



[Frequently Asked Questions]

Regression Analysis for Qualitative Variables

Subject:	Business Economics
Course:	B. A. (Hons.), 5 th Semester, Undergraduate
Paper No. & Title:	Paper – 531 Elective Paper Q1 – Advanced Econometrics
Unit No. & Title:	Unit – 4 Binary Data and Limited Dependent Variable Models
Lecture No. & Title:	Lecture – 1 Regression Analysis for Qualitative Variables

Frequently Asked Questions

Q1. Explain why dummy variables are necessary.

A1. To deal with qualitative data, we need to transform them into quantitative data. This is done by means of dummy variables.

e.g. Sex → Male or Female

We put $D = 1$ if Male

$= 0$ if Female

Thus we have a dummy variable D .

Q2. In practice, how dummy variables occur?

A2. Dummy variables can occur in different ways.

- (1) One dummy explanatory variable.
- (2) One dummy and one or more explanatory variables.
- (3) Dummy dependent variable.
- (4) Mixture of dummy dependent as well as dummy independent and explanatory variables etc.

Q3. What is dummy variable trap?

A3. If dummy variable is not properly defined, we have multicollinearity problem.

Rule is: If a qualitative variable has m categories (or classes) then we should include $(m-1)$ dummy variables. If this is not observed, we get dummy variables trap.

Q4. What are ANOVA model and ANCOVA model in reference to categorical variables?

A4. ANOVA (Analysis of variance) model contains regression of a dependent quantitative variable in terms of the regression with dummy explanatory variables only.

ANCOVA (Analysis of Covariance) model contains regression of a dependent quantitative variable in terms of a mixture of both dummy independent and explanatory variables.

Q5. How pooling of cross-section data and time series can be done using dummy variables?

A5. We can mention them as Sample I and Sample II respectively with n_1 and n_2 observations.

It is assumed that both have different intercepts but the same slope.

We can consider combined sample of size $(n_1 + n_2)$ and assign dummy variable representing Sample I and Sample II respectively. This has same slope as the two earlier regressions.

If we regress to third model for combined samples, we can obtain unbiased and efficient estimator for slope co-efficient. Hence use of dummy variables is advantageous.

Q6. How interaction effect can be represented using dummy variables?

A6. With one dummy and one explanatory variable the interaction effect can be represented by means of product of dummy with explanatory variable in the case of additive model.

$$\text{e.g. } y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + U_t$$

In the case of multiplicative model, the interaction effect can be represented by means of product of dummy variables.

$$\text{e.g. } y_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 (D_{2t} \cdot D_{3t}) + \beta X_t + U_t$$

Q7. How dummy variables method is useful in deseasonalising time series data ?

A7. We have 4 quarters and n years, thus having 4n observations. We define one dummy variable for each quarter so that we have a model $\underline{y} = D\underline{b} + \underline{y}^\alpha$

Where $\underline{y} : 4n \times 1$, $\underline{b} = 4 \times 1$, $D = 4n \times 4$, $\underline{y}^\alpha = 4n \times 1$

\underline{y}^α Represents disturbances.

We can find $\underline{b} = (D'D)^{-1} D'\underline{y}$

then $\underline{y}^\alpha = \underline{y} - D\underline{b} = M\underline{y}$

Where $M = [I - D(D'D)^{-1}D']$ is symmetric idempotent matrix.

\underline{y}^α is the deseasonalised series.

We can consider extension of this model to overcome the drawbacks of this method.

Q8. What is piecewise regression?

A8. Very often the sales commission given to sales persons depends upon their sales performance. There is a target for sales called threshold value. Before target, there is one stage and when the target is achieved and sales increase, it is second stage. We can have separate regression for each stage. Instead of that we can use dummy variables method which can help to consider both the stages together. This is called piecewise regression.

e.g. $Y_i = \alpha_1 + \beta_1 X_t + \beta_2 (X_i - X^*) D_i + U_i$

Where $D_i = 1$ if $X_i > X^*$

$= 0$ if $X_i < X^*$

$X^* =$ threshold value, D is dummy variable.