CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

BACHELOR THESIS

Impact of socio-demographic aspects on training and contest intensity among runners

2020 **Jan Klásek**

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 3, 2020

Signature

Abstrakt

Běh se stal v posledních letech populární volnočasovou aktivitou. Hlavní poslání této práce je odhalit efekt socio-demografických faktorů na objem tréninku amatérských bežců a následnou pravděpodobnost jejich účasti na závodech. Data pro tuto analýzu byla získána pomocí vlastního on-line dotazníku. Objem tréninku byl odhadnut pomocí metody nejmenších čtverců, pravděpodobnost účasti na závodech poté logistickou regresí. Individuální motivační faktory a zdravotní indikátory mají největší efekt. Stupeň vzdělání, ani demografické faktory nejsou statisticky signifikantní, příjem domácnosti má vliv pouze na účast na závodech.

Klíčová slova	běh, běžecké závody, fyzická aktivita, metoda nejmenších čtverců (OLS), logistická regrese

E-mail autora	`klasekjan@gmail.com`
E-mail vedoucího práce	`jana.votapkova@fsv.cuni.cz`

Abstract

Running has become very popular leisure time activity in the recent years. The main intention of this thesis is to uncover how various socio-demographic factors influence the training intensity among leisure time runners and their consecutive participation at the races. The data for the analysis ware collected through the self-developed online questionnaire. The volume of the training was estimated using OLS method, probability of participation was analyzed by the logistic regression. Individual motivational factors and health indicators have the biggest effect. Education nor demographic factors are not statistically significant. Household income has an impact only on the race participation.

Keywords	running, running races, physical activity, ordinary least squares (OLS), logistic regression

Author's e-mail	`klasekjan@gmail.com`
Supervisor's e-mail	`jana.votapkova@fsv.cuni.cz`

Bachelor's Thesis Proposal

Author	Jan Klásek
Supervisor	Jana Votápková
Proposed topic	Impact of socio-demographic aspects on training and contest intensity among runners

Motivation Over 26% of people in the Czech Republic are considered obese, which places our country among the top 10% of the most obese countries in the world (WHO, 2015). Moreover 15% of children aged 15 are considered overweight or obese (OECD, 2017). Men spend on average 3.6 hours per week doing sport or fitness activity, women spend 3.2 hours (ČSÚ, 2018).

However, we can observe increasing interest in running as a free-time activity, as running is one of the least demanding sports in terms of time and equipment. Consequently, every weekend several races are organized for amateur runners by major companies such as RunCzech, RunTour, Nightrun (city races) or Behej Lesy, Horska Vyzva, Spartan Race (trail races) as well as many local events.

This thesis is motivated by the interest in the sports industry and healthy lifestyle. The objective is to find out what socio- demographic factors drive people in the Czech Republic to show interest in healthier lifestyle, increased amount of time spent doing a physical activity and consequent attendance at the related sport events. An econometric analysis using the data collected from runners and supported by the data from the main event organizers will be carried out.

Contribution Several studies examining the determinants of exercise intensity and length have been conducted in the United States using data from the National Health and Nutrition Examination Surveys (NHANES), examples include Meltzer & Jena (2010) or Melayne & McInnes (2011). These papers are introducing a model of exercise behavior as a function of age, race, education etc. The purpose of this theses is to perform a similar research

for the Czech Republic and focus it on one specific type of exercise – long-distance running as well as adding some new potentially important variables into the model. Thesis will also consider questions regarding principles of healthy lifestyle.

Methodology This thesis aims to estimate the impact of several socio-demographic aspects on time spent doing physical activity per week (focused mainly on long-distance running) and the subsequent participation at the weekend races. Questionnaires will be distributed among at least 300 runners and additional data will be obtained from the RunCzech company as the major organizer of the races with the number of participants exceeding 50000 every year.

Two models will be built and will take the following form:

$$y_i = \alpha + \beta X_i + u$$

where the dependent variable y_i is a number of hours spent training per week, and alternatively, the number of races attended per year. X_i is a matrix of independent variables and will include for example, age, the level of education completed, number of hours at work a week, marital status, number of children living in a household, number of years since the individual started running, etc. Given the distribution of the variables, appropriate econometric models will be applied (OLS, logit, etc.)

Outline

1. Introduction
2. Literature review and hypotheses
3. Data
4. Methodology
5. Results
6. Discussion
7. Conclusion

Bibliography

Meltzer, D. O., Jena, A. B. (2010). The economics of intense exercise. *Journal of Health Economics*, 29(3), 347-352.

Global action plan on physical activity 2018-2030: more active people for a healthier world. Geneva: World Health Organization; 2018

Mitáš, J., Ding, D., Frömel, K., Kerr, J. (2014). Physical activity, sedentary behavior, and body mass index in the Czech Republic: A nationally representative survey. *Journal of Physical Activity and Health*, 11(5), 903-907.

Nejedlá, M. (2014). Zdravotní a hospodářské důsledky epidemie obezity a možnosti její prevence ve školách. *Česká antropologie,* 64(supplementum), 20-24

Contents

Chapter 1

Introduction

Running was for a long time considered as a purely competitive sport. For women even any long-distance running was considered as very dangerous. In fact, the American Kathrine Switzer was attempted to be stopped during the 1967 Boston Marathon and any distance longer than 1500 metres was not present at the Olympic Games up until 1984, when 3000 metres race was introduced together with the most distinguished marathon.

Since than, running or jogging (running at lower intensity) has become a worldwide spread leisure time activity with hundreds of runners filling the major city roads and parks as well as local forest trails and mountains. It is widely available type of activity due to nearly no equipment or sport facility needed. Races are since than not just for a few top athletes, on the contrary, tens of races are held every weekend all around each country. The most famous races nowadays attract over 50,000 mostly amateur runners from all around the world, creating an interesting economic opportunity for many cities. Solely the Czech Republic has the most highest ranked races by the World Athletics, meaning those races are attended by the top world athletes and attract significant attention.

However, despite the fact the Czech Republic is the nation of runners like Emil Zátopek or Jarmila Kratochvílová, it is also considered as one of the most unhealthy countries in Europe with obesity rate exceeding 26%. This arises from the fact that nowadays most of the adolescents do not have

enough physical activity and bad habits are created for their future.

Many studies examined the level and intensity of the physical activity undertaken by an individual based on several socio-demographic factors, but very few are focused on a specific type of activity or sport. Also previous studies do not take into account health or specific motivational factors. These findings would allow proposed policies to be focused more accurately.

This thesis has the following structure. Chapter 2 summarizes the plans and recommendations of the World Health Organization for the following decade. It provides guidelines to increase the level of physical activity across the population. Consequently, it acknowledges the reaction of the Czech government to the above mentioned document. Chapter 3 reviews the existent literature covering the both economic and health consequences of running particularly together with increased level of physical activity. In the second part of this chapter, the studies from different countries examining effect of various socio-demographic factors on the level of physical activity are summarised. Chapter 4 describes the process of data collection, reveals its structure and provides the descriptive statistics of the variables. Chapter 5 explains the two main methodological concepts used it the thesis. Results of the estimation are presented in Chapter 6, followed by discussion of the results. The last chapter concludes the thesis and suggests the further possible extensions.

Chapter 2

Physical activity as a policy priority

In 2018 the World Health Organization (WHO) published a document called *Global action plan on physical activity 2018-2030: more active people for a healthier world*. It summarizes current situation regarding the health of the population together with future goals and recommendations to be implemented at the national level. WHO estimates that 25% of adults and 75% of adolescents do not meet global recommendations for physical activity (150 minutes of moderate-intensity aerobic physical activity or 75 minutes of vigorous intensity aerobic physical activity per week for adults and 60 minutes of accumulated daily vigorous intensity aerobic physical activity for adolescents). Tightly connected with rising global inactivity are direct healthcare costs together with costs arising from lost labour productivity.

The plan induces policies that should promote healthy lifestyle in order to achieve many of the 2030 Sustainable Development Goals (SDGs). Those include primarily SDGs 2-5, 8-13, 15-17 (Appendix 1).
Four objectives were set to fulfill these goals:

- *Create active societies*
  "Create a paradigm shift in all of society by enhancing knowledge and understanding of, and appreciation for, the multiple benefits of regular physical activity, according to ability and at all ages."

- *Create active environments*

  "Create and maintain environments that promote and safeguard the rights of all people, of all ages, to have equitable access to safe places and spaces, in their cities and communities, in which to engage in regular physical activity, according to ability."

- *Create active people*

  "Create and promote access to opportunities and programs, across multiple settings, to help people of all ages and abilities to engage in regular physical activity as individuals, families and communities."

- *Create active systems*

  "Create and strengthen leadership, governance, multisectoral partnerships, workforce capabilities, advocacy and information systems across sectors to achieve excellence in resource mobilization and implementation of coordinated international, national and subnational action to increase physical activity and reduce sedentary behaviour."

According to WHO, the Czech Republic was the most obese country in the European region. Czech government, as a reaction, implemented the Health 2020 action plan with two strategic objectives: *Improving health for all and reducing health inequalities* and *Strengthening governance for health by the public sector and involving the whole of society, communities and individuals in decision-making* (Ministry of Health of the Czech Republic, 2014).

Strategic objectives should be tackled in four independent priority areas for policies:

- Investing in health and disease prevention through a life-course approach, empowering people, providing safe and sustainable environment for the attainment of their full health potential

- Tackling major health challenges of noncommunicable and communicable diseases, continuous health monitoring

- Strengthening people-centred health systems, ensuring the availability and accessibility of health services, focusing on health protection and promotion and disease prevention, strengthening public health capacities and emergency preparedness, continuous health monitoring and securing appropriate emergency response
- Creating resilient communities living in supportive environment

Chapter 3

Literature review and hypotheses

3.1 Economic and health consequences of running

Studies agree that aerobic exercise contributes to prevention, both primary and secondary, of several chronic diseases and reduces risk of premature death (Warburton et al., 2006; Pedisic et al., 2019). There is evidence that aerobic exercise provides significant improvements in depression, even with comparison to psychotropic treatment. (Penedo and Dahn, 2005)

As the lower threshold for the individual physical activity WHO suggests 150 minutes per week (WHO, 2010), however even activity groups exercising on average 92 min per week showed a 14% reduced risk of mortality and 3 year longer life expectancy compared with individuals in the inactive group. (Wen, 2011)

The Copenhagen City Heart Study (CCHS) (Schnohr et al., 2012) is considered to be the core paper on health consequences of running specifically. This study investigates association between jogging and predispositions to all-cause mortality on a random sample of 17,589 participants of ages between 20 and 98, 1,878 of them considered as joggers, others considered as healthy people but non-joggers. Those people were invited to the study between years 1976 and 2003. In this study the expected lifetime was cal-

culated by integrating the predicted survival curve estimated in the Cox model. This model is generally used for analysing the survival of patients in a clinical trials using modelling techniques for exploring the relationship between the survival of a patient and several explanatory variables. (Walters, 2009)

During the 35-year maximum period the age adjusted hazard ratios (risk of dying at a certain time, adjusted for age; value >1 suggest increased risk, <1 shows decreased risk) among joggers was 0.56 (95% CI; 0.46, 0.67) for men and 0.56 (95% CI; 0.4, 0.8) for women. The age adjusted survival increase for joggers was 6.2 years for men and 5.6 years for women in CCHS.

The optimal doses of running for the best potential health benefits were estimated at quite low doses of about 2.4 hours per week. At this amount the runners showed the minimal susceptibility to cardiovascular diseases (CVD) and all-cause mortality. (Lavie et al., 2015).

Mental benefits are also closely connected to distance running as runners showed the increase level of euphoria after exercise, supporting the theory of "runner's high", the positive state after long runs caused by the increased level of endorphins in the brain. (Boecker, 2008)

Several studies also reviewed the potential negative effect of high doses of running (marathons, ultra marathons, Ironmans), showing a visible increase in atrial size and dilation of the ventricular chambers, mainly the right ventricle, with a reduction in function, especially of the right ventricle and the ventricular septum. These consequences were typically resolved within 3 days after the race, with typically no obvious permanent effects. The death rate at marathons for CVD was estimated as 0.54 per 100,000 participants, although it was criticized by some experts, that the numbers were contaminated by other participants, e.g. half-marathoners and did not count for the periods after the race.

Economic consequences need to be divided into two parts. Firstly, the benefits for the whole society arise from lowering the necessary healthcare costs by improving the overall health of the population and interconnected

increase in labour productivity. Secondly, the economic benefits for places organizing the events are taken into consideration.

Global costs of physical inactivity were estimated to be INT$54 billion in direct healthcare per year with additional INT$14 billion in lost productivity in 2013, inactivity also accounts for 1-3% in national healthcare costs (Ding, 2016). In the Czech Republic it is estimated that due to obesity 48,192 years of life are lost annually, causing economic lost of CZK 37 - 81 billion as a sum of diagnostics and healthcare cost and consequences of lower tax payments and lost productivity (Nejedla, 2014). The economic benefits of exercise are mostly shown for physically active older adults (age 45+) as the disease-prevention healthcare costs avoided highly outweigh (about three times lower costs) the incurred costs compared to their non-active peers. For younger adults (15 to 44 years) the saved prevention costs are compensated by direct costs arising from the sports activities (injuries, prevention etc.) (Nichol et al., 1994).

The recent data of economic consequences of domestic races come from CzechTourism analyzing RunCzech, the main competition organizer in the Czech Republic. Numbers account for the year 2011 using input-output analysis of direct and indirect (multiplier) impacts. As many as 5 races took place in 2011 - the Prague International Marathon; half-marathons in Prague, Olomouc and Usti nad Labem and 10k race in Prague; together attended by around 82,000 people, 25,000 of whom were runners, others include family support, spectators, volunteers and the race staff. Total spending of all people were CZK 215 million, 138 million of which were spend by foreign visitors, therefore creating a new income for the Czech economy. The total impact of expenditures by visitors was an increase in turnover by CZK 333 million, of which 103.5 million contributed to an increase of GDP, 249 full-time jobs were created thanks to the expenditures.

Since half of the events were held in Prague, the events had a big impact on the local economy, increasing the local GDP by at least 41.6 million and created 153 permanent jobs.

In 2019 Runczech organized 8 races in 3 more cities compared to 2011 - Ceske Budejovice, Karlovy Vary and Liberec with the significant increase in attendance[1] and thus with expected economic benefits.

race		individuals	team races	overall
Prague Marathon	men	7114		10554
	women	3440		
Prague HM	men	5663	566	7866
	women	1637		
Karlovy Vary HM	men	1896	676	3459
	women	887		
Ceske Budejovice HM	men	1634	670	3030
	women	726		
Olomouc HM	men	3107	1142	5612
	women	1363		
Prague 10k	men	3873		6199
	women	2326		
Prague 5k	women	2207		2207
Usti nad Labem HM	men	1633	564	2913
	women	716		
Liberec Nature Run	men	1338		2116
	women	778		

Apart from RunCzech there are several more event organizers with thousands of participants - RunTour, NightRun (both 8 cities), Behej Lesy (8 locations) and hundreds of local event organizers.

Recently RunCzech tries to expand to another European countries. Since 2017, they also co-organize two races in Italy - Napoli City Half Marathon and Sorrento Possitano Ultra-marathon. In 2019 RunCzech became a part of a SuperHalfs project, series of 5 half-marathons around the Europe.

The biggest marathon in terms of number of participants in the world is the New York City Marathon (NYCM) with 50,000 runners and over 2 million spectators annually. The estimated economic impact of 2014 NYCM was $415 million, with estimated cost of only $35 million. One of the way to get a place in the race is the fundraising for the charity, $34.5 million was

[1] Data obtained by direct request to RunCzech in December 2019

raised for about 8,500 charities across the United States (Forbes, 2015).

The calculation of the economic impact of marathons and sporting events in general has been criticized by both Cobb and Olberding (2007) and Papanikos (2015). Cobb emphasizes the import substitution effect, an effect where local participants would otherwise spend the money outside their local area if the marathon did not take place. He performed a study of a 2006 Flying Pig Marathon in Cincinnati, finding out that over half local marathon participants would have run a different race outside their city had the marathon not taken place. He suggests better diversification of surveys addressed for local and non-local runners in order to estimate the economic impact more accurately. Papanikos makes a case study of The Athens Marathon concluding that even the marathon is a perfect example of an event with a real brand name, it provides no long-term positive effects for the city, although he does not blame the people connected with the organization of the event.

3.2 Summary of studies conducted in the USA, England and the Czech Republic

Becker (1965) builds a fundamental model of household behaviour and allocation of time among different activities. He puts together demand for goods, labor supply and time costs in the households, moreover he distinguishes between time in different part of the week or day as not having the same cost. Sleep-Leisure-Occupation-Transportation-Home-based activities (SLOTH) model (Pratt et al., 2004) is a time-budget model which has been discussing the potential interventions that might increase physical activity in four non-sleep domains, since the mean time slept has been found as non-changing in the past decades. Each policy is to be carefully and critically evaluated on four criteria: *economic efficiency* (measuring the ratio of economic costs to economic benefit of the selected policy), *equity* (the degree to which the policy redistributes benefits fairly across society), *effectiveness* (likelihood of reaching the goals and objectives of the selected policy) and *feasibility* (likelihood that the selected policy would be adopted by the elected officials).

For instance if the policymakers need to evaluate the paid advertisement through TVs and other mass media promoting activity in children. This intervention is said to be economically justifiable since it tackles information deficit and irrationality, it is equitable as most children have access to TV and time spent in front of is proven to be positively correlated with obesity. The feasibility is justifiable, however the efficiency of this type of intervention alone is questionable, nevertheless can be effective together with other programs that might provide places or community support for the physical activity.

Studies evaluating the relationship between several socio-demographic factors and physical activity and its intensity have been performed predominantly in the USA.

Both education and income have been found to be positively correlated with the level of physical activity by both genders. (McInnes and Shinogle,

2011) Effects of another several socioeconomic factors on leisure time physical activity, divided into three models according to intensity of exercise, have been estimated in their study across the USA collected via Behavioral Risk Factor Surveillance System (BRFSS) survey. This survey also provided country codes and therefore authors were able to add supplementary, country-specific data and evaluating the policy effects state by state. The effect of marriage decreases the level of exercise overall, however if married, men are more likely to exercise than women, independently on the income level. Working population has lower engagement in exercise than people who do not work for any reason and the effect is multiplied when the income is low. Yet their study lacks the measures of time constraints in family including the family size and number of dependent individuals to investigate the results deeper. Similar conclusions have been drawn among seven most popular sports in England (swimming, cycling, gym exercise, running, aerobics, football or rugby and racket sports)(Farrel and Shields, 2002). Positive relationship to education has been found for specific sports, particularly swimming, gym exercise, running, aerobics and racket sports.

The impact of area-specific variables, such as price of exercise (e.g. gym memberships), sporting goods and transportation costs proved to be economically significant only in case of transportation. Nevertheless it is necessary to interpret this results cautiously, since individuals interested in exercise may want to live in areas providing those facilities. In this case, it should be noted that American cities have historically different urban geography compared to cities in Europe. In Europe higher income groups tend to concentrate in the city centers, whereas in the USA they predominantly concentrate in suburbs.

As the policies across the USA differ, it was possible to measure the effect of sin taxes, which were found out not to have any significant effect, suggesting that smoking or drinking and inactivity to be substitutes. Overweight and obesity is generally associated with lower exercise level, although with some minor exceptions.

Although the clear positive relationship between participation in physical activity and increasing income has been proven, there occurs a negative relationship between time spent on physical activity and the individual income. (Humphreys and Ruseski, 2007)

Meltzer and Jena (2010) come to the similar findings based on a different dataset, putting and emphasis on the empirical results of relationship between income and exercise intensity stating that exercise intensity rises with increasing wage, supporting the hypothesis that increased time costs are compensated with higher exercise intensity.

The effect of drug intake has also been reviewed in their study. Drinking alcohol is positively related to sport participation, whereas smoking has negative impact. This result might emphasize the important social effect of team sports or group activities.

The studies conducted in the Czech Republic show similar results. The level of physical activity does not significantly differ across regions, however the level of activity significantly differs with size of the community (town), where smaller places show increased level of physical activity, which is in contrast to results from the USA (Mitáš et al., 2013). Men were more likely to be intensely active if lived in a place of less than 100,000 residents and also participated in some form an organised activity. Individuals in high Socioeconomic status (SES) quartile together with those living in a family with a children were more likely to meet recommended weekly amount of activity. In a contrast, females were likely to meet guidelines if they were in the low SES quartile. (Frömel et al., 2009). The model describing effect of physical activity and socio-demographic aspects on BMI showed no statistically significant relationship between occupational or household PA and BMI (Mitáš et al., 2014). Therefore potential policies should put stress on leisure time exercise to reach the desired effect of policies.

Chapter 4

Data

4.1 Data collection

The data for this thesis were obtained through an online questionnaire using the Google Forms service (with a translation in Appendix 3). This questionnaire was posted in December 2019 in the closed Facebook group "Běžci", which is the largest online community of runners of all performance levels in the Czech Republic. This group contains above 21,000 members (as for May 2020). Runners were motivated to fill in the questionnaire by providing 3 prizes:

- starting fee for one of the RunCzech race of one's own choice
- collapsible bottle
- 10 protein bars

The winners of the prizes were drawn 2 days after the first publishing of the questionnaire, however the questionnaire was thereafter re-posted several times in order to collect more answers. Altogether a total number of 468 filled-in questionnaires was collected.

Before advancing to the further analysis, the sample was considered for duplicates, outliers and other data disturbances. Eventually, the final sample used for the descriptive statistics and estimation contains 451 respondents.

The summary of all variables showed that the obtained sample contains

reasonable proportions across all age, income, educational and gender groups which should satisfy the random sample assumption in the OLS analysis. The question about membership in the running groups the author is himself also present in (Adidas Runners Prague, Mattoni Free Run) was included to avoid selection bias through FB algorithms which might highlight the post to the author's FB friends. The number of those observations in the sample was considered negligible, and those answers were kept in the sample, since the above mentioned groups have considerable number of members and their presence in the sample is expectable.

The only selection bias which appears to be present in the sample is the proportion of runners participating in the races, which is assumed to be overrated, since the membership in the FB group, following the sites and several interest groups might signal an increased interest in the topic and consequently participation in at least some number of races is expected.

4.2 Data structure and descriptive statistics

Based on the literature review and other consequent procedure, variables covering various aspects were obtained through the questionnaire. The complete statistics are provided in Table 8.1 and 8.2 in the Appendix 2.

Respondents were asked to rate subjective health on the scale from 1 to 5, with rating of 1 being the best. Intention was to compare the health before the individual started running and by the collection of the answers.

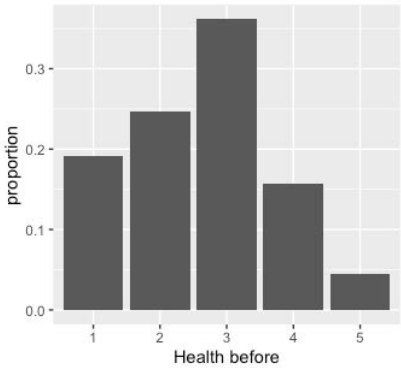

Figure 4.1: Subjective health before

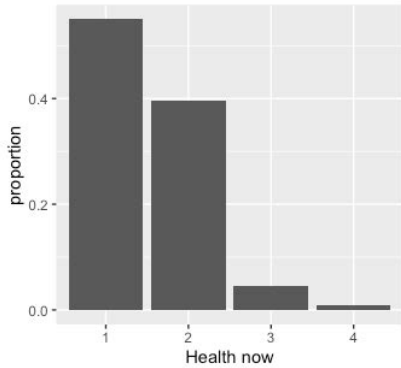

Figure 4.2: Subjective health now

Significant improvement can be observed with the worse ratings being barely present.

Other health and regimen variables also were considered for. Nearly 90% of runners reported themselves as non-smokers, however some heavy smokers can be observed in the sample with their consumption of more than 20 cigarettes per day, which is generally considered as incompatible with endurance sport. One fifth of the respondents admitted the regular consumption of alcohol every day or every other day, the interest was not put on whether the individual is an abstinent or drinks on rare occasions as this difference is not expected to have any significant effect on the performance. Majority of runners also reported an interest about their food intake which may take a form of any diet, counting of calories intake, dietary regimes or exclusion of some meals. BMI index, the common metrics allowing statistical comparison of the weight of people of different height, computed as $\frac{weight(kg)}{height(m)^2}$ ranges across all categories, some runners reported as underweight (BMI $\leq$ 18), but majority falling into the normal (healthy) weight category. As 75th percentile equals to 25.3 we can see that about 25% falls into either overweight or obese class I category. 13.5% reported taking regular medication, 7% being hospitalized in the past year for any cause. Injury in the past year was reported by 43.5% of respondents, this injury might be both running or non-running related, it was crucial to account for both, since even non-running injuries can be a consequence of bad movement stereotypes evolved by bad running technique and possible over-training.

The sample contains more women (53.9%) than men (46.1%), the same proportion is single as there are married individuals. The average age of runners is 35 years, as the distance running is considered to be for all age groups and the performance might peak at that age. Nevertheless, we can observe both teenage athletes as well as those with age over 70. The highest proportion of individuals has only high school diploma, but the share of all university degrees sums up to 46%. Working hours of individuals in the sample corresponds to the common 40 hours working week, 81.2% of

them having static or sedentary job. The potentially important financial situation was, based on previous research, researched by a household income per month. The respondents were to choose from the different net income ranges. As the number of the extreme values on both sides was small, the answers were divided to the three income classes: low (up to 30,000 CZK) - 20.2%, middle (30,001 CZK to 50,000 CZK) - 29.9% and high (over 50,000 CZK) - 33.5%. The remaining share (16.4%) did not know or were not willing to share their income. Some studies suggest the effect of the size of the community on the volume of the PA, therefore this variable was also obtained. The sample shows the numbers similar to the demography of the Czech Republic[1].

Since running and consequent training for races might be very time-consuming, it was necessary to account for the number of people living in the household together with the runner. Most of the runners reported living with at least one adult person (mean of 1.26), with an average of 1 child. But as older children can take care of themselves home alone for the duration of the training, point of interest was also a number of children below age of 10, which has only a mean of 0.54. As not only children might require increased attention, the number of dependent individuals (disabled, old, etc.) was surveyed, however this number was negligible.

In the sample, runners spend on average 4.64 hours on running training. This time variable was chosen instead of the common distance (kilometres per week), because of the diverse pace of running and the consequential disturbance and incomparability in the model. Nevertheless, depending on the pace, it correspond to 35 to 60 kilometers per week. The predominant motivation to start running was to relax, clear the head after work (54.8%), then active spending of the free time was a reason for 28.2%, cheap alternative to other sports for 4.7% and insufficient activity during the mere walks for 3.3%. Health issues were the main reason for only 9.1% of runners. Training in some group or community was reported by 38.6%, this number might be amongst other factors influenced by the fact that running groups exist

[1] Czech Statistical Office, 2019

mostly in bigger cities, and many runners do not have an approach to any of them. Average duration of the running career was 6 years, 26% reported repeated beginnings.

As running might not be only or main activity for the runners, time spend training might be influenced by sport/exercise activities (swimming, cycling, gym, team sports etc.). 81.4% reported doing such activity with a mean time of 3.26 hours.

Attendance to various races was reported by 87.6% of runners in the sample, however this percentage might be biased, since it is supposed to be correlated with the presence in the group the sample was taken from. The range differs between 1 and 50 (i.e. race every weekend during the year), with the mean number of 8 per year. The most preferred distance of the race was 10k to half marathon (48.6%) followed by distances up to 10k (29.2%). Unanticipated result, based on the participation at many city marathons was the last place of this distance, overtaken even by ultra-marathons. Main motivation to attending races was self satisfaction (54.1%) and active spending of the weekends (16.5%) or some appreciation of training km's (10.3%). Motivation to compare to other runners was chosen by 14.7% or win was only chosen by 4.4% and was the least preferred answer. This confirms the hypothesis of running as a free time activity rather than a competitive sport.

Chapter 5

Methodology

This chapter discusses appropriate methodological concepts. For the first model in this thesis, Ordinary Least Squares (OLS) estimation is used. In the second part of the chapter, logistic regression, intuition behind this methodology and interpretation of the results is explained.

5.1 Ordinary Least Squares Method

The first part of this thesis is to estimate how various socio-economic factors have an influence on the time spent doing physical activity, particularly running.

The OLS model is a basic linear regression tool, where estimates are obtained by minimising the sum of squared differences between predicted values and the actual values of the dependent variable.

Assuming linearity, this relationship can be described by the following multiple regression model equation:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \epsilon_i \tag{5.1}$$

where y_i denotes the dependent variable and $x_1, ..., x_k$ the independent variable, ϵ stands for the usual error term.

This model has to satisfy a set of assumptions to reach the desirable properties (Wooldridge (2012)):

MLR.1 - Linearity in parameters:

The model can be written as

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \epsilon$$

where $\beta_0, \beta_1, ..., \beta_k$ are the unknown parameters of interest and ϵ is an unobserved random error.

MLR.2 - Random Sampling

We have a random sample of n observations, $[(x_{i1}, x_{i2}, x_{ik}, y_i) : i = 1, ..., n]$ following the population model in MLR.1

MLR.3 - No Perfect Collinearity

In the sample none of the independent variable is constant and there are no perfect linear relationships among the independent variables.

MLR.4 - Zero conditional Mean

The expected value of the error term ϵ is zero given any values of the explanatory variables.

$$E(\epsilon|x_1, ..., x_k) = 0$$

MLR.5 - Homoskedasticity

Given any values of the independent variables, the error term ϵ has the same variance. Mathematically:

$$Var(\epsilon|x_1, ..., x_k) = \sigma^2$$

If the model satisfies assumptions MLR.1 - MLR.4, the estimates are unbiased, moreover if the MLR.5 is fulfilled, estimates are also efficient.

Having this fulfilled the problem lies in minimising the sum of square residuals (RSS):

$$\min_{\hat{\beta}_0 ... \hat{\beta}_k} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - ... - \hat{\beta}_k x_k)^2 \tag{5.2}$$

This minimization problem is solved using FOC, which leads to k + 1 linear equations with k + 1 unknowns $\hat{\beta}_0, ...\hat{\beta}_k$:

$$\begin{aligned} \sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - ... - \hat{\beta}_k x_k) &= 0 \\ \sum_{i=1}^{N} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - ... - \hat{\beta}_k x_k) &= 0 \\ &\vdots \\ \sum_{i=1}^{N} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - ... - \hat{\beta}_k x_k) &= 0 \end{aligned} \tag{5.3}$$

Solving for $\hat{\beta}_0, ...\hat{\beta}_k$ sample regression function for independent variables is obtained:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k \tag{5.4}$$

The estimates $\hat{\beta}_1, ..., \hat{\beta}_k$ are interpreted ceteris paribus. From the equation (5.4) we have:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + ... + \hat{\beta}_k \Delta x_k \tag{5.5}$$

where Δx_i denotes the change on variable x_i.

In this thesis assume time spent running depends on five matching factors, namely available time, motivational factors, family situation, health factors and financial situation.

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \epsilon_i \tag{5.6}$$

where y_i stands for hours per week running for an in individual i ($i = 1, ..., N$), x_{1i} stands for set of motivational factors, x_{2i} stands for set of family situation variables, x_{3i} stands for financial situation, x_{4i} stands for set of health factors and x_{5i} stands for set of time factors, ϵ stands for the standard error term.

5.2 Binary response model

In the second model the probability that the runner is attending running races based on the socio-economic variables is going to be estimated.

Such probability can be also obtained by linear regression, more specifically Linear Probability Model (LPM), but this technique has several drawbacks, when dealing with limited variables: 0 for not attending races, 1 if attending. This specific type is called binary variable. In this models, interest lies on the response probability:

$$P(y = 1|\mathbf{x}) = P(y = 1|x_1, x_2, ..., x_k), \tag{5.7}$$

where $\mathbf{x}$ denotes the full set of of explanatory variables, with the same structure as in the OLS model above. Variable y, the dependent variable, denotes the probability of attending the races.

Binary response models take form:

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) = G(\beta_0 + \mathbf{x}\beta) \tag{5.8}$$

where G is a function taking values strictly between 0 and 1:

$$G : 0 < G(z) < 1, \forall z \in R \tag{5.9}$$

which is satisfying the basic definition of probability, i.e. having its values between 0 and 1.

Several G functions exist, nevertheless the most preferred is the logistic function (model is thereafter called logit model)

$$G(z) = \frac{exp(z)}{1 + exp(z)} = \Lambda(z) \tag{5.10}$$

This function has the desirable properties: it is increasing with the biggest magnitude at $z = 0$, $G(z) \to 0$ as $z \to -\infty$ and $G(z) \to 1$ as $z \to \infty$. This function has a non-linear shape as shown in the figure below (blue line)[1]:

[1]Graph obtained from https://www.statisticssolutions.com/conduct-interpret-logistic-regression/

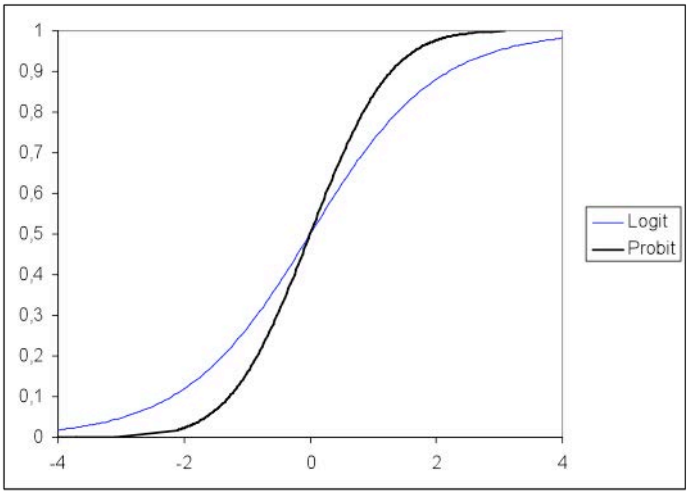

Figure 5.1: Logit model

Models can be derived from an underlying latent variable model. Let y^* be an unobserved (latent) variable such as

$$y^* = \beta_0 + \mathbf{x}\beta + u, y = 1[y^* > 0], \tag{5.11}$$

where the notation $1[\cdot]$ defines the binary outcome. The function $1[\cdot]$ is called the indicator function taking on following values:

$$y = \begin{cases} 1 & \text{if } y^* > 0 \\ 0 & \text{if } y^* \leq 0 \end{cases}$$

We assume u is independent of x and u has standard logistic distribution in case of logit model with mean of 0.

The response probability for y can than be derived as:

$$P(y = 1|\mathbf{x}) = P(y^* > 0|\mathbf{x}) = P[u > -(\beta_0 + \mathbf{x}\beta)|\mathbf{x}] = 1 - G[-(\beta_0 + \mathbf{x}\beta)] = G(\beta_0 + \mathbf{x}\beta) \tag{5.12}$$

In applications of the binary response models we are interested explaining effects of the x_j on the response probability $P(y = 1|\mathbf{x})$

In contrast to the LPM model, the magnitudes of each β_j are not representative by themselves due to the non-linear nature of the G function.

The partial effects of the explanatory variables can be obtained using calculus as:

$$\frac{\partial P(y=1|\mathbf{x})}{\partial x_j} = \frac{dG}{dx_j}(\beta_0 + \mathbf{x}\beta)\beta_j = g(\beta_0 + \mathbf{x}\beta)\beta_j \tag{5.13}$$

Since G is a strictly increasing function, the partial effect will always have the same sign as the coefficient β_j, therefore the positive or negative effect of the given explanatory variable can be determined directly from the regression table. However estimating the exact magnitude of the effect is not a straightforward process.

For binary variables the computation process is very simple. Let x_1 be a binary variable, than the partial effect of changing the value from 0 to 1 is equal to:

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + ... + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + ... + \beta_k x_k) \tag{5.14}$$

The difference above can also be used for other discrete variables (e.g. number of children). If x_1 denotes this variable, then the effect on the probability of x_1 changing from c to $c+1$ is simply

$$G(\beta_0 + \beta_1(c+1) + \beta_2 x_2 + ... + \beta_k x_k) - G(\beta_0 + \beta_1 c + \beta_2 x_2 + ... + \beta_k x_k) \tag{5.15}$$

But usually the interest lies in summarizing the partial effect of all variables, which is hard due to the fact, that the magnitude of the partial effect depends on x. There are two general methods for reporting the partial effects. The first one is to replace each explanatory variable with its sample average. Than the partial effect of x_j is equal to

$$g(\hat{\beta}_0 + \bar{x}\hat{\beta})\hat{\beta}_j = g(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + ... + \hat{\beta}_k\bar{x}_k)\hat{\beta}_j \tag{5.16}$$

where $g(z)$ stands for logit model. Using this approach we obtain the marginal effect of x_j of the "average" person in our sample, called Partial Effect at the Average (PEA). The second method called Average Partial Effect (APE) averages each individuals partial effects:

$$PEA = \frac{\sum_{i=1}^{n}[g(\hat{\beta}_0 + x_i\hat{\beta})\hat{\beta}_j]}{n} \tag{5.17}$$

The results of these two approaches provide different results. This is due to the fact that it might not make sense to interpret the effects for

the average person. Generally, having several non-continuous groups in the sample, the average does not represent any actual observation. Therefore having many binary or discrete variables the APE is more suitable method.

As mentioned above, due to the nonlinear nature of $E(y|x)$, OLS nor WLS cannot be used for the estimation. Instead, the method of Maximum Likelihood Estimators (MLE) is used. Under very general conditions, the MLE is consistent, asymptotically normal and asymptotically efficient. With standard errors provided by most econometric packages supporting probit of logit, (asymptotic) t tests and confidence intervals can be constructed similarly as with OLS or 2SLS.

The goodness of fit measure cannot be measured by R^2 as used with OLS, following from the process of obtaining the estimates. However several similar measures (usually called pseudo R^2) evaluating the model on a scale from 0 to 1 have been developed. Here, the McFadden's R^2 is used for evaluation.

$$\text{McFadden's } R^2 = 1 - \frac{\mathcal{L}_{ur}}{\mathcal{L}_o} \tag{5.18}$$

$\mathcal{L}_{ur}$ stands for the log-likelihood function for the estimated model and $\mathcal{L}_o$ denotes the log-likelihood function including intercept only.

Chapter 6

Estimation Results

In this part of the thesis the results of the estimations are presented, used methods and appropriate test are introduced. For the interpretation of the results, usual significance levels of 1%, 5%, 10% provided by the econometric packages were used. All estimations were performed using R software.

6.1 Training volume model

Results of the OLS regression are presented in the Table 6.1. The number of obtained independent variables from the survey is very high, many variables are categorical with several possible values. Including all of them in the model would lead to significant drop in the degrees of freedom in the model, even though the variables do not improve the model in any way. Therefore only independent variables that improve the model or are a point of research of this thesis are kept in the model. For the purpose of the estimation, the structure of some variables was changed to simplify its interpretation. The current health (even though subjective) is a crucial variable for the analysis. Due to low overall number of negative ratings (as shown in Figure 4.2, Chapter 4), the new variable was created (named $Health$ in the model), joining together values 1 and 2 - creating a level *good* health) and 3 and 4 ($worse$ health). Education was transformed using Likert Scale with five levels: 1 - primary education, 2 - high school, 3 - bachelor degree, 4 - master degree and 5 - doctoral degree or higher.

It can be supposed that presence of young children in the household has an effect on the amount of free time available, however it is necessary to account for the number of adult individuals living together with a runner, meaning the care of the children can be taken. Thus the interaction term uncovering the this relation is needed to be included in the model.

Table 6.1: OLS regression

	Dependent variable:
	Training hours
Constant	6.808*** (1.101)
Race motivation - Use of training km	1.694*** (0.387)
Race motivation - Win	2.560*** (0.623)
Motivation to begin - Insufficient activity	−0.757 (0.606)
Other activities - hours	0.052 (0.038)
Gender - Male	0.840*** (0.235)
Age	0.016 (0.013)
Education	0.003 (0.114)
Work hours	−0.002 (0.009)
Married	0.403 (0.267)
Sedentary job	−0.428 (0.294)
Worse health	−1.073** (0.509)
BMI	−0.147*** (0.041)
Injury	0.606*** (0.225)
Young children	−0.501** (0.239)
Young children:Adults	0.341** (0.151)
Observations	441
R^2	0.195
Adjusted R^2	0.167
F Statistic	6.874*** (df = 15; 425)

Standard errors in parantheses

Note: *p<0.1; **p<0.05; ***p<0.01

Income class of the household as a potential interest independent variable is not included in this model. Based on LR test it did not significantly improve the model and itself it did not have a statistically significant effect. The model shows the performance of R^2 reaches 19.5%, the F Statistic is equal to 6.874 on 15 and 425 d.f., meaning the null hypothesis of joint insignificance of the variables can be rejected. Overall this model is performed on 441 observations, the missing 10 observations arise from including the job type variable which effect is important for further policy targeting. This information was obtained from only those, who currently work for at least 20 hours per week.

Breusch-Pagan test is one of the tools to identify whether the heteroskedasticity is present in the linear regression model. Under the null hypothesis the error variances are all equal - homoskedastic, under the alternative as the value of the dependent variable increase, the variance also increases - the standard errors are said to be heteroskedastic and consequently the standard errors of the basic OLS estimators are not valid violating the MLR.5 assumption of the OLS regression. The result of this test shows the result of 55.659 on 15 *df*, thus heteroskedasticity is detected in the model and it is necessary to resolve this using heteroskedastic-robust standard errors, shown in Table 6.2.

Overall, on the respondents we can observe positive effect of some of the racing motivation, expectably with the highest magnitude with the motivation to win. Results suggest a significant effect of gender, men run nearly one hour more than women holding other factors constant. On the contrary, age has no effect as well hours spent in work or school per week. Some health variables are both statistically and economically significant, mainly the BMI coefficient and the subjective health rating of 3 and 4 joint together. Bit unexpected proves to be the coefficient of the *injury*, which is significant and highly positive. Education and demographic variables prove to be insignificant by now. Having each young children increases the time spent outside training, explained together with significant the interaction term.

Table 6.2: OLS regression with robust standard errors

	Dependent variable:
	Training hours
Constant	6.808*** (1.089)
Race motivation - Use of training km	1.694*** (0.515)
Race motivation - Win	2.560*** (0.973)
Motivation to begin - Insufficient activity	−0.757** (0.341)
Other activities - hours	0.052 (0.045)
Gender - Male	0.840*** (0.259)
Age	0.016 (0.013)
Education	0.003 (0.118)
Work hours	−0.002 (0.010)
Married	0.403 (0.253)
Sedentary job	−0.428 (0.306)
Worse health	−1.073** (0.461)
BMI	−0.147*** (0.041)
Injury	0.606*** (0.235)
Young children	−0.501** (0.217)
Young children:Adults	0.341** (0.157)
Observations	441
R^2	0.195
Adjusted R^2	0.167

Standard errors in parentheses

Note: *p<0.1; **p<0.05; ***p<0.01

6.2 Race participation model

As explained in the methodology, the second model in Table 6.3 uses the logistic regression to describe the relationship between the similar socio-demographic factors and likelihood of participation in races:

Table 6.3: Logistic regression

	Dependent variable:
	Races
Constant	5.090*** (1.729)
Motivation to begin - Insufficient activit	−3.031*** (0.781)
Motivation to begin - Active free time	−0.872* (0.450)
Group training	2.108*** (0.558)
Years running	−0.048 (0.031)
Gender - Male	1.231*** (0.438)
Age	0.034 (0.022)
Education	0.087 (0.222)
Income class - middle	−1.103** (0.531)
Income class - high	−1.249** (0.568)
Regular alcohol	1.100* (0.606)
Cigarettes	−0.093** (0.038)
Married	1.025** (0.490)
Sedentary job	0.008 (0.526)
Worse health	−1.792*** (0.660)
BMI	−0.172*** (0.064)
Injury	−0.122 (0.398)
Young children	−0.044 (0.279)
Observations	369
McFadden's R^2	0.404
Chi-square	67.934***

Standard errors in parentheses

Note: *p<0.1; **p<0.05; ***p<0.01

This model predicts the relationship between participation in races on socio-demographical factors on 369 respondents - those who were willing to share the income of their household. In comparison to the model predicting the training volume, only motivational factors to begin running were used, racing motivation would be multicollinear with the dependent variable. In addition, dummy variable *Group training*, indicating whether the individual trains in some group was added together with variables *Cigarettes* - number of cigarettes per day, and *Regular alcohol*, dummy variable indicating regular alcohol consumption in any amount. Performance of the model measured by McFadden's pseudo R^2 is 0.404, model is statistically significant with $\chi^2 = 67.934$ and $prob > \chi^2 = 0.000$

Due to non-linear nature of the logistic regression, coefficients obtained in Table 6.3 are not directly interpretable, only the partial effect will always have the same sign as the coefficient. The numbers used for the interpretation will be obtained using Average Partial Effects. This method is chosen since high number of categorical or dummy variables is present in the model. It is important to note that again due to non-linearity of the logistic regression partial effects differ for each individual, thus this method takes average across all observations.

Table 6.4: Average partial effect of the logistic regression

	APE
	Races
Motivation to begin - Insufficient activity	−0.42
Motivation to begin - Active free time	−0.08
Group training	0.14
Years running	−0.00
Gender - Male	0.10
Age	0.00
Education	0.01
Income class - high	−0.11
Income class - middle	−0.10
Regular alcohol	0.07
Cigarettes	−0.01
Married	0.08
Sedentary job	0.00
Worse health	−0.21
BMI	−0.01
Injury	−0.01
Young children	−0.00
Observations	369

Chapter 7

Discussion

The results uncover, apart from the direct interpretation of the regression coefficients, the interaction between factors driving training and contesting. Despite some unobservable factors, the core of this relationship lies in one's personal motivation. Aim for a personal achievement increases training volume by hours a week. Contrary to that, if motivation is just some active spending of free time or just some need for increase in physical activity, running is thereafter considered as just as a complementary activity and both training volume and race participation is decreased. The support in a form of some kind of group training is not reflected in training, however results on average in 14% higher participation in races. The same kind of relationship can be observed with respect to regular alcohol intake, which can also suggest the similar community element.

Men, holding other factors constant run 1 hour more than women and are 10% more likely to participate in the races, which corresponds to attendance in the most races. The effect of marriage, boosting race participation by 8%, but with no effect on training is likely correlated with some other motivational factors, e.g. spending weekends together, strengthened by the fact both are runners, which is not available in the sample. Furthermore, having young children (up to 10 years old) at home decreases time available for training, in this case it is necessary to evaluate this together with significant interaction term taking into account number of adults living in the

household together with the runner. Nevertheless it is not significant with respect to racing, might be an effect of most of the races organizing shorter complementary races for children together with the main events.

Since long distance running and endurance is supposed to develop to a certain degree with age, we do not observe observe any decrease with this variable. Neither does education or household income effect volume of training, only increased household income decreases participation in races, while bit counter intuitive, there is a big range of races with very small starting fees, therefore there are other unobserved factors behind this result.

Effect of size of the residence (therefore slightly distinguishing between city and countryside) is not present with relationship to running, supporting the hypothesis of running as a generally available activity with very low requirements for any equipment or sporting facility.

Highly important have shown to be the health variables which are significant in both models. BMI index, comparing weight of people with different height. while not having very high coefficient at the first sight, ranges from 17 to 32 and each point decreases the training time by 0.15 hours. Likewise, each cigarette decreases the likelihood of participation in the races by 1%. The health variables were summed up using subjective rating of individual's current health on a scale from 1 to 5, which might capture the health from the different point of view. Even wast majority reported better health compared to the time they had started running, rating of 3 or 4 significantly decreases both variables of interest. Most unanticipated result is regarding the highly positive coefficient of $Injury$ variable, suggesting the presence of some both running or non running-related bad stereotypes causing injuries in long time period.

Hypothesis of having sedentary job increasing leisure time physical activity is not supported at any significant level in neither model and could be a point of interest in the future, protecting for further deterioration in number of civilisation diseases in population.

Finally, there is a set of variables that are neither statistically nor eco-

nomically significant. Example of such a variable is number of hours spent doing other sport or exercise activities or having interest in some form of a diet. Those result might come from having two contrary groups in a sample - one more performance oriented, having other additional activities to even improve running form, while the second group of more hobby runners with a rather small training time cancels the effect of such variables out.

Chapter 8

Conclusion

This thesis studied the factors that drive the training intensity and participation in races among runners in the Czech Republic. Running as the leisure time activity has significantly evolved over the last two decades, also creating many new economic opportunities.

This study was motivated by the previous papers examining the demographic factors driving the amount and intensity of physical activity (e.g. McInnes and Shinogle, 2009 in the USA, Farrel and Shields; 2001 in England and Frömel et al., 2009 in the Czech Republic), but none of the studies was focused on this specific type of activity.

The volume of the training was estimated using OLS model which aimed to identify the factors that significantly influence the time spent running per week. The significant variables at 10% level included above all the motivational factors, both to begin training and to participate in races; than *Gender*, *Health* which describes the subjective health in the time of collection of the answers, *BMI* index, *Injury* indicating any kind of injury in the past year, and the presence of young children in the household, however with consideration of number of adults in the same household. Education or income level does not play a role in the training volume. In the contrary to the Frömel (2009), the size of the residence does not have any effect on running particularly.

In the second part of the thesis the same similar set of variables was used

to used to estimate the probability of participation in races. This analysis was performed using logistic regression. In addition to motivational factors as in the first model, group training appears to boost race participation. Analysis has shown the 10% higher probability of racing for males, which corresponds to the race attendance in major races. Marriage has a positive effect. In this model also *Cigarettes* and *Regular alcohol* variables are significant, however with an opposite sign. Participation decreases with moving to the higher income class. Size of the city nor education did not prove to be significant on the participation.

Results do not fully correspond to the results of the work performed on physical activity in general or on different sports. The theory of negative relationship between volume of running and income of the household as the relation to the opportunity costs is not supported, however this relationship can be observed in respect to race participation. In addition we do not observe any relationship between activity and size of the city the individual is living in, supporting the theory of general availability of running as a sport.

One point of interest was to evaluate whether sedentary job increases volume of training in a free time compared to manual workers. No significant change was found, therefore it could be a future policy intervention to promote activity at sedentary jobs, for example by providing additional job benefits.

Analysis has been performed on cross-sectional data collected at one time point. Moreover it did not take into account the intensity of the individual segments of the training, just its overall volume. Possible extension of this thesis could take use of panel data to see the changes of the dependent variables in time based on the socio-demographic factors and take into account the effect of the unobserved factors. Another extension could take use of the subjective ranking of the runners own health and observe its change in time, based on intensity of training.

Bibliography

BECKER, Gary S. "A Theory of the Allocation of Time". *The economic journal* (1965): 493–517. Print.

BOECKER, Henning, et al. "The runner's high: opioidergic mechanisms in the human brain". *Cerebral cortex* 18.11 (2008): 2523–2531. Print.

CHOU, Shin-Yi, Michael GROSSMAN, and Henry SAFFER. "An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System". *Journal of health economics* 23.3 (2004): 565–587. Print.

COBB, Steven, Douglas J OLBERDING, et al. "The importance of import substitution in marathon economic impact analysis". *International Journal of Sport Finance* 2.2 (2007): 108–118. Print.

DING, Ding, et al. "The economic burden of physical inactivity: a global analysis of major non-communicable diseases". *The Lancet* 388.10051 (2016): 1311–1324. Print.

FARRELL, Lisa and Michael A SHIELDS. "Investigating the economic and demographic determinants of sporting participation in England". *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165.2 (2002): 335–348. Print.

FRÖMEL, Karel, Josef MITÁŠ, and Jacqueline KERR. "The associations between active lifestyle, the size of a community and SES of the adult population in the Czech Republic". *Health & Place* 15.2 (2009): 447–454. Print.

HAMRIK, Zdenek, et al. "Physical activity and sedentary behaviour in Czech adults: results from the GPAQ study". *European journal of sport science* 14.2 (2014): 193–198. Print.

Health 2020 - National Strategy for Health Protection and Promotion and Disease Prevention. Ministry of Health of the Czech Republic, 2014. Print.

HUMPHREYS, Brad R and Jane E RUSESKI. "Participation in physical activity and government spending on parks and recreation". *Contemporary economic policy* 25.4 (2007): 538–552. Print.

LAVIE, Carl J, et al. “Effects of running on chronic diseases and cardiovascular and all-cause mortality”. *Mayo Clinic Proceedings*. Elsevier. 2015. 1541–1552. Print.

LEE, Duck-chul, et al. “Leisure-time running reduces all-cause and cardiovascular mortality risk”. *Journal of the American College of Cardiology* 64.5 (2014): 472–481. Print.

MCINNES, Melayne M and Judith A SHINOGLE. “Physical activity: Economic and policy factors”. *Economic aspects of obesity*. University of Chicago Press, 2011. 249–282. Print.

MELTZER, David O and Anupam B JENA. “The economics of intense exercise”. *Journal of Health Economics* 29.3 (2010): 347–352. Print.

MITÁŠ, Josef, et al. “Self-reported physical activity in perceived neighborhood in Czech adults-national study”. *Acta Gymnica* 43.2 (2013): 23–30. Print.

MITÁŠ, Josef, et al. “Physical activity, sedentary behavior, and body mass index in the Czech Republic: A nationally representative survey”. *Journal of Physical Activity and Health* 11.5 (2014): 903–907. Print.

NEJEDLA, Marie. “Health and economic consequences of obesity epidemic and possibilities of its prevention in schools”. *Ceska antropologie* 64 (2014): 20–24. Print.

NICHOLL, Jon P, Patricia COLEMAN, and John E BRAZIER. “Health and health care costs and benefits of exercise”. *Pharmacoeconomics* 5.2 (1994): 109–122. Print.

ORGANIZATION, World Health. *Global action plan on physical activity 2018-2030: more active people for a healthier world*. World Health Organization, 2019. Print.

PAPANIKOS, Gregory. “The economic effects of a marathon as a sport tourism event”. *Athens Journal of Sports* 1.225 (2015): 225–240. Print.

PEDISIC, Zeljko, et al. “Is running associated with a lower risk of all-cause, cardiovascular and cancer mortality, and is the more the better? A systematic review and meta-analysis”. *British journal of sports medicine* (2019). Print.

PENEDO, Frank J and Jason R DAHN. “Exercise and well-being: a review of mental and physical health benefits associated with physical activity”. *Current opinion in psychiatry* 18.2 (2005): 189–193. Print.

PRATT, Michael, et al. “Economic interventions to promote physical activity: application of the SLOTH model”. *American journal of preventive medicine* 27.3 (2004): 136–145. Print.

RAABOVÁ, Tereza. *Economic Impact of International Running Events of RunCzech.com 2011*. Economic impacT v.o.s., 2012. Print.

SARIS, WHM, et al. “How much physical activity is enough to prevent unhealthy weight gain? Outcome of the IASO 1st Stock Conference and consensus statement”. *Obesity reviews* 4.2 (2003): 101–114. Print.

SCHNOHR, Peter, et al. “Longevity in male and female joggers: the Copenhagen City Heart Study”. *American journal of epidemiology* 177.7 (2013): 683–689. Print.

WALTERS, Stephen J. *What is a Cox model?* Citeseer, 2009. Print.

WARBURTON, Darren ER, Crystal Whitney NICOL, and Shannon SD BREDIN. “Health benefits of physical activity: the evidence”. *Cmaj* 174.6 (2006): 801–809. Print.

WEN, Chi Pang, et al. “Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study”. *The lancet* 378.9798 (2011): 1244–1253. Print.

WOLF, Isabelle D and Teresa WOHLFART. “Walking, hiking and running in parks: A multidisciplinary assessment of health and well-being benefits”. *Landscape and Urban Planning* 130 (2014): 89–103. Print.

WOOLDRIDGE, Jeffrey M. *Introductory econometrics: a modern approach.* South-Western Cengage Learning, 2013. Print.

Appendix 1 - SDG list

Sustainable development goals (SDGs) by The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015:

SDG 1: No Poverty

SDG 2: Zero Hunger

SDG 3: Good Health and Well-being

SDG 4: Quality Education

SDG 5: Gender Equality

SDG 6: Clean Water and Sanitation

SDG 7: Affordable and Clean Energy

SDG 8: Decent Work and Economic Growth

SDG 9: Industry, Innovation and Infrastructure

SDG 10: Reduced Inequalities

SDG 11: Sustainable Cities and Communities

SDG 12: Responsible consumption and Production

SDG 13: Climate Action

SDG 14: Life below Water

SDG 15: Life on Land

SDG 16: Peace, Justice and Strong Institutions

SDG 17: Partnerships for the Goals

Appendix 2 - Descriptive statistics

Table 8.1: Summary of numerical variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Training hours	451	4.640	2.492	0	3	6	15
Numberof races	451	8.060	7.990	0	3	10	50
Years of running	451	5.931	6.302	0	3	7	60
Other activities - hours	451	3.264	2.901	0	1	4.5	15
Age	451	35.377	9.973	15	28	42	83
Work hours	451	40.022	13.266	0	38	45	75
Number of children	451	1.016	1.029	0	0	2	5
Young children	451	0.537	0.817	0	0	1	3
Dependent individuals	451	0.078	0.578	0	0	0	11
Adults	451	1.264	0.984	0	1	2	6
Cigarettes	451	1.033	3.672	0	0	0	30
Height	451	1.744	0.091	1.520	1.680	1.800	2.040
Weight	451	71.524	12.694	48	62	79	109
BMI	451	23.410	2.923	16.980	21.267	25.283	32.000

Table 8.2: Summary of categorical variables

variable	**value**	**n**	**%**
Motivation to begin	active free time	127	28.2
	cheap sport	21	4.7
	insufficient activity	15	3.3
	relax	247	54.8
	health issues	41	9.1
Group training	yes	174	38.6
	no	277	61.4
Races	yes	395	87.6
	no	56	12.4
Type of races	up to 10k	119	29.2

	10k to HM	198	48.6
	HM	42	10.3
	marathon	18	4.4
	ultra marathon	30	7.4
Motivation to race	self satisfaction	220	54.1
	active weekends	67	16.5
	win	18	4.4
	use of training km's	42	10.3
	comparison	60	14.7
Continuous training	yes	335	74.3
	no	116	25.7
Other activities	yes	367	81.4
	no	84	18.6
Gender	man	208	46.1
	woman	243	53.9
Education	primary	20	4.4
	high school	224	49.7
	bachelor	55	12.2
	master	133	29.5
	doctoral or higher	19	4.2
Family status	single/divorced	228	50.6
	married	223	49.4
Job type	phys. demanding	83	18.8
	sedentary/static	358	81.2
Income	up to 20,000 CZK	28	6.2
	20,001 to 30,000 CZK	63	14
	30,001 to 40,000 CZK	67	14.9
	40,001 to 50,000 CZK	68	15.1
	50,001 to 75,000 CZK	105	23.3
	75,001 to 100,000 CZK	28	6.2
	over 100,000 CZK	18	4

	don't know/don't share	74	16.4
Region	South Bohemian	23	5.1
	South Moravian	41	9.1
	Karlovy Vary	14	3.1
	Hradec Kralove	18	4
	Liberec	14	3.1
	Moravian-Silesian	25	5.5
	Olomouc	29	6.4
	Pardubice	32	7.1
	Pilsen	30	6.7
	Prague	101	22.4
	Central Bohemian	69	15.3
	Usti nad Labem	23	5.1
	Vysocina	16	3.5
	Zlin	16	3.5
Size of the city	up to 2,000 inhabitants	96	21.3
	2,000 to 5,000	52	11.5
	5,000 to 10,000	42	9.3
	10,000 to 100,000	118	26.2
	over 100,000 inhabitants	143	31,7
Health before	1	86	19.1
	2	111	24.6
	3	163	36.1
	4	71	15.7
	5	20	4.4
Health now	1	248	55
	2	179	39.7
	3	20	4.4
	4	4	0.9
	5	0	0
Injury	yes	196	43.5

	no	255	56.5
Hospital	yes	32	7.1
	no	419	92.9
Medication	yes	61	13.5
	no	390	86.5
Regular alcohol	yes	87	19.3
	no	364	80.7
Diet	yes	285	63.2
	no	166	36.8

Appendix 3 - Questionnaire

Dotazník pro běžce

Ahoj běžci, chtěl bych Vás všechny poprosit o vyplnění krátkého dotazníku pro moji bakalářskou práci na IES FSV UK na téma "Charakteristiky ovlivňující tréninkovou intenzitu a účast v běžeckých závodech". Vyplnění by Vám nemělo zabrat více než pár minut. Pokud nechcete soutěžit o ceny, Vaše odpověď je anonymní. O výsledky se poté velice rád podělím :)

*Povinné pole

1. Kolik hodin týdně trávíte běžeckým tréninkem? *
   How many hours per week do you spend running?

2. Motivace pro začátek běhu? *
   What was your motivation to start running?

   - Aktivní trávení volného času (Active free time)
   - Zdravotní důvody (Health issues)
   - Nedostačná aktivita při pouhé procházce (se psem, kočárkem, dítětem) (Insufficient activity just walking (with dog, chilc
   - Levné sportování (Cheap sport)
   - Relax, vyčištění hlavy (Relax, clearing your head)

3. Trénujete i ve skupině? * Do you also train with some group or community?

   - Ano (Yes)
   - Ne (No)

4. Účastníte se běžeckých závodů? * Do you participate in running races?

- Ano (Yes)
- Ne (No)

5. Pokud se účastníte závodů, kolik jich za rok absolvujete? If so, how many in how many races do you participate per year?

6. Jakých závodů se nejčastěji zúčastňujete? What is your favorite race distance?

- Do 10 km (Up to 10k)
- 10 km až půlmaraton (10k to HM)
- půlmaraton (Halfmarathon)
- maraton (Marathon)
- ultramaraton (Ultramarathon)

7. Jaká je vaše motivace k účasti v závodech? What is your motivation to participate in races?

- Výhra (To win)
- Zúročení treninkových kilometrů (Good use of training kilometres)
- Porovnání se s ostatními běžci (Comparison to other runners)
- Osobní uspokojení (Self-satisfaction)
- Aktivní trávení víkendu (Active weekends)

8. Před kolika lety jste začali běhat? * How many years ago have you started running?)

9. Běháte kontinuálně nebo jste opakovaně začínali? * Have you been running continuously or have you been satrting over again?

- Začal/a jsem a doteď běhám (Continuously)
- Opakovaně jsem začínal/a (Starting over)

10. Provozujete jiné sportovní aktivity kromě běhu? * Do you do any other sport activities?

- Ano (Yes)
- Ne (No)

11. Pokud ano, kolik hodin týdně jimi průměrně trávíte? (za poslední rok) If so, how many hours per week? (Over the past

12. Jaké je Vaše pohlaví? * Gender

- Žena (Woman)
- Muž (Man)

13. Jaký je Váš věk? * How old are you?

14. Jaké je Vaše nejvyšší dosažené vzdělání? * What is you highest level of education?

- Základní (Primary)
- Středoškolské (High school)
- Vysokoškolské - bakalář (Bachelor's Degree)
- Vysokoškolské - magisterské (Master's Degree)
- Vysokoškolské - PhD. a vyšší (PhD. or higher)

15. Jaký je Váš rodinný stav? * Are you married?

- Svobodná/ý / rozvedená/ý (Single/divorced)
- vdaná / ženatý (Married)

16. Kolik hodin týdně strávíte prací/studiem? * How many hours a week do you work/study?

17. Pokud je odpověď na předchozí otázku méně než 20, uveďte důvod. If the answer is lower than 20, share the reaso

- Mateřská dovolená,/Rodičovská dovolená/Péče o osobu blízkou (Maternity/family leave/caretaker)
- Nezaměstnaný - hledám práci (Unemployed - looking for a job)
- Invalidita (Disability)
- Starobní důchod (Retirement)
- Nepotřebuji více pracovat (I do not need to work anymore)

18. Jaká je povaha vašeho zaměstnání? * What is the nature of your job?

- Sedavé/statické (např. kancelář, studium, lékař apod.) (Sedentary/static)
- Fyzicky náročné zaměstnání (dělník, apod.) (Physically demanding)

19. Kolik dětí žije ve Vaší domácnosti? * How many children live in your household?

20. Kolik z počtu dětí uvedených výše je ve věku do 10 let? How many of the children mentioned above is under 10?

21. Kolik osob nad 10 let vyžadující zvýšenou péči (např. ze zdravotních důvodů) žije ve vaší domácnosti? * How many people over 10 years living in your require increase attention (e.g health issues)?

22. Kolik dospělých osob žije ve Vaší domácnosti (kromě Vás)? * How many adults (excluding you) live in your household?

23. Jaký je měsíční čistý příjem Vaší domácnosti? * What is the net income of your household?

- do 20 000 Kč
- 20 001 - 30 000 Kč
- 30 001 - 40 000 Kč
- 40 001 - 50 000 Kč
- 50 001 - 75 000 Kč
- 75 001 - 100 000 Kč
- nad 100 000 Kč
- (nechci uvádět/nevím) (I don't know/don't want to share)

24. V jakém kraji bydlíte? * In which region do you live?

- Praha
- Středočeský kraj
- Karlovarský kraj
- Plzeňský kraj
- Jihočeský kraj
- Liberecký kraj
- Ústecký kraj
- Královehradecký kraj
- Pardubický kraj
- Vysočina
- Zlínský kraj
- Moravskoslezský kraj
- Jihomoravský kraj
- Olomoucký kraj

25. Kolik má obec ve které bydlíte obyvatel? * How many inhabitants live in your town?

- do 2000 obyvatel
- 2000 - 5000 obyvatel
- 5000 - 10000 obyvatel
- 10000 - 100000 obyvatel
- nad 100000 obyvatel

26. Jak jste se zdravotně cítili než jste začali s běžeckým tréninkem? (stupnice jako ve škole)
How did you feel (from the medical perspective) before you started running? (1- best, 5 - worst)

- 1 - výborně
- 2
- 3
- 4
- 5 - velmi špatně

27. Jak se cítite zdravotně nyní? * How do you feel now (from the medical perspective)?

- 1 - výborně
- 2
- 3
- 4
- 5 - velmi špatně

28. Měli jste v poslední době nějaké zranění související i nesouvisející s během? *
Have you recently had any injury (even running non-related)?

- [ ] Ano (Yes)
- [ ] Ne (No)

29. Byli jste v poslední době hospitalizováni v nemocnici z jiných příčin? *
Have you recently been hospitalised from any other reason?

- [ ] Ano (Yes)
- [ ] Ne (No)

30. Užíváte dlouhodobě léky na nějakou chronickou nemoc? (cukrovka, vysoký tlak, štítná žlá astma atd.) * Do you take any medication (long-term) - e.g. diabetes, high blood pressure, asthma?

- [ ] Ano (Yes)
- [ ] Ne (No)

31. Kouříte? Pokud ano, kolik cigaret denně? * Do you smoke? If so, how many cigarettes a day?

- [ ] Ne (No)
- [ ] Jiné: _____ (How many?)

32. Pijete alkohol? Pokud ano, kolik jednotek týdně? * Do you drink alcohol? If so, how many units per week?

Specifikace jednotky alkoholu		
	1 jednotka orientačně	1 jednotka upřesnění
Pivo	1 sklenice	Pivo 12: 1 dávka = 250 ml (0,5 l = 2 dávky), Pivo 10: 1 dávka = 330 ml (0,5 l = 1,5 dávky)
Víno	1 sklenka	100 ml
Lihoviny	1 malá skleničkа	25 ml (malé štamprle), 50 ml = 2 dávky (velké štamprle)

33. Platí následující? Dávám si menší množství alkoholu každý den či obden? *
Is the following true? I drink a small amount of alcohol every day or every other day.

- Ano (Yes)
- Ne (No)

34. Jaká ve Vaše výška? * What is you height?

35. Jaká je Vaše váha? * What is your weight?

36. Zajímáte se o Vaši stravu (skládání jídelníčku, diety, počítání kalorií,...)? *
Do you take care about your diet (calories intake, diets etc.)?

- Ano (Yes)
- Ne (No)

37. Patříte do nějaké běžecké skupiny? * Are you member of any running group?

- [ ] MFR
- [ ] Adidas Runners
- [ ] Jiná nebo žádná. (Other or none)