Speeding problem detection in business surveys: benefits of statistical outlier detection methods

Berislav Žmuk [1,†]

[1]*Department of Statistics*, *Faculty of Economics and Business, University of Zagreb,*
J. F. Kennedy 6, 10 000 Zagreb, Croatia
E-mail: ⟨*bzmuk@efzg.hr*⟩

Abstract. Speeding describes the unusually fast responses provided to survey questions. A characteristic of speeders is that answers by-pass cognitive process. Consequently, this low respondent engagement results in the poor quality and validity of data.

The issue at hand is how to detect speeders in a survey. The presumption is the use of different statistical outlier detection methods. This paper presents graphical methods for outlier detection, such as: dot-plot diagrams, scatter diagrams, histograms and box-plot diagrams. Furthermore, the quantitative methods for outlier detection in this paper are the z-score, modified z-score, Dixons' test, Grubbs' test, Tietjen-Moore test, Rosners' or the generalized extreme studentized deviate (ESD) test. The performance of these outlier detection methods was observed on completion times of 217 surveys from enterprises which participated in a web survey on the use of statistical methods, and which use them in their business processes.

The analysis has shown that none of the observed outlier detection methods were able to detect speeders in an appropriate and satisfactory way as shown by the threshold. The main reasons for this can be found in slowers, the violations of assumptions on normal distribution and in masking. Hence, existing outlier detection methods should be improved and adjusted in future research in order to detect speeders. The introduction of novel speeders detection methods would be a good choice for future research.

Keywords: speeding, outlier detection methods, statistical method, web survey

1. Introduction

Worse than no answer is getting a wrong answer in a survey question. Here, a wrong answer is when a reply does not correlate to a respondent's attitude, thin-

---

† Corresponding author

king or facts. Such answers are considered invalid or incorrect. Web surveys are prone to such situation. The main reason is that these surveys are self-administered surveys (Bowling, 2005). Hence, no one directly observes respondents. Furthermore, no assistance from an interviewer is possible. Therefore, web surveys should be well designed and be interesting to respondents, not too long, nor be technically demanding (Fowler, Cosenza, 2008). The progress of respondents can be monitored using (para)data on survey processes and contexts. Paradata enables researchers to see what device respondents use in completing a survey, the time needed to complete the entire questionnaire and even time necessary to complete each question, the respondent's location when they took the survey, and similar data (Heerwegh, 2003, Kreuter, 2013). Paradata can be also used to study non-response bias (Durrant, D'Arrigo, Steele, 2011). The availability of such data depends on the software used to devise a web survey, and the actual design of the survey. Hence, questionnaire design is important.

Questionnaire design is also important in preventing respondents from satisficing (Krosnick, 1991). Satisficing is when respondents choose an answer solely for the purpose of satisfying the form or completing the questionnaire without cognitive effort (Krosnick, 2002, Lenzner, Kaczmirek, Lenzner, 2010). During this process, the likelihood that a respondent will provide an answer that is not valid and correct increases. The more questions in the questionnaire, the greater the likelihood of providing such answers. The problem of satisficing is in later question even more expressed (Galesic, 2006).

In web surveys, this satisficing is observable when a respondent completes a questionnaire too quickly (Zhang, Conrad, 2013). This is known as *speeding*. However, the question remains as to how to detect respondents - speeders. Speeders are respondents who complete questionnaires in significant less time than other respondents. Accordingly, speeders may also be treated as outliers. An outlier is a data value which is considerably and significantly different from other values (Acuña, Rodriguez, 2004, Ben-Gal, 2005). Outliers can have a significant impact on statistical analyses (Rasmussen, 1988). Generally speaking, outliers increase data variability levels which in turn reduce the power of statistical tests (Zimmerman, 1994). Furthermore, non-randomly distributed outliers disrupt the normal distribution of data, which in turn affects the probabilities of Type I and Type II errors in statistical tests (Peña, Pireto, 2001).

However, the biggest danger of outliers is their impact on statistical results which could lead to misleading conclusions. In the business world, decision making which is based on misleading conclusions can be very costly and devastating for an enterprise. Therefore, it is crucial that outliers be detected first and, subsequently, decide on how to treat them. Outliers might be just mistyped values, but they may also point to important factors in business processes. It then becomes important to careful delve into the source of outliers and the picture they are presenting. Only then is a decision made whether to keep or omit outliers in the

analysis. The basis for this is successful detection of real outliers. Accordingly, appropriate outlier detection methods are crucial.

Speeders also can be treated as outliers as the survey times are considerably lower compared to that of other respondents. Hence, the logical assumption is that outlier detection methods can help in detecting speeders. This is the research hypothesis of this paper. Previous studies have examined response times as a proxy for respondent burden with the assumption that faster response times are linked to more efficient questionnaire designs (Bassili, 1996, Draisma, Dijkstra, 2004, Tourangeau, Couper, Conrad, 2004). However, few studies have questioned the appropriateness of utilized speeder detection methods nor have they investigated the negative effects of speeding on data quality (Zhang, Conrad, 2013). Hence, the question of how to determine the speeding threshold remains open. Accordingly, this paper deals with the problem of determining the speeding threshold using different outlier detection methods.

It should be pointed out that the researcher has to be careful when classifying a respondent to as a speeder. In other words, a respondent may quickly fill out a questionnaire if reading fast or possessing a lot of experience as a respondent, one who has participated in numerous surveys and is familiar with survey processes. For this reason, a researcher should apply an appropriate outlier detection method. If not careful in performing outlier detection and selection, the research may haphazardly classify a respondent as a speeder despite that not being the case. Furthermore, some real speeders may be missed if the outlier detection and selection process is not taken seriously.

This paper will observe the issue of speeder detection for the case of business surveys where the respondents are actual employees in the respective enterprises. Respondents in business surveys exhibit different characteristics than other respondents. The main problem with business surveys is to convince respondents to set aside the appropriate time and participate properly in the survey. This would mean that they should take part in surveys during work hours, which might be a problem due to their job requirements. Hence, these respondents do not have the spare time and are forced to finish the questionnaire as fast as possible. Also, respondents in enterprises are often interrupted by telephone calls, e-mails and their colleagues. They usually have to take time out when answering surveys and continue later. All this means that they need more time to complete the questionnaire. Here arises the issue of those respondents who are relatively slow in completing web surveys. This has to be taken into account when conducting speeder detection.

This paper is organized as follows. After a brief introduction, where the research problem is presented, the second chapter provides an overview of the data used in the analyses. The third chapter treats graphical and quantitative outlier detection methods used in the paper. This is followed by the fourth chapter which provides an analysis and discussion. First descriptive statistics methods are used

in the analyses in order to gain insight into the data, and subsequently the selected graphical and quantitative outlier detection methods are applied and the results are compared and discussed. The fifth chapter provides the final conclusions and recommendations for further research.

2. Data

This paper treats data from a web survey on statistical methods used in Croatian enterprises. The survey was conducted in 2013, in which more than 600 Croatian enterprises participated (Žmuk, 2015). The length of questionnaire depended on whether the particular enterprise utilized statistical methods in their business processes. If so, the questionnaire was comprised 28 questions (including the filter question). On the other hand, if an enterprise did not use statistical methods, it was required to answer 11 questions in order to complete the questionnaire (including the filter question). For the purpose of this paper, only the times needed for completing the questionnaire by enterprises using statistical methods are observed. The sample include 217 such enterprises. Accordingly, 217 pieces of data on questionnaire completion times by enterprises were observed in the analysis.

3. Methods

Detecting speeders or enterprises that completed the questionnaire significantly faster than others will rely on using different outlier detection methods. These methods are divided into groups for graphical and quantitative outlier detection methods. The group relating to graphical outlier detection methods used in this paper are: dot-plot diagram, scatter diagram, histogram, and box-plot diagram. Quantitative outlier detection methods applied in the paper are: Chebyshev's inequality, z-score, modified z-score, Dixons' test, Grubbs' test, Tietjen-Moore test, Rosners' or the generalized extreme studentized deviate (ESD) test. Importantly, using these methods means that not only speeders but also slowers are detected.

The results of all these methods are then benchmarked to a threshold. The threshold is determined by assuming that an average respondent will require about 300 milliseconds per word in a question to completely understand the question, perform the thinking process and provide an answer (Zhang, Conrad, 2013). In that way, the efficacy of each outlier detection method will be determined and compared to other methods.

3.1. Graphical methods for outlier detection

Graphical methods techniques for outlier detection are based on a graphical approach to determining whether certain data points are anomalies. Sometimes this is not an easy task. At times, outliers are difficult to identify, and at other times it remains unclear whether a data point is in fact an outlier. Furthermore, a number of researchers observing the same graphs may result in a different number of outliers in the analysis. For this reason, graphical methods techniques for outlier detection are very subjective, and should be viewed as tools for preliminary outlier detection, which are subsequently verified by quantitative methods for outlier detection.

The dot-plot diagram is useful tool in outlier detection when a relatively small number of data points are observed. It can also be used for a large number of data, but in that case graphical representation might be complex and ambiguous. As its name suggests, the dot-plot diagram relies on dots. The vast majority of the dots are found together in a single group. However, some dots or data points are found at a greater distance from the group on the left or right side. These data points are treated as outliers.

A scatter diagram uses dots to present data points also in the form of a dot-plot diagram. However, scatter diagrams introduce another dimension of data, two dimensions of data. The technique of determining outliers therefore is somewhat different compared to the dot-plot technique. Here outliers are data points that do not follow a certain pattern of other data points, and outliers can also be found in any direction, as long as the distance of data points from the pattern is considerable.

Instead of using dots, histograms use columns but the principle is the same. Columns which are more distanced from other columns are suspect of containing outliers. In most cases, when a histogram is used, data are in forward by a researcher grouped into frequency distribution. In that case, research must be careful done in determining the number of classes and their width as such values have a strong impact on detecting real outliers.

The box-plot diagram originally consisted of a figure that showed the lowest and highest data points, and between them a rectangle was placed for which limits were determined by the value of lower and upper quartiles (Hoaglin, 2003). However, it is very difficult to detect outliers in this way. Hence, Tukey improved the box-plot by introducing the so-called fences based on the interquartile value multiplied by a certain value (Choonpradub, McNeil, 2005). In the meantime, there have appeared numerous versions of the box-plot diagram (Wickham, Stryjewski, 2011).

3.2. Quantitative methods and tests for outlier detection

After a preliminary analysis of outlier detection is conducted using graphical methods, the final conclusion on outliers is made using quantitative methods and a statistical test for detecting outliers. Chebyshev's inequality and the z-score methods are the best known classical quantitative methods for outlier detection. Both methods consider an outlier to be a data point lying far from the mean of all data points. However, Chebyshev's inequality is based on the confidence interval and is calculated as follows:

$$P\left(\overline{x} - ks < \mu < \overline{x} + ks\right) \le 1 - 1/k^2, k > 1 \tag{1}$$

where $\overline{x}$ is the sample mean, k is a number higher than one, and s is the sample standard deviation. According to Chebyshev's inequality, outliers are data points with values outside of the calculated interval. Still, given that Chebyshev's inequality method can be used for any data distribution, even if unknown, the main drawback is that it shows only the lowest, but not precise, data points share which could be found in the calculated interval (Seo, 2002). On the other hand, the z-score method inspects individual data points if they are outliers. Z-score values are calculated using the following equation:

$$z_i = \left(x_i - \overline{x}\right)/s, i = 1, 2, ..., n \tag{2}$$

where z_i is the z-score for an i-th data point, x_i is the original value of an i-th data point, $\overline{x}$ is the sample mean, s is the sample standard deviation and n is the sample size. Contrary to Chebyshev's inequality method, the z-score method is recommended only if data points are normally distributed, which is the main drawback of this approach. Furthermore, this method is not recommended for considerably small sample sizes, i.e. smaller than 10 units, as the danger of erroneous outlier detection is possible. On the other hand, some real outliers may possible be hidden, as their values are also taken into account when calculating the sample mean and sample standard deviation (Seo, 2012). Avoiding a situation where some outliers are hidden, requires calculating the modified z-score. Hence, the median may be used instead of the sample mean, and the median absolute deviation instead of the sample standard deviation (Habib, 2012). In that case, Equation 2 is transformed into:

$$M_i = \left[0.6745 \cdot \left(x_i - Me\right)\right]/MAD, i = 1, 2, ..., n \tag{3}$$

where M_i is a modified z-score for the i-th data point, x_i is the original value of an i-th data point, Me is the sample median, MAD is the median absolute deviation. The constant value of 0.6745 was added in Equation 3 to take into

account the fact that $E(MAD) = 0.6745\sigma$ for a large sample size (Iglewicz, Hoaglin, 1993).

Different statistical tests can be applied to detect outliers. There are a lot of statistical tests along with the different versions. However, they can be roughly divided depending on the conditions for which they are applied. Statistical tests for detecting outliers exist and which do not rely on knowing the distribution of data points or distribution of data parameters. There are also statistical tests for inspecting whether one or more outliers exist. In general, there are six basic types of statistical tests of outlier detection (Barnett, Lewis, 1983). The best known and most frequently used statistical tests for outlier detection are: Dixons' test, Grubbs' test, Tietjen-Moore test, Rosners' or the generalized extreme studentized deviate (ESD) test. This paper will utilize these tests. However, it must be emphasized that, generally speaking, almost all statistical tests for outlier detection have two major drawbacks. The first drawback is the fact that almost all statistical tests for outlier detection are appropriate only for univariate analysis, while in reality it would be better to observe outlier presence over several variables which requires multivariate analysis. Another drawback is that all tests assume a certain shape of data distribution, which is, in reality, rarely known.

Dixons' test of outlier detection is known also as Dixons' Q test because it uses the Q test statistical value (Dean, Dixon, 1951). The main advantage of Dixons' test is that it can be applied when sample size is very small, between 3 and 25 units, but it can also be applied for larger samples as well. Dixons' test is a univariate test used only if data is normally distributed. Furthermore, Dixons' test can be used to test the presence of only one outlier at a time. These characteristics of Dixons' test may seem as drawbacks but the main advantages, and the reason that Dixons' test is popular, are its convenience and ease of application. Namely, a Q test statistical value is calculated as a ratio of the difference between a potential outlier and the nearest data point and the difference between a potential outlier and the highest or the lowest data point. In order to avoid two or more outliers in row on the same side, what would result in decision that there are no outliers, test statistic is, depending upon sample size, calculated different (Thompson, Lowthian, 2011). As is the case with other statistical tests, test statistics is compared to theoretical values which are tabled and shown for a certain level of significance (see Rorabacher, 1991). If test statistics are lower than the theoretical value, the conclusion is that there are no outliers. To be more precise, regardless of what is tested, the conclusion is that the largest or smallest data point is not an outlier. If a researcher wants to verify the presence of more outliers in the sample, Dixons' test can be repeated for the next data points in the row. However, in that case, the power of the test is reduced and another statistical test for outlier detection should be considered.

Given that Dixons' test may result in a hidden outlier, sometimes preference is given to Grubbs' test or the maximum normed residual test (Hund, Massart,

Smeyers-Verbeke, 2000). Just like Dixons' test, Grubbs' test can also test whether only one outlier is exists at the time. Hence, Grubbs' test can be used to inspect whether the smallest or largest data point in the sample is an outlier. However, to get valid test results, the recommendation is a sample size larger than 6 (Dol, Verhoog, 2010). Grubbs' test is a univariate test and is used if data is normally distributed, but it can also be used on logarithmic values of data points (Grubbs, 1969). The null hypothesis of Grubbs' test contains the assumption that there are no outliers in the sample, whereas the alternative hypothesis holds the assumption that there is exactly one outlier. If there is doubt as to the existence of more outliers, another statistical test for outlier detection should be considered. After calculating the sample mean and sample standard deviation, G test statistics is calculated as follows:

$$G = \left[\max|x_i - \bar{x}|\right]/s, \tag{4}$$

$$G = \left(\bar{x} - x_{min}\right)/s, \tag{5}$$

$$G = \left(x_{max} - \bar{x}\right)/s, \tag{6}$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, x_{min} is the smallest value in the sample, x_{max} is the largest value in the sample, x_i is a data point value in the sample with the largest absolute difference from the sample mean. Equation 4 is used when a two-sided Grubbs' test is conducted for inspecting the presence of any outliers. On the other hand, Equations 5 and 6 are used for a one-sided Grubbs' test to inspect whether the smallest or largest value is an outlier. After calculating G test statistics, critical value is then calculated. The critical value for a two-sided Grubbs' test is as follows:

$$G^* = \left[(n-1)/\sqrt{n}\right] \cdot \sqrt{\left(t_{\alpha/(2n);n-2}\right)^2 / \left[n - 2 + \left(t_{\alpha/(2n);n-2}\right)^2\right]} \tag{7}$$

where n is the sample size, $t_{\alpha/(2n);n-2}$ is the theoretical value of the Students' distribution at a significance level of $\alpha/2n$ and for n-2 degrees of freedom. In case of a one-sided test, the significance level in Equation 7 is replaced with α/n. If G test statistic is higher than the G* critical value, the null hypothesis is rejected and the conclusion is that there is an outlier in the sample. In that case, the outlier is omitted and Grubbs' test is applied again as long as the conclusion is that there is no outlier in the sample.

The Tietjen-Moore test (Tietjen, Moore, 1972) represents a generalization of the Grubbs' test for the case of multiple outliers. Namely, in case testing for a single outlier, the Tietjen-Moore test is equivalent to the Grubbs' test (NIST/SE-MANTECH, 2012). However, contrary to Grubbs' test, the Tietjen-Moore test

can detect multiple outliers at the time. Still, the Tietjen-Moore test assumes approximately-normal distributed univariate data points. The emphasis is that for the Tietjen-Moore test, the number of suspected outliers must be specified in forward and exactly. Here, the normal probability plot can be helpful. The null hypothesis of the Tietjen-Moore test holds the assumption of no outliers, whereas the alternative hypothesis assumes the existance of k outliers. Calculate test statistics requires sorting the data points from the smallest to highest value. After that, the following test statistics can be calculated:

$$L_k = \frac{\sum_{i=1}^{n-k}\left(y_i - \overline{y}_k\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \quad , \tag{8}$$

$$L_k = \frac{\sum_{i=k+1}^{n}\left(y_i - \overline{y}_k\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \quad , \tag{9}$$

where y_i is the i-th largest value, $\overline{y}$ is the sample mean, $\overline{y}_k$ is the sample mean without k the largest values (Equation 8) or the sample mean without k the smallest values (Equation 9). If the aim is to test for the presence of outliers on both sides of the distribution, the first absolute value of differences between actual data points values and average values should be calculated. After that, data points are sorted again from the smallest to largest value, but according to the value of the differences. In this case, the test statistics is calculated using Equation 8. The test statistics value is between zero and one. If the test statistics is close to 1, the conclusion is that there are no outliers. On the other hand, a test statistics close to 0 suggests the presence of outliers. The Tietjen-Moore test is always a one-tailed test. Formulating a conclusion requires comparing the test statistics to the theoretical one which is determined by simulation. The simulation is based on 10,000 standard normal random samples of size n (NIST/SEMANTECH, 2012). A table with theoretical values can be found at ASTM International (2008).

Rosners' test or Generalized Extreme Studentized Deviate (ESD) test can be used to verify the presence of k outliers in the sample (Rosner, 1975). The k value must be chosen before starting the testing process. Rosners' test is always two-sided, meaning that it can be inspected if outliers are coinddently the smallest and highest values. However, this test can be applied only if a sample is larger than 24 (Gibbons, 1994). Furthermore, Rosners' test can be used only if data points are normally distributed. The test can be used not only on logarithmic data values but also only if normally distributed. After calculating the sample mean and

sample standard deviation, the value of a data point farthest from the sample mean must be determined. This data point is then used to calculate test statistics using following equation:

$$R_m = |x_m - \overline{x}_m| / s_m, m = 1,2,...,k \qquad (10)$$

where R_m is Rosners' test statistics for m-th data point, x_m is the value of the data point farthest away from the sample mean, $\overline{x}_m$ is the sample mean with the included m-th data point, s_m is the sample standard deviation with the included m-th data point, and k is the presumed number of outliers. After calculating the test statistics, the m-th data point is then omitted from further analysis and the procedure is repeated and additional k-1 times. Consequently, there is in all k test statistics values which should compared with critical values be one after the other, and which are calculated as follows:

$$R_{c(m)} = 2.295 + 0.02734n - 0.0002175n^2 - 0.03786m + 0.0009356nm -$$
$$- 0.0000793m^2 - 0.00006973nm^2 + 0.00000874m^3 + 0.000003943nm^3 \quad (11)$$
for $25 \le n \le 50$ *and* $1 \le m \le 10$

$$R_{c(m)} = 2.8613 + 0.006641n - 0.00001521n^2 - 0.003914m + 0.001207nm -$$
$$- 0.0006349m^2 - 0.00002405nm^2 + 0.00003133m^3 + 0.00000163nm^3$$
for $51 \le n \le 200$ and $1 \le m \le 10$ $\qquad (12)$

where n is starting sample size and m is the presumed number of outliers. Using Equations 11 and 12, the critical values at a significance level of 5% are calculated (McCuen, 2003), but these values can also be found in corresponding tables (Gilbert, 1987). The next step involves comparing test statistics and critical values in reserve order, which means that the last calculated value is first compared with the critical values. So, first is test statistics R_k compared with critical value $R_{c(k)}$. If test statistics R_k is greater than the critical value $R_{c(k)}$, the null hypothesis can be rejected and the conclusion is that the k-th value is an outlier. Moreover, it can also be concluded that all previous examined data points are also outliers. On other hand, if test statistics R_k is less than the critical value $R_{c(k)}$, the k-th value cannot be considered an outlier. In that case, the process of testing continues and test statistics R_{k-1} is compared with the critical value $R_{c(k-1)}$. This process continues until an outlier is found or the beginning of the series m is reached.

4. Analysis and discussion

4.1. Threshold and descriptive statistics analysis

The respective questionnaire contained a total of 28 questions on statistical methods used in enterprises which had been grouped into five question groups. These groups are shown in Table 1, together with number of questions in each group, number of words and number of signs for each question group.

Question group	Number of questions	Number of words	Number of characters (with no spaces)	Threshold in seconds		
				250ms per word	300ms per word	600ms per word
Filter	1	81	519	20.25	24.30	48.60
Statistical methods used in enterprises	8	390	2,528	97.50	117.00	234.00
Perception of statistical methods used in enterprise	10	458	2,836	114.50	137.40	274.80
Impact of statistical methods used	3	150	864	37.50	45.00	90.00
Enterprise and respondent	6	138	953	34.50	41.40	82.80
Total	28	1,217	7,700	304.25	365.10	730.20

Table 1: *Content of the questionnaire content and the defined threshold*

According to Table 1, there are in all 1,217 words in the questionnaire. If a threshold of 300 milliseconds is taken into account, it is estimated that an average respondent needs about 365 seconds (slightly more than 6 minutes) to complete the questionnaire. The vast majority of questions are a closed question type, hence this estimation is correct for the average respondent. However, there are some remaining issues in the questionnaire for which a respondent may provide a shorter or longer answer or comment. For this reason, an upper threshold of 600

milliseconds per word was introduced. In other words, if a respondent needed more than 730.20 seconds, they are considered to be a slower. On the other hand, experience and the ability of respondents to read and understand quickly should also be taken into account. This led to defining also a lower threshold of 250 milliseconds per word. So, if a respondent completes the questionnaire in less than 304.25 seconds, they are considered to be a speeder.

The analysis of outliers is conducted on the entire sample of those 217 enterprises that use statistical methods, but also on a strata level. In that way, a distinction is made between small, medium-sized and large enterprises. The enterprises are classified into these categories based on total value of assets, the amount of revenue and the average number of employees during the financial year. The critical limits for classification are taken from the Accounting Act (Official Gazette, 109/07). The basic descriptive statistics results of survey completion times for the entire sample and at a strata level are given in Table 2.

Enter-prises	Sample size	Ave-rage	Standard deviation	Coefficient of variation	Medi-an	Min	Max	Skew-ness
Small	196	784	453	58	652	278	2598	1.95
Medium-sized	12	794	390	49	687	333	1672	1.15
Large	9	1000	660	66	684	504	2299	1.42
All	217	794	459	58	654	278	2598	1.90

Table 2: *Basic descriptive statistics results of survey completion times, in seconds*

According to the results provided in Table 2, it becomes evident that the sample sizes between different strata vary. Accordingly, the performance of the observed outlier detection method for large and small samples is observable and comparable. Furthermore, the variation level for each observed category of enterprise seems to be exceptionally high. Consequently, none of the observed category of enterprises follows a normal distribution. It would be interesting to see how all these mentioned data characteristics affect detection outlier power for the observed outlier detection methods. Table 3 reveals that there are many more slowers than speeders in the sample. This is as expected due to fact that respondents in enterprises were often interru-pted by other employees, phones, e-mails and tasks, disruptions that prolong the time of completing the survey.

Enterprises	Speeders		Slowers	Total number of outliers
	Less than 250ms per word	Less than 300ms per word	More than 600ms per word	
Small	2	12	81	95
Medium-sized	0	1	5	6
Large	0	0	4	4
All	2	13	90	105

Table 3: *Number of enterprises that are speeders or slowers based on the thresholds*

The results in Table 3 will be used as benchmarks for other outlier detection methods.

4.2. Graphical outlier analysis

The first graphical method used for outlier detection is the dot-plot diagram. As the diagram's name suggests, it is based on dots. Each dot represents a data value. In this particular case, each dot represents an on-time completed questionnaire by an enterprise. The dot-lot diagrams for all enterprises together and for small, medium-sized, and large enterprises are shown separately in Figure 1.

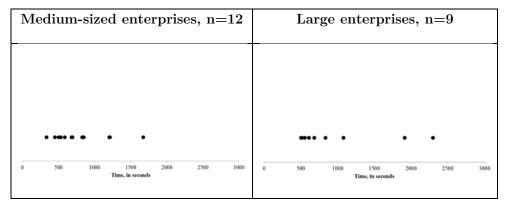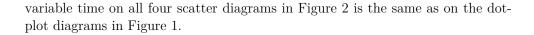| Medium-sized enterprises, n=12	Large enterprises, n=9

Figure 1: *Dot-plot diagrams for times needed to complete the questionnaire, in seconds*

For the purpose of comparing the four observed dot-plot diagrams in Figure 1, the x-axis is the same for all diagrams. All x-axes range from 0 to 3000 seconds. According to the dot-plots in Figure 1, outliers (here dots) not grouped together with other dots are easily detected. However, according to the dot-plots it seems that there are no speeders presents at any of the observed enterprises. Still, strictly speaking, the conclusion may be that there is one speeder if only medium-sized enterprises are observed. But this conclusion is somewhat questionable, and it could be further discussed. On the other hand, there are no such problems in detecting slowers. At all enterprise levels, the dot-plot suggests that there are 11 enterprises which are slowers and which required 1,794 seconds or more to complete the questionnaire. Those 11 enterprises are represented by the 11 dots to the right in Figure 1 and that seem to be distanced from the others dots on the left. Similar, for small enterprises, there are detected 9 dots distanced from the others. Those 9 slowers required 1,794 seconds or more to complete the survey. Because of the relatively smaller sample sizes, smaller number of slowers can were also detected at medium-sized and large enterprises. Thus, at medium-sized enterprises, 3 slowers were found, at enterprises that required 1,201 seconds or more to complete the questionnaire, whereas there were 2 slowers at large enterprises that required 1,914 seconds or more to complete the survey.

In Figure 2 shows scatter diagrams of questionnaire completion times for all four observed types of enterprises. On the x-axes enterprises according theirs ordinal number in the survey are shown. Here the horizontal position of the dots is not determined by the time needed to complete the questionnaire as was the case with the dot-plot diagrams, but the first dot represents the first enterprise that completed the questionnaire, the second dot refers to the second enterprise that completed the questionnaire, and so on. However, the vertical position of the dots is determined by the time required to complete the questionnaire. Thus, time in seconds is shown on the y-axis. Again, for comparison purposes, the range of the
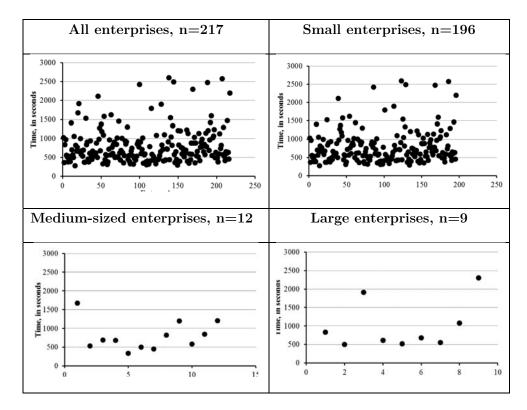
variable time on all four scatter diagrams in Figure 2 is the same as on the dot-plot diagrams in Figure 1.

Figure 2: *Scatter diagrams for times required to complete the questionnaire, in seconds*

Detecting speeders requires looking at the dots that are considerably closer to the x-axis than other dots. However, hardly any speeders can be detected as shown in the scatter diagrams in Figure 2. On the other hand, slowers are observable, but the problem is how to exactly ascertain which dots represent slowers and which do not. Meaning, the dots are in order of value and hence searching for slowers in the scatter diagrams it is a little bit ambiguous and confusing. The hardest way to determine slowers is across all enterprises, whereas for the other three levels it is obvious which dots represent slowers. Nevertheless, a possible conclusion is that the number of slowers across all enterprise, small enterprise and large enterprises, is the same as in the earlier analysis that was based on the dot-plot diagrams. As shown in the scatter diagram for medium-sized enterprises, it seems there is only one slower, i.e. one dot is positioned significantly above the others, and which required 1,672 seconds to complete the survey.

Histograms of questionnaire completion time for all four types of enterprises are given in Figure 3. The histograms are based on frequency distribution data where

data points are grouped into certain classes. The frequency shows the number of enterprises in each class. It was determined that class sizes be 200 seconds and a time of 200 seconds was chosen as the starting point for forming classes. Due to the different ranges of completion times, the number of classes may be different for the observed enterprise types. Consequently, for all enterprises and small enterprise, there are 11 classes in all, 8 classes for medium-sized enterprises, and 10 classes for large enterprises. If each class contains at least one frequency or one data point, the number of classes is equal to the number of columns on the histogram. However, if a class contains no data, there will be an empty space instead of a column on the histogram. That empty space can be used to find outliers more easily.
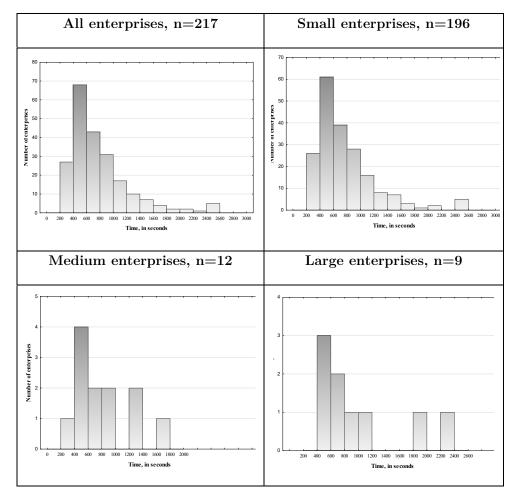

Figure 3: *Histograms for times needed to complete the questionnaire, in seconds*

As before, none of the histograms in Figure 3 reveal the presence of speeders. The first columns, containing enterprises with the lowest questionnaire completion times, are well connected to the other columns without any space between them. The heights of the first columns are decent in comparison to the heights of other columns, which shows that these columns contain a significant number of enterprises. However, for medium-sized enterprises, the first column has a height of only one, suggesting that there is one speeder for that type of enterprise. If slowers are observed, based on the histograms the conclusion is that they can be found in the last five classes across all enterprises and small enterprises. The columns representing those five last classes are either very low compared to the other columns or are separated from the other columns. Therefore, the conclusion is that there are 14 slowers, when observing all types of enterprises, and 11 slowers, when small enterprises observed. For medium-sized and large enterprises, the slowers are easier to detect due to quite a number of empty spaces or "zero" columns, thus making a distinction between the columns. Consequently, the number of slowers in medium-sized enterprises is three, whereas it is two for large enterprises.

Different versions of the box-plot diagrams can be used. Here is used the version which under outliers considers data points which values are either lower than median value reduced by 1.5 distance of the difference between the third quartile and the first one or higher than median value increased by 1.5 distance of the difference between the third quartile and the first one. Because of this quantitative background, detection of outliers using box-plots is a straightforward process. According to Figure 4, where box-plots are shown for all four observed enterprise types, there are no speeders. In addition, there are no slowers either in medium-sized nor in large enterprises. However, the box-plots shown in Figure 4 indicate that there are 12 slowers across all enterprises and 9 slowers in small enterprises.
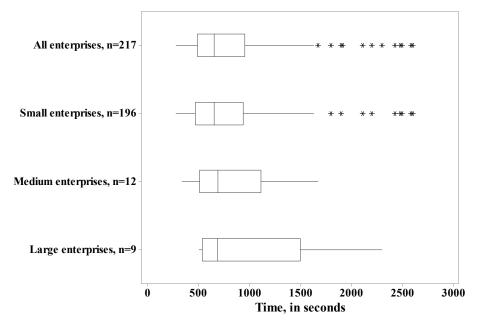
Figure 4: *Box-plots for times required to complete the questionnaire, in seconds*

If the results in Table 4 are compared to those in Table 3, the conclusion is that there are very large differences in the numbers of detected speeders and slowers among the observed graphical outlier detection approaches and threshold results.

Enter-prises	Outliers category	Graphical outlier detection approach			
		Dot-plot diagram	Scatter diagram	Histo-gram	Box-plot diagram
All	Number of speeders	0	0	0	0
	Number of slowers	11	11	14	12
	Total number of outliers	11	11	14	12
Small-sized	Number of speeders	0	0	0	0
	Number of slowers	9	9	11	9
	Total number of outliers	9	9	11	9

Medium	Number of speeders	1	0	1	0
	Number of slowers	3	1	3	0
	Total number of outliers	4	1	4	0
Large	Number of speeders	0	0	0	0
	Number of slowers	2	2	2	0
	Total number of outliers	2	2	2	0

Table 4: *Number of detected speeders and slowers using different graphical outlier detection approaches*

The main problem here is that the observed graphical outlier detection approaches were not able to detect speeders and separate them from the others. This problem was more prevalent in larger samples. In that case, enterprises with "normal" times, in some way, covered up the speeders, and consequently graphical outlier detection methods were not able to detect them clearly. Consequently, the conclusion is that graphical outlier detection methods are not sensitive enough to detect speeders.

4.3. Quantitative outlier analysis

In quantitative outlier analysis, Chebyshev's inequality method will not be observed separately, due to its imprecision stemming from the way it defines its inequality confidence interval. Therefore, quantitative outlier analysis will begin with the z-score and modified z-score approaches. The number of speeders and slowers based on these two outlier detection approaches are shown in Table 5.

Enter-prises	z-score				Modified z-score			
	No. of speeders		No. of slowers		No. of speeders		No. of slowers	
	z < -3	z < -2	z > 2	z > 3	M < -3	M < -2	M > 2	M > 3
Small	0	0	9	6	0	0	22	10
Medi-um-sized	0	0	1	0	0	0	2	1
Large	0	0	0	0	0	0	2	2
All	0	0	11	7	0	0	26	14

Table 5: *Number of outliers based on the z-score and modified z-score approaches*

Detecting outliers (speeders and slowers) required using two intervals for the both z-score and modified z-score approaches. In the first interval, speeders have a z-score or modified z-score less than -2, while the slowers have a z-score or modified z-score greater than 2. In the second interval, the limits were set to -3 and 3, respectively. According to the results in Table 5, both intervals for both observed outlier detection approaches resulted in no speeders. On the other hand, slowers were detected. Given that modified z-score account no normal data distribution, this approach resulted in a higher number of slowers compared to the number of slowers detected using the z-score approach. However, the number of detected slowers is far less than defined for the threshold provided in Table 3. As no speeders were detected using the z-score and modified z-score approaches, the conclusion is that their position and role in the process of detected speeders in the survey is questionable.

Table 6 shows Dixons' test results for outlier presence across all four observed types of enterprises. The Dixons' test can be used if only one outlier is present. Hence, the analysis was used to inspect either the smallest or largest values as outliers, respectively. Furthermore, Section 3.2 describes the core version of Dixons' test for outlier detection. However, Dixon emphasized that the larger samples from a normal population are selected the more likely is that extremes or outliers are included (Dixon, 1953). For that reason, Dixon introduced ratios, which should be used in line with sample sizes. If the sample size ranges from three to seven, the ratio r10 (original Dixons' test statistics) should be used. If the sample size varies from eight to 10, the ratio r11 is recommended. If the sample size ranges from 11 to 13, a ratio of r21 is recommended. If the sample size is greater than 13, ratio r22 should be used (Dixon, 1953). These ratios differ in that they take into account the difference between the second and the first data point, the third and first ones, the penultimate and last ones, or it can observe the difference between the data before the penultimate and the last ones (Rorabacher, 1991). According to these recommendations, different Dixons' ratios were used in the analysis.

Enter-prises	Tested value	Sample size	Dixon's ratio	Test statistics	p-value
All	Minimum	217	r22	0.01	0.981
	Maximum	217	r22	0.05	0.804
Small	Minimum	196	r22	0.01	0.982
	Maximum	196	r22	0.05	0.813
Medium-sized	Minimum	12	r21	0.19	0.737
	Maximum	12	r21	0.38	0.266
Large	Minimum	9	r11	0.01	0.977
	Maximum	9	r11	0.22	0.442

Table 6: *Dixons' outlier tests results for the observed enterprises levels*

The null hypothesis of Dixons' test is based on the assumption that all data values come from the same normal population, whereas the alternative hypothesis assumes that the smallest or the largest (in dependence which one is tested) data point value is an outlier. According to the results in Table 6, at any usually used significance levels, the conclusion is that neither the lowest nor highest values at each of the four types of observed enterprise are outliers. In other words, Dixons' test failed to detect any speeders or slowers. The results are astonishing, especially given that all previous observed outlier detection methods at least pointed out the presence of slowers. There could be two reasons for this. The first is that the observed data distributions do not exhibit a normal distribution (see Table 2), which could have impact on Dixons' test results. The second reason could be the presence of more than just one outlier, which only Dixons' test is capable of detecting, and there the smallest or largest values cannot be declared outliers.

The conducted Grubbs' outlier test results for the observed enterprise types are shown in Table 7. The Grubbs' test can also detect the presence of just one outlier in the dataset and therefore it was the test of choice for the smallest or largest values as outliers, respectively. The contents of the hypotheses for Grubbs' test are identical to the contents of hypotheses for the Dixons' test. According to the results in Table 7, Grubbs' test results for any frequently used significance levels show that the smallest data point value cannot be considered an outlier. Consequently, Grubbs' tests show that there are no speeders. However, at a significance level of 1%, Grubbs' tests show that the largest values across all enterprises and small enterprise types are outliers. Furthermore, at a significance level of 10%, the conducted Grubbs' test shows that the largest value in medium-sized enterprises can be considered an outlier. Finally, at a significance level of 15%, Grubbs' test points out that the largest value for large enterprises can be considered an outlier. Therefore, regardless of assuming a violation to the normal

distributed data, Grubbs' test was able to point out the presence of slowers in the dataset.

Enterprises	Value	G test statistics	p-value
All	Minimum	1.12	1.000
	Maximum	3.93	0.007
Small	Minimum	1.12	1.000
	Maximum	4.01	0.004
Medium-sized	Minimum	1.18	1.000
	Maximum	2.25	0.058
Large	Minimum	0.75	1.000
	Maksimum	1.97	0.104

Table 7: *Grubbs' outlier test results for the observed enterprises*

The Tietjen-Moore test is used to detect more than just one outlier. In the analysis, a separate test was conducted to see whether the smallest data point values were outliers (speeders) and whether the largest data point values were outliers (slowers). In both cases, it is assumed that one-tail statistical test on the lower limit is conducted where in the null hypothesis is the assumption that there are no outliers whereas in the alternative hypothesis the assumption of certain number of outliers presence is assumed. The analysis was conducted so that the number of observed outliers were increasing until the null hypothesis could no longer be reject for all four observed significance levels (10%, 5%, 2.5%, and 1%). The results are shown in Table 8.

Enterprises	Outliers	Significance level			
		10%	5%	2.5%	1%
All	Number of speeders	0	0	0	0
	Number of slowers	>31*	>31*	>31*	>31*
Small	Number of speeders	0	0	0	0
	Number of slowers	>31*	>31*	>31*	>31*
Medium-sized	Number of speeders	0	0	0	0
	Number of slowers	5	5	3	0
Large	Number of speeders	0	0	0	0
	Number of slowers	3	3	3	3

*The statistical software stopped working if the number of tested outliers was set to more than 31 (NIST, 2013).

Table 8: *Tietjen-Moore outlier test results for the observed enterprises*

According to the results in Table 8, it becomes obvious that even the Tietjen-Moore test was unable to detect the presence of speeders at any of the four observed significance levels. On the other hand, the detected number of slowers were close to the number of given thresholds. However, this is true for medium-sized and large enterprises. Unfortunately, for all enterprises and small enterprises, a full comparison with the threshold cannot be done due to limitation of the statistical software, but it was assumed that the final number of slower's would be nearly identical to the number of thresholds.

The null hypothesis of the Rosners' test contains the assumption that there are no outliers, whereas the underlying assumption of the alternative hypothesis is the existence of a certain number of outliers. The test is a two-sided test and observes absolute differences. Hence, it was necessary to observe which data point values were included in the analysis. This enables us to determine the number of observed smallest data point values (speeders) and largest data points (slowers), respectively. The results of the conducted Rosners' tests or Generalized Extreme Studentized Deviate (ESD) tests are given in Table 9.

Enterprises	Outliers	Significance level			
		10%	5%	2.5%	1%
All	Number of speeders	0	0	0	0
	Number of slowers	10	8	8	6
Small	Number of speeders	0	0	0	0
	Number of slowers	8	7	7	7
Medium-sized	Number of speeders	0	0	0	0
	Number of slowers	0	0	0	0
Large	Number of speeders	0	0	0	0
	Number of slowers	2	2	2	0

Table 9: *Rosners' outlier test results for the observed enterprises levels*

According to the results in Table 9, neither was Rosners' test capable of detecting speeders. Some slowers were detected using Rosners' tests. But taking into account the threshold results, the performance of the Rosners' tests is worse than the Tietjen-Moore tests.

5. Conclusions

Speeding in surveys is a significant issue and is becoming a growing problem with the increasing popularity and use of web surveys. The main problem posed by speeding is that respondents fail to invest ample time in completing surveys, and as a consequence reduces time for cognitive processes. Consequently, this may

lead to respondents providing wrong answers not in line with his attitudes and beliefs.

However, the question remains as to how to identify speeders. The assumption in this paper was that this can be done using different outlier detection methods. Accordingly, the aim of the paper was to select the most precise outlier detection method from the observed outlier detection method. The graphical outlier analysis has shown that the dot-plot diagram, scatter diagram, histogram and box-plot diagram were able to detect outliers. However, they were able to detect only slowers but not speeders. A similar problem with detecting speeders relied on applying quantitative outlier detection methods. They too were able to detect slowers, but detecting speeders was a lot more difficult. Unfortunately, it has been shown that all observed outlier detection methods had problems in detecting speeders. This was due to their inadequate sensitivity and were thus unable to recognize speeders in appropriate way either for large or small samples. Therefore, the conclusion is that all observed outlier detection methods are ineffective in detecting speeders.

Three possible reasons exist as to why the respective outlier detection methods exhibited poor performance in detecting speeders. The first one reason can be found in slowers who have taken full attention of outlier detection methods. The second reason for the poor performance in detecting speeders is the fact that the observed outlier detection methods assume, more or less, that data is normally distributed, which here was not the case. Hence, one of the assumptions for their application has been violated. On the other hand, if the data contains speeders, then the expectation is that the normality assumption has been violated. The third reason could possibly be the large number of speeders in the distribution. This meant that they had masked each other, meaning that the observed outlier detection methods were unable to carry out detection .

Given the above, future research should incorporate additional efforts in finding a method that detect speeders precisely and reliably. To detect speeders properly, the existing outlier detection method should be improved significantly. However, perhaps the development of new specialized techniques or methods for detecting speeders is the best solution.

Acknowledgement

References

[1] Acuña, E. and Rodriguez, C. (2004). On detection of outliers and their effect in supervised classification. Available at: http://academic.uprm.edu/~eacuna/vene31.pdf [Accessed 21/05/16].

[2] ASTM International. (2008). Standard practice for dealing with outlying observations. Available at: http://www.astm.org/FULL_TEXT/E178/HTML/E178.htm [Accessed 10/06/16].

[3] Barnett, V. and Lewis, T. (1983). Outliers in statistical data. Chichester: John Wiley & Sons.

[4] Bassili, J. N. (1996). The how and why of response latency measurement in telephone surveys. In Schwarz, N. and Sudman, S. (Eds.). Answering questions: Methodology for determining cognitive and communicative processes in survey research. San Francisco: Jossey-Bass.

[5] Ben-Gal, I. (2005). Outlier detection. In Maimon, O. and Rockach, L. (Eds.). Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers (pp. 131–146). New York: Springer.

[6] Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health, 27(3), 281–291.

[7] Choonpradub, C. and McNeil, D. (2005). Can the box plot be improved? Songklanakarin Journal of Science and Technology, 27(3), 649–657.

[8] Dixon, W. J. (1953). Processing data for outliers. Biometrics, 9(1), pp. 74–89.

[9] Dean, R. B. and Dixon, W. J. (1951). Simplified statistics for small numbers of observations. Analytical Chemistry, 23(4), 636–638.

[10] Dol, W. and Verhoog, D. (2010). Methodology for data quality management. Available at: http://www.ilr.uni-bonn.de/agpo/rsrch/caprird/docs/d2.3.5.pdf [Accessed 28/05/16].

[11] Draisma, S. and Dijkstra, W. (2004). Response latency and (para)linguistic expressions as indicators of response error. In Presser, S. et al. (Eds.). Methods for testing and evaluating survey questionnaires. New York: Wiley.

[12] Durrant, G.B., D'Arrigo, J. and Steele, F. (2011). Using field process data to predict best times of contact conditioning on household and interviewer influences. Journal of the Royal Statistical Society: Series A (Statistics in Society), 174(4), 1029–1049.

[13] Fowler, F. J. and Cosenza, C. (2008). Writing effective questions. In de Leeuw, E. D., Hox, J. J. and Dillman, D. A. (Eds.). International Handbook of Survey Methodology (pp. 136-160). New York: Psychology Press, Taylor & Francis.

[14] Galesic, M. (2006). Dropouts on the Web: Effects of interest and burden experienced during an online survey. Journal of Official Statistics, 22(2), 313–328.

[15] Gibbons, R. D. (1994). Statistical methods for groundwater monitoring. New York: John Wiley & Sons.

[16] Gilbert, R. O. (1987). Statistical methods for environmental pollution monitoring. New York: John Wiley & Sons.

[17] Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. Technometrics, 11(1), 1–21.

[18] Habib, E. A. E. (2012). Mean absolute deviation about median as a tool of explanatory data analysis. International Journal of Research and Reviews in Applied Sciences, 11(3), 517–523.

[19] Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a Web Survey. Social Science Computer Review, 21(3), 360–373.

[20] Hoaglin, D. C. (2003). John W. Tukey and data analysis. Statistical Science, 18(3), 311–318.

[21] Hund, E., Massart, D. L. and Smeyers-Verbeke, J. (2000). Inter-laboratory studies in analytical chemistry. Analytica Chimica Acta, 423(2), 145–165.

[22] Iglewicz, B. and Hoaglin, D. C. (1993). How to detect and handle outliers. ASQC Quality Press.

[23] Kreuter, F. (Ed.) (2013). Improving surveys with paradata: Analytic uses of process information. New York: John Wiley & Sons.

[24] Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology, 5(3), 213–236.

[25] Krosnick, J. A. (2002). The causes of no-opinion responses to attitude measures in surveys: They are rarely what they appear to be. In Groves, R. M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (Eds.). Survey Nonresponse. New York: John Wiley & Sons.

[26] Lenzner, T., Kaczmirek, L. and Lenzner, A., (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. Applied Cognitive Psychology, 24(7), 1003–1020.

[27] McCuen, R. H. (2003). Modeling hydrologic change: statistical methods. Boca Raton: CRC Press.

[28] NIST/SEMATECH. (2012). e-Handbook of statistical methods: 1.3.5.17.2. Tietjen-Moore Test for Outliers. Available at: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h2.htm [Accessed 09/06/16].

[29] NIST. (2013). Dataplot. Available at: http://www.itl.nist.gov/div898/software/dataplot/homepage.htm [Accessed 16/06/16].

[30] Official Gazette. (2007). The accounting act. 16(109), Zagreb: Narodne novine d.d., 8130–8137.

[31] Peña, D. and Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. Technometrics, 43(3), 286–310.

[32] Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis D squared and Comrey Dk. Multivariate Behavioral Research, 23(2), 189–202.

[33] Rorabacher, D. B. (1991). Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level. Analytical Chemistry, 63(2), 139–146.

[34] Rosner, B. (1975). On the detection of many outliers. Technometrics, 17(2), 221– 227.

[35] Seo, S. (2002). A review and comparison of methods for detecting outliers in univariate Data Sets. Available at: http://d-scholarship.pitt.edu/7948/1/Seo.pdf [Accessed 28/05/16].

[36] Tietjen, G. L. and Moore, R. H. (1972). Some grubbs-type statistics for the detection of outliers. Technometrics, 14(3), 583–597.

[37] Thompson, M. and Lowthian, P. J. (2011). Notes on statistics and data quality for analytical chemists. London: Imperial College Press.

[38] Tourangeau, R., Couper, M. P. and Conrad, F. G. (2004). Spacing, position, and order: interpretive heuristics for visual features of survey questions. Public Opinion Quarterly, 68(3), 368–393.

[39] Wickham, H. and Stryjewski, L. (2011). 40 years of boxplots. Available at: http://vita.had.co.nz/papers/boxplots.pdf [Accessed 27/05/16].

[40] Zhang, C. and Conrad, F. G. (2013). Speeding in web surveys: the tendency to answer very fast and its association with straightlining. Survey Research Methods, 8(2), 127–135.

[41] Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and nonparametric tests. Journal of General Psychology, 121(4), 391–401.

[42] Žmuk, B. (2015). Adoption and benefits of statistical methods in enterprises: differences between Croatian regions. South East European Journal of Economics and Business, 10(1), 55–65.