the total distance to be covered in the same amount of time."

A common example is averaging a set of speeds tabulated for a set of uniform distances.

For example, suppose that your trip has four 100 km segments. You travel: 100 km/hr for the first 100 km, 110 km/hr for the second 100 km, 90 km/hr for the third 100 km, and 120 km/hr for the last 100 km. What is your average speed? In total, you travel 400 km. The time it took to travel 400 km is: 100km / 100km/hr + 100km / 110km/hr + 100km / 90km/hr + 100km / 120km/hr, or 1.000 hr + 0.909 hr + 1.111 hr + 0.833 hr = 3.854 hr. Therefore the average speed is 400 km / 3.854 hr = 103.8 km/hr.

For speeds tabulated over uniform distances, (100 km in this example), the average speed may be computed directly from the harmonic mean equation:
4/(1/100+1/110+1/90+1/120)=103.8 km/hr.

As another example, suppose you travel for four hours. In the first hour you drive 100 km/hr, in the second hour you drive 110 km/hr, in the third hour you drive 90 km/hr, and in the fourth hour you drive 120 km/hr. What is your average speed for the trip? In total, you travelled for 4 hours. The total distance you travelled in the four hours is 100+110+90+120 = 420 km. So your average speed for the trip would be 420km/4hr = 105.0 km/hr. Note that in this case, we use the arithmetic mean.

Exercizes

1. Consider two random normally distributed variables, A and B, with means $\mu_A = \mu_B = 10$ cm and variances $\sigma_A = 2cm$ and $\sigma_B = 5cm$. Generate 100 random samples of 30 elements each of the random variables A and B. Compute the arithmetic mean, the harmonic mean, and the geometric mean of each sample. Compute the mean of the sample means and the variance of the sample means. How do the variance of the sample means of A and B compare? You may use the built-in Matlab function `randn.m` for generating the samples.

2. A bicyle speedometer computes speed by measuring the time required for one wheel revolution and uses the "known" circumference of the wheel. Given the times for n wheel revolutions ($n \gg 1$), $t_1, \ldots, t_n$, and the circumference of the wheel, c, what is the right way to compute the average speed for the n cycles? How would the values of the arithmetic mean, geometric mean, and hyperbolic mean over-estimate, under-estimate, or correctly-estimate the actual average speed?

References on-line

http://maxvalue.com/tip104.htm
http://en.wikipedia.org/wiki/Geometric_mean
http://en.wikipedia.org/wiki/Harmonic_mean
http://www.math.toronto.edu/mathnet/questionCorner/geomean.html
http://mathforum.org/library/drmath/view/52804.html
http://www.buzzardsbay.org/geomean.htm