CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2018 Monika Neufussová

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Monika Neufussová

Lifestyle-related behaviour: an analysis of gender differences in the Czech Republic

Bachelor thesis

Prague 2018

Author: Monika Neufussová

Supervisor: PhDr. Jana Votápková

Academic year: 2017/2018

Bibliographic note

NEUFUSSOVÁ, Monika. *Lifestyle-related behaviour: an analysis of gender differences in the Czech Republic.* Prague 2018. 57 pages. Bachelor thesis (Bc.). Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor: PhDr. Jana Votápková.

Abstract

This thesis studies gender differences in health-related behaviour in the Czech Republic. It analyses approaches of various socio-demographic groups to smoking, drinking alcohol, performing physical activities, nutrition and stress. The cluster analysis was performed to divide the respondents into healthy and unhealthy segments. Using the principal component analysis, the risk factors were composed to create a single lifestyle variable which was further regressed on gender, age, partnership, education level and body mass index in order to find out the effects of these characteristics on lifestyle. Besides, Logit, Poisson and linear regressions were used to test the impacts of socio-demographic variables on individual habits.

Althought it was not possible to divide the sample into healthy and unhealthy parts, both principal component analysis and individual regressions provided evidence that women live healthier. Furthermore, the analyses revealed that older and less educated people live healthier compared to the younger and more educated individuals. Living in a partnership has no significant effect.

Abstrakt

Tato bakalářská práce se zabývá genderovými rozdíly v životním stylu a s tím souvisejícím chováním v České republice. Analýza je založena na zkoumání přístupů různých socio-demografických skupin ke kouření, alkoholu, fyzickým aktivitám, stravě a stresu. K rozdělení respondentů do zdravého a nezdravého segmentu byla použita shluková analýza. Pomocí analýzy hlavních komponent byla z jednotlivých rizikových faktorů vytvořena jediná proměnná *lifestyle*, která byla dále použita jako závislá proměnná v modelu, kde nezávislými proměnnými byly pohlaví, věk, partnerství, dosažená úroveň vzdělání a index tělesné hmotnosti. Jednotlivé zdravotní návyky byly také testovány zvlášť. S ohledem na jejich rozdělení byly použity logistická regrese, poissonovy a lineární regresní modely.

Ačkoliv shluková analýza neposkytla postačující výsledky, analýza hlavních komponent i jednotlivé regrese vedly ke zjištění, že ženy žijí zdravěji než muži. Kromě toho analýza ukázala, že starší a méně vzdělaní lidé také žijí zdravěji. Soužití s partnerem nemá žádný významný vliv na způsob života.

Keywords

lifestyle, gender differences, Czech Republic, principal component analysis

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 9, 2018

Signature

Acknowledgement

I would like to express my gratitude to my supervisor PhDr. Jana Votápková for her guidance and inspiring comments. I would also like to thank my family and especially my boyfriend for his huge emotional support.

Project of Bachelor Thesis

Author of the bachelor thesis:	**Monika Neufussová**
Supervisor of the bachelor thesis:	**PhDr. Jana Votápková**
Academic year:	**2017/2018**

Theme: **Lifestyle-related behaviour: an analysis of gender differences in the Czech Republic**

Research question and motivation:

Besides socioeconomic determinants and physical environment, lifestyle is one of the major determinants of health. As Von Bothmer and Fridlund (2005) claim in their paper concerning the gender-based health habits of Swedish university students, people's approach to their health is found to be a fine predictor of mortality. Self-rated health has a strong predictive power. Previous studies from the United States demonstrate that women care more about their diet (they are more interested in nutrition) but they are also more stressed than men. On the other hand, men are not that concerned with nutrition and tend to consume more alcohol, but at the same time they are likely to do more exercise.

In the past decades, the demand for healthy products and wellness expanded - it barely existed 30 years ago. To be able to target health education programs and promotions it is necessary to know people's health-related habits. In this bachelor thesis, my objective will be to investigate whether there are any gender differences in health-related habits of Czech people. I will compare their overall approach both to healthy lifestyle and to individual habits. I assume that women will show a higher tendency to maintain a healthy lifestyle.

Contribution

Both Czech and international studies on lifestyle approach were focused mainly on children and adolescents. Von Bothmer and Fridlund (2005) focused on Swedish university students, Steptoe and Wardle (2001) compared

Western and Eastern Europe students without discussing gender differences, and Haase et al. (2004) investigated gender differences but only regarding leisure time physical activities. I will try to add the missing research and I will focus on lifestyle-related behaviour of all cohorts, emphasising differences between genders in the Czech Republic.

Methodology

I will use data from the World Health Survey done by the World Health Organization in 2003 in the Czech Republic. The data were collected in face-to-face interviews and contain information about people's health habits, such as smoking, alcohol consumption etc. In addition, they also contain socio-demographic characteristics.

The descriptive part will compare women's and men's approach to individual habits which include daily consumption of fruit and vegetables, number of cigarettes smoked, quantity of alcohol consumed, attitude to physical activity and stress. A cluster analysis will be used to divide respondents into groups based on their behaviour. Logit and Poisson regressions will be used to test the effects of gender on individual habits econometrically. Additionally, other socio-demographic variables will be controlled for. These include age, marital status, education, job, weight, height etc. Since the overall lifestyle cannot be observed directly, individual habits will then be weighted using a principal component analysis and a single lifestyle variable will be obtained. An additional regression on the obtained composite variable will be carried out.

Outline

1. Introduction
2. Literature Review
3. Data Description
4. Methodology
5. Results

6. Conclusion

References:

[1] Divine, R.L. and Lepisto, L. (2004): Analysis of the healthy lifestyle consumer, *Journal of Consumer Marketing* **22/5**, 275-283.

[2] Denton, M. and Walters V. (1999): Gender differences in structural and behavioural determinants of health: an analysis of the social production of health, *Social Science & Medicine* **48**, 1221-1235.

[3] Denton, M.; Prus, S. and Walters, V. (2004): Gender differences in health: a Canadian study of the psychosocial, structural and behavioural determinants of health. *Social Science & Medicine* **58**, 2585-2600.

[4] Ministry of Health of the Czech Republic (2014): Zpráva o zdraví obyvatel České republiky, Praha.

In Prague on .....

Signature of the supervisor Signature of the author

Contents

Introduction **2**

1 Healthy product markets **4**

2 Literature Review **6**

3 Data Description **11**

4 Methods **15**

4.1 Cluster analysis 15

4.2 Principal component analysis 16

4.3 Logit model 17

4.4 Poisson model 18

4.5 Multiple linear model 19

5 Results **21**

5.1 Cluster analysis 21

5.2 Principal component analysis and a corresponding regression 22

5.3 Individual habits 24

6 Discussion **31**

6.1 Gender differences 31

6.2 Limitations and improvements 32

Conclusion **34**

References **36**

List of tables **41**

List of figures **42**

Appendix **43**

Introduction

According to a World Health Organisation's report, a healthy lifestyle can be defined as *"a way of living that lowers the risk of being seriously ill or dying early, helps people enjoy more aspects of life longer and helps the whole family"* (World Health Organisation Regional Office for Europe, 1999). Besides socioeconomic determinants and physical environment, lifestyle is one of the major determinants of health and an effective predictor of mortality (World Health Organisation, 2003). Even though forty years ago the healthy products market hardly existed (Berry, 2004), since then this sector has enlarged very quickly and research has been done to find out its peculiarities.

In view of the fact that habits are shaped mainly during childhood, many researchers have focused on analysing children and adolescents (von Bothmer and Fridlund, 2005; Steptoe et al., 2002). Moreover, a vast majority of studies concentrated either on one or on just a few habits instead of examining the lifestyle maintenance as a whole. Plotnikoff et al. (2004) explored imbalances in performing physical activities; Wardle, Haase and Steptoe (2004) focused on food choices and McDonough and Walters (2001) analysed stress. Usually the studies emphasize gender differences, but they also control for age, education, family status or income inequalities. Nevertheless, there is a lack of Czech studies in this field. The existing Czech research concerning health stresses differences in self-rated health rather than analysing health-related behaviour (Hraba et al., 1996). Although the Czech Health Report contains important and sufficiently detailed facts regarding contrasts in gender-related behaviour, it is merely descriptive. The percentages of gender representatives who incline to certain habits are provided, however, the report does not investigate the causes (Ministry of Health of the Czech Republic, 2014).

The aim of this thesis is to fill in this gap by analysing Czech adults' approach to risk factors, such as smoking, alcohol and stress, and to other health-related behaviours, like nutrition or exercising. The effect of gender will be of primary interest, while other socio-demographic variables will be

controlled for, such as age, family status and education level. The hypothesized effect of gender is that women will have a healthier lifestyle than men. In the analyses, various lifestyle factors are used as dependent variables. In addition, a joint dependent variable is created using a principal component analysis. Besides, the possibility of dividing the sample into healthy and unhealthy segments is tested using cluster analysis.

This bachelor thesis is organized as follows. In section 1 the recent changes of the healthy products market are described and the reasons for research are highlighted. In section 2 the most relevant studies concerning lifestyle are summarized, both foreign and Czech. There are mostly international studies and only a few from the Czech Republic, which provides room for the current and further research. In chapter 3 the data used in this thesis are described. The section 4 introduces the methods which are used in the analysis and explains their theoretical background. Chapter 5 presents the results, discusses the differences between genders and compares socio-demographic groups. It describes the relations between particular risk factors as well. Finally, chapter 6 highlights gender differences, discusses the implications and limitations of the analyses performed and proposes possible improvements for future research.

1 Healthy product markets

Regardless of which gender attaches more importance to maintaining healthy lifestyle, it is indisputable that this relatively new sector influences many industries (Divine and Lepisto, 2005). It is still quite a young field, because it emmerged and started to grow in the eighties (Berry, 2004). In the Czech Republic it was probably even later due to different social and political status.

In fact, the healthy products market was not needed in the past, since one of the causes of its formation was the change in food industry. The number of chemicals added to the food products has increased during recent years. Higher industrial production of food means that nutritional additives, colouring, flavouring and texturizing agents are added in order to increase the nutritional value of food products or to create so-called functional food. But on the other hand the usage of additives led to higher availability of junk food that replaced more nutritious snacks. Lower prices of food products containing chemicals pushed the organic food to its own market (Branen et al., 2001). This organic market is still relatively small, but it is growing and increasingly more popular. Also organic farming is on the rise and it is going to be more supported by governments (Hughner et al., 2007).

As reducing obesity became one of the national aims in the United States, food and beverage companies selling unhealthy goods committed to limit calories in their products (Slining et al., 2013). Consequently, fast food chains such as Subway or even McDonald's started to target consumers maintaining healthy lifestyle and offered healthier menus (Barrier, 2004; Anderson et al., 2003). Decreasing obesity rates may also decrease the healthcare costs associated with it. These were more than 5 million pounds in the United Kingdom in 2006-2007 and are expected to double if the contemporary trends are not ceased (Weaver, 2017).

Nevertheless, the food industry is not the only one affected. In 2008 the World Health Organisation recommended to national governments to decrease smoking in their countries by 25% until 2025 (World Health Or-

ganisation, 2013). This requires increasing excise taxes, banning tobacco advertising or stricter smuggling controls. Notwithstanding the fact that rising specific excise taxes is a key component for smoking cessation, governments avoid it, which causes programs to be ineffective (Jha and Peto, 2014). But Uruguay, France or South Africa implemented the recommendations and succeeded to significantly reduce cigarette consumption (Abascal et al., 2012; Walbeek, 2006; Hill, 2014).

With an increasing interest in living healthy, people download many motivational applications into their smart phones, focused mainly on weight loss, nutrition and physical activities. Advantageously these applications are free or very cheap and do not require much time as opposed to appointments with an adviser. They also enable their users to share their progress with friends on social sites (Riley et al., 2011). Even though the evidence supporting their effectiveness is not very strong, the mobile health (or mHealth) market had an estimated income of 13,674 million USD in 2015 and is presumed to still grow (BusinessWire, 2016). Naturally the mHealth market in the Czech Republic is smaller. The reasons might be that the Czech Republic is ranked as a country with poor market conditions by publishers and developers and has the lowest usage of mobile internet in the European Union (Research2guidance, 2015).

To conclude, the recent rise of interest and changes in lifestyle-related behaviour affect many sectors; including food industry, telecommunications market or government aims and policies. Therefore it is necessary to understand how exactly the citizens approach to health changed. This knowledge may be gathered from research. Some relevant studies are summarized in the next chapter.

2 Literature Review

The topic of primary interest of this chapter is lifestyle-related behaviour and its differences among particular socio-demographic groups. It reviews the most important studies performed in this field and summarizes their results.

It is generally believed that men are healthier than women (Arber and Cooper, 1999). However, in 2003, when the data used in this thesis were collected, life expectancy of women in the Czech Republic was 6.5 years higher than the one of men (Czech Statistical Office, 2018). This paradox was widely discussed in the past. Several studies confirmed that besides other factors it is also lifestyle behaviour that influences health status and diseases (Blaxter, 1990; Denton and Walters, 1999). The other factors are genetic and biological features, inequalities in socio-economic sphere like employment (Arber, 1997; Blaxter, 1990; Denton and Walters, 1999) and demographical factors such as age (Arber and Cooper, 1999) or marital status (Walters et al., 2002).

Lifestyle is a good predictor of mortality (World Health Organisation, 2003) and maintaining healthy lifestyle can help prevent illnesses such as diabetes, high blood pressure or obesity. However, inequalities between individual groups exist. Usually researchers concentrated on gender differences (Denton et al., 2004; von Bothmer and Fridlund, 2005; Denton and Walters, 1999), age groups (Plotnikoff et al., 2004) or education groups (Divine and Lepisto, 2005). Most of the studies agreed that the most relevant factors influencing healthy lifestyle are nutrition, stress, physical activities, tobacco products and alcohol consumption (Divine and Lepisto, 2005; von Bothmer and Fridlund, 2005; Denton and Walters, 1999; Steptoe et al., 2002).

Even though it is important to build healthy eating habits already during childhood, in the Czech Republic half of the children eat neither fruit nor vegetable at least once a day, they do not have breakfasts and eat too much salt, sugar and fat. Virtually all age groups were found to eat fewer minerals than the World Health Organisation recommends (Ministry of Health of the

Czech Republic, 2014). There was no study focused on gender differences in nutrition of the Czech adults, however, Wardle, Haase and Steptoe (2004) analysed gender differences in people's approach to dieting, avoiding fat and cholesterol, eating fruit and adding salt, in 23 countries. Significant differences between men and women were found in avoiding fat (in 22 out of 23 countries), in eating fibre (in 20 out of 23 countries) and in eating fruit (in 18 out of 23 countries). Significant gender differences in limiting salt were found in only six countries. Interestingly, in all 23 countries women were more likely to limit fat and eat fibre and in 21 countries they tended to eat more fruit than men. This phenomenon was explained by a theory that women attach greater importance to weight control and to eating healthy. Another possible explanation given is that men prefer taste or easy cooking to health (Wardle and Griffith, 2001; Steptoe et al., 2002).

Kim & Jang (2017) analysed food choices in relation to the level of stress. They found out that people who are stressed tend to prefer unhealthy food. This was already claimed by previous studies which showed that higher level of stress is connected to increased consumption of fat and decreased consumption of fibre, fruit and vegetable (Rutters et al., 2009; O'Connor et al., 2008). Nevertheless, Kim and Jang (2017) showed that it is not only stress, but also gender that determines food choices, because this "mindless" eating in stress is more typical for women. This finding contradicts the results of the mentioned Wardle's et al. (2004) study which suggest that women tend to eat more healthily. This might be caused by limitations of both studies. Wardle et al. (2004) did not analyse limiting sugar and asked people about their usual behaviour, so they did not think about the states of stress. On the other hand in the study of Kim and Jang (2017) unhealthy food was represented by sweet chocolate cake.

Stress was analysed also by the Canadian study of McDonough and Walters (2001) who claim that women are sicker than men as they are exposed to more stressful situations and life events, which is besides other things given by the gender division of labour. As the main stressors social life, re-

lationships, children and family health were identified . The evidence from the Czech Republic is different. Although there are not sufficient data about exposure to stressors for the Czech population, Daňková (2010) made a simple survey finding out that women are in poorer mental health and are less vital than men. Unlike McDonough and Walters, Kebza distinguishes also between short-term, long-term and "returning" stressors, highlighting their very different impacts (Kebza, 2014).

However, it is quite difficult to define the state of stress, because perception is very individual. Also the exhibitions of distress vary. According to Mirowsky and Ross (1995), men express their stress by aversion and fury; on the other hand women are more emotive and anxious which can lead them to depressive states.

Exercise positively influences health and prevents obesity, cardiovascular diseases, some types of cancer and has many other benefits. Physical activity is recommended by many organisations, including the World Health Organisation (Ministry of Health of the Czech Republic, 2014; Plotnikoff et al., 2004). Notwithstanding these recommendations and increasing number of opportunities for sports, the amount of middle-aged Czech men who do not do any sport rose between 2005 and 2010 by seven percentage points, up to 38% (National Institute of Public Health, 2012). In the Western countries less than 40% of adult individuals exercise regularly (Seefeldt et al., 2002). The lack of physical activity leads to illnesses and obesity. More than 57% of adult population in the Czech Republic is overweight or obese and with such statistics the Czechs are one of the leading nations in Europe (Ministry of Health of the Czech Republic, 2014).

Participating in a physical activity is influenced by several factors, such as the amount of friends exercising, work type, injury from previous physical activity, educational level and so on. The differences between men and women were found to be quite small, but an interesting finding was that current alcohol consumers tended to exercise more regularly than the abstainers. Other significant factors influencing individual's frequency of physical activ-

ity were age, marital status, urban/rural environment or smoking (Plotnikoff et al., 2004).

The negative effects of smoking are very well known. Tobacco is probably the biggest threat to health, it causes many diseases, including lung cancer or cardiovascular diseases. It is dangerous not only for those who smoke directly, but also for their surroundings. Annually, it kills more than seven million people, while more than 10% of those are non smokers exposed to second-hand smoking (World Health Organisation, 2017). In the Czech Republic the number of smokers oscillates around 30% and the trend is stable in the past few years. The daily smokers are mainly men and they also smoke more cigarettes than women (Ministry of Health of the Czech Republic, 2014), which is consistent with international results, even though there are some exceptions. For example in Northern Europe the gender differences were negligible (Payne, 2004). The evidence from South Korea showed that the share of smoking women was approximately ten times lower than the share of smoking men. All socio-economic categories of women smoke less than men, the greatest difference between them was in the group of 30-39 years of age. Smoking was positively associated with lower education and higher levels of stress and negatively connected to being married (Chung et al., 2010).

As far as alcohol consumption is concerned, it has also well known negative effects. Besides the physical diseases and mental and behavioural disorders, it causes addictions and many injuries. In 2012, almost six percent of deaths in the world were attributable to drinking alcohol (World Health Organisation, 2015). The region with the highest intake of pure alcohol per capita is Europe (World Health Organisation, 2014) and the Czech Republic has the highest consumption per capita in Europe (Ministry of Health of the Czech Republic, 2014).

Several studies showed that men's inclination to alcohol is higher than the women's one. McDonough and Walters (2001) showed that men are 83% more likely to drink more than two alcoholic beverages a day and five times

more likely to drink more than fourteen alcoholic drinks per week. But interestingly, women reported higher levels of distress, so drinking is apparently not associated with stress. Also von Bothmer and Fridlund (2005) support the hypotheses that men drink more alcohol than women. Plotnikoff et al. (2004) found out that being a current alcohol consumer increases the chance that a person will exercise regularly. They explained this phenomenon by a theory that adults do sports in order to socialize and commonly gather after physical activities.

Czech women also drink significantly less alcohol than men and are also more likely to abstain. The greatest problem concerning alcohol in the Czech Republic is youth drinking, because the Czech adolescents drink much more alcohol than other Europeans of the same age (Ministry of Health of the Czech Republic, 2014). However, Divine and Lepisto (2005) who analysed healthy lifestyle consumer discovered that alcohol is not a variable that helps to distinguish healthy and unhealthy segments. This is probably due to the fact that drinking moderately is far less serious problem than alcoholism. Even people maintaining healthy lifestyle drink wine which is sometimes recommended for fighting heart diseases (Shapiro and Biddle, 1996).

In this part the most relevant studies were summarized. The research focused on this behaviour has many forms. Often the researchers concentrate on particular behaviour instead of the whole lifestyle performance. Even though the findings of existing studies are sometimes similar, they are usually done for specific countries and their results cannot be extended to the Czech Republic, where the research concerning this topic is missing. Therefore, the aim of this thesis is to fill this gap by analysing the data from the Czech Republic and using them to investigate the Czech citizen's health habits. Men and women will be compared and their individual approaches to lifestyle will be highlighted.

3 Data Description

In this section the variables used in the later econometric analysis are described. Firstly, the data and methods of measurements are described and the expected effects are stated. There is also a Correlation matrix (Tables 8 and 9) showing correlations between the variables and a descriptive statistics with focus on gender differences (see Table 5, Table 6 and Table 7).

The World Health Survey data were used. The data were collected by the World Health Organisation in 2003 in the Czech Republic through face to face interviews. There were 935 adult individuals questioned and the data contain information about their self-rated health, health habits, inclination to risk factors and their socio-demographic characteristics.

Lifestyle variables represent dependent variables in the analysis. These include the following:

Smoking dummy variable which is equal to 1 for smokers regardless of the intensity and 0 otherwise. Unfortunately, the information about number of tobacco products smoked per day was very often missing thus only the former dummy variable could be used. The respondents were further asked whether they smoke daily, non-daily or not at all. As many as 651 out of 935 participants of this study qualified themselves as non-smokers. The most favourite tobacco product seems to be manufactured cigarettes, the daily smokers smoked on average 13.5 cigarettes a day. It is worth noticing that 126 out of 223 daily smokers were men and for 41 daily smokers the highest level of education achieved is secondary school or lower.

The variable *Alcohol* measures for the amount of drinks per week. The respondents were asked about whether they had ever drunk alcohol and about the frequency and amount drunk during the week before the interview. There were 126 respondents who had never drunk alcohol, 106 of them were women and only 20 were men. On average, women drink 2.4 drinks a week, while men have on average 8.9 drinks.

The variables *Problem sleep*, *Tired*, *Depressed* and *Worried* used in the analysis take values from 1 to 5. The respondents described their problems

on five-degree scales where the levels were labeled as follows: 1 = none, 2 = mild, 3 = moderate, 4 = severe, 5 = extreme problems. One quarter of the respondents admitted moderate or more severe problems with sleeping, 70% of them were older than fifty and 67.5% of them were women. One fifth of the respondents felt at least moderately depressed or sad and almost 72% of them were women.

The variables *Fruit* and *Vegetable* represent how many servings of fruit and vegetable the respondents eat. In case of fruit one serving refers to e.g. one apple or one banana, one serving of vegetable means half a cup of carrot, corn, beans or others. The respondents report to eat on average 1.8 servings of fruit and 1.3 servings of vegetables a day. There were significant differences between consumption of fruit and vegetables - almost 75% of respondents ate none or only one serving of vegetables a day, while 43% of respondents consumed more than one serving of fruit.

Walking is a variable representing weekly time spent with walking. *Vigorous activity* and *Moderate activity* are variables created to measure the week frequency of these physical activities. Vigorous activities represent sports like aerobics or fast cycling, while moderate activities stand for playing double tennis or for example light weight-lifting. An information about the time spent with these activities was very often missing thus only frequency could be used in the analysis. The data show that most of the individuals asked do not practice vigorously on more than two days per week, only 17.2% of them perform vigorous activity more than twice a week. On the other hand, more than a half of the respondents perform moderate activity on three or more days a week. Almost 44% of women reported to perform moderate activity more than 5 times per week.

Besides the risk factors the socio-demographic characteristics were observed as well. The crucial variable of primary interest is the *gender* variable. It is a dummy variable coded 1 for female and 2 for male. In the sample there was 55.19% of women and 44.81% of men. Based on the literature reviewed, it is expected that women will attach higher importance

to healthy nutrition, i.e. eat more fruit and vegetable and at the same time they are expected to drink less alcohol and to smoke less. On the other hand, it is presumed that women will report higher levels of distress and tiredness. The effect of gender on exercising is quite unpredictable, since there are many contradictions. Men are more likely to exercise vigorously, while women tend to prefer moderate activities. It depends which habit prevails. It is thus hypothesized that women will maintain healthier overall lifestyle. These expectations are supported by the correlation matrix.

Age was measured by an open response question.The expected effect on healthy lifestyle is mixed. Unlike the research of Divine and Lepisto (2005), who proved that age has a positive impact on maintaining healthy lifestyle, the data from the Czech Republic suggest that with increasing age people tend to feel more stressed and worried, eat less vegetables and exercise less. But they also smoke less and drink less alcohol, thus it depends what dominates.

For the purpose of the analysis a dummy variable *Partnership* was created, equal to one if a person is currently married or cohabitating and zero otherwise. Answering about their marital status the respondents had six possibilities: never married, currently married, separated, divorced, widowed, cohabitating. A majority of respondents were currently married (58.82%). Based on the correlation matrix performed, living with a partner is expected to have positive effect on mental health and on pursuing moderate physical activities. People living with a partner are also expected to smoke less.

The variable *Education level* ranges from 1 to 7 depending on the highest level of education completed. The highest level of education completed is one of the following possibilities: no formal schooling, less than primary school, primary school completed, secondary school completed, high school (or equivalent) completed, college/pre-university/university completed, post graduate degree completed. The variable *Education years* measured the length of completed education in years as answered by the respondents. For most of the respondents high school was the highest level of education they

had completed (36.04%). Based both on the literature (Plotnikoff et al., 2004; Divine and Lepisto, 2005) and the Czech data, it is expected that more educated respondents will maintain healthier lifestyle.

The *BMI* is calculated as

$$BMI = \frac{weight[kg]}{(height[m])^2}, \tag{1}$$

using individual's weight in kilos and height in centimetres which were measured by open response questions. BMI is an index often used as a measure of obesity, sometimes known as a Quetelet index (Frankenfield et al., 2001). It takes values between 15.57 and 45.17. Apparently, most of the participants (40.64%) belong to the category 18.5-25 characterised as ideal, healthy ratio between weight and height. However, almost one fifth of individuals have BMI values that signify obesity.

In this chapter I summarized what the data look like, how they are distributed between particular socio-demographic categories and how the particular variables used in later econometric analysis were created. The detailed description of the data is shown in Appendix A in tables 5, 6 and 7 and correlations between individual variables are provided in Correlation matrix in tables 8 and 9.

4 Methods

In this section the methods of analysis are described. Three techniques were used and are explained in this chapter. Firstly, a cluster analysis was done to find out whether it is possible to divide respondents into healthy and unhealthy segments. This division could furhter be used to discover some characteristics of these groups and patterns of their behaviour. Secondly, a principal component analysis was performed to get a single lifestyle variable. This lifestyle variable was regressed on socio-demographic variables which enabled me to find out which gender tends to live healthier. Finally, various regressions were done to analyse indivudual habits.

4.1 Cluster analysis

Cluster analysis is a method allowing a researcher to divide individuals into specified number of segments. In this thesis k-means clustering method was used. There are k segments (clusters) grouped such that the variation inside clusters is as low as possible, in other words particular segments are as different as possible. Number of segments, k, is specified by analyst in advance.

In order to decide how many segments should be used, an elbow method was performed. At first, the respondents who did not answer all questions about their habits had to be removed from the sample. The outliers of individual variables were identified and removed from the sample, since they might have destroyed the structure of clusters. To be more specific, respondents who reported to drink more than 90 drinks per week, to eat more than 22 portions of fruit per day or to walk more than 87 hours per week were excluded, as these values are highly unlikely to be true and they probably mean a measurement error. Additionally, the variables *fruit, vegetable, vigorous activity, moderate activity* and *walking* were multiplied by minus one in order that higher values always signify worse lifestyle maintenance. Finally, the data were scaled because of different units of variables. The total

within sums of squares for each cluster were computed as follows:

$$\sum_{k=1}^{K} \sum_{i \in S_k} \sum_{j=1}^{p} (x_{ij} - \overline{x}_{kj})^2$$

where S_k stands for a set of observations in the k-th cluster and $\overline{x}_{kj}$ is the j-th variable of the cluster center in the k-th cluster (Jain, 2010). The total within sum of squares were depicted graphically to identify ideal number of clusters. The point of break (elbow) is the point of optimal number of clusters, since marginal decrease of total within sum of sqaures is lower. In other words, the added value of having one more cluster is low.

4.2 Principal component analysis

Principal component analysis is a dimensionality reduction method that linearly combines existing variables to create new ones (so called principal components). The essence of it is to find the direction (or directions) with the largest variance. This will provide the eigenvectors, which are the vectors indicating the direction, and eigenvalues, which are the numbers indicating the magnitude of variance explained by corresponding eigenvector. The eigenvectors are then used to reorient the data in a new coordinate system, where the new axes are determined by the eigenvectors (Abdi and Williams, 2010).

In this thesis, the variables used in the principal component analysis were *smoking dummy, alcohol, fruit, vegetables, vigorous activity, moderate activity, walking, problem sleep, tired, depressed* and *worried*. Firstly, the data had to be adjusted so that the lower values always meant living healthier. Secondly, the respondents who did not convey all their lifestyle habits were removed from the sample. Afterwards, only one direction with the highest variance was searched for and the principal component resulting from it was used as a lifestyle variable. This newly created lifestyle variable was analysed using a multiple regression model. It was regressed on the socio-demographic characteristics in order to find their effects on maintaining healthy lifestyle measured by the composed variable.

The assumptions for this model were verified; the zero conditional mean and normal distribution of errors was controlled by a plot and homoskedasticity was tested by a Breusch-Pagan test. Fulfiling these assumptions, the estimator is the best unbiased estimator.

4.3 Logit model

Since smoking was measured as a dummy variable, a logit regression was used for the analysis. Let $\mathbf{x}$ be a vector of explanatory variables, i.e. *gender, age, partnership, education level, BMI, alcohol, fruit, vegetables, problem sleep, depressed, vigorous activity, moderate activity* and *walking*. As in other models, it is necessary to avoid high correlations between explanatory variables. Correlation of 0.5 was set as a minimum required for excluding a variable from regression. This was a reason the variables weight, occupation, tired and worried had to be removed (see Correlation matrix in Table 8 and Table 9). The variable *education level* was chosen to represent education attained, since additional years of education do not always mean more educated people (e.g. university students who need more than standard three years to obtain the bachelor degree) and the interpretation would not be clear.

Then the probability that a person smokes is defined as:

$$P(smoking\_dummy = 1|\mathbf{x}) = G(\beta_0 + \mathbf{x}\beta), \tag{2}$$

where G is a cumulative distribution function of standard logistic random variable defined as:

$$G(z) = \frac{e^z}{1 + e^z} \tag{3}$$

which assures that $0 < G(z) < 1$ for all z real. The effect of gender on smoking is obtained from the partial derivative:

$$\frac{\delta P(smoking\_dummy = 1|\mathbf{x})}{\delta gender} = g(\beta_0 + \mathbf{x}\beta)\beta_1, \tag{4}$$

where β_1 is a coefficient explaining the effect of gender and g is a partial derivative of G. Since G is a strictly increasing function, its derivative g is always positive. Thus the direction of the gender effect depends on sign of

β_1. And since gender is also a dummy variable, equal to 1 for female and 2 for male, the magnitude of the effect (it means the difference between a man and a woman) can be computed as a difference

$$G(\beta_0 + 2\beta_1 + \beta_2 x_2 + ... + \beta_k x_k) - G(\beta_0 + \beta_1 + \beta_2 x_2 + ... + \beta_k x_k),$$

where $x_2, ..., x_k$ represent age, partnership and other explanatory variables (Wooldridge, 2015). The exact magnitude of the effect depends on the values of $x_2, ..., x_k$, thus it is necessary to plug them in. In this analysis, the average values of explanatory variables were used, except for *parthership* and *smoking dummy* variables, where average values are not interpretable.

4.4 Poisson model

For the variables that take only a few non-negative values, a Poisson model is convenient. It is used for count data that represent number of occurences of a phenomenon per a given time period. In this thesis, a Poisson model was applied to analyse variables measured with a five-point scale (*problem sleep, tired, depressed* and *worried*), variables representing frequencies (*vigorous activity* and *moderate activity*) and amounts (*fruit*, *vegetables* and *alcohol*).

Let's consider *vigorous activity* to be a dependent variable and $\mathbf{x}$ the vector of explanatory variables (as already noted, the highly correlated variables need to be excluded). The idea of a Poisson model is based on the assumption that:

$$E(vig\_activity|\mathbf{x}) = e^{(\beta_0 + \mathbf{x}\beta)}. \tag{5}$$

To interpret the model, the natural logarithm of the equation is taken:

$$log[E(vig\_activity)|\mathbf{x}] = \beta_0 + \mathbf{x}\beta. \tag{6}$$

Now the changes in *vigorous activity* and the coefficients β_j can be easily interpreted as:

$$\%\Delta E(vig\_activity|\mathbf{x}) \approx (100\beta_j)\Delta x_j \tag{7}$$

and

$$\frac{\delta E(vig\_activity|x)}{\delta x_j} = e^{\beta_j} \tag{8}$$

is a multiplicative effect of one unit increase in x_j (Wooldridge, 2015).

Since Poisson regression is a non-linear model, the residuals are always heteroskedastic and thus the robust form has to be used (Cameron and Trivedi, 1998).

The BMI variable was suspected of endogeneity caused by reversed causality. A high value of BMI suggests either being overweight or a muscular figure and it is not clear whether BMI is determined by the frequency of exercising, or vice versa. A regression-based tests were performed to test endogeneity. It was detected in the model where *vigorous activity* was a dependent variable, i.e. in this model error term is correlated with explanatory variable and the zero conditional mean assumption is violated:

$$E(u|\mathbf{x}) \neq 0. \tag{9}$$

The endogeneity issue causes the model to be biased and inconsistent, thus the results are not trustworthy as they might be incorrect. As a solution, it was necessary to find an instrumental variable (let us call it z) strongly correlated with BMI such that:

$$Cov(z, BMI) \neq 0 \wedge Cov(z, u) = 0. \tag{10}$$

The *weight* variable was selected as a proper instrument and used in the regression.

4.5 Multiple linear model

Even though the variable *walking* takes on only positive values, there are too many different values, thus a Poisson model is not appropriate for an analysis and multiple regression was used instead. Firstly, the distribution of the residuals and the other key assumptions need to be controlled for. The model needs to be linear in parameters and the sample has to be random. None of the explanatory variables must be constant and there must not be a perfect linear relationship among the explanatory variables. Homoskedasticity is tested by a Breusch-Pagan test. Further, exogeneity is assumed, i.e.:

$$E(u|\mathbf{x}) = 0. \tag{11}$$

Under these assumptions the ordinary least square estimator is unbiased and the interpretation of beta coefficients is simply:

$$\Delta week\_walking = \beta_j \Delta x_j, \tag{12}$$

other explanatory variables being fixed (Wooldridge, 2015).

5 Results

In the following chapter the results of the analyses are presented. It is divided into three parts. In the first section the results of a cluster analysis are shown. In the second section, the principal component analysis and the results of a corresponding regression are presented. The last section describes the impacts of socio-demographic variables on individual habits.

5.1 Cluster analysis

Firstly, the elbow method was performed to find out the ideal number of clusters. The results are shown in Figure 1. However, the elbow method may provide ambiguous results. In this case, both two and four clusters may be convenient. However, creating four segments would not have a good logical interpretation, thus two clusters were created.

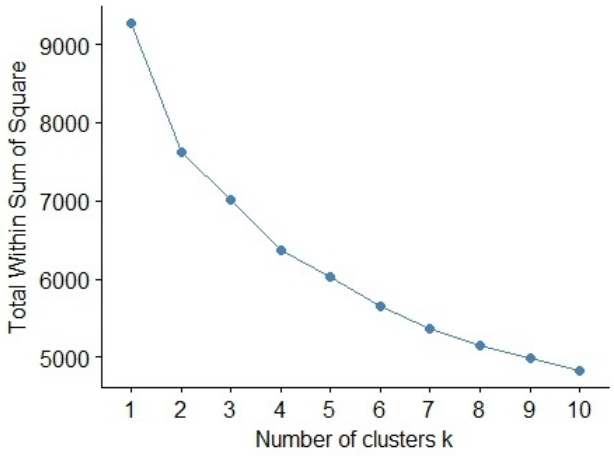

Figure 1: Optimal number of clusters

The means for all variables in each cluster were then computed. Initially scaled units were returned back to their unscaled form so that the means could be logically interpreted. The cluster means are reported in Table 1. The red numbers signify the poorer group from the health perspective. However, the t-test revealed that means of the first four variables are not statistically different on 95% confidence level, i.e. the groups are basically the same in their approaches to smoking, alcohol, eating vegetables

and performing vigorous activities. Thus the respondents cannot be clearly divided into *healthy* and *unhealthy* segments as in Divine and Lepisto (2005) and the results cannot be used for deeper analysis as in Divine and Lepisto. It may be concluded that the group which is in poorer mental health eats significantly more fruit, spends less time walking and exercises moderately less frequently than the other group. Hence the results show that doing less sport is associated with more stress and with being tired. However, no other characteristics of these segments cannot be inferred. It is not possible to decide whether they maintain healthy or unhealthy lifestyle and certainly it cannot be claimed which gender is more likely to live healthier.

	Cluster 1	Cluster 2
Smoking dummy	0.315	0.270
Alcohol	5.132	4.293
Vegetables	-1.282	-1.351
Vig. activity	-1.208	-0.977
Fruit	-1.704	-2.066
Mod. activity	-3.623	-2.927
Week walking	-18.829	-15.680
Problem sleep	1.358	2.896
Problem tired	1.488	2.884
Depressed	1.240	2.911
Worried	1.330	2.958

Table 1: Cluster means

5.2 Principal component analysis and a corresponding regression

A principal component analysis was used to create a new variable *lifestyle* which was composed of the individual health habits. The proportions of the original data variance explained by principal components are shown in Table 2. The single component used as a lifestyle variable explains 25.7 % of variance.

The classic linear model assumptions were tested. The errors seemed to

	Proportion of Variance	Cumulative Proportion
Component 1	25.7%	25.7%
Component 2	13.1%	38.8%
Component 3	12.3%	51.1%
Component 4	9.5%	60.6%
Component 5	8.5%	69.1%
Component 6	7.3%	76.4%
Component 7	6.9%	83.2%
Component 8	6.5%	89.7%
Component 9	5.8%	95.5%
Component 10	3.2%	98.7%
Component 11	1.3%	100%

Table 2: Proportion of Variance explained by components

be approximately normally distributed, nevertheless heteroskedasticity was detected, thus heteroskedasticity-robust standard errors were applied. All explanatory variables were exogeneous, including BMI.

The lifestyle variable was then regressed on the socio-demographic variables, i.e. the model was as follows:

$$\text{lifestyle} = \beta_0 + \beta_1 \text{gender} + \beta_2 \text{age} + \beta_3 \text{partnership} + \beta_4 \text{education\_level} + \beta_5 \text{BMI} + u, \tag{13}$$

where u stands for the error term.

Since lifestyle is a dimensionless variable, the exact numerical interpretation was not available. Notwithstanding this disadvantage, the results provided evidence that there were significant differences between genders. Women live healthier than men, which is in accordance with my original hypothesis. The results also showed that older people tend to live healthier than the young. There were no significant differences detected between people living with a partner and single people. Neither body mass index has a significant effect on lifestyle. However, the lifestyle significantly worsens with higher education level attained. I consider this finding very surprising, as I expected that more educated people will live healthier. On the other hand, people with higher level of education usually have more demanding

jobs. Perhaps they tend to spend a lot of time working and they do not have enough time to exercise or to eat healthy. Moreover, their job is probably stressful as well. Plotnikoff et al. (2005) found a similar inverse relationship between education and sport though only in the young age cathegory. He assumes that the reason why this was detected only in the young age group might be that present-day access to information about health benefits of exercising is easier, while for older generations school was the only source of such information.

The coefficients and their significances are provided in Table 3. Three stars stand for significance at 0.1% level.

Variable	Coefficient
Gender	0.644***
	(0.105)
Age	-0.031***
	(0.003)
Partnership	0.151
	(0.105)
Education level	0.200***
	(0.055)
BMI	0.019
	(0.013)

Table 3: PCA results

5.3 Individual habits

The lifestyle habits were regressed on the socio-demographic variables and the other explanatory variables representing habits. All the results are summarized in Table 4 and the most important results are discussed model by model.

As far as smoking is concerned, a logit regression was used, where *smoking dummy* was a dependent variable. The variables that had statistically significant effect on whether a person smokes or not were gender, age, ed-

ucation level, alcohol consumption and moderate activity. The signs of the coefficients suggest that men, younger people, less educated and those who drink more alcohol and perform more moderate activities tend to be smokers.

These trends might be explained as follows. People who are less educated smoke more than the educated, because they are not fully aware of, or they ignore, the health consequences. Besides, people tend to socialize with similarly educated people. Hence their bad habit of smoking is supported by their social group. The reason why younger respondents smoke more might be again being part of a social group, where smoking is popular. Either at school or at work, having a cigarette may present a moment of relax while talking to classmates or colleagues during a break. Furthermore, middle-aged people usually have a child, which may prevent them from smoking during the period of pregnancy, and afterwards they do not start again.

To analyse the *alcohol* consumption, a Poisson model was used. The variables having statistically significant effect on the amount of alcohol consumed were gender, education level, smoking and moderate activity. The results provide evidence that men and people who are more educated drink more alcohol. People who exercise more often drink significantly more alcohol as well, even though the effect is not large - people who perform moderate activity once per week drink 1.06 times more alcohol than those who do not exercise. This might be caused by the social gatherings of adult people after sport, where alcohol is typically consumed.

Also, people who smoke drink more alcohol. Being a smoker increases the amount of alcohol consumed per week by 56.1%. The effect is quite strong, nevertheless, I suppose it might have been weakened after the smoking ban in restaurants in the Czech Republic.

However, it is necessary to be cautious when interpreting the results. While the information about the number of drinks the week before questioning is available, it does not say much about respondents' habits. This week might have been exceptional because of a birthday party, promotion, passing exams and so on. Besides, an accusation of alcoholism ought not

be based on the amount of alcohol drunk. Having a glass of wine few times a week cannot be interpreted as alcoholism. Hence nothing certain can be concluded about what social groups tend to be habitual drinkers.

The *fruit* consumption was explained with a Poisson model as well. There were no significant differences between men and women found in this analysis and also a vast majority of the other explanatory variables were insignificant. The only variables influencing fruit consumption significantly were vegetable consumption and a frequency of moderate activity, both with positive signs. This implies that people who eat large portions of fruit also tend to eat a lot of vegetables and do sports frequently, which suggests that they probably care about their health.

Vegetable consumption was analysed with a Poisson model. Similarly as in the case of fruit, there were neither age nor partnership differences detected. However, there were statistically significant differences found between genders. The results suggest that men eat 15.4% less vegetables than women. Since vegetables is the most recommended food when dieting, this finding is in accordance with the international studies which found that women were more likely to care about nutrition (Wardle et al., 2004). A positive coefficient of education provides evidence that people who attain higher levels of education are aware of vegetable benefits and consume it more than less educated people.

The frequency of performing *vigorous activities* was analysed with a Poisson model as well and the explanatory variable BMI was instrumented by weight. There were significant differences detected between genders. Other significant variables were age, weight and moderate activity. The sign of the coefficient suggests that men are more vigorously active than women and the difference is large. Men exercise vigorously 2.05 times more often than women. This is probably caused by the fact that women prefer moderate activities, as shown in the next model. Also, women may not have enough time for exercising as they care about house and children. People who are older tend to be naturally less physically active, which includes also performing

vigorous activities. An increase in the frequency of vigorous activity was also explained by higher weight. This finding suggests that heavy people tend to exercise more, probably in order to lose weight. Surprisingly, another variable explaining frequent vigorous exercising was performing moderate activities. I expected to find out that respondents face a trade-off between vigorous and moderate activities, but evidently they perform both. Perhaps they exercise vigorously to be in a good shape and then moderately to relax and have fun with friends.

Another variable explained by a Poisson model was *moderate activity*. The analysis showed there were significant differences between genders. Virtually all explanatory variables were significant as well, except for education level, vegetables consumption, problems with sleeping and depressed. As already mentioned, moderate activities are prefered by women. Similarly as in case of vigorous activities, older people exercise less. Moderate activity is the only variable where significant differences between people living with a partner and people living alone were detected. It showed up that people living in a partnership tend to exercise 30.1% more frequently. Perhaps the fact that a partner exercises pushes the respondent to perform physical activity, or they even exercise together. Drinking alcohol was a variable increasing frequency of performing moderate physical activity, however, the effect of social gatherings with alcohol was already discussed. Interestingly, smokers are more moderately active than non-smokers. Again, this might be caused by get-togethers after sports, where people smoke.

As far as the *walking* variable is concerned, it was analysed with a linear regression model. There were no significant gender differences detected. The only significant variables explaining the changes in the time spent by walking were age and moderate activity. The results provide evidence that older people tend to walk less than the younger ones. An additional 10 years of age cause that a person walks 90 minutes per week less. This is probably due to the fact that the elderly do not go to work, while younger people spend a lot of time walking at work or to work. As opposed to age,

performing moderate activities has positive effect on walking. It is supposed that when a person exercises, he then finds it contraproductive to travel short distances by car or to use public transport and rather walks.

Since *problems with sleeping* were measured on a scale, this variable was explained with a Poisson model. There were slightly significant differences between men and women detected, however, the practical size of the difference is negligible. On the other hand, age appeared to be a significant explanatory variable. People who are older tend to suffer from problems with sleeping, which seems quite natural. Besides age, only problems with feeling tired and being depressed were significant in explaining variations in problems with sleeping. These problems are related to each other, and hence this relationship is quite straightforward; when people are depressed, they are nervous and wake up often during the the night. Nevertheless, I believe that feeling tired does not cause problems with sleeping, but vice versa.

Feeling *tired* was also a scale variable, thus a Poisson model was used again. There were no significant differences among groups based on the socio-demographic characteristics. The explanation why age was not significant in the current model, even though it significantly affected problems with sleeping in the previous model, might be that all age groups feel tired, though from different reasons. While young people tend to live in a hurry and are tired of it, the tiredness is also one of the signs of ageing and might be caused by illnesses as well. The variables having significant effect were problems with sleeping and being depressed, both with positive signs. Naturally, as people have problems to fall asleep and wake up often at night, they feel more tired during a day. Similarly, when people experience a state of stress or even depression, it tires them. Interestingly, being a smoker slightly decreases the level of tiredness. Walking variable was significant as well. The sign of its coefficient suggest that people who spend more time walking tend to feel a little less tired. Even though the opposite might be expected (walking is a physical activity and as such it should be tiring), this direction of the effect is also explicable. As a light physical activity, walking is rather regenerative

and helps to take one's mind off things, which causes feeling relaxed, not tired.

Being *depressed* was measured on a scale as well, hence a Poisson model was used for the analysis. As in the previous model, there were no significant imbalances among groups based on age, education level attained and partnership. Nevertheless, there were minor gender differences detected that show that men feel less depressed than women. I suppose this is probably caused by the division of labour, as women usually have a job and at the same time keep the house and care about children. The only other explanatory variables which were statistically significant were depressed and being worried. As already stated, these problems are closely connected to each other.

As far as the variable *worried* is concerned, it was analysed with a Poisson model. As well as in case of being tired, there were no significant gender differences detected. Neither age, education level attained nor being in a partnership influences significantly the level of respondent's worries. Having problems with sleeping and being depressed has a statistically significant impact on being worried. This is again caused by the fact that these issues are tightly related to each other and they are usually felt together.

Explanatory variables	Dependent variable										
	Smoking dummy	Alcohol	Fruit	Vegetables	Vig activity	Mod activity	Walking	Problems sleeping	Tired	Depressed	Worried
Model	Logit	Poisson	Poisson	Poisson	Poisson	Poisson	Linear	Poisson	Poisson	Poisson	Poisson
Gender	0.623**	1.418***	-0.044	-0.154**	0.720***	-0.291***	-0.478	-0.073*	-0.006	-0.056*	-0.025
	(0.191)	(0.113)	(0.075)	(0.048)	(0.141)	(0.060)	(1.445)	(0.032)	(0.030)	(0.025)	(0.025)
Age	-0.023***	-0.006	0.003	-0.001	-0.020***	-0.006***	-0.148***	0.006***	-0.001	0.001	0.000
	(0.005)	(0.003)	(0.002)	(0.001)	(0.004)	(0.002)	(0.038)	(0.001)	(0.001)	(0.001)	(0.001)
Partnership	0.005	-0.001	0.013	-0.006	0.037	0.301***	-0.363	0.002	-0.005	-0.038	0.022
	(0.175)	(0.111)	(0.068)	(0.045)	(0.126)	(0.060)	(1.277)	(0.028)	(0.026)	(0.022)	(0.021)
Education level	-0.446***	0.143*	-0.007	0.056**	0.049	-0.021	-0.956	-0.021	0.002	-0.019	0.012
	(0.089)	(0.061)	(0.035)	(0.020)	(0.06)	(0.030)	(0.619)	(0.16)	(0.015)	(0.012)	(0.011)
BMI	-0.016	-0.002	-0.008	-0.009*	-	0.013*	0.142	-0.003	0.001	-0.003	0.003
	(0.019)	(0.013)	(0.008)	(0.005)	(-)	(0.006)	(0.141)	(0.003)	(0.003)	(0.002)	(0.002)
Weight	-	-	-	-	0.009*	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(0.004)	(-)	(-)	(-)	(-)	(-)	(-)
Smoking dummy	-	0.561***	-0.110	0.003	-0.120	0.127*	1.442	0.023	-0.061*	0.002	-0.011
	(-)	(0.122)	(0.093)	(0.050)	(0.129)	(0.057)	(1.505)	(0.033)	(0.030)	(0.023)	(0.023)
Alcohol	0.042**	-	-0.005	-0.002	0.004	0.007**	-0.092	-0.000	0.001	0.001	0.001
	(0.015)	(-)	(0.004)	(0.003)	(0.003)	(0.002)	(0.069)	(0.002)	(0.001)	(0.001)	(0.001)
Fruit	-0.075	-0.043	-	0.058***	0.011	0.022**	0.280	0.004	0.008	-0.009	0.003
	(0.010)	(0.037)	(-)	(0.015)	(0.029)	(0.008)	(0.374)	(0.010)	(0.005)	(0.006)	(0.005)
Vegetables	0.036	-0.049	0.248***	-	0.090	0.014	-0.590	-0.009	-0.016	0.023	0.009
	(0.108)	(0.084)	(0.026)	(-)	(0.060)	(0.031)	(0.790)	(0.018)	(0.014)	(0.013)	(0.011)
Vig activity	-0.050	0.032	0.006	0.016	-	0.060***	0.388	-0.008	0.009	0.008	0.002
	(0.045)	(0.025)	(0.021)	(0.012)	(-)	(0.012)	(0.397)	(0.008)	(0.007)	(0.005)	(0.005)
Mod activity	0.060*	0.057**	0.024*	0.004	0.104***	-	1.035***	-0.007	-0.002	-0.001	-0.002
	(0.028)	(0.021)	(0.011)	(0.008)	(0.024)	(-)	(0.232)	(0.005)	(0.005)	(0.004)	(0.004)
Walking	0.004	-0.004	0.002	-0.001	0.004	0.006***	-	0.000	-0.003***	0.000	0.000
	(0.004)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)	(-)	(0.001)	(0.001)	(0.001)	(0.001)
Problems with sleeping	-0.051	-0.016	0.049	-0.041	-0.041	-0.067	-1.323	-	0.219***	0.043***	0.048***
	(0.097)	(0.061)	(0.068)	(0.027)	(0.076)	(0.034)	(0.711)	(-)	(0.014)	(0.013)	(0.013)
Tired	-	-	-	-	-	-	-	0.255***	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.018)	(-)	(-)	(-)
Depressed	0.083	0.122	-0.053	0.049*	0.104	-0.038	0.806	0.073***	0.107***	-	0.348***
	(0.090)	(0.062)	(0.054)	(0.024)	(0.056)	(0.030)	(0.751)	(0.016)	(0.014)	(-)	(0.010)
Worried	-	-	-	-	-	-	-	-	-	0.374***	-
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.010)	(-)

Table 4: Results of individual results regressions

*** = Statistically significant at 5% level ** = Statistically significant at 1% level * = Statistically significant at 0.1% level

Heteroskedasticity-robust standard errors are reported in parantheses

6 Discussion

This chapter is divided into two parts. Firstly, gender differences found in the analyses are presented and summarized. Secondly, the limitations of the methods used in the analyses are discussed and the possible improvements are presented. It is also suggested what further research is needed to deepen the analysis and the knowledge of gender differences in health-related behaviour.

6.1 Gender differences

Since gender was a crucial variable, the differences between men and women are emphasized in the following paragraphs. The results of the models provide sufficient evidence to claim that men smoke significantly more and drink more alcohol than women, while women eat significantly more vegetables, perform moderate activities more frequently and are less vigorously active. Men have slightly less problems with sleeping and feel a little less depressed than women.

Other things being equal, the exact difference in probability of smoking between a man and a woman is 62.3%. Men also drink 141.8% more alcohol per week than women. Both of these results suggest that women tend to live healthier. This conclusion is further supported by the finding that women care more about nutrition as they eat 15.4% more vegetables than men. Moreover, women perform moderate activities 29.1% more often than men. The only relevant habit in which men prevailed in terms of health is exercising vigorously, since women perform vigorous acitivities 72% less frequently than men. Although men also statistically predominate in the mental health, the actual size of the gender difference is very small (in both cases not even 8%) and thus it can be concluded that both genders are almost equal, concerning their sleeping and stress.

As far as the other lifestyle habits are concerned, it cannot be concluded which gender lives healthier, as there were no significant differences found. Thus we can suppose that men and women are equally tired and worried, eat

comparable amounts of fruit and spend approximately the same week time by walking. Nevertheless, taking all the differences found between genders into account, it can be concluded that women tend to maintain healthier lifestyle than men. This conclusion is in accordance with the results of principal component analysis and supports the original hypothesis that women live healthier than men. This result is similar to the ones of Wardle et al. (2004) or von Bothmer et al. (2005).

6.2 Limitations and improvements

Even though the results of performed regressions provided evidence for claiming that women live healthier than men and thus I did not reject my original hypothesis, it is necessary to consider the limitations of the models. The crucial problem that threatens the quality of the analyses performed is endogeneity, which means that the error term in the model is correlated with an explanatory variable:

$$E(u|\mathbf{x}) \neq 0. \tag{14}$$

The endogeneity issue causes the models to be biased and inconsistent, hence the results are not trustworthy as they might be incorrect. One of the causes of endogeneity is an omitted variable. Concerning the current thesis, this is the case of regressing lifestyle variable obtained by principal component analysis on the socio-demographic variables. Explanatory variables used in this model explain 17.25 % of variation in the lifestyle variable.

Should the data allow for it, the analysis could be enriched for further variables including income, living in a city center or village, decision to lose weight, being overweighted in the past and being surrounded by people living healthily. For example decision to lose weight should have significantly positive impact on lifestyle, while people living in a rural area are probably limited in their food choices and in number of physical activities. People who have friends living healthy might be inspired and more informed, thus they should live healthier. Plotnikoff et al. (2004) found out that proportion of friends who exercise has a positive effect on performing physical activity.

Another improvement of the present study would be a detection of possible cointegration in health habits. Availability of time series data would enable the researchers to find out whether there is a common trend in the health-related behaviour.

The analyses performed in this thesis suggest the importance of deeper exploration of health-related behaviour in the Czech Republic. Nevertheless, to analyse gender differences and imbalances among age or education groups with more advanced models, it is crucial to have more complex dataset. The data should contain more information about habits and their lagged values as well, as they might be important for respondent's behaviour and his decision-making.

Conclusion

This thesis examines the gender differences in lifestyle-related behaviour in the Czech Republic. The hypothesis that women maintain healthier lifestyle than men was analysed using the data from the World Health Survey done by the World Health Organisation in 2003. The data contain information about respondents' health-related habits, such as smoking, drinking alcohol, performing physical activities, nutrition and stress. Besides these, the data also contain socio-demographic characteristics, like gender, age, family status and education level attained. A principal component analysis, Logit, Poisson and multiple linear regressions were used in order to identify the differences between genders and to compare socio-demographic groups. Moreover, a cluster analysis was performed, however, its results were not sufficient to be used in further analysis that would answer the research question (see subsection 5.1).

The results of the analyses performed provide evidence which support the original hypothesis that women tend to live in a healthier way than men. This conclusion is based both on the results of the principal component analysis and on the results of the regressions analysing particular habits. It was proved that men smoke and drink alcohol significantly more than women, whilst women eat more vegetables and perform moderate activities more frequently. Although men are significantly more vigorously active, it can be concluded that women's lifestyle maintenance is better. Furthermore, the results suggest that older people tend to live healthier than younger ones. These findings are in accordance with the previous research done by e.g. Wardle et al. (2004) who found that women attach higher importance to nutrition, or von Bothmer et al. (2005) who identified men as more habitual drinkers.

However, as far as mental health is concerned, there were not many differences detected between men and women. This is in contrast to the findings of McDonough and Walters (2001) who found Canadian women to be significantly more stressed. It also contradicts the findings of Daňková in the Czech

Republic. The regression corresponding to the principal component analysis revealed that more educated people care about their health less. This is a startling finding, contrasting with previous studies (Plotnikoff et al., 2004; Chung et al., 2010).

To summarize, lifestyle is a major determinant of health and as such a very important feature of life. Since there is a lack of Czech studies concerning lifestyle, this thesis contributes by filling this gap through describing the differences in approaches among socio-demographic groups with a focus on gender. Nevertheless, lifestyle is a complex issue and a more detailed analysis might overcome the limitations of this thesis. Availability of more thorough data containing detailed information about respondents in a longer time span might enable researchers to perform a deeper analysis.

References

Abascal, W., E. Esteves, B. Goja, F. G. Mora, A. Lorenzo, A. Sica, P. Triunfo, and J. E. Harris (2012). Tobacco control campaign in Uruguay: a population-based trend analysis. *The Lancet 380*(9853), 1575–1582.

Abdi, H. and L. J. Williams (2010). Principal component analysis. *Wiley Interdisciplinary Review: Computational Statistics 2*(4), 433–459.

Anderson, M., A. Baer, K. Stampley, and A. McMains (2003). Fast food nation. *Adweek 44*(38), 26.

Arber, S. (1997). Comparing inequalities in women's and men's health: Britain in the 1990s. *Social Science & Medicine 44*(6), 773–787.

Arber, S. and H. Cooper (1999). Gender differences in health in later life: the new paradox? *Social Sciences & Medicine 48*(1), 61–76.

Barrier, B. (2004). McDonald's, Greene go the distance to strike healthful balance. *Nation's Restaurant News 38*(24), 40.

Berry, D. (2004). In pursuit of wellness. *Dairy Foods 105*(5), 34–38.

Blaxter, M. (1990). *Health and lifestyles*. Routledge.

Branen, A. L., P. M. Davidson, S. Salminen, and J. Thorngate (2001). *Food Additives*. CRC Press.

BusinessWire (2016, June). Global mHealth Market Forecast to 2022 - Increasing Demand for Patient Centric Healthcare Models - Research and Markets.

Cameron, A. and P. Trivedi (1998). *Regression Analysis of Count Data*. Cambridge University Press.

Chung, W., S. Lim, and S. Lee (2010). Factors influencing gender differences in smoking and their separate contributions: Evidence from South Korea. *Social Science & Medicine 70*(12), 1966–1973.

Czech Statistical Office (2018). Life expectancy by sex and age - time series.

Daňková, Š. (2010). European Health Interview Survey in CR - EHIS CR. Mental health, vitality and cognitive abilities. *Aktuální informace Ústavu zdravotnických informací a statistiky České republiky 12*, 1–12.

Denton, M., S. Prus, and V. Walters (2004). Gender differences in health: a Canadian study of the psychosocial, structural and behavioural determinants of health. *Social Science & Medicine 58*(12), 2585–2600.

Denton, M. and V. Walters (1999). Gender differences in structural and behavioural determinants of health: an analysis of the social production of health. *Social Sciences & Medicine 48*(9), 1221–1235.

Divine, R. L. and L. Lepisto (2005). Analysis of the healthy lifestyle consumer. *Journal of Consumer Marketing 22*(5), 275–283.

Frankenfield, D. C., W. A. Rowe, R. N. Cooney, J. S. Smith, and D. Becker (2001). Limits of body mass index to detect obesity and predict body composition. *Nutrition 17*(1), 26–30.

Hill, C. (2014, April). Impact de l'augmentation des prix sur la consommation de tabac. Institut Gustave Roussy.

Hraba, J., F. Lorenz, G. Lee, and Z. Pechačová (1996, November). Gender differences in health: evidence from the czech republic. *Social Science & Medicine 43*(10), 1443–1451.

Hughner, R. S., P. McDonagh, A. Prothero, C. J. S. II., and J. Stanton (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. *Journal of Consumer Behaviour 6*(2-3), 94–110.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters 31*(8), 651–666.

Jha, P. and R. Peto (2014). Global Effects of Smoking, of Quitting, and of Taxing Tobacco. *The New England Journal of Medicine 370*, 60–68.

Kebza (2014). *Zpráva o zdraví obyvatel České republiky*, Chapter 5.8, pp. 118–120. Ministry of Health of the Czech Republic.

Kim, D. and S. Jang (2017). Stress and food choices: Examining gender differences and the time horizon framing effect. *International Journal of Hospitality Management 67*, 134–142.

McDonough, P. and V. Walters (2001). Gender and health: reassessing patterns and explanations. *Social Science and Medicine 52*(4), 547–559.

Ministry of Health of the Czech Republic (2014). Zpráva o zdraví obyvatel České republiky. Praha.

Mirowsky, J. and C. E. Ross (1995). Sex Differences in Distress: Real or Artifact? *American Sociological Review 60*(3), 449–468.

National Institute of Public Health (2012). Studie helen.

O'Connor, D. B., F. Jones, M. Conner, B. McMillan, and E. Ferguson (2008). Effects of daily hassles and eating style on eating behavior. *Health Psychology 27*(1), 20–31.

Payne, S. (2004). Gender in lung cancer and smoking research. World Health Organisation.

Plotnikoff, R. C., A. Mayhew, N. Birkett, C. A. Loucaides, and G. Fodor (2004). Age, gender, and urban-rural differences in the correlates of physical activity. *Preventive Medicine 39*(6), 1115–1125.

Research2guidance (2015, May). Eu countries' mhealth app, market ranking 2015.

Riley, W. T., D. E. Rivera, A. A. Atienza, W. Nilsen, S. M. Allison, and R. Mermelstein (2011). Health behavior models in the age of mobile interventions: are our theories up to the task? *Translational behavioral medicine 1*(1), 53–71.

Rutters, F., A. G. Nieuwenhuizen, S. G. T. Lemmens, J. M. Born, and M. S. Westerterp-Plantenga (2009). Acute stress-related changes in eating in the absence of hunger. *Obesity 17*(1), 72–77.

Seefeldt, V., R. M. Malina, and M. A. Clark (2002). Factors Affecting Levels of Physical Activity in Adults. *Sports Medicine 32*(3), 143–168.

Shapiro, L. and N. Biddle (1996). To your health. *Newsweek 127*(4), 52.

Slining, M. M., S. W. Ng, and B. M. Popkin (2013). Food companies' Calorie-Reduction Pledges to Improve U.S. Diet. *American Journal of Preventive Medicine 44*(2), 174–184.

Steptoe, A., J. Wardle, W. Cui, F. Bellisle, A.-M. Zotti, R. Baranyai, and R. Sanderman (2002). Trends in smoking, diet, physical activity and attitudes to health in young adult Europeans from 13 countries, 1990-2000. *Preventive Medicine 35*(2), 97–104.

von Bothmer, M. and B. Fridlund (2005). Gender differences in health habits and in motivation for a healthy lifestyle among Swedish university students. *Nursing and Health Sciences 7*(2), 107–118.

Walbeek, C. V. (2006). Industry responses to the tobacco excise tax increases in South Africa. *South African Journal of Economics 74*(1), 110–122.

Walters, V., P. McDonough, and L. Strohschein (2002). The influence of work, household structure, and social, personal and material resources on gender differences in health: An analysis of the 1994 Canadian national population health survey. *Social Science & Medicine 54*(5), 677–692.

Wardle, J. and J. Griffith (2001). Socioeconomic status and weight control practices in British adults. *Journal of Epidemiology & Community Health 55*(3), 185–190.

Wardle, J., A. M. Haase, A. Steptoe, M. Nillapun, K. Jonwutiwes, and F. Bellisie (2004). Gender differences in food choice: The contribution of health beliefs and dieting. *Annals of Behavioral Medicine 27*(2), 107–116.

Weaver, J. (2017). *Practical Guide to Obesity Medicine*. Elsevier Health Sciences.

Wooldridge, J. M. (2015). *Introductory Econometrics: A modern approach*. Nelson Education.

World Health Organisation (2003). *The European Health Report 2002*. World Health Organisation Regional Office for Europe.

World Health Organisation (2013). Draft: comprehensive global monitoring framework and targets for the prevention and control of non-communicable diseases.

World Health Organisation (2014). Total alcohol per capita (15+) consumption in litres of pure alcohol, 2010.

World Health Organisation (2015, January). Alcohol fact sheet.

World Health Organisation (2017, May). Tobacco fact sheet.

World Health Organisation Regional Office for Europe (1999). Healthy living: What is a healthy lifestyle? techreport, World Health Organisation.

List of Tables

1 Cluster means 22
2 Proportion of Variance explained by components 23
3 PCA results 24
4 Results of individual results regressions 30
5 Descriptive Statistics - the whole sample 43
6 Descriptive Statistics - women 44
7 Descriptive Statistics - men 45
8 Correlation Matrix - part 1 46
9 Correlation Matrix - part 2 47

List of Figures

1 Optimal number of clusters 21

Appendix A

	Min	Max	Mean	Median	Standard deviation
Age	18.00	93.00	47.85	47.00	18.36
Weight	40.00	140.00	76.02	75.00	15.47
Height	132.00	205.00	170.80	170.00	9.26
BMI	15.57	45.17	26.01	25.54	4.64
Partnership	0.00	1.00	0.61	1.00	0.49
Education level	1.00	7.00	4.42	4.00	0.96
Education years	0.00	28.00	12.39	12.00	2.73
Occupation	1.00	10.00	5.08	5.00	2.34
Problem sleep	1.00	5.00	1.82	1.00	1.04
Problem tired	1.00	5.00	1.90	2.00	0.96
Depressed	1.00	5.00	1.92	1.00	1.04
Worried	1.00	5.00	1.80	1.00	1.02
Fruit	0.00	30.00	1.80	1.00	1.89
Vegetables	0.00	7.00	1.30	1.00	0.86
Smoking dummy	0.00	1.00	0.30	0.00	0.46
Alcohol	0.00	120.00	5.63	2.00	9.64
Vig. activity	0.00	7.00	1.13	0.00	2.01
Mod. activity	0.00	7.00	3.41	3.00	2.90
Week walking	0.00	98.00	17.69	10.50	19.13

Table 5: Descriptive Statistics - the whole sample

	Min	Max	Mean	Median	Standard deviation
Age	18.00	89.00	48.57	49.00	18.31
Weight	40.00	120.00	69.23	68.00	13.43
Height	132.00	186.00	165.30	165.00	6.52
BMI	15.57	45.17	25.38	24.81	4.85
Partnership	0.00	1.00	0.61	1.00	0.49
Education level	1.00	7.00	4.35	4.00	0.97
Education years	0.00	20.00	12.11	12.00	2.72
Occupation	1.00	10.00	4.56	4.00	2.13
Problem sleep	1.00	5.00	1.97	2.00	1.09
Problem tired	1.00	5.00	2.01	2.00	0.99
Depressed	1.00	5.00	1.90	1.00	1.112
Worried	1.00	5.00	1.96	2.00	1.10
Fruit	0.00	30.00	1.94	1.00	1.88
Vegetables	0.00	7.00	1.40	1.00	0.92
Smoking dummy	0.00	1.00	0.23	0.00	0.42
Alcohol	0.00	28.00	2.41	1.00	3.83
Vig. activity	0.00	7.00	0.73	0.00	1.62
Mod. activity	0.00	7.00	3.56	3.00	2.94
Week walking	0.00	98.00	17.80	14.00	18.42

Table 6: Descriptive Statistics - women

	Min	Max	Mean	Median	Standard deviation
Age	18.00	93.00	46.95	45.00	18.39
Weight	50.00	140.00	84.30	83.00	13.66
Height	158.00	205.00	177.50	178.00	7.53
BMI	16.60	44.19	26.79	26.17	4.24
Partnership	0.00	1.00	0.61	1.00	0.49
Education level	2.00	7.00	4.50	4.00	0.94
Education years	8.00	28.00	12.74	12.00	2.71
Occupation	1.00	10.00	5.58	7.00	2.42
Problem sleep	1.00	5.00	1.63	1.00	0.93
Problem tired	1.00	5.00	1.76	1.50	0.90
Depressed	1.00	5.00	1.50	1.00	0.87
Worried	1.00	5.00	1.60	1.00	0.89
Fruit	0.00	24.00	1.62	1.00	1.89
Vegetables	0.00	6.00	1.17	1.00	0.75
Smoking dummy	0.00	1.00	0.38	0.00	0.49
Alcohol	0.00	120.00	8.96	6.00	12.32
Vig. activity	0.00	7.00	1.635	0.00	2.31
Mod. activity	0.00	7.00	3.22	3.00	2.85
Week walking	0.00	84.00	17.57	10.50	19.99

Table 7: Descriptive Statistics - men

Appendix B

	Gender	Age	Partnership	Education level	Education years	Occupation	Weight	Height	BMI	Smoking dummy
Gender	1.000	-0.044	0.006	0.078	0.114	0.219	0.485	0.657	0.151	0.168
Age		1.000	0.136	-0.191	-0.224	-0.066	0.127	-0.304	0.334	-0.184
Partnership			1.000	0.057	0.062	-0.075	0.140	-0.047	0.177	-0.034
Education level				1.000	0.870	-0.679	-0.036	0.178	-0.156	-0.122
Education years					1.000	-0.600	-0.010	0.226	-0.152	-0.104
Occupation						1.000	0.163	0.094	0.135	0.215
Weight							1.000	0.493	0.843	0.018
Height								1.000	-0.045	0.083
BMI									1.000	-0.034
Smoking dummy										1.000

Table 8: Correlation Matrix - part 1

	Alcohol	Fruit	Vegetable	Problem sleep	Problem tired	Depressed	Worried	Vig. activity	Mod. activity	Week walking
Gender	0.340	-0.084	-0.132	-0.164	-0.128	-0.193	-0.176	0.225	-0.058	0.008
Age	-0.091	0.006	-0.081	0.349	0.232	0.217	0.209	-0.215	-0.172	-0.145
Partnership	-0.028	0.021	-0.003	-0.003	-0.009	-0.017	0.016	-0.008	0.158	0.019
Education level	0.032	0.018	0.094	-0.178	-0.129	-0.184	-0.152	0.051	0.010	-0.080
Education years	0.041	0.021	0.079	-0.183	-0.139	-0.203	-0.164	0.073	0.005	-0.023
Occupation	0.061	-0.040	-0.079	0.007	-0.071	0.018	-0.029	0.088	0.127	0.190
Weight	0.172	-0.048	-0.109	-0.076	-0.068	-0.097	-0.065	0.163	0.052	0.031
Height	0.230	-0.004	-0.006	-0.246	-0.177	-0.209	-0.192	0.261	0.040	0.035
BMI	0.057	0.049	-0.121	0.065	0.038	0.020	0.046	0.026	0.034	0.019
Smoking dummy	0.196	-0.061	-0.024	-0.078	-0.088	-0.017	-0.022	0.068	0.102	0.082
Alcohol	1.000	-0.079	0.058	-0.092	-0.041	-0.039	-0.035	0.147	0.076	-0.007
Fruit		1.000	0.285	0.022	0.030	-0.016	0.002	0.011	0.073	0.045
Vegetable			1.000	-0.031	-0.023	0.030	0.028	0.041	0.049	-0.009
Problem sleep				1.000	0.642	0.479	0.475	-0.123	-0.144	-0.058
Problem tired					1.000	0.499	0.505	-0.049	-0.118	-0.086
Depressed						1.000	0.863	-0.030	-0.069	0.015
Worried							1.000	-0.030	-0.068	0.007
Vig. activity								1.000	0.200	0.078
Mod. activity									1.000	0.150
Week walking										1.000

Table 9: Correlation Matrix - part 2