

## A COEFFICIENT OF VARIABILITY

Manuel Martinez-Pons

Brooklyn College, City University of New York, USA

Received 2012-06-15, Revised 2012-09-05; Accepted 2013-04-12

### ABSTRACT

A coefficient of variability is described. It is shown that conducting all possible paired comparisons in a group of scores, the investigator can ascertain the group's degree of variability as well as its converse, the group's degree of homogeneity. The coefficient, useful in determining whether to divide a class into subgroups for the purpose of instruction, has the advantage that it is in the form of a proportion and that it thus provides a common ground for comparison of variability across measures.

**Keywords:** Variability, Variability Coefficient, Homogeneity Coefficient

### 1. INTRODUCTION

Two issues often arise in the course of an investigation: the degree of homogeneity a group exhibits on some measure and the difference in homogeneity the group exhibits across two or more measures. These issues assume particular relevance when the interest lies in deciding whether to subdivide the group on the basis of the information at hand. As an example in education, while seeking to ascertain the degree of homogeneity of a class on each of two tests, an instructor may also want to compare the class' homogeneity levels across the two exams (in pursuing the matter, having administered tests of, say, learning style and learning readiness, he or she may find that while all class members share a collaborative learning style, only half the class is prepared to undertake instruction in, say, the multiplication of fractions).

Often, a starting point in the treatment of such issues is an examination of group variability and the variance ( $s^2$ ) and standard deviation ( $s$ ) may come to mind as the statistics applicable to the task.  $s^2$  and  $s$ , however, are poorly equipped for the purpose: first,  $s^2$  and  $s$  are not comparable across measures when the means or the measurement scales at hand differ (although standardized scores may afford some degree of comparison under such circumstances, their use for present purposes requires relatively large normally distributed samples, conditions not always present in the classroom). In addition, lack of a natural ceiling in  $s^2$  and  $s$  limits their usefulness in determining variability's converse, i.e.,

homogeneity, or in providing an intuitively meaningful interpretation of their size. While the Coefficient of Variation (Croxtton *et al.*, 1967), defined as  $s/M$ , is often used to compare two standard deviations when their means differ substantially, it, too, is inadequate for present purposes: because  $s$  is not always smaller than the mean, it is possible for CV to be greater than 1-lack of a natural ceiling which, as in the case of  $s^2$  and  $s$ , makes a definitive interpretation of the size of CV impossible.

The purpose of this study is to describe a measure of variation that (a) has a natural ceiling, affording a universally interpretable index of variation and providing a clear basis for comparing variability across measures; and (b) is easily converted to a natural-ceiling measure of group homogeneity. The study begins with a definition of variability and then presents measures of variability and homogeneity based on this definition.

#### 1.1. Variability

Variability consists of the differences in magnitude that exist in a set of occurrences of some measure. If at least one occurrence differs in magnitude from the others, the set exhibits variability; if no difference occurs, then the set does not exhibit variability. When only one occurrence differs in size from the others, the set exhibits minimum variability; and the greater the total difference in magnitude among the occurrences, the greater the variability exhibited by the set. The following fictitious test scores manifest variability because differences in magnitude occur among them: 4, 3, 7, 9, 5 and 2.

### 1.2. A Coefficient of Variability

If variability is seen in this light, then its measure can be formulated as the sum of the observed differences among occurrences of a measure divided by the maximum possible sum of the differences. This operation can be expressed as follows Equation 1:

$$vc = OV/MPV \tag{1}$$

Where:

- vc = Variability coefficient
- OV = Sum of obtained absolute comparisons
- MPV = Maximum possible variation

The numerator (OV) in this expression consists of the sum of the absolute differences among occurrences of the measure at hand. A matrix arrangement of the differences among a group of scores is helpful in visualizing the calculations used to derive OV. For the six scores examined earlier, the matrix is displayed in **Table 1**.

The scores in **Table 1** appear vertically along the table's left as well as horizontally along its top. For each row, the cells represent the difference between the score on the left column and the other scores in the set. Each score on the horizontal list is subtracted from each of the scores on the vertical list and the remainder for each subtraction is recorded as an absolute value in the intersecting cell. If no difference emerges, a 0 is recorded. For example, the comparison of the first and second scores yields a difference (i.e.,  $4-3 = 1$ ), which is recorded as an absolute value in the second cell from the left on the first row. The triangular half above the diagonal is used to carry out the OV calculations. In **Table 1**, OV, or the sum of the absolute cell values, is 48.

The derivation of the denominator (MPV) in (1) is based on the following reasoning: the maximum sum of differences in a set of scores will occur if half the scores have the lowest value contained in the set and the other half carry the highest value. For example, for a group of six scores with a low value of 2 and a high value of 9, the highest variation will occur if the data take the following values: 2, 2, 2, 9, 9, 9 ( $Sum_{diff} = 63$ ). Compare this set with the following, which will yield a lower sum: 2, 2, 3, 9, 9, 9 ( $Sum_{diff} = 62$ ) or the following, which will yield a lower sum still: 2, 2, 3, 8, 9, 9 ( $Sum_{diff} = 61$ ). The more the upper and lower halves of the set depart from the extreme values, the lower the sum of differences exhibited by the group. This reasoning can be extended to the derivation of MPV on the basis of the number of comparisons possible between the high and low scores in a two-value data set. For a comparison matrix of a data set half of which consists of one uniform value and half of which consists of a different uniform value, only comparisons of the two different values will yield non-zero remainders.

**Table 1.** Matrix arrangement of differences in a group of scores

	4	3	7	9	5	2
4	0	1	3	5	1	2
3		0	4	6	2	1
7			0	2	2	5
9				0	4	7
5					0	3
2						0

**Table 2.** Matrix arrangement of the calculation of MPV

	2	2	2	9	9	9
2	0	0	0	7	7	7
2		0	0	7	7	7
2			0	7	7	7
9				0	0	0
9					0	0
9						0

For an even number of cases, the number of such comparisons is the number of scores in the group's lower half multiplied by the number of scores in the group's upper half, that is  $(N/2) (N/2)$  and thus, the number of non-zero comparisons will equal the square of half the cases in the data set that is,  $(N/2)^2$ . The highest possible variability will consist of the product of this square and the sum of the comparisons of the two values. Thus, for a group of scores consisting of an even number of cases, MPV can be calculated as follows Equation 2:

$$MPV = (N/2)^2 R \tag{2}$$

where, N = group size and R = the range, or the difference between the highest and lowest scores.

**Table 2** illustrates the calculation of MPV for a hypothetical data set in which half the scores carry the low value of 2 and half carry the high value of 9 and which thus exhibits maximum variability.

The comparisons of the two values in this table, each yielding a remainder of 7, appear enclosed within the heavy-lined rectangle. The number of such comparisons is  $(6/2)^2 = 9$ . Hence, for the two-valued data set with an even number of cases in **Table 2**:

$$\begin{aligned}
 MPV &= (N/2)^2 R \\
 &= 3^2 (9-2) \\
 &= 9 \times 7 \\
 &= 63
 \end{aligned}$$

For a group of scores consisting of an odd number of cases, MPV can be calculated as follows Equation 3:

$$MPV = [(N-1)/2] [(N + 1)/2] R \tag{3}$$

For the data appearing in **Table 1**, going by (1):

$$\begin{aligned} vc &= OV/MPV \\ &= 48/63 \\ &= 0.762 \end{aligned}$$

Because *vc* is in the form of a proportion, it is comparable across measures. Also because of its proportional form, *vc* has an advantage over the Coefficient of Variation (CV), alluded to earlier, defined as  $s/M$ , often used to compare two standard deviations when their means differ substantially. As already noted, the standard deviation is not always smaller than the mean and thus, it is possible for CV to be greater than 1-lack of a natural ceiling which, as in the case of  $S^2$  and  $S$ , makes a definitive interpretation of its size impossible. Such interpretation is possible for *vc*, since it can never exceed 1. For example, the above *vc* value of 0.762 can be interpreted as a level of variability on this test that is 76.2% of the maximum possible variation for the sample at hand. By contrast, a group with a *vc* of .20 would show less than one third of the variability of the group in **Table 1**, regardless of the means or scales of measurement involved.

The range, *R* in (2) and (3), can be unduly affected by outliers. When this problem occurs, Stevens (1996) recommendation can be followed for dealing with outliers in general: perform two analyses, one including the outlier and one excluding it and report both findings.

### 1.3. A Coefficient of Homogeneity

A coefficient of homogeneity (*hc*) can be derived as the converse of *vc*, that is Equation 4:

$$hc = 1-vc \tag{4}$$

For the scores in **Table 1**:

$$\begin{aligned} hc &= 1-0.762 \\ &= 0.238 \end{aligned}$$

This value can be interpreted as a level of homogeneity on this measure that is 23.8% of the group's maximum possible homogeneity.

The more closely *vc* approximates 1 (and hence, the more closely *hc* approximates 0), the more likely it is that a group consists of two distinct sub-groups of approximately equal size relative to the issue at hand and that therefore, the investigator may want to address the issue by attending to each sub-group's unique characteristics. Conversely, the more closely *vc*

approximates 0 (and hence, the more closely *hc* approximates 1), the more likely it is that the group consists of one set relative to the issue at hand and that the investigator may thus be able to address the issue by approaching the group as a cohesive whole. In the example in the introduction involving learning style and learning readiness, a *vc* of 0 for the former would be indicative of a collaborative learning style shared by the entire class and a *vc* of 1 for the latter would be indicative of two distinct sub-groups regarding readiness to undertake instruction in the multiplication of fractions.

It is important to note that a *vc* approximating 1, suggesting the likelihood of two distinct sub-groups, is not necessarily also indicative that the sub-groups are substantially different. For example, the following scores will yield a *vc* of 1 while representing what might be considered a negligible difference between the sub-groups' means:

$$1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ (M_1-M_2 = 1)$$

In general, the substance of *vc* = 1 is an issue involving a judgment regarding the magnitude of the difference between the two means. In the example in the introduction, the instructor would have to decide whether the difference between the sub-groups' means on the learning readiness measure is large enough to merit attention.

## 2. CONCLUSION

In conclusion, a coefficient of variability, *vc*, is presented. It is in the form of a proportion and hence, if overcomes the limitations of  $s$  and  $s^2$  in providing clear indexes of variability and homogeneity due to their lack of a natural ceiling.

## 3. REFERENCES

- Croxton, F.E., D.J. Crowden and S. Klein, 1967. Applied General Statistics. 3rd Edn., Prentice-Hall, New York, pp: 754.
- Stevens, J., 1996. Applied Multivariate Statistics for the Social Sciences. 3rd Edn., NJ, Erlbaum, Mahwah, ISBN-10: 0805816704, pp: 659.