DE GRUYTER OPEN

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 50 (63) 2017
DOI: 10.1515/slgr-2017-0018

Adam Kęska
Institute of Political Studies
Polish Academy of Sciences

AXIOMATIC DETERMINATION OF A CLASS OF ORDINAL VARIATION MEASURES

Abstract. The article deals with the problem of the dispersion of ordinal variables. At first, it specifies the very concept of dispersion for this type of scale. Then some of the most known measures that fit to the concept of ordinal variation are recalled. They are constructed with two different types of statistical models: using loss functions and using distance functions. Finally, a new approach, which is the use of an axiomatic method for the construction of a dispersion measure, is proposed. Some relations and comparisons between different measures and between different approaches are shown.

Keywords: descriptive statistics, ordinal variables, ordinal variation, dispersion measures, axiomatic method.

Introduction

Identifying the basic parameters describing the distribution of an ordinal variable is not as trivial as in the case of other types of scales. Particularly problematic is the issue of dispersion, because there is no universally accepted and applied parameter of dispersion for ordinal variables. It is both needed to specify the concept of dispersion in the case of ordinal variables and to define a measure that has adequate properties. There have already been some attempts to introduce measures of dispersion for ordinal variables. I will skip *ad hoc* propositions and at first I will refer to those that are based on some statistical models. I will show some relations between them and compare their behavior in practice. Then I turn to a new approach, which is the use of an axiomatic method for the construction of a dispersion measure. In the end, it will be apparent that this new approach is also connected to earlier designs.

1. Concepts of dispersion

Dispersion may be understood so differently that you can actually define it as a group of distributions' characteristics. What members of this group have in common is that they always refer to the extent to which cases are clustered in a "typical" value and its surroundings.

1.1. Variation and diversity

The above observation that dispersion is a group of characteristics and not just one property, results from the fact that a concentration in the typical value can be understood differently. If we only consider just how frequencies (and thus all the cases) are distributed among the categories, we would then mean the concept of **diversity**. The more the distribution is concentrated on one category, the lower the diversity. And the more even the distribution is between categories – the greater the diversity.

Diversity merely assumes a distinctiveness between the different categories, and so it requires only a nominal level of measurement. In turn, the **variation** is the type of dispersion that takes into account the size of the differences between the different categories (variable values). This involves the assumption that these differences are meaningful. With this approach, one is able to consider not only the concentration exactly in a certain category but also around a certain value. Concentration becomes gradable – any value may be closer to, or further away from, a typical value. Variation is low in a population, where most cases have an assigned value close to the center. As the assigned values move away from the center, the variation rises.

1.2. The concept of "ordinal variation"

The above two concepts: of diversity and variation – are not adequate to describe the distribution of an ordinal variable. The second of them violates the assumptions of measurement – the difference between the values of an ordinal variable is not interpretable. The first one is allowed from the measurement point of view, but does not take into account any information about order.

Using a diversity parameter to describe an ordinal variable could bring counterintuitive results. In particular, the value of these parameters would be the same for all the distributions which have the same set of frequencies. Meanwhile, the order in which the frequencies are assigned to categories should be taken into account. There still is a center of distribution of an ordinal variable, and for many reasons the median may be considered this natural center. The first intuition is thus that the greater the share of the median category the lower the dispersion of an ordinal variable. This type of dispersion will be called the "**ordinal variation**". Robert Leik[1] called the opposite of this variation the "consensus" and hence, the parameters of consensus differ from the parameters of ordinal variation only in their direction, which is opposite.

It is also not difficult to identify the natural extremes of this ordinal variation; that is, when it is the lowest, and when it is the highest. Minimal variation naturally occurs when, as in the case of all other concepts of dispersion, all the cases belong to one category. For such distribution, the corresponding parameter of ordinal variation should therefore take the zero value.

The other extreme, the maximum ordinal variation, is somewhat more difficult to identify. A simplifying assumption is needed; that is, the assumption that there exist minimum and maximum values (categories) that are impassable extremes. Then the maximum variation corresponds to the most "polarized" distribution – one in which half of the population is concentrated in the lowest category, and the other half in the top category. Several authors[2] discussing ordinal variation have agreed on this. Such a distribution will be consequently called a "two-point extreme distribution". Without the above assumption about extreme categories there is no maximum ordinal variation. This is actually just like with conventional variation – its indices do not have maximal values.

These insights lead to a fairly unequivocal concept of diversity and to consistent expectations from possible parameters associated with it. We should expect that this diversity will be the greater, the more the distribution will deviate from that which is concentrated in one (median) category towards the extreme two-point distribution. The task now is therefore to find a parameter that represents the described property.

2. Descriptive parameters for ordinal variables

It is not uncommon that different statistical measures for ordinal variables are proposed as ad hoc ideas. In other words, these measures are not derived from a theoretical model nor axiomatically, but are rationalized with a superficial compound with certain intuitions only. Rather than looking for any formula that seems to give results consistent with certain intuitions, it is more desirable to derive a parameter from independent theoretical criteria. In other words, only a theoretical model of the statistical description should generate the parameters and methods. In this section two types of criteria will be considered: loss functions and distance function. Both lead us to derivation of some dispersion measures. The next section will present a different approach to the parameter derivation – it will show the consequences of the adoption of certain axioms that can be postulated for a measure of ordinal variation.

2.1. Loss function for the ordinal scale

2.1.1. Optimal description model

Many of the widely used descriptive parameters (both location and dispersion parameters) can be interpreted in terms of the optimal description. However, this has not been satisfactorily shown for methods designed specifically for ordinal variables. The reason for this is the lack of a universally accepted loss function for the ordinal scale.

The optimal description model can be viewed as a decision theory model. The problem of an optimal description of the distribution of one variable can be represented as follows: we are looking for such a value x_a, which in some sense can be considered "typical" for the whole population, because we want to describe it by a single value. It should vary as little as possible – according to some criteria – from the values assigned to all cases. This criterion is a loss function defined on a pair of values: the predicted value and the actual value, formally:

$$l(x_a, X(\omega_i))$$

Since we can compute the value of this function for any case in the population, the loss function itself is also a variable defined in the same population. And it has its own distribution. The ultimate criterion that decides what should be the value x_a is a parameter (usually the arithmetic mean) of the distribution of a loss function. The X_a value will therefore be referred to as the optimal description (or prediction) for the variable distribution relative to the loss function adopted. It will also determine a location parameter, the "typical" value in the sense given by a specific loss function. In turn, the mean loss function indicates how much all assigned values differ (in the sense defined by a loss function) from the value that is considered typical. It is thus a measure of dispersion.

Three classic loss functions: binary, modular, and squared, lead accordingly to the pair of location parameter-dispersion parameter: modal-modal error, median-average deviation from the median, and arithmetic mean-variance. Of these three loss functions the only one applicable for an ordinal scale is the binary function. It is still inadequate, however, because it ignores the information about the order of categories. Therefore there is a need for another loss function that suits the nature of ordinal variables.

2.1.2. Number of categories as a loss function

In the context of a set of ratings Grzegorz Lissowski[3] proposed the adoption of a number of categories between the predicted category and the prediction as a loss function:

$$l(x_j, x_a) = |j - a|.$$

Mean loss function is minimized when the median category is used as a description. This mean loss function when an optimal description is used can be considered as a measure of ordinal variation $Z(X)$:

$$Z(X) = \frac{m}{2} - \sum_{i=1}^{m} |F(x_i) - 0{,}5|,$$

where $F(x_i) = P(X \le x_i)$, and m stands for number of categories.

Some specific properties of this measure will be analyzed later. One can easily note, however, that it meets the basic expectations of a measure of ordinal variation: it takes the value "0" for the distribution concentrated in one category and the maximum value for the extreme two-point distribution.

The measure $Z(X)$ turns out to be closely linked to the measure "D" introduced by Robert Leik[4] without such justification with the model of optimal description:

$$D = \frac{2\sum_{i=1}^{m} d_i}{m - 1},$$

where $d_i = \min\{F(x_i), 1 - F(x_i)\}$; m – number of categories.

To demonstrate this relationship it should be first noted that the value d_i can also be presented in the following form:

$$d_i = \min\{F(x_i), 1 - F(x_i)\} = 0{,}5 - |F(x_i) - 0{,}5|.$$

Therefore, the sum of values d_i, which is present in the numerator of measure D, equals:

$$\sum_{i=1}^{m} d_i = \sum_{i=1}^{m} (0{,}5 - |F(x_i) - 0{,}5|) = \sum_{i=1}^{m} 0{,}5 - \sum_{i=1}^{m} |F(x_i) - 0{,}5|$$
$$= \frac{m}{2} - \sum_{i=1}^{m} |F(x_i) - 0{,}5|,$$

which equals the measure $Z(X)$. Further conversion of this sum was explained by Robert Leik with the necessity of standardization to the interval $[0,1]$ in order to enable comparability for different numbers of categories. The maximum value of the sum (and, consequently, also of measure $Z(X)$) is dependent on m and equals $(m-1)/2$. Dividing the sum by this value gives the final form of the measure D. It is identical to the measure $WZ(X)$, introduced by Grzegorz Lissowski as a relative measure of ratings variation.

However, one can argue that the loss function "number of categories" violates strict ordinal assumptions. Since we do not know the distance between categories, we have no reason to conclude, for example, that the 3 categories above is a greater error than 1 category below. Also, even if errors are in the same direction, we only know that with more categories there is greater error, but we do not know how many times greater.

One can also show the link between the adoption of such a loss function and the descriptive statistical method reserved for stronger levels of measurement. Note that for each ordinal variable, there is a monotonic transformation of its values, that all the distances between successive categories equal 1. After such a transformation this loss function is equivalent to the modular function. The mean loss function (and thus the value of the proposed measure of dispersion) is then equal to the average deviation from the median. From this it follows that the adoption of this loss function is equivalent to the adoption of an assumption about equal distances (all equal 1) between categories and using the modular loss function.

2.2. Functions of distance between distributions

To compare the distributions of two variables or the same variable in two different populations we often use location, dispersion or other parameters describing certain distribution characteristics. Comparison of two distributions can also be performed by calculating a certain distance between them. This way you can answer the question: to what extent are these distributions similar? Using the same procedure, you can also determine the distance of a given distribution from some comparison distribution. If we assume a distribution corresponds to a certain trait, the distance from this distribution may be treated as a parameter of the said characteristic. If we have a natural comparison distribution corresponding to the maximum or minimum dispersion, the distance to it will attest to the dispersion of the given distribution.

2.2.1. Distances between cumulative distribution functions

A function of distance as a base to build a measure of dispersion was used openly and directly by Julian Blair and Michael Lacy in the article "Statistics of Ordinal Variation". They assume that the ordinal variable has a fixed, specified number of categories. They note then, that the distribution of such a variable is equivalent to the vector of cumulative frequencies $F(x_i)$. With the assumption of a fixed number of categories it is established that the number of the coordinates of such a vector is constant. All possible distributions can thus be represented as points in a space of as many dimensions as the number of categories. Therefore some distance function can be used to calculate the distance between any two distributions.

The two authors suggest the distance from the extreme two-point distribution as a measure of dispersion. They agree with the proposition that such a distribution represents the greatest dispersion. They also recognize that the distribution concentrated in one category is naturally the least dispersed. However, calculating distances from the least dispersed distribution would be problematic as there are as many of them as the number of possible categories. Therefore it would not be clear, from which one we should calculate the distance. Thus, the ordinal variation in this approach is defined as the distance of the given vector $F(x_i)$ from the vector $\langle 0.5, 0.5, \ldots, 0.5, 1\rangle$.

In this theoretical framework, there are many possible criteria leading to different results. Similarly to the optimal description model, where different loss functions can be selected – here different distance functions may be selected. The authors mainly propose the use of the Euclidean distance. Then the distance of a given vector $F(x_i)$ from the vector $\langle 0.5, 0.5, \ldots, 1\rangle$ is (due to the fact that the last coordinate is always 1 it may be omitted):

$$d^2 = \sum_{i=1}^{m-1} (F(x_i) - 0{,}5)^2$$

However, this (and any other thus constructed) is a parameter which is directed in an opposite way to what we expect from a parameter of ordinal variation. Using the terminology of Robert Leik, it is a parameter of consensus. Its value is zero for extreme two-point distribution (representing the greatest dispersion and the lowest consensus). A simple correction by subtracting the d^2 from its maximum possible value results in a parameter with a correct polarity, the parameter of ordinal variation:

$$(m-1)/4 - d^2$$

This measure also has an interpretation in terms of $m-1$ dummy variables representing the ordinal variable X. Let's take a series of $m-1$

binary variables X_i such that $X_i = 1$ if and only if $X \leq x_i$. Thus, each of these variables has the following distribution $P(X_i = 1) = F(x_i)$; $P(X = 0) = 1 - F(x_i)$. The variance of each individual dummy variable is $D^2(X) = F(x)(1 - F(x_i))$. As shown by Blair and Lacy[5], the sum of the variances turns out to be equal to the above ordinal variation measure:

$$BL = \sum_{i=1}^{m} F(x_i)(1 - F(x_i)) = (m - 1)/4 - d^2.$$

It turns out this measure proves to be closely related to a measure introduced by Berry and Mielke[6]:

$$IOV = T/T_{max},$$

where $T = \sum_{i<j} N(x = x_i)N(X = x_j)(j - i)$, $1 \leq i < j \leq m$ and T_{max} is the maximum value of T achievable for a given number of N.

The IOV measure differs from measure BL only with normalization – it is a relative measure, taking values from the range $[0, 1]$. Otherwise, it has the same properties as the parameter Blair and Lacy obtained by using the Euclidean distance from extreme two-point distribution.

Also interesting is the application of the city block distance function instead of the Euclidean distance. Then the distance of a given distribution from extreme two-point distribution is as follows:

$$\sum_{i=1}^{m-1} |F(x_i) - 0{,}5|.$$

Since we know that the maximum distance in this case is $(m - 1)/2$, analogous polarity correction can be used. Subtraction from the maximum value results in the following parameter of ordinal variation:

$$(m - 1)/2 - \sum_{i=1}^{m-1} |F(x_i) - 0{,}5|.$$

Notice that $F(x_m) = 1$, therefore $|F(x_m) - 0{,}5| = 0{,}5$. So the above formula can be transformed to:

$$(m - 1)/2 - \sum_{i=1}^{m} |F(x_i) - 0{,}5| - 0{,}5,$$

which in turn reduces to:

$$m/2 - \sum_{i=1}^{m} |F(x_i) - 0{,}5|,$$

which is identical to the previously defined measure $Z(X)$. As it turns out, this parameter is rationalized not only with the model of optimal description, but also as a closeness to extreme two-point distribution.

2.3. Comparison of measures of ordinal variation

The performance in practice of all the mentioned measures is illustrated by the following table. The table shows their values in some particularly interesting cases. All of them equal zero for distribution concentrated in one category and their maximal values for extreme two-point distribution. Measures D and IOV are, accordingly, normalizations of Z and BL so they take values from the interval $[0, 1]$.

Example 1

Distribution	distribution 1	distribution 2	distribution 3	distribution 4	distribution 5
Type of distribution	extreme two-point	uniform	binomial with $p = 0{,}5$	binomial $p = 0{,}8$	one-point
$P(X = 0)$	0.5	0.25	0.125	0.008	0
$P(X = 1)$	0	0.25	0.375	0.096	1
$P(X = 2)$	0	0.25	0.375	0.384	0
$P(X = 3)$	0.5	0.25	0.125	0.512	0
Z	1.5	1	0.75	0.6	0
D	1	2/3	0.5	0.4	0
BL	0.75	0.625	0.4688	0.351	0
IOV	1	5/6	0.625	0.468	0

3. Axiomatic determination of a class of measures

In the case of nominal variables – where dispersion can be interpreted only as diversity – thanks to the formulation of some natural demands we get a measure of diversity with good properties. That is the entropy[7]. This result is so valuable that it inspires to analogous search for a parameter of ordinal variation. The following section will be an attempt to formalize natural requirements from a measure of ordinal variation. Then it will be found that a set of postulates generates not one measure, but a certain class of parameters.

3.1. Formulation of the problem

In order to be able to formulate mutually non-exclusive postulates it is necessary to adopt certain assumptions, which are to some extent arbitrary, and may not be natural in every situation. First of all, it is inevitable to limit the perspective to the case of a fixed number of categories of the variable. Variables with a different number of categories are not directly comparable when it comes to ordinal variation. A fixed number of categories also involves the possibility of empty categories (with zero frequency). Even if no case belongs to some category, this category is not excluded and the number of categories remains the same.

Another difficulty arises from the fact that it will be needed to pick one location parameter as a natural center of distribution. However, there seems to be an agreement that the median is a good and natural location parameter for all ordinal variables. The set of axioms will therefore assume the naturalness of the median.

The assumption of a fixed m number of categories leads to the conclusion that the cumulated distribution vector (fixed length m) contains all relevant information. Thus in general the problem is to find such measure ZP, which is the function of the cumulative distribution vector. Such defined ZP is automatically independent of the strictly increasing monotonic transformations of the variable, since such a transformation does not change the distribution vector. This automatically ensures the applicability of the measure for ordinal variables and makes additional postulates involving the scale of measurement unnecessary.

When we take into account all the natural intuitions concerning ordinal variation we can formulate the following desirable properties of a measure:

- Independence of the values' direction:

The ordinal variables contain information on the order of objects, while the direction in which the values increase is essentially arbitrary. For example – the same information would hold on a school grade scale of 1 to 6, where 1 would be the worst grade and a reversed scale: from 6 to 1, where the worst grade would be 6. The measure of ordinal variation should be resistant to this arbitrariness. This property therefore implies a stronger requirement for a parameter than just independence from the increasing transformation – it should be independent of any strictly monotonic transformation.

- The minimum value (zero) should be taken for the one-point distribution:

Concentration of distribution in a single category in a natural way corresponds to the total absence of dispersion. The measure should then indicate the lack of dispersion by taking the zero value.

- The maximum value for the extreme two-point distribution:

Another natural end of the continuum is the extreme dispersion at the extreme two-point distribution, so that for this distribution the measure should take its maximum value.

- ZP function should be continuous:

Small changes in the distribution of the variable (and thus – in distribution vector) should not lead to abrupt changes in the value of measure.

- Transfer of cases to the more extreme category should result in an increased value of measure of ordinal variation:

If you move a certain number of objects from one category into the next, which is closer to the extreme category, we arrive at a more dispersed distribution as it will be closer to the extreme two-point distribution. Such a transfer in the bottom half of the distribution (the half of population with the lowest values of the variable) is in the direction of the lowest category. Whereas in the top half it is in the direction of the highest category. Additionally, the transfer cannot cross the border between the lower and the upper half, so it does not change the median.

- Reduction in the value of measure ZP after such a transfer should be the greater the larger the transfer (the greater the number of transferred cases).

3.2. Formalization of postulates

Let us introduce the following denotations:

$F_x = \langle F(x_1), F(x_2), \ldots, F(x_m)\rangle$ – vector of values of the cumulative distribution function for all the m variable categories;

$\mathbb{F}$ – set of all possible vectors F_x

Therefore, the target measure ZP is formally a function: $ZP : \mathbb{F} \to \mathbb{R}$, so it assigns real numbers to cumulative distribution vectors.

Previously formulated postulates can be formalized as follows:

A1. Independence of the values' direction.

$$ZP(\langle F(x_1), F(x_2), \ldots, F(x_m)\rangle) \\ = ZP(\langle 1 - F(x_{m-1}), 1 - F(x_{m-2}), \ldots, 1 - F(x_1), 1\rangle)$$

Note: The argument of ZP on the right side of the equation is a cumulative distribution vector we get from the "mirror image" of the distribution introduced on the left side of the equation.

A2. Minimal value for one-point distribution.

$$\forall F_x \in \mathbb{F} : \forall i : F(x_i) = 0 \vee F(x_i) = 1 \rightarrow ZP(F_x) = 0$$

Note: the implication predecessor describes the one-point distribution – for such distribution the cumulative frequencies are always "0" or "1" (the "1" values start from the category to which all cases belong, any previous vector values are "0 ").

A3. Maximal value for extreme two-point distribution.

$$\forall F_x \in \mathbb{F} : ZP(F_x) \leq ZP(\langle 0.5, 0.5, \ldots, 0.5, 1\rangle)$$

Note: the vector $\langle 0.5, 0.5, \ldots, 0.5, 1\rangle$ represents extreme two-point distribution. Probability at the first category equals 0.5; all middle categories are empty, therefore the cumulative distribution does not change until the last category, which also has the probability of 0.5.

A4. Transfer towards extreme category.

$$ZP(\langle F(x_1), \ldots, F(x_j){+}\varepsilon, F(x_{j+1}), \ldots, F(x_{m-1}), 1\rangle){-}ZP(F_x){=}t(\varepsilon, F(x_j)),$$

where $\varepsilon \in [0, F(x_{j+1}) - F(x_j)]$, t is a continuous function increasing with ε, $t(0, F(x_j)) = 0$, $F(x_j) + \varepsilon \leq 0.5$.

Note: The axiom describes a transfer of some cases from a category closer to the center (x_{j+1}) to her nearest category (x_j), which is closer to the extreme category. Because of the property $F(x_j) + \varepsilon \leq 0.5$ the described transfer can only occur in the lower half of distribution (behaviour of the measure after a transfer in the upper half will be determined with the addition of axiom A1, which will be explained later). Such a transfer is valued positively; that is, the value of ZP after the transfer should rise. The rise can depend on, besides the size of the transfer, the cumulative distribution $F(x_j)$, so on the "place" in distribution where the transfer occurs. The continuity of function t ensures, that small changes in distribution will not result in sudden rises in ZP.

3.3. Observations on relations between axioms

With the above formalizations of postulates, it is possible to make a few observations that allow for the simplification of the set of axioms and assist in obtaining the final form of the measure of ordinal variation.

Observation 1. For formal simplicity, Axiom A4 involves only transfers in the lower part of distribution. The axiom A1 complements the A4 as for transfers in the upper half of the distribution. If the parameter is to have the same value for reversed distributions (one distribution being a mirror reflection of the other), the change in its value must be carried out similarly also in the case of transfers in the upper half of the distribution. According to A1 we have: $ZP(F_x) = ZP(\langle 1-F(x_{m-1}), 1-F(x_{m-2}), \ldots, 1-F(x_1), 1\rangle)$. With inserting the vector $\langle 1-F(x_{m-1}), 1-F(x_{m-2}), \ldots, 1-F(x_1), 1\rangle$ in place of F_x in A4 we arrive at:

$$\begin{aligned} &ZP(\langle 1-F(x_{m-1}), 1-F(x_{m-2}), \ldots, 1-F(x_{j+1}), 1-(F(x_j)+\varepsilon), \ldots, \\ &\quad 1-F(x_1), 1\rangle) - ZP(\langle 1-F(x_{m-1}), 1-F(x_{m-2}), \ldots, 1-F(x_1), 1\rangle) \\ &= t(\varepsilon, 1-F(x_j)), \end{aligned}$$

where $\varepsilon \in [0, F(x_{j+1}) - F(x_j)]$, t is a continuous non-negative function increasing with ε, $t(0, 1-F(x_j)) = 0$, $1-(F(x_j)+\varepsilon) \leq 0.5$.

Therefore it is apparent, that after reversing the situation given in axiom A5 the transfer towards the extreme category is a transfer to the higher category. This is so due to the fact that cumulative frequency $1-F(x_j)$ from before the transfer turns into $1-(F(x_j)+\varepsilon) = 1-F(x_j)-\varepsilon$, so it is smaller by ε. In both cases (transfers in the top or bottom half of distribution) the transfer cannot cross the middle of distribution. That is, it is not allowed to transfer a case from the top half to bottom half or the other way around. In the case of the upper half, it is assured by the condition $F(x_j) - \varepsilon \geq 0.5$.

In order to analogically value similar transfers in the upper half of the distribution the following condition must be fulfilled:

$$t(\varepsilon, F(x_j)) = t(\varepsilon, 1-F(x_j))$$

Since the function t has the same value for the argument of $F(x_j)$ and for $1-F(x_j)$, i.e., therefore its value depends only on the module of the difference: $|F(x_j)-0.5|$. The value of transfer must therefore be equal to the value of a function $v(\varepsilon, |F(x_j)-0.5|)$, which is a non-negative continuous function increasing with ε, wherein $v(0, |F(x_j)-0.5|) = 0$.

Observation 2. The minimum value for the one-point distribution is a consequence of the transfer postulate (A4, supplemented by A1). Take any distribution, which is not one-point. Using reverse transfers rather than transfers towards the extremes can lead to one-point distribution, always reducing (in accordance with A4 supplemented with A1) with each transfer the value of ZP. However, any transfer made when the starting point is a one-point distribution can only increase this value. Of course the A4 postulate does not imply that the minimum is zero (as proclaimed by A2) – this value is just a matter of normalization.

Observation 3. The maximum value for the extreme two-point distribution (A3) results from the A4. Take any distribution other than extreme two-point distribution. Then always at least one cumulative frequency is different from 0.5. A transfer towards the extreme categories can therefore always be found, which according to A4 (supplemented with A1) will increase the value of the parameter ZP. If after such a transfer, distribution is still not extreme two-point, then there is a possibility of a subsequent transfer. By a number of such transfers you can achieve the extreme two-point distribution, increasing the value of the measure in every step. Only for extreme two-point distribution are all cumulative frequencies (except for the last category, which is always equal to 1) equal to 0.5. It is then impossible to make a further transfer as specified in the A4. Thus, the ZP reaches the maximum value at the extreme two-point distribution.

For instance, take frequency distribution $\langle 0.1, 0.2, 0.3, 0.4\rangle$, for which the vector of cumulated frequencies is $F_x = \langle 0.1, 0.3, 0.6, 1\rangle$. In this it could be a series of the following transfers:

1) Since $0.6-0.1 \geq 0.5$, it is possible to transfer 0.1 from the third category to the last category. We arrive at frequencies $\langle 0.1, 0.2, 0.2, 0.5\rangle$ and cumulated frequencies $\langle 0.1, 0.3, 0.5, 1\rangle$. Cumulated frequency in the third category became 0.5, so further transfer (such that the value of ZP is increased as stated in A4) from this category to the higher one is not possible.
2) Since $0.3 + 0.2 \leq 0.5$, the transfer of 0.2 from the third to the second category is possible. We get frequency distribution $\langle 0.1, 0.4, 0, 0.5\rangle$ and cumulated frequencies $\langle 0.1, 0.5, 0.5, 1\rangle$. Now only a transfer from the second to the first category remains feasible.
3) The transfer of 0.4 from the second to the first category results in frequency distribution $\langle 0.5, 0, 0, 0.5\rangle$ and cumulated frequencies $\langle 0.5, 0.5, 0.5, 1\rangle$, that represent the extreme two-point distribution. No further transfers that would increase ZP are possible.

Observation 4. The extreme two-point distribution is the only distribution that can be achieved through a series of transfers (that would increase ordinal variation according to A4) from any other distribution. For any other distribution, it is mutually exclusive that you could achieve it through transfers from one-point distribution in the first category and from one-point distribution in the last category. In one case of these starting points there would always be the necessity to take more than half of the population from the extreme category. Such a transfer is median-changing and is not a transfer described in A4 (it violates the requirement $F(x_j)+\varepsilon \leq 0.5$). You can achieve the extreme two-point distribution even through transfers from one-point distributions from both extreme categories. This would require transferring half the population at a time.

These observations reduce the set of significant axioms. The axiom of maximum A3 is not needed, whereas the role of A2 is just normalization, as it sets the minimal value of ZP to 0.

3.4. The Result

The set of axioms determines a class of measures of ordinal variation.

Theorem

In order to fulfill properties A1–A4 the function ZP must form:

$$(m-1)w(0.5) - \left(\sum_{i=1}^{m-1} w(|F(x_i) - 0.5|)\right),$$

where $w(\cdot)$ is a continuous non-negative increasing function and $w(0) = 0$.

Proof

Observation 1 entails that if we can transform vector F_{X^1} into vector F_{X^2} with a transfer towards an extreme category, then $ZP(F_{X^1}) - ZP(F_{X^2}) = v(\varepsilon, |F(x)j) - 0.5|)$, where x_j is the category, for which the cumulative frequency does change.

Let us introduce another observation that is crucial for the proof.

Observation 5. Note, that if we consider a transfer such as in A4, there is only one change after it in the cumulative frequencies vector and that change equals the quantity of the transfer. Therefore, the transferred quantity ε (that allows us to transform vector F_{X^1} into F_{X^2}) is:

$$\varepsilon = F^1(x_j) - F^2(x_j), \text{ when } F^1(x_j) \leq 0.5$$

$$\varepsilon = F^2(x_j) - F^1(x_j), \text{ when } F^2(x_j) \geq 0.5,$$

so in general:

$$\varepsilon = |F^1(x_j) - F^2(x_j)|$$

If there will be consequential transfers of the same type between other categories, then certain coordinates of vectors will change accordingly by transfer quantities. So the coordinates of the vector of differences $\langle |F^1(x_i) - F^2(x_i)| \rangle$ (where $i = 1, 2, \ldots, m$) will equal quantities of the performed transfers. Any given vector of differences may be interpreted as a result of a series of transfers of quantities, which equal the vector coordinates.

In accordance with A4, a single transfer should result in a change in ZP by value $v(\varepsilon, |F(x_j) - 0.5|)$. Every consequent transfer causes another analogous change in ZP. So if we can transform vector F_{X^2} into vector F_{X^1} through a series of transfers towards an extreme category then:

$$ZP(F_{X^1}) - ZP(F_{X^2}) = \sum_{i=1}^{m-1} v(|F^1(x_i) - F^2(x_i))|, |F^2(x_i) - 0.5|).$$

From A4 complemented with A1 and from Observation 3 we have that for all F_X (for all distributions): $ZP(F_X) \leq ZP(\langle 0.5, 0.5, \ldots, 0.5, 1\rangle)$, whereas vector F_X can be achieved from vector $\langle 0.5, 0.5, \ldots, 1\rangle$ through transfers directed from extreme categories. In addition, vector $\langle 0.5, 0.5, \ldots, 1\rangle$ is the only one that can be transformed into any other vector through a series of such consequent transfers (not crossing the border between the top and bottom half of the population). Therefore only a comparison of coordinates of a given vector F_x and vector $\langle 0.5, 0.5, \ldots, 0.5, 1\rangle$ guarantees that differences in coordinates are interpretable as quantities of individual transfers. In accordance with Observation 5, consequent modules of differences between coordinates of a given vector and vector $\langle 0.5, 0.5, \ldots, 0.5, 1\rangle$ correspond to a series of transfers that lead to extreme two-point distribution. These modules equal $|F(x_i) - 0.5|$, except for the last category, where the cumulated frequency is always 1, so there is no difference between coordinates.

For instance, take the distribution already mentioned in Observation 3: $\langle 0.1, 0.2, 0.3, 0.4\rangle$. For this distribution the vector F_x is $\langle 0.1, 0.3, 0.6, 1\rangle$. The differences between its coordinates and those of extreme two-point distribution are: $|0.1 - 0.5| = 0.4$, $|0.3 - 0.5| = 0.2$, $|0.6 - 0.5| = 0.1$, $1 - 1 = 0$ and correspond with transfers that lead to extreme two-point distribution, as described previously in Observation 3.

Therefore, in accordance with Observation 5, the vector of differences $\langle |F(x_i) - 0.5| \rangle$ is a vector of transfers that are needed to transform F_X into vector $\langle 0.5, 0.5, \ldots, 1\rangle$. The difference between the value of the measure ZP for extreme two-point distribution and for a given distribution is:

$$ZP(\langle 0.5, 0.5, \ldots, 0.5, 1\rangle) - ZP(F_X) = \sum_{i=1}^{m-1} v(|F(x_i) - 0.5|, |F(x_i) - 0.5|)$$

It is apparent, that in this case the function v is a function of two identical arguments. Therefore it has to be equivalent to some one-argument function $w(|F(x_i) - 0.5|)$, which is continuous, non-negative, and increasing. So the above difference can also be presented as:

$$ZP(\langle 0.5, 0.5, \ldots, 1\rangle) - ZP(F_X) = \sum_{i=1}^{m-1} w(|F(x_i) - 0.5|) \qquad (*)$$

If the given distribution is one-point, then the vector F_X is $\langle 0, 0, \ldots, 0, 1, \ldots, 1, 1\rangle$. Since all cumulative frequencies are 0 or 1, in this case the sum takes its maximum value (function w is increasing), which is:

$$\sum_{i=1}^{m-1} w(0.5) = (m-1)w(0.5)$$

In accordance with A3, the value of the measure for one-point distribution has to be 0. Therefore equation (*) in the case of one-point distribution takes the form:

$$ZP(\langle 0.5, 0.5, \ldots, 0.5, 1\rangle) = (m-1)w(0.5),$$

which determines the maximum value of ZP, which has to be its value for extreme two-point distribution.

Back again to equation (*), if we leave sole $ZP(F_X)$ on the left side, we arrive at:

$$ZP(F_X) = ZP(\langle 0.5, 0.5, \ldots, 0.5, 1\rangle) - \sum_{i=1}^{m-1} w(|F(x_i) - 0.5|).$$

If we now insert the known value of ZP for extreme two-point distribution, we finally get:

$$ZP(F_X) = (m-1)w(0.5) - \sum_{i=1}^{m-1} w(|F(x_i) - 0.5|).$$

The proof has to be now supplemented with showing that the measure of this form always fulfills the conditions A1–A4.

Axiom A3 is always fulfilled for this form of function ZP, because the sum has its minimal value for extreme two-point distribution. This value is zero, since all differences $|F(x_i) - 0.5|$ are zero and $w(0)$ and w cannot take negative values.

Axiom A1 is fulfilled, because the set of differences $|F(x_i) - 0.5|$ (where $i = 1, 2, \ldots, m$) is the same for reversed distributions. For any given vector F_x this set of differences is: $|F(x_1) - 0.5|, |F(x_2) - 0.5|, \ldots, |F(x_{m-1}) - 0.5|$. Then for F'_x, which is a "mirror reflection" of F_x it is: $|1 - F(x_{m-1}) - 0.5|, |1 - F(x_{m-2}) - 0.5|, \ldots, |1 - F(x_1) - 0.5|$. That can be rewritten as: $|0.5 - F(x_{m-1})|, |0.5 - F(x_{m-2})|, \ldots, |0.5 - F(x_1)|$, so it is the same set as that corresponding to F_x (except for the reversed order). The value of measure ZP depends only on this set of differences (not their order) and function $w(\cdot)$.

For the one-point distribution all of the differences $F(x_i) - 0.5$ are 0.5. Since cumulated frequency is always in $[0, 1]$, the value "0.5" is the largest possible argument of function $w(\cdot)$. Therefore the sum is at a maximum for one-point distribution. That also means the value of measure ZP is minimal. That means the axiom A2 is fulfilled.

In the case of A4 we have to consider two cases of a transfer towards the extreme category: the one in the lower half of the population which is towards the bottom category and the one in the top half which is towards the highest category.

- Transfers in the lower half of the population:

Distributions that are one transfer apart will differ in just one coordinate of cumulated frequencies vector. After the transfer one value becomes $F(x_j)+\varepsilon$ instead of $F(x_j)$. The sums before and after the transfer will differ in just one value: after the transfer it is $w(|F(x_j) + \varepsilon - 0.5|$ instead of $w(|F(x_j) - 0.5|)$. Therefore the difference between the value of ZP before and after the transfer will equal $w(|F(x_j) - 0.5|) - w(|F(x_j) + \varepsilon - 0.5|)$. As $F(x_j) + \varepsilon \leq 0.5$ the above difference is a non-negative function increasing with ε. That means A4 is fulfilled.

- Transfers in the upper half:

Distributions that are one transfer apart will also differ in just one coordinate of cumulated frequencies vector. After the transfer one value becomes $F(x_j) - \varepsilon$ instead of $F(x_j)$. The sums before and after the transfer will differ in just one value: after the transfer it is $w(|F(x_j) - \varepsilon - 0.5|$ instead of $w(|F(x_j) - 0.5|)$. Therefore the difference between the value of ZP before and after the transfer will equal $w(|F(x_j) - 0.5|) - w(|F(x_j) - \varepsilon - 0.5|)$. As this time we have $F(x_j)-\varepsilon \geq 0.5$, the above difference is also a non-negative function increasing with ε. That means A4 is fulfilled also for the upper half of the population.

There are an infinite number of possible ZP functions that fulfill all the conditions. They differ in function $w(\cdot)$, so in an evaluation of transfers. Depending on whether this function rises faster or slower with increasing arguments, the transfers close to the extremes or to the center of distribution will be valued differently. Note that the value of measure depends on the module of difference: $|F(x_j) - 0.5|$. The cumulative frequency at the very center of an ordered population would be 0.5, and towards the extremes the difference from 0.5 becomes bigger, so the module can be interpreted as a "distance" to the center. If the $w(\cdot)$ function is convex it will overrate transfers near the extremes and underrate that close to the center. If it is concave – then just the contrary. This "distance" will not matter, if $w(\cdot)$ is linear.

3.5. Further normalization

On all of the ordinal variation measures you can put one further constraint. So far their values only have the lower limit (zero). However, depending on the given number of categories, their maximum values will differ. To introduce some degree of comparability for different populations and variables it is required to normalize the measure to a constant range of possible values. For interpretational simplicity usually the normalization is a linear transformation of values so that they have a range $[0, 1]$. After such a transformation, every value can be interpreted as a percent of the maximum possible ordinal variation. Measures "D" and "IOV", introduced in the previous section, are normalized this way.

3.6. Examples of measures complying with the axioms

The measure $Z(X) = m/2 - \sum_{i=1}^{m} |F(x_i) - 0.5|$, mentioned earlier, does comply with the axioms. It was introduced as a mean loss function, but notice it can be rewritten in the following form:

$$(m-1)/2 - \sum_{i=1}^{m} |F(x_i) - 0.5|,$$

so it is clear it matches the general formula, determined by the axioms. The function $w(\cdot)$ is the identity function in this case. Transfers are thus valued proportionally to the quantity of transfer. It also means that Leik's measure D and Lissowski's WZ (which are in fact the same) also comply with all the postulates. They are both Z's normalization into range $[0, 1]$.

Another example would also be Blair and Lacy's measure d^2, but only after a correction of values' direction, that is:

$$BL = (m - 1)/4 - d^2,$$

so

$$BL = (m - 1)/4 - \sum_{i=1}^{m-1} (F(x_i) - 0.5)^2.$$

This version of their measure has a minimum for one-point distribution and a maximum for extreme two-point distribution. It is apparent that the $w(\cdot)$ function is here the squared function. After a transfer below the median and towards the center there is one change in the sum. There is a value $(F(x_i) - \varepsilon - 0.5)^2$ instead of $(F(x_i) - 0.5)^2$, which makes a difference of $\varepsilon^2 - 2\varepsilon(F(x_i) - 0.5)$. The above function is (as it should be) increasing with ε and also decreasing with $F(x_i)$. It means that the transfers are valued higher the closer they are to the lowest category. It would be similar in the case of transfers in the upper half. The transfers that are farther away from the center are more important. The same can be said about Berry and Mielke's IOV, which is actually a normalization of BL.

On the other hand, a measure based on distances, one that does not take into account the number of categories, cannot comply with the axiom of transfer. Empty categories cannot be ignored. For example, the following two distributions: $\langle \frac{1}{2}, \frac{1}{2}, 0, 0 \rangle$ and $\langle \frac{1}{2}, 0, 0, \frac{1}{2} \rangle$ cannot be treated as equivalent, even though the structure of relations between cases could be the same. You would need transfers that change the value of variation measure to transform one of these distributions into another. The value of ordinal variation measure is different for these two distributions and it makes sense only with the assumption that the categories are fixed.

In the previous section I recalled some of the previously known measures of ordinal variation. As it turned out, these measures belong to the class of measures determined axiomatically. It means that the axiomatic method gives the existing measures an additional substantiation. But this method also shows that those measures are just examples and there are more sensible parameters to be considered.

NOTES

[1] Robert K. Leik, *A measure of ordinal consensus*, "The Pacific Sociological Review", Vol. 9, 1966, p. 85–90.

[2] See Robert K. Leik, *A measure of ordinal consensus*, "The Pacific Sociological Review", Vol. 9, 1966, p. 85–90, also Kenneth J. Berry, Paul W. Mielke, *Indices of ordinal variation*, "Perceptual and Motor Skills", Vol. 74, 1992, p. 576–578 and Julian Blair, Michael G. Lacy, *Statistics of ordinal variation*, "Sociological Methods and Research", Vol. 28, 2000, p. 251–280.

[3] Grzegorz Lissowski, *Miara zróżnicowania ocen*, "Prakseologia", Nr 141, 2001, p. 45–56.

[4] Robert K. Leik, *A measure of ordinal consensus*.

[5] Julian Blair, Michael G. Lacy, *Statistics of ordinal variation*.

[6] Kenneth J. Berry, Paul W. Mielke, *Indices of ordinal variation*, "Perceptual and Motor Skills", Vol. 74, 1992, s. 576–578.

[7] Janos Aczel, *On different characterizations of entropies*, [in:] *Probability and Information Theory*, Minaketan Behara, Klaus Krickeberg, Jacob Wolfowitz (ed.), New York 1969, p. 1–11 and Grzegorz Lissowski, Jacek Haman, Mikołaj Jasiński, *Podstawy statystyki dla socjologów*, p. 134.