ORDANOVA: Analysis of Ordinal Variation

Gadrich, Tamar
ORT Braude College, Industrial Engineering and Management Department
51 Snunit
KARMIEL (21982), Israel
E-mail: tamarg@braude.ac.il

Bashkansky, Emil
ORT Braude College, Industrial Engineering and Management Department
51 Snunit
KARMIEL (21982), Israel
E-mail: ebashkan@braude.ac.il

Consider an object that is measured using an ordinal scale with *K* ordered categories. The common approach when dealing with ordinal scale values is to convert the ordinal estimation into a numerical one by assigning numerical values to each category of the ordinal variable. This procedure is undesirable because it can lead to misunderstanding and misinterpretation of the measurement results [Blair & Lacy (2000), Franceschini et al. (2005), Bashkansky & Gadrich (2008)]. Since ordinal variable values can only be comparisons of "greater than", "less than", "equal to" and "unequal to", all statistical measures of such ordinal variables must reflect these limitations. The focus of this article is on ORDANOVA (**Or**dinal **d**ata **an**alysis **o**f **va**riation); i.e., the analysis of ordinal data variation. Once we understand how to make this analysis, the knowledge can be utilized for practical engineering applications such as quality/failure classification, uncertainty evaluation, repeatability and reproducibility (R&R) analysis, distinguishing feature identification and so on.

Description of ordinal data

The complete description of the set of n ordinal data based on an ordinal scale with K ordered categories *coded* by integers k =1,2,...,K is given by the set $(n_1, n_2, ..., n_K) \equiv \{n_k\}_{k=1}^{K}$, where n_k is a quantity of data belonging to the k-th category $\left(\sum_{k=1}^{K} n_k = n\right)$. Let $\hat{p}_k = \frac{n_k}{n}$ $(k = 1, 2, ..., K)$, $\left(\sum_{k=1}^{K} \hat{p}_k = 1\right)$ and $\hat{F}_k$ $(k = 1, 2, ..., K)$, $0 \le \hat{F}_1 \le \hat{F}_2 ... \le \hat{F}_K (=1)$ be the proportion and cumulative frequency of data belonging to/(up to) the k-th category. Denote by p_k and F_k the theoretical distributions describing the studied phenomenon. The set $\{p_k\}_{k=1}^{K}$ as well as the set $\{F_k\}_{k=1}^{K}$ provide the full description of random variable distribution, in the same manner as the sets $\{\hat{p}_k\}_{k=1}^{K}$ or $\{\hat{F}_k\}_{k=1}^{K}$ provide the full description of sample distribution. It is well known [Agresti (2002)] that for a random sample of size n, the probability of receiving the set $\{n_k\}_{k=1}^{K}$ is given by the following multinomial distribution:

$$P(n_1, n_2, \ldots, n_K) = \frac{n!}{\prod_{k=1}^{K} n_k!} \prod_{k=1}^{K} \hat{p}_k^{n_k} \tag{1}$$

$$E(\hat{p}_k) = p_k, \quad Var(\hat{p}_k) = \frac{p_k \cdot (1 - p_k)}{n}, \quad Cov(\hat{p}_k, \hat{p}_{k'}) = -\frac{p_k p_{k'}}{n} \tag{2}$$

and also simple calculation shows that $Var(\hat{F}_k) = [F_k \cdot (1 - F_k)]/n$. The last is obvious if we consider all categories from 1 to *k* as one category and from *k+1* up to *K* as another category.

For the aim of the following analysis we will also briefly consider some metrological aspects of ordinal measurement/classification [Bashkansky & Gadrich (2010b)].

Measurement errors resulting from misclassification are described using a „*K* by *K*" stochastic classification matrix $\hat{P}$. Its components, $P_{j|i}$ (where $1 \le i, j \le K$), are the conditional probabilities that an object will be classified as relating to level j, given that its actual/true level is i (clearly, $\sum_{j=1}^{K} P_{j|i} = 1 \quad 1 \le i, j \le K$). For every i, j; $P_{j|i}$ contains all the necessary information about the measurement/classification errors.

The most exact error-free ordinal measurement means that $\hat{P}$ is the unit matrix. For the "blind", most disordered sorting (MDS), when classification is made without any correlation between true and measured categories, all the components of the classification matrix $\hat{P}_{MDS}$ are equal: $P_{j|i} = 1/K \quad 1 \le i, j \le K$.

Choosing an ordinal dispersion measure

In this paper we focus on an ordinal dispersion measure that allows practical engineering applications such as quality/failure classification, uncertainty evaluations, repeatability and reproducibility (R&R) analysis, distinguishing feature identification and so on. In order to fulfill these targets the desirable properties of such a variation measure for a given { p_k } (or equivalently { F_k }) distribution, are as follows:

(1) It equals zero when all the data relate to a single category.
(2) It has a maximal value for the most polarized distribution (50% of data belong to the lowest category and 50% belong to the highest category).
(3) It is reversible, i.e., invariant to category order inversion.
(4) It is similar to a Bernoulli distribution variation, when applied to only two categories.
(5) It is similar to existing measures of inequality (Gini [Yitzhaki (1998)], for example).
(6) The measure can be decomposed to "between" and "within" components.

Based on a literature survey, we assembled a number of ordinal dispersion measures. Our study showed that all the above mentioned properties are best satisfied by Blair and Lacy"s (2000) measure defined for population as in (3) and for sample distribution as in (4):

$$h_{(T)}^2 = \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} F_k (1 - F_k) \tag{3}$$

$$\hat{h}_{(T)}^2 = \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} \hat{F}_k (1 - \hat{F}_k) \tag{4}$$

(the sub-index $T \equiv$ Total means that the measure is calculated based on the all data). These measures are normalized to the $[0 \div 1]$ range.

Applying (3) to every ordinal level i $(i = 1,2,...,K)$ of the classification matrix, one receives K repeatability dispersion (D_i) measures [Bashkansky & Gadrich (2010b)]:

$$D_i = \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} F_{k|i} \left(1 - F_{k|i}\right) \qquad (1 \le i \le K), \tag{5}$$

where $F_{k|i} = \sum_{j=1}^{k} P_{j|i}$ is the cumulative frequency of data belonging up to the k-th category, given that its actual/true level is i. Note: repeatability does not necessarily mean accuracy. For example, for fully opposite classification, repeatability is perfect ($\forall i : D_i = 0$).

ORDANOVA decomposition

Let us consider M ordinal samples of different sizes – n_m $(\sum_{m=1}^{M} n_m = N)$ as a subject for variation analysis. Denote $\pi_m = \frac{n_m}{N}$ $(\sum_{m=1}^{M} \pi_m = 1)$.

Let $\hat{p}_{km} = \frac{n_{km}}{n_m}$ $(k = 1,2,...K; m = 1,...,M)$ and $\hat{F}_{km}$ $(k = 1,2,...K; m = 1,...,M)$ be the proportion and cumulative frequency of data belonging to/(up to) the k-th category in the m-th sample.

Denote by $\hat{p}_{k.}$ the ***total*** proportion of items belonging to the k-th category. It follows that:

$$\hat{p}_{k.} = \frac{1}{N} \sum_{m=1}^{M} n_{km} = \sum_{m=1}^{M} \frac{n_m}{N} \cdot \frac{n_{km}}{n_m} = \sum_{m=1}^{M} \pi_m \cdot \hat{p}_{km} \,, \tag{6}$$

while the ***total*** cumulative frequency of items belonging up to the k-th category $\hat{F}_{k.}$ is equal to

$$\hat{F}_{k.} = \sum_{j=1}^{k} \hat{p}_{j.} = \sum_{m=1}^{M} \pi_m \cdot \hat{F}_{km} \,. \tag{7}$$

Now define the ***total*** variation $\hat{h}^2_{(T)}$ according to (4) as:

$$\hat{h}^2_{(T)} = \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} \hat{F}_{k.} \cdot (1 - \hat{F}_{k.}) \,, \tag{8}$$

and the ***within*** m-th sample variation as:

$$\hat{h}^2_{m\,(W)} = \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} \hat{F}_{km} \cdot (1 - \hat{F}_{km}) \,. \tag{9}$$

The classic variation ***between*** the k-th category sample‟s cumulative frequencies is averaged by:

$$s^2_{k(B)} = \sum_{m=1}^{M} \pi_m \cdot (\hat{F}_{km} - \hat{F}_{k.})^2 \,. \tag{10}$$

The ORDANOVA decomposition allows the total variation $\hat{h}^2_{(T)}$ to be split into the “within” and “between” components as follows:

$$\hat{h}^2_{(T)} = \sum_{m=1}^{M} \pi_m \cdot \hat{h}^2_{m(W)} + \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} s^2_{k(B)} \quad . \tag{11}$$

Due to space limitations, we omit the proof of this decomposition.

Using the ORDANOVA decomposition in engineering applications

Inference about the multinomial proportions from small samples

Assume that some entity (such as quality, for example) is measured on an ordinal scale basis with K levels. M samples of equal size n are drawn at different times or from different places or from different processes and so on. Each sample contains data distributed among K categories. A common question is: whether all samples were drawn from the **same** population characterized by some set of $\{p_k\}$, or from different populations that differ in time, place, and process and so on? Such an approach is typical, for example, in Statistical Process Control (SPC) interested in checking process stability, when comparing process lines or raw material suppliers, as well as in many other engineering applications.

Different approximations for large samples are not valid for small samples, whereas "exact" solutions [Agresti (2002)] are usually intended for the binary case only. Using the above ORDANOVA decomposition (11), some auxiliary insight can be concluded based on the following considerations.

If the assumption H_0 that all samples come from the same population $\{p_k\}$ is true, then:

$$E[\hat{h}^2_{(T)}] = h^2_{(T)}\left(1 - \frac{1}{n \cdot M}\right) = h^2_{(T)}\left(1 - \frac{1}{N}\right) = h^2_{(T)} \cdot \frac{(N-1)}{N} \qquad \left(d.f._{(T)} = N-1\right) \tag{12}$$

$$E\left(S^2_{(B)}\right) = h^2_{(T)} \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{M}\right) = h^2_{(T)} \cdot \frac{(M-1)}{N} \qquad \left(d.f._{(B)} = M-1\right) \tag{13}$$

$$E\left(h^2_{(W)}\right) = h^2_{(T)} \cdot \left(1 - \frac{1}{n}\right) = h^2_{(T)} \cdot \frac{M \cdot (n-1)}{N} \qquad \left(d.f._{(W)} = M \cdot (n-1)\right) \tag{14}$$

$$E\left(S^2_{(B)}\right) / d.f._{(B)} = E\left(h^2_{(W)}\right) / d.f._{(W)} \quad , \tag{15}$$

and therefore, we can roughly estimate in what proportion the total variation is expected to be split among its "between" and "within" components (and this splitting does not depend on the amount of categories!). Once more we omit the proof due the size limitations. Fluctuations in the splitting are, of course, possible; for each case they may be simulated and their likelihoods can be assessed using (1), but it is reasonable to assume that any dramatic change in the expected splitting will make us suspicious of the H_0 hypothesis.

For a ternary scale (K=3) and samples larger than 100, [Bashkansky & Gadrich (2010a)] provided a complete solution based on asymptotical delta method. In the same manner, a solution for more than three categories can be reached. For small samples we propose to base our answers on (12-15) and see (11) as a challenge to the further advancement of the problem‟s research.

Analyzing the accuracy of the ordinal measuring/classification system

Another important area, where a study of the variation is essential, is accuracy analysis of measuring systems (MSs), especially their precision component, such as repeatability and reproducibility (R&R) [Bashkansky et al. (2011)]. It is clear that when the assessed entity is an ordinal quantity with K levels, we need to use the appropriate ordinal measure of dispersion (3,4) to analyze the total variation change after classification.

Let us suppose that N objects (e.g., items produced by some manufacturing line) are submitted to ordinal MS for aim of its metrological analysis. Note by n_k the amount of objects belonging to the k-th ordinal category (item quality level, for example), and by $p_k = n_k / N$ their ***a priori*** proportion of the total amount N (the true values of these proportions are known). After passing through the measuring process objects may be redistributed to the categories due to MS errors. The latter are expressed with the help of a classification matrix, as described in the section titled "Description of ordinal data". In order to evaluate the dispersion after implementing the MS, we applied (3) to q's proportion vector $(q_1, q_2, \ldots, q_K)$ where $q_j = \sum_{i=1}^{K} p_i \cdot P_{j|i}$ denotes the ***a posteriori*** probability that the classified item belongs to the j-th category (obviously, $\sum_{j=1}^{K} q_j = 1$):

$$h^{2(q)}_{(T)} = \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} F_k^{(q)} \left(1 - F_k^{(q)}\right), \tag{16}$$

where $F_k^{(q)} = \sum_{j=1}^{k} q_j = \sum_{j=1}^{k} \left\{ \sum_{i=1}^{K} p_i \cdot P_{j|i} \right\} = \sum_{i=1}^{K} p_i \cdot F_{k|i}$

is the cumulative frequency of data belonging up to the k-th category.

Similar to the ORDANOVA decomposition, this yields:

$$h^{2(q)}_{(T)} = \sum_{i=1}^{K} p_i \cdot D_i + \frac{1}{(K-1)/4} \sum_{k=1}^{K-1} \sum_{i=1}^{K} p_i \cdot \left(F_{k|i} - F_{k.}\right)^2, \tag{17}$$

where

$$F_{k.} = \sum_{i=1}^{K} p_i \cdot F_{k|i}$$

is the cumulative frequency of items belonging up to the k-th category.

After classification the variation may increase, decrease or remain unchanged relative to the incoming variation.

Distinguishing factor identification

Following (15), if the factor under study does not yield to segregation between groups formed according to different levels of the factor, the "between" to "within" variation parts ratio is expected to be no more significant than the "between" to "within" ratio of degrees of freedom:

$$S^2_{(B)} / h^2_{(W)} \approx d.f._{(B)} / d.f._{(W)} \tag{18}$$

It is plausible to assume that the more the LHS of (18) exceeds its RHS, the greater is the segregative effect of the factor studied. With some caution we may even suggest that the LHS/RHS ratio can serve as an indicator of the segregation power (discrimination significance) by this factor. An example of such an approach is the analysis of university course evaluations, based on a simplified, but similar considerations taking into account the ordinal nature of student ratings [Rampichini et al. (2004)]. The satisfaction survey questionnaire items with higher values of variation between courses were considered to be the most reliable indicators of course quality.

REFERNCES

Agresti A, Categorical Data Analysis. (2nd edn), Wiley: New York, 2002.

Blair J, Lacy MG. Statistics of ordinal variation. Sociological Methods & Research 2000; 28: 251-280.

Bashkansky E, Gadrich T. Evaluating quality measured on a ternary ordinal scale. Quality and Reliability Engineering International 2008; 24:957-971.

Bashkansky E, Gadrich T. Statistical Quality Control for Ternary Ordinal Quality Data. Applied stochastic models in business and industry DOI: 10.1002/asmb.868 (On line from December 6, 2010a)

Bashkansky E, Gadrich T. Some Metrological Aspects of Ordinal Measurements. Accreditation and Quality Assurance 2010b; 15(6): 331-336.

Bashkansky E, Gadrich T, Knani D. Some metrological aspects of the comparison between two ordinal measuring systems. Accreditation and Quality Assurance 2011; 16:63-72.

Franceschini F, Galetto M, Varetto M. Ordered samples control charts for ordinal variables. Quality and Reliability Engineering International 2005; 21:177-195. DOI: 10.1002/qre.614

Rampichini C, Grilli L, Petrucci A. Analysis of university course evaluations: from descriptive measures to multilevel models. Statistical Methods & Applications 2004; 13:357-373.

Yitzhaki S. More than a dozen alternative ways of spelling Gini. Economic Inequality 1998; 8:13-30.

RÉSUMÉ

Dr. Tamar Gadrich serves as the Head of the Industrial Engineering and Management Department at Ort Braude College, Karmiel, Israel. Tamar has a D.Sc. in statistics from the Technion—Israel Institute of Technology. Her current research interests are in the area of evaluating quality measured on an ordinal scale and applied probability. Dr. Gadrich has published her research in journals such as Quality and Reliability Engineering International, Methodology and Computing in Applied Probability, and Applied Stochastic Models in Business and Industry.