Applied Mathematical Sciences, Vol. 3, 2009, no. 3, 119 - 128

# Numerical Relations within Cubes

# and the Estimation of Central Tendency

**G. L. Silver**

Los Alamos National Laboratory*
P.O. Box 1663, MS E502
Los Alamos, NM 87545, USA
gsilver@lanl.gov

*Los Alamos National Laboratory is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. DE-AC52-06NA25396.

**Abstract**

The discovery of relationships connecting data in cubical array is pertinent to data analysis. New equations that interrelate data at the vertices, edges, and sides of a cube are obtained by the shifting operator. A new center point estimator is described. It competes with the mean and the median as illustrated by examples.

**Mathematics subject classifications:** 65D07, 65D17

**Keywords:** Interpolation, Experimental design, Shifting operator, Operational equations, Central tendency

## 1. Introduction

The identities of trigonometry have been known for centuries as relationships among angles. It has recently been pointed out that they also represent formerly unknown relationships among points in two- and three-dimensional space [3]. This paper illustrates the derivation of new relationships among data at the vertices and center points of the faces of a cube. It also describes a new sequence of expressions that estimate the centers of ordered numbers. Its members compete with the mean and the median as estimators of central tendency.

## 2. Relations among data at the vertices and face-center points of the cube

A trigonometric identity expressed in the three letters x, y, z appears as Eq. (1). Let Eq. (1) and subsequent identities be converted to their Euler forms. Let the symbol F(x,y,z) denote an unknown function at the center point E of the cube in Fig. 1. Eq. (1) can be multiplied by the third-power of the function, $F^3(x,y,z)$ or $E^3$, as illustrated in Ref. [3]. The result of this operation simplifies to Eq. (2). The center point of a face of the cube is given a four-letter designation such as BDIG. See Fig. 1.

$$2\sin(x)\cos(x)\cos(x-y-z) - 2\cos(x)^2\sin(x-y-z) + \sin(x-y-z) + 2\sin(x)\cos(x)\cos(x+y-z) - 2\sin(x+y-z)\cos(x)^2 + \sin(x+y-z) = \sin(x+y+z) + \sin(x-y+z) \quad (1)$$

$$(B - H + D - F - I + A - G + C)E^2 + (BDIG + ACHF)[(H + F)(BDIG) - (B + D)(ACHF)] = 0 \quad (2)$$

Another three-parameter identity appears as Eq. (3). The foregoing procedure converts it into Eq. (4).

$$2\sin(x)\cos(x)\cos(x+y+z) - 2\sin(x+y+z)\cos(x)^2 + 2\sin(x)\cos(x)\cos(x-y-z) - 2\cos(x)^2\sin(x-y-z) + 2\sin(x)\cos(x)\cos(x+y-z) - 2\sin(x+y-z)\cos(x)^2 + \sin(x+y-z) - \sin(x-y+z) = 0 \quad (3)$$

$$(D - F - G + C)E^2 + (BDIG + ACHF)[(H + A + F)(BDIG) - (ACHF)(D + B + I)] = 0 \quad (4)$$

Eqs. (2) and (4) are exact on linear numbers in cubic array as in Fig. 1. They are also exact when these numbers appear as arguments of simple functions like $2^x$, as well as sin(x), cos(x), their two-member linear combinations, and sinh(x), cosh(x) and their two-member linear combinations. Eqs. (2) and (4) are used to estimate the datum at the center of one face of the cube given the vertex data A-I and the datum at the center of the opposing face. The cube can be rotated so that Eqs. (2) and (4) can be applied to the estimation of other face-center points [4].

Eqs. (2) and (4) are sensitive to translation of the data. Eq. (2) can be augmented by adding a parameter *t* to each datum. Eq. (5) results when the parameter is allowed to increase indefinitely. Eq. (6) results when Eq. (4) is similarly treated. Both equations are exact on trilinear numbers and their squares arranged as in Fig. 1. Elimination of BDIG and ACHF yields Eq. (7), a simple relationship among the numbers at the vertices of the cube. It is exact on trilinear numbers and their squares but otherwise it is not a good predictor.

$$4(BDIG - ACHF) - B + H - D + F - I + A - G + C = 0 \quad (5)$$

$$6(BDIG - ACHF) - 2B + 2H - D + F - 2I + 2A - G + C = 0 \quad (6)$$

$$B - H - D + F + I - A - G + C = 0 \quad (7)$$

A three-parameter identity can be multiplied by the third power of the unknown function at the center point of the cube as illustrated by Eq. (8). The result of this operation is Eq. (9). It can be rewritten as Eq. (10).

$$E^3[\cos(x+y+z) + \cos(x+y-z) + \cos(x-y+z) + \cos(y+z-x) - (4)\cos(x)\cos(y)\cos(z)] = 0 \quad (8)$$

$$E^2[(A+I)/2 + (D+F)/2 + (G+C)/2 + (H+B)/2] - 4[((BDIG + ACHF)/2)((CDIH + ABGF)/2)((FGIH + ABDC)/2)] = 0 \quad (9)$$

$$(A + B + C + D + F + G + H + I)E^2 - (BDIG + ACHF)(CDIH + ABGF)(ABDC + FGIH) = 0 \quad (10)$$

Let a term *t* be added to every datum in Eq. (10). Let *t* increase without limit. The result of this operation is Eq. (11), another equation connecting the face-center points to the center and corner points of the cube. Eq. (11) is exact on the first, second, and third powers of linear numbers as applied to the vertices of the cube. It is insensitive to translation of the data. Eq. (11) is not new but its connection to operational equations like Eq. (10) has not been previously demonstrated.

$$(A + B + C + D + F + G + H + I + 16E) - 4(BDIG + ACHF + CDIH + ABGF + ABDC + FGIH) = 0 \quad (11)$$

Let an arbitrary, three-parameter sum or product be selected for examination. One choice is cos(x+y+z)cos(x–y–z)cos(x–y+z)cos(x+y–z). It can be expanded into a sum of products of the powers of the cosines of one letter as in Eq. (12).

$$\cos(x+y+z)\cos(x-y-z)\cos(x-y+z)\cos(x+y-z) = (4)\cos(x)^2\cos(y)^2\cos(z)^2 - (2)\cos(x)^2\cos(y)^2 - (2)\cos(y)^2\cos(z)^2 - (2)\cos(x)^2\cos(z)^2 + \cos(x)^4 + \cos(y)^4 + \cos(z)^4 \quad (12)$$

Eq. (12) can be multiplied by the sixth power of an arbitrary function: $F^6(x,y,z)$ or $E^6$. The result is Eq. (13). It relates the numbers at the vertices, center point, and face-center points of the cube as in Fig. 1. Eq. (11) results when Eq. (13) is treated by the parameter-addition and limit process so Eq. (11) has at least two precursor equations.

$$
\begin{aligned}
&E^2((I + A)/2)((B + H)/2)((C + G)/2)((D + F)/2) \\
&\quad - (4)((BDIG + ACHF)/2)^2((CDIH + ABGF)/2)^2((FGIH + ABDC)/2)^2 \\
&\quad + 2E^2((BDIG + ACHF)/2)^2((ABGF + CDIH)/2)^2 \\
&\quad + 2E^2((ABGF + CDIH)/2)^2((ABDC + FGIH)/2)^2 \\
&\quad + 2E^2((BDIG + ACHF)/2)^2((ABDC + FGIH)/2)^2 \\
&\quad - E^2((BDIG + ACHF)/2)^4 - E^2((ABGF + CDIH)/2)^4 - E^2((ABDC + FGIH)/2)^4 \\
&\quad = 0 \qquad (13)
\end{aligned}
$$

## 3. New measure of central tendency

The most popular measure of central tendency is the arithmetic average or mean. It has many advantages: familiarity, simplicity, exactness on linear numbers, invariance under translation of the data, and a differentiable connection to each datum. A disadvantage is its sensitivity to an aberrant datum or “outlier.” The second popular choice of center point estimator is the median. It is less familiar than the average but it supplies invariance under translation of the data and it is insensitive to an outlier. The median has two disadvantages: it requires the data to be sorted by magnitude and it has no connection to the data by means of an easily differentiable expression.

Two sequences of operational estimators of central tendency have recently been illustrated [5]. The members of one sequence are denoted (P4E), (P5E), (P6E) and (P8E) in Ref. [5]. Rousseeuw suggests a desirable property of center point estimators is invariance under translation of the data [2]. If Rousseeuw is correct, the sequence of center point estimators denoted by suffix E is flawed because it does not provide the desirable property of translational invariance.

Another sequence of center point estimators in Ref. [5] is denoted by the suffix P. That sequence meets Rousseeuw’s criterion of invariance under translation of the data. It has the advantages of exactness on linear numbers and their squares. No other common estimators of central tendency combine translational invariance with exactness on linear numbers and their squares. The members of the sequence of terms suffixed by P are also more resistant than the mean to the adverse effects of an aberrant datum or “outlier.” The described attributes have been illustrated by Tables 1-6 in Ref. [5]. Note that the third and fourth column headings in Table 5 of Ref. [5] should read Eq. (6) and Eq. (7), respectively. The title of Table 6 should read Eqs (8) and (10).

The presentation of the first three members of the P sequence in Ref. [5] is awkward and their properties were not described. To make the subject clearer, the first three members of the sequence are rewritten as Eqs. (14)-(16). The symbols P4P, P6P, and P8P represent estimates of the centers of the four, six, and eight numbers, respectively. They apply to numbers ordered by magnitude denoted z1, z2, z3… .

$$P4P = [(z4–z1)^2(z3+z2) – (z4+z1)(z3–z2)^2] / [2(z4–z1)^2 – 2(z3–z2)^2] \quad (14)$$

$$P6P = [2(z6–z1)^2(z5–z2)^2(z4+z3) – (z6–z1)^2(z5+z2)(z4–z3)^2 – (z6+z1)(z5–z2)^2(z4–z3)^2] / [4(z6–z1)^2(z5–z2)^2 – 2(z6–z1)^2(z4–z3)^2 – 2(z5–z2)^2(z4–z3)^2] \quad (15)$$

$$P8P = [3(z8–z1)^2(z7–z2)^2(z6–z3)^2(z5+z4) – (z8–z1)^2(z7–z2)^2(z6+z3)(z5–z4)^2 – (z8–z1)^2(z7+z2)(z6–z3)^2(z5–z4)^2 – (z8+z1)(z7–z2)^2(z6–z3)^2(z5–z4)^2] / [6(z8–z1)^2(z7–z2)^2(z6–z3)^2 – 2(z8–z1)^2(z7–z2)^2(z5–z4)^2 – 2(z8–z1)^2(z6–z3)^2(z5–z4)^2 – 2(z7–z2)^2(z6–z3)^2(z5–z4)^2)] \quad (16)$$

The formulas display patterns in their numerators. For example, the numerator of P4P contains two members. The first one has a positive sign and coefficient 1. The remaining member has a negative sign and coefficient 1. Each member consists of two terms. The numerator of P6P contains three members. The first one has a positive sign and coefficient 2. The remaining members have negative signs and coefficients 1. Each member consists of three terms. The numerator of P8P has four members. The first one has a positive sign and coefficient 3. The remaining members have negative signs and coefficients 1. Each member consists of four terms. The positive sign migrates within the numerators of the individual members. The described rules apply to successive formulas that apply at least up to 16 ordered numbers.

The denominators also display patterns. The denominator of P4P contains two members. The first one has a positive sign and coefficient 2. The other one has a negative sign and coefficient 2. Each member contains one term. The denominator of P6P contains three members. The first one has a positive sign and coefficient 4. The remaining members have negative signs and coefficients 2. Each member has two terms. The denominator of P8P has four members. The first one has a positive sign and coefficient 6. The remaining members have negative signs and coefficients 2. Each member has three terms. The rules apply to expressions up to at least 16 ordered numbers.

The numbers in P4P, P6P, and P8P are denoted zX where z represents a sorted number and suffix X is its ordered position rank. The suffixes X in the numerators and denominators exhibit patterns. The first number in each term has the largest value of X.

The sum of the two X suffixes inside each term is the same. In the numerators of P6P and successive estimators, the plus sign migrates from right to left. If N data are concerned, the numerators contain N/2 members of N/2 terms each. The denominators contain N/2 members of N/2–1 terms each. Terms that represent squared differences remain the same or decrease from left to right.

Successive members of the suffix-P sequence are formed by induction based on generalizing observed patterns in numerators and denominators. Generation of the successive estimators by induction applies up to expressions for 16 ordered numbers at least. This observation suggests the series can be extended indefinitely. It also suggests the existence of a general expression for the estimators based on the sign of summation (Sigma notation) and another general expression based on the sign of integration. These presumed expressions are not known so they are opportunities for research. An additional challenge is the presentation of analogous formulas for the suffix-E series of operational center point estimators [5].

As an example of an operational formula, choose the ten numbers suggested by Rousseeuw [2]. After ordering them, they are: 40,75,80,83,86,88,90,92,93,95. He gives the mean of the numbers as 82.2, the median as 87.0, the LMS estimate of their center as 90.5, and the reweighted mean as 86.9 [2]. The operational estimate of the center of the ten numbers, using a ten-number, suffix-P formula, is 87.0. The operational estimate agrees with the median.

The mean and the median are easier to apply than the operational formulas. On the other hand, computers make the operational estimators easy to store and easy to apply. The mean and the median do not fail when encountering identical numbers as data. The operational estimators fail when the data are all equal numbers. They may also fail if several data are equal. The mean and the operational formulas can be differentiated with respect to any datum. The median is not readily differentiated with respect to any datum. The mean, median, and the operational estimators require a minimum of two, three, and four numbers, respectively.

For purposes of illustration, let eight basis numbers be chosen as the first eight ordered integers. Their mean and median are both 4.5. Operations M can be applied to 4.5 as well to the basis numbers. Thus, if $M^2$ is the squaring operation, the "true" center of the eight data is M(4.5) or 20.25. Table 1 illustrates that both the median and the operational formula P8P render estimates of the center of the data that are closer to the "true" values than the mean. That is, the median and the operational estimate are often more accurate than the mean in the context of the illustration.

In order to estimate the center of an odd-numbered set of ordered data by an operational formula, remove the median. An even number of ordered data remains. Estimate the center of remaining data with an operational formula taken from the described sequence. Call this estimate Q. The center of the original set of ordered numbers is the arithmetic average of its median and Q. That is, the expression (median + Q)/2 is the operational estimate of a set of an odd number of ordered data [5].

Other sequences of operational center point formulas have been described [5,6,7]. Two of them yield estimators that are sensitive to translation of the data. In spite of their potential accuracies, they are deficient from Rousseeuw's point of view. One approach depends on the repetitive application of a four-point operational formula. Computers can be programmed to execute those operations. However, the approach can fail on two or more equal, successive numbers or four successive, linear numbers. Sorted data often contain such sequences so the cited method is more difficult to apply.

## 4. Discussion

The series typified by Eqs. (14)-(16) represents the operational analog of the arithmetic mean. The series denoted by the suffix E is the operational analog of the geometric mean [5]. Unlike the geometric mean, the operational formulas are not reduced to zero when they encounter one zero as a datum. This advantage partially compensates for the disadvantage of their complexity. The ten-number, suffix-E center-point formula, P10E, renders 87.0 as the center of Rousseeuw's trial data cited above [2,5]. Copies of the first seven members of each series are available from the author.

The identity in Eq. (17) can be converted to its Euler form and multiplied by $E^2$ as described above. This converts it into Eq. (18), an equation connecting data at the centers of the faces of the cube (four letter notations) and at two of its vertices (single letter notations) as in Fig. 1. Two-letter notations like GI represent a datum at the midpoint of edge GI in Fig. 1. Eq. (18) is exact on linear numbers and simple expressions like $2^x$, sin(x), sinh(x), and $x^2$. It is invariant under translation of the data. Eq. (18) is an example of how the shifting operator can be used to relate data in a cubical array.

$$\sin(x)\sin(y) + \sin(z)\sin(x+y+z) - \sin(x+z)\sin(y+z) = 0 \tag{17}$$

$$(BDIG - ACHF)(CDIH - ABGF) + (FGIH - ABDC)(I - A) - (GI - AC)(HI - AB) = 0 \tag{18}$$

Eq. (19) is a relationship connecting the midpoints of the faces of the cube in Fig. 1. It is exact on simple forms like $2^x$, sin(x), and sinh(x). It is sensitive to translation of the data. A term *t* can be added to each term in Eq. (19). A second relationship is obtained when *t* is permitted to increase indefinitely. The new form is Eq. (20). It is exact on linear data and there squares, as illustrated in Fig. 1, and it is invariant under their translation. Eqs. (18)-(20) can be used to estimate data that are missing from prismatic arrays [8,9].

$$\begin{aligned}&[(FGIH)(ABDC) – (ACHF)(BDIG)](CDIH)^2\\&+ (ABGF)(CDIH)[(BDIG)^2 + (ACHF)^2 – (FGIH)^2 – (ABDC)^2]\\&+ (FGIH)^2(BDIG)(ACHF) – (FGIH)(ABDC)(BDIG)^2 – (FGIH)(ABDC)(ACHF)^2\\&– (ABGF)^2(BDIG)(ACHF) + (FGIH)(ABDC)(ABGF)^2 + (ABDC)^2(BDIG)(ACHF) = 0\end{aligned} \tag{19}$$

$$\begin{aligned}&[(2((ACHF) + (BDIG) – (ABGF) – (CDIH))(FGIH)\\&+ ((BDIG) – (ACHF))^2 – ((ABGF) – (CDIH))^2)](ABDC)\\&+ [(CDIH) – (BDIG) – (ACHF) + (ABGF)](ABDC)^2\\&+ [((BDIG) – (ACHF))^2 – ((CDIH) – (ABGF))^2](FGIH)\\&+ [(CDIH) – (BDIG) – (ACHF) + (ABGF)](FGIH)^2\\&– [((BDIG) – (ACHF))^2 + 2(ABGF)((ACHF) + (BDIG))](CDIH)\\&+ [(ACHF) + (BDIG)][(ABGF)^2 + (CDIH)^2]\\&– [(BDIG) – (ACHF)]^2(ABGF) = 0\end{aligned} \tag{20}$$

Eq. (21) is an identity whose operational interpretation is Eq. (22). That equation relates data at the midpoints of the edges of the cube in Fig. 1. Eq. (22) is exact on trilinear numbers and their squares and it is invariant under translation of the data. It is likewise exact on trilinear numbers as exponents in simple expressions like $2^x$, and in cos(x) and cosh(x). Eq. (23) yields Eq. (24) as its operational interpretation. Eqs. (22) and (24) have similar properties.

$$\sin(x – z)\sin(x + z) – \sin(y – z)\sin(y + z) – \sin(x – y)\sin(x + y) = 0 \tag{21}$$

$$(BD – FH)(GI – AC) – (CD – FG)(HI – AB) – (BG – CH)(DI – AF) = 0 \tag{22}$$

$$\sin(y – z)\sin(x)^2\sin(y + z) – \sin(x – z)\sin(y)^2\sin(z + x) + \sin(x – y)\sin(z)^2\sin(x + y) = 0 \tag{23}$$

$$\begin{aligned}&(CD – FG)(BDIG – ACHF)^2(HI – AB) – (BD – FH)(CDIH – ABGF)^2(GI – AC)\\&\quad + (BG – CH)(FGIH – ABDC)^2(DI – AF) = 0\end{aligned} \tag{24}$$

Operational equations are part of a subject called symbolic methods. They are based on application of the shifting operator to trigonometric identities. They are new instruments for data analysis. A brief introduction to the history of symbolic methods, including their reception by society, can be found in the book by E. T. Bell [1].

Table 1. Comparisons of the centers of eight ordered data generated by applying selected functions M to the integers 1-8. The centers of the data are estimated by the mean, the median, and formula P8P in the text. The true value is taken as M(4.5).

| Function | Mean | Median | Formula P8P | True value |
|---|---|---|---|---|
| M | 4.5 | 4.5 | 4.5 | 4.5 |
| $M^2$ | 25.5 | 20.5 | 20.3 | 20.3 |
| $M^3$ | 162 | 94.5 | 91.8 | 91.1 |
| $2^M$ | 63.8 | 24.0 | 23.3 | 22.6 |
| Ln(M) | 1.33 | 1.50 | 1.50 | 1.50 |
| (M)Ln(M!) | 28.4 | 18.3 | 17.9 | 17.8 |
| 100/M | 34.0 | 22.5 | 22.3 | 22.2 |
| $100/M^2$ | 19.1 | 5.13 | 5.04 | 4.94 |
| $M^M/100$ | 22064 | 16.9 | 16.6 | 8.70 |
| $Ln(M^M)$ | 7.41 | 6.80 | 6.76 | 6.77 |

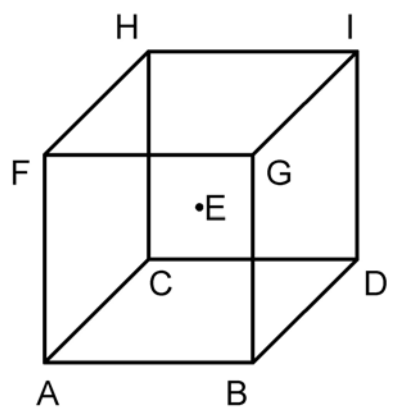


Fig. 1. The nine-point cube.

## References

1. E. T. Bell, The Development of Mathematics, McGraw-Hill Book Company, New York, 1940. pp. 387-389.

2. P. J. Rousseeuw, A. M. Leroy, Robust Regression and Outlier Detection. Wiley-Interscience, New Jersey, 2003. See pp. 158, 165, 169.

3. G. L. Silver, Operational method of data treatment. J. Comput. Chem. 6 (1985) 229-236.

4. G. L. Silver, Analysis of three-dimensional grids: Sine-and-cosine equation for the nine-point prismatic array. Appl. Math. Comput. 183 (2006) 617-621.

5. G. L. Silver, Operational measures of central tendency. Appl. Math. Comput. 186 (2007) 1379-1384.

6. G. L. Silver, Operational center points. Appl. Math. Comput. 188 (2007) 1042-1048.

7. G. L. Silver, Center points and grid equations from sub-grid equations. Appl. Math. Comput. 189 (2007) 1467-1475.

8. G. L. Silver, Analysis of three-dimensional grids: Interpolation of data in prismatic array and the estimation of a missing datum. Appl. Math. Comput. 174 (2006) 1398-1407.

9. G. L. Silver, Analysis of three-dimensional grids: The estimation of two missing data. Appl. Math. Comput. 184 (2007) 743-747.

Received: June 7, 2008