1-2017

An Average Lesson

Charles M. Biles
Humboldt State University, nia@suddenlink.net

An Average Lesson

An important aspect of education is to learn how to do **research**. A research project investigates a general field of interest along with its components. Research includes the process of asking questions of interest and obtaining answers to some of those questions. In today's world the public often focuses simply on answers. However, much work must be done before one can get useful answers. Research encompasses the entire process.

1. Background

In doing research one often encounters classic problems. For example, how can one say something informative about a group when the individuals in the group are different? This problem is known as the **Problem of Variation** or the **Problem of Diversity**. *Problem* here is used in a positive sense meaning a situation or reality, not in a negative sense meaning something disgusting or to be avoided.

Along with classic research problems come classic solutions; for example, **denial**. One way to face the problem of variation is to ask yourself, "What variation—how would the members look if they all looked the same?" This is precisely the thinking behind the statistical concept of **average** and embodied in the English word **typical**. By appealing to a typical member you can begin to talk about the group in an organic living sense. Of course denial is unrealistic if you stop there. But it can be an informative and useful beginning. To solve a problem (negative sense), ask yourself what the world would be like without the problem. For example, imagine a world without war, hate, disease, or poverty.

Imagine now a research project where you have a group of interest and also a question of interest that can be posed to each member of the group. Statisticians call such questions **random variables** because answers vary arbitrarily among members. For example, my group of interest could be all the students enrolled spring semester 2017 at Humboldt State University. Questions of interest (random variables) could include: What is your name? What level student are you (freshman, sophomore, junior, senior, graduate)? How old are you? What is your serum tetrahydrocannabinol level?

Some questions of interest must be answered with a number obtained by counting or measuring something. Statisticians call such questions **quantitative random variables**. Statisticians calculate a mean in order to obtain an average for a list of numbers obtained from such variables. A **mean** is an average calculated in a way that preserves some rationale for those numbers. In this way, rather than being boggled by a bewildering array of numbers, one bypasses the variation and simply refers to this typical value. For example, HSU's website says that the average cost for a resident student at HSU is $23,954.[1] Although costs vary among students, this reference is helpful when planning finances.

2. Means

We now formalize the computational aspects of a mean. Recall that a mean of a set of numbers is an average calculated in a way that preserves some rationale for those numbers. Since numbers may be used for a variety of different purposes, accordingly there are a variety of different means. The three most common means are the arithmetic mean, the geometric mean, and the harmonic mean. We begin with the simplest kind of group, a group with only two members. So our current quest is: given two positive numbers a and b, how and why do we calculate the various means?

---

[1] https://www.humboldt.edu/cost. Accessed January 2017.

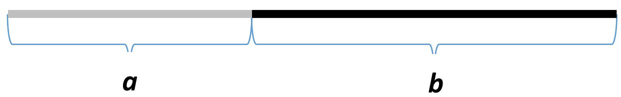
Arithmetic Mean

The **arithmetic mean** is the ordinary average of two numbers usually taught in the fifth grade: add the two numbers and divide by 2. For example, the arithmetic mean of 4 and 16 is (4 + 16)/2 = 10. We denote the arithmetic mean (AM) of 4 and 16 by AM(4,16) = 10. In general, suppose *a* and *b* are two different positive numbers. Then, the arithmetic mean of *a* and *b* is given by

$$\text{AM}(a,b) = \frac{a+b}{2}$$

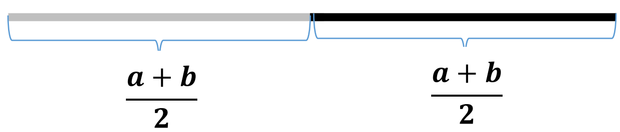
The arithmetic mean is an additive mean. Think of *a* and *b* as the length of two boards that we join linearly to make a longer board. The combined length is *a* + *b*. The thing we want to preserve is the combined length of the boards. Now consider making a final board with two components each having the same length. How long should each component be so that the result, a final board of length *a* + *b*, is preserved? For this rationale, each component needs to have length (*a* + *b*)/2; i.e., the arithmetic mean. Voila! Same result, but this time without the variation in the components.

Hence, the general rationale for the arithmetic mean is as follows. Suppose I have two different positive numbers and add them together to get their sum (notice the variation in the numbers you are adding together).



Total length *a* + *b*.

Now, remove the variation. What would you use if both numbers were replaced by the same quantity but still gave the same sum? Answer: replace *a* and *b* by their arithmetic mean.



Total length *a* + *b*.

Voila! Same result without the variation in the components.

Geometric Mean

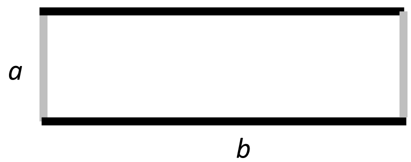
The **geometric mean** (GM) views the numbers *a* and *b* multiplicatively instead of additively. The geometric mean of two numbers is obtained by multiplying them together and then taking the square root of the resulting product. For example, GM(4,16) = $\sqrt{4 \times 16}$ = $\sqrt{64}$ = 8. In general, the geometric mean of two positive numbers *a* and *b* is given by

$$\text{GM}(a,b) = \sqrt{a \times b}$$

The obvious question is, who would want to do such a thing? The thinking behind the geometric mean is similar to the thinking behind the arithmetic mean, but the geometric mean focuses on multiplication rather than addition. When I multiply 4x16 = 64, I notice that I am multiplying two different numbers to get the answer of 64. What if I wanted to replace the two different numbers by a single number to get the same answer? Hence, ?x? = 64. To this end we need the square root of 64, which is 8.

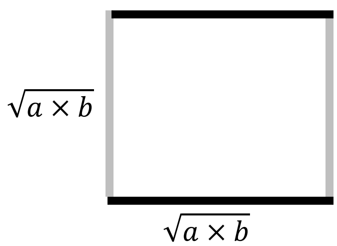
Suppose I made a rectangular laboratory space that is 4 feet by 16 feet. The resulting working space is obtained by multiplication yielding a working area of 64 square feet. So, what would be the dimensions of the room which had the same work space, 64 square feet, but sides of equal length? For this rationale, we need the geometric mean. The arithmetic mean will not work.

Hence, the general rationale for the geometric mean is as follows. Suppose I have two different positive numbers and multiply them to get their product (notice the variation in the numbers you are multiplying together).



Total area $a \times b$.

Now, remove the variation. What would you use if both numbers were replaced by the same quantity but still gave the same product? Answer: replace *a* and *b* by their geometric mean.



Total area $a \times b$.

Voila! Same result without the variation in the components.

One can obtain a variety of applications of the geometric mean by googling “applications of the geometric mean,” especially “applications of the geometric mean in business.”

Harmonic Mean

An understanding of the harmonic mean begins with a classic formula from elementary school: distance = rate x time (in symbols, $d = rt$). For example, if I travel 240 miles in 6 hours, then my average speed is 40 miles per hour. Of course this does not mean that the car was in cruise control and I traveled 40 miles per hour the entire trip. What is so amazing about this averaging formula is that you don’t even know what speeds are being “averaged.”

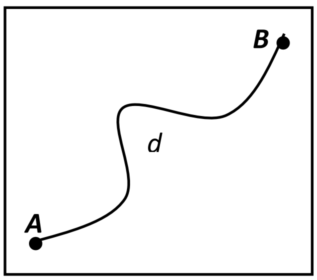
Now suppose I make the return trip along the same route but it takes me 4 hours. My average speed for the return trip is (240 miles)/(4 hours) = 60 miles per hour. So the arrival trip was made at 40 mph and the return trip at 60 mph. What was my average speed for the entire trip? How would I average 40 and 60 to obtain the average trip speed? To answer these questions, we need to determine what rationale of the numbers we want to preserve. For an average total trip speed, the total distance is 480 miles (arrival plus return distances). The total time for the trip is 10 hours. Hence, the average speed for a 480 mile trip in 10 hours is 48 miles per hour. But, AM(40,60) = 50, so the arithmetic mean will not do. Also, GM(40,60) = $\sqrt{40 \times 60}$ = $\sqrt{2400}$ = 48.98979⋯, so the geometric mean won’t do either. We need the harmonic mean.

The **harmonic mean** (HM) of two positive numbers *a* and *b* is given by

$$\text{HM}(a,b) = \frac{2}{\frac{1}{a}+\frac{1}{b}} = \frac{2\times a\times b}{a+b} = \frac{a\times b}{(a+b)/2} = \frac{(\text{GM})^2}{\text{AM}} \quad (*)$$

Hence, the harmonic mean is a hybrid mean being equivalent to the square of the geometric mean divided by the arithmetic mean. In our example, HM(40,60) = 48.

A formal explanation for the harmonic mean formula in (*) may be given as follows. Suppose we want to make a round trip from A to B with distance d between the two points. The total round-trip distance is then $2d$.



Suppose that the time for the arrival trip from A to B is t_1. Since, in general, $d = rt$, then also $r = d/t$ and $t = d/r$. So, the arrival trip speed, r_1, is calculated by $r_1 = d/t_1$. Thus, $t_1 = d/r_1$. Similarly, suppose that the time for the return trip is t_2. Then the return speed, r_2, is calculated by $r_2 = d/t_2$. Thus, $t_2 = d/r_2$.

Now let r be the average speed and t be the total time for the whole round-trip. The round-trip distance is $2d$ and the round-trip time is given by $t = t_1 + t_2$. Then, for the entire round-trip, we have:

	$2d = rt$	
$\Leftrightarrow$	$2d = r(t_1 + t_2)$	substitute: $t = t_1 + t_2$
$\Leftrightarrow$	$2d = r\left(\frac{d}{r_1} + \frac{d}{r_2}\right)$	substitute: $t_1 = d/r_1$ and $t_2 = d/r_2$
$\Leftrightarrow$	$2d = rd\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$	factor out d on the right hand side
$\Leftrightarrow$	$2 = r\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$	divide through by d
$\Leftrightarrow$	$r = \frac{2}{\frac{1}{r_1}+\frac{1}{r_2}}$	algebra
$\Leftrightarrow$	$r = \text{HM}(r_1,r_2)$	definition of harmonic mean

We conclude that the average speed for the round-trip is the harmonic mean of the arrival trip speed and the return trip speed.

3. Averaging

Suppose we are given two positive numbers, a and b. We want to report an average. What are our options? Here are five options that have applications. Only context and rationale can determine which one to use.

min(a,b)	This is the minimum of a and b. It is the "at least" option.
max(a,b)	This is the maximum of a and b. It is the "at most" option.
AM(a,b)	The arithmetic mean.
GM(a,b)	The geometric mean.
HM(a,b)	The harmonic mean.

The following interesting **mean inequality** relationship always occurs for positive numbers where $a < b$.

$$a = \min(a,b) < \mathrm{HM}(a,b) < \mathrm{GM}(a,b) < \mathrm{AM}(a,b) < \max(a,b) = b$$

For example, consider 4 and 16. Then,

$$4 = \min(4,16) < \mathrm{HM}(4,16) = 6.4 < \mathrm{GM}(4,16) = 8 < \mathrm{AM}(4,16) = 10 < \max(4,16) = 16$$

Further, the only way that two of the quantities can be equal happens if and only if the two numbers a and b were the same to begin with.

4. Extending the Means

Thus far we have provided mechanisms for calculating the AM, GM, and HM of two *positive* numbers. The formulas may be extended to include the special case that $a = 0$ and $b > 0$.

The formula for the arithmetic mean applies immediately:

$$\mathrm{AM}(0,b) = \frac{0+b}{2} = \frac{b}{2}$$

Similarly for the geometric mean: $\mathrm{GM}(0,b) = \sqrt{0 \times b} = \sqrt{0} = 0$.

However, there is a situation with the harmonic mean. The formula for HM(a,b) involves $1/a$. When $a = 0$ we have a problem. Fortunately, the algebraic simplification displayed for the harmonic mean formula (see formula * on the previous page) computes for $a = 0$ and $b > 0$:

$$\mathrm{HM}(0,b) = \frac{2\times 0\times b}{0+b} = \frac{0}{b} = 0$$

Although the formulas for the arithmetic, geometric, and harmonic means can be expanded to accommodate $a = 0$, the strict mean inequality does not carry over for $a = 0$. For $a = 0$ we have

$$0 = \min(0,b) = \mathrm{HM}(0,b) = \mathrm{GM}(0,b) < \mathrm{AM}(0,b) = b/2 < \max(0,b) = b$$

However we can expand the scope of the inequality to cover $0 \le a < b$. In this event, we always have

$$a = \min(a,b) \le \mathrm{HM}(a,b) \le \mathrm{GM}(a,b) < \mathrm{AM}(a,b) < \max(a,b) = b$$

Finally we merely note that one can also extend calculating these means for more than two numbers. Suppose that $a_1, a_2, \ldots, a_n$ are n positive numbers. Then,

$$\mathrm{AM}(a_1, a_2, \ldots, a_n) = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

$$\mathrm{GM}(a_1, a_2, \ldots, a_n) = \sqrt[n]{a_1 \times a_2 \times \cdots a_n}$$

$$\text{HM}(a_1, a_2, \dots, a_n) = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}$$

We caution that the alternative forms for HM(a,b) given in (*) do not apply beyond two numbers.

5. An Application: Rounding Decimal Numbers

Suppose that we have a positive decimal number that is not a whole number; i.e., a positive number with a genuine decimal part. Now, decimal fractions can be annoying or impractical. Hence, applying the classic research solution *denial* (what decimal fraction?) may provide convenience. Suppose that the integer part of our decimal number is n. We can then round the decimal down to n or round it up to n+1 to eliminate the decimal fraction. We will denote the rounding of a positive decimal number q by round(q). Note that round(q) will be either n or n+1.

There are five ways, depending on the application, to round a positive decimal number, q, with integer part n. In particular, round(q) =

1. min(n, n+1) = n — The "round down" option.
2. max(n, n+1) = n + 1 — The "round up" option.
3. n + 1 $\Leftrightarrow$ $q \geq n + ½$ $\Leftrightarrow$ $q >$ AM(n, n+1) — The usual method of rounding a decimal.
4. n + 1 $\Leftrightarrow$ $q \geq$ GM(n, n+1)
5. n + 1 $\Leftrightarrow$ $q \geq$ HM(n, n+1)

To illustrate, let's work through an example. We apply the five methods to illustrate how to round 2.437. Should I round up to 3 or round down to 2?

1. round(2.437) = 2 — min option: round down
2. round(2.437) = 3 — max option: round up
3. round(2.437) = 2 — usual rounding method: 2.437 < AM(2,3) = 2.5
4. round(2.437) = 2 — geometric mean option: 2.437 < GM(2,3) = 2.4494⋯
5. round(2.437) = 3 — harmonic mean option: 2.437 > HM(2,3) = 2.4

Exercises

1. Determine the following averages.

   a. min(7,30)
   b. max(7,30)
   c. AM(7,30)
   d. GM(7,30)
   e. HM(7,30)

2. Compute the arithmetic, geometric, and harmonic mean for each pair of numbers.

   a. 3 and 48 b. 4 and 7 c. 8 and 20 d. 8 and 9 e. 6 and 12

3. Complete the table by rounding each of the given decimal numbers using the indicated method.

Decimal	Method				
	min	max	AM	GM	HM
3.2					
1.4					
1.463					
5.6					

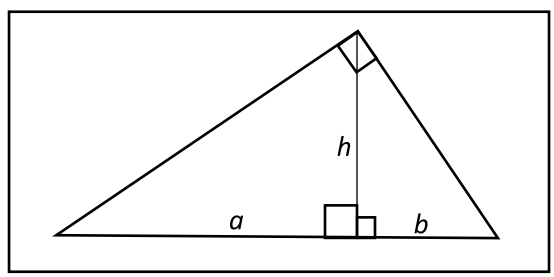
4. **Sabermetrics** is the mathematical and statistical analysis of baseball records.[2] Most baseball fans are familiar with routine measures such as batting average, on base percentage, and even the power batting average where each hit is weighted by the number of bases reached by the hit. Some baseball aficionados find the **power-speed number** for a player a useful statistic to measure a player's clutch offensive productivity. The power-speed number is the harmonic mean of a player's home runs and stolen bases. The lifetime major league leaders for the power-speed index are the Giants legend Barry Bonds (613.90) and the *A's* Ricky Henderson.[3]

   a. Lifetime, Ricky Henderson hit 297 home runs and had 1406 stolen bases. What was Ricky Henderson's lifetime power-speed number?
   b. Ricky Henderson played 14 years for the Oakland *A's*. During this time with the *A's*, Henderson hit 167 home runs and stole 867 bases. What was Henderson's power-speed number at Oakland?

5. Compute the arithmetic, geometric and harmonic mean for each set of numbers.

   a. 1, 3, 5.
   b. 10, 12, 15, 18, 20.

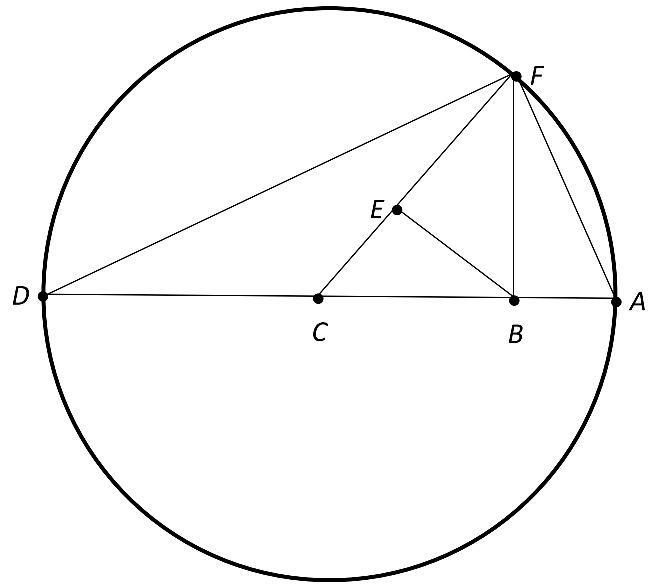
6. Use the adjacent figure to show that h is the geometric mean of a and b. The three indicated angles in the figure are all right angles. [Hint: The Pythagorean Theorem can come in handy.]



[2] http://www-math.bgsu.edu/~albert/papers/saber.html.
[3] http://www.baseball-reference.com/leaders/power_speed_number_career.shtml.

1. (Challenge) Consider the circle shown at right. Let $|\overline{AB}| = a$ and $|\overline{BD}| = b$. The center of the circle is at *C*. $\overline{FB} \perp \overline{AD}$ and $\overline{BE} \perp \overline{CF}$.

   a. Verify that AM(*a*,*b*) = $|\overline{AC}|$.
   b. Verify that GM(*a*,*b*) = $|\overline{BF}|$.
   c. Verify that HM(*a*,*b*) = $|\overline{EF}|$.
   d. Explain why the diagram is a "proof without words" for the mean inequality.



Charles M. Biles, Ph.D.
Professor of Mathematics, Emeritus
Humboldt State University
Arcata, CA

Open source reference: http://www.nia977.wix.com/drbcap. This *Lesson* is available under Resources.

You are welcome and encouraged to please email me any comments, critique, suggestions, etc. Professors: please feel free to adapt to your own classroom use.

Thank you for your considerations.

I am especially grateful to Professors Dan Munton and John Martin from Santa Rosa Junior College, Robin Carter from College of the Redwoods, and Kamila Larripa, Tim Lauck, Holland Heese, and Steven Margell for their helpful collaborations. I also thank Grace Kimura, University of New Mexico, for her thoughtful review.

Revised 13 February 2017