

OPTIONS OF EVALUATION OF IDENTIFICATION QUESTIONS IN MARKETING RESEARCH

[Světlana Myslivcová, Otakar Ungerman]

Abstract—Marketing research is one of the basic methods how to identify changes in economy. Many methods are used to collect data and these methods are accompanied by the identification questions. These identification questions aim to categorize respondents according to selected criteria. Sociodemographic characteristics are used in the research of individuals, whereas characteristics identifying companies are used in the research of legal entities. The theoretical part of this paper deals with explanation of the role of the identification question in the research and with statistical methods used to evaluate these questions. First, ways how to process data using descriptive statistics are introduced. The second main part deals with the data processing using mathematical statistics, testing of statistical hypothesis. In the following practical part the quantitative research made using questionnaire is presented. The aim of the evaluation is to verify a hypothesis of existence of statistically significant differences between groups of respondents who are divided into groups according to the identification questions. The statistical evaluation was made using several different tests. Advantages and disadvantages of these tests are specified after the comparison. This paper is of benefit to marketers advising on how to deal with identification questions)

Keywords—identification question, marketing research, mathematical statistics)

I. Introduction

Market knowledge is currently a prerequisite for companies that want to enjoy long-term success. To record changes in the market environment, marketing research, which can have a qualitative or quantitative form, is used. Especially with quantitative methods of data collection, the goal is to obtain information from as many respondents as possible. This group of respondents is often very heterogeneous, so researchers attempt to divide respondents according to certain parameters into groups matching the given characteristics. For this distribution of respondents additional questions are used, which most often appear in the literature under the name of "identification questions".

Otakar Ungerman
Technical University of Liberec,
Czech Republic

Světlana Myslivcová
Technical University of Liberec,
Czech Republic

These identification questions supplement the basic set of questions that address the identified problem. In order to subsequently identify questions included in the evaluation, they must be properly defined. If the identification questions are properly defined, it is necessary to determine how they affect the research results, i.e. to find out if the responses of one group of respondents are different from those of another. To accurately determine how significant the difference between the groups is, the methods of mathematical statistics are used under the name "statistical hypothesis testing." The selection of the exact method is always dependent on the type of stated hypothesis, the size of the test set and the type of distribution. When evaluating the role of identification questions in marketing research, three assessments are most often used: the Pearson's chi-squared test, the Wilcoxon test and one-way analysis of variance. The resulting acceptance or rejection of the null hypothesis is different due to various test criteria. This paper aims to identify the similarities and differences in testing with the null hypothesis, resulting from the identification questions. This paper presents testing of non-parametric hypotheses that arose from the conducted primary research.

II. Literature Overview

Identification questions, sometimes also called analytic questions in the literature [1], classify and identify the most important features of the basic set. In their work the authors, Ivanová and Olecká [2], illustrate the importance of identification questions in connection with operationalization, the achievement of the highest possible reliability of the research. In connection with the identification questions, the term indicator is used, which is a parameter that characterizes the objectively observable indication of certain characteristics. The most common indicators in the division of natural persons are demographic characteristics, namely age, sex, family, education and income. There are various indicators for research purposes in industrial markets. The most important indicators for classifying businesses are industry classification, number of employees and ownership form [1]. Overall it can be summarized that if the basic set of data is not completely homogeneous, in the data collection process identification questions must be used to divide the respondents into smaller albeit homogeneous groups.

III. Methodology and goals

The methodology is divided into two parts. The first part is focused on case studies and the second part on the methods of evaluation:

A) A case study is one of the qualitative research methods and ranks among the non-experimental methods. Most often these are methods that are used in research in the

social sciences, which include marketing. Case studies explain the problems and phenomena both in the past and in the present, which occurred in one organization or across a group of organizations [3]. In this paper the most important enterprises in the Liberec Region are surveyed. The purpose of the research was to find formulas of corporate behaviour. The primary research consisted of quantitative data collection, which aims to determine the frequency of the behaviour of the analysed subjects. [4]. For the quantitative research, the method of simple random sampling was chosen, where each unit of the basic set has the same probability of being selected [5]. For the quantitative research, electronic polling was used, which followed a previous personal or telephone agreement.

B) The methods of data analysis were derived from the research goals and the type of data analysed. The data were analysed using the statistical program Statgraphics 16. The methods of evaluation can be divided into two groups:

- Descriptive statistics identifies and summarizes the information, processes it in the form of tables and graphs, and calculates their numerical characteristics. In the research methods of processing of mean values are used: average, variance, confidence interval and median.

- Mathematical statistics – testing of statistical hypotheses is a decision-making process, which on the basis of the results obtained from the random sample objectively prescribes a decision on whether the tested hypothesis should be rejected [6]. When testing statistical hypotheses, a null (tested) hypothesis, which is indicated as H_0 , is distinguished from an alternative hypothesis, which is indicated as H_1 . A null hypothesis is a hypothesis about which the test has to decide whether to reject it or not. An alternative hypothesis is the hypothesis that is accepted if the null hypothesis is rejected [7]. The testing of statistical hypotheses has clearly established procedures.

1. Formulation of research questions as null and alternative statistical hypotheses

2. Selection of acceptable level of erroneous decision-making (choice of level of significance)

3. Choice of test criteria

4. Calculation of the value of the test criteria

5. Determination of critical values of the test criteria

6. Recommendations – acceptance or rejection of the null hypothesis H_0

When choosing the level of significance, $\alpha=0.05$ (0.1, 0.025) is most commonly used, to achieve 95% (90%, 97.5%) certainty of correct decisions. The choice of test criteria follows, to which the values from the research are achieved. After calculation the result is compared with a critical value and recommendations are made. To determine the strength of dependence, the p-value is calculated, which is the lowest level at which the hypothesis can be rejected. In this paper three types of tests were used:

A. Wilcoxon nonparametric test

The Wilcoxon nonparametric test is used when an adequate sample size cannot be ensured or the normality of the data cannot be verified [7]. The Wilcoxon test is used for one choice and is a test based on the median. Its only

prerequisite is symmetrical division of the random variable X , from which comes a random selection. Test criteria.

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (1)$$

B. Analysis of variance

Analysis of variance (ANOVA) is a method which tests whether the variable Y (continuous) depends on the variable X (categorical). For an explanatory variable, that acquires "to" the categories and is also indicated as a factor, the name "one-way analysis of variance" is used [8]. Analysis of variance defines the overall variability of the variables Y as the sum of intragroup and intergroup variability of this variable, wherein the individual groups are formed according to the categories of factor X . When comparing the components of variance the fact that the value of the denominator is always the same (n) is exploited. It suffices therefore to compare only the value in the numerator, i.e. the sums of squared deviations. Test criteria:

$$\begin{aligned} SS_W &= \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \\ SS_B &= n \cdot \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 \\ SS_{TOTAL} &= SS_W + SS_B \end{aligned} \quad (2)$$

C. Pearson's chi-squared test

Pearson's chi-squared test is used to examine variability of dispersion in probability considerations. The chi-squared test is a statistical non-parametric method that is used to determine whether there is a proven strong relationship between two indicators. The test is often used to verify hypotheses in a contingency table. The calculation is based on a comparison of the expected frequency and the actual frequency using the formula:

$$\frac{(n_1 - n_2)^2}{n_1 + n_2} \geq X_{\alpha,1}^2 \quad (3)$$

This methodology is used to fulfil the main goal: To verify in three ways the hypothesis of the existence of statistically significant differences between groups of respondents, which are divided according to identification questions.

IV. Evaluation Of Case Studies

For the purposes of further evaluation is first presented a case study on the topic "Use of personnel marketing in the Liberec Region". The basic group for primary data collection consisted of major industrial enterprises in the Liberec Region, of which there were 102. The sample group included 58 fully completed and checked questionnaires, the rate of return was 56.9 %. Despite the high return, the number of responses was small. The limited number of responses was due to the fact that in the Liberec Region there are only a limited number of major companies. The

evaluation included one scale-question, where one meant maximum importance and six meant maximum unimportance. This question is confronted with an identification question, which is the size of the enterprise. The identification question divided the respondents into two groups. The first group consisted of medium-sized companies, which represented 26.4 % of the sample. The second group consisted of large enterprises, which represented 73.6 % of the sam-ple. The division was based on the number of employees in the company, where according to an EU metric the border is 250 employees.

A. Descriptive statistics

Descriptive statistics quantitatively describe the main characteristics of the measurement results. The aim is to concisely depict in numbers essential information on measured data. The basic characteristics include average, confidence interval, standard deviation (sd) and median, which describe the mean value of the group. The resulting values resulting from the research are pre-sented in Table 1.

TABLE I. DESCRIPTIVE STATISTICS

Attribute	Company size	\bar{x}	sd	Med-ian
1. Analysis of competitors' recruitment strategy	medium	3.8	1.619	3
	large	3.3	1.090	3
2. Research aimed at identifying decision-making criteria of candidates	medium	3.4	1.878	3
	large	3.3	1.614	3
3. Career pages for display on mobile phones	medium	4.5	1.581	4.5
	large	3.6	1.690	4
4. Career profiles on social networks	medium	3.8	1.549	4
	large	2.5	1.703	2
5. Incorporating game elements into the recruitment process and communication	medium	4.5	1.509	4.5
	large	3.4	1.577	3
6. Recruiting at universities	medium	3.8	2.150	3.5
	large	2.0	1.089	2
7. Targeted recruitment campaigns for specific positions	medium	2.2	1.922	1
	large	1.6	0.934	1
8. Independent job websites	medium	4.1	1.912	4
	large	2.4	1.289	2
9. Plan to build corporate culture of the employer	medium	3.2	1.932	2.5
	large	2.0	0.964	2
10. Video presentation of corporate culture	medium	4.8	1.229	5
	large	3.1	1.274	3
11. Welcome package for new employees	medium	2.9	1.853	2
	large	2.6	1.660	2
12. Tools supporting employee satisfaction	medium	2.4	1.776	1.5
	large	1.5	0.693	1
13. BigData	medium	4.9	1.287	5.5
	large	4.1	1.349	4

If the values are regarded as ordinal variables, a tool for evaluating differences between enter-prises of different sizes is the median. The distribution of scale response to the intervals can be presented so that the effective attributes in the interval <1;2>, neutral in interval <3;4> and inef-fective then have a median in interval <5;6>. From the results it can be concluded that the attrib-utes 4, 8 and 10 are assessed differently by each group of respondents. For further evaluation, the variables are further regarded as cardinals

and the median values are the basis for statistical hypothesis testing.

B. Mathematical statistics

Mathematical statistics is a scientific discipline related to descriptive statistics. It analyses data derived from empirical studies that contain an element of randomness. Mathematical statistics incorporates statistical hypothesis testing, which can also be included in applied mathematics.

A presentation of methods follows in which the hypothesis is determined as non-parametric. Three tests were used that are most frequently employed in marketing research for identification questions. For all methods, only the p-value is presented, which is the lowest possible level of significance at which the null hypothesis can still be rejected. The reason is that the p-value gen-erally provides more information about the outcome of statistical analysis than mere rejection or non-rejection of the null hypothesis. During testing the significance level $\alpha=0.05$ is used. Subse-quently, this level is supplemented by two variants higher $\alpha=0.1$, and lower $\alpha=0.025$. The same tested hypothesis is set for all methods, which is H_0 : among the addressed companies, which are divided by number of employees, there are no statistically significant differences in the answers.

1) Wilcoxon test

This test is in many cases associated with a limited sample size. It is also applicable in cases where normality cannot be unequivocally determined. The null hypothesis of the one-sample Wilcoxon test applies to the median. Table 2 shows the p-values for the conclusions of the Wil-coxon non-parametric test of conformity of average values for the size of the companies.

TABLE II. RESULTING VALUES OF WILCOXON'S TEST

Attribute	Company size	Wilcoxon test = p-value
1. Analysis of competitors' recruitment strategy	medium	0,600
	large	
2. Research aimed at identifying decision-making criteria of candidates	medium	0,800
	large	
3. Career pages for display on mobile phones	medium	0,167
	large	
4. Career profiles on social networks	medium	0,031
	large	
5. Incorporating game elements into the recruitment process and communication	medium	0,082
	large	
6. Recruiting at universities	medium	0,018
	large	
7. Targeted recruitment campaigns for specific positions	medium	0,529
	large	
8. Independent job websites	medium	0,016
	large	
9. Plan to build corporate culture of the employer	medium	0,091
	large	
10. Video presentation of corporate culture	medium	0,002
	large	
11. Welcome package for new employees	medium	0,670
	large	
12. Tools supporting employee satisfaction	medium	0,167
	large	
13. BigData	medium	0,121
	large	

The analysis showed that at a significance level of $\alpha=0.05$ there is a statistically significant difference in the evaluation of the effectiveness attributes of 4, 6, 8 and 10 between groups of respondents from medium and large enterprises. For these attributes H1 was adopted: among the addressed companies, which are divided by number of employees, there are statistically significant differences in the responses.

2) One-way analysis of variance

The second method for evaluating the impact of company size on the research results was one-way analysis of variance (ANOVA). It is a simple method based on analysis of variance, in which the effect of one factor on the investigated dependent variable is analysed. Basic statistics computed in the analysis of variance is the calculation of F-statistics in the analysis of variance, which can be written in a general form as the ratio of variance between groups and the variance within the group. The results are presented in Table 3.

TABLE III. RESULTING VALUES OF ANOVA TESTS

Attribute	Company size	
	Statistic F	HV
1. Analysis of competitors' recruitment strategy	1,091	0,303
2. Research aimed at identifying decision-making criteria of candidates	0,073	0,789
3. Career pages for display on mobile phones	1,999	0,166
4. Career profiles on social networks	4,410	0,043
5. Incorporating game elements into the recruitment process and communication	3,341	0,076
6. Recruiting at universities	11,677	0,002
7. Targeted recruitment campaigns for specific positions	1,953	0,171
8. Independent job websites	9,530	0,004
9. Plan to build corporate culture of the employer	6,647	0,014
10. Video presentation of corporate culture	13,795	0,001
11. Welcome package for new employees	0,167	0,686
12. Tools supporting employee satisfaction	5,615	0,023
13. BigData	2,655	0,113

If $HV \leq 0.05$, it can be concluded that the answers to the question, depending on the size of the company, are different in a statistically significant way. It can be stated that the variance analysis confirmed previous findings for attributes 4, 6, 8 and 10. In addition, differences in the assessment of attributes 9 and 12 were identified as statistically significant. In these six attributes H1 as adopted: among the addressed companies, which are divided by number of employees, there are statistically significant differences in the responses.

3) Pearson's chi-squared test

The third way to evaluate the effect of identification questions for research is Pearson's chi-squared test, which is the fundamental and most widely used test of independence by means of a contingency table. It is a non-parametric statistical method, which consists in comparing observed and expected frequencies. The condition for the realization of ($n > 30$) is met; the critical value amounted to $\chi(1-\alpha)$; $df =$

11.07. Evaluation of the Pearson chi-squared test is presented in Table 4.

TABLE IV. RESULTING VALUES OF PEARSON'S CHI-SQUARED TEST

Attribute	Test criteria (G)	p-value
1. Analysis of competitors' recruitment strategy	7.073	0,215268544
2. Research aimed at identifying decision-making criteria of candidates	3.82	0,575580001
3. Career pages for display on mobile phones	4.069	0,539480167
4. Career profiles on social networks	10.14	0,071372982
5. Incorporating game elements into the recruitment process and communication	5.253	0,385773397
6. Recruiting at universities	15.992	0,006867054
7. Targeted recruitment campaigns for specific positions	4.431	0,489171327
8. Independent job websites	11.327	0,045260871
9. Plan to build corporate culture of the employer	10.222	0,069190771
10. Video presentation of corporate culture	12.062	0,033943797
11. Welcome package for new employees	4.395	0,49401597
12. Tools supporting employee satisfaction	9.437	0,092854593
13. BigData	2.715	0,743858634

From the results it follows that at significance level $\alpha=5\%$ there are statistically significant differences in three attributes. After performing Pearson chi-squared test attributes 6, 8 and 10 had a final p-value below the critical level. All three attributes already emerged in the previous two evaluations. For these three attributes H1 was adopted: among the addressed companies, which are divided by number of employees, there are statistically significant differences in the responses.

C. Method comparison

Overall, a statistically significant difference was demonstrated in mean values when evaluating attributes 6, 8 and 10 among a group of large enterprises and a group of medium-sized enterprises. This assertion proved statistically significant in all three methods. The most statistically significant differences were calculated using one-way analysis of variance, where there were six. Conversely, when using the Pearson chi-squared test, a difference was identified in only three attributes. During testing a significance level of $\alpha=0.05$ was employed. If the significance level was increased to $\alpha=0.1$, statistically significant differences could be confirmed for six attributes by all three methods. Conversely, if the level of significance was reduced to $\alpha=0.025$, a consensus of all three methods would only occur in one attribute. This comparison leads to the conclusion that the method used can significantly affect the evaluation of statistically significant differences in identification questions.

V. CONCLUSION

This paper was written to help marketers evaluate the identification questions that are used in marketing research. These questions divide respondents into homogeneous segments. In market-ing, market segmentation is one of the fundamental practices and is subsequently applied in determining the appropriate strategy. Identifying differences between the segments is therefore very useful, as it can increase or reduce the potential target group. The paper presents the most frequently used methods that are useful for identifying differences between groups of respondents. By means of a case study one scale question is presented to, which is then faced with an identification question. The research entailed identification of the importance of marketing tools in personnel marketing and identification of the effect of the size of the enterprise on the response. Four methods were used for the evaluation.

a) If the answer was regarded as ordinal variables, the main role was played by the median. Evaluation takes place solely on a comparison of median enterprises with different sizes. The evaluation revealed that there are significant differences among the three attributes 4, 8 and 10. The advantage of this method is its simplicity and the ability to use it without knowledge of advanced methods of evaluation.

b) When using the Wilcoxon test, differences in attributes 4, 6, 8 and 10 were identified. The evaluation is based on the median and its use is suitable in cases where the hypothesis is non-parametric. This method can be used for samples with low frequency or if normality is not guaranteed.

c) When using one-way analysis of variance (ANOVA) differences in attributes 4, 6, 8, 9, 10 and 12 were identified. The evaluation is based on testing of intragroup and intergroup variance. The usability of methods in cases where normal distribution and variances are demonstrated is identical.

d) When using the Pearson chi-squared test differences in attributes 6, 8 and 10 were identified. The evaluation is based on a comparison of measured and expected frequencies. The method is suitable for non-parametric hypotheses and its advantage is relatively low necessity of frequency $n_1+n_2>30$. The method is suitable for nominal variables.

The main objective of this paper was to present the role of identification questions in marketing research and subsequent statistical evaluation. The evaluation concluded that it highly depends on the chosen method. Although the chosen methods were not in direct conflict with the terms of usage, all of the methods confirmed differences in only two attributes (8 and 10), although in some of the methods six differences were discovered. The paper did what it set out to achieve and can be a great asset for marketers engaged in marketing research.

This document was created with financial support from TUL, as part of the grant scheme supporting specific university research projects.

References

- [1] Machková, H., "Mezinárodní marketing". Grada Publishing, Praha, 2009.
- [2] Ivanová, K. a Olecká, I. "Metodologie vědecko-výzkumné činnosti". Moravská vysoká škola Olomouc, Olomouc, 2010.
- [3] Molnár, Z., "Pokročilé metody vědecké práce". Profess Consulting, Zeleněč, 2012.
- [4] Creswell, John W., "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches". Sage Publication Inc, London, 2008.
- [5] Meloun, M. a Militký J. "Kompendium statistického zpracování dat". Academia, Praha, 2006.
- [6] Hendl, J., "Přehled Statistických metod. Analýza a metaanalýza dat". Portál, Praha, 2012.
- [7] Budíková, M., Králová, M. a Maroš, B., "Průvodce základními statistickými metodami". Grada Publishing, Praha, 2010.
- [8] Řezánková, H. a Löster, T., "Úvod do statistiky". Oeconomica, Praha, 2012.