On means, averages, their calculation and interdependence

Several summary statistics exist for the description of samples from a distribution. Most commonly known is the “average” or arithmetic mean. Less well known are the geometric mean and the harmonic mean.

History

Pythagoras described the different means (Cantrell 2003), but they were obviously well known before by the Babylonians.

Calculation

Pythagoras showed a geometric representation (Fig 1.), and the algebraic representation for two numbers \(a, b\) is:

\[
A = \frac{a + b}{2} \\
G = ab \\
H = \frac{1}{\frac{1}{a} + \frac{1}{b}}
\]

The relationship among the 3 means is always \(H \leq G \leq A\) which can be seen rather clearly with geometric representation. With more than two values we get a set of values \((a_1, a_2, \ldots, a_j, \ldots, a_n)\):

![Figure 1: Phytagorean means: with two values \(a\) and \(b\) one can construct the arithmetic mean (A), the geometric mean, and the harmonic mean (H) geometrically.](image-url)
Additional material

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

$$G = \prod_{i} a_i$$

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i}}$$

The 3 means are closely related and Havil (2003) showed that

$$H = \frac{G^2}{A}$$

Examples

Calculate the means from a data set that has 6 numbers drawn from a Normal distribution with mean 100 and standard deviation 10. The expectation of this distribution is the same as the arithmetic mean. data: 103, 86, 98, 111, 86, 111 → $H = 98.0606, G = 98.6168, A = 99.1667$

With many data points the difference between the means gets smaller and smaller, because the distribution is symmetric.

Data drawn from a Gamma distribution with parameters $\alpha = 2, \beta = 50$ would look like this: data: 78, 19, 255, 150, 96, 80 → $H = 60.6324, G = 87.0576, A = 113$

The Gamma distribution with the parameters specified above has expected value of 100. But with 6 sample points we can not expect lots of accuracy. The Gamma distribution is skewed and has a heavy right tail, whereas the Normal distribution is symmetric.

Data drawn from a distribution with a very heavy right tail will make the difference between the mean more and more obvious. Six points from a Gamma distribution with $\alpha = 0.1$ and $\beta = 1000$ (expected value is 100) [824, 2, 5, 411, 163, 17] give these values: $H=7.80635, G=45.9269, A=237$.

Reference
