https://doi.org/10.31272/jae.i151.1538
https://admics.uomustansiriyah.edu.iq
P-ISSN: 1813–6729 E-ISSN: 2707-1359

# Binary Logistic Regression Analysis to Study Factors Affecting Addiction

**Suhad Ahmed Ahmed**
Dept. of Computer Science, College of Education for Pure Sciences (Ibn Al-Haitham), University of Baghdad, Baghdad, Iraq.
Email: suhad.a.a@ihcoedu.uobaghdad.edu.iq , ORCID: https://orcid.org/0000-0002-0072-6990

**Article Information**

**Article History:**
Received: 12 / 12 / 2025
Revised: 28 / 01 / 2026
Accepted: 03 / 02 / 2026
Available Online: 01 / 03 / 2026
Pages no: 142 – 149

**Keywords:**
Binary Response, Logistic Regression Model, Questionnaire form, Addiction.

**Correspondence:**
Researcher name :
Suhad Ahmed Ahmed

Email:
suhad.a.a@ihcoedu.uobaghdad.edu.iq

## Abstract

*The problem of drug addiction is considered one of the most dangerous health, social, and psychological issues facing the world.*

*The research aims to study the effect of some variables on addiction. Factors influencing addiction using the logistic regression model. Cases were classified into addicted and non-addicted based on variations in the level of addiction and the number of times the individual visited the hospital, which represents the dependent variable. The independent variables represent the factors influencing addiction.*

*The sample consists of 160 cases of both genders and various ages, based on a questionnaire that includes 20 questions about the causes of addiction. The results showed that the most significant factors leading to addiction were chronic diseases, imprisonment or detention, as well as smoking and hookah use. The SPSS statistical software was applied to obtain and analyse the results.*

## 1. Introduction

There are many factors that influence the tendency toward drug addiction, along with a strong and urgent desire that drives the addict to obtain drugs by any means and to increase the dose from time to time, with difficulty or even impossibility of quitting. There is a clear and dangerous increase in addiction in Iraq, particularly among young people. Addiction is a state of physical and psychological dependence in which an individual loses control over the use of a substance, continuing to use it repeatedly despite knowing its negative effects.

According to official statistics, the number has reached more than a million people, including both drug users and addicts, according to estimates by global health organisations.

There are many factors that influence the tendency toward addiction. To study these factors, according to official statistics, the number has reached more than 800 million people, including both drug users and addicts, according to estimates by global health organisations.

Especially when the dependent variable is binary logistic (e.g., success/failure, yes/no, or addicted/not addicted), the binary logistic regression model is used. This model analyses the relationship between a binary dependent variable and independent variables that may be numerical or categorical. It is widely used in health sciences, finance, marketing, and biostatistics to forecast the likelihood of an event occurring.

## 2. Paper importance

The importance of this research stems from addressing the phenomenon of addiction as one of the complex social and health problems that directly affect individuals and society. The use of the logistic regression model to analyse the factors influencing addiction provides a rigorous statistical framework for explaining the relationship between the independent variables and addiction status, classified as (addicted, non-addicted). Moreover, the results of this study help identify the most

influential factors, which can support policymakers and health institutions in developing scientifically based preventive and therapeutic strategies, and enrich statistical and applied studies in binary data analysis.

## 3. Paper aims

1- To study the factors influencing addiction and determine the nature of their relationship with an individual's addiction status.
2- To apply the logistic regression model to estimate the probability of addiction based on the studied variables.
3- To identify the most statistically significant variables affecting addiction.
4- To provide statistical results that can be utilised in addiction prevention and treatment strategies.

## 4. Logistic Regression Model

Logistic regression is used to analyse the relationship between a binary dependent variable and one or more independent variables, which can be either continuous or categorical. The dependent variable y is modelled as a Bernoulli variable, taking the value 1 if the event occurs and 0 if it does not; 1 represents the positive outcome and 0 the negative outcome.

Since the dependent variable is binary, a traditional linear regression model is not appropriate because its predicted values can fall outside the [0,1] range. Therefore, logistic regression is used, linking the probability of the outcome to the predictors via the logit function, ensuring that predicted probabilities remain in the [0,1] interval.[1],[2]

In linear regression:

$$Y= B_0+B_0X_1 + \varepsilon \quad (1)$$

1- This can predict any real number $(-\infty,+\infty)$, but probabilities must be between 0 and 1.
2- So we need a function that maps real numbers to [0,1].

For a single independent variable x, the logistic regression model is expressed as:

$$p(x) = E(y/x) =\frac{e^{B_0+B_1x}}{1+e^{B_0+B_1x}} \quad (2)$$

where $B_0$ , $B_1$ are The logistic regression model's parameters in the case of a single independent variable.[3]،[4]

If there are k independent variables, the mathematical form becomes:

$$f(x) = E(y/x) =\frac{e^{B_0+B_1x_1+\cdots+B_kx_k}}{1+e^{B_0+B_1x_1+\cdots+B_kx_k}} \quad (3)$$

where $B_0$,$B_1$ ,…,$B_k$ represent the parameters to be estimated.[5]

### *4.1. Estimating the Logistic Regression Model*

The model parameters are estimated using Maximum Likelihood Estimation (MLE), which identifies the set of coefficients that maximises the likelihood of observing the actual data.
The odds of the outcome are defined as:

The odds are the ratio of the probability of success to the probability of failure as follows:

$$\text{Logit } (f(x)) = \ln \left(\frac{f(x)}{1-f(x)}\right) = e^{B_0+B_1x_1+\cdots+B_kx_k} \quad (4)$$

The logit (log-odds) is the natural logarithm of the odds: [6]

$$\text{Odds}=\frac{f(x)}{1-f(x)} = \frac{e^{B_0+B_1x_1+\cdots+B_kx_k}}{1} = e^{B_0+B_1x_1+\cdots+B_kx_k} \quad (5)$$

This transformation makes the model linear with respect to the parameters.

The parameters $B_0, B_1, \dots, B_k$ are estimated using Maximum Likelihood Estimation (MLE)**.** The goal is to find the parameters that maximise the likelihood of observing the actual data. [6]

In this form, the model is linear in terms of the parameters$B_0$,$B_1$ ,…,$B_k$ allowing Maximum Likelihood Estimation to be used to estimate the coefficients so that the probability of observing the data is maximised.

### *4.2. Hosmer and Lemeshow Test*

The Hosmer–Lemeshow test is commonly used to assess the goodness-of-fit of a logistic regression model, i.e., whether the model adequately represents the observed data. The method involves grouping predicted probabilities into deciles (or other quantiles) and comparing the observed and expected frequencies of the outcome within each group using a chi-square statistic:[7],[8]

$$H = \sum_{k=1}^{n} \frac{(o_k - n_k p_k)^2}{n_k p_k (1-p_k)} \quad (6)$$

where: $n_k$ : represents the total number of observations. $o_k$ : is the total number of observed responses. $p_k$ : is the average of the predicted probabilities.

After calculating H, it is compared to a significance level of 0.05 using n−1 degrees of freedom.

A non-significant Hosmer–Lemeshow test ($p > 0.05$) indicates that there is no strong evidence of lack of fit, suggesting that the model's predictions are consistent with the observed data. It is important to note that a non-significant result does not confirm that the model is correctly specified, especially if there are issues such as separation or overfitting.

This test is used to evaluate the goodness of fit, i.e., whether the model represents the data well. It is similar to the Chi-square test, and its mathematical formula is:[7]،[8]

### *4.3. Pseudo-R² Measures*

In logistic regression, the traditional coefficient of determination ($R^2$) used in linear regression is not directly applicable because the dependent variable is binary. Instead, $R^2$-like statistics, often called pseudo-$R^2$, are used to indicate model fit.

#### *4.3.1 Cox & Snell R²*

1- Approximates the proportion of variance explained by the model.
2- Its maximum value is less than 1, even if the model fits perfectly.[11]

#### *4.3.2 Nagelkerke R² (adjusted Cox & Snell)*

1- Adjusts Cox & Snell $R^2$ so that the maximum possible value is 1.
2- widely used as a measure of model goodness-of-fit, analogous to $R^2$ in linear regression.[12],[13]

The Nagelkerke $R^2$ is calculated as:

$$R^2 = 1 - \left(\frac{L_0}{L_m}\right)^{2/n} \quad (7)$$

$L_o$: Weighted Function with intercept. $L_m$ : Weighted Function with all independent variables.

Pseudo-$R^2$ values provide an approximate measure of explanatory power, but they should not be interpreted in the same way as $R^2$ in linear regression.

## 5. Applied Aspect

Society is increasingly facing the phenomenon of drug addiction, which is influenced by a wide range of factors. These factors were identified among addicted patients in order to determine the main causes leading to addiction. To identify the most significant of these factors, SPSS was used to analyse the data.

## 6. Data Description

The study data consisted of a sample of patients addicted to substances from Ibn Rushd Teaching Hospital for Psychiatry. Data were collected using a questionnaire distributed to 160 patients. The environmental factors surrounding the patients that led to substance use were considered as the independent variables, while the dependent variable represented the patient's status, either addicted or non-addicted, coded as (1, 0), respectively. The classification of a patient as addicted was based on the number of follow-up visits, the dosage percentages, and, in cases requiring the patient to be hospitalised for several days, below is a clarification of the independent variables:

1- Patients' Ages (x1)
2- Marital Status (x2)

(Single male, Single female, Married, Divorced male, Divorced female, Widow)

3- Occupation (x3)
(Housewife, Employee (female), Student (female), Student (male), Employee (male), Self-employed, Unemployed)
4- Gender (x4)
(Male, Female)
5- Educational Level (x5)
(Bachelor's degree, Secondary school, Intermediate school, Primary school, Illiterate)
6- Causes Leading to Addiction (x6)
(Financial problems, Social problems)
7- Causes of Job Loss, Family Loss, or Other Causes (x7)
8- Attempt to Quit Addiction (x8)
(Yes, No)
9- Awareness or Counselling to Quit Addiction (x9)
(Yes, No)
10 - Experimentation, Bad Companions, Unemployment, Other Reasons (x10)
11- Addiction Due to Chronic Diseases (x11)
12- Desire to Quit Drugs (x12)
(Yes, No)
13- Family Support (x13)
(Yes, No)
14- Psychological Problems (x14)
(Yes, No)
15- Physical Problems (x15)
(Yes, No)
16- Influence of Relatives or Close People (x16)
(Yes, No)
17- Addiction Due to Imprisonment or Detention (x17)
(Yes, No)
18- Smoking Cigarettes or Hookah (x18)
(Yes, No)
19- Loss of One or Both Parents (x19)
(Yes, No)
20- Lack of Attention and Awareness (Absence of Family Support) (x20)
Dependent Variable (y)
The dependent variable was classified into two categories:
(Addicted, Not addicted)

## 7. Binary Logistic Regression Model [7]

To identify the variables that affect addiction, the statistical software SPSS was used.

$H_0$ : No differences were found in the effect of independent variables.

$H_1$ : found different effects of independent variables.

The results are shown below:

**Table (1)** Omnibus Test of Model Coefficients

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 30.278 | 15 | .011 |
| | Block | 30.278 | 15 | .011 |
| | Model | 30.278 | 15 | .011 |

The table (1) illustrates the efficiency and significance of the model using the maximum likelihood ratio. The degrees of freedom (15) indicate the number of independent variables.

**Table (2)** Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 180.374 | .172 | .236 |

The $R^2_{cox\ \&\ snell}$value indicates that 17.2% of the variation in the dependent variable is explained, while the $R^2_{Nagelkerke}$value indicates that 23.6% of the variation in the dependent variable is explained

**Table (3)** Hosmer and Lemeshow Test

| Step | Chi-square | df | Sig. |
|---|---|---|---|
| 1 | 17.609 | 8 | .024 |

The results of the chi-square ($\chi^2$) test indicate the goodness-of-fit of the logistic regression model to the data, with a significance value (sig) = 0.024**,** which exceeds 0.05. This means there is no significant difference between the predicted and actual values, indicating that the model is appropriate.

Additionally, the chi-square value = 17.609 suggests that there is no problem with the model's fit.

**Table (4)** Variables in the Equation

| case | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | x1 | .017 | .041 | .172 | 1 | .678 | 1.017 | .939 | 1.101 |
| | x2 | -.213 | .146 | 2.132 | 1 | .144 | .808 | .607 | 1.076 |
| | x5 | .672 | .325 | 4.282 | 1 | .039 | 1.957 | 1.036 | 3.698 |
| | x6 | .108 | .196 | .303 | 1 | .582 | 1.114 | .759 | 1.635 |
| | x7 | -.577 | .461 | 1.567 | 1 | .211 | .562 | .228 | 1.386 |
| | x9 | -.510 | .716 | .508 | 1 | .476 | .600 | .148 | 2.443 |
| | x10 | .162 | .686 | .055 | 1 | .814 | 1.175 | .306 | 4.509 |
| | x11 | -.798 | .591 | 1.822 | 1 | .177 | .450 | .141 | 1.434 |
| | x14 | -.016 | .424 | .001 | 1 | .971 | .985 | .429 | 2.260 |
| | x15 | .467 | .627 | .555 | 1 | .456 | 1.595 | .467 | 5.450 |
| | x16 | -.271 | .439 | .381 | 1 | .537 | .763 | .323 | 1.803 |
| | x17 | .667 | .509 | 1.720 | 1 | .190 | 1.949 | .719 | 5.283 |
| | x18 | 1.119 | .430 | 6.780 | 1 | .009 | 3.063 | 1.319 | 7.112 |
| | x19 | -.371 | .445 | .697 | 1 | .404 | .690 | .288 | 1.649 |
| | x20 | -.666 | .618 | 1.161 | 1 | .281 | .514 | .153 | 1.725 |
| | Constant | .098 | 1.441 | .005 | 1 | .946 | 1.103 | | |

a. Variable(s) entered on step 1: x1, x2, x5, x6, x7, x9, x10, x11, x14, x15, x16, x17, x18, x19, x20.

The table indicates the variable x18 has a significance value (sig = 0.009), less than 0.05, showing that the likelihood of drug addiction increases due to smoking cigarettes or hookah, with an odds ratio of (Exp(B) = 3.063).

The variable x17 has a significance value (sig = 0.039), which is also less than 0.05, indicating that the probability of drug addiction due to the level of education increases, with a probability ratio (Exp(B) = 1.957)., The other variables do not have a significant effect on drug addiction.

The results indicate that chronic diseases, imprisonment or detention, and smoking cigarettes or hookah are the most significant factors influencing drug addiction.

**Table (5)** Classification

| Observed | Predicted | | |
|---|---|---|---|
| | Not Addicted | Addicted | Percentage |
| Not Addicted | 26 | 33 | 44.1 |
| Addicted | 20 | 81 | 80.2 |
| Overall Percentage | | | 66.9 |

The table (5) represents the actual category, whether a person is addicted or not. The model predicted that 44.1% of people were not addicted, while 80.2% were predicted to be addicted.

## 8. Conclusions

These results are based on the findings of the study, which included a sample of patients, and after considering all the factors influencing addiction, the study revealed that the most significant causes of drug addiction are chronic diseases, imprisonment or detention, and smoking and hookah use, as these factors were found to be the major contributors to addiction.

## 9. Conclusion

The main points of this study involve the use of a binary logistic regression model to identify the key factors influencing drug addiction. The research analysed data from 160 patients attending the hospital for addiction treatment and examined the underlying causes. The results indicated that chronic diseases, imprisonment or detention, and smoking (cigarettes and hookah) are the most significant factors associated with drug addiction, while other demographic and social variables showed no statistically significant effect. These findings highlight the multifaceted nature of addiction and emphasise the importance of addressing both health-related and behavioural factors in prevention and treatment strategies.

## 10. Recommendations

Based on the findings of this study, the following recommendations are proposed:

1- Greater attention should be given to individuals suffering from chronic diseases, as they are more vulnerable to drug addiction and require targeted medical and psychological support.
2- Preventive and rehabilitative programs should be implemented for individuals exposed to imprisonment or detention to reduce the risk of addiction.
3- Awareness campaigns should be intensified to highlight the risks of smoking cigarettes and hookah as gateway behaviours leading to drug addiction.
4- Policymakers and health institutions are encouraged to strengthen preventive programs, family support systems, and community-based interventions aimed at reducing addiction rates, particularly among young people.

## 11. Supplementary Material

(None).

## 12. Author's Contributions

Suhad Ahmed Ahmed: Designed the research and developed the conceptual framework. Wrote and edited the manuscript. Conducted the statistical analyses using SPSS. Interpreted the results and contributed to the discussion.

## 13. Funding

(None).

## 14. Data Availability Statement

The data used in this study were collected through questionnaires distributed to patients at Baghdad Medical City Hospital during field visits. Information was gathered regarding the reasons for visits and the number of times the patients had visited.

## 15. Acknowledgements

The author would like to express their sincere gratitude to Ibn Rushd Teaching Hospital for Psychiatry for facilitating data collection. They also extend their thanks to all the doctors who contributed their time and information to support the completion of this research.

## 16. Conflict of Interest

The authors declare no conflict of interest.

## References

[1] Al-Rawi, Khasha'a Mahmoud (1987). "Introduction to Regression Analysis." Directorate of Dar Al-Kutub for Printing and Publishing / University of Mosul.
[2] Abd Al-Mun'im, Dr. Tharwat Mohammed (2005). "Regression Book." Anglo-Egyptian Library.
[3] Anthony Hayter, Wei Liu, & Chanita Tantipoj (2022). "Binary Response Analysis Using Logistic Regression in Dentistry." Department of Anatomy, Faculty of Dentistry, Mahidol University, Bangkok 10400, Thailand.
[4] Hajar, F. T., & Ibrahim, W. S. (2019). Use the logistic regression model to find the most important factors affecting lung cancer in Iraq in 2017. Journal of Administration and Economics, 44, 121.

[5] Ibrahim, W. S. (2020). To determine the most important factors affecting pancreatic cancer in Iraq using the logistic regression model. Periodicals of Engineering and Natural Sciences (PEN), 8(1), 277-283. DOI: 10.21533/pen.v8i1.1189

[6] Irfan, P., & Heman, D. (2006). "A Logistic Regression Analysis of Ischemic Heart Disease." Journal of Applied Sciences. o DOI: 10.3923/jas.2006

[7] Mohammed, A. J., & Ibrahim, W. S. (2020). Study of the probability of change in share prices based on the value of the Babylon Hotel using logistic regression. Journal of College of management and Economics, 43(123), 434-446.

[8] Sekulovski, N., Bartoš, F., Bergh, D., Arena, G., Godmann, R., Goyal, V., Julius M. Pfadt, Marsman, M., & Adrian E. Raftery (2025). "Comparing Variable Selection and Model Averaging Methods for Logistic Regression." December 1. 10.31234/osf.io/7g29a.

[9] Ghanem, Dr. Adnan, & Al-Jaouni, Dr. Fereed Khalil (2011). "Using Binary Logistic Regression Technique in Studying the Economic and Social Determinants of Household Income Sufficiency (An Empirical Study on a Random Sample of Families in Damascus Governorate)." Journal of Damascus University for Economic and Legal Sciences, Volume 9.

[10] Ibrahim, W. S., Majeed, G. H., & Anber, J. A. (2022). Determine the most important factors affecting breast cancer in Iraq. Periodicals of Engineering and Natural Sciences, 10(2), 260-266. https://doi.org/10.21533/pen.v10.i2.604

[11] Daeesh, A. Mohammed Amin, & Mohammed Sari (2017). "Logistic Regression Model: Its Concept, Characteristics, and Applications." Algeria.

[12] Hosmer, D. W., & Lemeshow, S. (2000). "Applied Logistic Regression" (Second Edition). New York: John Wiley & Sons, Inc. DOI: 10.1002/0471722146

[13] Talib, Haider Raed, & Mazloom, Sara Adel (2020). "Using Binary Logistic Regression Technique to Analyze Factors Affecting Heart Disease Mortality." Applied Study on a Sample of Patients in Dhi Qar Governorate. Journal of Al-Rafidain University College for Sciences, Issue (46). DOI: 10.55562/jrucs.v46i1

https://doi.org/10.31272/jae.i151.1538
https://admics.uomustansiriyah.edu.iq
P-ISSN: 1813–6729 E-ISSN: 2707-1359

# تحليل الانحدار اللوجستي الثنائي لدراسة العوامل المؤثرة على الإدمان

**سهاد أحمد أحمد**
قسم علوم الحاسوب، كلية التربية للعلوم الصرفة (ابن الهيثم)، جامعة بغداد، بغداد، العراق.
Email: suhad.a.a@ihcoedu.uobaghdad.edu.iq, ORCID: https://orcid.org/0000-0002-0072-6990

## المستخلص

*تعد مشكلة الادمان على المخدرات من اخطر المشاكل الصحية والاجتماعية والنفسية التي تواجه العالم.*

*لذا سيتم في هذا البحث دراسة تأثير العوامل المؤثرة في الادمان باستعمال أنموذج الانحدار اللوجستي حيث صنفت الحالات الى (مدمن، غير مدمن) حسب تفاوت مستوى حالة الادمان وعدد مرات مراجعته للمستشفى والتي تمثل متغير الاستجابة اما المتغيرات المستقلة تمثلت بالعوامل المؤثرة على الادمان.*

*تتكون العينة من (160) حالة ولكلا الجنسين ومختلف الاعمار بموجب استمارة استبيان تتكون من (20) سؤال حول اسباب الادمان، اثبتت النتائج اهم الاسباب المؤدية الى الادمان الاصابة بالامراض المدمنة والحكم بالسجن اوالتوقيف وكذلك بسبب التدخين والاركيلة حيث تم تطبيق البرنامج الاحصائي SPSS في ايجاد النتائج.*

## معلومات البحث

**تواريخ البحث:**
التقديم: 12 / 12 / 2025
المراجعة: 28 / 01 / 2026
قبول النشر: 03 / 02 / 2026
نشر الكتروني: 01 / 03 / 2026
تسلسل الصفحات: 142 – 149

**الكلمات المفتاحية:**
ثنائي الاستجابة، أنموذج الانحدار اللوجستي ، استمارة استبيان ، الادمان.

**المراسلة:**
أسم الباحث: سهاد أحمد أحمد
Email:
suhad.a.a@ihcoedu.uobaghdad.edu.iq