Progress in Nutrition 2022; Vol. 24, N. 4: e2022117 DOI: 10.23751/pn.v24i4.12170

Original article

Logistic regression analysis of finding associated factors to predict loss weight adults in Erbil City (2018)

İlker Etikan, Hewir Abdulqadir Khidir
Department of Biostatistics, Near East University, Cyprus

Abstract. One of the threatened risk factors which leads to numerous health issue is overweight or obesity. Many studies have been carried out about this problem and yet not exact cause have been found. Numerous factors are highlighted as primary ones such as, not doing exercise, unorganized daily meal, medical condition and etc. In addition, statistical analysis plays an essential role in finding the most effective factors linked to reducing body weight and since the values of response variables lies under two levels which shows no linear relationship between the outcome and explanatory variables, thus Binary Logistic Regression a family of Generalized Linear Model was performed. Logistic regression analysis is a very common tool and serves great part in health science due to the fact that most of the phenomena's outcome have only two values (alive/dead, exposed/not exposed, presence/absence, and etc.). The overall adults who underwent losing weight and succeeded was 57.7% and 42.3% who failed. Frequently medical visits and exercise were highly significant and odds of one-unit increase, has about 3 times more chance to lose. Gender, eating-out, overeating and irregular eating were all highly significant. However, diet and number of healthy meals were not found to be associated.

Key words: Overweight, GLMs, Reducing Weight, Logistic Regression Model, Explanatory Analysis, Odds Ratio.

Background to the study

People are getting more and more obese around the world and this directly affects their health and well-being (1). As a result, several severe diseases occur frequently such as heart disease, hypertension, diabetes, stroke, osteoarthritis and cancer (2,4). Nevertheless, researches have shown that a change in eating behavior and doing regular exercises are considered as obesity treatment (5,6). In this research, a doctor, a dietitian, and an exercise programmer have directed the weight reduction program. Therefore, the patient's diet, physical activity and medical condition are studied for the weight reduction program. To have an active behavioral therapy, sibutramine or orlistat were prescribed (7,8). It is of paramount importance to predict which factors are mainly relevant to decrease in the weight after weight loss intervention.

The disparity between current practice and ideal management of adult obesity is one of the factors that contribute to being resistant to weight loss schemes. Other factors can be lack of knowledge, time and reimbursement (1,2). In addition, physicians are often cynical about how hard headway patients are trying to lose weight. This is due to the fact that the average of weight loss is relatively modest whereas the repetition rate is extremely high in a well-controlled clinic (3). However, sometimes a significant number of patients achieve long term weight loss which in turn makes the physicians optimistic about these clinical trials.

Interestingly, weight loss can be clinically significant regardless to whether the rate is small or huge.

A modest loss (5%-10% of total body weight) will have its major health benefits. Studies have shown that people lose weight much more when they are guided by their primary care physician (5,8). This is due to the fact each patient needs different approach to lose weight and there is no fixed approach that works for all. Therefore, the primary care physicians can suggest a variety of treatments such as dietary change, increased physical activity, medication for selected patients, and surgery for severely obese adults. Moreover, in 2019 The National Committee on Quality Assurance included body mass index (BMI) to the list of effective care measures which physicians are rated on (9). Further to this, previous researches have indicated that obesity accounts for more than 9% of annual health care costs, consequently obesity should be dealt with as a medical condition not only as a risk factor.

In the practical phase, we usually face analysis based on linear relationship between variables and the assumption of normality distribution of the response variable and the error term of the fitted model. However, robust to normality assumption very often the case of concerning and usually transformation is applied to overcome problems with non-normal error bit. While it is usually concerned where errors normal distribution is not achieved via transformation, for instance when the outcome variable is categorical. Thus, in this situation, it is better to fit a model on the actual data rather than transforming. New techniques were introduced by (10) and further improved by (11) and is called generalized linear modeling (GLM).

Generalized linear models

In order to make the generalized linear models (GLMs) more applicable than the general linear model, we need to show some important characteristics and above all is that the technique of least square estimation is no longer valid and instead maximum likelihood manner is applied. The most vital components are described below (12);

1. Random Component: It is related to the probability distribution of the response variable and must be from the exponential family (Normal, Binomial, Poisson, Gamma and Negative Binomial). Each specifies along with level values of the response variable.
2. Systematic Component: This one represents the independent (covariates) variables in the model, as they might be continuous and/or categorical, plus interaction terms between predictors might be included along with polynomial functions of predictors.
3. Link Function: It is generally directed to connection between the random component (response variable) and the systematic component (predictors). The expected value of Y is linked to the covariates:

$$g(x)=b0+b1X1+b2X2+\ldots$$

As g(m) is the link function and b0, b1, etc are the parameters need to be estimated. We show the three most common link functions;

a. Identity Link: Mathematically it is written as g(m)=m and is used in standard linear models.
b. Log Link: Mathematically it is written as g(m)=log(m) and is used mostly for count data (non-negative values) in log-linear models.
c. Logit Link: Mathematically it is written as g(m)=log[m/(1-m)] and used for binary data (Logistic Regression).

Logistic regression

A very common application of GLMs in medicine or health science is to build a responsive model between response variable with only two outcome values such as (e.g. alive/dead, presence/absence, etc.) and explanatory variables which can be any type of data either continuous or categorical. As (13) mentioned that two situations where GLM application were required due to the nature of the response variable (binary type) as seen in Figure 1. First, it was to predict the probability of survival (survived or did not survived), and second was about the burrowing probability. Also, (14)

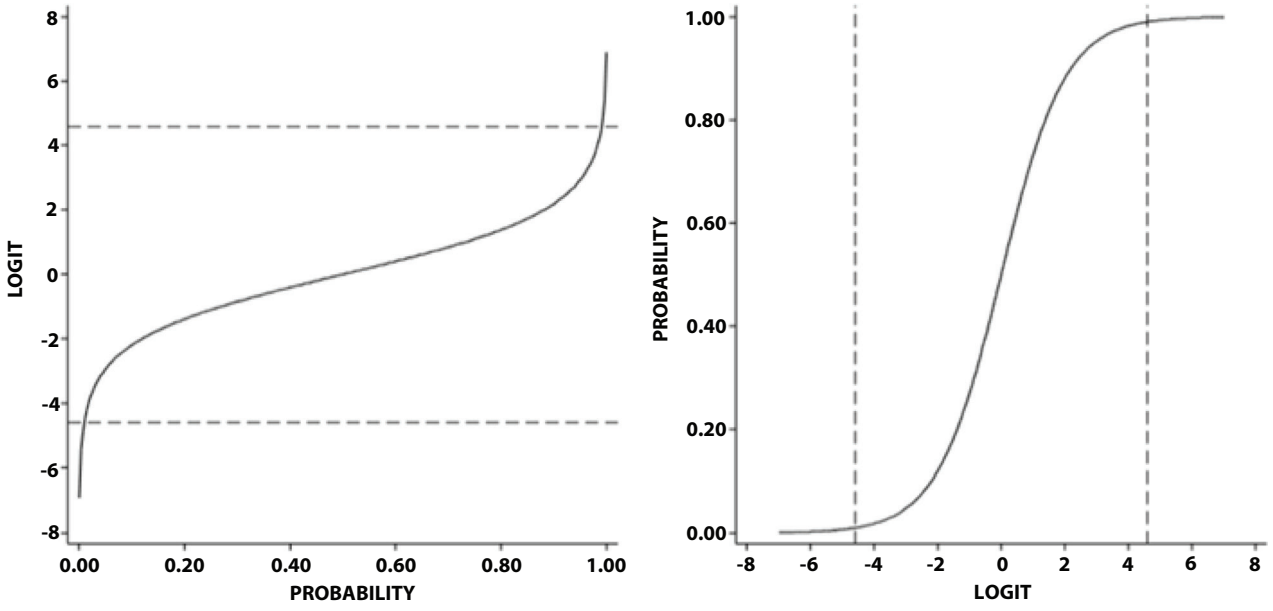

Figure 1. Logit and inverse logit functions

studied the relationship between the presence/absence of species in forest shrubs as response variable against a set of predictors such as, distance to nearest woodland, stand age, stand area, etc.

Logistic model and parameters

The model for logistic regression can be written as follows:

$$\pi(x) = \frac{g(x)}{1 - g(x)}$$

$$g(x) = e^{\beta_0 X_1 + \beta_1 X_2 + \ldots + \beta_p X_p}$$

where β_s are parameters need to be estimated, $\pi(x)$ is the probability that $y_i = 1$ (i.e. the event is present) for a set of given x_i's. In order to make the interpretation of the parameters easier, we simply can transfer them into odds with the same interpretation as done in 2×2 contingency table which is the probability of an event occurs relative to its converse.

$$Odds = \frac{\pi(x)}{1 - \pi(x)}$$

The probability of $Y_i = 1$ is greater that its converse if the odds are >1, and vice versa. To calculate odds through logistic regression, we can easily convert it to odds and odd ratio by powering the coefficient values on exponential function. Then, to make the modeling easier the natural logarithm can be taken to the odds formula;

$$\ln(Odds) = ln\ \frac{\pi(x)}{1 - \pi(x)}$$

Then $g(x)$ can be driven as below:

$$g(x) = \beta_0 X_1 + \beta_1 X_2 + \ldots + \beta_p X_p$$

It is also worth mentioning that the maximum likelihood estimation is processed and of course due its complexity the values are computed through computation. To test an overall model whether significant, a test is used called likelihood ratio test (LRT), as it is based on comparison the likelihood of the full model against the null model (i.e. a model with only intercept) (15). It is most likely the F-test in linear regression. However, we use a special case of likelihood ratio test to investigate the full and saturated model

using deviance (SSE in linear regression) (16). The deviance is an unexplained measure of variation in the data. Moreover, regarding to testing the significance of parameters, Wald test is an approach to test the null hypothesis as written below (4).

Illustrating and reporting logistic regression analysis result

The dataset was collected by a group of undergraduate students in Salahaddin University – Erbil/Iraq in 2018 academic year and consisted of a combination of characteristics (11 predictors) in relation to weight change (loss weight was event of interest) as response variable (1= failure to loose, 1= success to loose) for 125 inner city adults at several sport centers and physician centers. It is worth mentioning that data was only collected to those underwent weight reduction process for more than 3 months. Of these adults, 71 (57.7%) were lost their weight at the time of collecting the data and 52 (42.28%) were not.

It is for the best to generate the descriptive statistics of the data and see how the covariates are related to outcome variable. Table 1 guide us that there is potential relationship among between some independent variables and dependent binary variable. Whereas no significant effect of both age and height was found. This is good step to go further in detail afterwards.

Logistic regression analysis

A multiple logistic regression analysis was applied to the dataset in order to examine the likelihood effect of the some biographic, health as well as other characteristics on losing weight. The analysis was carried out by R version 3.6.2 program. The fitted model was turned out as:

$$\begin{aligned} &\textit{Prediction of logit}\,(\textit{loss weight}) \\ &\quad = -12.687 + 1.984 * \textit{Gender} + 1.508 * \\ &\quad \textit{FMV} + 2.814 * \textit{IE} + 1.265 * \textit{Exercise} + \\ &\quad 3.564 * \textit{Overeating} \end{aligned}$$

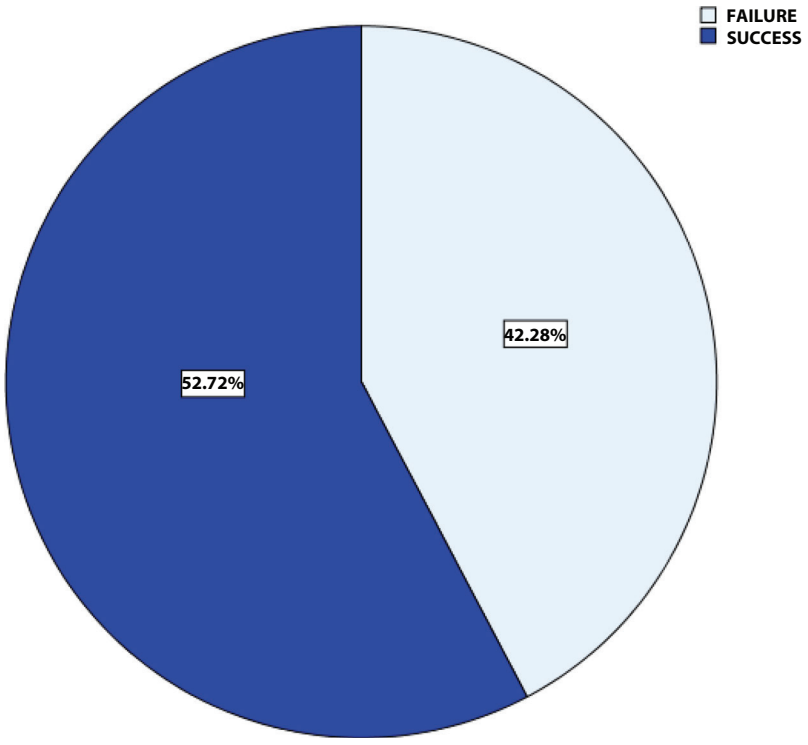

Figure 2. Pie chart of response variable

Table 1. Descriptive Statistics for the dataset.

Variables	Failure	Success
Age	25.154 ± 5.795	25.958 ± 8.591
Hight	166.365 ± 10.041	168.521 ± 8.833
Exercise	1.923 ± 0.860	2.915 ± 0.967
Frequency of Medical Visit	3.038 ± 0.625	5.000 ± 1.309
Gender		
Male	33 (0.56)	26 (0.44)
Female	19 (0.30)	45 (0.70)
Diet		
Yes	13 (0.19)	54 (0.81)
No	39 (0.70)	17 (0.30)
Meal		
1	4 (0.13)	27 (0.87)
2	12 (0.27)	32 (0.73)
>= 3	36 (0.75)	12 (0.25)
Irregular Eating (Yes/No)		
Yes	41 (0.77)	12 (0.23)
No	11 (0.16)	59 (0.84)
Overeating		
Yes	40 (0.77)	12 (0.23)
No	12 (0.17)	59 (0.83)
Eating-out		
Yes	42 (0.81)	10 (0.19)
No	10 (0.14)	61 (0.86)
Smoking		
Yes	32 (0.59)	22 (0.41)
No	20 (0.29)	49 (0.71)

The result of Table 2 is probably the most important outcome in logistics where we can learn how the outcome variable and covariates are associated. To make the interpretation simple, only significant coefficients were retained, and to understand and interpret the parameters value, odds ratio was computed by exponentiating B's. Referencing to Gender, the odds (7.275) indicates that the model predicts that the odds of losing weight are 7.275 times higher for women than that for men. This is in fact a good result and shows that women are more likely to intend to lose weight than men.

The odds ratio for frequency of medical visit (FMV/times per month) was calculated as 4.517 with 95% CI with range (1.861 to 10.960). This leads us to say that for one unit increase in FMV, an adult has 3.886 times more chance to lose weight compared to not to. Exercise also played an important role to be effective in losing weight with one hour increase in doing exercise per day this would lead to have odds (3.543) times chance to lose weight comparing to not to increase.

Regarding to Irregular eating, eating out and over eating, they all showed to have great impact in

Table 2. Result of Estimated parameters, Odds Ratio in Logistic Regression.

	B	S.E.	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Gender (1)	1.984	0.924	4.613	0.032	7.275	1.189	44.503
Frequency of Medical Visit	1.508	0.452	11.112	0.001	4.517	1.861	10.960
Irregular Eating (1)	2.814	0.932	9.112	0.003	16.679	2.683	103.680
Exercise	1.265	0.455	7.726	0.005	3.543	1.452	8.645
Overeating (1)	3.564	1.072	11.054	0.001	35.298	4.319	288.509
Constant	-12.687	2.676	22.470	0.000	0.000		

Table 3. Log-Likelihood value of full model and reduced model.

Models	Chi-square
Full model log-likelihood	127.325
Reduced model (constant only) log-likelihood	167.567
-2 Log likelihood	40.242
P-value	0.000

Table 4. Goodness of fit output for the full model.

Statistic	Value	df	P-value
Hosmer & Lemeshow	1.165	8	0.997
Nagelkerke R Square	0.867		
Cox & Snell R Square	0.645		

losing weight and they need to be taken into account for further works. Those who avoid irregular eating has 16.679 more chance of losing weight than those not. In relation of over eating association with losing weight was beyond of our expectation as it turned out to have a significant effect with 35.298 odds in favor to lose weight.

The table 3 proves that the full model is reduced the log-likelihood value and hence it is improved with significant difference between the two models.

To assess the fitted model, goodness of fit was performed as shown in Table 4. The statistical test of Hosmer-Lemeshow was not significant as it was true for Pearson χ^2 or G^2 (deviance), so no tangible evidence for lack of fit. In addition, Nagelkerke R Square indicates a good feel about the model which provided 86.7% the uncertainty in success of losing weight and can be explained by the entered variables in the model.

Conclusions

In this study, several important factors were found to be highly effective on reducing body weight and it might be not be identical to other regions. Since the food in Iraq is full of fat, salt and other unhealthy materials and also almost all nutrients are imported from abroad. Hence, not a good quality control is undertaken, and those attributes related to this factor were highlighted as significant with large effect. Exercise (hours/day) and number of medical visits (visit/month) had great impact on reducing weight and as a result of increasing one-unit at each, on average there is an odd of more chance to lose weight with 3 times. The effect parameters of Irregular eating and overeating were also high and significant, and on average those who stopped such habits while underwent in program of losing, had 16 times more chance to lose compared to others who were not. In addition to this, eating-out was playing substantial to lose weight with 7.951 odds ration in favor not to have meal out. The study found out that women were more interested in reducing weight than men with odds ratio 7.275. However, diet, number of healthy meals in a day, smoking and age were not found to be statistically significant and that was they were removed in the final model to simple the interpretation. It is worth mentioning that this study had its own limitations such as, the number of participants, this study was not a clinical trial and one can have with clinical trial design and end up with a better result. It is recommended to generate classification table and calculate sensitivity, specificity and plot ROC to see how effective the covariates influenced the outcome variable.

Highlight

Aim of the Study To examine and analyze the relationship between some factors such as gender, age, exercise, number of meals per day, eating out, over eating, irregular eating, medical visit, smoking and adult's weight in Erbil City.

Research Questions: 1) Are the factors significantly correlated to adult asthma? 2)What is the greatest procedure to investigate the connection between those factors and losing weight's adult?

Objectives of the study: 1) To assess the contribution of attributers that are linked to the intention of losing weight. 2) To use Logistic regression models to analyze results. 3) To interpret the estimate parameters and assess the fitted model.

Conflicts of Interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article.

References

1. Hedley AA, Ogden CL, Carroll MD, Curtin LR, Flegal KM, Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. JAMA 2004; 291:2846-50
2. Pi-Sunyer FX Medical hazards of obesity. Ann Intern med 1993; 119:655-60.
3. Eckel RH, Obesity and heart disease: a statement for healthcare professionals from the Nutrition Committee, American Heart Association 1997. Circulation; 96:3248-50.
4. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH, The disease burden associated with overweight and obesity. JAMA 1999; 282:1523-
5. Zachwieja JJ, Exercise as treatment for obesity. Endocrinol Metab Clin North Am 1996; 25:965-88
6. Flechtner-Mors M, Ditschuneit HH, Johnson TD, Suchard MA, Adler G, Metabolic and weight loss effects of long-term dietary intervention in obese patients: four-year results. Obes Res 2000; 8:399-402.
7. James WP, Asturp A, Finer N, et al., Effect of sibutramine on weight maintenance after weight loss: a random trial. STORM Study Group. Sibutramine Trial of Obesity Reduction and Maintenance. Lancet 2000; 356:2119-25.
8. Davidson MH, Hauptman J, DiGirolamo M, et al., Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat a randomized controlled trial. JAMA 1999; 281:235-42.
9. Whately JE, Gillespie WJ, Honig J, Walsh MJ, Blackburn AL, Blackburn GL, Does the amount of endurance exercise in combination with weight training and a very-low-energy diet affect resting metabolic rate and body composition? Am J Clin Nutr 1994; 59:1088-92.
10. Nelder, J. & Wedderburn, R.W.M, Generalized linear models. Journal of the Royal Statistical Society 1972; A135: 370–384.
11. McCullough, P. & Nelder, J.A., Generalized Linear Models, 2nd edition. Chapman & Hall, New York 1989.
12. Hilbe, J.M., Generalized additive models software. The American Statistician 1993; 47: 59–64.
13. Beck, M.W. Size-specific shelter limitation in stone crabs: a test of the demographic bottleneck hypothesis. Ecology 1995; 76: 968–980.
14. Matlack, G.R., Plant species migration in a mixed history forest landscape in eastern North America. Ecology 1994; 75: 1491–1502.
15. Menard, S. Applied logistic regression analysis (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–106). Thousand Oaks, CA: Sage 1995.
16. Menard, S. Coefficients of determination for multiple logistic regression analysis. The American Statistician 2000; 54(1), 17–24.

Correspondence;
Prof. Dr. İlker Etikan,
Faculty of Medicine, Department of Biostatistic
University of Near East – North Cyprus
E-mail: ietikan@gmail.com