

**PROPERTIES OF NORMALIZATION METHODS USED
IN THE CONSTRUCTION OF AGGREGATE MEASURES**

Kesra Nermend, Ph.D. Eng.

*University of Szczecin
Faculty of Economics and Management
Mickiewicza 64, 71-101 Szczecin, Poland
e-mail: kesra@szafir.univ.szczecin.pl*

Received 10 October 2012, Accepted 23 April 2013

Abstract

This article presents a study on the impact of missing objects, untypical objects and the impact of displacing coordinates of the object on the results of normalization in the construction of aggregate measures for ranking socio-economic objects. The study was conducted on simulated data sets generated in order to investigate the properties of normalization methods. This article focuses on the standardization formula responsible for moving the set of objects.

Keywords: normalization, synthetic measure, taxonomy, methods for linear ordering.

JEL classification: C38.

Introduction

When ranking socio-economic objects it is essential that the values of aggregate measures are affected only by the properties of a given object expressed by means of its coordinates (i.e. indicators that describe the object). It is helpful if the measure values are not influenced, or influenced to a small extent only, by the changes in coordinates of other objects, the completeness of a set of objects or by the presence of non-typical objects. The situation is particularly disadvantageous when whichever object from the set of study determines the order of other objects to be ranked. For example, the appearance of non-typical objects can radically affect the ranking.

The values of the aggregate measure which was proposed by Z. Hellwig¹, the so called Hellwig's synthetic measure, are affected by the methods used at subsequent stages of its evaluation. It found several applications, such as: construction of synthetic variables in the process of econometric modeling², evaluation of product quality³, studies on regional growth⁴, evaluating stock investment attractiveness⁵, measurement of social status⁶, poverty⁷, famine⁸, the ease of doing business⁹, competitiveness¹⁰, peace¹¹. In case of evaluating the regional growth it is also possible to apply the vector measure¹² whose advantage is its flexibility in applying economics, especially in its variation where the scalar product is used exclusively¹³.

All the methods of measuring the socio-economic growth require proper normalization of indicators. Accurate normalization determines the order of objects in the set. Theoretical and practical studies on a small set of objects were conducted by Kukuła and Pawełek. This paper presents the results of studies on simulated data where the author takes into consideration different size of data found in the economic studies in order to specify to what extent the choice of a method to determine indicators of the object set displacement affects the stability of the objects' positions.

1. The test set and the analyzed normalization methods

In order to analyze the properties of the normalization methods the author generated test sets consisting of 6 (NUTS1 in Poland), 16 (NUTS2 in Poland), 66 (NUTS3 in Poland), 150, 380, 700 (approximately NUTS4 in Poland) objects. The point (object) coordinates of these sets have a normal distribution whose average value is 2 for the axes 0X and 0Z and 1 for the axis 0Y. Figure 1 shows the example of a set of 380 elements.

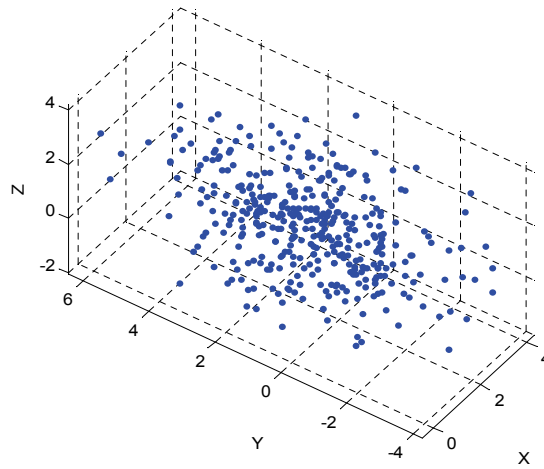


Fig. 1. A test set of a size of 380 elements

Source: own compilation.

In most normalization methods two operations are performed: the set of objects is moved and the scale of coordinates is changed. In general, the formula of property normalization is written¹⁴ as:

$$x'_{ij} = \left(\frac{x_i - A_i}{B_i} \right)^p \tag{1}$$

where:

- x'_{ij} – the value of the i -th variable for the j -th object after normalization,
- B_i – the normalization base of the i -th variable ($B_i \neq 0$),
- A_i, p – parameters.

The counter of this formula is responsible for the displacement of a set of points. The aggregate measures, which are counted in relation to the reference point which has been determined on the basis of the provided data set, are insensitive to this displacement. For example, in aggregate measures constructed according to Hellwig’s method the reference point is a model determined on the basis of data obtained from the matrices of closed test observations. When $p = 1$ the value A_i does not affect the calculation result. Thus it is possible to take it as zero so that the above formula is reduced to the form:

$$x'_{ij} = \frac{x_i}{B_i} \tag{2}$$

There are measures, however, for which the value A_i is relevant, such as non-model methods. Therefore it is important to find out to what extent A_i has effect on the result of normalizing these measures. With a view to that the author defines the coefficient of change in the value of coordinates which describes how the coordinates' values of the normalized objects after normalization that has been limited solely to these coordinates differ from the coordinates of the objects that have been subjected to an operation:

$$w = \frac{\sum_{i=1}^N \sum_{j=1}^M \left| \text{norm} \left(x_{ij} \right) - \text{norm} \left(\hat{x}_{ij} \right) \right|}{NM} \quad (3)$$

where:

- norm – a selected normalizing function,
- \hat{x}_{ij} – coordinates subjected to an operation,
- M – the number of objects,
- N – the number of indicators (coordinates).

The operations performed on the coordinates will simulate certain changes in indicators which are describing the objects. Let the value of N be 3. For the sake of this study on the object set displacement it has been assumed that seven variants of the normalizing function serve as the normalizing function it self¹⁵.

$$1. \text{norm} \left(x_{ij} \right) = x_{ij} - \bar{x}_i \quad (4)$$

$$2. \text{norm} \left(x_{ij} \right) = x_{ij} - \min_j \left(x_{ij} \right) \quad (5)$$

$$3. \text{norm} \left(x_{ij} \right) = x_{ij} - \max_j \left(x_{ij} \right) \quad (6)$$

$$4. \text{norm} \left(x_{ij} \right) = x_{ij} - \text{median}_j \left(x_{ij} \right) \quad (7)$$

$$5. \text{norm} \left(x_{ij} \right) = \begin{cases} 0 & \text{for } x_{ij} < \bar{x}_i - a\sigma_i \\ x_{ij} - \bar{x}_i + a\sigma_i & \text{for } x_{ij} \in \langle \bar{x}_i - a\sigma_i; \bar{x}_i + a\sigma_i \rangle \\ 2a\sigma_i & \text{for } x_{ij} > \bar{x}_i + a\sigma_i \end{cases} \quad (8)$$

$$\begin{aligned}
 6. \text{ norm} \left(x_i \right) &= \begin{cases} 0 & \text{for } x_i < \text{percentile}_a \left(x_j \right) \\ x_i - \text{percentile}_a \left(x_j \right) & \text{for } x_i \in \left(\text{percentile}_a \left(x_j \right); \text{percentile}_{100-a} \left(x_j \right) \right) \\ \text{percentile}_{100-a} \left(x_j \right) - \text{percentile}_a \left(x_j \right) & \text{for } x_i > \text{percentile}_{100-a} \left(x_j \right) \end{cases} \quad (9) \\
 7. \text{ norm} \left(x_i \right) &= x_j \quad (10)
 \end{aligned}$$

2. Testing sensitivity to missing objects

In the research process not always is the diagnostic test full. There can be some data or an object missing. For instance, when preparing rankings of the European Union states we often find out that data concerning some countries are not available.

In such a case the objects for which data are missing are eliminated from the set. However, the missing data can become available later, therefore a question arises how dismissing the object from the set can affect the normalization results. With a view to that a test set was randomly chosen from which subsequent objects were eliminated. In the process of the elimination the indicator of changes in the coordinate values was observed. The result is shown in Figure 2.

When reducing objects the best results are achieved by means of a method in which the mean value is used. When the reduction is very limited, e.g. 0.5%, the methods based on maximum and minimum values are better. However, the obtained results are not significantly better than the results for the mean value. In all the methods the indicator of changes in the point coordinate values is growing as the number of objects is increasing. The reduction of the number of objects to 10% gives the indicator value of less than 5%. Taking into consideration the fact that on further stages of the study the reduced number of objects affects as well the functioning of the methods a conclusion has been drawn that the number of missing objects should be as small as possible.

What is more, the method relying on a mean value provides the highest repeatability of results. The results obtained in subsequent computations (based on the same parameters) differ from one another to a small extent only – as seen in Figure 3. The lowest repeatability is presented by the methods basing on a minimum and a maximum.

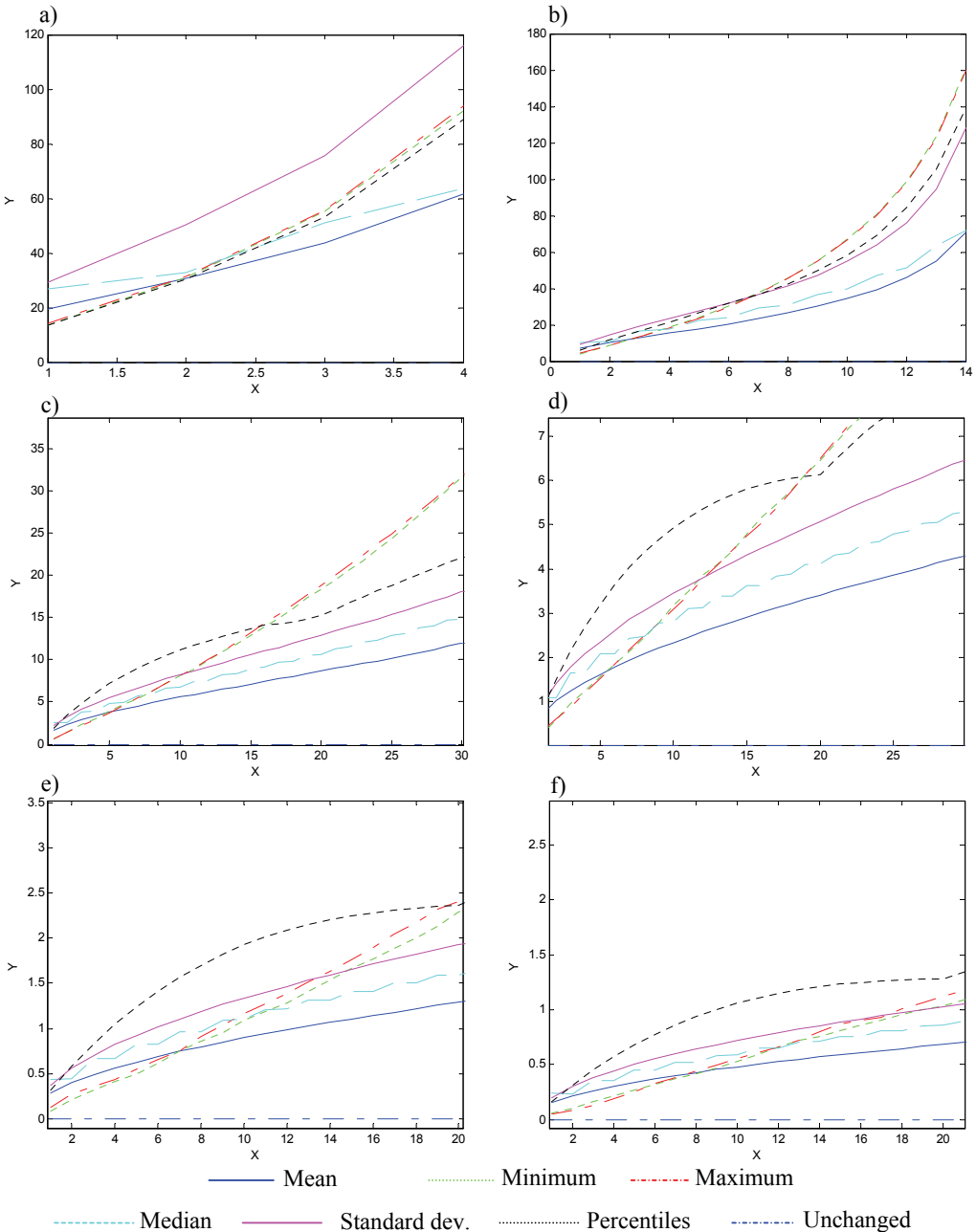


Fig. 2. Dependency of the indicator of change in the point coordinate values on the number of reduced objects in the set of a) 6, b) 16, c) 66, d) 150, e) 380 and f) 700 objects

Source: own compilation.

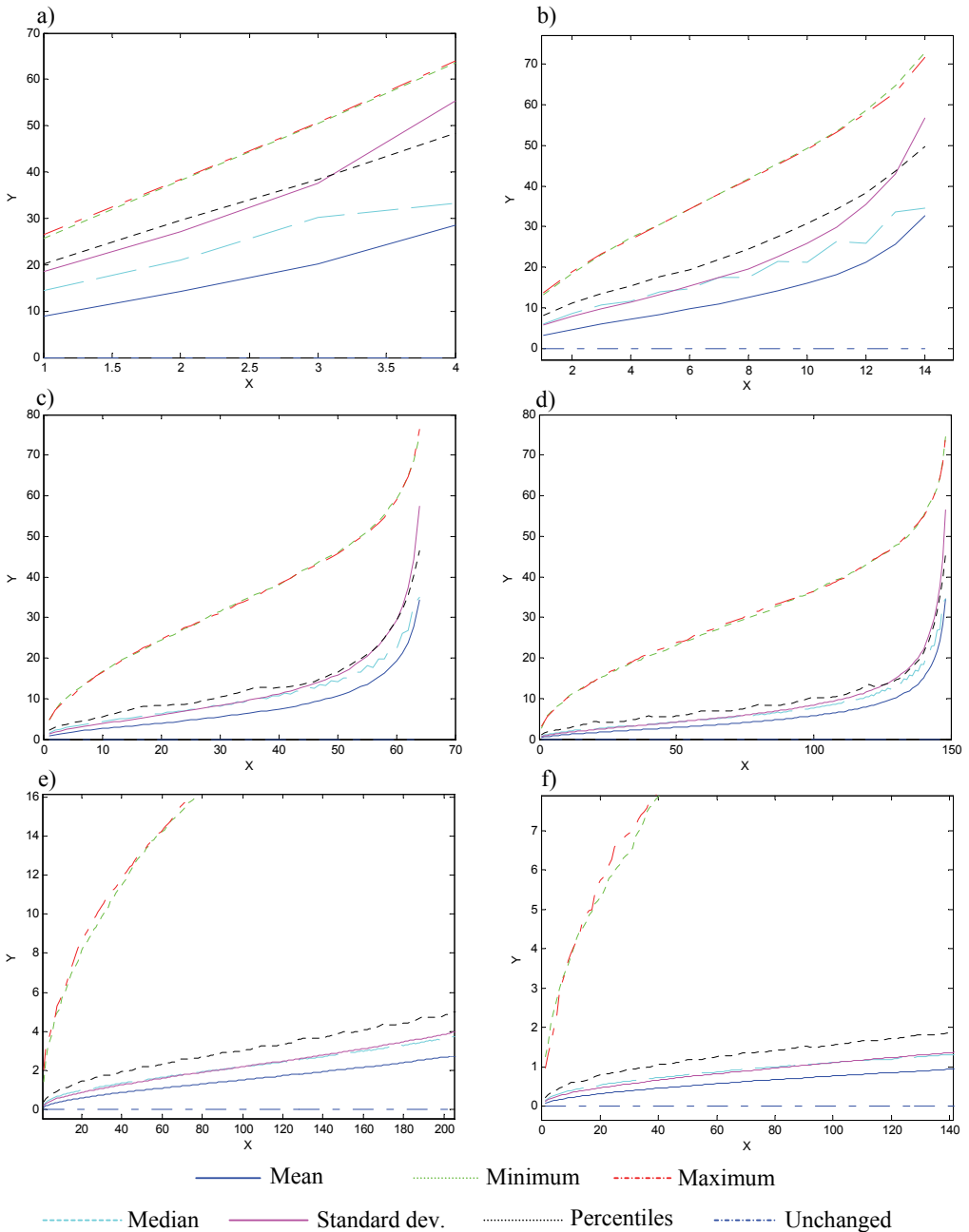


Fig. 3. Dependency of the repeatability coefficient on the number of reduced objects in the set of a) 6, b) 16, c) 66, d) 150, e) 380 and f) 700 objects

Source: own compilation.

3. Testing sensitivity to non-typical objects

In the course of our studies we happen to come across non-typical objects whose indicators' values can be several, sometimes tens or, on rare occasions, hundreds of times larger than in case of other objects. The example of such an object is the district of Bełchatów whose indicators vary much from other districts due to a brown coal mine and a big power plant operating in the area, which significantly influences those indicators that are related to environmental protection and the local government revenue¹⁶. The purpose of the next study was to evaluate the effects the non-typical objects had on normalization results. To that end one of the objects was being moved along the one of the ox coordinate axis. The standard deviation of the set of objects along that axis was 2. The movement commenced at a distance of 8 along the ox axis from the center of the set. It ended at the same distance of 8 of the ox axis from the set centre. The objects was moving across the set.

The study results are presented in Figure 4. Methods that are the most insensitive to non-typical objects are the ones that rely on a median and a mean value. Sensitivity of the majority of methods to single non-typical objects decreases as the number of objects grows. It is very small in case of 150 objects. When the number of objects is 380 and more we can consider it its effect on the test results practically insignificant. The exception here are the methods using a maximum and a minimum which are very sensitive to non-typical objects.

The former is sensitive to non-typical objects whose coordinates are considerably smaller, while the latter – to non-typical objects with coordinates considerably larger than the coordinates of the remaining objects. Their sensitivity to non-typical objects is so significant that it makes them practically useless when a given set contains any such objects. It results from the fact that the emergence of non-typical objects leads to the substantial displacement of the whole set. In such a situation the minimum method can be replaced by the percentile method. Its properties are similar, but it is much less sensitive to non-typical objects.

When testing the repeatability of results the repeatability coefficient was replaced with the standard deviation of the indicator of change in the point coordinate values. The repeatability turned out relatively high.

4. Testing sensitivity to the displacement of one of the objects

Objects are evolving in time, therefore the values of their indicators are changing in time as well, which leads to their displacement. However, such displacement should not affect the positions of other objects.

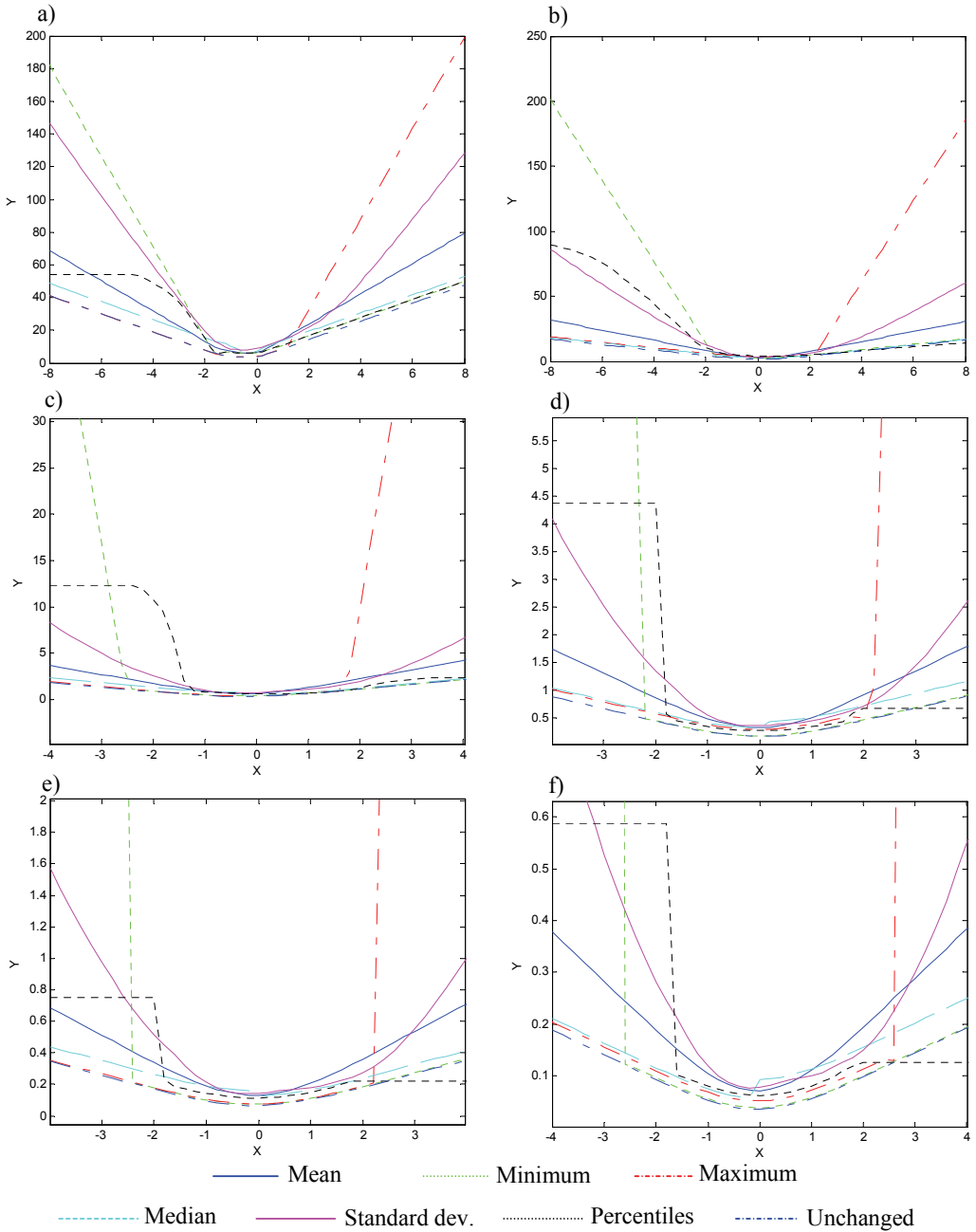


Fig. 4. Dependency of the indicator of change in the point coordinate values on non-typical objects: a) 6, b) 16, c) 66, d) 150, e) 380, f) 700

Source: own compilation.

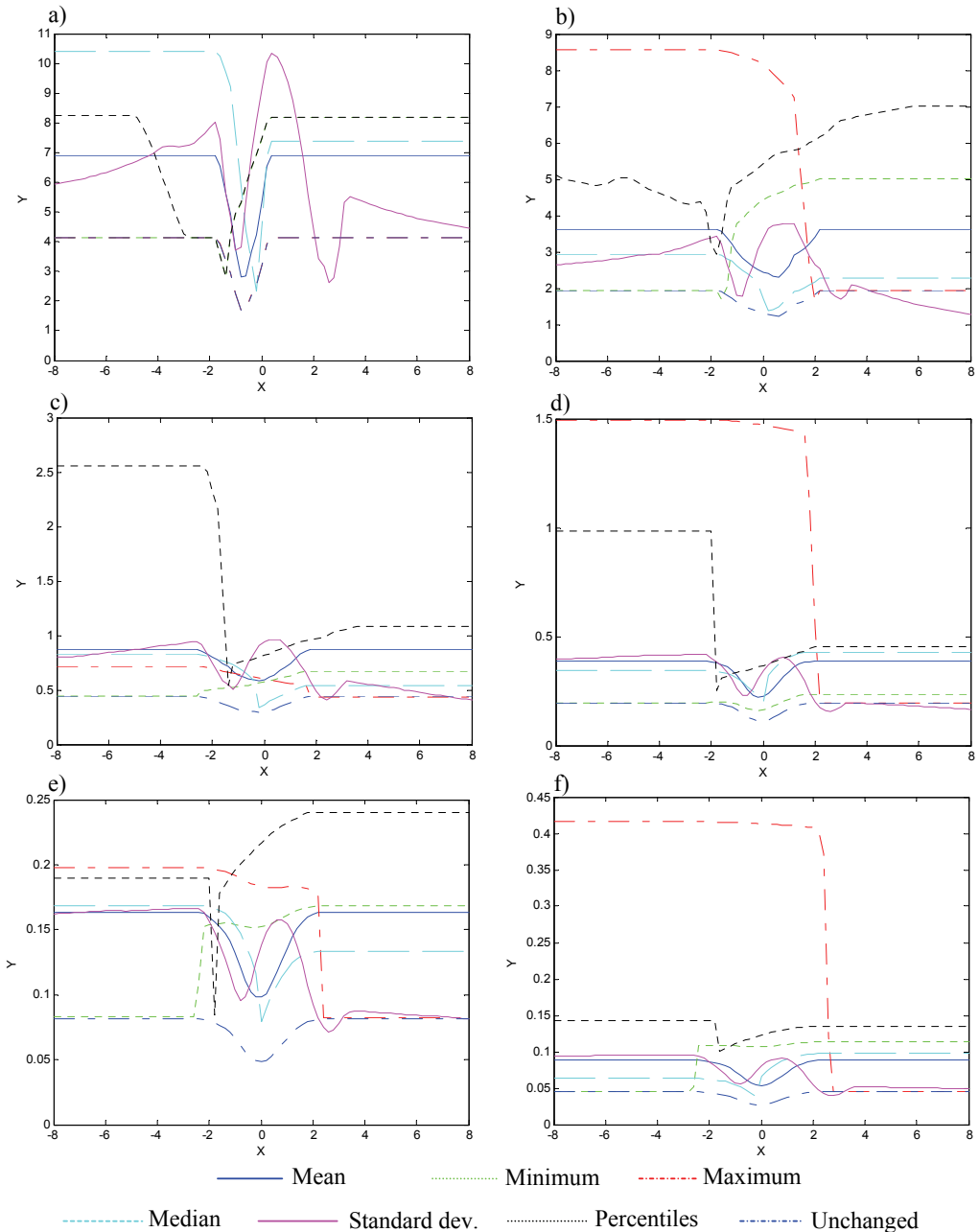


Fig. 5. Dependency of the standard deviation of the indicator of change in the point coordinate values on non-typical objects: a) 6, b) 16, c) 66, d) 150, e) 380, f) 700

Source: own compilation.

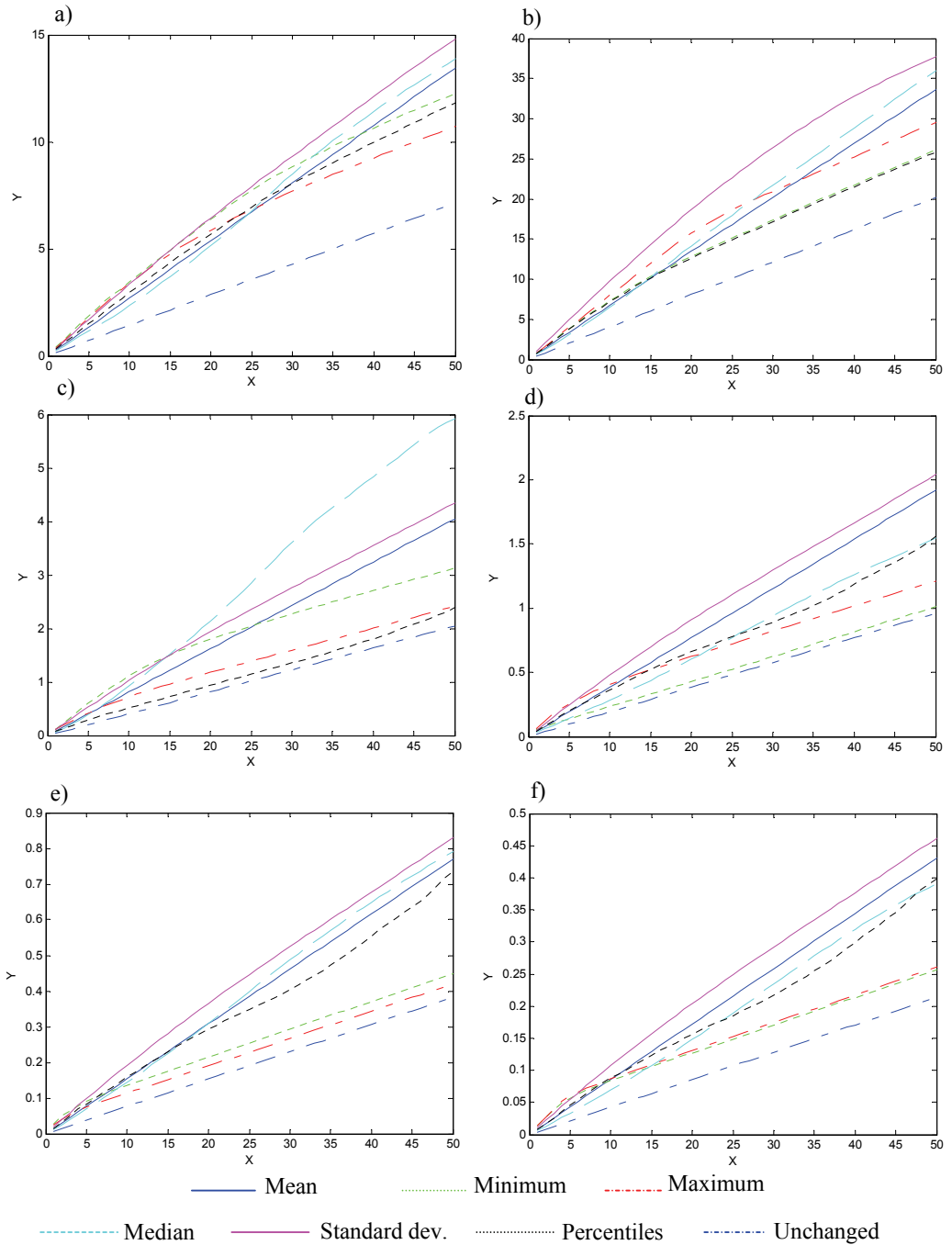


Fig. 6. Dependency of the indicator of change in the point coordinate values on the displacement of one of the objects : a) 6, b) 16, c) 66, d) 150, e) 380, f) 700

Source: own compilation.

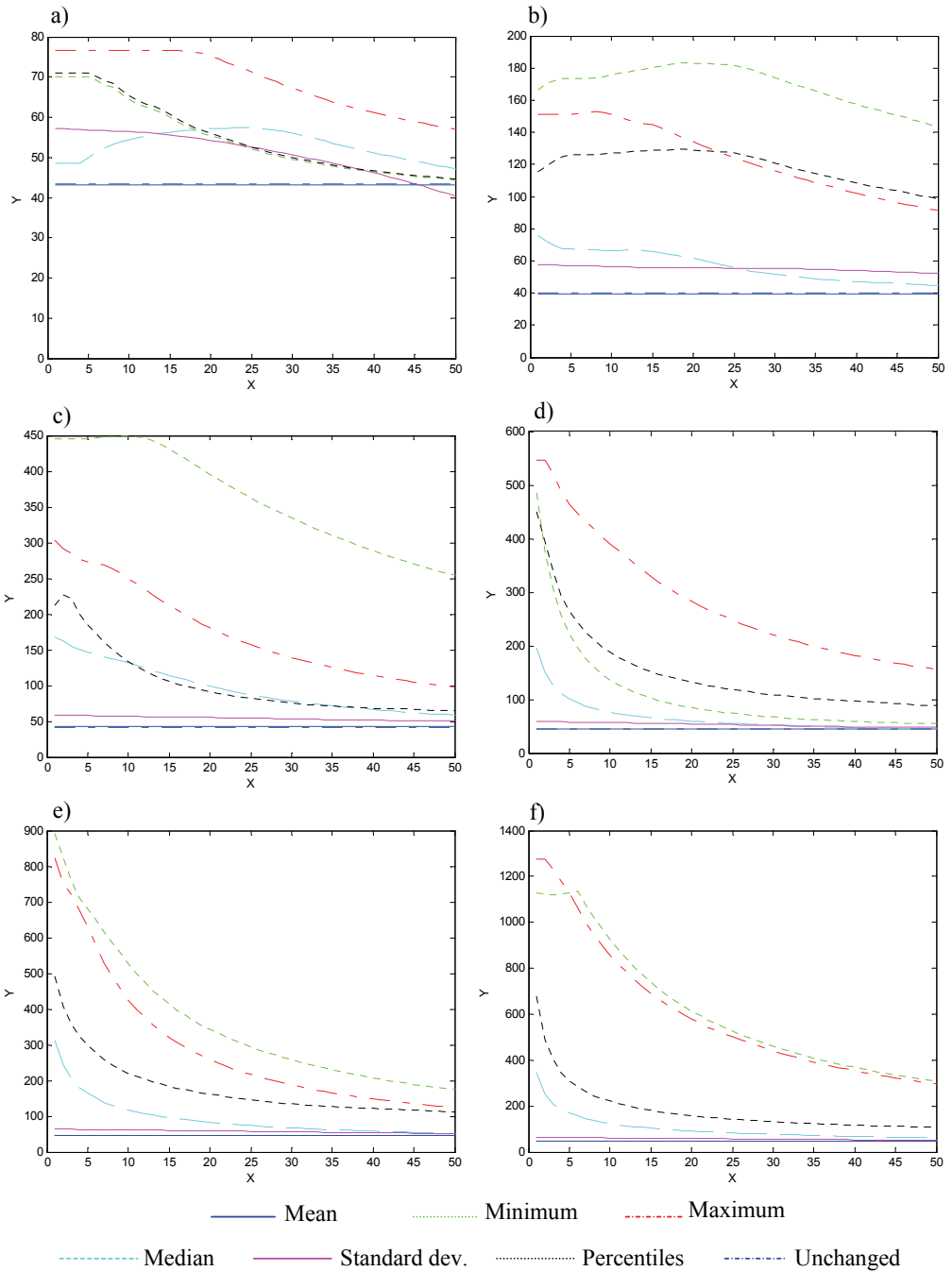


Fig. 7. Dependency of the repeatability coefficient on the object displacement in the set of a) 6, b) 16, c) 66, d) 150, e) 380 and f) 700 objects

Source: own compilation.

Due to the fact that during normalization the set of objects is moved according to the value calculated on the basis of coordinates of all the objects, it seems worthwhile to find out if the displacement of one object has an impact on the coordinates of all the objects.

In order to determine the degree of that impact one of the objects was randomly chosen from the set and then it was moved to a randomly chosen position in space. In the course of the displacement the indicator of change in the point coordinate values was determined. The results are shown in Figure 6. On the oy axis there is the indicator of change in the point coordinate values, while on the ox axis – the displacement step. Zero denotes the original object position, while 50 – the target position. Figure 7 presents the repeatability of the obtained results. Because the position in space towards which the object was moving was chosen at random, the repeatability coefficient does not take high values. The best results are achieved when applying the method based on a mean value. When the set consists of more objects the standard deviation method has also proved useful. When the number of objects is medium or large and the object displacement is significant the percentile and median methods are recommendable.

Conclusions

In the process of normalization the displacement and scaling indicators are determined on the basis of coordinates of all the objects, in consequence of which the changes in coordinates of other objects, the completeness of the object set and the presence of non-typical objects will in a way have an impact on the position of other objects. The scope of this impact will depend on the normalization method that we have chosen. In this paper the author tested the effect of missing objects, of non-typical objects and of the displacement of one of the objects on normalization results. All the above factors influence the displacement of the object set in space. In case of non-typical objects the minimum and maximum methods turned out the least useful. Their sensitivity to non-typical objects was so high that it practically eliminated them from the process of constructing measures that are invariant when facing the object set movements. One of them are the without pattern. measures. methods that are the least sensitive to non-typical objects are the methods where a mean value and a median are applied. The mean value method has proven most useful in case of an incomplete set of objects. The methods which are the least sensitive to displacement of one of the objects are the ones basing on a minimum and a maximum.

Notes

- ¹ Hellwig (1968).
- ² Bartosiewicz (1984).
- ³ Borys (1984).
- ⁴ See Nermend (2006a); Nermend (2008a), pp. 87–97.
- ⁵ Łuniewska, Tarczyński (2006).
- ⁶ *Social Watch Annual Report...* (2010).
- ⁷ Ibidem.
- ⁸ *Global Hunger Index...* (2012).
- ⁹ *Making a difference...* (2012).
- ¹⁰ *Global Competitiveness Report...* (2012).
- ¹¹ *Global Hunger Index...* (2012).
- ¹² See Nermend (2006a), pp. 127–129; Nermend (2006b), pp. 123–126; Nermend (2008b); Kolenda (2006).
- ¹³ Nermend (2009).
- ¹⁴ Grabiński, Wydymus, Zeliaś, (1989).
- ¹⁵ See Gatnar, Walesiak (2004), p. 35; Panek (2009), pp. 37–41.
- ¹⁶ Nermend (2009), pp. 114–120.

References

- Bartosiewicz, S. (1984). *Zmienne syntetyczne w modelowaniu ekonometrycznym*. Prace naukowe AE, nr 262, Wrocław.
- Borys, T. (1984). *Kategoria jakości w statystycznej analizie porównawczej*. Prace naukowe AE, seria: monografie i opracowania, nr 23, Wrocław.
- Gatnar, E. & Walesiak, M. (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Global Competitiveness Report 2010–2011*. World Economic Forum, www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2010-11.pdf (15.10.2012).
- Global Hunger Index. The challenge of hunger: focus on the crisis of children undernutrition*. International Food Policy Research Institute, www.ifpri.org/sites/default/files/publications/ghi10.pdf (15.10.2012).
- Global Peace Index Methodology, Results and Findings*. Institute for Economics & Peace, www.visionofhumanity.org/wp-content/uploads/PDF/2010/2010%20GPI%20Results%20Report.pdf (15.10.2012).
- Grabiński, T., Wydymus, S. & Zeliaś A. (1989). *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*. Warszawa: PWN.

- Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny* 1968, No. 4.
- Human Development Report*. United Nations Development Programme. New York: Oxford University Press, <http://hdr.undp.org/en/reports/global/hdr2010> (15.10.2012).
- Kolenda, M. (2006). *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Kukuła, K. (2000). *Metoda unitaryzacji zerowanej*. Warszawa: PWN.
- Łuniewska, M. & Tarczyński W. (2006). *Metody wielowymiarowej analizy porównawczej*. Warszawa: PWN.
- Making a difference for entrepreneurs*. World Bank, www.doingbusiness.org/reports/doing-business/doing-business-2011 (15.10.2012).
- Nermend, K. (2006a). A Synthetic Measure of Sea Environment Pollution. *Polish Journal of Environmental Studies*, Vol. 15, No. 4b, 127–129.
- Nermend, K. (2006b). Using average-variance representation in economic analyses. *Polish Journal of Environmental Studies*, Vol. 15, No. 4C, 123–126.
- Nermend, K. (2008a). Employing similarity measures to examine the development of technical infrastructure in Polish counties. *Folia Oeconomica Stetinensia*, Vol. 15, No. 7, 87–97.
- Nermend, K. (2008b). Zastosowanie rzutu wektora do budowy miernika syntetycznego. *Przegląd Statystyczny*, No. 3.
- Nermend, K. (2009). Vector Calculus in Regional Development Analysis. Series: *Contributions to Economics*, Springer.
- Panek, T. (2009). *Statystyczne metody wielowymiarowej analizy porównawczej*. Warszawa: SGH, 37–41.
- Pawełek, B. (2008). *Metody normalizacji zmiennych w badaniach porównawczych złożonych zjawisk ekonomicznych*. Kraków: Wydawnictwo UE w Krakowie.
- Social Watch Annual Report 2010: Time for a new deal after the fall*, Montevideo, Social Watch, www.socialwatch.org/sites/default/files/Social-Watch-Report-2010.pdf (17.10.2012).