A Work Project, presented as part of the requirements for the Award of a Masters Degree in Economics from NOVA School of Business and Economics

Obesity and Socioeconomic Gradient

Constança Youd David

Student number - 527

A Project carried out with the supervision of

Professor Pedro Pita Barros

January 7th, 2013

Abstract

This work project studies the determinants of body weight for Portuguese population with special incidence in the socioeconomic dimension through a comparison between 2006 and 2012. More specifically how is body weight influenced by a set of socioeconomic variables that may vary throughout the years. In other words, we set out to understand if socioeconomic status (SES) is a real contributor regarding weight. This study entails a thorough analysis of these variables with the intention of understand which of them remain significant and have real impact on the weight of a human body. Literature points to a confirmation of this real impact of SES status in the body weight and, according to our results, income represents the strongest variable to explain changes in human body weight, coinciding with previous findings. In quantitative terms, an increase of 250€ in the household net income for 2012 is translated into an increase of 0.272 in the BMI. Data used was provided by National Health Institute (INSA) along with National Institute of Statistics (INE) and also collected through a survey.

Keywords:

Body Mass Index

Body weight

SES

1. Introduction

The International Obesity Taskforce considers obesity one of the most serious medical and public health problems of our times. Obesity refers to an increase in body weight beyond the limitation of skeletal and physical requirements, as the result of excessive accumulation of body fat, mainly in the viscera and subcutaneous tissues of the body. It is associated with increased disability and many potentially life-threatening health problems, including hypertension, diabetes, increased risk of coronary diseases, increased unexplained heart attack, hyperlipidemia, infertility and a higher prevalence of colon, prostate, endometrial and breast cancer. In a more extreme scenario, we should add death to the above list since approximately 300,000 deaths a year are attributed to obesity. This means that obesity goes beyond health, it is also an economic phenomenon, entailing both direct and indirect costs related with chronic diseases and mortality.

The main purpose of this work project is answering the following research question – Does the increase in SES lead to an increase of the prevalence of obesity? Are SES inequalities a real contributor to the increase in obesity? Or environment and other social factors represent a better explanation? Our analysis shows us that SES seems to have a great contribution to explain changes in human body weight, and its impact has been increasing over the years.

According to Olshansky et al. (2005), "unless effective population-level interventions to reduce obesity are developed, the steady rise in life expectancy observed in the modern era may soon come to an end and the youth of today may, on average, live less healthy and possibly even shorter lives that their parents". Taking into consideration the impact of excess body height, it is crucial that we measure it accurately. The most common measure is the body mass index (BMI) defined by the ratio of weight divided by height squared (note that weight is measured in kilos and height in metres). This proxy of body weight is less accurate than laboratory measures of body composition because its formula is limited since it does not account for variations in muscle mass or in the distribution of body fat (e.g. an individual who practice a lot of sports tends to transform more easily body fat into muscle mass but may maintain or even increase his overall weight). Nevertheless, BMI is a favoured method of assessing excess weight – it is simple, rapid and inexpensive to calculate. The World Health Organisation defines different categories according to the level of BMI – An individual that presents a BMI lower that 18.4 is consider to be "underweight"; a BMI between 18.5 and 24.9 means that the individual is "normal weight";

an “overweight” person needs to present a BMI between 25 and 29.9; and an individual with a BMI higher than 30 is said to “obese”. Obesity is further divided into three groups: class 1 (BMI 30 to <35), class 2 (BMI 35 to <40) and class 3 (BMI ≥40).

There are two main important issues regarding this measure of weight. The first has to do with the idea that self-reported data on height and weight is usually measured with error. According to Strauss (1999), Goodman et al. (2000), and Kuczmarksi et al. (2001), there is a tendency for individuals to over-report height and understate weight. This phenomenon leads to an underestimation of BMI. A number of regression-based procedures have been proposed and performed to correct the self-report errors but in accordance with Baum II and Ruhm (2007) the results may not be substantively altered by doing so regarding limited data. Thus, we will present our results and conclusions based on the uncorrected BMI. Second, official statistics use a more complex criterion for children (under 21 years old) under the Center for Disease Control and Prevention’s National Center for Health Statistics. We decided to use the same criterion for all respondents in order to provide some consistency across individuals.

Worldwide obesity has more than doubled since 1980. In 2008, more than 1.4 billion adults, 20 and older, were overweight. Of these, over 200 million men and nearly 300 million were obese. More than 40 million children under the age of five were overweight in 2010. These statistics show the severe impact of obesity in the world population and that is the reason why this disease should not be neglected.

Portugal is not immune to this global problem of excessive body weight. According to recent studies, more than half of the Portuguese population (53%) are overweight and many are already suffering from morbid obesity. In fact, Portugal is part of the group of countries with the highest percentage of young people who are overweight or obese. In a ranking of 39 European and North America states, the country appears in the fifth place considering students aged 11. This data comes from the Health Behaviour in School-aged Children – a large survey of behaviours and lifestyles of adolescents conducted every four years in Europe and North America in cooperation with the World Health Organization. These studies point out poor eating habits that escalated in recent decades in the life of the Portuguese and the lack of exercise as responsible for the cause of this ”silent scourge”. Alerts are made with a special preoccupation of re-educate the population concerning its eating habits in order to avoid the proliferation of diseases associated with excess weight, with special incidence in children. Preventing childhood obesity is acting directly on the root of the problem through initiatives responsible for enhancing the health of children and to promote fight against physical inactivity and poor nutrition. But the truth is that there is no interest in changing the situation because economic interests always end up being above the health and quality of life.

This work project attempts to test whether some determinants have real impact in BMI or not and what is their relationship with human body weight. For that we use data from an inquiry from National Health Survey along with INE (Statistics Portugal) and from a survey in which I included a limited number of variables and compare values between both, defining the mean of BMI as a starting point.

In the second part of this paper we will describe the raw data we worked with and how it was transformed in order to obtain the final database that was used.

In the third part we will discuss the methodology used, essentially in terms of econometric and statistical analysis.

Subsequently, we present and analyze the statistical results of our model and discuss its implication. This section is divided into two distinct parts: an analysis that will be done separately regarding the three years and a comparison between samples within the databases regarding the same range of ages, which will be explained later.

Finally, we move to the conclusions where we will summarize and highlight the most important findings.

1.1. Literature Review

There are a lot of factors that contribute to overweight and obesity. According to Madden (2010), the incidence of obesity is more pronounced for women than for men and it has declined between 2002 and 2007 due to an increased obesity amongst the better-off instead of lower obesity amongst the less well-off. Using the data collected through Slán Surveys, "the socioeconomic gradient in obesity is exclusively confined to the incidence of obesity rather than what we might call the intensity of obesity." For men, the biggest contribution for obesity comes from education and self-assessed health while in females' case it comes from education, self-assessed health and, with some relevance, from equivalised income.

In a complementary line of thought Devaux et al. (2009) consider that large and persistent social inequalities in obesity and overweight by education level and socioeconomic status exist in OECD countries and these are consistently larger in women than in men.

As pointed out by Baum and Ruhm (2007), body weight rises with age but is negatively correlated with SES (socioeconomic status) meaning that the prevalence of excess body weight is more common in older people since levels of overweight increase from childhood up to age 75 years and also, obesity tends to decrease with income and education. This finding suggests that efforts to prevent or reduce obesity need to start early in life and continue as least until retirement time. Socioeconomic inequalities in body mass are marked although data from the HSE suggest these do not appear to have widened over the last 15 years. There is a clear need to focus on these inequalities in body mass, according to the authors.

Another important aspect highlighted by the Public Health Research Consortium is genetics. Does the weight gain among parents influence the weight gain in children? Based on its results, intergenerational effects (composed by genetics and food habits) may be responsible for amplifying the growth of the obesity epidemic through the generation of a repeating cycle. It will be beneficial to help parents to adopt lifestyle changes that can provide consistent models for their children and shape the family's environment. A genetic predisposition to weight gain, however, does not automatically mean that a person will be obese. Eating habits and patterns of physical activity also play a significant role in the amount of weight a person gains.

Beyond this intergenerational effect, Wang and Zang (2004) introduce the impact of ethnicity in one of their studies. They found remarkable ethnic differences in the relationship between SES and obesity. Although the extant literature documented a higher prevalence of obesity among minorities than in whites, their results presented a lower socioeconomics inequality in obesity within minority groups. This analysis suggested that besides gender and age, also ethnicity could be an important factor on socioeconomic inequality in overweight and obesity.

The idea of an inverse relation between SES and overweight or obesity in most cases is also supported by Stamatakis et al. (2010). According to the authors, 41% of the reviewed studies show this relation opposed to 31% reporting a mixture of inverse and no associations. This has to do with the importance of taking into account multiple SES indicators since studies concentrated on a single SES indicator fail to fully describe correct interactions. The presented results show that, in this case, children with less advantaged socioeconomic backgrounds tend to display higher rates of prevalence than children with higher income or better support. The collected data also show that the obesity epidemic has been slowing in recent years, which is supported by reversed or stabilized trends in overweight or obesity in some countries such as France, Switzerland and Sweden. This fact can be explained by an extensive media attention around the issue of obesity and the rise in aesthetics awareness and health consciousness along with some anti-obesity policies and established targets.

The relationship between obesity and socioeconomic status varies remarkably across countries with different socioeconomic development levels as it is defended by Wang et al. (2002). “Higher SES subjects were more likely to be obese in China and Russia, but in the low-SES groups were at a higher risk. Obesity was more prevalent in urban areas in China but in rural areas in Russia”. In other words, different SES groups from different countries are at different risks. Considering the example of China and its son preference towards daughters, men are expected to achieve higher SES than women and so tend to present lower rates of prevalence of obesity. This scenario support an important result pointed out by Wang - the association between obesity and SES weakened over time and SES inequality was not an important contributor to the dramatic increase in the prevalence of obesity. Findings suggest that other social and environment factors, which have influenced changes in people’s lifestyle, might explain the increasing overweight problem. Effective intervention efforts for the prevention and management of obesity should target all SES groups from a population perspective.

2. Data

2.1. Sources

The information used in this work project was obtained through two separate sources: from the National Health Survey from 1999 and 2006, and from a survey created on purpose for this work project (2012). Both sources provide information about habits and behaviours of the population regarding health care, nutrition and extra activities. This information will be divided into 1999, 2006 and 2012, representing the years in which the inquiries were answered.

2.1.1. National Health Survey

The National Health Survey (INS) is conducted in partnership between the National Institute of Health Dr. Ricardo Jorge (INSA) and the National Institute of Statistics (INE). So far have been conducted four times using probability samples representative of the population in mainland Portugal and the Autonomous Regions of the Azores and Madeira.

The information provided by INS consists in two inquiries from 1998/99 and 2006. The first database is composed by 257 variables divided into 12 categories such as socio-demographical characterization, general information about health, chronic diseases, tobacco consumption, expenses and income, medical care, among others. The inquiry from 2006 is much more complete since it is composed by 430 variables divided into 18 categories – the same 12 categories of the previous database and 6 additional ones like use of medication, life quality and preventive care. It is important to emphasize that the variable "age" is described according to 19 categories, each one representative of one interval of ages.

2.1.2. Own Survey

The later source, created in order to compare years of 1999 and 2006 with the present one, is a survey in which we included just 27 variables. The idea was creating a more succinct database representing almost all the categories mentioned above through one or two variables. Besides the fact that it is more limited in terms of variables, this database is restricted to a range of ages from 18 to 33 years.

Our intent is that the included behaviours capture the effects of a broad range of lifestyle factors even though they are not all represented in our model. For example, we make no attempt to resolve the debate over whether smoke quitting play a role in explaining the growth in obesity. Instead, we incorporate information on tobacco use in hope that it proxies the effects of a constellation of health inputs that may be related to obesity. The variables that compose this database will be considered to be the focus of our work and will be these variables that we will account for regarding 1999 and 2006.

2.2. Building the data base

The purpose in which we set out with this work project obviously implies a comparison between 1999, 2006 and 2012. However, the ranges of ages of the initial databases are different, making a direct relationship impossible. Instead, it is necessary to create samples within databases in order to make a comparison given the same age group. For 1999 we have ages varying between 0 and 103 years old. In the 2006 inquiry the range of ages goes from 0 to more than 85 years old and the survey of the present year accounts for people between 18 and 33 years old. We intend to use the latest survey to define the ages of our samples, since it represents the smallest range. The consequence of this fact is that we come up with three samples corresponding to the three databases restricted to the interval of ages of 18 to 33 years old. This approach naturally has a tendency to mitigate the impact of age in the body weight.

3. Methodology

The model that we have built uses as dependent variable the weight definition of a certain individual, given by the body mass index. As mentioned before, this index obeys to a specific scale so that our focus goes to people with a BMI equal or higher than 30, considered to be obese. Note that it is also important to take into consideration individuals which index goes from 25 to 29.9 since they belong to the category “overweight”. Here, the body mass index, as explained above, is obtained through the weight in kilos divided by height squared in metres. Given that, we perform the calculation of this formula in all databases, adding two new variables – height in metres and height in metres squared - in order to create our continuous dependent variable called BMI (See Appendix 1, representing the distribution of our dependent variable for the three years).

One of the problems of these inquiries is that almost all questions can be answered with “don’t know” or “don´t answer”. We decided to treat them as missing values.

The variable that we have chosen to work with is continuous. Its value is not limited regarding the superior limit (it can’t be a negative value) even though it is more likely to be between a specific interval. Because of that, estimating standard linear regressions for this type of variable seems suitable. We could also use the analysis of covariance or even the ANOVA but OLS is simpler and it is able to provide the same qualitative results.

3.1. Ordinary Least Squares method (OLS)

OLS is a method for estimating the unknown parameters in a linear regression model. This method minimizes the sum of squares vertical distances between the observed responses in the dataset and the responses predicted by the linear approximation. Since the linear regression model is well-unknown, we can mention the professional approach of Wooldridge (2009)[1].

3.2. Model

Before stating the model it is important to discuss the intuition behind the factors that affect our dependent variable. It can useful to separate these factors into main

[1] Wooldridge, Jeffrey M. (2009), “Introductory Econometrics: A Modern Approach”, 4th edition.

categories, since the amount of variables differs from database to database. Variables will be sorted according to eight categories: socio demographic characterization, education, economic situation, health, physical activities, eating habits, tobacco consumption and genetics. Education and economic situation represent our proxies to SES.

We will be using body mass index as a dependent variable and our main focus is to study the impact of socio-economic status in this variable. It is important to highlight that the impact of some variables may not be as simple as it was expected.

Age and gender are the most relevant variables we need to account for regarding the socio demographic characterization. We expect older people to present higher levels of body weight since this age group is associated with more health problems, mainly specific diseases and physical and mental disabilities, as mentioned by Baum and Ruhm (2007).
This can reduce the ability to perform daily tasks by their own, leading to a more sedentary lifestyle. In our particular case, this relationship is not very relevant because we are restricting the range of age (18-33 years old), but we will still analyze this variable. Gender is also an important variable, as women are known to be less athletic than men since they tend to easily accumulate more fat mass. Pregnancy may also play an important role.

Regarding education sector we need to take into consideration factors like years of schooling and high degree of education (under degree, master's degree, PhD, bachelor, etc). These two variables together are the main proxy for SES. We can use the level of education to understand who well-informed an individual is about nutrition, it will be expected that people with more years of schooling and higher degrees present a lower body mass index.

The economic condition includes income and expenditures. Our focus will be on income. In a more basic approach, an individual with a higher income will have a higher purchasing power, being able to acquire any type of food especially the so-called fast food, which is not considered to be cheap food anymore. This means that this individual will tend to present a higher body mass index than an individual with less purchasing power. On the other hand, high income usually means high education and since, as mentioned before, people with a higher level of studies tend to present a lower body mass index, we could also consider that people in a good economic situation are associated with lower body weight. In the initial dataset of all three years, income is a categorical variable with different magnitude of intervals for each year, which will enable an economic interpretation. In order to solve this problem we created a new variable (called "rendimento2") in which we used the average value of each interval instead of categorical values (1, 2, 3, 4, 5). This way, we can now interpret the impact of an increase of a particular value in Euros in our dependent variable. For the specific case of 1999, year in which we were still using “escudos”, we decided to convert escudos in euros, in order to simplify the analysis. It is important to highlight that we are not taking into consideration the household, meaning that we will not be able to adjust the household’s net income to the number of people that compose that particular household. In this case, we do not have that information but supposing we could access it, we would use OECD equivalence scales. The idea is based on a technique in which members of a household receive different weightings. According to OECD, to the first adult is assigned a weight of 1.0, for the second and each subsequent person aged 14 and over is attributed a weight of 0.5 and to each child aged under 14 is assigned a weight of 0.3.

Another crucial determinant is health. In this category we are including possible diseases like diabetes, asthma, depression, chronic anxiety among others, the consumption of any type of medication or treatment, physical and mental disabilities. Note that the INS also provides several other health related variables like oral health and preventive care that we will not use in the model since they are not statistically relevant. Variables that express diseases or the use of medication or treatments are dummy variables indicating the presence of the disease or the consumption of medication/treatment, rather than not. The intuition is that individuals with no diseases or disabilities tend to be healthier and so present an average body mass index. In the presence of some diseases the relation may not be so simple – being obese is a risk factor to the development of some diseases meaning that we have an association between the disease and the body weight (presence of inverse causality). Some medication can have impact on our dependent variable as well. Antidepressants present an association with human body weight.

Regarding physical activities we are considering the extra activities, practice of sports and its frequency. People who practice sports or other activity that involves a certain level of effort have more likely to present lower body weight. However, it is important to highlight that the practice of exercise usually means more muscular mass which contributes to a weight increase. In other words, the daily practice of exercise may also lead to an increase of the levels of body mass index. Once again, this is not an obvious relationship and may influence the signal of the coefficient. In our survey the information about physical exercise is provided through a dummy variable indicating if the respondent practices exercise or any type of sports and two categorical variables regarding the frequency per week and the time spend in each time.

Eating habits represent an extremely important category to explain changes in body weight. Concerning the nutrition field, we are taking into consideration not only food but also alcoholic beverages - the number of daily meals, the type of food in which meals are based on and the quality of diet, the frequency of fast food and the consumption of alcohol mainly the type of beverages and their frequency. The better the eating habits, the healthier the individual leading to an average BMI and the probability of being below the average is higher than be above that value.

Another included variable will be the consumption of tobacco. This factor may be able to explain some eating habits such as the lack of some type of food or the incomplete meals. The number of smoked cigarettes per day can be a good proxy to understand the impact of this factor, determining the level of addiction as well as tentative to quit or reduce. These variables are mainly categorical with exception to the dummy variable indicating if the respondent is a smoker or not.

The last determinant has to do with genetics. As stated by some authors, intergenerational effects may amplify the tendency to be overweight or obese. We will expect individuals with obesity cases in the family to present a higher tendency to be over weighted or obese, corresponding to a higher body mass index. The relationship is explained by a dummy variable indicating if the individual has any case of obesity running in the family or not.

These categories are only used to simplify the approach and represent a guiding line to the 2012 survey. As mentioned in the data section, all three databases differ from the number of variables and it is not possible to take into consideration 430 variables. Thus, we will only consider variables that fit, at least, in one of these categories. Basically, this means that we will not be able to present a model that is exactly the same for all the three years. The year of 1999 will contains variables from which we have no information in 2006 and the same will happen in reverse.

We can represent a simplified model using regression analysis to explore the association between these categories and body weight:

Equation 1

$$Bmi_i = \beta_0 + \beta_1 sociodem + \beta_2 income + \beta_3 educ + \beta_4 health + \beta_5 physact + \beta_6 eathabits + \beta_7 tobacco + \beta_8 genetics + u_i$$

, where β_i represent the coefficients of each variable and u_i is the error term.

This is the model that we will use in the analysis of 2012. The other two reviews will contain more variables in relation to the categories mentioned above but they are all based on this model.

Before we move on, it is important to characterize our dependent variable so we know what type of variation we are trying to explain. The mean of this measure or body weight for 1999, 2006 and 2012 is 25.4481, 25.0671 and 22.5192, respectively. The table in appendix 2 provides this information. Based on these numbers, we could state that there is a tendency for the decrease of BMI over the years, but if we take into consideration that we have different characteristics for different datasets, this comparison may not be accurate. In fact, appendices 6, 9 and 12 illustrate the BMI mean for the three age limited samples, and the tendency is not the same anymore.

4. Results

The presentation of the results for the three years will be done separately. We will first present the regression with the initial variables and then the regression without the variables that are not statistically significant.

As we state before, we will compare the three years in the next section. For this purpose, and to simplify the comparison later on, we need to guarantee that we are working with populations with the same characteristics, mainly age, years of schooling and income. The table below shows us the population characteristics of each year dataset:

Table 1 – Population characteristics and selected age analysis interval for 1999, 2006 and 2012

Variables	1999	2006	2012
Age	[0,103]	[0,>85]	[18,33]
Age analysis interval	[18,33]	[18,33]	[18,33]
Years of schooling	[0,24]	[0,24]	[5,23]
Income	[0,>2000]	[0,>2000]	[0,>5000]

As we can see, the interval of income seems to coincide in all years but regarding the other two variables, the year of 2012 presents different ranges of ages and years of schooling. This means that we cannot take conclusions by directly comparing the 2012 survey with the other two inquiries, which were created to represent the population. Instead, we will have to choose two samples, one within the 1999 population and the other within the 2006 population, according to the characteristics of the 2012 population. Note that, since differences in years of school are very small, we decided not to restrict samples regarding this variable because the results will be the same.

4.1. Analysis for 1999:

In appendix 4 we can see the Stata output of the initial regression presented in the previous section using the Ordinary Least Squares method already described. Almost all variables appear not to be statistically significant which means that they don't explain body weight. Note that in order to avoid omitted variables due to collinearity, we eliminated a few variables from the initial regression.

The next step was dropping the non significant variables one by one until we reach a set of variables that really affects the dependent variable. This selection process will influence the number of observations. We are considering a significant level of 10%, meaning that all p-values of variables have to be below 10% otherwise we will not reject the null hypothesis that states $\beta_i=0$. Thus, the table below represents the final choice of variables that are statistically significant:

Table 1 – final regression for 1999

BMI	Coef.	t	P>|t|
sexo	-2.000542	-13.5	0
idade	0.1590384	10.72	0
anosescola	-0.060848	-2.82	0.005
dorescostas	-0.348779	-2.01	0.045
nradiografias	-0.292354	-1.76	0.078
rendimento2	0.0003147	-2.54	0.011
comerforaref	-0.162211	-2.33	0.02
carne	-0.387863	2.84	0.005
legumes	-0.447844	-2.63	0.009
ndias	0.0427895	1.03	0.005
_cons	23.62093	35.93	0

Focusing on our main variables, gender, age, years of schooling and income, we can observe that both gender and age have p-values equal to zero, and income presents a p-value of 1.1%. Furthermore, the sign of the coefficient is the one expected for age and income. According to this regression, increased age leads to an increase of the body weight. The relationship between body weight and income is positive which means that people in a more comfortable economic situation present higher body weight. As we explained in the model section, this may not be an obvious relation but in a theoretical approach it seems very intuitive. Although, according some authors, this relationship should be negative. Regarding the variable “gender”, the relationship with the dependent variable appears to be the opposite of what should be expected. The coefficient of this variable has a negative sign, and taking into consideration that the respondent chooses 1 if male and 2 if female, this means that males tend to be characterized by a higher body mass index. The variable that represents years of schooling is also statistically significant and it presents a negative coefficient which corresponds to a consistent conclusion – more educated people tend to present lower levels of body weight.

In health category, we came up with one relevant disease to explain changes on body weight, back pain (“dorescostas”), which is a binary variable that takes the value 1 for the presence of the disease and the value 2 for the absence of it. In order to understand the association between some diseases and the body weight, we might need scientific explanation, but we must keep in mind a possible inverse causality for some diseases. This means that being obese is a factor of risk for the presence of some diseases. The variable presents a negative sign and so people that suffer from this disease are associated with high values of body mass index. There are also a variable represented in the regression concerning medical treatments and tests – the number of radiographies done in the last three months (“nradiografias”) that we will disregard.

The consumption of certain types of food is represented by the consumption of meat, which is not very significant to explain the variation of human body weight. In nutrition terms, in order to obtain a balanced diet, individuals should avoid red meat and consume white meat instead. Supposing that this variable corresponds to red meat, the negative coefficient sign makes sense, since this is a binary variable taking the value of 1 for the consumption of meat and the value 2 for the absence of it. Thus, people who consume red meat are more likely to have higher BMI that people who do not consume this good.

Information about physical activity is also provided through the number of days per week in which the individual practise any type of regular activity, which is a continuous variable. The data shows a positive relation between the variable and the variation of human body weight. The regression states that individuals who practise regular activities more times per week tend to have a higher BMI. Once again, we must consider the creation of more muscular mass since the individual is very active regarding regular physical activities. This means that, in some cases, we are not sure about which is the strongest argument – if more regular activity leads to an increase of muscular mass or if more physical activity lead to a decreasing body weight process.

It is important to emphasize the overall significance of the model. We can do it directly through the table above or we can perform the command "test" using all independent significant variables. The value of the F-statistics is zero, which implies that we reject the null hypothesis H_0: $\beta_1=\beta_2=...=0$, and so the model is significant, overall. Additionally, the model is able to explain 20% of the variation of our dependent variable, a value provided by the Adjusted R-squared., used to allow a comparison between the three years.

Source	SS	df	MS
Model	2796.56892	10	279.656892
Residual	11134.3375	1605	6.93728192
Total	13930.9064	1615	8.62594824

Number of obs =	1616
F(10, 1605) =	40.31
Prob > F =	0.0000
R-squared =	0.2007
Adj R-squared =	0.1958
Root MSE =	2.6339

4.2. Analysis for 2006:

The appendix 7 represents the Stata output of the initial regression we run for the year of 2006 using the OLS method previously described. The independent variables are quite similar to the regression above and, once again, the majority of them seem to be not statistically relevant since their p-values are higher than 10%.

As explained before, variables regarding physical activities appear to be relevant to justify variations in body weight but looking through the regression we can see that they are not included. This happens because the database has a lot of missing values which lead to regressions with no observations. In other words, we are not able to join certain variables like regular activity because the pool of observations of this variable does not coincide with the pool of observations regarding other variables.

The next step is repeating what we have done for 1999 – drop non significant variables, one by one until we reach a set of variables able to influence the dependent variable. We came up with the following table, representing the final choice of variables. Note that we are still working with a level of significance of 10%.

Table 2 – final regression for 2006

BMI	Coef.	t	P>|t|
idade	-1.30554	-3.16	0.002
comer	0.613252	1.9	0.058
diabetes	-1.82185	-3.34	0.001
tensaoalta	-0.67032	-1.66	0.098
osteoporose	1.714708	2.84	0.005
obesidade	-5.40063	-4.73	0
rendimento2	0.000746	2.09	0.037
vinho	-0.98109	-2.37	0.018
_cons	61.53859	7.56	0

Regarding to our primary variables – age, gender, years of schooling and income – we are forced to drop the variable gender ("sexo") and the variable years of schooling ("anosescola") because according to their p-value, these were not relevant to the model, which was not expected. Despite these results, both coefficient signs were in perfect concordance with what should be expected: gender was characterize by a positive sign since it is a dummy variable meaning that respondents could answers 0 to male and 1 for female and the coefficient of years of schooling had a negative sign. Both relationships were explained for 1999. About age and income, both present low p-values of 0.2% and 3.7%, respectively, but coefficient sigs are not so clear. The above data shows a negative relationship between age and BMI which means that younger people are more likely to have higher levels of body weight. This result does not make sense and according to the authors of our Literature Review, the relation should be exactly the opposite, which happens when we consider the total sample without restricting it in terms of age. But in this case, we should not take it too serious since we are leading with a small range of ages (See Table 1). We will have a tendency to disregard this variable. Concerning income, the regression presents a positive relationship with BMI. Thus, people in a better economic situation tend to present higher body weight.

Diseases, disabilities and the health situation are the three components of this regression representing the health category. In the topic of disabilities we can highlight one expected relationship regarding the variable of speaking help to eat ("comer"). This variable has to do with the ability of eating and it can take the value of 1 for people who can eat by their own, 2 for people who can do it but with some limitations and 3 for the ones who can only do it with help. We expect people who can eat by their own to present higher levels of body mass index than people who need help to do it. This relationship is not supported by the positive sign of the coefficient, representative of the opposite relation in which people who need help to eat present higher BMI. Once again, this relationship may not be very relevant since young people (18- 33 years old) are not likely to present this type of disability. The sector of diseases is composed for diabetes ("diabetes"), osteoporosis ("osteoporose") and high blood pressure ("tensaoalta"). All three variables are binary taking the value 1 for the presence of the disease and 2 for the absence of it. Osteoporosis is represented by a positive coefficient sign meaning that people with this disease tend to have lower levels of body mass index. Once again, we should take into consideration the possibility of having inverse causality since relationships between diseases and BMI are not obvious. Nevertheless, we can try to use logic arguments to explain part of these relations. Considering the example of osteoporosis – since this disease makes part of aging process, people suffering from it should be associated with high levels of body mass index due to the impact on age in body weight. Instead, the impact appears to be the opposite. As we are limited in terms of age, this relationship ends up being less relevant that it should be, especially because age also presents an unexpected association. The association with diabetes is characterized by a negative sign which lead us to the conclusion that people who suffer from this are more likely to have high body weight. The link between body weight and diabetes is a good example of inverse causality and that is the reason why we decide to use the word "association" – the fact of being overweight or obese may lead to the appearance of disease. So, we should say that, according to our analysis, the association between BMI and disease is negative. The high blood pleasure is characterized by a negative association with the dependent variable.

The eating habits category is represented by consumption wine. Other goods such as hydrates, vegetables, fruit (representing the food consumption) or whisky do not explain the model. It is important to remind that these variables are binary in which 1 represents the consumption of the good and 2 represents the absence of consumption. Wine, representing the consumption of alcoholic beverages, has a negative coefficient sign which makes sense, meaning that individuals who consume alcoholic beverages tend to have higher body height, as explained for the analysis of 1999.

This set of significant variables is composed by one more factor that wasn't discussed yet, the binary variable obesity ("obesidade"). It stands for previous cases of obesity in the family, including the respondent. Thus, the value 1 corresponds to individuals who have/had someone obese in the family or to individuals who had already been obese and the value 2 represents people who don't have any case of obesity. According to this model, genetics can be responsible for the amplification of excess body weight since the coefficient is negative. We conclude that individuals with cases of obesity in the family or who already suffered from being obese are more likely to present higher levels of body mass index.

Finally, it is important to refer that the model is relevant regarding the overall significant since the value of the F-statistics is 0 < 10% and we can prove it by performing the "test" command. Additionally, the model is able to explain approximately 13% of the variation of body weight.

Source	SS	df	MS
Model	1123.73182	8	140.466478
Residual	6732.93204	412	16.3420681
Total	7856.66386	420	18.7063425

Number of obs =	421
F(8, 412) =	8.60
Prob > F =	0.0000
R-squared =	0.1430
Adj R-squared =	0.1264
Root MSE =	4.0425

4.3. Analysis for 2012:

The appendix 10 represents the Stata output of the initial regression regarding the year of 2012 using the OLS method. In this case, instead of choosing the set of variables and perform an initial regression, we decided to add variables one by one until we reach the regression shown in the appendices section otherwise, we would have an huge amount of variables omitted by colinearity. Once again, we can see that we have plenty of variables that appear to be not statistically significant.

Although the initial process was quite different, the next step remains the same – dropping non statistical variables one by one in order to obtain a relevant set of variables with impact on the dependent variable. The table 3 shows the result of this process leading to our final regression for 2012:

Table 3 – final regression for 2012.

BMI	Coef.	t	P>|t|
yearschooling	-0.85143	-6.57	0
rendimento2	0.001088	4.87	0
_cons	33.8703	13.91	0

As described before, this database is limited in terms of observations and variables and that is the reason why its final regression is composed for just 2 variables - both variables used as proxies to SES, education and income.

The coefficient signs coincide with the expected ones for both variables. Regarding education, “yearschooling”, which stands for total years of schooling, has also the expected negative sign reinforcing that more educated people are associated with lower body weight. The other variable, “rendimento2”, presents the positive coefficient observed in the last two analysis (1999 and 2006) meaning that individuals in a better economic position have a higher purchasing power and so they are able to afford a privileged range of goods, making them more likely to present higher levels of body mass index.

It is important to underline that although this model is composed by 2 variables, they are able to explain approximately 78% of the variation of our dependent variable, a much higher percentage when compared with the above mentioned years, which are more complete in terms of significant variables. Its overall significance is reinforced by the “test” command since the value of F-statistics is zero.

Note that factors like the gender, diseases, if the respondent is a smoker or not and the quality of diet are considered to be not relevant, a not expected conclusion. Age is also a variable that doesn’t make part of the final regression but there is a reason for that to happen – as we mentioned in the data section, this information comes from a survey with a very limited number of answers and very restricted in terms of age groups. We are considering a range of ages that goes from 18 to 33 years old and so, it doesn’t even make sense to study the impact of age in this situation.

```
      Source |       SS       df       MS              Number of obs =     122
-------------+------------------------------           F(  2,   119) =  221.11
       Model |  2142.99076     2  1071.49538           Prob > F      =  0.0000
    Residual |  576.659838   119  4.84588099           R-squared     =  0.7880
-------------+------------------------------           Adj R-squared =  0.7844
       Total |   2719.6506   121  22.4764512           Root MSE      =  2.2013
```

4.4. Comparison between years:

After the above interpretation of each year, we must now compare values between 1999, 2006 and 2012.

The table below represent coefficients regarding age, gender, SES status, sports, genetics, if the individual is a smoker and the consumption of alcoholic beverages for all the three years. The lack of information means that the variable is not statistically significant for that year and the choice of variables is based on their importance.

Table 4 – coefficients of all significant variables for the three years

Variables	1999	2006	2012
Age	0.1590384	-1.30554	-
Gender	-2.000542	-	-
Years of schooling	-0.060848	-	-0.85143
Income	**0.0003147**	**0.0007461**	**0.001088**
Sports	0.04279	-	-
Smoker	-	-	-
Diseases	-0.29235(back pain)	-1.822(diabetes) -0.6703(high blood pressure) 1.715(osteoporosis)	-
Alcohol consumption	-	-0.9810937	-
Genetics	-	-5.400629	-
Number of observations	1616	421	122

According to this data, we are able to analyze the economic significance of results. For this purpose, we must focus just on income. We are considering an increase of 500€ in the households net income for the year of 2012. The first step is adjusting this value according to the annual inflation rate for the remaining two years. The Appendix 3 illustrates inflation rates, and the value of 500€ in each year according to the corresponding inflation. Considering all variables constant, an increase of 360€ in the households net income for the year of 1999 will be translated into a BMI increase of 0.113292. In the exact same conditions, an increase of 447€ in the household net income for the year of 2006 will lead to a 0.33351 increase of our index. Regarding the year of 2012, increasing the net income by 500€ will result in a BMI increase of 0.544. According to these results, the impact of income is increasing over time, reinforcing the importance of this variable.

Despite being composed only by both variables that represent SES, the final regression for the year of 2012 presents an Adjusted R-squared of 0.784, approximately.

This means that these variables are able to explain 78% of the variations of BMI. Note that the Adjusted R-squared for 1999 and 2006 is 20% and 13%, respectively. Discrepancies in the value of Adjusted R-squared, taking into consideration the number of explanatory variables, confirm the great contribution of income and years of schooling, although with few observations is easier to achieve higher values of R-squared. Still, discrepancies are very significant. Another important conclusion is that genetics does not explain our dependent variable. According to our analysis, in 1999 and 2012 this variable is not significant enough and the coefficient sign presented in 2006 does not make sense. Other variables related with consumption (food, alcohol, tobacco) had lost their impact over the years, and are not even significant nowadays.

5. Conclusions

In this work project we set out to study the impact of socioeconomic factors in the human body weight by performing regressions for the three different years. Our main concern was to understand if SES was a real contributor or if there were other variables able to better explain changes in our dependent variable. For that purpose, we choose a set of variables and by performing sequential regressions we ended up with a final one, composed just by statistically significant variables for each year. The result of this process is illustrated in the table of section 4.4. As we can see, the amount of significant variables tends to decrease over the years since that, the regression of 1999 is composed by five explanatory variables and for the last year of study we ended up with only two variables. Thus, we are dealing with a consequent reduction of significant variables that leads us to our main focus, SES. The analysis shows us that age and gender are losing their impact on BMI which means that some intuitions may be changing over time. Disregarding the impact of age, gender is not a primary variable to explain changes in body weight anymore.

As we explained before, years of schooling and income are the variables used to represent SES. We found that income is the more consistent variable and that its contribution is increasing over the years, representative of its real importance to explain variations in BMI. On the other hand, years of schooling didn’t seem to be a great contributor in the first two years but it became a very important variable for the last year. This phenomenon can be explained by the growing concern on welfare and aesthetics – leading to recent studies regarding eating habits, the importance of regular physical activity and other ways of prevent obesity and assure a good quality of life. Thus, the more educated people are the one who have easy access to this type of information, explaining why years of schooling became so important to explain variations in body weight.

In a more summarized approach, we can conclude that socioeconomic status is, with no doubt, a real contributor to explain changes in BMI, since it is represented by income, the most consistent variable over time, and it is also represented by years of schooling, a variable that became very relevant in the last year of analysis. Furthermore, our results confirm the expected relationship – income is positively correlated with BMI and years of schooling is negatively correlated with BMI.

Based on our analysis, we can conclude that variables related with SES are real contributors to explain changes in BMI. When compared with other mentioned variables related with health conditions, genetics and eating and drinking habits, income and years of schooling are definitely more significant in statistic terms, and allow a better explanation of variations in our dependent variable.

6. References

Angrist, Joshua D. and Jorn-Steffen Pischke (2008), “Mostly Harmless Econometrics: An Empiricist’s Companion”;

Baum, Charles L. and Christopher J. Ruhm (2007), “Age, Socioeconomic Status and Obesity Growth”;

Devaux, Marion, Franco Sassi, Jody Church, Michele Cecchini and Francesca Borgonovi (2009), “Education and Obesity in Four OECD Countries”, OECD Education Working Papers, No. 39, OECD Publishing;

Garcia, Jaume and Climent Quintana-Domeque (2005-2006), “Obesity, Employment and Wages in Europe”;

International Obesity Taskforce - www.iaso.org

Maitra, Professor Pushkar and Dr. Anurag Sharma (2007), “Does Exercise Reduce Obesity? Evidence from Australia”; Centre for Health Economics, Research paper 2007 (20);

Madden, David (2006), “Body Mass Index and the Measurement of Obesity”; Health, Econometrics and Data Group, working paper 06/11;

Madden, David (2010), “The socioeconomic Gradient of Obesity in Ireland”; UGD Geary Institute, Discussion Paper Series;

Porqueras, Pere Gomis and Adrian Peralta-Alva (2008), “A Macroeconomic Analysis of Obesity”; Federal Reserve Bank of St. Louis, Research Division, Working paper 2008-017A;

Sanz de Galdeano, Anna (2005), “The Obesity Epidemic in Europe”; Centre for Studies in Economics and Finance, Working paper no. 143;

Stamatakis, E. and J. Wardle and TJ Cole (2010), “Childhood obesity and overweight prevalence trends in England: evidence for growing socioeconomic disparities”; International Journal of Obesity, 2010;

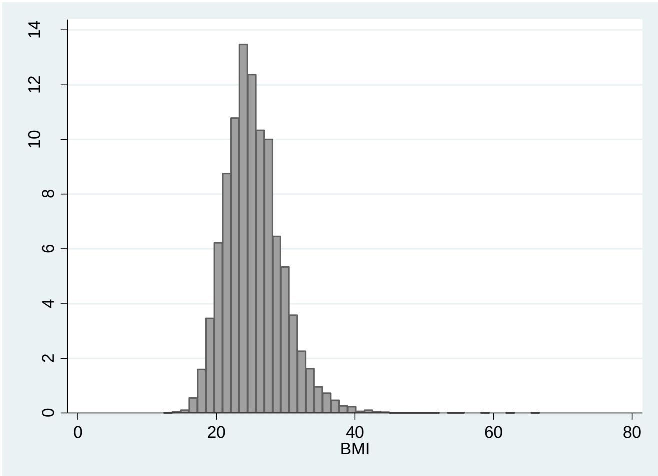
Wang, Youfa, Carlos Monteiro and Barry M. Popkin (2002), “Trends of obesity and underweight in older children and adolescents in the United States, Brazil, China and Russia”; The American Journal of Clinical Nutrition, 2002;

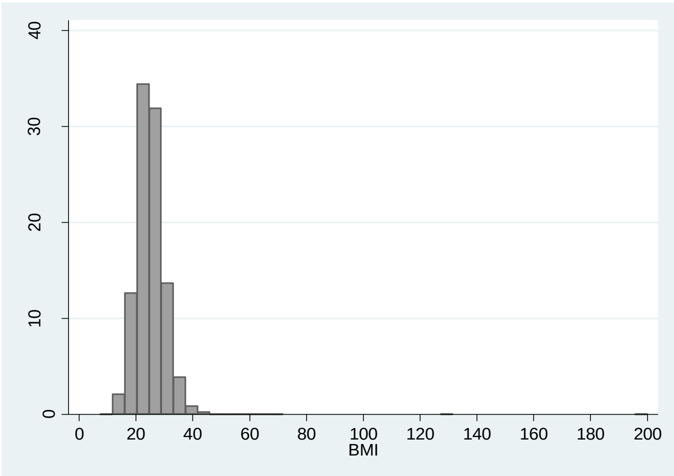
Wang, Youfa and Qi Zhang (2004), “Socioeconomic inequality of obesity in the United States: do gender, age and ethnicity matter? “, Social Science & Medicine, Vol 58(6);

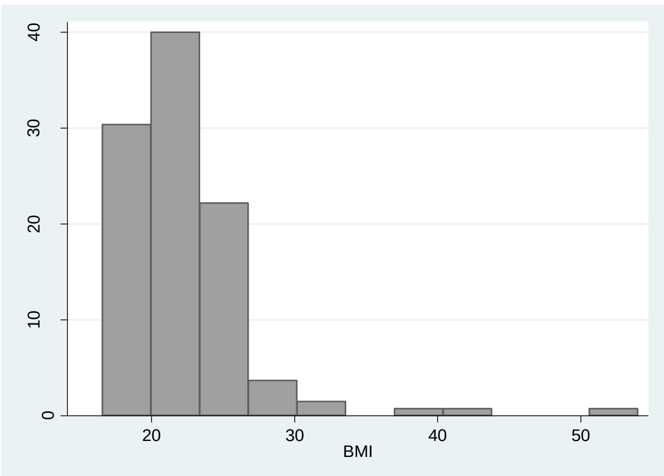
World Health Organisation - www.who.int

7. Appendices

Appendix 1 – histograms of the dependent variable, BMI, for 1999, 2006 and 2012.







Appendix 2 – Descriptive statistics of dependent variable for 1999, 2006 and 2012, respectively.

Variable	Obs	Mean	Std. Dev.	Min	Max
BMI	38688	25.44809	4.134426	12.48699	66.66666

Variable	Obs	Mean	Std. Dev.	Min	Max
BMI	38274	25.06706	4.944534	7.52757	200

Variable	Obs	Mean	Std. Dev.	Min	Max
BMI	135	22.51917	4.592427	16.56065	53.97924

Appendix 3 – Inflation rates and the correspondent value of 500€ (in 2012)

Years	Inflation rate	Correspondent value
1999	0.023	**360.4236066**
2000	0.029	370.8758912
2001	0.044	387.1944304
2002	0.036	401.1334299
2003	0.033	414.3708331
2004	0.024	424.3157331
2005	0.023	434.074995
2006	0.031	**447.5313198**
2007	0.025	458.7196028
2008	0.026	470.6463125
2009	-0.008	466.881142
2010	0.014	473.4174779
2011	0.0366	490.7445576
2012	0.01886	**500**

Sources/Entities: Pordata, INES

A Work Project, presented as part of the requirements for the Award of a Masters Degree in Economics from NOVA School of Business and Economics

Obesity and Socioeconomic Gradient

[Extra Appendices]

Constança Youd David

Student number - 527

A Project carried out with the supervision of

Professor Pedro Pita Barros

January 7th, 2013

Appendix 4 – The initial regression relative to the year of 1999

Source	SS	df	MS
Model	744.754331	43	17.3198682
Residual	680.963727	109	6.24737365
Total	1425.71806	152	9.37972407

Number of obs	=	153
F(43, 109)	=	2.77
Prob > F	=	0.0000
R-squared	=	0.5224
Adj R-squared	=	0.3339
Root MSE	=	2.4995

idade	0.124134	2.32	0.022
anosescola	-0.13175	-1.52	0.131
Estadosaude	-0.34378	-0.77	0.444
doente	0.479546	0.69	0.495
baixarse	-0.9559	-0.54	0.592
ouvir	-0.06276	-0.03	0.974
ver	0.853737	0.52	0.607
falar	5.044425	1.71	0.09
asma	0.33362	0.28	0.781
Alergia	-0.53467	-0.77	0.442
tençaoalta	-1.03285	-0.39	0.698
dorescostas	-1.08309	-2.08	0.04
nanalises	0.047326	0.08	0.933
nbiopsias	-1.52914	-0.44	0.658
nradiografias	-0.81033	-1.61	0.11
	0.630368	0.73	0.468
nmamografias	0.38711	0.13	0.898
nTAC	-1.54361	-0.41	0.684
nelectrocardio	2.282476	1.48	0.143
nendoscopias	-0.10599	-0.02	0.983
nfisioterapia	-0.0527	-1.05	0.297
comprimidosdormir	2.65037	0.91	0.363
rendimento2	0.000696	1.4	0.163
fuma	4.527148	2.54	0.012
ncigarrosdia	-0.04092	-1.31	0.193
nrefeiçoes	-1.19091	-1.37	0.172
comerforaref	-0.50547	-2.1	0.038
sopa	-0.44379	-0.9	0.37
peixe	0.301195	0.61	0.541
carne	-0.97236	-1.03	0.307
hidratos	0.687824	0.47	0.636
legumes	-0.90859	-1.49	0.139
fruta	0.184449	0.27	0.79
pao	0.968917	1.14	0.258
vinho	-0.52295	-0.79	0.43
cerveja	1.708553	2.24	0.027
bagaço	-0.07639	-0.11	0.912
martini	0.35742	0.67	0.506
whisky	-0.54773	-1	0.318
nbebidas	0.095799	0.37	0.709
acttemposlivres	0.05486	0.21	0.831
ndias	0.390706	2.6	0.011
_cons	12.6054	1.07	0.285

Appendix 5 – Descriptive summary of variables for the final regression of 1999

```
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------------------
BMI             float  %9.0g
sexo            byte   %8.0g                  SEXO
idade           int    %8.0g                  Idade
anosescola      byte   %10.0g                 Quantos anos de escolaridade completou com
                                                aproveitamento?
dorescostas     byte   %8.0g                  Costuma ter dores nas costas?
nradiografias   byte   %8.0g                  Quantas vezes fez radiografias?
rendimento2     float  %9.0g                  Rendimentos total ganho pela familia no mes passadc
comerforaref    byte   %8.0g                  Come fora das refeições? Quantas vezes?
carne           byte   %8.0g                  Comeu carne ontem?
legumes         byte   %8.0g                  Comeu legumes e hortaliças ontem?
ndias           byte   %10.0g                 Quantos dias por semana
```

Appendix 6 - Descriptive statistics for the significant variables of 1999

Variable	Mean	Std. Dev.	Min	Max
BMI	23.05552	2.936996	14.04082	38.10395
sexo	1.301361	0.458991	1	2
idade	23.63119	4.563293	18	33
anosescola	10.98205	3.391528	1	24
dorescostas	1.811881	0.390928	1	2
nradiografias	0.108911	0.398784	0	8
rendimento2	997.4093	565.8846	108.5	2131
comerforaref	1.309406	0.957065	0	6
carne	1.472153	0.499379	1	2
legumes	1.201733	0.401418	1	2
ndias	2.459158	1.585793	1	7

Appendix 7 – The initial regression relative to the year of 2006.

Source	SS	df	MS
Model	912.172277	36	25.3381188
Residual	3125.72408	190	16.4511794
Total	4037.89636	226	17.8667981

Number of obs	=	227
F(36, 190)	=	1.54
Prob > F	=	0.0347
R-squared	=	0.2259
Adj R-squared	=	0.0792
Root MSE	=	4.056

BMI	Coef.	t	P>|t|
sexo	0.178944	0.24	0.81
idade	-1.33089	-1.99	0.048
anosescola	0.060226	0.16	0.87
estadosaude	0.063751	0.14	0.888
limitadocasa	-0.15375	-0.13	0.895
distanciaandar	-0.23578	-0.39	0.694
apanhardochao	0.528216	0.9	0.372
deitarlevantar	-0.21631	-0.28	0.78
comer	1.324457	1.82	0.071
ver	-1.21052	-1.52	0.129
diabetes	-1.9969	-2.45	0.015
asma	0.07845	0.06	0.954
tensaoalta	-0.54185	-0.87	0.384
doençareumatica	0.12704	0.19	0.846
osteoporose	2.488467	2.8	0.006
cancro	-1.0264	-0.73	0.466
pedrasrins	0.890893	0.94	0.348
insufrenal	1.020347	0.59	0.554
ansiedade	0.445226	0.32	0.752
bronquitecronica	0.363268	0.31	0.761
avc	1.183509	0.95	0.345
obesidade	-5.02354	-3.33	0.001
depressao	-0.2558	-0.17	0.869
medicacaodormir	0.028643	0.04	0.97
rendimento2	-0.00039	-0.53	0.597
fuma	0.172882	0.18	0.856
nrefeiçoes	-0.9245	-0.84	0.401
laticinios	0.021159	0.03	0.977
sopa	-0.70104	-0.96	0.34
hidratos	0.978158	1.2	0.231
verduras	-0.06886	-0.11	0.911
vinho	-1.68562	-2.36	0.019
bagaço	1.121044	0.78	0.434
whisky	1.246257	0.98	0.328
martini	-0.72232	-0.75	0.454
cerveja	0.862158	0.84	0.402
_cons	53.6452	3.22	0.002

Appendix 8 – Descriptive summary of variables for the final regression of 2006

variable name	storage type	display format	value label	variable label
BMI	float	%9.0g		
idade	double	%10.0g		idade
comer	double	%10.0g		CONSEGUE COMER (CORTAR A COMIDA, LEVAR OS ALIMENTOS E BEBIDAS À BOCA) ?
diabetes	double	%10.0g		TEM OU JÁ TEVE DIABETES (AÇÚCAR NO SANGUE) ?
tensaoalta	double	%10.0g		TEM OU JÁ TEVE TENSÃO ARTERIAL ALTA (HIPERTENSÃO ARTERIAL) ?
osteoporose	double	%10.0g		OSTEOPOROSE, tem ou ja teve?
obesidade	double	%10.0g		OBESIDADE, tem ou ja teve?
rendimento2	float	%9.0g		
vinho	double	%10.0g		VINHO, tomou nos ultimos 12 meses?

Appendix 9 - Descriptive statistics for the significant variables of 2006

Variable	Mean	Std. Dev.	Min	Max
BMI	25.4406	4.325083	15.4315	37.77778
idade	18.3658	0.482226	18	33
comer	1.327791	0.630126	1	3
diabetes	1.84323	0.364016	1	2
tensaoalta	1.543943	0.498658	1	2
osteoporose	1.864608	0.342549	1	2
obesidade	1.969121	0.173195	1	2
rendimento2	630.7589	557.4571	75	2499
vinho	1.605701	0.489281	1	2

Appendix 10 - The initial regression relative to the year of 2012

Source	SS	df	MS
Model	973.470017	18	54.0816676
Residual	145.888253	74	1.97146288
Total	1119.35827	92	12.1669377

Number of obs =	93
F(18, 74) =	27.43
Prob > F =	0.0000
R-squared =	0.8697
Adj R-squared =	0.8380
Root MSE =	1.4041

BMI	Coef.	t	P>|t|
gender	-0.42599	-1.03	0.308
age	-0.0248	-0.41	0.682
yearschooling	-0.54898	-4.64	0
rendimento2	0.00111	5.46	0
timespentinpcperday	-0.11841	-0.71	0.477
diet	0.116141	0.57	0.569
dailymeals	-0.1679	-0.84	0.403
fastfood	0.123443	0.59	0.556
frequencyofpractice	0.19361	0.85	0.397
timespenteachtime	-0.2559	-0.81	0.419
physicalbarriers	0.496173	0.82	0.417
chronicanxiety	-0.76901	-1.57	0.121
stomachprob	-0.05362	-0.1	0.924
asma	-0.80938	-1.39	0.17
medication	0.570779	1.48	0.143
obesity	0.020006	0.04	0.97
smoker	-0.47206	-1.33	0.189
freqalcohol	0.077464	0.38	0.704
_cons	29.84407	9.53	0

Appendix 11 – Descriptive summary of variables for the final regression of 2012

variable name	storage type	display format	value label	variable label
BMI	float	%9.0g		
yearschooling	byte	%8.0g		How many years of complete schooling have you had since primary school?
rendimento2	float	%9.0g		Household's income net of taxes

Appendix 12 - Descriptive statistics for the significant variables of 2012

Variable	Mean	Std. Dev.	Min	Max
BMI	22.51917	4.592427	16.56065	53.97924
yearschooling	15.81481	2.868213	5	23
rendimento2	1997.951	1709.386	250	7000