

[Frequently Asked Questions]

Regression Analysis for Dummy Dependent Variable

Subject:

Business Economics

Course:

Paper No. & Title:

B. A. (Hons.), 5th Semester, Undergraduate

Paper – 531 Elective Paper Q1 – Advanced Econometrics

Unit No. & Title:

Unit – 4 Binary Data and Limited Dependent Variable Models

Lecture No. & Title:

Lecture – 2 Regression Analysis for Dummy Dependent Variable

Frequently Asked Questions

Q1. What is meant by dummy dependent variable?

A1. Many times the dependent variable occurs as a qualitative variable.

e.g. $Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + U_i$

Here $Y_i = 1$ If the person is employed

= 0 Otherwise

X = income of the person

 $D_i = 1$ for Male

= 0 for Female

Q2. What is LPM?

A2. When the dependent variable has two categories, it is probabilistic in nature. Here Y as well as stochastic disturbance term U has its probability distribution.

Q3. What is the estimation problem for LPM?

A3. (i) since dependent variable *Y* has only two categories, normality assumption would be incorrect.

(ii) Disturbances are heteroscadastic in nature. Hence directly OLS cannot be applied.

(iii) When we find estimated value of Y from the fitted model, it is only either 0 or approximated as 1. This is a limitation of the model.

Q4. What is logistic regression?

A4. Regression based upon LOGIT ratio can be called logistic regression.

We consider c.d.f. of logistic distribution and define the odd ratio

 $L_i = L_n \left(\frac{P_i}{1-P_i}\right)$, then the regression $L_i = \beta_1 + \beta_2 X_i + U_i$ is termed as logistic regression.

Q5. What are features of LOGIT model?

A5. (i) LOGIT (denoted by L_i) is the natural logarithms of odds ratio (i.e. Prob. of success divided by probability of failure).

(ii) Thus L is linear in X but the probabilities by themselves are not.

(iii) We can include many explanatory variables in the model by defining the logit ratio accordingly.

Q6. How would you estimate LOGIT model?

A6. For individual data, we have to use maximum likelihood method.

For grouped data, we can estimate P_i by $\hat{P}_i = p_i = \frac{n_i}{N_i}$

 n_i = Sample observations (Number of families owning house) N_i = Population observations (Number of families in the population)

Then we write $\hat{L}_i = L_n \left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Since $V(U_i) = \frac{1}{NP_i(1-P_i)}$, there is heteroscadasticity, we can transform the model by dividing both the sides by $V(U_i)$ and hence run regression for the transformed model.

The estimates of Y will be represented in terms of logarithm of odds ratio.

Q7. What is PROBIT model?

A7. Unlike LOGIT model, here we use the concept of c.d.f. of normal distribution and then define utility Index I_i and write $I_i = \beta_1 + \beta_2 X_i$. Hence we can estimate this model.

Q8. What are further extensions for this special category of models?

A8. They are called TOBIT model, Poisson Regression Model, ordinal logit and probit models, Multinomial logit and probit models etc.