

Simplest Description of Binary Logit Model

Dr. Rashmi

Assistant Professor
D.A.V. College, Bathinda, Punjab

Abstract: *This article makes an attempt to explain the logit model in a very simple manner and suggests how it is applied to the qualitative data. Logit model is explained with the help of an example which contains data from a college. Pass/faliure status of the students is determined which is the dependent variable of the study. Marks of the students in internal exams, gender of the students and background of the students is taken as independent variables. E-Views is used to run the model and finally the results of the example are made clear.*

Keywords: *example, logit, marks*

1. SIMPLEST DESCRIPTION OF BINARY LOGIT MODEL

The past few decades have seen a quiet revolution in the development of quantitative business tools. No longer the province of back-office operations, new techniques have emerged for solving million and even billion-dollar problems.

In recent years, discrete-choice analysis has emerged as a powerful tool in the market-research. Based on consumers' hypothetical or real-world choices, discrete-choice models help the product managers in the following tasks.

- Positioning products in a competitive marketplace
- Developing strategic and tactical pricing strategies
- Sizing market for the new product
- Determining the optimal mix of product features
- Isolating market segments based on product preferences
- Estimating brand values among competing products

The various tools of discrete-choice model are binary logit, binary prbit, multinomial logit model, multinomial probit model.

2. BINARY LOGIT

Logit model is used to model the relationship between a dependent variable Y and one or more independent variables X. The dependent variable, Y, is a discrete variable that represents a choice or category from a set of mutually exclusive choices or categories. For instance, an analyst may wish to model the choice of travel mode. The various means of transport may be airways, rail, bus, etc. Now the choice of travel mode may depend upon several factors like cost of mode, time availability, facilities provided by the mode of transport and demographics factors, etc. If one assumes that there are only three modes of travel that is airways, bus and rail. Then the consumer has the following choices of travel.

- Travel by airways or bus
- Travel by airways or rail
- Travel by rail or bus.

Such type of model is known as multinomial logit model where more than two choices are available. Such a model takes the form of binary logit when Y represents only two choices. For example,

whether to buy or not to buy, whether to operate or not, etc. In such cases, there are only two options, out of which respondent selects only one.

To operate such models, one of the choice is taken as base category and if the respondent belongs to that base category, he/she is assigned a value of '0' and if he/she belongs to category other than base category, he/she gets a value of '1'. For example, in the above instance, there is a choice between 'buy' or 'not to buy'. If we set 'not to buy' as base category, then respondents who 'buy' get the value of '1' and who mark 'not to buy', get '0' value. Binary logit model takes the following form:

$$\text{Log} [P_i/1 - P_i] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \mu_n$$

where α is constant; X_n is a vector of independent variables; β is the coefficient of X_n ; μ_i refers to the error term which captures all misspecifications associated with a given individual and alternative. P_i is the probability of buying; and $1 - P_i$ is the probability of not buying. $\text{Log} [P_i/1 - P_i]$ is the log-odds ratio.

3. ASSUMPTIONS OF BINARY LOGIT MODEL

The various assumptions of binary logit model are discussed as under:

1. The observations on dependent variable Y are randomly sampled from the population of interest.
2. There is uncertainty in the relation between Y and X's, as reflected by a scattering of observations around the functional relationship.
3. Y is caused by or associated with the X's and X's are determined by influences (variables) 'outside' of the model.
4. The distribution of error terms must be assessed to determine if a selected model is appropriate.

4. INPUTS FOR BINARY LOGIT MODEL

As already stated, it is a binary choice model, thus Y will have only two type of values either '1' or '0'. On the other side, X may take continuous or discrete value. X describes the attributes of the choice maker or event and/or various attributes of the choices thought to be causal or influential in the decision or classification process.

5. OUTPUT OF BINARY LOGIT MODEL

- Functional form of relation between Y and X's.
- Strength of association between Y and X's.
- Probability of choosing a particular option by the respondent (if marginal effects are calculated).
- Comparison in terms of how many times a particular respondent will do a particular task in comparison to other respondent (if odds ratio is calculated).

Let us understand the logit model with the help of an example. Suppose we want to determine pass/failure status of the students of MBA in the final exams. The results/performance of the students will depend upon a large number of factors like their marks in internal exams, internal assessment, intelligence level, attendance status and even on gender also as it is generally said that girls more concentrate on their studies as compared to boys.

In order to understand that how logit model is run in E-Views, data has been collected from a college situated in New Delhi¹. Following information has been collected.

- 1) Pass/failure status of the students of MBA in final exams.
- 2) Marks of the students in internal exams.
- 3) Gender of the students.
- 4) Background of the students that is, either they are commerce graduates, +2 commerce or non-commerce

The other variables are ignored to reduce the complexity involved in the system.

¹ For maintaining secrecy, name of the college is not disclosed.

6. DEPENDENT VARIABLE

In this analysis, we have to determine the pass/failure status of the students which is a dependent variable. Now the students will either pass or fail. Thus there are only two choices, 1) pass and 2) fail. Thus this analysis will take the form of binary logit because of only two choices. In case of qualitative analysis, when there are only two options, we use the dummy variable. One thing which must be remembered while using dummy variable is to take $m-1$ variables where m represents the number of choices. In the present case, there are only two choice that is, either pass or fail. Thus only one dummy variable will be used which will depict both status of the students. Now to cover up both these status, one choice is taken as base category. For example, in this case, fail status is taken as base category. Further pass/fail status is qualitative aspect. It has no numerical value. To convert the qualitative aspects into quantitative terms, the choice option which is taken as base category is assigned a value of zero and the other category, which is pass status is assigned value one. This variable is denoted as 'Result'.

7. INDEPENDENT VARIABLES

The independent variables used in this case to have a clear understanding of logit model are discussed as under.

8. MARKS OF THE STUDENTS IN INTERNAL EXAMS

The marks obtained by students in their internal exams were collected from the college. Obviously the students who perform well in internal exams are going to secure good marks in their final exams. The marks of the students are in quantitative terms. Thus the value of this variable was used as it was. This variable is denoted as 'Internal exams marks'.

9. GENDER OF THE STUDENT

The gender of the students was asked. Again gender is a qualitative aspect. The student will either a male or a female. Thus there are only two choices. In case of two choices, only one variable will be used which will depict both genders. Here female is taken as base category. Thus if a student is female, she will get a value of zero. On the other side, if the student is male, he will be assigned a value of one. This variable is denoted as 'Gender'.

10. BACKGROUND OF THE STUDENTS

Background of the students' means that what was the stream of a MBA student in his/her graduation. There are three choices to this question, 1) BBA, B.Com means commerce at graduation level, B.Sc, BCA means no-commerce at graduation level, 3) commerce studied at +2 level. It is obvious that if the students have done commerce at either +2 level or in their graduation, they are going to perform better in their final exams of MBA. Now it is qualitative variable and has three choices. Thus two dummy variables will be introduced. Students who have neither done commerce at +2 nor at graduation is taken the base category. Two dummy variables introduced are graduate commerce students and +2 commerce students. Base category means graduate commerce students and +2 commerce students will be compared with non-commerce students. The student who is commerce graduate or +2 commerce will be allotted the value of 1. In case when both of these categories have value 0, it means that student belongs to non-commerce category.

11. DATA COLLECTION AND ANALYSIS

For proper understanding of logit model, data regarding 48 students of MBA of a college is obtained. However to maintain secrecy, the whole data is not disclosed. A part of the data is shown in Table 1 to have a clear perception of the variables.

Table1. *Data of the students*

Sr. No.	Result of the students	Internal exam marks (out of 40)	Gender	Background of the student
1	Fail	15	Male	+2 commerce
2	Pass	22	Female	+2 commerce
3	Pass	24	Male	Commerce graduate
4	Fail	12	Female	Non-commerce
5	Pass	23	Female	Commerce graduate
6	Pass	23	Female	Non-commerce

7	Pass	25	Female	+2 commerce
8	Pass	19	Female	Commerce graduate
9	Fail	13	Male	Non-commerce
10	Pass	25	Male	Commerce graduate
11	Fail	12	Male	Non-commerce
12	Pass	28	Male	Commerce graduate

As explained above, the above mentioned qualitative data is converted into quantitative terms and is shown in Table 2.

Table 2 shows how the qualitative data is to be converted into quantitative terms.

Table2. Data in Quantitative Terms

Sr. No.	Result of the students	Result	Internal exam marks (out of 40)	Gender	Gender	Background of the student	Commerce graduate	+2 commerce
1	Fail	0	15	Male	1	+2 commerce	0	1
2	Pass	1	22	Female	0	+2 commerce	0	1
3	Pass	1	24	Male	1	Commerce graduate	1	0
4	Fail	0	12	Female	0	Non-commerce	0	0
5	Pass	1	23	Female	0	Commerce graduate	1	0
6	Pass	1	23	Female	0	Non-commerce	0	0
7	Pass	1	25	Female	0	+2 commerce	0	1
8	Pass	1	19	Female	0	Commerce graduate	1	0
9	Fail	0	13	Male	1	Non-commerce	0	0
10	Pass	1	25	Male	1	Commerce graduate	1	0
11	Fail	0	12	Male	1	Non-commerce	0	0
12	Pass	1	28	Male	1	Commerce graduate	1	0

E-Views has been used to determine the logit results of above mentioned students. Following results are obtained.

Table3. Results of Logit Model (Dependent variable: Results of the students)

Variable	Parameter estimate	T-ratio
Constant	5.182	2.696*
Internal exam marks	3.107	2.745*
Gender	-2.817	2.171**
+2 Commerce	1.288	2.784*
Commerce graduate	2.857	2.106**

Note: *** means significant at 1% and 5% respectively.

Table 3 shows that all variables are significant. A new variable named constant is added while analysing. No model can be run without constant variable. Coefficient value of constant states that by how much amount the dependent variable changes, if the value of independent variables is zero.

The coefficient value of internal exams marks is positive which means that if the students have obtained more marks in internal exams, there is more probability that they will pass in their final exams.

Further gender is possessing negative value. Gender is a dummy variable which means either the respondent is male or female. In this study, female is taken as base category which means result of the students is calculated for males as compared to females. Thus if the value of coefficient is positive, it means there is more probability that males will pass as compared to females. Whereas if the coefficient value is negative, it means there is more probability that females will pass as compared to males.

Remaining two variables are +2 commerce and commerce graduates. These two variables represent the background of the students which is measured in three categories: +2 commerce, commerce

Simplest Description of Binary Logit Model

graduate and non-commerce. Non-commerce is taken as base category. The remaining two variables +2 commerce and commerce graduates are compared with non-commerce category. The coefficient value of these two variables is positive. It means as compared to non-commerce students,+2 commerce students have more probability that they will pass. Same is true for commerce graduate students. Further coefficient value of commerce graduate students is more than +2 commerce students, which means there is more probability that commerce graduate students will pass as compared to +2commerce students.

12. CONCLUSION

Logit model is simplest method for an analysis of qualitative data. Software like E-Views and Microfit has an option of logit model. These software give a value of marginal effect. This marginal effect can be multiplied with the coefficient value of variables to obtain the probability value. Thus exact probability of the change in value of variables can also be obtained to have a clear perception of the problem.

REFERENCES

Gujarati, D.N. (2004), *Basic Econometrics*, Ed. 4th, New Delhi, Tata McGraw-Hill Publishing Company Limited.

<http://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/v2chapter5.html>

AUTHOR'S BIOGRAPHY



Dr. Rashmi, is working as an Assistant Professor in D.A.V. College, Bathinda. She is M.Com, PhD. Her one book titled *Consumer Choice Process: An Empirical Study of Indian Consumers* is released. Her **fourteen** articles have been published in Vikalapa The Journal of Decision Makers, IUP Journal of Marketing Management, Asia Pacific Business Review, XLRI Management and Labour Studies, Management and Change, Journal of Management and Information Technology, Vishwakarma Business Review (2) IITM Journal of Management and IT, Global Advanced Research Journal of Management and Business Studies, Tecnia Journal of Management Studies, Gian Jyoti E-Journal, Journal of Research in Marketing, SOP Transaction on Marketing Research. Her two articles are in process of communication and recently she is working on another article. She has participated in national conferences and also one paper has been presented at an international conference. She has two years and eight months of research experience and four years of teaching experience. She has acted as a reviewer for Journal of Promotion Management, SOP transactions on Marketing Research and Journal of Social Science Research. She is editorial board member of SOP transactions on Marketing Research, Global Research Journal of Business Management, Global Research Journal of Marketing Management, Blue Ocean Research Journal.