Volume 1 Issue 1

Journal of
FOOD SCIENCE RESEARCH

Full Paper

JOFSR, 1(1), 2016 [039-048]

Regression model of the factors that influence weight of young adolescents

Raquel P.F.Guiné[1]*, Sofia R.Fernandes[2]
[1]Research Centre CI&DET, Polytechnic Institute of Viseu, Campus Politécnico, Viseu, (PORTUGAL)
[2]Food Engineering Department, Agrarian School of Viseu, Quinta da Alagoa, Viseu, (PORTUGAL)
E-mail: raquelguine@esav.ipv.pt

ABSTRACT

Excess of weight is a preoccupation for public health worldwide and is common among children and adolescents. This work intended to produce a model to predict weight of young people between 10 and 15 years depending on several tested independent variables: height, age, gender, physical activity, frequency of eating in Fast Food restaurants, the habits of eating at home, the eating of healthy or non-healthy foods. For this diverse tools for regression analysis were employed. The results allowed encountering a model that included four from the eight variables tested and excluded the other four, for not being explanatory of weight under the conditions tested. Finally, the regression model was subject to verification of the assumptions for application of this kind of model, and in light of the results obtained, it was concluded that the model can be applied as a predictive model for weight, for the population of young adolescents aged between 10 and 15 years.

KEYWORDS

Body weight;
Healthy food;
Fast food;
Physical activity;
Regression model;
Sport.

INTRODUCTION

Obesity is a major public health problem worldwide and is the most common nutritional problem among children and adolescents. The prevalence of overweight and obesity has increased dramatically in recent years, being quite preoccupying[1].

For children and adolescents, the limits of underweight, normal weight, overweight and obesity are dependent on the age and gender[2,3].

It has been postulated that changes in lifestyles and environmental conditions, such as electronic entertainment media and lack of access to public transport, have increased the level of risk for vulnerable young people[4]. Some primary prevention approaches include community-based interventions aimed at increasing the adoption of lifestyles to reduce weight, and more particularly on the importance of physical activity and care with diet, supplemented with contingency plans at the level of government schools and public health departments[4].

Many factors were found to contribute for the weight of the population in general, and adolescents in particular, and these include genetics, family environment, education, food habits, fast food, sports, physical activity, or sedentarism[5].

Strong relationships have been demonstrated between an excessive time devoted to screen sedentary activities (television, computer, console, etc ...) and the prevalence of overweight[6]. A study conducted in the

United States[7] concluded that children with a combination of high sedentary behavior and lower physical activity were twice as likely to be overweight than their counterparts, less sedentary and more active. In another study[8] was reported that higher television viewing time increased the likelihood of being overweight regardless of the adolescent being active or inactive, but that those who watched more television and exercised less were three times more likely to have excess weight than their counterparts, less sedentary and more active.

The objective of the present work was to derivate a regression model of the weight of young adolescents as a function of some sociodemographic variables, as well as some variables associated with eating habits and physical activity.

MATERIALS & METHODS

Data collection

This survey was undertaken by means of a questionnaire, which constitutes one of the privileged ways of collecting data refereeing to social behaviors. The questionnaire was developed for this study and previously submitted to the approval of the competent authority (The Service of Educational Projects of the Portuguese General Direction of Education) for application in school context.

The questionnaire was structured so as to collect data on sociodemographic variables; the practice of sports and physical activity; the level of knowledge relating to healthy eating; and finally the habits regarding eating and food.

Sampling and procedure

A multistage stratifed sampling method was used in the selection of schools. From all the public schools in Viseu, Portugal, four were randomly selected, and hence the collection of data was held in the schools Azeredo Perdigão, Grão Vasco, Mundão and Repeses. The sample was exclusively constituted by students attending the 2nd and 3rd cycle (5th to 9th grades) of basic education in Portugal. The sample was selected among all classes of the years considered and from all the schools included in the study. There was a random selection of classes in each school and the questionnaires were delivered to the corresponding directors of class so they could distribute them during their lessons. The participation of the respondents was voluntary and proceeded by written authorization from their parents, authorizing their children to participate in the survey. In the end, 720 consented valid questionnaires were obtained for young adolescents aged between 10 and 15 years old.

Sample characterization

The sample was constituted by 720 students, from which 49.4% were male and 50.6% were female, thus representing evenly both sexes.

The ages of the enquired varied from 10 to 15 years, with a mean of 12.43±1.52 years. The age distribution can be seen in Figure 1.

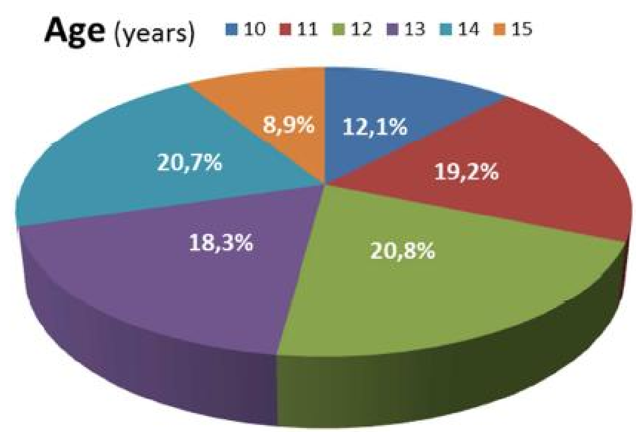

Figure 1 : Age distribution of the enquired

The students attended the 2nd and 3rd cycles of the Basic Education in Portugal, being 20.7% in the 5th grade, 23.3% in the 6th grade, 21.5 in the 7th grade, 16.1 in the 8th grade and 18.3 in the 9th grade.

Regression modeling

The objective of this study was to develop a multiple linear regression model that relates weight with some of the variables measured in the study. Some questions allowed in the survey allowed obtaining information such as height, gender or age, which is intended to incorporate into the model.

Other factors that are intended to try incorporating in the model are associated with practices related to physical exercise, leisure activities and free time occupation habits, including physical inactivity, or feeding practices. Information about sports practice was collected in the questionnaire through questions such as: Are you federated in any sport?, Do you practice school sports? or Do you practice high competition sports? (All of these should be answered yes or no). Information concerning the leisure activities and free time occupation habits was obtained through question: What of the following activities did you practice yesterday? (with multiple answer options: Playing outdoor, Internet, Watch TV, Play computer, Playing sports, Reading books). Information relating to habits of eating out was obtained by question: How many times did you go to the following convenience restaurants during the last week? (with multiple answer options: Mcdonalds, Pizza Hut, Vitamins, Telapizza, H3, Subway, Pans & Company, Piantella, House of Soups, Fish Times). The information about the frequency of eating at home was obtained through questions about frequency of eating at home the following meals: breakfast, lunch or dinner.

Also the habits of consumption of certain foods were evaluated through questions about the types of food eaten according to categories: healthy foods included milk, yogurt, bread, cereals, soup, vegetables and salads, fruits and natural fruit juices, fish or meat, non-healthy foods included chocolate, chips, carbonated sweetened drinks, sweets, cookies and sweet deserts.

The variables considered potentially explanatory to try to incorporate into the regression model to describe the weight of pre-teens and young teenagers were as follows:

- H : height (quantitative variable, expressed in meters),
- A : age (quantitative variable),
- G : gender (dichotomous qualitative variable: male = 0 female = 1),
- SPA : sports and physical activity (whether school, federated or high competition sports and other physical activities) (dichotomous qualitative variable: if not practicing any kind of sport or physical activity = 0, if practicing some kind of sport or physical activity = 1),
- EO : frequency of eating out in fast food restaurants (quantitative variable resulting from the sum of the number of times that the adolescents went to fast food restaurants to eat pizza, hamburgers and similar types of meals),
- EH : frequency of eating at home (dichotomous qualitative variable: if usually eat at home = 0, if usually eat out = 1),
- HF : eat healthy foods (quantitative variable resulting from the sum of the number of times that the adolescents eat foods considered healthy),
- NHF : eat non-healthy foods (quantitative variable resulting from the sum of the number of times that the adolescents eat foods considered non-healthy).

Hence the regression model to study could be written as:

$$W = f(H, A, G, SPA, EO, EH, HF, NHF) \quad (1)$$

where W is the body weight, in kg.

Development methodology

Previously to the model development, single linear regression models were tested between the dependent variable (weight) and each of the independent variables considered, independently, to verify if the variable alone would seem to influence the weight or not.

In the first stage a simple linear regression equation was considered to evaluate the relation between the weight and the 1st variable that was considered to be potentially explanatory, the age. Then each of the other variables was successively incorporated into the model, in order to verify if it would add explanatory power to the model. It was also used the partial F test to test the introduction of a set of variables that constituted a complete model to compare with the reduced model developed up until that point of the work.

The overall significance of each model was evaluated using the F-test and the significance of each of the estimated coefficients for the different independent variables was evaluated by the t-test.

The goodness of the fit was evaluated by the value of the coefficient of determination (R^2) when it came to simple linear regression and the value of the adjusted coefficient of determination in cases of multiple regression.

In all tests a level of significance of 5% was considered, unless when specified a different significance level, and software SPSS version 22 was used to treat the data.

RESULTS & DISCUSSION

Previous analysis of the relations between the weight and each of the considered variables

Before starting to build the multiple linear regression

Full Paper

model a preliminary analysis of the potential links between the independent variables and the dependent variable weight was undertaken, using for such models of simple linear regression the general equation as described by:

$$\mathbf{Y} = b_0 + b_1\mathbf{X} \quad (2)$$

where: Y = dependent variable (weight), X = independent variable, b_0 = constant (intercept of the straight line), b_1 = coefficient of the variable (slope of the straight line).

The different variables considered here were those included in Equation (1): H (height), A (age), G (gender), SPA (sports and physical activity), EO (frequency of eating out in fast food restaurants), EH (frequency of eating at home), HF (eat healthy foods, NHF (eat non-healthy foods).

TABLE I presents briefly the results for the different linear regressions, only to allow observe the kind of influence that each of the variables considered would have by itself on the dependent variable weight. Note that there is not a comprehensive discussion of these results, since they were only designed to guide the following work. The results in TABLE 1 show that while the variables weight and height have some explanatory power on weight, the rest have a very limited explanatory power, attending to the values of the coefficient of determination which is very low. Still, the regression is significant in most whole of the cases (except for variables EO, EH and HF) given the significance of the F test. For this test to the significance of the regression, when p-value <0.05 the null hypothesis

TABLE 1 : Principal results of the independent simple linear regression models

Linear regression	R^2	F *Sig.*	Constant (b_0) *Sig.*	Constant (b_1) *Sig.*
W = *f*(H)	0.557	901.338 0.000	-74.966 0.000	0.782 0.000
W = *f*(A)	0.375	431.700 0.000	-3.391 0.001	20.777 0,000
W = *f*(G)	0.014	10.428 0.001	81.931 0.000	-3.229 0.001
W = *f*(SPA)	0.024	17.732 0.000	50.257 0.000	-1.367 0.008
W = *f*(EO)	0,003	2.167 0.141	48.041 0.000	-0.222 0.141
W = *f*(EH)	0.005	3.286 0.070	46.440 0.000	1.761 0.070
W = *f*(HF)	0.003	2.508 0.114	49.773 0.000	-0.381 0.114
W = *f*(NHF)	0.005	3.843 0.050	47.260 0.000	0.966 0.050

(H0: $\beta_1 = \beta_2 = \beta_3 = 0$, the regression is not significant) is rejected and in that case a linear relation exists between the independent variable and the dependent variable considered. Thus, it can be observed that for variables EO, EH and HF the p-value is greater than 0.05, hence not rejecting H0, i.e., there is no evidence that these variables influence weight. As regards all other linear regressions they have a significant F, also being significant the coefficients associated with the independent variable (b_1) in each case.

Linear regression model between weight and height

The linear regression model between weight and height can be expressed by the following equation:

$$\mathbf{W} = -74.966 + 0.782\ \mathbf{H} \quad (3)$$

This model estimates that, on average, for each centimetre more in height the weight increases by 0.782 kg. The overall significance of the regression model was evaluated by the F test, which tests the null hypothesis of no relationship between the independent variable and the dependent. The F statistic is the ratio of the variance explained by the model to the variance that was explained (F = SSR/SSE), indicating that when F is high the proportion of the variance explained by the model is significantly greater than the variation not explained, and therefore leads to the rejection of the null hypothesis. In this case the F statistic is 901.338 and has a significant p-value lower than 0.001. Thus, at the level of significance of 5%, because $p < 0.05$ H0 is rejected, and hence it is concluded that height influences weight. To test the significance of the coefficient that multiplies the independent variable the t-test was used. This tests the null hypothesis (H0: $\beta_i = 0$) that the independent variable is not explanatory of the dependent variable. This procedure, in the present case of simple linear regression, only serves to confirm the conclusion of the F test, as for simple regression the results of F-test and t-test coincide. From TABLE I both the constant and the coefficient of the independent variable height have significant p-values at a significance level of 5%, so the variable height is significant for the model, confirming what was previously observed.

The value of the correlation coefficient in this case was 0.746, which is indicative of a strong association between these two variables. The correlation coefficient measures the association between two variables according to the magnitude of the absolute value [9–11]. If $0.00 < |R| < 0.10$ the association is very week, if $0.10 \leq |R| < 0.30$ the association is week, if $0.30 \leq |R| < 0.50$ the association is moderate, if $0.50 \leq |R| < 0.70$ the association is strong and if $0.70 \leq |R| < 1.00$ the association is very strong. For R = 0 there in no association and for R = 1 or R = -1 the association is perfect. When the value of R is negative it indicates an inverse correlation between the two variables.

The determination coefficient (R^2), when expressed in percentage, gives the explanatory capacity of the model, which in this case is 55.7%, thus explaining more than half of the total variance of the dependent variable (weight), so that the remaining 44.3% would be explained by other factors that not the height.

Multiple linear regression model between weight and height, age and gender

The previous model was extended including more independent variables, namely age and gender, in order to verify if the model became more robust with the introduction of these new variables. The results show that the adjusted determination coefficient increased from 0.556 to 0.583, indicating that the introduction of the new variables increases just slightly the robustness of the model, now explaining 58.3% of the variance in the dependent variable weight.

Testing the overall significance of the model revealed that in this case the regression is significant at a significance level of 5%, since the p-value of the F test is less than 0.05 (F = 335.498, p-value < 0.000), leading to the rejection of the null hypothesis that there is no relationship between the independent variables and the dependent variable.

TABLE 2 shows the results obtained for the model coefficients and allows to verify that the variable gender does not have explanatory power over weight, contrarily to the variables height and age. This variable was significant when a simple linear regression was made between weight and gender alone (results in TABLE I), but when the model already incorporates the variables height and age, it becomes redundant. In fact, the p-value for variable gender is not significant at a significance level of 5% (p-value = 0.075), but could be significant at the 10% significance level, and so, given the relevance of the variable in question the option is to keep it in the model, that in this case is translated by equation (4). This option is justified because it is known that for children and adolescents, percentile tables are determined by gender and age[2,3], and hence it is of interest to keep in the model the variable gender.

$$W = -70.193 + 0.623\,H + 1.655\,A - 0.993\,G \qquad \textbf{(4)}$$

The observation of the standardized coefficients beta (TABLE II) also shows that the value for the variable gender is very low (beta = -0,044), confirming the relative unimportance when compared to the other variables, and particularly age (beta = 0.595).

TABLE 2 : Coefficients of the multiple linear regression model between weight and height, age and gender

Model	Unstandardized Coefficients	Standardized Coefficients	t Statistic	Sig.
	B	Beta		
(constant)	-70.193		-17.150	0.000
H	0.623	0.595	18.288	0.000
A	1.655	0.221	6.838	0.000
G	-0.993	- 0.044	-1.783	0.075

Because gender is a dichotomous variable assuming the values 0 and 1 for genders male and female, respectively, the model from equation (4) can be decomposed in two, one for each gender according to equations (5) and (6):

$$W = -70.193 + 0.623\,H + 1.655\,A \;\; \text{(for male)} \qquad \textbf{(5)}$$

$$W = -71.186 + 0.623\,H + 1.655\,A \;\; \text{(for female)} \qquad \textbf{(6)}$$

Complete multiple linear regression model for weight

At this stage a multiple linear regression model was obtained that already includes three explanatory variables (height, age and gender), and which reflect intrinsic characteristics of the young people surveyed. However, it is intended also to incorporate into the model five more independent variables related to practices and behaviors of young people who might influence their body weight, which are:

- SPA: sports and physical activity (whether school, federated or high competition sports and other physical activities),
- EO: frequency of eating out in fast food restaurants,
- EH: frequency of eating at home,
- HF: eat healthy foods,
- NHF: eat non-healthy foods.

Note that, although in the preliminary analysis it was observed that the variables EO, EH and HF appeared to have no explanatory capacity towards the independent variable weight, yet they were included in the analysis to confirm if this assumption was definite. The fundamentals for the decision of not immediately abandoning these variable based only on the results of the simple regression analysis (TABLE I) consisted on the fact that there are many references in the literature about the influence of these types of behavioural factors in body weight and consequently in corporal body mass index (BMI)[12–16].

Considering a more efficient methodology in alternative to the successive addition of each of the independent variables, at this stage the partial F-test was used to compare the full model (comprising eight independent variables) with the reduced model (comprising only three independent variables). To do this, was used the stepwise regression method. The partial F-test tests the null hypothesis that the reduced model is appropriate (H0: $\beta_{r+1} = \ldots = \beta_k = 0$; where $\beta_{r+1} \ldots \beta_k$ are the coefficients of the variables added to the complete model and not included in the reduced model). By rejecting H0, there is evidence that among the $x_{r+1} \ldots x_k$ variables at least one it useful in the model.

The results show that the final model (model 2) incorporates just one from the five variables that were added to the test block in the complete model, which is the variable associated with sports and physical activity (SPA), as can be seen in TABLE 3.

Thus, the final solution (model 2) excludes the variables EO (frequency of eating out in fast food restaurants), EH (frequency of eating at home), HF (eat healthy foods) and NHF (eat non-healthy foods), as indicated in TABLE IV. These variables have p-values of the t test considerably high (0.601, 0.349, 0.903, 0.330, respectively for EO, EH, HF and NHF) (TABLE 4), leading in all cases to not rejecting the null hypothesis (H0: the independent variable has no influence on the dependent variable, weight).

TABLE 3 : Variables added to the complete model

Model	Variables entered	Variables removed	Method
1	H, A, G		Enter
2	SPA		Stepwise (Criteria: Probability of F to enter ≤ 0.05; Probability of F to remove ≥ 0.100)

TABLE 4 : Variables excluded from the complete model

Model	Var.	Beta In	t	Sig.	PC[a]	CS[b] Tolerance
1	SPA	-0.064	-2.595	0.010	-0.097	0.933
	EO	-0.011	-0.448	0.654	-0.017	0.987
	EH	0.019	0.763	0.446	0.029	0.980
	HF	0.000	0.009	0.993	0.000	0.986
	NHF	-0.038	-1.575	0.116	-0.059	0.996
2	EO	-0.013	-0.523	0.601	-0.020	0.987
	EH	0.023	0.937	0.349	0.035	0.976
	HF	0.003	0.121	0.903	0.005	0.984
	NHF	-0.024	-0.975	0.330	-0.036	0.937

[a]PC = Partial correlation; [b]CS = Collinearity statistics

These results are somewhat expected considering what had been observed during the preliminary analysis by simple linear regression between these variables and weight. In fact the preliminary results had indicated that the variables EO, EH and HF had no influence on the weight of the young adolescents, and for variable NHF it was on the limit of significance (p-value = 0.05), but this was for a model with only one explanatory variable and in the case of this complete model this variable no longer had significance. These results are contradictory to some extent with some information in the scientific literature[12–15], but may be due to the fact that these students reside in Viseu, a city where there are still healthily eating habits and hence this was not a differentiator factor for body weight. Effectively in the sample under study was observed that 56% do not eat at Fast Food Restaurants even once a week, 24% go only once a week and only 20% go more than once a week, indicating that these restaurants are still a rarity in the feeding habits of these young people. With regard to activities in leisure time, only 27% practice exclusively sedentary activities compared to 72% who have other activities in their spare time, such as playing outside and playing sports. These young people are still mostly practicing sports (57% practice school, federated or high competition sports).

TABLE 5 shows the coefficients of the variables that were considered to build the complete model. The results show that all variables have a significant p-value in the t test at the 5% significance level, leading to the rejection of the null hypothesis that the variables are not explanatory. Hence all variables showed to have influence on the dependent variable weight, including the variable gender, which at a certain point was questioned if it should or not be included in the model. On the other hand, the analysis of the standardized coefficients beta reveals that the variables that have a highest influence on the weight are height (coeff. beta of H = 0.594) followed by age (coeff. beta of A = 0.217), while variables gender and physical activity have a minor influence on weight (coef. beta of G and SPA are -0.056 and -0.064, respectively).

With the values of the coefficients from TABLE V, one can write the final model equation:

$$W = -67.804 + 0.622\,H + 1.573\,A - 1.265\,G - 0.568\,SPA \quad (7)$$

TABLE 5 : Coefficients of the variables of the complete model

Model Var.	Unstandardized Coefficients	Standardized Coefficients	t Statistic	Sig.
	B	Beta		
(constant)	-67.804		-16.224	0.000
H	0.622	0.594	18.324	0.000
A	1.573	0.210	6.471	0.000
G	-1.265	-0.056	-2.241	0.025
SPA	-0.568	-0.064	-2.595	0.010

The results also showed that the p-value of the partial F-test (0.015) is less than alpha (0.05) leading to rejection of the null hypothesis at a significance level of 5%. Thus, the introduced variable SPA showed some explanatory capacity on the dependent variable weight. Since in this case the complete model differs from the reduced model on only one variable, it is verified that the partial F-test is an equivalent to the T-test for this inserted variable, i.e., the value of the statistic of the partial F-test (6.736) is equal to the square of the statistic of the t-test for the coefficient of that variable ($-2.595^2 \cong 6.766$) and the p-value is also equal (0.010).

Also the results of the F-test to the final model (model 2) shows that the model is significant at a significance level of 5% because the value of F is high (255.323) and has a p-value (<0.001) less than 5%. When the statistic of the F-test is high it means that the ratio of the variation in weight explained by the model is considerably larger than the variation that is not explained. Since p<0.05 the null hypothesis of the independent variables not having influence on the variable weight is rejected.

As regards the explanatory capacity of the model, it was observed that the determination coefficient increased in the complete model, although very slightly, compared to the reduced model, so that this model (corresponding to equation (7)) can explain 58.5% of the variance in the dependent variable weight.

Validation of assumptions

For applying linear regression models the following assumptions should be verified:

- The errors must obey the normal distribution;
- The errors must be independent, have zero mean and constant variance;
- The independent variables are not correlated, i.e. there is no multicollinearity.

Once regression was performed using the method of least squares, the zero-mean of the errors is automatically ensured.

Normality of the errors

To test the normality of the errors the standardized residuals were stored in the database to test. Figure 2 shows that the dots are more or less aligned with the diagonal but with some deviation, which is an indicator that maybe the normality of the distribution of the errors might be violated.

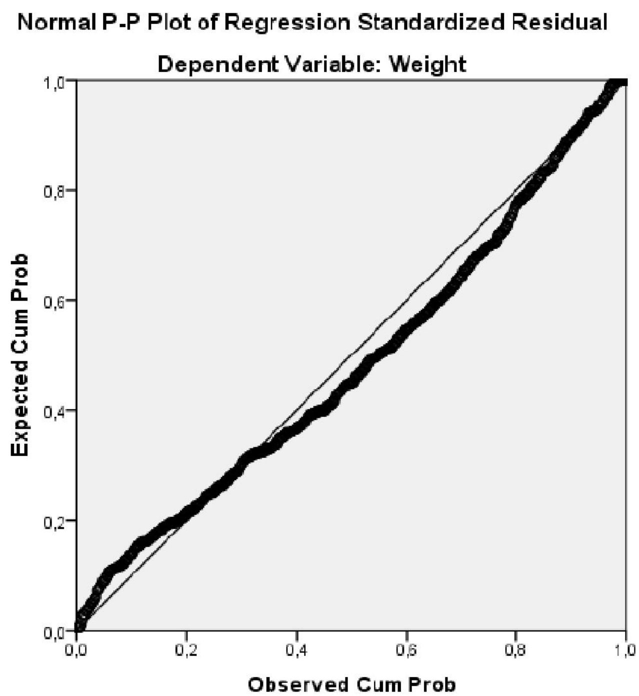

Figure 2 : Normal probability plot

For confirmation the normality of the residuals was tested by the Shapiro-Wilk normality test, for being the most adequate when the sample sized exceeds 30, and which results are presented in TABLE 6. The result of the test to normal distribution of the errors indicates that the p-value is significant, i.e. $p < 0.05$, which determines the rejection of the null hypothesis (H0: The distribution is normal), and therefore the distribution of the errors does not obey normality at a level of significance of 5%. Nevertheless it is important notice that this assumption is important but not critical, especially in the case of large samples, as it is the case in the present study, where N = 720.

TABLE 6: Normality test to the errors

	Shapiro-Wilk		
	Statistic	**df**	**Sig.**
Standardized residuals	0.963	720	0.000

Homocedasticity and independence of the errors

Figure 3 shows the graph of the residuals as a function of the predicted values, both standardized, in order to eliminate the effects of the magnitude of the measurement scale for the errors. The observation of the points suggests that the errors are distributed randomly around the y=0 line, with maybe just some outliers, thus indicating that the variance is constant.

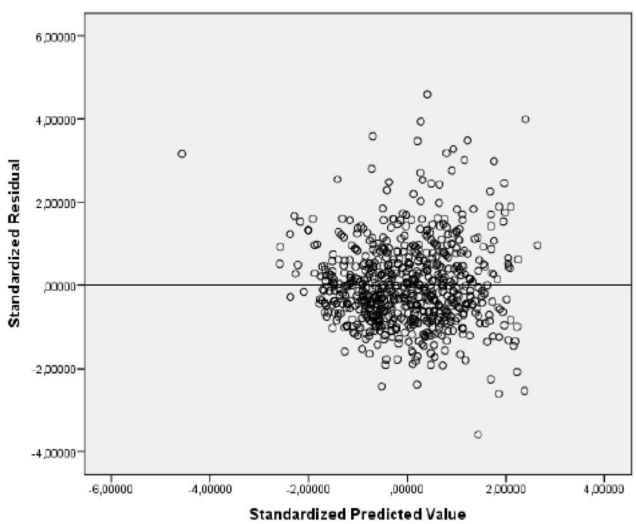

Figure 3 : Residuals distribution

For confirmation, verification of the independence of errors was made by the Durbin-Watson test, whose results are shown in TABLE 7.

TABLE 7 : Durbin-watson test to the independence of the errors

R	R^2	Adjusted R^2	Standard error of estimate	Durbin-Watson
0.767	0.588	0.586	7.328	2.066

The value of the observed statistic for the Durbin-Watson test (d) was 2.066. In order to make any conclusion there must be determined for the present case study the values DU and DL, which allow to set the decision intervals of Durbin-Watson table (TABLE 8). In this case n = 720 observations and k = 4 independent variables (Height, Age, Gender, and Physical activity). The critical limits were found in Durbin Watson tables for these n and k values, although the exact values for this sample size in particular were not found. However, it was possible to find tables for large samples, in particular for N = 700 and k = 5 (because in this case the tables explicitly indicated that the value of k included the independent term (intercept)), whereby these figures are used as an approximation. Thus, the values of critical limits are: DL = 1.864 and DU = 1.887, from the Durbin-Watson tables for k = 5 and n = 700 for a significance level of 5% [17]. From these critical limits the decision intervals were calculated and are shown in TABLE VIII. Since the observed value of the test statistic is in the range [DU, 4-DU[, the condition of independence of the errors is verified. Note that this condition is essential for validation of the regression model.

TABLE 8 : Durbin-watson decision intervals

Interval	d = 2.066	Decision[a]
[0 , DL]	[0 , 1.864]	Reject H0 / Dependence of errors
[DL , DU]	[1.864 , 1.887]	Inconclusive
[DU , 4-DU]	[1.887 , 2.113]	Not reject H0 / Independence
[4-DU , 4-DL]	[2.113 , 2.136]	Inconclusive
[4-DL , 4]	[2.136 , 4]	Reject H0 / Dependence of errors

[A]**H0: There is no correlation between successive residuals**

Multicollinearity

For the analysis of multicollinearity were used for each variable the values of tolerance and VIF (variance inflation factor) shown in TABLE 9. For absence of collinearity the values of tolerance should be greater than 0.1 and the values of VIF should be less than 10, correspondingly, since VIF is the inverse of tolerance. From observing the values in TABLE 9 it is seen that in all cases the tolerances are greater than 0.1 and the VIF are lower than 10, so there is indication of absence of multicollinearity. Thus, with respect to this assumption it is not violated, i.e., it is accepted that the variables are not correlated.

TABLE 9 : Ealuation of multicollinearity

Variables	t	Sig.	Collinearity Statistics	
			Tolerance	VIF
H	18,324	0,000	0,548	1,824
A	6,471	0,000	0,548	1,824
G	-2,241	0,025	0,937	1,068
SPA	-2,595	0,010	0,933	1,071

For confirmation it was further determined the correlation matrix between the independent variables in the model (TABLE 10). The results show that in all cases the values are lower than 0.7, the strongest correlation was found between the variables height and age, with a value of $R = 0.659$, indicating that no strong correlations exist between the variables.

TABLE 10 : Pearson correlations between the independent variables in the regression model

	H	A	G	SPA
H	1			
A	0.659**	1		
G	-0.129**	0.002	1	
SPA	-0.106**	-0.182**	-0.183**	1

**Corelation is significant at the 0.01 level

CONCLUSION

This work allowed building by successive steps a multiple linear regression model to describe the variation in the weight of young people between 10 and 15 years depending on the explanatory variables H (height), A (age), G (gender), and SPA (sports and physical activity). It was also found that the introduction of other variables which initially were thought to have some explanatory power over weight, such as the frequency of eating in Fast Food restaurants (EO), the habits of eating at home (EH), the eating of healthy (HF) or non-healthy foods (NHF), revealed after all not having explanatory power in the model, meaning that these variables are not important when the model already incorporated the other four.

After finding the regression model this was subject to verification of the assumptions, having been observed that there was a deviation with respect to normality, which however would not be critical taking into account the sample size in the study (720 cases). With regards to the multicollinearity this did not occur, so the variables were uncorrelated with each other. As regards the homoscedasticity, it was found that the variance of the errors was approximately constant and these were independent, as demonstrated by the result of Durbin-Watson test.

In light of these results, and because the most important principles for the application of regression were assured, the model that was found can be applied as a predictive model for weight depending on the variables considered, for the population of young adolescents aged 10 and 15 years.

REFERENCES

[1] F.C.Woon, Y.S.Chin, M.T.Mohd Nasir; Obes.Res.Clin.Pract. (in press).

[2] R.J.Kuczmarski, C.L.Ogden, S.S.Guo, L.M.Grummer-Strawn, K.M.Flegal, Z.Mei, R.Wei, L.R.Curtin, A.F.Roche, C.L.Johnson; '2000 CDC Growth Charts for the United States' CDC - Department of Health and Human Studies, Washington DC., **(2002)**.

[3] M.Onis, A W Onyango, E Borghi, A.Siyam, C Nishida, J.Siekmann; Bull.WorldHealthOrgan., **85**, 660 **(2007)**.

[4] G.Slaney, J.Salmon, P.Weinstein; Med.Hypotheses, **79**, 68 **(2012)**.

[5] K.Campbell, N.Andrianopoulos, K.Hesketh, K.Ball, D.Crawford, L.Brennan, N.Corsini, A.Timperio, Appetite, **55**, 84 **(2010)**.

[6] S.Allender, P.Kremer, A.Silva-Sanigorski, K.Lacy, L.Millar, L.Mathews, M.Malakellis, B.Swinburn; J.Sci.Med. Sport, **14**, 512 **(2011)**.

[7] S.B.Sisson, S.T.Broyles, B.L.Baker, P.T.Katzmarzyk; J.Adolesc.Health, **47**, 309 **(2010)**.

[8] J.C.Eisenmann, R.T.Bartee, D.T.Smith, G.J.Welk, Q.Fu; Int.J.Obes., **32**, 613 **(2008)**.

[9] J.Maroco; Análise Estatística Com O SPSS Statistics, 5th (Ed.), Report number; Brazil **(2012)**.

[10] M.H.Pestana, J.N.Gageiro; 'Análise de Dados Para Ciências Sociais – A Complementaridade Do SPSS', 6th Ed., Edições Sílabo; Brazil, **(2014)**.

[11] J.P.Stevens; Applied Multivariate Statistics for the Social Sciences, 5th (Ed.), Routledge; New York, **(2009)**.

[12] P.A.Alviola, R.M.Nayga, M.R.Thomsen, D.Danforth, J.Smartt; Econ.Hum.Biol., **12**, 110 **(2014)**.

[13] A.C.McClure, S.E.Tanski, D.Gilbert-Diamond,

A.M.Adachi-Mejia, Z.Li, J.D.Sargent; Am.J.Prev.Med., **45**, 560 **(2013)**.
[14] C.L.Newman, E.Howlett, S.Burton; J.Bus.Res., **67**, 1573 **(2014)**.
[15] B.Mandal, L.M.Powell; Econ.Hum.Biol., **14**, 50 **(2014)**.
[16] K.A.McGeary; Soc.Sci.Med., **82**, 67 **(2013)**.
[17] N.E.Savin, K.J.White; Econometrica, **45**, 1989 **(1977)**.