

Unraveling the relationship between Sports Engagement and Life Satisfaction: A study for all Ages

Emma Bakker | s4514513 | e.n.bakker@student.rug.nl

*Rijksuniversiteit Groningen | Bachelorthesis | 05/06/2024
Accompanist: Francesca Giardini & Rita Smaniotto*

Abstract

In the Netherlands, many people spend their free time doing sports. It is a rather popular activity to do in one's leisure time. Since so many people are active in sports, it would be interesting to know if this might affect life in a positive way. Many researchers and sociologists have tried to study the concept of life satisfaction and figure out what factors in life are influential in predicting life satisfaction. In this study, the main goal was to figure out whether sports and life satisfaction are positively related to each other. In addition, the effect of age on this relationship was also examined. For this research, the LISS panel was used as our dataset. Since this study focused on Dutch individuals of all ages, a subset of the dataset that consists of Dutch participants only was used. The participants completed several questionnaires and from these, we created several variables to test the effect of sport on life satisfaction. Beforehand, it was expected that people who spend more time on sports would show a higher level of satisfaction than people who spend less time on sports. We also expected that this positive effect would be higher for younger people than for older people. However, these results turned out to be incorrect. The opposite was found. People who spend more time on sports would as a consequence show a small decrease in their life satisfaction than people who spend less time practicing sports. Also, no effect of age was found on the relationship between sport and life satisfaction. The effect of sport on life satisfaction was the same for all ages. The unexpected results invite us to conduct this study again, but this time taking into account the violation of certain assumptions and having a more accurate sample where all ages are equally presented.

Index

1. Introduction	3
2. Theory	5
3. Methods.....	10
4. Results	13
5. Conclusion and discussion	19
6. Literature.....	21
Appendix I.....	26
Appendix II	35
Appendix III.....	39
Appendix III: AI software.....	45

1. Introduction

Over the years, a lot of research has been done to determine what influences life satisfaction. Life satisfaction could be described as one's global assessment of their overall quality of life (Malvaso & Kang, 2022). Research shows that many aspects in life influence life satisfaction, like socioeconomic status, health, social relationships, lifestyle and leisure activity involvement (Procter et al, 2017). One of many things people might enjoy doing in their leisure time is sports. In the Netherlands over 50% of the population engages in sports activities weekly (CBS, 2022). If this many people spend their free time doing sports, it could be interesting to investigate whether engaging in sports positively influences life satisfaction. Participating in sports has long been recognized for its many benefits, like improved physical or mental health (Siefken et al, 2019). However, while these effects have been well-documented, the effect of sports on overall life satisfaction warrants further research.

In the past, studies regarding sports have mainly focused on the physical aspect of sports. A lot of the literature is written from a physiological perspective, where the relationship between physical activity and cardiovascular functioning is the center of attention (Waddington, 2000). However, in the last few years more research has been done on the sociological aspects of sports. There are numerous different ways to practice sports, from organized team activities to individual engagements. These activities can evoke a variety of social or psychological emotions. Participating in sports facilitates social connections, relationships and interactions with other people. These interactions can create camaraderie, team building, and feelings of belongingness. In addition to this, sports can shape identity formation and improve social cohesion (Sabin & Marcel, 2014).

The effect of sports on one's judgement of their overall life satisfaction requires further investigation. Since life satisfaction is a subjective concept, people assess their quality of life based on their own criteria and needs. Few studies have focused on this subjective side. Additionally, not a lot of research has differentiated between age (An et al, 2020). People of different ages acquire different resources in life, and these resources can fulfill specific material, social, or personal goals (Bartram, 2021). Therefore, looking at the relationship between sports and life satisfaction, while also taking the effect of age into account, could give helpful insights to the field of sociological research that focuses on improving overall life satisfaction.

When looking at this relationship it is important to take other factors into account. Sports engagement intersects with a lot of different social determinants, like gender, age, race, socioeconomic status and health. These determinants can shape an individual's experience and opportunity within sports. To gain insights into the relationship between sport and life satisfaction, it is important to reflect upon the impact of these factors. Exploring this relationship can improve the understanding of the holistic benefits of sports engagements and provide further implications for overall life satisfaction.

Studying the outcomes of sports from a sociological perspective can provide valuable insights for policymakers. Since over half of the Dutch population engages in sport activities weekly, it would be interesting to know whether more exercise leads to more life satisfaction. If sports engagement is positively related to life satisfaction, then it may be interesting for policymakers to actively promote participating in sports and make it more accessible for those who could benefit from it but do not have the adequate resources to engage in sports. By uncovering factors that might influence this relationship, interventions and policies aimed at improving life satisfaction can be adjusted accordingly.

From this, the following research question can be formulated: ***Does an increase in sports engagement have a positive effect on the life satisfaction of Dutch individuals, and is there a difference in terms of age?*** In this research we will focus on the Dutch population since the country has shown to be actively engaging in sports. The participants are drawn from a LISS panel. Participants had to answer multiple questionnaires, and from this, a statistical analysis will be conducted to research the relationship between sport and life satisfaction.

2. Theory

The main goal of this research is to look at the effect of sports on life satisfaction, and whether spending more hours on sports leads to an improvement in life satisfaction. Engaging in sports can have numerous beneficial outcomes for individuals and these outcomes will likely influence whether an individual is satisfied with their life or not (Wheatley & Bickerton, 2017). In addition to this, the factor of age will also be taken into account. People of different ages have different priorities, goals and needs. As a result, they might draw different levels of satisfaction from an increased engagement in sports. When researching the effect of sport on life satisfaction, it is important to define a few concepts first. In this research sport has a rather broad definition. It can be defined as any type of game or activity needing physical effort and skill that is done for enjoyment or as a job (Hammerschmidt & Jones, 2022). These physical activities can be done on an individual level or as a team. This research does not distinguish between the different types of sports. For example, it does not differentiate between sports like soccer, running, tennis, fitness, or horseback riding. In addition, no distinction is made between competitive and non-competitive sports. However, for this research it will be interesting to look at the amount of sports that someone does, and whether more sports engagement leads to more life satisfaction.

Life satisfaction is another broad and abstract concept that warrants further clarification. Literature shows that life satisfaction has been studied from two perspectives: objective and subjective. Objective life satisfaction focuses on external factors like income, access to healthcare and having opportunities to engage in leisure activities. Contrary, subjective life satisfaction mainly focuses on the internal evaluation of life events and the circumstances that influence these events (Huebner, 2004). In this study, life satisfaction is measured subjectively. Participants of the study define their subjective life satisfaction on a scale level. Subjective life satisfaction is therefore defined as a process in which an individual assesses the quality of their life by looking at their own set of criteria (Suldo et al, 2009).

Effect of sport on life satisfaction

There are several arguments as to why sports would have a positive effect on life satisfaction. Firstly, it is expected that sports can improve overall health. Distinctions can be made between physical and mental health. For physical health, sports engagement can improve fitness and mobility, increase muscle strength and reduce the risk of various conditions such as cardiovascular disease and obesity (Ministerie van Algemene Zaken, 2023). Research shows that there is a positive relationship between physical health and life satisfaction (An et al, 2020). When people are in good health, they will be more satisfied with the life they lead. Since physical health is a big contributor to overall life satisfaction, understanding how sports impact health is crucial.

Engaging in sports can also have positive outcomes for mental health. In turn, mental health can be a significant indicator of life satisfaction (Guzmán et al, 2020). One important outcome of sports is that it can stimulate the release of endorphins and stimulate neurotransmitters like serotonin and dopamine, which are associated with feelings of happiness. Sport engagement can therefore reduce symptoms of depression and anxiety (Ooms et al, 2020). Sports also provide a healthy outlet for stress relief. It can reduce levels of cortisol which are related to the bodily stress hormones. This reduction of cortisol can decrease feelings of tension and stress (Harvard Medical School, 2020). Engaging in sports can therefore serve as a coping mechanism during times of distress. It provides a productive way to manage negative emotions and regulate moods. Managing stress through sports can therefore contribute to better mental health and improve overall life satisfaction.

Sports can also serve as a contributor to socialization. Socialization is a process where individuals find acceptance among a group or society through the internalization of values, norms, behavior patterns and social roles (Weiss & Norden, 2021). When sports are played as a team, you have many interactions with other people. Members of a sports team might adopt group norms, values and behaviors. These interactions can create camaraderie, team-building and social relationships. Through this process of socialization, friendships can arise. Having high-quality friendships is an important determinant of life satisfaction (Rohrer et al, 2018). Friendships can fill basic psychological needs, such as relatedness, sharing feelings, and knowing that one can depend on others (Li & Kanawaza, 2016). Therefore, engaging in sports can serve as a way to create and enhance friendships, which in turn can influence life satisfaction. If people were to increase the amount of time spent playing sports, this would ultimately result in them seeing their friends and teammates more often. Spending more time with friends could increase the quality of their friendship, and consequently increase life satisfaction.

For many individuals, participating in sports can also provide a sense of purpose and identity. This phenomenon can be described as athletic identity: the degree of personal connection to sports (Edison et al, 2021). This connection can give people a sense of belonging and meaning, which are important aspects of life satisfaction. The greater emphasis one ascribes to their role in sports, the more likely one's motivation or outlook may be influenced by certain competencies, performances and achievements. According to this, sports can also influence the development of various skills and capabilities, such as leadership, communication, discipline, goal setting and emotional control (Duz & Aslan, 2020). By learning new skills through sports, a person might gain a sense of achievement. Whether it's mastering a new task, improving a performance, or winning a sports game, being successful in sports boosts confidence in many ways. These achievements will therefore have a positive effect on self-esteem which can positively influence life satisfaction (Romunde, 2023). By spending more hours on sports, these aspects can develop and improve even more, and as a result, life satisfaction will possibly also increase.

If there is a causal relationship between sports and life satisfaction, it would also be interesting to know if this causality runs in both directions. Research has shown that sport has a positive effect on happiness and that happiness has a positive impact on sport (Frey & Gullo, 2021). However, the effect sizes on this reverse causality are very small. An increase in happiness only shows a slight increase in the probability of sports participation, whereas an increase in sports participation significantly improves happiness. If an increase in sports improves happiness, will an increase in sports then also improve life satisfaction? If this is the case, then it would be incredibly beneficial for people to exercise more and perhaps join a new sports team. The arguments above have shown that there are many ways in which sports can influence life satisfaction. For this research, it will be interesting to look at the amount of time an individual spends practicing sports and if a difference in duration shows substantially different effects on life satisfaction. According to this, the first hypothesis can be formulated: *People who spend more time doing sports are more likely to experience a higher life satisfaction than people who spend less time doing sport.*

The role of age

While many arguments can be stated to support the hypothesis that sport has a positive effect on life satisfaction, it may also be interesting to look at the influence of age on this relationship. In life, people go through different developmental stages and transitions. Individuals' needs and preferences can change throughout these developmental stages, and as a result, individuals across different age groups may draw different levels of satisfaction from sports engagement. Therefore, it could be interesting and valuable to research age as a mediator in the relationship between sports participation and life satisfaction.

For younger individuals, like children, teenagers, and adolescents, there are many ways in which sports can influence life satisfaction. Engaging in sports at a young age can be very helpful in developing basic motor skills and coordination. These skills can consequently influence future health, by decreasing the risk of obesity and cardiovascular disease. Engaging in sport however does not limit itself to physical development. There also seem to be correlations between sports and academic achievements. Younger people who play sports make more use of skills such as planning, performance evaluation and self-reflection. These are skills that can also be useful in school (Logan et al, 2019). Lastly, playing sports offers a great opportunity to meet new people and make friends. Social relationships can play a big role in shaping one's self-esteem and overall satisfaction. Therefore, playing sports at a young age can uncover many positive effects that are important contributors to life satisfaction.

In adulthood, individuals come across a variety of life-changing transitions, whether it's becoming a parent, having an intensive job, or buying a house. According to research, life satisfaction shows a U-shaped curve, where life satisfaction starts to decline after adolescence and starts to increase again during older adulthood (Orben et al, 2022). This can be explained by growing social,

financial, or family pressures to do well in life. As a result, this can induce stress. Engaging in sports can serve as a coping mechanism to manage this stress or improve the overall work-life balance (Carraro et al, 2018). Subsequently, it is noted that physical activity in adulthood is associated with better cognitive function in later life (Hoang et al, 2016). Research found that high levels of cardiovascular fitness were associated with a decreased risk of dementia (Hörder et al, 2018). Engaging in sports during midlife can therefore have a lot of beneficial outcomes for a healthier lifestyle and consequently improve life satisfaction. However, for adults, one of the biggest indicators of life satisfaction is to have a strong social capital, for example having close relationships with peers (Orben et al, 2022). Therefore, the effect of sports on life satisfaction for adults might be positive, but the effect will most likely be smaller in comparison to younger people since other factors that influence life satisfaction seem to have bigger priority.

People who have reached retirement age have more time to engage in leisure activities like sports participation. Even though aging is an inevitable part of human life, it is noted that engaging in physical activity can increase functional ability in daily activities and slow down the process of cognitive decline (Andrieieva et al, 2019). These sports-related consequences can make life more enjoyable and valuable. Therefore, even for the elderly, sports could have a positive effect on life satisfaction. However, at this age, other factors are also significant for life satisfaction, like spending time with family. A few factors that highly influence life satisfaction for older people are social support and interaction (Papi & Cheraghi, 2021). Even though sports offer opportunities to increase social interaction, people from older ages might not feel the need to engage in sports. They may be more focused on engaging in activities that are more focused on rest and stability. Therefore, older people might draw a smaller level of satisfaction from engaging in sports than younger people do.

According to this, sport seems to affect life satisfaction across various age groups in various ways. The specific mechanisms through which sport influences life satisfaction may differ based on the specific needs and preferences that are set at different stages of life. For younger people, there seems to be much more to gain from an increased engagement in sports than for older people. At this young age, sports can influence and shape life in many aspects, and these aspects tend to be very influential in measuring life satisfaction. Understanding these differences is essential for designing sport-related policies and activities. From this, the second hypothesis can be formulated: *The positive effect of sports engagement on life satisfaction is stronger for younger people.*

Controlling for other influences

By looking closely at the relationship between sport and life satisfaction, it is important to take other influences into account. Other significant indicators might influence a person's life satisfaction. By controlling for these factors, it will be easier to look at the impact of sports on life satisfaction. Failure to control for these indicators may lead to confounding.

One important indicator of life satisfaction is health. Research found that life satisfaction is negatively related to obesity and bad lifestyle behaviors such as smoking, heavy drinking, and overeating (Strine et al, 2008). In addition, people who suffer from chronic diseases also show low levels of life satisfaction (Bri et al, 2022). Due to chronic illness, people are not able to participate in valued activities within the family, workplace, or community. These limitations in life due to illness can influence whether someone is satisfied with their life or not (Ambrosio et al, 2016). According to this, health can affect a person's life satisfaction.

Another factor that can influence life satisfaction is income. Studies show that individuals with high incomes experience a higher level of life satisfaction than those with low incomes (Cheung & Lucas, 2015). Having a higher income provides opportunities for individuals to satisfy their desires and can provide more ways for them to be satisfied with their lives. However, according to the *Easterlin paradox*, it is noted that once income covers basic needs in life, life satisfaction is mostly driven by non-monetary factors. The paradox states that even though at a certain point in time happiness is directly associated with income, over time this association does not grow as income continues to grow (Masuda et al, 2021). Nevertheless, it will be interesting to look at income as a controlling factor when looking at the relationship between sport and life satisfaction.

The third factor to control for is gender. Some researchers claim that women report lower life satisfaction as compared to men. However, this research remains inconclusive, since other researchers argue that there are no differences in the level of life satisfaction across gender (Chui & Wong, 2016). These disagreements do not rule out the fact that there is a possibility that gender might condition the effect of other variables. In the case of sports engagement, men might draw more life satisfaction from sports participation than women, or vice versa. For this research, it is interesting to take gender differences into account.

3. Methods

The LISS-panel

This research uses data from the Longitudinal Internet studies for the Social Sciences panel (LISS). This panel has existed since 2007 and focuses on scientific and social research. The collected data is public for non-commercial purposes such as scientific, social, or policy-relevant research (Centerdata, n.d.). The panel is drawn from a population register by Statistic Netherlands and consists of 5000 households, comprising a total of approximately 7500 individuals. This panel can be divided into two different datasets, one that consists of a Dutch panel and one that consists of an immigrant panel. Since this research focuses on Dutch individuals, we will look at the panel that only consists of Dutch people. This data is useful for research where a valid representation of the Dutch population is of interest. The data is retrieved through online questionnaires that are filled out on a monthly basis. Individuals from the panel receive a monetary incentive for each completed questionnaire. To maintain a high response and ensure that every individual has the opportunity to fill out the questionnaires, computers and internet access will be provided for those who are not able to provide this themselves (Centerdata, n.d.).

In this research, the dataset consists of several questionnaires merged together. According to this dataset, the panel consists of 6795 people all aged 16 years or older. For this research, data from the following filled-out questionnaires will be used:

- Personality (Wave 13)
- Social integration and leisure (Wave 13)
- Health (Wave 13)

The survey focused on personality has a sample of 6514 and a non-response rate of 17,6%. The data was collected in May and June 2021. The survey focused on social integration and leisure has a sample of 6680 and a non-response rate of 10,6%. The data was collected in October and November 2020. The survey focused on health has a sample of 6832 and a non-response rate of 16%. The data was collected in November and December 2020. Since the questionnaires were conducted in quite recent times, the data will still be relevant for this research. The sample is quite large, and the response rate is sufficient. As mentioned above, the LISS panel consists of Dutch people. For our research, it is important to look at the Dutch population and to know how they spend their leisure time. Therefore, the sample in the dataset seems qualified to represent our population.

Operationalization

After inspecting the dataset, questionnaires, and codebooks, the following variables have been selected to conduct the research:

Life satisfaction. This variable is measured with the following question: How satisfied are you with the life you lead at the moment? Respondents could answer this question on a scale from 0-10, where 0 equals 'not at all satisfied' and 10 equals 'completely satisfied'. If respondents were not sure how to answer the question, they could rate it -9 which equals 'I don't know'. To get a clear view of our respondents' level of life satisfaction, the option -9 ('I don't know') is recoded into 'missing'.

Health. This variable is measured with the following question: How would you describe your health, generally speaking? Respondents could answer this question on a scale from 1-5, where 1 equals 'poor', 2 equals 'moderate', 3 equals 'good', 4 equals 'very good', and 5 equals 'excellent'. Therefore, the variable is measured subjectively. This variable does not explicitly state whether it concerns physical health or mental health. To answer the question, respondents can give their own meaning to the term 'health'. This may cause problems with validity.

Sport. This variable is measured with the following question: How many hours do you spend on sports per week, on average? It can be seen as a continuous variable where respondents report the number of hours they spend on sports weekly. This variable will give a clear view of the amount of time participants engage in sports.

Gender. Respondents were asked what their gender was through a binary variable, where 1 equals 'male' and 2 equals 'female'. This variable was recoded to a dummy variable where 0 equals 'male' and 1 equals 'female'. According to the dataset there are slightly more women than men. This difference will most likely not have a significant impact on the reliability of the research.

Age. In this dataset, the participants on the LISS panel are all aged 16 years or older. The dataset shows a fairly even distribution of ages (Appendix I). Respondents aged 80 or older appear to be slightly underrepresented in this sample. Since there are no respondents younger than 16, this might cause some issues with the validity of the research.

Income. This is measured with an open-answer question where respondents could write down their net monthly income in Euros. For our analysis, this variable has been recoded to a different variable where a 1-point increase in income equals a €100 income increase, instead of a €1 increase. This adjustment has been made to get a clear view of the effect of income on life satisfaction. A €1 increase in income will most likely not affect life satisfaction, but a €100 income increase probably will.

Analytical strategy

To answer the research question and figure out whether the formulated hypotheses are supported or not, a statistical analysis must be performed. For our analysis, we need to focus on the respondents who have answered all of the questions from the variables that have been chosen above. As a result, the statistical analysis will be carried out on a subset of our sample. This subset consists of 2257 individuals. In this research *Life satisfaction* is the dependent variable and *Sport* is the independent variable. Since *Life satisfaction* is a continuous variable, a linear regression can be used. *Age* is also an independent variable that will be used as a moderator. In this research, we are interested in the effect of *Age* in the relationship between *Sport* and *Life satisfaction*. In addition, three control variables will be added to the research model to control for other factors that might influence life satisfaction. The control variables are: *Health*, *Income*, and *Gender*. These are also independent variables. Because multiple independent variables are used to predict or explain the dependent variable Y (*Life satisfaction*), a multiple regression analysis will be performed.

The regression analysis consists of the following 4 steps:

Model 1: In this model, the dependent y (*Life satisfaction*) is predicted by the control variables: *Health*, *Income* and *Gender*

Model 2: In this model, the independent x (*Sport*) is added. This is a centered variable. Y is explained by both x and the control variables.

Model: In this model, the moderation variable (*Age*) is added. This is a centered variable. Y is explained by x, the moderator, and the control variables

Model 4: An interaction variable (*Sport*Age*) is added to this model. Y is explained by x, moderator, interaction and control variables.

4. Results

4.1 Descriptive statistics

Table 1: Description of variables included in the analysis: mean (standard deviation), minimum and maximum value and total number of respondents

<i>Variabele</i>	<i>Mean (Standard deviation)</i>	<i>Minimum</i>	<i>Maximum</i>	<i>N Total</i>
Life satisfaction	7,37 (1,41)	0	10	2257
Age	51,00 (18,52)	16	93	2257
Sport	4,57 (4,36)	0	130	2257
Health (Scale 5 items)	3,30 (0,78)	1	5	2257
Gender (Male=0; Female=1)	46,9% Male 53,1% Female	0	1	2257
Income	19,65 (34,12)	0	1456,66	2257

As mentioned in the methods chapter, the statistical analyses have to be carried out on a subset of respondents who have a valid score on all variables. Therefore, table 1 displays the subset rather than the full sample. This explains why the N is equal for all variables. For each variable, the mean and associated standard deviation, and the minimum and maximum value have been reported. *Life satisfaction* has a mean of 7,37. This number might seem high, but according to literature, this number is relatively normal. For this subset of data, the oldest person is 93 years old. In the total sample, it showed that the oldest person is 103 years old. Since the youngest age reported is 16, there seems to be an absence of younger people like children and teenagers. This underrepresentation of the younger age group could cause problems with the validity of this research since it is not possible to clearly visualize what the effect is for the younger group. The mean of the total sample (50,41) is not much different from the one reported in Table 1 (51,00), according to Appendix I. Therefore, the ages reported in the subset seem to be rather equal to the ones in the total sample. As mentioned in the methods chapter, *Income* has been recoded to a new variable where a score of 1 represents an income of €100. Since the average score is 19,65 this equals an average income of approximately €1965. Looking at the maximum value it seems that there might be a few scores that are much higher. The highest income is €145.666, as reported in table 1. As a result, the standard deviation has been pulled upwards and is relatively high. In paragraph 4.2 we will discuss whether these high incomes are considered outliers. What is interesting about the variable *Sport*, is that the maximum value reported is 130. This would mean that a person would spend 130 hours per week on sports. Since this would average about 18 hours of sports engagement per day, it seems very unlikely that this is an accurate reported score. When taking a closer look at the data, it turns out that this particular individual is a 71-

year-old female. Considering her older age, we can assume that this score may be a typo. The second highest score reported on the variable *Sport* is 40 hours per week. This seems like a more appropriate number. The maximum value also affects the standard deviation of the variable *Sport*. The standard deviation is rather high in comparison to the mean. Paragraph 4.2 will take a closer look into possible outliers. If this score turns out to be an outlier, it is important to determine whether this individual might affect the statistical analysis of this research, and if it's necessary to exclude the participant from the dataset.

Table 2: Correlations between variables

	1. Life satisfaction	2. Sport	3. Age	4. Health	5. Income	6. Gender (0=Male; 1=Female)
1. Life satisfaction	-					
2. Sport	0,032	-				
3. Age	0,130**	0,049*	-			
4. Health	0,259**	0,063**	-0,277**	-		
5. Income	-0,005	0,025	0,112*	0,003	-	
6. Gender (0=Male; 1=Female)	0,002	-0,124	-0,56**	-0,059**	-0,099**	-

* significant at $p < 0,05$, ** significant at $p < 0,01$; N is 2257

In Table 2, the correlations between the variables are displayed. Correlation indicates the degree of association between two variables, or the extent to which two variables influence each other. When looking at the first column it is noted that *Life satisfaction* and *Health* are positively correlated. In the theory chapter, it was already discussed that having good health seems to be an important indicator of having high life satisfaction. Therefore, it makes sense that these variables are somewhat highly correlated with each other. A higher score on *Health* seems to influence *Life satisfaction* positively. *Sport* and *Life satisfaction* are also positively correlated. Even though this correlation coefficient is small, it can be assumed that an increase in sports engagement will positively influence life satisfaction. Another interesting correlation is that between *Health* and *Age*. The variables show a negative correlation of -0,227. A negative correlation means that one variable will increase while the other variable decreases. In this case, it makes sense that when an individual ages, their health will

subsequently decline. The correlation between income and life satisfaction is somewhat surprising. According to the literature it was expected that income would be an important indicator for life satisfaction. However, in this table, the correlation coefficient is rather small and also negative. If one's income increases, their life satisfaction decreases to a small extent.

4.2 Model evaluation

Table 3: Regression analysis

	Model 1		Model 2		Model 3		Model 4	
	b	SE	b	SE	b	SE	b	SE
Life satisfaction	5,800**	0,131	5,803**	0,131	5,415**	0,133	5,411**	0,133
Gender (0=man, 1=women)	0,047	0,058	0,053	0,058	0,088	0,057	0,084	0,057
Health (Scale 5 items)	0,468**	0,037	0,466**	0,037	0,581**	0,037	0,583**	0,037
Income	<0,0005	0,001	<0,0005	0,001	-0,001	0,001	-0,001	0,001
Sport			0,006	0,007	0,002	0,006	-0,003	0,007
Age					0,017**	0,002	0,017**	0,002
Sport*Age							<0,0005	<0,0005
R^2 adjusted	0,066		0,066		0,111		0,111	
F -change	54,321		0,813		114,839		1,886	
	**				**			
N	2257		2257		2257		2257	

* significant at $p < 0,05$; ** significant at $p < 0,01$

Table 3 shows the different regression coefficients that have been calculated while performing a multiple regression analysis. As mentioned in the methods chapter, the multiple regression analysis consists of four steps where four regression models have been computed. In the bottom rows, the R^2 adjusted and F -change are also mentioned. To evaluate the model fit of this regression we can look at the adjusted R^2 . The adjusted R^2 (also known as the coefficient of determination) is a number between 0 and 1 that indicates the extent to which a statistical model can predict a certain outcome. Adjusted R-squared is an adjusted version of R-squared that is adjusted for the number of predictors in the model. The adjusted R^2 increases when the new variable improves the model more than would be expected by chance. It decreases when a predictor improves the model less than expected. In Model 2

we see that the adjusted R^2 is equal to the adjusted R^2 in Model 1. The addition of sport has therefore not explained more variance. In model 3 the variable age is added. This caused the adjusted R^2 to increase. The addition of age therefore explains more variance. In model 4 the interaction is added. This addition does not show an increase in the adjusted R^2 and therefore the interaction does not explain more variance.

Another way to check if the model predicts the dependent variable in a significant way is to look at the F -change. The F -change is a test that is used to determine whether the variable added significantly improves the prediction of the regression model. It is related to the R^2 change and determines whether its change is significant. In Table 3, the F -change in model 1 is significant (54,321; $p < 0,01$). The variables that were added in this model are the control variables and can significantly improve the model prediction. In model 2, the variable *Sport* is added. However, the F -change is not significant here. In model 3 the moderator variable *Age* is added. Here it shows that the F -change is significant again (114,839; $p < 0,01$). The addition of the variable *Age* significantly improved the prediction. In model 4, also the complete model with all the variables and interaction added, the F -change is not significant, meaning that the interaction variable is not significantly improving the model prediction. If the variable were to be removed from the model, then it would not significantly reduce the predictive power of the model.

Subsequently, another important aspect of the model evaluation is the checking of assumptions, multicollinearity, and outliers. Since a linear regression is performed, it is important to check whether the assumptions are violated. In Appendix III, the following assumptions have been checked: 1) linearity, 2) independence, 3) homoscedasticity, and 4) normality. These inspections showed that only the assumption of independent observations may have been violated. This is because the panel used as data for this study is on a household basis. Therefore, the participants of this study are not fully independent from each other. Since some respondents are related to each other, there may be a coherence between the observed scores of these respondents. The standard errors are possibly smaller than they actually should be and as a result, some of the answers participants gave in the survey might show bias. The sample can still be used for estimating parameters. However, it will possibly not give optimal results and the generalizability of this research can therefore not be fully guaranteed. Multicollinearity does not seem to be an issue. A careful inspection of the VIF scores showed that none of the independent variables are highly correlated with each other (Appendix III), meaning that they do not mirror each other and predict the same thing.

Lastly, it is important to take outliers into account. Outliers are extreme values that differ from most other observation points in a data set. They can have a big impact on the statistical analysis and influence the results of the hypothesis testing. In paragraph 4.1 it was shown that one of the

participants in this research has an extremely high value on the variable *Sport* (130 hours of sport engagement per week). It became a priority to investigate whether this score affected the regression analysis. In SPSS, we computed a table (*casewise diagnostics* table) with cases that have residuals further than 3 standard deviations (Appendix III). The cases mentioned in this table are considered outliers, and the individual mentioned above was not mentioned in this table. The next step was to figure out whether the outliers in the casewise diagnostics table had a problematic effect on the regression analysis. This was checked by looking at the DFBETA. The DFBETA is a tool to measure whether the removal of an influential observation would noticeably change the effect of the coefficient. The outliers mentioned in the casewise diagnostics table do not show high DFBETA values. Therefore, the outliers are not that problematic and do not have to be removed from the dataset.

4.3 Hypothesis testing

To test the hypotheses, a stepwise regression analysis was performed. This regression can be seen in Table 3. The regression analysis consists of 4 steps. In Model 1, the dependent variable *Life satisfaction* is only predicted from the control variables. In model 2, the independent variable *Sport* was added. Since this is a regression analysis with a moderation effect, some of the variables had to be centered beforehand to control for multicollinearity in the interaction terms. The variables that have been centered are *Sport* and *Age*. In model 3, the moderator variable *Age* is added. Lastly, in model 4 the interaction term of *Sport* and *Age* has been added. The control variables in model 1 show interesting results. According to literature research, these variables may be good predictors of life satisfaction. Looking at this model it is shown that *Health* has a positive effect on *Life satisfaction* ($\beta=0,468, p<0,01$). From this we can indicate that an increase in health will increase life satisfaction significantly. The regression coefficient for *Income* is very small. Therefore, an increase in *Income* does not really have an effect on life satisfaction, in this regression. This is somewhat compliant with what was written in the theory section about the effect of income on life satisfaction. According to the *Easterlin paradox*, it is noted that once income covers basic needs in life, life satisfaction is mostly driven by non-monetary factors. Therefore, an increase in income will no longer increase life satisfaction. Gender is not significant as a predictor ($\beta=0,047$). The slope is positive, so women will on average report a higher level of life satisfaction than men.

In this study, we want to know whether sport has a positive effect on life satisfaction and if an increase in sports engagements will increase life satisfaction. According to our first hypothesis, we expect that individuals who spend more time doing sports will show higher levels of life satisfaction than individuals who spend less time doing sports. In Model 2, the control variables indicate approximately

the same regression coefficient. The added variable *Sport* has a slope of 0.006. This slope is not significant, but it is positive. If a person spends 1 hour doing sport, their life satisfaction will increase by 0,006 points (given that the other variables are equal). However, this effect seems too small to be considered as proof that doing more sports will increase life satisfaction. In Model 3 the moderator *Age* is added and as a result the regression coefficient of *Sport* decreases ($\beta=0,002$). The moderator *Age* has a positive regression coefficient and is also significant ($\beta=0,017$, $p<0,01$). Older respondents will show a higher degree of life satisfaction than younger people do. In Model 4 the interaction variable is added. Here, the effect of *Sport* on life satisfaction even turns negative ($\beta=-0,003$). Therefore, hypothesis 1 (*People who spend more time doing sports are more likely to experience a higher life satisfaction than people who spend less time doing sports*) can be rejected. According to this regression analysis, it seems that doing more sports does not have a positive effect on one's life satisfaction. This is not what was expected beforehand. Perhaps other variables have more impact on life satisfaction that have not been addressed in this research. As mentioned in paragraph 4.2, the addition of *Sport* in Model 2 did not affect the adjusted R^2 . The variable cannot improve the predictive power of the model. *Sport* is therefore not a good predictor of life satisfaction.

In this research, we also want to know whether people of different ages who participate in sports show differences in their level of life satisfaction. An interaction occurs when the effect on the dependent variable of an independent variable depends on the value of another independent variable. Looking at model 4, the regression coefficient of the interaction in the model is very small and not significant ($\beta < 0.0005$) This means that there is no visible interaction effect. The effect of *Sport* on *Life satisfaction* does not depend on the different values in *Age*. Therefore, we also reject hypothesis 2 (*The positive effect of sports engagement on life satisfaction is stronger for younger people*). This is also not what we expected. *Age* does not seem to have an effect on the relationships between *Sport* and *Life satisfaction* and younger people do not draw a higher level of satisfaction from practicing sports than older people do. The effects of sport on life satisfaction seem to be consistent for all ages. The variable *Age* does seem to have a significant effect on the dependent variable life satisfaction in both Model 3 and 4 ($\beta=0,017$, $p<0,01$). So even though *Age* seems to be a significant predictor for life satisfaction in this regression, it does not affect the relationship between *Sport* and *Life satisfaction*

5. Conclusion and discussion

The statistical analysis of this study showed that an increased engagement in sports did not result in increased life satisfaction for our respondents. People who spend more time on sports are not necessarily more satisfied with their lives than people who spend less time on sports. This is not what was expected beforehand. In the theory section, there were a lot of different arguments claiming that doing sports could be beneficial for improving life satisfaction. Nevertheless, in this study the result seems to be different and this raises the question of whether there are other explanations or underlying mechanisms that can clarify this unexpected result. According to the literature, practicing sports is mostly related to increased health. By engaging in sports, fitness, mobility and muscle strength can be increased and the risk of various conditions such as cardiovascular disease and obesity can be reduced (Ministerie van Algemene Zaken, 2023). Since health is an important indicator of how satisfied someone is with their life, this may have given a false impression that sport influences life satisfaction, while health actually does. This also seems to be reflected in the regression analysis, since health was included as a control variable and had a significantly large effect on life satisfaction, while sport did not.

As mentioned previously, life satisfaction is a very broad and abstract concept. Subjective life satisfaction is defined as a process in which an individual assesses the quality of their life by looking at their own set of criteria (Suldo et al, 2009). Several factors can influence someone's life satisfaction and perhaps sports are not considered an important factor in how someone assesses their life. If one were to increase the amount of time spent in sports, it could take away from spending time doing other valuable things. It is the balance and variety in activities that often contribute to higher life satisfaction. Prioritizing sports and exercise over other commitments might lead to tension and conflict in other domains of life if not enough attention is paid to this. This might explain why increasing the number of hours spend on sports did not influence life satisfaction positively.

Beforehand, it was also expected that the positive effect of sport on life satisfaction would be stronger for younger people than for older people. Age would have a moderating effect on the relationship between sports and life satisfaction. However, the statistical analysis showed that there was not a visible interaction effect present, and therefore age did not moderate the relationship between sports and life satisfaction. This also raised the question of why this surprising result was found, since it did not match up to our theory. Perhaps there were big differences within age groups themselves. For example, due to personal circumstances (such as chronic illness or disabilities), some young people may benefit less from practicing sports. At the same time, some older people, despite their age, might be very active and fit and benefit greatly from sports. This variety within age groups may prevent age from showing a clear interaction effect in the relationship between sport and life staisaction. The lack of interaction effect may also be due to a lack of young people in the dataset. The sample showed ages between 16 and 93 years old. Since there were no respondents present under 16

years of age, the young age group was not accurately portrayed. In the literature, the effects of sport on life satisfaction for younger people were mostly written from the perspective of children and teenagers. This group of people was not correctly represented in this study, since the youngest age was 16. Another explanation for the lack of interaction effect may be that sport was too broadly defined. No distinction was made between the different kinds of sports, even though people of different ages may play different sports. At a younger age, people often still participate in team sports, while at an older age, people more often participate in fitness independently. This is partly because middle-aged people are relatively busier with other responsibilities and therefore want to plan specific times to exercise. This is easier to manage if they exercise independently. If we were to only focus on one type of sport in this research, then the result might have been different.

Looking at the performance of the research it seems that a few implications have to be mentioned. First of all, the sample that was used for this research seemed to show a disturbance. The sample was drawn from a panel that was based on a household level. Therefore, the assumption of independent cases was violated. Participants of this study were not fully independent from each other, since some respondents are related to each other. As a result, there may have been a coherence between the observed scores of the respondents, which could cause the standard errors to possibly be smaller than they actually should be. This bias could have influenced the overall results of this research. One way to solve this in the future is to do the statistical analysis again, but then select people who are the head of the household. This gives you a sample where the participants are not dependent on each other. However, this would also cause the sample to become a lot smaller. Another issue previously mentioned with the sample was the lack of individuals under 16 years of age. If this study were to be done again, it would be very important to obtain a representative sample of the population, and this would therefore include children and teenagers below the age of 16.

Although this research did not give the results that were expected in advance, it did provide certain insights. For example, it showed that being in good health seems to be a big indicator of high life satisfaction. It also showed that practicing sports at a higher intensity or duration will not increase life satisfaction, instead, it will slowly decrease life satisfaction. Practicing sports itself will not necessarily indicate whether someone is satisfied with their life or not. However, it would not seem wise to convince people to spend less time doing sports. Spending time on sports as a leisure activity will still provide many positive outcomes. To get a clearer answer to the question of whether sport can positively influence life satisfaction, we recommend conducting the study again. This time with a sample that is better tailored to the target group, which does not consist of people who are dependent on each other.

6. Literature

- Ambrosio, L., Portillo, M. C., Rodríguez-Blázquez, C., Rodríguez-Violante, M., Castrillo, J. C. M., Arillo, V. C., ... & Martínez-Martín, P. (2016). Living with chronic illness scale: international validation of a new self-report measure in Parkinson's disease. *npj Parkinson's Disease*, 2(1), 1-6. <https://doi.org/10.1038/npjparkd.2016.22>
- An, H. Y., Chen, W., Wang, C. W., Yang, H. F., Huang, W. T., & Fan, S. Y. (2020). The relationships between physical activity and life satisfaction and happiness among young, middle-aged, and older adults. *International journal of environmental research and public health*, 17(13), 4817. <https://doi.org/10.3390/ijerph17134817>
- Andrieieva, O., Hakman, A., Kashuba, V., Vasylenko, M., Patsaliuk, K., Koshura, A., & Istyniuk, I. (2019). Effects of physical activity on aging processes in elderly persons. DOI:10.7752/jpes.2019.s4190
- Bartram, D. (2021). Age and life satisfaction: Getting control variables under control. *Sociology*, 55(2), 421-437. <https://doi.org/10.1177/0038038520926871>
- Bi, K., Chen, S., Yip, P. S., & Sun, P. (2022). Domains of life satisfaction and perceived health and incidence of chronic illnesses and hospitalization: evidence from a large population-based Chinese cohort. *BMC Public Health*, 22(1), 1703. <https://doi.org/10.1186/s12889-022-14119-3>
- Bondin, V. I., Lyubetsky, N. P., Schalohyan, S. I., Gerasimov, M. V., & Samygin, S. I. (2020). Sport in the context of globalization. In *Artificial intelligence: Anthropogenic nature vs. social origin* (pp. 237-247). Springer International Publishing. https://doi.org/10.1007/978-3-030-39319-9_28
- Carraro, E., Schilirò, T., Biorci, F., Romanazzi, V., Degan, R., Buonocore, D., ... & Gilli, G. (2018). Physical activity, lifestyle factors and oxidative stress in middle age healthy subjects. *International journal of environmental research and public health*, 15(6), 1152. <https://doi.org/10.3390/ijerph15061152>

- CBS. (2022). *Sportdeelname wekelijks: Het aandeel van de Nederlandse bevolking van 4 jaar en ouder dat één keer per week of vaker sport*. RIVM.
<https://www.sportenbewegenincijfers.nl/kernindicatoren/sportdeelname-wekelijks>
- Centerdata. (n.d.). About the panel. lissdata.nl. <https://www.lissdata.nl/how-it-works>
- Cheung, F., & Lucas, R. E. (2015). When does money matter most? Examining the association between income and life satisfaction over the life course. *Psychology and aging, 30*(1), 120. <https://doi.org/10.1037/a0038682>
- Chui, W. H., & Wong, M. Y. (2016). Gender differences in happiness and life satisfaction among adolescents in Hong Kong: Relationships and self-concept. *Social Indicators Research, 125*, 1035-1051. <https://doi.org/10.1007/s11205-015-0867-z>
- Duz, S., & Aslan, T. V. (2020). The Effect of Sport on Life Skills in High School Students. *Asian Journal of Education and Training, 6*(2), 161-168.
DOI:10.20448/journal.522.2020.62.161.168
- Edison, B.R.; Christino, M.A.; Rizzone, K.H. Athletic Identity in Youth Athletes: A Systematic Review of the Literature. *Int. J. Environ. Res. Public Health* 2021, 18, 7331.
<https://doi.org/10.3390/ijerph18147331>
- Frey, B. S., & Gullo, A. (2021). Does sports make people happier, or do happy people more sports?. *Journal of Sports Economics, 22*(4), 432-458. <https://doi.org/10.1177/1527002520985667>
- Guzmán, J., Green, J. G., Oblath, R., & Holt, M. K. (2020). Life satisfaction mediates the association between mental health risk and perceptions of school functioning among children and adolescents. *Contemporary school psychology, 24*(4), 389-399.
<https://doi.org/10.1007/s40688-019-00257-w>
- Hammerschmidt, J., Kraus, S., & Jones, P. (2022). Sport entrepreneurship: Definition and conceptualization. *Journal of Small Business Strategy, 32*(2), 1.
<https://doi.org/10.53703/001c.31718>
- Harvard Medical School. (2020, 7 juli). *Exercising to relax: How does exercise reduce stress? Surprising answers to this question and more*. Harvard Health Publishing.
<https://www.health.harvard.edu/staying-healthy/exercising-to-relax>

- Hoang, T. D., Reis, J., Zhu, N., Jacobs, D. R., Launer, L. J., Whitmer, R. A., ... & Yaffe, K. (2016). Effect of early adult patterns of physical activity and television viewing on midlife cognitive function. *JAMA psychiatry*, 73(1), 73-79.
doi:10.1001/jamapsychiatry.2015.2468
- Hörder, H., Johansson, L., Guo, X., Grimby, G., Kern, S., Östling, S., & Skoog, I. (2018). Midlife cardiovascular fitness and dementia: a 44-year longitudinal population study in women. *Neurology*, 90(15), e1298-e1305.
<https://doi.org/10.1212/WNL.0000000000005290>
- Huebner, E.S. Research on Assessment of Life Satisfaction of Children and Adolescents. *Social Indicators Research* 66, 3–33 (2004).
<https://doi.org/10.1023/B:SOCI.0000007497.57754.e3>
- Li, N. P., & Kanazawa, S. (2016). Country roads, take me home... to my friends: How intelligence, population density, and friendship affect modern happiness. *British Journal of Psychology*, 107(4), 675-697. <https://doi.org/10.1111/bjop.12181>
- Logan, K., Cuff, S., LaBella, C. R., Brooks, M. A., Canty, G., Diamond, A. B., ... & Stricker, P. R. (2019). Organized sports for children, preadolescents, and adolescents. *Pediatrics*, 143(6) <https://doi.org/10.1542/peds.2019-0997>
- Malvaso, A., & Kang, W. (2022). The relationship between areas of life satisfaction, personality, and overall life satisfaction: An integrated account. *Frontiers in psychology*, 13, 894610.
<https://doi.org/10.3389/fpsyg.2022.894610>
- Masuda, Y. J., Williams, J. R., & Tallis, H. (2021). Does life satisfaction vary with time and income? Investigating the relationship among free time, income, and life satisfaction. *Journal of happiness studies*, 22, 2051-2073. <https://doi.org/10.1007/s10902-020-00307-8>
- Ooms, L., Dool, R. van den, & Elling, A. (2020). Psychische gezondheid, bewegen en sport: de rol van bewegen en sport voor mensen met (ernstige) psychische klachten, factsheet 2020/9. Utrecht: Mulier Instituut.
<https://studenttheses.uu.nl/handle/20.500.12932/38481>
- Orben A, Lucas RE, Fuhrmann D, Kievit RA. (2022). Trajectories of adolescent life satisfaction. *R. Soc. Open Sci.* 9: 211808. <https://doi.org/10.1098/rsos.211808>

- Papi, S., & Cheraghi, M. (2021). Multiple factors associated with life satisfaction in older adults. *Menopause Review/Przegląd Menopauzalny*, 20(2), 65-71. DOI: <https://doi.org/10.5114/pm.2021.107025>
- Proctor, C., Linley, P. A., Maltby, J., & Port, G. (2017). Life satisfaction. *Encyclopedia of adolescence*, 2(1), s2165-s2176. DOI 10.1007/978-3-319-32132-5_125-2
- Rohrer, J. M., Richter, D., Brümmer, M., Wagner, G. G., & Schmukle, S. C. (2018). Successfully striving for happiness: Socially engaged pursuits predict increases in life satisfaction. *Psychological science*, 29(8), 1291-1298. <https://doi.org/10.1177/0956797618761660>
- Rosenzweig, L. R., & Zhou, Y. Y. (2021). Team and nation: Sports, nationalism, and attitudes toward refugees. *Comparative Political Studies*, 54(12), 2123-2154. <https://doi.org/10.1177/0010414021997498>
- Sabin, S. I., & Marcel, P. (2014). Study regarding the impact of sport competitions on students socialization. *European scientific journal*, 10(26). POSDRU/159/1.5/S/141086.
- Shah, E. J., Fransen, K., Slater, M. J., & Barker, J. B. (2023). The impact of intra-team communication and support relationships on team identification and collective efficacy in elite team sport: a social network analysis. *International Journal of Sport and Exercise Psychology*, 21(4), 638-659. <https://doi.org/10.1080/1612197X.2022.2084761>
- Siefken, K., Junge, A., & Laemmle, L. (2019). How does sport affect mental health? An investigation into the relationship of leisure-time physical activity with depression and anxiety. *Human Movement*, 20(1), 62-74. DOI: <https://doi.org/10.5114/hm.2019.78539>
- Strine, T. W., Chapman, D. P., Balluz, L. S., Moriarty, D. G., & Mokdad, A. H. (2008). The associations between life satisfaction and health-related quality of life, chronic illness, and health behaviors among US community-dwelling adults. *Journal of community health*, 33, 40-50. <https://doi.org/10.1007/s10900-007-9066-4>

- Suldo, S. M., Huebner, E. S., Friedrich, A. A., & Gilman, R. (2009). Life satisfaction. In *Handbook of positive psychology in schools* (pp. 27-36). Routledge.
<http://ndl.ethernet.edu.et/bitstream/123456789/47718/1/128.pdf#page=46>
- Waddington, I. (2000). Sport and health: A sociological perspective. *Handbook of sports studies*, 408-421. <http://ndl.ethernet.edu.et/bitstream/123456789/76391/1/84.pdf#page=446>
- Weiss, O., & Norden, G. (2021). Socialization and Sport. In *Introduction to the Sociology of Sport* (pp. 52-98). Brill. https://doi.org/10.1163/9789004464711_006
- Wheatley, D., & Bickerton, C. (2017). Subjective well-being and engagement in arts, culture and sport. *Journal of cultural economics*, 41, 23-45.
<https://doi.org/10.1007/s10824-016-9270-0>
- World Health Organization, *Mental Health: A State of Well-Being* (World Health Organization, 2014).
<https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>

Appendix I

Variable overview

This appendix shows the descriptive statistics of the variables that are used in this research. The data is drawn from a sample. Some of the variables had to be recoded or adjusted in our analysis. The sub-headings are organized by variable and show their descriptive statistics and possible recoding's and adjustments.

1.1 Life satisfaction: original variable

```
FREQUENCIES VARIABLES=cp21m011
/STATISTICS=MEAN
/BARCHART FREQ
/ORDER=ANALYSIS.
```

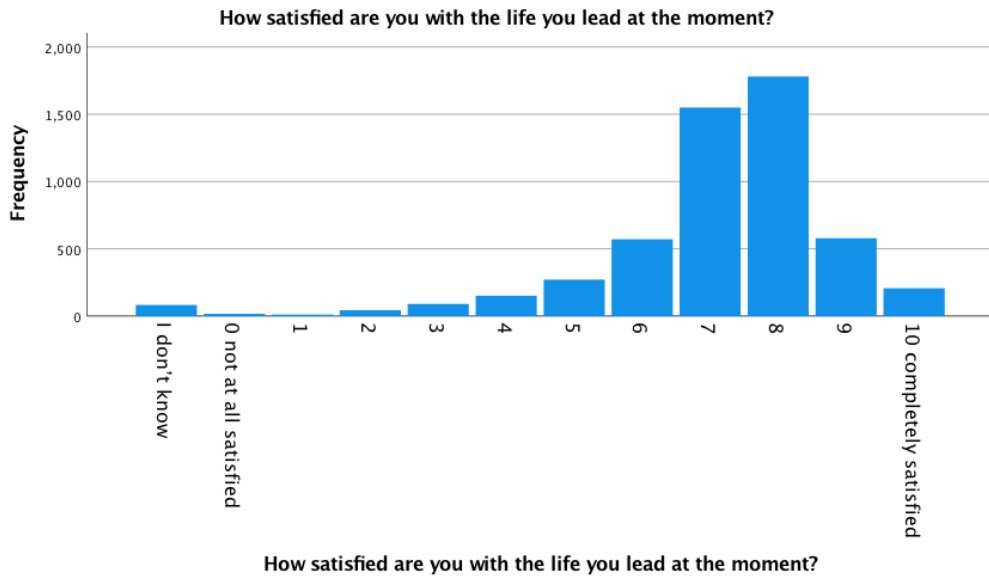
Statistics

How satisfied are you with the

N	Valid	5362
	Missing	1433
Mean		6.98

How satisfied are you with the life you lead at the moment?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I don't know	83	1.2	1.5	1.5
	0 not at all satisfied	18	.3	.3	1.9
	1	12	.2	.2	2.1
	2	44	.6	.8	2.9
	3	91	1.3	1.7	4.6
	4	152	2.2	2.8	7.5
	5	272	4.0	5.1	12.5
	6	572	8.4	10.7	23.2
	7	1551	22.8	28.9	52.1
	8	1781	26.2	33.2	85.3
	9	579	8.5	10.8	96.1
10 completely satisfied	207	3.0	3.9	100.0	
	Total	5362	78.9	100.0	
Missing	System	1433	21.1		
	Total	6795	100.0		



1.2 Life satisfaction: recoding's and adjustments

A small adjustment has been made with the variable *life satisfaction*. The option -9 (“I don’t know”) was recoded into ‘missing’. This recoding gives a clear overview on what scores respondents have answered on the question: How satisfied are you with the life you lead at the moment?

```
RECODE cp21m011 (-9=SYSMIS) (0 thru 10=Copy) INTO Life_satisfaction.
EXECUTE.
```

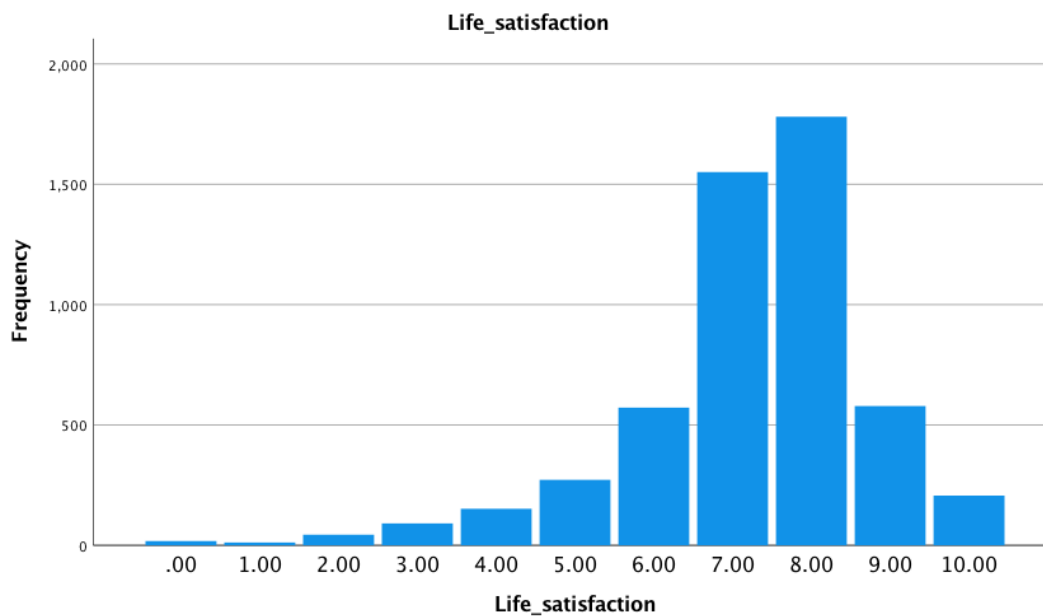
1.3 Life satisfaction: descriptive statistics of final variable

```
FREQUENCIES VARIABLES=Life_satisfaction
/STATISTICS=MINIMUM MAXIMUM MEAN
/BARCHART FREQ
/ORDER=ANALYSIS.
```

Statistics		
Life_satisfaction		
N	Valid	5279
	Missing	1516
Mean		7.2285
Minimum		.00
Maximum		10.00

Life_satisfaction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	18	.3	.3	.3
	1.00	12	.2	.2	.6
	2.00	44	.6	.8	1.4
	3.00	91	1.3	1.7	3.1
	4.00	152	2.2	2.9	6.0
	5.00	272	4.0	5.2	11.2
	6.00	572	8.4	10.8	22.0
	7.00	1551	22.8	29.4	51.4
	8.00	1781	26.2	33.7	85.1
	9.00	579	8.5	11.0	96.1
	10.00	207	3.0	3.9	100.0
		Total	5279	77.7	100.0
Missing	System	1516	22.3		
Total		6795	100.0		



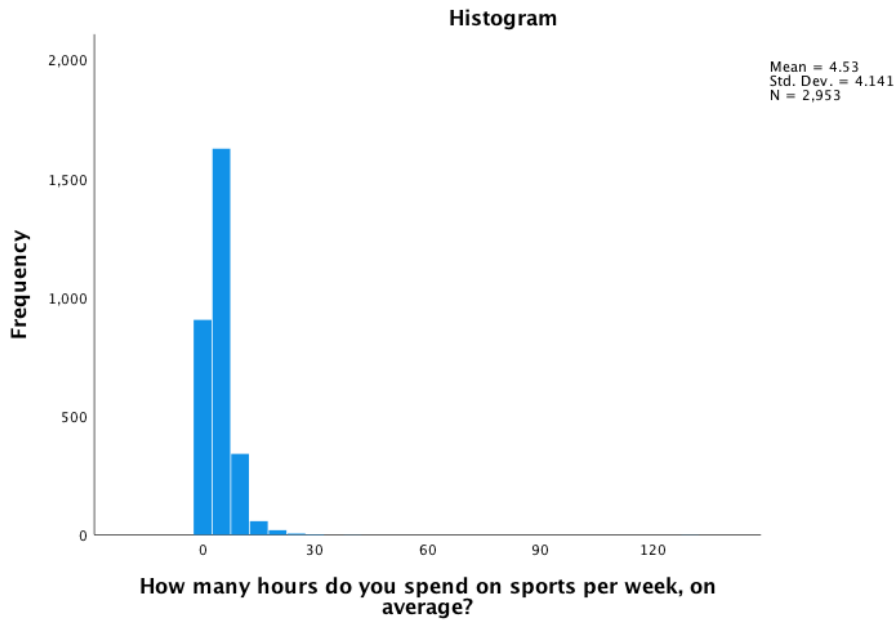
2.1 Sports: original variable

```
FREQUENCIES VARIABLES=cs20m105
/STATISTICS=MINIMUM MAXIMUM MEAN
/HISTOGRAM
/ORDER=ANALYSIS.
```

Statistics

How many hours do you spend

N	Valid	2953
	Missing	3842
Mean		4.53
Minimum		0
Maximum		130



How many hours do you spend on sports per week, on average?

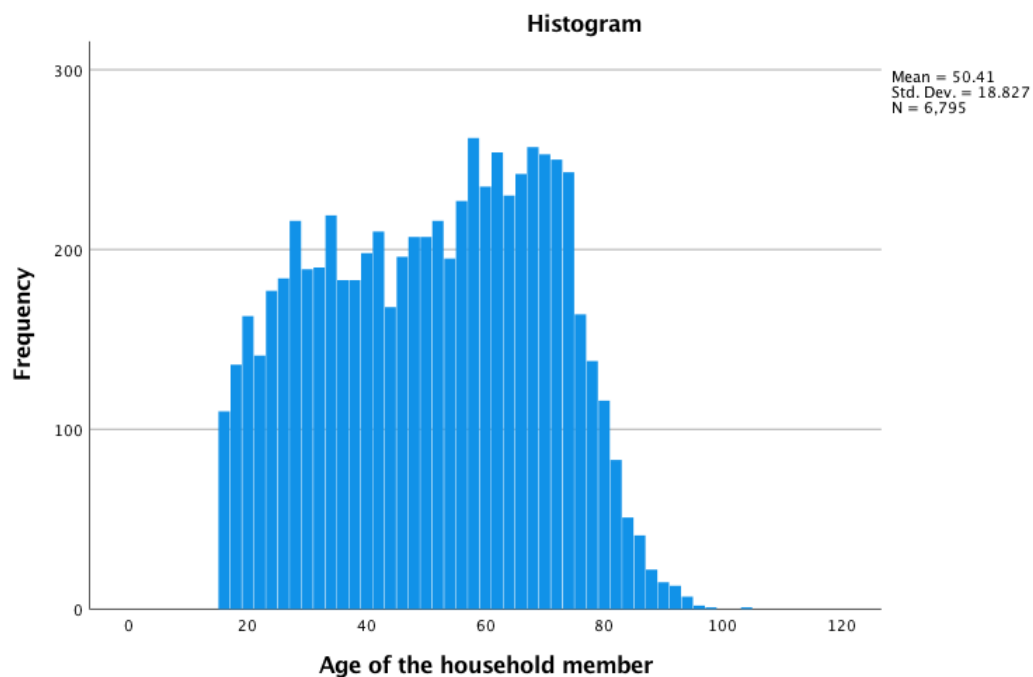
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	13	.2	.4	.4
	1	307	4.5	10.4	10.8
	2	584	8.6	19.8	30.6
	3	549	8.1	18.6	49.2
	4	425	6.3	14.4	63.6
	5	293	4.3	9.9	73.5
	6	232	3.4	7.9	81.4
	7	125	1.8	4.2	85.6
	8	127	1.9	4.3	89.9
	9	26	.4	.9	90.8
	10	125	1.8	4.2	95.0
	11	11	.2	.4	95.4
	12	51	.8	1.7	97.1
	13	1	.0	.0	97.2
	14	17	.3	.6	97.7
	15	27	.4	.9	98.6
	16	11	.2	.4	99.0
	17	1	.0	.0	99.1
	18	1	.0	.0	99.1
	19	1	.0	.0	99.1
20	16	.2	.5	99.7	
21	1	.0	.0	99.7	
24	2	.0	.1	99.8	
25	3	.0	.1	99.9	
30	2	.0	.1	99.9	
40	1	.0	.0	100.0	
130	1	.0	.0	100.0	
	Total	2953	43.5	100.0	
Missing	System	3842	56.5		
Total		6795	100.0		

2.2 Sports: recoding's and adjustments

No adjustments have been made to this variable.

3.1 Age: Original variable

```
FREQUENCIES VARIABLES=leeftijd  
/FORMAT=NOTABLE  
/STATISTICS=MINIMUM MAXIMUM MEAN  
/HISTOGRAM  
/ORDER=ANALYSIS.
```



Statistics

Age of the household member

N	Valid	6795
	Missing	0
Mean		50.41
Minimum		16
Maximum		103

3.2 Age: recoding's and adjustments

No adjustments have been made to this variable.

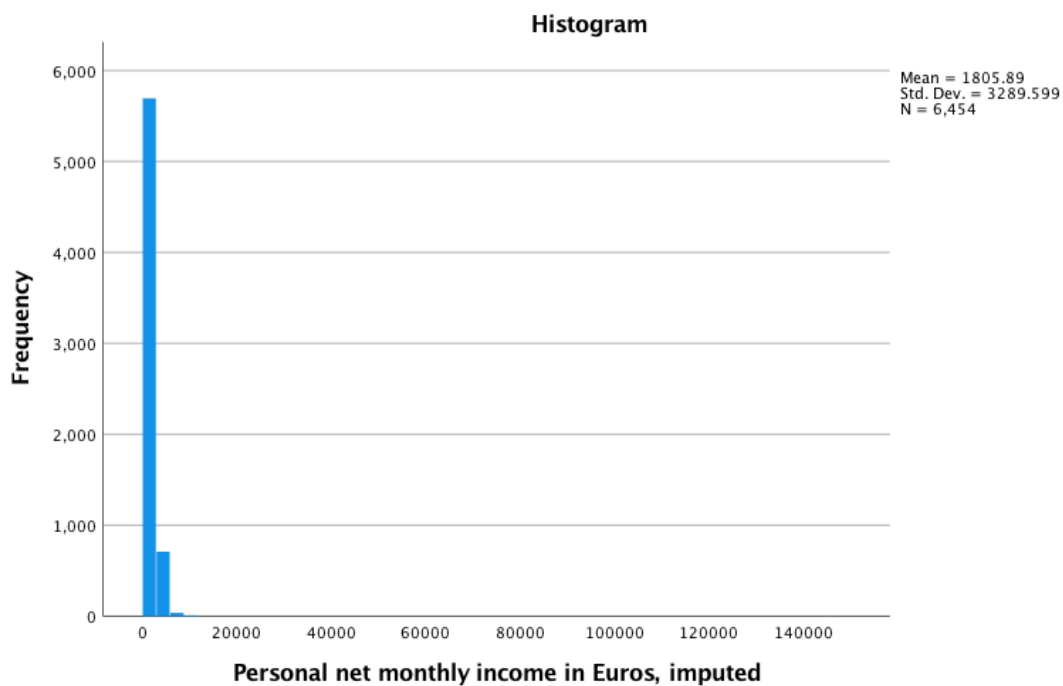
4.1 Income: Original variable

```
FREQUENCIES VARIABLES=nettoink_f  
/STATISTICS=MINIMUM MAXIMUM MEAN  
/HISTOGRAM  
/ORDER=ANALYSIS.
```

Statistics

Personal net monthly income in

N	Valid	6454
	Missing	341
Mean		1805.89
Minimum		0
Maximum		146652



4.2 Income: recoding's and adjustments

There has been a small recoding made to the variable *Income*. This adjustment has been made to show what the effect is for *Income* in the regression analysis. Now, a 1-point increase in the variable income equals a €100 income increase, instead of a €1 increase.

```
COMPUTE Inkomen_Honderden=nettoink_f / 100.  
EXECUTE
```


4.3 Income: descriptive statistics of final variable

```
FREQUENCIES VARIABLES=Inkomen_Honderden
/STATISTICS=MINIMUM MAXIMUM MEAN
/HISTOGRAM
/ORDER=ANALYSIS.
```

Statistics

Inkomen_Honderden

N	Valid	6454
	Missing	341
Mean		18.0589
Minimum		.00
Maximum		1466.52

5.1 Health: Original variable

```
FREQUENCIES VARIABLES=ch20m004
/STATISTICS=MINIMUM MAXIMUM MEAN
/BARCHART FREQ
/ORDER=ANALYSIS.
```

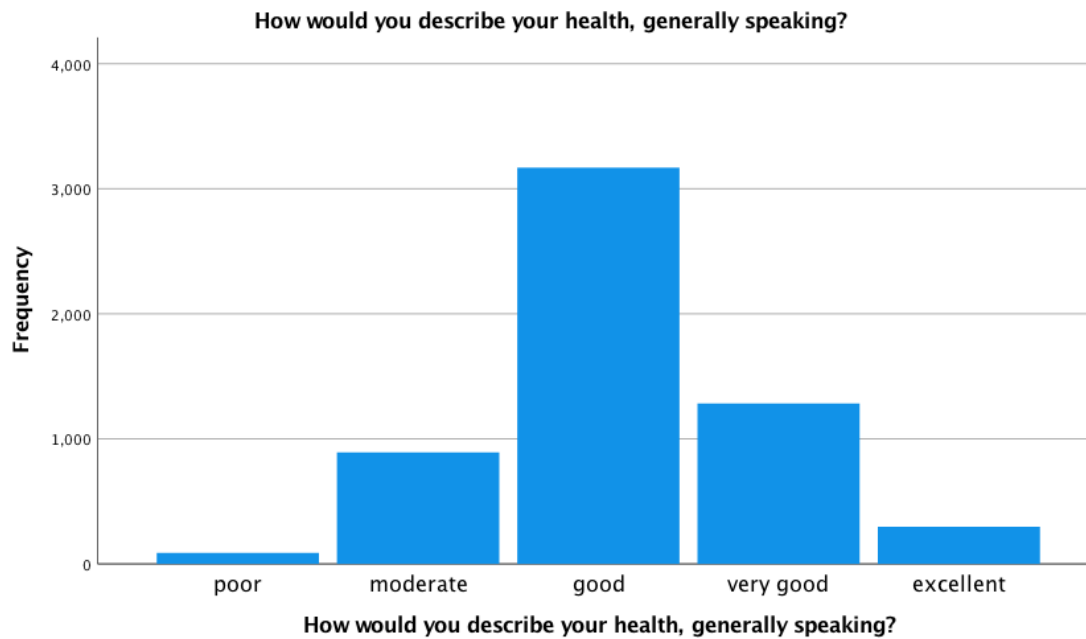
Statistics

How would you describe your

N	Valid	5730
	Missing	1065
Mean		3.14
Minimum		1
Maximum		5

How would you describe your health, generally speaking?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	poor	89	1.3	1.6	1.6
	moderate	892	13.1	15.6	17.1
	good	3169	46.6	55.3	72.4
	very good	1283	18.9	22.4	94.8
	excellent	297	4.4	5.2	100.0
	Total	5730	84.3	100.0	
Missing	System	1065	15.7		
Total		6795	100.0		



5.2 Health: recoding's and adjustments

No adjustments have been made to this variable.

6.1 Gender: Original variable

```
FREQUENCIES VARIABLES=geschlacht
/STATISTICS=MEAN
/ORDER=ANALYSIS.
```

Statistics

Gender

N	Valid	6795
	Missing	0
Mean		1.54

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	3107	45.7	45.7	45.7
	Female	3688	54.3	54.3	100.0
	Total	6795	100.0	100.0	

6.2 Gender: Recoding's and adjustments

The variable gender had to be recoded to a dummy-variable. The options for male or female have been recoded from 1 and 2, to 0 and 1.

```
RECODE geschlact (1=0) (2=1) INTO New_Geschlact.  
EXECUTE.
```

6.3 Gender: descriptive statistics of final variable

```
FREQUENCIES VARIABLES=New_Geschlact  
/STATISTICS=MEAN  
/ORDER=ANALYSIS.
```

Statistics

New_Geschlact

N	Valid	6795
	Missing	0
Mean		.5428

New_Geschlact

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	3107	45.7	45.7	45.7
	1.00	3688	54.3	54.3	100.0
Total		6795	100.0	100.0	

Appendix II

In appendix I, the descriptive statistics of the variables represented the scores that respondent gave in the full sample. For our regression analysis, the analysis is carried out on a subset of respondents who have a valid score on all variables. Therefore, the table with descriptive statistics in the result paragraph displays the subset rather than the full sample. The N will then be equal for all variables.

The following syntax was run to filter out missing data:

```
DATASET ACTIVATE DataSet2.
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Life_satisfaction
/METHOD=ENTER cs20m105 leeftijd New_Geslacht Inkomen_Duizenden ch20m004
/SAVE RESID.
```

```
RECODE RES_1 (SYSMIS=0) (ELSE=1) INTO obs.
EXECUTE.
```

```
USE ALL.
COMPUTE filter_$=(obs = 1).
VARIABLE LABELS filter_$ 'obs = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

1.1 Descriptive statistics

```
DESCRIPTIVES VARIABLES=cs20m105 Inkomen_Honderden Life_satisfaction New_Geslacht ch20m004
leeftijd
/STATISTICS=MEAN STDDEV MIN MAX.
```

```
FREQUENCIES VARIABLES=New_Geslacht
/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN
/HISTOGRAM
/ORDER=ANALYSIS.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
How many hours do you spend on sports per week, on average?	2257	0	130	4.57	4.357
Inkomen_Honderden	2257	.00	1456.66	19.6513	34.12815
Life_satisfaction	2257	.00	10.00	7.3673	1.40517
New_Geslacht	2257	.00	1.00	.5312	.49913
How would you describe your health, generally speaking?	2257	1	5	3.30	.781
Age of the household member	2257	16	93	51.00	18.519
Valid N (listwise)	2257				

New_Geslacht

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid .00	1058	46.9	46.9	46.9
1.00	1199	53.1	53.1	100.0
Total	2257	100.0	100.0	

1.2 Correlations

CORRELATIONS

/VARIABLES=Life_satisfaction cs20m105 leeftijd ch20m004 New_Geslacht Inkomen_Honderden
 /PRINT=TWOTAIL NOSIG FULL
 /MISSING=PAIRWISE.

Correlations

		Life_satisfaction	How many hours do you spend on sports per week, on average?	Age of the household member	How would you describe your health, generally speaking?	New_Geslacht	Inkomen_Honderden
Life_satisfaction	Pearson Correlation	1	.032	.130**	.259**	.002	-.005
	Sig. (2-tailed)		.124	<.001	<.001	.938	.809
	N	2257	2257	2257	2257	2257	2257
How many hours do you spend on sports per week, on average?	Pearson Correlation	.032	1	.049*	.063**	-.124**	.025
	Sig. (2-tailed)	.124		.019	.003	<.001	.238
	N	2257	2257	2257	2257	2257	2257
Age of the household member	Pearson Correlation	.130**	.049*	1	-.277**	-.056**	.112**
	Sig. (2-tailed)	<.001	.019		<.001	.008	<.001
	N	2257	2257	2257	2257	2257	2257
How would you describe your health, generally speaking?	Pearson Correlation	.259**	.063**	-.277**	1	-.059**	.003
	Sig. (2-tailed)	<.001	.003	<.001		.005	.879
	N	2257	2257	2257	2257	2257	2257
New_Geslacht	Pearson Correlation	.002	-.124**	-.056**	-.059**	1	-.099**
	Sig. (2-tailed)	.938	<.001	.008	.005		<.001
	N	2257	2257	2257	2257	2257	2257
Inkomen_Honderden	Pearson Correlation	-.005	.025	.112**	.003	-.099**	1
	Sig. (2-tailed)	.809	.238	<.001	.879	<.001	
	N	2257	2257	2257	2257	2257	2257

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

2.1 Regression analysis

Since our regression analysis involves a moderator effect, a few arrangements have to be made to the variables. First, our two independent variables *Sport* and *Age* have to be centered. This is done by recoding the variables to a new variable where the average has been subtracted. Secondly, an interaction variable has to be made of the two centered variables. This is done by multiplying the two centered variables into a new interaction variable.

```
COMPUTE Sport_c=cs20m105 - 4.53.  
EXECUTE.
```

```
COMPUTE Age_c=leeftijd - 50.41.  
EXECUTE.
```

```
COMPUTE SportxAge=Sport_c * Age_c.  
EXECUTE.
```

```
REGRESSION
```

```
  /MISSING LISTWISE
```

```
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE
```

```
  /CRITERIA=PIN(.05) POUT(.10)
```

```
  /NOORIGIN
```

```
  /DEPENDENT Life_satisfaction
```

```
  /METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden
```

```
  /METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden Sport_c
```

```
  /METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden Sport_c Age_c
```

```
  /METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden Sport_c Age_c SportxAge
```

```
  /SCATTERPLOT=(*ZRESID,*ZPRED)
```

```
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)
```

```
  /SAVE MAHAL COOK LEVER.
```

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	5.800	.131		44.167	<.001		
	New_Geslacht	.047	.058	.017	.815	.415	.987	1.014
	How would you describe your health, generally speaking?	.468	.037	.260	12.763	<.001	.996	1.004
	Inkomen_Honderden	.000	.001	-.004	-.208	.835	.990	1.010
2	(Constant)	5.803	.131		44.174	<.001		
	New_Geslacht	.053	.058	.019	.916	.360	.973	1.028
	How would you describe your health, generally speaking?	.466	.037	.259	12.692	<.001	.993	1.007
	Inkomen_Honderden	.000	.001	-.004	-.220	.826	.990	1.010
	Sport_c	.006	.007	.019	.902	.367	.981	1.019
3	(Constant)	5.415	.133		40.662	<.001		
	New_Geslacht	.088	.057	.031	1.551	.121	.970	1.031
	How would you describe your health, generally speaking?	.581	.037	.323	15.530	<.001	.912	1.096
	Inkomen_Honderden	-.001	.001	-.028	-1.405	.160	.978	1.023
	Sport_c	.002	.006	.006	.279	.780	.978	1.023
	Age_c	.017	.002	.224	10.716	<.001	.903	1.107
4	(Constant)	5.411	.133		40.628	<.001		
	New_Geslacht	.084	.057	.030	1.487	.137	.968	1.034
	How would you describe your health, generally speaking?	.583	.037	.324	15.571	<.001	.911	1.098
	Inkomen_Honderden	-.001	.001	-.030	-1.482	.138	.975	1.026
	Sport_c	-.003	.007	-.008	-.374	.708	.776	1.288
	Age_c	.017	.002	.222	10.623	<.001	.900	1.111
	SportxAge	.000	.000	.031	1.373	.170	.785	1.273

a. Dependent Variable: Life_satisfaction

Model Summary^e

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			
						F Change	df1	df2	Sig. F Change
1	.260 ^a	.067	.066	1.35786	.067	54.321	3	2253	<.001
2	.260 ^b	.068	.066	1.35792	.000	.813	1	2252	.367
3	.336 ^c	.113	.111	1.32484	.045	114.839	1	2251	<.001
4	.337 ^d	.114	.111	1.32458	.001	1.886	1	2250	.170

a. Predictors: (Constant), Inkomen_Honderden, How would you describe your health, generally speaking?, New_Geslacht

b. Predictors: (Constant), Inkomen_Honderden, How would you describe your health, generally speaking?, New_Geslacht, Sport_c

c. Predictors: (Constant), Inkomen_Honderden, How would you describe your health, generally speaking?, New_Geslacht, Sport_c, Age_c

d. Predictors: (Constant), Inkomen_Honderden, How would you describe your health, generally speaking?, New_Geslacht, Sport_c, Age_c, SportxAge

e. Dependent Variable: Life_satisfaction

Appendix III

In this research, the dependent variable *Life satisfaction* is a continuous variable. As a result, a linear regression can be performed. In order to do this, the data must meet a few conditions. These conditions are:

1. **LINEARITY:** The relationship between the explanatory and dependent variables is linear
2. **INDEPENDENCE:** Independence ensures that the observations are not related, The data were obtained from a random sample of the population
3. **HOMOSCEDASTICITY:** the variance of the error term is the same for all values of the explanatory variable.
4. **NORMALITY:** The distribution of the residuals should be bell-shaped and symmetrical. This ensures that the errors are normally distributed.
5. **NO MULTICOLLINEARITY:** The explanatory variables you include in the regression do not have a linear relationship

1.1 Linearity

Linearity means that the relationship between the explanatory and dependent variables is linear. Linearity can be checked by looking at the correlation table in Appendix II. If two variables show a correlation coefficient of 0, then there is no linearity between these variables. Fortunately, in our correlation table, all correlation coefficients seem to have a value. Therefore the relationship between the explanatory and dependent variable is linear.

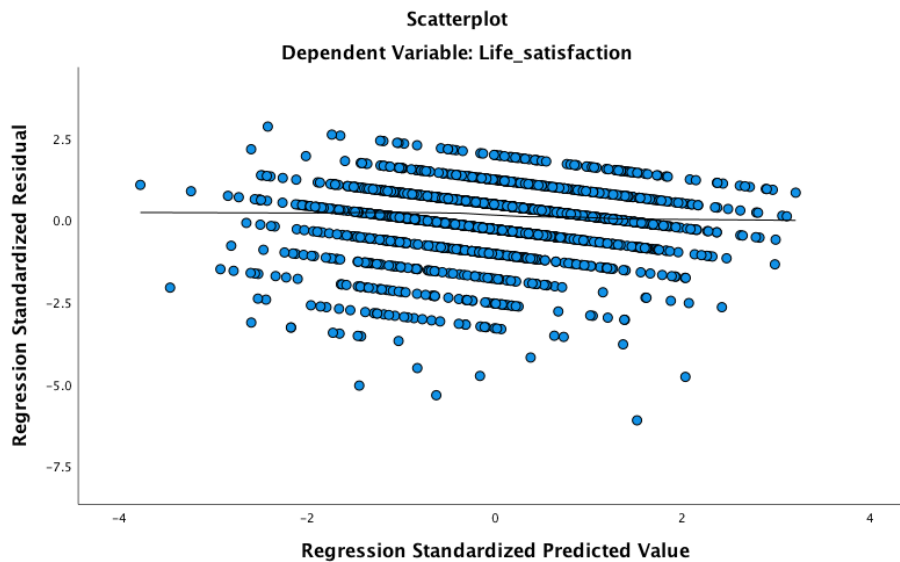
1.2 Independence

When conducting research, it is common to collect data through a sample. This way, the results can be generalized from the sample to the entire population. In order to do this, it is important that the sample is random and has the same characteristics as the population. In addition to the fact that the respondents are chosen at random, it is also important that the observations are independent of each other. This means that one observation should not influence the other.

The panel that is used for this research consists of households. This means that not all participants are fully independent from each other. This can have consequences for the results of this research. For estimating parameters in the research model, the sample can still be used. However, it is possible that it will not give optimal result. Generalizability cannot therefore be fully guaranteed.

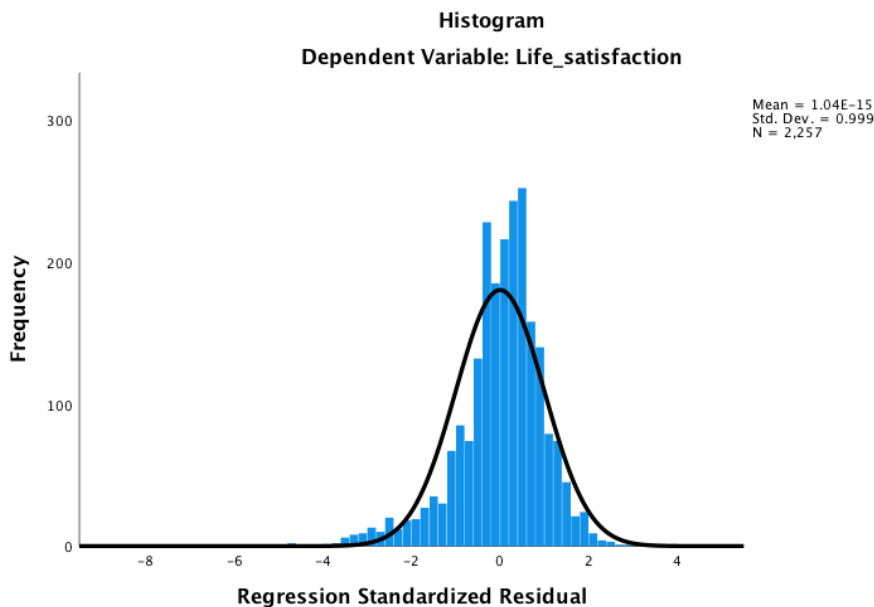
1.3 Homoscedasticity

The assumption of homoscedasticity states that the variance of the error term is the same for all values of the explanatory variable. Two figures can be looked at. With the scatterplot, the variance of the error term should be equal for all values on the line. In the figure below the variance does not seem to be perfectly equal, however is also does not seem to be very crooked. Therefore, the assumption is not heavily violated and will likely not influence the regression analysis.



1.4 Normality

The assumption of normality says that the distribution of the residuals should be bell-shaped and symmetrical. This ensures that the errors are normally distributed. The figure below shows the residuals. Here we see that the residuals are somewhat normally distributed. The residuals follow the line of the normal distribution curve. Therefore, this assumption is not violated



1.5 No multicollinearity

Multicollinearity is a statistical phenomenon that occurs when two or more independent variables in a regression model are highly correlated with each other. In other words, multicollinearity indicates a strong linear relationship between the predictor variables. This can be checked by looking at the VIF scores. In the table below it shows that the VIF scores average around 1,00. These are relatively low VIF scores, meaning that the assumption of multicollinearity has not been violated.

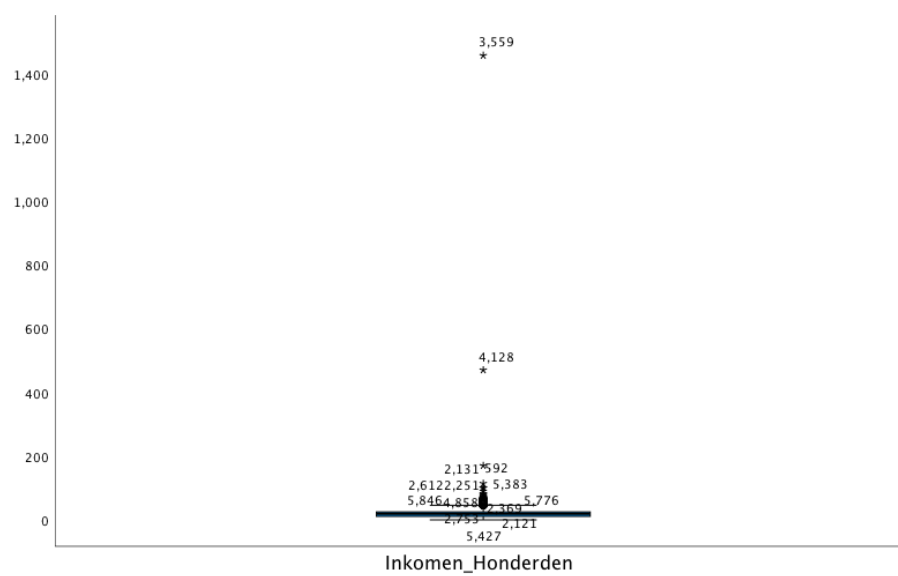
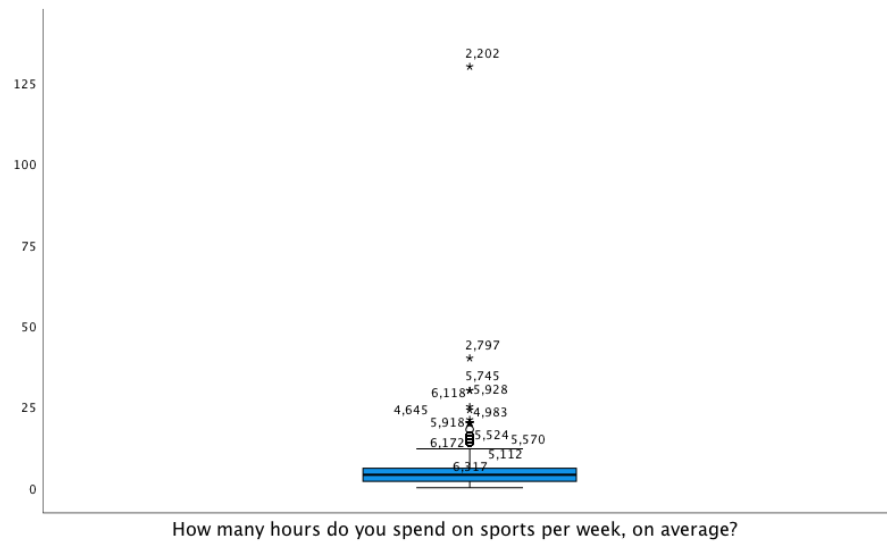
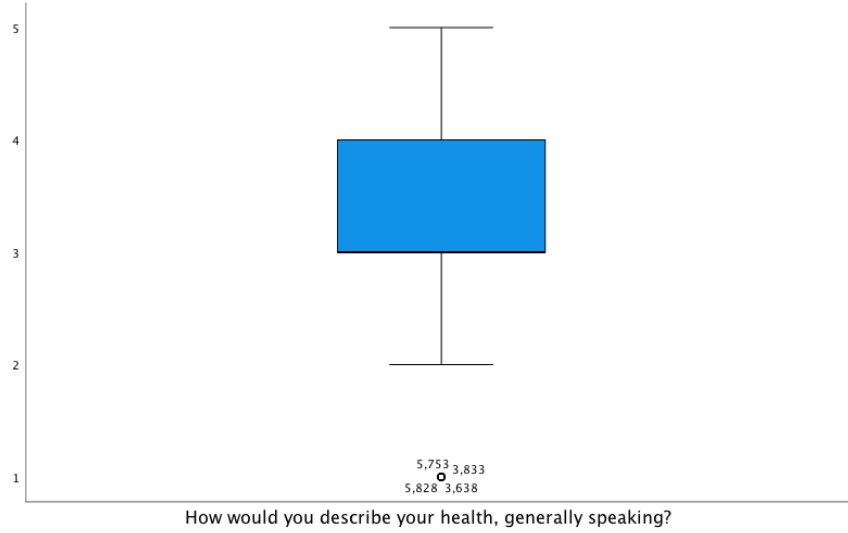
		Coefficients ^a						Collinearity Statistics	
Model		Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	Sig.	Tolerance	VIF	
1	(Constant)	5.800	.131		44.167	<.001			
	New_Geslacht	.047	.058	.017	.815	.415	.987	1.014	
	How would you describe your health, generally speaking?	.468	.037	.260	12.763	<.001	.996	1.004	
	Inkomen_Honderden	.000	.001	-.004	-.208	.835	.990	1.010	
2	(Constant)	5.803	.131		44.174	<.001			
	New_Geslacht	.053	.058	.019	.916	.360	.973	1.028	
	How would you describe your health, generally speaking?	.466	.037	.259	12.692	<.001	.993	1.007	
	Inkomen_Honderden	.000	.001	-.004	-.220	.826	.990	1.010	
3	(Constant)	5.415	.133		40.662	<.001			
	New_Geslacht	.088	.057	.031	1.551	.121	.970	1.031	
	How would you describe your health, generally speaking?	.581	.037	.323	15.530	<.001	.912	1.096	
	Inkomen_Honderden	-.001	.001	-.028	-1.405	.160	.978	1.023	
	Sport_c	.006	.007	.019	.902	.367	.981	1.019	
4	(Constant)	5.411	.133		40.628	<.001			
	New_Geslacht	.084	.057	.030	1.487	.137	.968	1.034	
	How would you describe your health, generally speaking?	.583	.037	.324	15.571	<.001	.911	1.098	
	Inkomen_Honderden	-.001	.001	-.030	-1.482	.138	.975	1.026	
	Sport_c	-.003	.007	-.008	-.374	.708	.776	1.288	
	Age_c	.017	.002	.222	10.623	<.001	.900	1.111	
SportxAge	.000	.000	.031	1.373	.170	.785	1.273		

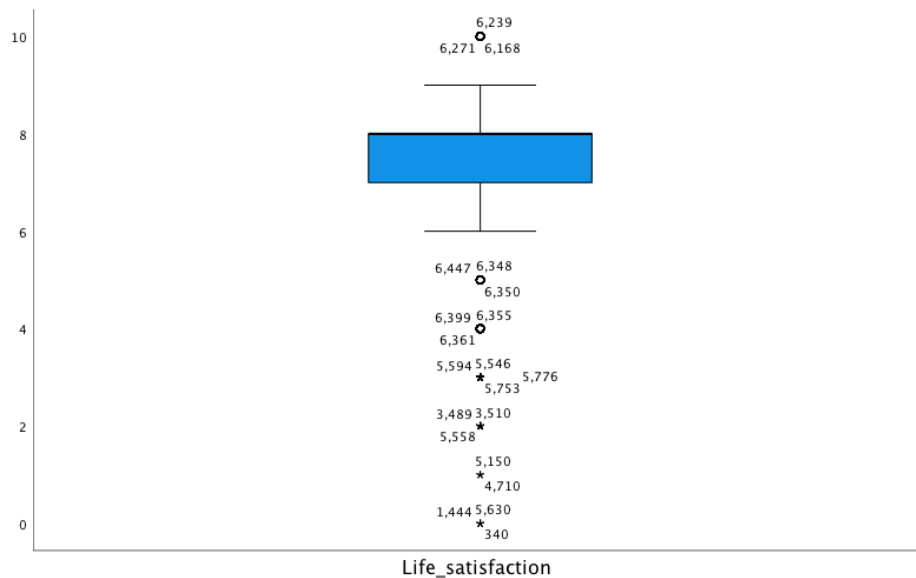
a. Dependent Variable: Life_satisfaction

2.1 Outliers

Outliers are extreme values that differ from most other observation points in a data set. They can have a big impact on the statistical analysis and influence the results of the hypothesis testing. There are a few ways to detect outliers. First, let's take a closer look at the boxplots:

```
EXAMINE VARIABLES=ch20m004 cs20m105 Inkomen_Duizenden New_Geslacht leeftijd Life_satisfaction
/PLOT BOXPLOT STEMLEAF
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES EXTREME
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```





The boxplots of these variables display which of the participants represent outliers. The circled numbers are considered big outliers, but the numbers with an asterisk are considered problematic. The variable Sport, Income and Life satisfaction seem to have a few problematic outliers. Let's put these outliers together in a casewise diagnostics model.

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Life_satisfaction
/METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden
/METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden Sport_c
/METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden Sport_c Age_c
/METHOD=ENTER New_Geslacht ch20m004 Inkomen_Honderden Sport_c Age_c SportxAge
/SCATTERPLOT=(*ZRESID,*ZPRED)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE DFBETA.

```

Casewise Diagnostics^a

Case Number	Std. Residual	Life_satisfaction	Predicted Value	Residual
77	-3.787	3.00	8.0163	-5.01626
115	-3.685	2.00	6.8807	-4.88072
340	-5.045	.00	6.6821	-6.68209
378	-4.780	2.00	8.3316	-6.33158
410	-3.185	3.00	7.2186	-4.21863
813	-3.186	3.00	7.2205	-4.22049
997	-3.264	3.00	7.3236	-4.32363
1001	-3.297	3.00	7.3668	-4.36681
1167	-3.460	2.00	6.5824	-4.58241
1274	-3.542	2.00	6.6914	-4.69136
1444	-6.105	.00	8.0869	-8.08687
1943	-3.322	3.00	7.4003	-4.40032
2444	-3.160	3.00	7.1853	-4.18535
2597	-3.529	2.00	6.6745	-4.67448
3012	-3.524	3.00	7.6684	-4.66838
3098	-3.275	2.00	6.3373	-4.33735
3430	-3.560	3.00	7.7159	-4.71589
3489	-3.122	2.00	6.1358	-4.13583
3510	-3.274	2.00	6.3362	-4.33624
3705	-3.306	3.00	7.3785	-4.37853
3941	-3.039	4.00	8.0254	-4.02540
4710	-4.751	1.00	7.2926	-6.29257
4739	-3.242	3.00	7.2949	-4.29492
4896	-3.056	3.00	7.0475	-4.04750
5083	-3.089	3.00	7.0910	-4.09099
5150	-4.511	1.00	6.9756	-5.97557
5490	-3.258	3.00	7.3160	-4.31599
5558	-3.434	2.00	6.5488	-4.54875
5630	-5.339	.00	7.0715	-7.07149
5733	-3.037	4.00	8.0221	-4.02212
5776	-3.021	3.00	7.0010	-4.00104
6037	-4.189	2.00	7.5488	-5.54876

a. Dependent Variable: Life_satisfaction

What is interesting about this table is that some of the outliers detected in the boxplot, are not even mentioned in this table, like cases 3559 and 2202 who seem to be far off from the boxplots (3559 for *Income* and 2202 for *Sport*). Apparently, their residuals do not seem to be further than 3 standard deviations and are therefore not technically classified as outliers. The outliers mentioned in the table above do have a residual that is further than 3 standard deviations away. Since there seem to be a lot of outliers, it is important to determine whether these outliers are actually problematic. This can be done by looking at the DFBETA. The DFBETA is a tool to measure whether the removal of an influential observation would noticeably change the effect of the coefficient.

The outliers mentioned in the table above do not show high DFBETA values. Therefore, the outliers are not that problematic and do not have to be removed from the dataset.

Appendix III: AI software

While writing this research, I used an AI website that assists with English writing skills. The website I used is called Grammarly. Grammarly is a free AI-operated website that improves academically written texts by checking sentences, spelling and punctuation. I used this website because I wrote this thesis in English and I have less experience in writing academic texts in English. Grammarly is only used as a tool to improve grammar. No other inspiration was taken from this program.