11 Data analysis

This chapter discusses methods used to prepare and analyse data obtained by the research. The collected data are called “raw data” and as such, they do not solve the problem. It is necessary to give then a meaning or make a sense of collected data by processing and analysing the data. The question also is how should be the data analysed to achieve objectives of the research.

11.1 Data processing

Before analysing data, it is important to process the data and prepare data for analysis and interpretation. Data processing includes editing, entering coded data into a computer, verifying data and creating variables and data files for analysis (Simova, 2010). Before the data are inputted into the computer, they are edited.

Editing ensures that the information from questionnaire is complete, accurate and consistent. It is important to ensure that the data is “clean”, i.e. free from possible errors, misclassification and gaps. Questions should be checked to see that all questions were answered. Researcher should decide what to do if answers have been omitted. If the missing answers are important and there is not chance to ask the respondent again, it is better to exclude the questionnaire from data analysis.

Accuracy and consistency are another important aspects of editing. Obvious inaccuracies in answers should be rejected. Doubtful answers should be checked against other sections of the questionnaires or other questions. Editing procedures relate to checking for errors and omissions on the data collection forms. A preliminary editing can be done by the data collectors. Its purpose is to ensure completeness, consistency, and reliability of the data. Rigorous editing is performed by the researcher before the data are entered into the computer. The questionnaire data collected on collection forms can be keyed into a statistical software manually.

Coding is another important task to do before analysing data. Since the statistical software deals with numbers only, the answers must be precoded. Coding is usually printed on questionnaire forms and answers are recorded as codes. Most questions in the questionnaires are closed-ended questions with pre-coded response categories. Answer categories are designed to be exhaustive and mutually exclusive, and to make it easy for respondents and data collectors to select appropriate options. A residual “other” category with space for the interviewer to write in responses is included as well. The

range of possible responses is tested before data are collected. As with response categories, codes are designed to be exhaustive, mutually exclusive, meaningful, consistent and easy to use in data analysis.

Tabulation is the next stage in processing data. The objective is to prepare quantitative data so that they are readily understandable. This entails counting the frequency of certain cases within classifications relevant to particular surveys. Data are organised into a data matrix, in which columns represent variables and rows represent records for individual cases (i.e. stores, retailers or consumers). Variables reflect questions in the questionnaires and are put in the same order. The collection forms and respondents are sequentially numbered so in case of typing mistakes the right answers are easily identified. The data in files are checked for errors and verified again.

Sometime data collection forms can be used to minimise the amount of missing data as well as errors in data recording and for entering data into the computer for analysis. They also minimise questionnaire cost (paper and photocopies). In case data is collected electronically, the procedure of preparing and processing data is much easier, simpler and faster. Possibilities of making mistakes are to a great extent eliminated (Simova, 2010).

11.2 Methods of data analysis

The methods of data analysis are selected according to the nature of the data (the type of measurement), research questions and objectives, and the number of variables used in the analysis. The steps related to data processing and methods of data analysis are presented in Figure 1.

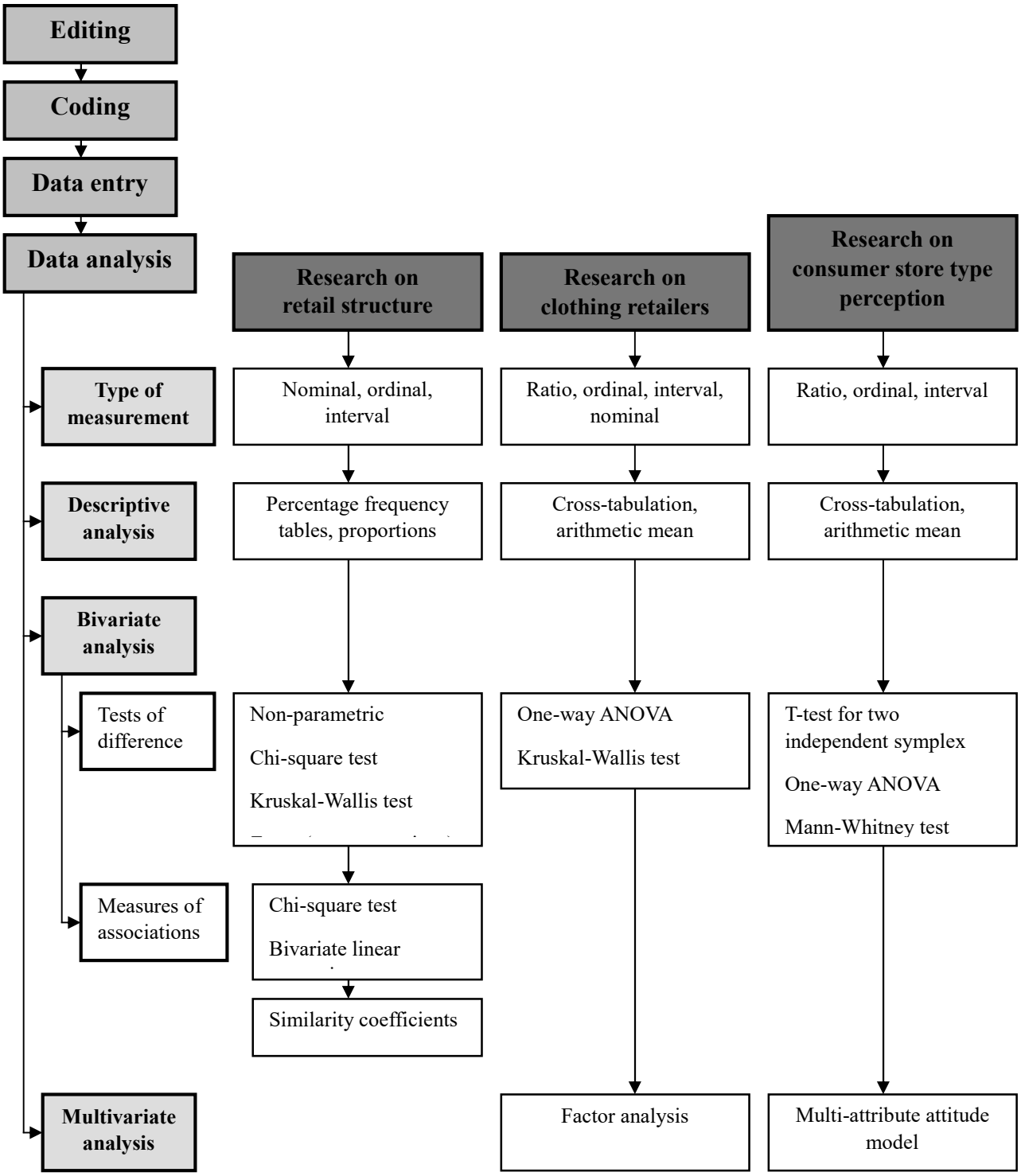

Figure 1 Data processing and analysis

Source: Simova (2010)

Descriptive analysis is used to understand and interpret the results of the research. The calculation of percentage distribution, averages and percentage cross-tabulations are used to summarise the data, get descriptive information and understand the nature of the relationships by making relative comparisons.

Bivariate data analysis is used to test the significance of differences and measure the association between variables. Parametric hypothesis tests are used for interval-scaled or ratio-scaled data; non-parametric statistical procedures are applied for nominal- or ordinal-scaled data.

Multivariate data analysis is conducted for simultaneous investigation of more than two variables. For example, factor analysis is used to identify the main factors affecting changes in clothing retailing and the multi-attribute attitude model can be used to explore consumers' perceptions and attitudes towards different types of clothing stores. Sometime analysis can be more complicated as the objective is to find a cause-effect relationship.

The next step is an interpretation of the results, which gives the meaning to findings. Interpretation leads to conclusion that represent simple generalization of analytical and interpretational results. At the end, there is a clear statement of findings without facts, tables and diagrams. Conclusion may lead towards recommendations to a problem solution.

11.3 Examples of methods used in data analysis

The methods of data analysis used the study of the research on retail structure, research on retailers themselves and research on consumer store type perceptions and attitudes are discussed in more detail below.

Methods of Data Analysis Used in the Research on the Structure of Clothing Retailing

Data on the structure of clothing retailing and its changes were obtained by repeated surveys conducted annually in the period 1994-1999.

There are at least two basic approaches to analysing repeated surveys: either to analyse them separately (which is the same as analysing a single survey except that the procedures are repeated for several surveys) or to cumulate the surveys. Because the objective of the study was to determine whether the retail structure changed, then to test for statistical significance, the cumulated approach

was used to analyse the surveys. The cumulate-data approach does not involve any special effort in data set up, is flexible and provides a basis for testing for change.

In determining the most appropriate analysis techniques, the hypotheses of the retail change and the nature of the available data were considered. It should be noted that the data was nominal (categorical) obtained at discrete points in time, which limited the sophistication of the statistical test that could be undertaken. Most literature on longitudinal analysis deals mostly with the analysis of interval data and proposes techniques (trend analysis, cohort analysis, time series analysis[1]) that require assumptions that have little or no plausibility in the case of categorical data. Although some variants of these models for categorical data analysis have appeared, the literature in this field is not very accessible and the majority of these procedures are not included in standard computer programs. As a starting point for the analysis of categorical variables, percentage frequency tables and percentage cross-tabulations were used to describe the retail structure and changes in the retail characteristics in the period 1994-1999. Because most of the data were collected on a nominal scale, non-parametric statistical tests were used to examine significant differences in changes in all retail characteristics among years (from 1994 to 1999) and five town categories. To test the significance of differences for more than two pairs of grouping variables (for six years and five town categories), the non-parametric Chi-square test for more than two independent samples for categorical data and one-way analysis of variance (ANOVA) for interval data were used. In particular, the Kruskal-Wallis test that is the nonparametric analogue for a one-way analysis of variance (ANOVA) was used. These tests provide protection from identifying too many differences as significant when they are not[2]. Associations between retail characteristics and the size of the population of towns were measured by Chi-square and bivariate linear regression. The extent of retail changes was measured by similarity coefficients.

Methods of data analysis used in the research on retailers themselves

The starting point of the analysis of the impact of environmental factors, consumers, competition and sourcing on the development of clothing retailing in the Czech Republic was descriptive analysis using mainly cross-tabulation and arithmetic means. Ranking was used to identify the main (most important) factors associated with changes in clothing retailing as

---

[1] These analyses are based on the regression model that is not suitable for the analysis of categorical data. The standard regression model assumes linear relations between variables measured at the interval level, underlying both, normal distribution and no interaction effects, which is not applicable to categorical data measured at the nominal or ordinal level. The language of regression analysis was not developed for the analysis of categorical data.

[2] Testing for example three pairs of means, the probability of a difference falling in the critical region is not 0.05 but $1-0.95^3=0.0975$. With the increasing number of tested pairs of means, the critical region increases. In this case, the critical region would be 10 percent instead of the desired 5 percent.

perceived by retailers. The difference in retailers' perceptions of the impact of mentioned factors on changes in clothing retailing by retail formats and the size of population (town categories) were examined by the one-way ANOVA test and the non-parametric Kruskal-Wallis test.

As the development of clothing retailing is a result of the simultaneous effect of a few factors, multivariate data analysis was used to investigate the impact of environmental factors, consumers, competition and sourcing on changes in clothing retailing. In particular, factor analysis was used to explore the impact of a set of variables on development of clothing retailers as perceived by different groups of clothing retailers (Simova, 2010).

The purpose of factor analysis is to summarise the information contained in a large number of variables and reduce them into a smaller number of factors in order to identify the underlying dimensions (structure) in examined phenomena. Factor analysis begins by constructing a new set of factors based on the calculation of the intervariable correlation matrix. Although there are a few factor analysis techniques that look at the degree of association among all of the variables, a principal components analysis (which is probably the most frequently used approach) was used in the analysis. Five factors with eigenvalues larger than or equal to one explaining about 60 percent of the total variance in the data were extracted. As usually the initial factor extraction does not give interpretable factors, orthogonal Varimax rotation (assuming that resulting factors are uncorrelated) was used to obtain factors that could be named and interpreted. The factors were named and described according to the highest (rotated) factor loadings that showed relatively high correlations of the variables with the factors. In a subsequent analysis, mean factor scores[3] were used to explore how identified factors were related to the size of towns and retailers' background characteristics (SPSS, 1998).

Methods of data analysis used in the research on consumers' store type perception and attitudes

Two surveys were conducted in 1995 and 1999 in both, small (below100,000 inhabitants) and large (more than 100,000 inhabitants) towns in order to examine the relationship between consumers and different retail formats as well as the impact of consumers' demographics

[3] Factor scores represent an individual's combined response to the several variables representing a particular factor.

(gender, age and income), consumers' preferences and store perceptions on the development of clothing retailing in the Czech Republic (if any). Having consumer store type perception measured at two different points of time (in 1995 and 1999) in small (below100,000 inhabitants) and large (more than 100,000 inhabitants) towns, the study attempted to identify patterns of relationships between consumers' attitudes, preferences and shopping frequency, and changes in clothing store characteristics (determinants) in terms of the size of population and consumer characteristics. An overall consumers' store type perception and attitudes towards four retail formats were measured by rating scales concerning retailing mix (price, merchandise, in-store ambience, staff, service and location).

Cross-tabulations and arithmetic means were the starting point of the analysis. Gender, age, and income level as an indicator of social status as well as an important factor in clothing shopping, were used as segmentation factors for analysis of consumer store type perception and store attribute preferences. Differences in store type evaluations and perceptions by consumers' demographics (gender, age and income) in relation to the character of data were tested by one-way ANOVA and Kruskal-Wallis tests.

Consumers' attitudes towards different retail formats considering the importance of eight selected store attributes in consumer shopping for clothing were analysed using the multi-attribute attitude model as it proved to be more reliable in predicting consumer preferences and measuring store attitudes than demographic, personality or general attitude models.

The model is presented by the equation:

$$A = \sum_{i=1}^{n} W_i * B_i \qquad (1)$$

Where A = the attitude towards a particular store

W = the weight or importance of attribute i

B = the evaluative aspect or belief toward attribute i for a particular store

n = number of attributes important in the selection of a given store (James et al., 1976).

It measures consumer attitudes as a function of the strength of beliefs about an object and an evaluation of these beliefs. Attitudes of consumers towards stores are reinforced by their beliefs and often attract strong feelings about retailers that may lead to particular intents of shopping behaviour.

The multi-attribute model was performed for all income categories as the level of income determines consumers' shopping behaviour and patterns to a great extent. Generally, higher income consumers spend more and buy higher quality clothing than those from lower income groups who buy rather lower-quality clothes. A comparison of consumers' store type perceptions segmented by income level at the different points of time (in 1995 and 1999) was used to identify how (if at all) consumers' shopping preferences in 1995-1999 influenced the development of clothing retailing in terms of changes in store attributes measured by consumers' perception (Simova, 2010).

Summary

To get a meaning of the collected data, it is necessary to process and analyse the data. Data processing includes editing, entering coded data into a computer, verifying data and creating variables and data files for analysis. The methods of data analysis are selected according to the nature of the data (the type of measurement), research questions and objectives, and the number of variables used in the analysis. **Descriptive analysis** is used to understand and interpret the results of the research., to summarise the data, get descriptive information and understand the nature of the relationships by making relative comparisons. **Bivariate data analysis** is used to test the significance of differences and measure the association between variables. **Multivariate data analysis** is conducted for simultaneous investigation of more than two variables. The next step is an interpretation of the results, which gives the meaning to findings. Interpretation leads to conclusion that represent simple generalization of analytical and interpretational results.

References:

1. SIMOVÁ, J.: *Retailing in a Transitional Central European Economy: A model of Clothing Development in the Czech Republic*. Lambert Academic Publishing, 2010. ISBN 978-3-8383-4333-4.