ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

*Andrzej Mantaj, Wiesław Wagner**

ORDERING THE SPATIAL UNITS BY THE NON-STANDARD METHOD AT VARIOUS STANDARDIZATION TRANSFORMATIONS

Abstract

Applying various standardization transformations for the set of characteristics we receive various ordering of spatial units. In the paper for the purpose of ordering we used the non-standard method in two variants: of the mean and of the median from the modules of transformed sample standardization values. The results of research on ordering were illustrated on the numerical material concerning the communes of the Podkarpackie Province for the year 2001.

Key words: Standardization transformations, numerical characteristics, ordering of spatial units, Podkapackie Province.

1. Introduction

Standardization transformations are commonly applied in the analysis of socio-economic phenomena of ordering the spatial units. Their basic purpose is to reduce the characteristics of different denomination to direct comparability. These are transformations of the type $(x-a)/b$, where a, b belong to sets of numerical characteristics of position and variability respectively. There exist many proposals of applying such transformations depending on the selection of characteristics a and b. The robust characteristics are also used among them. In practice there is most often applied the standardization transformation of the type $(x-\overline{x})/s$, at the arithmetic mean and standard deviation.

* The University of Information Technology and Management in Rzeszów.

Applying various standardization transformations for the set of characteristics we receive various ordering of spatial units. In the paper for the purpose of ordering we used the non-standard method in two variants: of the mean and of the median from the modules of transformed sample standardization values. The results of research on ordering were illustrated on the numerical material concerning the communes of the Podkarpackie Province for the year 2001.

2. Designations

Let X be the examined characteristic on the set of n statistical units $J_1, J_2, ..., J_n$. The set of observations of the characteristic X is expressed by n--element sample $x_1, x_2, ..., x_n$, which includes the set $P_n = \{x_1, x_2, ..., x_n\}$. The sequence of elements arranged from smallest to largest is presented by the set $P_{(n)} = \{x_{(1)}, x_{(2)}, ..., x_{(n)}\}$. Its elements are ordered statistics of the sample, where the index in brackets expresses the rank of observation in the sample P_n.

Through *MEAN*, *SD*, *AD*(a) *QUk* and *MED* we express operators (numerical modules) of determining the arithmetic mean, standard deviation, average deviation from the set a, quartiles of k-th degree and medians.

The following numerical characteristics were used in the paper:

A. Characteristics of positions:

A1. $\bar{x} = MEAN\{P_n\}$ – arithmetic mean,

A3. $\bar{x}_{HL} = MED\left\{\left(\frac{x_i + x_j}{2}\right); 1 \le i < j \le n\right\}$ – mean of Hodges-Lehmann (see e.g. Hampel *et al.*, 1986),

A4. $Q_k = QUk\{P_{(n)}\}$ – quartile of k-th series, at k = 1, 2, 3,

A5. $Me = MED\{P_{(n)}\}$ – median,

B. Characteristics of variability:

B1. $-s = SD(P_n)$ – standard deviation,

B2. $Q = \frac{Q_3 - Q_1}{2}$ – quartile deviation,

B3. $D(\bar{x}) = AD(\bar{x})$ – average deviation from the mean,

B4. $D(Me) = AD(Me)$ – average deviation from the median,

B5. $Mad = 1.4826 \cdot MED\{|x_i - Me| : x_i \in P_n, i = 1, 2, ..., n\}$ – median absolute deviation (median distance from the median), where the constant equals

$\frac{1}{\Phi^{-1}(0,75)}$, whereas $\Phi^{-1}(.)$ means the function inverse to the distribution function $\Phi(.)$ of the standardized normal distribution $N(0, 1)$ (Rousseeuw and Leroy, 1987),

B6. $s_{RC} = 1,1926 \cdot MED_i\{MED_j \,|\, x_i - x_j \,|\}$ – double median of absolute deviations, where the constant $c = 1,1926$ is the solution $\Phi\left[\Phi^{-1}\left(\frac{3}{4}\right)+\frac{1}{c}\right] - \Phi\left[\Phi^{-1}\left(\frac{3}{4}\right)-\frac{1}{c}\right]=\frac{1}{2}$ (Rousseeuw and Croux, 1993),

B7. $Q_{RC} = 2,2219 \cdot \{|\, x_i - x_j \,|; 1 \le i < j \le n\}_{(h)}$ – lower quantile of h-th rank, where $\{.\}_{(h)}$ means h-th term from the ordered sample including of absolute differences $|\, x_i - x_j \,|$, but $h = \frac{1}{4}\binom{n}{2}$, and the constant corresponds to the value $\frac{1}{\sqrt{2} \cdot \Phi^{-1}(0,625)}$, (Rousseeuw and Croux, 1993).

We set up the mentioned numerical characteristics into two sets of characteristics of position $A = \{\bar{x}, \bar{x}_{HL}, Q_1, Q_3, Me\}$ and variability $B = \{s, Q, D(\bar{x}), D(Me), Mad, s_{RC}, Q_{RC}\}$.

Some of the mentioned characteristics are very sensitive to occurrence of outlier observations in the sample. We classify to them: $\bar{x}$, s, $D(\bar{x})$. They are expressed by a very low point of breakdown (Rousseeuw and Leroy, 1987). In their place there are proposed many so called robust characteristics, i.e. calculated most often with the use of MED module (e.g. $\bar{x}_{HL}, Mad$).

3. Standardization transformations

Standardized sample statistics (*SSS*), expressing transformed observations of the sample through numerical characteristics $T_1(x) \in A$ and $T_2(x) \in B$, are called quotient statistics $u_i = \frac{x_i - T_1(x)}{T_2(x)}$, at $i = 1, 2, ..., n$. The statistics u expresses the value of deviation $x - T_1(x)$ falling on the unit $T_2(x)$. There exist many possible *SSS*. Their number is expressed by the product $(\#A) \cdot (\#B)$, where # means the size of the set. Not all of them are recommended. There are most often proposed *SSS* of the uniform type, i.e. when T_1 and T_2 are both classical or positional characteristics. Sometimes there are applied *SSS* of the mixed type.

Numerical characteristics T_1 and T_2 for the transformed sample values u equal $T_1(u) = 0$ and $T_2(u) = 1$. It results from the fact that the statistics u are linear transformations of the type $ax + b$, but $a = \frac{1}{T_2(x)}, b = -\frac{T_1(x)}{T_2(x)}$, and it means that $T_1(ax+b) = aT_1(x) + b$ and $T_2(ax+b) = aT_2(x)$. In particular this property occurs for the most-often-applied standardization transformation z $T_1(x) = \overline{x}$ and $T_2(x) = s$.

There are proposed 9 standardization transformations, given in Table 1. Standardization transformations S_1, S_2, S_3, S_4, S_5, S_7 are uniform, but S_1, S_3 are classical uniform transformations, and the remaining ones are positional. In the first case *SSS* are called classical standardized sample statistics, and in the second one, positional standardized sample statistics.

Let us yet notice that in the transformations S_1, S_2, S_3, S_4, S_5, the characteristics $T_2(x)$ are functions of the characteristics $T_1(x)$, i.e. arithmetic mean and median respectively. In the case S_2 it can be directly expressed through $Q_3 - Q_1 = (Q_3 - Me) + (Me - Q_1)$.

Table 1

Standardized sample statistics

Number	T_1	T_2	Number	T_1	T_2
S_1	$\overline{x}$	s	S_6	$\overline{x}$	s_{RC}
S_2	Me	Q	S_7	Me	Q_{RC}
S_3	$\overline{x}$	$D(\overline{x})$	S_8	$\overline{x}_{HL}$	s_{RC}
S_4	Me	$D(Me)$	S_9	$\overline{x}_{HL}$	Q_{RC}
S_5	Me	Mad			

Source: own study.

Properties $T_1(u) = 0, T_2(u) = 1$ will be illustrated on the example of the given data: 23.1, 19.4, 25.8, 22.6, 23.4, 18.7, 19.4, 23.6, 22.8, 20.8, 17.6, 24.5. Let us take here into consideration the transformations: S_1, S_2, S_3, S_4 *and* S_8.

The numerical characteristics calculated for them are:

$\overline{x} = 21.8083$, $\overline{x}_{HL} = 21.7$, $Me = 21.7$, $s = 2.5618$, $Q = 2.025$, $D(\overline{x}) = 2.1903$ and $D(Me) = 2.0583$, and the values of transformed statistics are given in the Table 2.

Table 2

Values of the selected standardization transformations

No.	x	S_1	S_2	S_3	S_4	S_8
1	23.1	0.5042	0.1975	0.5897	0.1943	0.5465
2	19.4	–0.9401	–1.6296	–1.0996	–1.6032	–0.8978
...	...	...	...	...	...	...
11	17.6	–1.6427	–2.5185	–1.9214	–2.4777	–1.6005
12	24.5	1.0507	0.8889	1.2289	0.8745	1.0930

Source: own study.

From the presented results in the Table 2 we notice that numerical characteristics that we take into account in a given standardization transformation always equal zero or one.

Following among others the presented results for *SSS* we can show that:

a) $r(S_{k,}S_{k'})=1$ – linear correlation coefficients equal 1, which results from properties for $r(ax+b,cx+d)=\dfrac{Cov(ax+b,cx+d)}{D(ax+b)D(cx+d)}=\dfrac{ac(Cov(x,x)}{acD(x)D(x)}=\dfrac{D^2(x)}{D^2(x)}=1$, where a, b, c, d are constants, but a, c are non-zero constants,

b) $g_1(x)=g_1(u)$, $g_2(x)=g_2(u)$ – coefficients of skewness and kurtosis are invariant, which results from $g_1(u)=g_1(ax+b)=\dfrac{\mu_3(ax+b)}{\mu_2^{3/2}(ax+b)}=$ $=\dfrac{a^3\mu_3(x)}{a^3\mu_2^{3/2}(x)}=g_1(x)$. It is similarly shown for g_2.

4. Non-standard method

In multi-characteristic taxonomic research of the set of n units (including spatial units) $J_1, J_2, ..., J_n$, for their ordering there is commonly used the non-standard method. It is determined with the use of: means or medians from absolute values of *SSS*. Its construction in the first case includes the following steps:

a) for each characteristic from the set sample there are determined characteristics $T_{1j}(x)$, $T_{2j}(x)$, at $j = 1, 2, ..., p$, where p is the number of characteristics,

b) for i-th unit of j-th characteristics there is applied standardization transformation $u_{ij}=\dfrac{x_{ij}-T_{1j}(x)}{T_{2j}(x)}$, for $i = 1, 2, ..., n$,

c) there is made averaging of absolute values u_{ij} for each unit of the sample at its characteristics, i.e. $U_i = \acute{S}R(|u_{i1}|, |u_{i2}|, ..., |u_{ip}|)$,

d) units $J_1, J_2, ..., J_n$ are ordered successively, in accordance with the ordered sample $U_{(1)} \leq U_{(2)} \leq ... \leq U_{(n)}$.

In the case of non-standard method of medians in step c) there are applied medians in the place of means, i.e. $U_i = MED(|u_{i1}|, |u_{i2}|, ..., |u_{ip}|)$.

In our case we apply the non-standard method for each of standardization transformations mentioned in the Table 1, for its two variants. After orderings according to particular standardization transformations, for each of the units there will be determined its average position.

5. Ordering the communes of podkarpackie province

We will present ordering the communes with the use of standardization transformations for numerical data concerning 185 spatial units – rural and municipal communes of Podkarpackie Province. We take the names of communes in alphabetical order according to the names of districts, and communes within them (*Statistical Yearbook of Podkarpackie Province*, 2002). There was taken the following set of 14 characteristics (Mantaj and Wagner, 2007):

X_1 – area in sq. km,
X_2 – population,
X_3 – population for 1 sq. km,
X_4 – population of productive age (%),
X_5 – mployed in total,
X_6 – registered unemployed,
X_7 – employed in total for 100 people,
X_8 – unemployed to employed (%),
X_9 – consumption of water from water supply systems for 1 inhabitant,
X_{10} – average usable area in sq. m for 1 flat,
X_{11} – area of arable land in sq. km,
X_{12} – area of forests and forest land in sq. km,
X_{13} – area of arable land to total area (%),
X_{14} – area of forests and forest land to total area (%).

Among the mentioned characteristics there are absolute characteristics: X_1, X_2, X_5, X_6, X_{11} i X_{12}, and the remaining ones are relative. Numerical data of characteristics X_1, ..., X_6, X_9, X_{10}, X_{11} i X_{12} for communes are included in the statistical yearbook. In the Table 3 there are given data for first 8 communes, i.e.:

Table 3

Numerical data for selected communes

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	15	6 713	447.5	64.1	1851	684	27.6	15.9	24.3	57.9	4.3	5.1	28.7	34.0
2	13	10 274	790.3	64.7	2075	1228	20.2	18.5	29.3	54.7	1.9	3.9	14.6	30.0
3	158	3 237	20.5	59.9	200	412	6.2	21.2	11.7	68.8	17.0	10.7	10.8	6.8
4	287	1 610	5.6	65.1	200	213	12.4	20.3	2.7	60.8	1.4	25.3	0.5	8.8
5	185	2 366	12.8	57.8	116	322	4.9	23.5	10.1	70.0	6.0	11.4	3.2	6.2
6	97	5 376	55.4	58.7	314	666	5.8	21.1	34.9	73.9	26.2	4.7	27.0	4.8
7	476	2 432	5.1	65.9	320	333	13.2	20.8	14.3	62.5	3.2	389.7	0.7	81.9
8	167	6 639	39.8	58.9	764	836	11.5	21.4	5.9	72.0	20.4	96.1	12.2	57.5
...	..	...	...	...	..	...	...	...	...	...	...	...	...	...

1 – Lesko (town), 2 – Ustrzyki Dolne (town), 3 – Baligród, 4 – Cisna, 5 – Czarna, 6 – Lesko 7 – Lutowiska, 8 – Olszanica (village).

Source: own study.

The presented data were used for ordering the communes with the non-standard method of ordering the communes in the variant of means and medians from modules of values of standardization transformations given in Table 2. All calculations were done according to our own recalculation formulas in the calculation sheet EXCEL. There was also used the procedure *Rank and Percentile* in the menu Tools | Data Analysis.

Date in the Table 3, although fragmentary, indicate very strong diversification of values of characteristics for particular spatial units (communes). It concerns the majority of characteristics: $X_1, X_2, X_3, X_5, X_6, X_7, X_{11}, X_{12}, X_{13}$ and X_{14}.

The values of number characteristics of position and variability are given in the Table 4.

Table 4

Numerical characteristics of characteristics

Characteristics	$\overline{x}$	$\overline{x}_{HL}$	Me	s	Q	$D(\overline{x})$	$D(Me)$	Mad	S_{RC}	Q_{RC}
X_1	95.67	87.50	85.00	75.07	42.50	53.55	52.67	62.27	60.82	59.99
X_2	9 744.22	8 269.50	7 573.00	8 970.97	2 781.00	4 974.17	4 621.05	3 976.33	3 900.99	4 283.82
X_3	198.71	121.95	105.97	260.68	62.41	167.60	198.71	73.90	77.20	79.06
X_4	57.95	57.57	57.23	3.05	1.80	2.37	2.28	2.36	2.40	2.52
X_5	1 375.23	624.00	507.00	3 170.43	303.00	1 429.18	1 040.96	336.55	355.39	366.61
X_6	932.66	812.50	757.00	757.42	300.50	465.11	441.55	449.23	423.37	431.05

Table 4 (contd.)

Characteristics	$\overline{x}$	$\overline{x}_{HL}$	Me	s	Q	$D(\overline{x})$	$D(Me)$	Med	S_{RC}	Q_{RC}
X_7	11.24	8.30	6.49	10.19	3.80	7.68	6.48	3.63	3.67	3.91
X_8	17.35	17.08	16.77	4.37	2.79	3.24	3.21	4.06	3.82	3.76
X_9	18.00	17.60	17.60	10.41	6.90	8.14	8.14	10.82	9.90	10.22
X_{10}	69.20	71.20	71.70	12.39	4.00	7.39	6.96	6.23	6.44	6.67
X_{11}	34.22	33.85	34.30	21.14	15.50	17.01	17.01	22.68	20.99	20.89
X_{12}	32.88	23.55	17.10	49.27	18.70	29.80	26.75	21.65	19.92	19.33
X_{13}	42.08	42.66	43.64	16.47	12.06	13.52	13.47	16.86	17.29	16.71
X_{14}	26.12	25.15	24.00	18.03	14.20	14.80	14.67	20.99	18.46	17.88

Source: own study.

The analysis of numerical characteristics from the Table 4:

- for many characteristics there are noticed considerably higher values of the mean than the median (X_1, X_2, X_3, X_5, X_6, X_7, X_{11}, X_{12}),
- the mean of Hodges-Lehmann – most often takes the values between the median and the arithmetic mean,
- standard deviations are generally the highest among the mentioned variation measures,
- average deviations $D(\overline{x})$, $D(Me)$ take the values little differing between each other, and for some characteristics (X_8, X_9, X_{11}, X_{13}, X_{14}) they are almost identical.

Table 5

Order of communes at various standardization transformations by the nonstandard method – mean variant

No.	S1	S2	S3	S4	S5	S6	S7	S8	S9	AP
1	40	25	44	35	23	39	27	24	27	32
2	24	16	26	19	18	23	17	18	18	20
3	57	69	59	72	82	57	84	77	78	71
4	13	26	16	21	33	25	33	32	32	26
5	37	62	36	58	66	42	67	58	57	54
6	94	122	97	118	126	84	122	114	113	110
7	4	14	7	7	11	7	11	11	10	9
8	55	72	61	50	61	78	55	65	63	62
9	51	67	54	47	51	65	47	57	56	55
10	19	46	15	20	31	18	23	28	25	25
...	...	...	...	...	...	...	...	...	...	...
180	22	36	22	31	38	30	38	35	36	32
181	30	17	28	23	17	21	18	17	17	21
182	180	150	180	161	144	182	149	171	171	165
183	93	28	92	46	27	61	32	33	34	50
184	54	66	56	53	67	71	65	67	68	63
185	107	131	108	106	112	101	104	108	105	109
SS	112.4	818.6	246.1	270.8	1 275.0	820.1	538.3	560.2	529.5	574.6

Source: own study.

In the Table 5 there were placed ranks of communes received from ordering with the non-standard method according to the variant of means for particular methods of standardization transformations. This is a fragment of 10 first and 6 last communes. In the table there were used designations: *SS* – square sum and *AP* – average position. The size of *SS* was calculated from standardization values, and *AP* from values of position (ranks) of particular standardization transformations.

On the basis of results presented in the Table 5, concerning the applied standardization transformations it can be stated that:

- for some communes all ranks are identical, which concerns the communes of numbers: 18 – rank 3; 92 – 2; 159 – 5 and 168 – rank 1,
- two communes, i.e. 26 and 124, have almost identical ranks 5, 6 and 10, 11,
- for the majority of communes, there occur very high differences in ranks (e.g. the commune number 6 has ranks from the lowest 84 to the highest 126),
- square sums of standardized sample statistics (*SSS*) can be ranked into the increasing sequence: 1 || 3, 4 || 8, 9, 7 || 2, 6 || 5, i.e. the strongest diversification in the values of *SSS* is expressed by S_5, at calculation of which the median and median absolute deviation were used.

In the table 6, there is placed a fragment of the table of ranks for communes at various standardization transformations in the non-standard method at the median variant.

Table 6

Order of communes at various standardization transformations by the non-standard method – median variant

No.	*S*1	*S*2	*S*3	*S*4	*S*5	*S*6	*S*7	*S*8	*S*9	*AP*
1	36	43	50	29	40	37	39	44	40	40
2	15	16	26	12	15	16	11	13	11	15
3	42	52	40	51	56	69	54	54	55	53
4	21	13	19	22	10	8	8	6	6	13
5	28	36	28	44	36	20	38	25	28	31
6	87	106	77	120	112	71	118	81	86	95
7	26	10	16	19	8	2	6	9	10	12
8	52	47	66	58	42	62	43	56	60	54
9	53	37	60	46	45	60	50	46	49	50
10	159	119	141	143	131	80	135	162	162	137
...	...	...	...	...	...	...	...	...	...	...
180	24	41	22	37	53	23	42	30	31	34
181	18	22	24	9	16	22	15	20	16	18
182	181	148	182	131	150	178	142	167	167	161
183	104	24	79	59	48	120	51	64	61	68
184	82	83	122	55	88	45	87	111	108	87
185	81	120	117	99	113	140	110	118	114	112
SS	71.2	210.8	153.7	127.3	274.3	293.6	128.8	149.2	145.3	172.7

Source: own study.

The results presented in the Table 6 allow stating that:

- some communes have little differing ranks, which concerns the communes of numbers: 26 – rank from median 4; 69 – 5, 92 – 3; 134 – 53; 145 – 22; 159 – 7; 168 – 1,
- ranks for the majority of communes considerably differ,
- there occurs considerable diversification of ranks at the variant of means and median within the set standardization transformation,
- square sums in all cases are lower than the ones given in the Table 5 and they can be ranked 1 || 4, 6, 7, 3, 9 || 2 || 5, 6,
- correlation coefficient between mean (Table 5) and median (Table 6) ranks is 0,9145.

6. Recapitulation

The presented numerical characteristics of position and variability from the samples show diversification depending on the kind of characteristics. However, there is noticed high usability of medians and median absolute deviation. Both characteristics do not cause greater difficulties in numerical calculations, which cannot be said of the mean of Hodges-Lehmann and s_{RC} and Q_{RC}, particularly in the case of big samples. Among standardization transformations used in the non-standard method of ordering spatial units at the variant of means there is recommended the transformation S_5, and at the variant of median, transformations S_5 and S_6. Square sum from Standardized Sample Statistics is an important meter of effeciency of standardization transformations.

Reference

Hampel F. R., Ronchetti E. M., Rousseeuw P. J., Stahel W. A. (1986), *Robust statistics: The approach based on influence function*, Wiley&Sons, New York.

Mantaj A., Wagner W. (2007), *Comparative analysis of classical and position number characteristics of selected social-economic characteristics of communes of Podkarpackie province* "Acta Universitatis Lodziensis", Folia Oeconomica **206**, 445–461.

Statistical *Yearbook of Podkarpackie Province 2002*, (2002), US, Rzeszów.

Rousseeuw P. J., Croux C. (1993), *Alternatives to the median absolute deviation.* "J. Amer. Statist. Assoc.", **88**, 1273–1283.

Rousseeuw P. J., Leroy A. M. (1987), *Robust regression and outlier detection*, Wiley&Sons, New York.

Andrzej Mantaj, Wiesław Wagner

Porządkowanie jednostek przestrzennych metodą bezwzorcową przy różnych przekształceniach standaryzacyjnych

Przekształcenia standaryzacyjne są powszechnie stosowane w analizie zjawisk społeczno-gospodarczych porządkowania jednostek przestrzennych. Ich podstawowym celem jest sprowadzenie cech o różnych mianach do bezpośredniej porównywalności. Są to przekształcenia typu $(x - a)/b$, gdzie a, b należą do zbiorów charakterystyk liczbowych, odpowiednio położenia i zmienności. Istnieje wiele propozycji stosowania takich przekształceń w zależności od doboru charakterystyk a i b. Wśród nich korzysta się także z charakterystyk odpornych. W praktyce najczęściej stosowane jest przekształcenie standaryzacyjne typu $(x-\bar{x})/s$, przy średniej arytmetycznej i odchyleniu standardowym.

Stosując różne przekształcenia standaryzacyjne do zbioru cech, otrzymuje się różne uporządkowanie jednostek przestrzennych. W pracy do porządkowania użyto metodę bezwzorcową w dwóch wariantach: średniej i mediany z modułów przekształconych próbkowych wartości standaryzacyjnych. Wyniki badań porządkowania zilustrowano na materiale liczbowym dotyczącym gmin woj. podkarpackiego za rok 2001.