Childhood obesity and ADHD: Logistic Regression Model

Raquel Maria Jacinto Escola

December 13, 2009

Abstract

In recent years, ADHD (Attention Deficit and Hyperactivity Disorder) has been considered as one of the most common psychiatric disorders in school age children. At the same time, the number of overweight children has been increasing. Therefore, it is important to identify groups of children who are at increased risk of overweight. Thus, the aim of this study is to find relationships between obesity and ADHD in school-age children in a certain urban area. For instance, we intend to evaluate the existence of lifestyle factors related to both disorders (sleep, television, sport).

For that purpose we applied logistic regression analysis, both classical (dichotomous and polytomous) and robust. As a result of these analysis, we found the expected relationship between obesity and nutritional habits. Also we have found that children with ADHD are at increased risk for being overweight, when we included this variable into a binary (obese, not obese) logistic model.

Keywords: Dichotomous Logistic Regression Model, AIC and BIC criteria, Odds Ratio, Robust Logistic Regression Model, Bootstrap Test, Quasi-deviance Test, Polytomous Logistic Regression Model.

1 Introduction

This study aims to identify obesity risk groups and seeks to find possible links between obesity and Attention Deficit and Hyperactivity Disorder (ADHD). Moreover, the current study intends to evaluate the existence of associated factors to both disorders related to lifestyle, as sleep habits or sport activity. Thus, this study seeks to find a model to establish an association between these habits, as well as the ADHD, and the obesity.

This work was made under a partnership between the Pediatrics of Hospital D. Estefânia (Development Center of Hospital D. Estefânia) and the Probability and Statistics Section of the Instituto Superior Técnico. The present data were collected in 2008/2009, to parents and teachers of basic schools in an urban area. For this study, only children aged between 6 and 8 years (without any cognitive disorder or identified genetic disease) were considered, in a total of 403 children.

1.1 Childhood Obesity

Obesity is the most prevalent nutritional disease in the world and is considered the epidemic of the century. In recent years, childhood obesity has increased. In Portugal, more than 30% of children aged between 7 and 9 years are overweight and about 11% are obese ([9]).

Usually, the screening of obesity is made according to the Body Mass Index (BMI), that is given by:

$$\mathbf{BMI} = \frac{Weight}{Height^2}, \tag{1}$$

with weight in kilograms and Height in meters.

In order to get a BMI classification (as underweight, normal weight, overweight and obese) we considered the percentile curves calculated by *National Centre for Health and Statistics* (NCHS). `http://www.cdc.gov`

After computing the BMI for each child and using the percentile curves, the classification in Table 1 is considered.

Table 1: BMI classification in four categories.

Classification	**BMI**
Underweight	Smaller than $5th$ percentile
Normal Weight	Between $5th$ and $85th$ percentile
Overweight	Between $85th$ and $95th$ percentile
Obese	Greater than $95th$ percentile

Then Table 2 gives the BMI classification by gender and age.

Table 2: BMI classification by gender and age.

Age	**Girls BMI**	**Boys BMI**	**Classification**
6 years old	$BMI < 13,4$	$BMI < 13,7$	
7 years old	$BMI < 13,4$	$BMI < 13,7$	Underweight
8 years old	$BMI < 13,5$	$BMI < 13,8$	
6 years old	$BMI < 17,2$	$BMI < 17,2$	
7 years old	$BMI < 17,6$	$BMI < 17,4$	Normal Weight
8 years old	$BMI < 18,3$	$BMI < 17,9$	
6 years old	$BMI < 18,8$	$BMI < 18,4$	
7 years old	$BMI < 19,6$	$BMI < 19,2$	Overweight
8 years old	$BMI < 20,7$	$BMI < 20,3$	
6 years old	$BMI > 18,8$	$BMI > 18,4$	
7 years old	$BMI > 19,6$	$BMI > 19,2$	Obese
8 years old	$BMI > 20,7$	$BMI > 20,3$	

According to the research, the most important causes of childhood obesity are the eating behaviours (fast food and overfeeding), the lack of sport (lifestyle) and genetic predisposition (5 to 25% of cases).

1.2 Childhood ADHD

The ADHD or Attention Deficit and Hyperactivity Disorder is a disorder characterised by hyperactive behaviour, attention deficit and impulsivity. This disorder is more common in boys than in girls and affects 3% to 6% of school-age children [8].

There are several types of ADHD, depending of the child symptoms. Some children are predominantly hyperactive or impulsive, whereas others have significant difficulties to pay attention, without being hyperactive or impulsive. Nevertheless, most children with ADHD manifest a mixture of these characteristics. A recent study, [10], found that children and adolescents with ADHD who are not currently taking medication may be at increased risk for overweight compared with children and adolescents without ADHD.

2 Logistic Regression Model

Regression models are one of the most important methods used in the statistical data analysis to model relationships between variables. The main objective of these models is to explore the relationship between one or more predictor variables (or independent), $\mathbf{X}$, and a response variable (or dependent), Y (see, for example, [5]).

2.1 Dichotomous Logistic Regression Model

Considering that

$$\mathbf{X}'\boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + ... + \beta_{p-1} X_{p-1}$$
$$\mathbf{X}'_i\boldsymbol{\beta} = \beta_0 + \beta_1 X_{i,1} + ... + \beta_{p-1} X_{i,p-1},$$

where $\mathbf{X}'$ is the $\mathbf{X}$ transpose and $\boldsymbol{\beta}$ are model parameters. Thus, the multiple logistic regression model is given by

$$E[Y|\mathbf{X}] = \pi(\mathbf{X}) = \frac{e^{\beta_0+\beta_1 X_1+...+\beta_{p-1}X_{p-1}}}{1+e^{\beta_0+\beta_1 X_1+...+\beta_{p-1}X_{p-1}}} = \frac{e^{\mathbf{X}'\boldsymbol{\beta}}}{1+e^{\mathbf{X}'\boldsymbol{\beta}}} = \frac{1}{1+e^{-\mathbf{X}'\boldsymbol{\beta}}}. \tag{2}$$

A crucial transformation in logistic regression models studies is the *logit* transformation, which is defined by

$$\pi^* = ln\left[\frac{\pi}{1-\pi}\right] = \mathbf{X}'\boldsymbol{\beta}. \tag{3}$$

The ratio $\frac{\pi(.)}{1-\pi(.)}$ is called the *odds*.

After fitting the model and assessing the estimated coefficients significance it is important to interpret their values.

When X_j, jth predictive variable, is coded 0 or 1 the coefficient interpretation associated to X_j can be made using the odds ratio (*odds ratio*), w, defined by,

$$w(1,0) = \frac{\left(\frac{e^{\pi^*_{j1}}}{1+e^{\pi^*_{j1}}}\right)\left(\frac{1}{1+e^{\pi^*_{j0}}}\right)}{\left(\frac{1}{1+e^{\pi^*_{j1}}}\right)\left(\frac{e^{\pi^*_{j0}}}{1+e^{\pi^*_{j0}}}\right)} = \frac{e^{\pi^*_{j1}}}{e^{\pi^*_{j0}}} = e^{\beta_j}. \tag{4}$$

To find the best model, we applied the AIC (*Akaike's Information Criterion*) and BIC (*Bayesian Information Criterion*) criteria,

$$AIC_q = -2\ln L(\hat{\boldsymbol{\beta}}) + 2q \tag{5}$$

$$BIC_p = -2\ln L(\hat{\boldsymbol{\beta}}) + q\ln(n). \tag{6}$$

2.2 Robust Logistic Regression Model

In robust statistics, robust regression is a way to overcome some limitations of the classic methods. For more details on robust regression see, for example, [7].

In logistic regression there are two main approaches to find robust parameters estimators, the first one minimizes the process likelihood (for example, see [3], [1] and [2]) and the second is based on influence functions (*eg*, [4]). Since the procedures for robust estimates proposed by [3], [1] and [2], are implemented in R, we thought to compute all these approaches. Unfortunately many numerical problems occurred, mainly due to obtaining singular covariance matrices estimates. Despite several attempts it was only possible to compute robust estimates for the logistic regression model when considering the method proposed by [2].

2.3 Polytomous Logistic Regression Model

The logistic regression models can be extended allowing handle polytomous responses (that have more than 2 categories).

Consider that the response variable, Y, has J categories. Then, for ith observation, there are J binary response variables denoted by $Y_{i1}, ..., Y_{iJ}$ where:

$$Y_{ij} = \begin{cases} 1, & \text{if the response of the } i\text{th observation is } j \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

As was seen for the dichotomous case, the *logit* for jth comparison is given by:

$$\pi_{ij}^* = \ln\left[\frac{\pi_{ij}}{\pi_{iJ}}\right] = \mathbf{X}_i'\boldsymbol{\beta}_j, \tag{8}$$

for $j = 1, 2, ..., J-1$ and

$$\pi_{ij} = \frac{e^{\mathbf{X}_i'\boldsymbol{\beta}_j}}{1 + \sum_{k=1}^{J-1} e^{\mathbf{X}_i'\boldsymbol{\beta}_k}}, \tag{9}$$

for $j = 1, 2, ..., J-1$.

3 Results

Exploratory data analysis was performed before logistic regression analysis. We looked at the distribution of variables (graphical) and some descriptive statistics. We also performed chi-square tests of independence. We only present here two plots, Figure 1 that represents the prevalence of children with ADHD and obesity

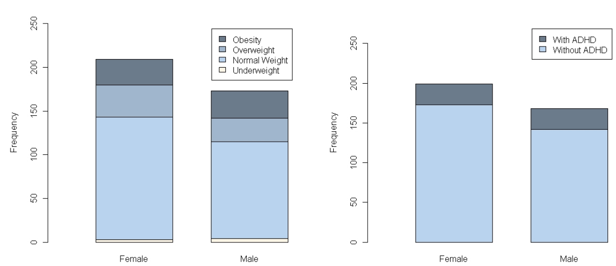

Figure 1: BMI and ADHD charts by gender.

for boys and girls. These plots do not show a strong prevalence of obesity gender.

The Figure 2 summarises the food frequency habits that were considered in the current study. Notice that the majority of children eat candies and chips more than three times a week.

3.1 Dichotomous Logistic Regression Model

Consider the dichotomous variable, Y (response variable) defined as follows:

$$Y = \begin{cases} 1, & \text{the child is obese} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

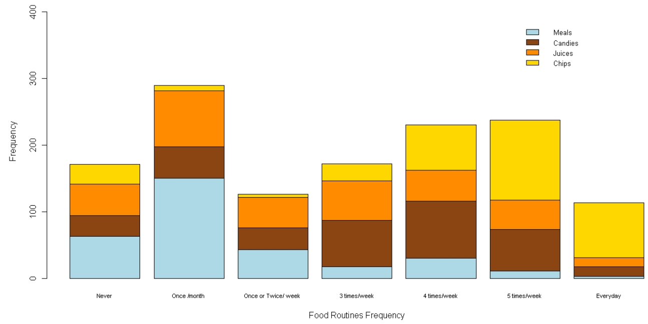

Figure 2: Bar chart of the frequencies of the four considered food routines.

Table 3: Estimated Odds ratio and their $CI_{95\%}$.

Variable	**Odds Ratio**	**Lower Bound**	**Upper Bound**
Age7	1.003	0.982	1.025
Age8	0.482	0.469	0.496
Sleep2	0.464	0.464	0.478
Sleep3	0.236	0.227	0.246
EatCandies1	4.563	4.252	4.896
FreqChips1	0.779	0.719	0.844
FreqChips2	1.100	1.013	1.195
FreqChips3	1.452	1.381	1.526
FreqChips4	1.700	1.629	1.775
FreqChips5	1.898	1.822	1.976
FreqChips6	2.508	2.407	2.613
ADHD1	1.069	1.039	1.099

After finding the best model according to AIC and BIC criteria, we obtained the estimated odds ratios and their confidence intervals of an (approximate) confidence level 95% present in Table 3.

Note that 7 years old children have a obesity risk close to that of children under 6 years (reference age). Moreover, 8 years old children seem to have a lower risk compared with 6 years old children.

Regarding the sleep variable children who sleep less than 8 hours per day (Sleep1 - reference category) seem to have a higher risk of obesity since other categories have a chance smaller than 1. For instance, comparing children who sleep less than 8 hours a day with those who sleep between 10 and 12 hours per day (Sleep3) note that the chance of obesity falls by about 76% for the latter.

Analysing the EatCandies variable note that children who eat candies appear to be a risk group (for obesity) since their risk of being obese is much higher than for children who do not eat candies.

For the frequency of eating chips, apparently children who never eat chips (reference category) have a lower risk of obesity than children who eat chips regulary (more than once a week). For instance, children who eat chips every day (FreqChips6) seems to be at highest risk of obesity, followed by the group of children group that eat chips five times a week (FreqChips5), for which the risk of becoming obese is about 90% longer than for children who never eat chips. Moreover, it seems that the risk of being obese increases with the increase of frequency of children eating chips.

Finally, for the ADHD variable, apparently, children suffering ADHD are more likely to be obese. In particular, children suffering from ADHD have an increased risk of being overweight of about 7% over children without this disorder.

After checking the model adequacy, and in order to infer the existence of outliers, we obtained the leverage points and the Cook distance for n observations:

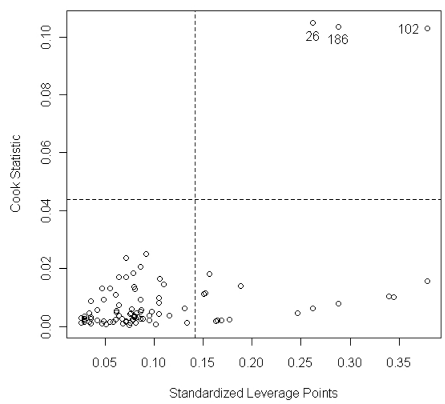

Figure 3: Standardised Leverage points and Cook statistic chart for the model.

Looking at Figure 3, we can identify 3 outliers and, after a detailed analysis, we could conclude that the outliers are influential observations. This calls for the use of robust methods.

3.2 Robust Logistic Regression Model

Due to numeric problems, we only considered the following variables: Age, Gender, Sleep, EatCandies, FreqCandies, DrinkJuices, EatChips and ADHD. Applying the bootstrap and *quasi-deviance* tests to the independent variables of the model we get:

Table 4: Estimates of the p-values using the tests *Bootstrap* e *Quasi-Deviance*.

Variable	***Bootstrap*Test**	***Quasi-Deviance* Test**
Age	0.48	0.45
Gender	0.50	0.67
Sleep	0.09	0.03
EatCandies	0.03	0.07
FreqCandies	0.13	0.07
DrinkJuices	0.99	0.99
EatChips	0.23	0.25
ADHD	0.66	0.72

Table 4 shows that only the variables Sleep, EatCandies and FreqCandies are statistically significant.

3.3 Polytomous Logistic Regression Model

Due to the small number of observations we considered 3 instead of 4 categories in the response variable (grouping the underweight and normal weight categories into a single category).

We applied the stepwise method to find the best model and obtained the estimated odds ratios in Table 5.

Regarding to Sleep there seems to be evidence that children who sleep less (Sleep1 - reference category) have a higher chance of being obese or overweight.

Table 5: Odds Ratio estimated for the polytomous logistic regression model.

	Overweight	Obesity
(Intercept)	0.02	0.90
Sleep2	0.76	0.50
Sleep3	0.63	0.08
FreqCandies2	0.92	0.34
FreqCandies3	1.05	0.96
FreqCandies4	1.14	1.64
DrinkJuices1	5.46	0.52
EatChips1	3.49	1.18
ADHD1	1.53	0.61

Children who eat candies 4 times a week or more (FreqCandies4) are the group where overweight and obesity is more frequent when compared with normal weight and children who eat candies once a month or less (FreqCandies1).

Regarding the DrinkJuices variable, children who drink juices appear to be at risk of being overweight 5.5 times higher than children who do not drink juice, compared to having a normal weight. The relative risk between being obese and having normal weight is expected to fade about 48%, that is, it seems that there is evidence that children who drink juice have less chance to be obese, compared with children with normal weight. This apparent contradiction may be explained by the lack of observations of overweight children who do not drink juices as shown in the Table 6.

Table 6: DrinkJuices relative frequency.

		Normal Weight	Overweight	Obesity
DrinkJuices	0	0.0857	0.0048	0.0286
	1	0.5619	0.1952	0.1238

Children who eat chips appear to be at risk of being overweight 3.5 times higher than children who do not eat chips, compared to having a normal weight. Analogous, there seems to be evidence that the obesity risk is higher (18.5%) for children who eat chips than for those who do not eat chips, compared to a normal weight.

Finally, regarding ADHD it is expected that the relative risk of being overweight, increase 53% in children with ADHD. Moreover, the relative risk of suffering obesity problems, for children with ADHD comparing with having a normal weight, is expected to fade from about 39%. This means that there seems to be evidence that children with ADHD have a lower chance of being obese, compared to having a normal weight. Once again, due to the small number of children with ADHD in each response variable category, as shown in Table 7, the above may seem contradictory.

Table 7: ADHD relative frequency.

		Normal Weight	Overweight	Obesity
ADHD	0	0.5619	0.1619	0.1381
	1	0.0857	0.0381	0.0143

The results presented in this study were obtained by using the statistical software R, [6], either using R functions or functions implemented in the same computational language.

4 Conclusions

In this study the logistic regression analysis is carried out as the BMI response variable and independent variables that reflect nutritional and lifestyle habits and also features like birth weight and parent's educational levels.

From preliminary data analysis we observed, for instance, that the percentage of obese child (or overweight children) is in reverse to the number of hours that a child sleeps per day. When was considering a dichotomous response variable (obese and not obese) for the logistic regression model, we found that, despite of the descriptive analysis did not show a relationship between age and BMI, children under 8 years seem to have a lower risk of obesity. Furthermore, apparently if the child sleeps few hours, the chance to be obese increases. As expected, the children that eat candies can be considered an obesity risk group and that the risk of being obese increases with the increase of the frequency at which children eat chips. Also, we have found that children with ADHD are at increased risk for being overweight, when we included this variable into a dichotomous logistic model. In general, the results obtained with polytomous logistic regression model confirmed the results of the dichotomous model.

ADHD and obesity are important topics in the field of childhood research and health care. This study provides some findings in these topics but future work is needed to better understand the other factors that influence the relationship between these variables.

References

[1] Bianco, A., Yohai, V.J. (1996). *Robust estimation in the logistic regression model.* In: Rieder, H. (Ed.), Robust Statistics, Data Analysis, and Computer Intensive Methods, Lecture Notes in Statistics, **109**. Springer, New York.

[2] Cantoni, E. e Ronchetti, E. (2001). Robust Inference for Generalized Linear Models. *Journal of The American Statistical Association*, **96**, 1022–1030.

[3] Copas, J. B. (1988). Binary Regression Models for Contaminated Data. *Journal of the Royal Statistical Society. Series B*, **50**, 225–265.

[4] Kordzakhia N., Mishra, G. D., Reiersølmoenand, L. (2001). Robust estimation in the logistic regression model. *Journal of Statistical Planning and Inference*, **98**, 211–223.

[5] Hosmer, D. e Lemeshow, S. (1989). *Applied Logistic Regression.* John Wiley and Sons, Inc.

[6] R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

[7] P. J. Rousseeuw and A. M. Leroy (2003). *Robust regression and outlier detection.* Wiley and Sons, New York.

[8] Silva, A. B. B, (2003). *Mentes Inquietas: Compreender o Distúrbio do Défice de Atenção (DDA).* Rio de Janeiro: Napads.

[9] Padez et. al. (2004). Prevalence of overweight and obesity in 7-9-year-old Portuguese children: trends in body mass index from 1970-2002. *American Journal Human Biology*, **16**, 670–678.

[10] Waring, Molly E., Lapane, Kate L. (2008). Overweight in Children and Adolescent in Relation to Attention-Deficit/Hyperactivity Disorder: Results from a National Sample. *Pediatrics*, **122**, e1–e6.