UNIT 9 EXTENSION OF REGRESSION MODELS: DUMMY VARIABLE CASES*

Structure

9.0 OBJECTIVES

After reading this unit, you will be able to:

- define a qualitative or dummy variable;
- discuss the ANOVA model with a single dummy as exogenous variable;
- specify an ANCOVA model with one quantitative and one dummy variable;
- interpret the results of dummy variable regression models;
- differentiate between 'differential intercept coefficient' and 'differential slope coefficient;
- describe the concepts of 'concurrent, dissimilar and parallel' regression models that you encounter while considering 'differential slope dummies'; and
- explain how more than two dummies and interactive dummies can be formulated into a regression model.

9.1 INTRODUCTION

In real life situations, some variables are qualitative. Examples are gender, choices, nationality, etc. Such variables may be dichotomous or binary, i.e., with responses limited to two such as in 'yes' or 'no' situations. Or they may have more than two categorical responses. We need methods to include such variables in the regression model. In this unit, we consider some such cases. We limit this unit to consider regressions in which the dependent variable is quantified. You may note in passing that when the dependent variable itself is a dummy variable, we have to deal with them by models such as Probit or Logit. In such models, the

* Dr. Pooja Sharma, Assistant Professor, Daulat Ram College, University of Delhi and Prof. B S Prakash, Indira Gandhi National Open University, New Delhi

OLS method of estimation does not apply. In this unit, we will not consider such cases. You will study about them in the course 'BECE 142: Applied Econometrics'.

In this unit, we consider only such cases in which the independent variable is a dummy variable. Qualitative variables are not straightaway quantified. By treating them as dummy variables we can make them quantified (or categorical). For instance, consider variables such as male or female, employed or unemployed, etc. These are quantifiable in the sense that by treating them as 1 if 'female, and 0 if 'male'. Similar examples could be 1 if yes and 0 if no; 1 if employed and 0 if unemployed, etc. In the above, we have converted a qualitative response into quantitative form. Thus, the qualitative variable is now quantified. Such regressions could be a simple regression, i.e., there is only one independent variable which is qualitative and treated as dummy variable. Or there could be two independent variables, one of which can be treated as dummy and the other is its covariant, i.e., there is a close relationship with the variable treated as dummy. For instance, pre-tax income of persons can be classified above a threshold level and treated as dummy variable, i.e., above or below the threshold level income with response taken as 1 or 0. Now, the post-tax income, which is a co-variant of pre-tax income, can be considered by its actual quantified value. There could be similar extension of situations where you have to consider multiple dummies and cases where you have to consider interactive dummies. The nature of such regressions, particularly for their inference or interpretational interest, is what we consider in the present unit.

9.2 THE CASE OF SINGLE DUMMY: ANOVA MODEL

We first consider a simple regression model with only one independent variable. Further, this independent variable is a dummy variable such as:

$$Y_i = \beta_1 + \beta_2 D_i + u_i \qquad \ldots (9.1)$$

Here, we take Y as the annual expenditure on food and D_i as gender taking the values 0 if the person is male and 1 if female. The D_i's are thus fixed and hence non-stochastic. Now, if we assume that $u_i \sim N(0, \sigma^2)$, the OLS method can be applied to estimate the parameters in (9.1). If we do this, the mean food expenditure for males and females are respectively given by:

$$E(Y_i \mid D_i = 0) = \beta_1 + \beta_2(0) = \beta_1 \qquad \ldots (9.2)$$

$$E(Y_i \mid D_i = 1) = \beta_1 + \beta_2 \qquad \ldots (9.3)$$

Here, β_1 gives the average or mean food expenditure of males. It is the category for which the dummy variable is given the value 0. The slope coefficient β_2 tells us by how much the mean food expenditure of females differ from that of the mean food expenditure of males. Hence, $\beta_1 + \beta_2$ gives the mean food expenditure for females. In view of this, it is not correct to call β_2 as the slope coefficient since there is no continuous regression line here. Hence, β_2 is the 'differential intercept coefficient'. It tells us by how much the value of intercept term differs between the two categories. A question that arises now is, what would have happened if we had interchanged the assignment of '0' between the two categories of males and females (i.e., if we had assigned the value '0' to females). You may note that, so long as we have only two categories as in the present instance, i.e., it is a case of simple regression with only one independent variable taken as a dummy variable D_i with the category of responses dichotomous or binary, it basically does not matter which category gets the value of 1 and which gets the value 0. However, some minor difference would be there. Let us see what this is.

The category to which we assign the value 0 is called as the base category. It is also called by alternative names such as reference or benchmark or the comparison category. In such an assignment, the intercept value represents the mean value of the category that gets the value 0 (which is males in our case above). What equation (9.3) tells us is, depending on such an assignment, the mean value of expenditure on food for females is to be obtained by adding the 'slope coefficient to the intercept value'. If the assignment of dummy is made the other way, i.e., females 0 and males 1, we see a change in the numerical value of the intercept term and its t value. Barring this, the R^2 value, the absolute value of the estimated dummy variable coefficient and its standard error, will remain the same. Let us see this with the help of an example for better understanding.

Consider the data on 'expenditure on food' and income for males and females as in Table 9.1. The data are averages based on the actual number of people (who are in thousands) in different age groups. We first construct Table 9.2 from the data in Table 9.1 as below.

Table 9.1: Data on Income and Food Expenditure by Gender

(Figures in $)

Age	Food Expenditure (female)	Income (female)	Food Expenditure (male)	Income (male)
< 25	1983	11557	2230	11589
25-34	2987	29387	3757	33328
35-44	2993	31463	3821	36151
45-54	3156	29554	3291	35448
55-64	2706	25137	3429	32988
> 65	2217	14952	2533	20437

Source: Table 6-1, Chapter 6, Gujarati.

Table 9.2: Food Expenditure in Relation to Income and Gender

Observation	Food Expenditure ($)	Income ($)	Gender
1	1983	11557	1
2	2987	29387	1
3	2993	31463	1
4	3156	29554	1
5	2706	25137	1
6	2217	14952	1
7	2230	11589	0
8	3757	33328	0
9	3821	36151	0
10	3291	35448	0
11	3429	32988	0
12	2533	20437	0

Source: Table 6-2, Chapter 6, Gujarati.

Results of food expenditure regressed on the gender dummy variable (without taking into account the income variable at this stage) presents the following results.

$$\hat{Y}_i = 3176.833 - 503.1667\, D_i$$

$$se = (233.0446) \qquad (329.5749)$$

$$t = (13.6318) \qquad (-1.5267) \qquad R^2 = 0.1890$$

The results show that the mean expenditure of males is 3177 $ and that of females is (3177 – 503 = 2674 $). The estimated D_i is not statistically significant (since its t value is only –1.53). This means that the difference in the food expenditure between gender is not statistically significant. Recall that we have assigned the value '0' to males. Hence, the intercept value represents the mean value for males. In this assignment, to get the mean value of food expenditure of females, we add the value of the coefficient of the dummy variable to the intercept value. Now, let us re-assign the value '0' to females and '1' to males. The regression results that we get are the following:

$\hat{Y}_i =$ 2673.667 + 503.1667 D_i

se = (233.0446) (329.5749)

$t =$ (11.4227) (–1.5267) $R^2 = 0.1890$

Thus, we notice that the mean food consumption expenditures of the two genders have remained the same. The R^2 value is also the same. The absolute value of the dummy variable coefficient and their standard errors are also the same. The only change is in the numerical value of the intercept term and its t value.

Another question that we may get is: since we have two categories, male and female, can we assign two dummies to them? This means we consider the model as:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_i + u_i \quad \ldots (9.4)$$

where Y is expenditure on food, $D_2 = 1$ for female and 0 for male and $D_3 = 1$ for male and 0 for female. Essentially, we are trying to see whether we can assign two dummies for male and female separately? The answer is 'no'. To know the reason for this, consider the data for a sample of two females and three males, for which the data matrix is as in Table 9.3. We see that $D_2 = 1 - D_3$ or $D_3 = 1 - D_2$. This is a situation of perfect collinearity. Hence, we must always use only one dummy variable if a qualitative variable has two categories, such as the gender here.

Table 9.3: Data Matrix for the Equation

Gender	Intercept	D_2	D_3
Male Y_1	1	0	1
Male Y_2	1	0	1
Female Y_3	1	1	0
Male Y_4	1	0	1
Female Y_5	1	1	0

A more general rule is: if a model has the common intercept β_1, and the qualitative variable has m categories, then we must introduce only $(m - 1)$ dummy variables. If we do not do this, we get into a problem of estimation called as the 'dummy variable trap'. Finally, note that when we have a simple regression model with only one dummy variable as considered here, the model considered is also called as the ANOVA model. This is because there is no second variable from which we are seeking to know the impact or variability on the dependent variable. When we have this, we get what we call as an ANCOVA model. We take up such a case in the next section.

9.3 ANALYSIS OF COVARIANCE (ANCOVA) MODEL

In economic analysis, it is common to have among explanatory variables some of which are qualitative and some others quantitative. Such models are called as Analysis-of-Covariance (ANCOVA) models. Here, we shall consider a model that has both a quantitative and a dummy variable among the regressors. In general, regression models containing a combination of quantitative and qualitative variables are called ANCOVA models. Here, the quantitative variables are called covariates or control variables. ANCOVA models are an extension of the ANOVA models. They provide a method of statistically controlling the effects of covariates (i.e., a quantitative explanatory variable) in a model that includes both the type of variables with the qualitative variable treated as a dummy variable. The quantitative variable considered is usually a covariate in the sense that it bears close association with the main variable. Because of this, exclusion of covariates from a model results in model specification error. In the example considered above, we regressed 'food expenditure' on only gender dummy [$Y_i = \beta_1 + \beta_2 D_i + u_i$]. Now, let us consider another variable, 'income after taxes', i.e., disposable income (a covariate of food expenditure) as an explanatory variable (X_i). The model now is

$$Y_i = \beta_1 + \beta_2 D + \beta_3 X_i + u_i \qquad \dots (9.5)$$

where Y = expenditure on food (\$), X = after tax income (\$), D = 1 for female and = 0 for male. Let us now consider, for better appreciation, the result for the regression in equation (9.5) obtained from the data in Table 9.2 as follows:

$$\hat{Y}_c = 1506.244 - 228.9868 D_i + 0.0589 X_i$$
$$t = (8.0115) \qquad (-2.1388) \qquad (9.6417)$$
$$R^2 = 0.9284$$

The dummy variable coefficient is statistically significant. Therefore, we reject the null hypothesis that there is no difference in the average value of expenditure on food for male and female. In other words, we conclude that gender has a significant impact on consumption or food expenditure. Note that this difference in consumption expenditure is inferred holding the effect of after-tax income constant. Likewise, holding the gender differences constant, the after tax income coefficient is significant. The slope coefficient for 'after tax income' indicates that the mean food expenditure [i.e., the marginal propensity to consume (MPC)] increases by 6 cents for every additional dollar of increase in the disposable income. Note that since we have taken '0' for males, the intercept term relates to the MPC for males. For female MPC, we have to add the intercept value to the coefficient of gender dummy (i.e., 1506.2 – 228.9 = 1277.3). Thus, the equations for the MPC of females and males can be respectively written as:

Mean food expenditure for females: $\hat{Y}_i = 1277.2574 + 0.0589 X_i$

Mean food expenditure for males: $\hat{Y}_i = 1506.2440 + 0.0589 X_i$

Since the MPC or the slope is same for both the gender, the two regressions are parallel as in Fig. 9.1 below.

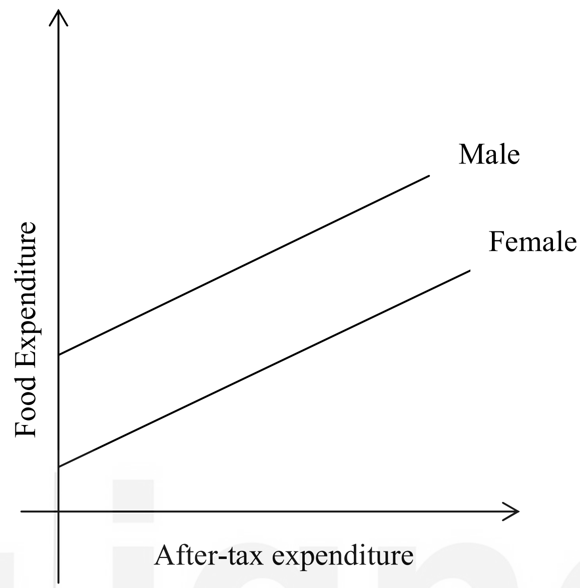


Fig. 9.1 Mean Food Expenditure for Male and Female

The model signifies the role and the impact of both the type of variables (quantitative and qualitative) in explaining a dependent variable. Specifically, in the example considered, the after tax expenditure is seen to affect the food expenditure of both males and females.

Check Your Progress 1 [answer questions in about 50-100 words]

1) Define a qualitative variable.

....

....

....

....

....

2) Specify a regression model with a single dummy variable. Mention its features from the point of view of interpretation of estimated coefficients.

....

....

....

....

....

3) What happens if the base value is reassigned for the dummy variable, say gender, in a simple regression model as in equation (9.1)?

....

....

....

....

....

4) What is meant by ‘dummy variable trap’? How do we avoid it?

....

....

....

....

....

5) Distinguish between an ANOVA model and an ANCOVA.

....

....

....

....

....

6) What is an advantage of ANCOVA model? What is a consequence of omitting the inclusion of a covariant in an ANOVA model?

....

....

....

....

....

7) Specify the general form of an ANCOVA model with one qualitative and one quantitative variable. What does the slope oefficient for the quantitative variable considerd indicate in general?

....

....

....

....

....

9.4 COMPARISON BETWEEN TWO REGRESSION MODELS

In the example considered above, i.e., for both the ANOVA and the ANCOVA models, we saw that the slope coefficients were same but the intercepts were different. This raises the question on whether the slopes too could be different? How do we formulate the model if our interest is to test for the difference in the slope coefficients too? In order to capture this, we introduce a 'slope drifter'. For the example of consumption expenditure for male or female considered above, let us now proceed to compare the difference in the consumption expenditure by gender by specifying the model with dummies as follows:

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i X_i) + u_i \qquad \ldots (9.6)$$

Note that the additional variable added is $D_i X_i$ which is in multiplicative or interactive form. In (9.6), we have taken $D_i = 0$ for males and $D_i = 1$ for females. Now, the 'mean food expenditure' for males is given by:

$$E(Y_i \mid D_i = 0, X_i) = \beta_1 + \beta_3 X_i \qquad \ldots (9.7)$$

{since $D_i = 0$}

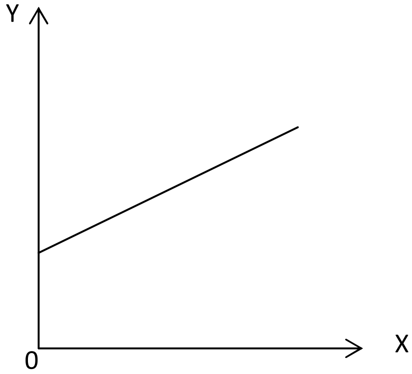
The 'mean food expenditure' for females is given by:

$$E(Y_i \mid D_i = 1, X_i) = \beta_1 + \beta_2 D_i + (\beta_3 + \beta_4 D_i) X_i$$

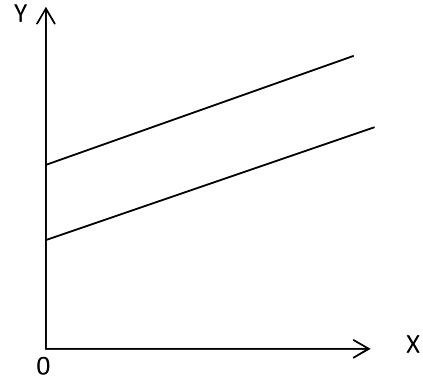
$$= (\beta_1 + \beta_2) + (\beta_3 + \beta_4) X_i \qquad \ldots (9.8)$$

{since $D_i = 1$}

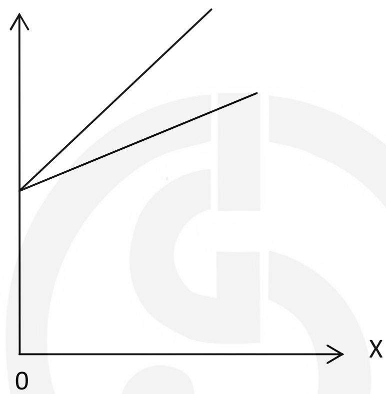
In equation (9.8), $(\beta_1 + \beta_2)$ gives the mean value of Y for the category that receives the dummy value of 1 when X is zero. And, $(\beta_3 + \beta_4)$ gives the slope co-efficient of the income variable for the category that receives the dummy value of 1. Note that the introduction of the dummy variable in the 'additive form' enables us to distinguish between the intercept terms of the two groups. Likewise, the introduction of the dummy variable in the interactive (or multiplicative) form (i.e., $D_i X_i$) enables us to differentiate between the slope coefficients (or terms) of the two groups. Depending on the statistical significance of the differential intercept coefficient, β_2, and the differential slope coefficient, β_4, we can infer whether the female and male food expenditure functions differ in their intercept values, or their slope values, or both. There can be four possibilities as shown in Fig. 9.2. Fig. 9.2 (a) shows that there is no difference in intercept or the slope coefficient of the two food expenditure regressions. Such regression equations are called 'Coincident Regressions'.



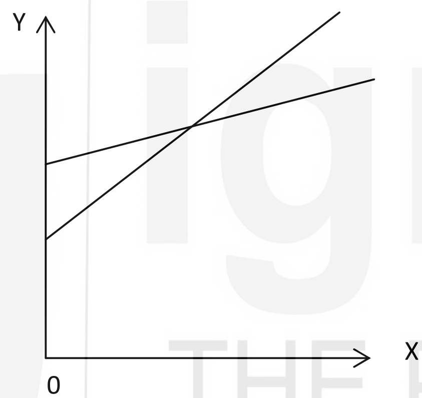
(a) Coincident Regressions



(b) Parallel Regressions



(c) Concurrent Regressions



(d) Dissimilar Regressions

Fig 9.2 Comparison of Regression Equations

Fig. 9.2 (b) shows that the two slope coefficients are the same but intercepts are different. Such regressions are referred to as 'Parallel Regressions'. Fig. 9.2 (c) shows that the two regressions have the same intercepts but

different slopes. Such regressions are referred as 'Concurrent Regressions'. Fig. 9.2 (d) shows that the two intercepts and the two slope coefficients are both different. Such regressions are called 'Dissimilar Regressions'.

9.5 MULTIPLE DUMMIES AND INTERRACTIVE DUMMIES

We often might require to consider more than one dummy variables. Besides, there could be cases where we might be interested in seeing for the impact of dummy variable interactions. Let us consider a case as given below.

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i \quad \dots (9.7)$$

where Y is income, X is education measured in number of years of schooling, D_2 is gender (0 if male, 1 if female), D_3 is if in reserved segment or group (e.g. SC/ST/OBC) taking the value 0 if 'not in reserved segment', i.e., in general segment and 1 if 'in reserved segment'. Here, gender (D_2) and reservation (D_3) are qualitative variables and X is quantitative variable. In this formulation (for example, equation 9.7) we have made an implicit assumption that the differential effect of gender is constant across the two segments of reservation. We have likewise assumed that the differential effect of reservation is constant across the two genders. This means if the average income is higher for males than for females, it is so whether the person is in the general segment or in the reservation segment. Likewise, it is assumed here that if the average income is different between the two reservation segments, it is so irrespective of gender. However, in many cases, such assumptions may not be tenable. This means, there could be interaction between gender and reservation dummies. In other words, their effect on average income may not be simply additive as in (9.7) but could be multiplicative. If we wish to consider for this interactive effect, we must specify the model as follows:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 (D_{2i}\, D_{3i}) + \beta_5 X_i + u_i \qquad \dots (9.8)$$

In equation (9.8), the dummy variable $D_{2i}D_{3i}$ is called as 'interactive or interaction dummy'. It represents the joint or simultaneous effect of two qualitative variables. Taking expectation on both sides of equation (9.8), i.e., by considering the average effect on income across gender and reservation, we get:

$$E(Yi \mid D_{2i} = 1, D_{3i} = 1, X_i) = \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 X_i \qquad \dots (9.9)$$

Equation (9.9) is the average income function for female reserved category workers where β_2 is the differential effect of being female, β_3 is the differential effect of being in the reserved segment and β_4 is the interactive effect of being both a female and in reserved segment. Depending on the statistical significance of various dummies, we need to make relevant inferences. The specification can easily be generalized for more than one quantitative variable and more than two qualitative variables.

Check Your Progress 2 [answer questions within the given space in about 90-100 words]

1) What is meant by a 'slope drifter'? When is it introduced and for what use? Specify a general model with such a 'slope drifter' and comment on the additional variable introduced.

....

....

....

....

....

2) Differentiate between the four type of regressions that we might get when considering a model of the type in equation (9.6) with two slope drifters β_2 and β_4 as therein.

....

....

....

....

....

3) List the four types of regression models, with dummy variables to accommodate different cases or situations, as we have considered in this unit. Specify their difference by their name and features.

....

....

....

....

....

9.6 LET US SUM UP

This unit makes a distinction between qualitative and quantitative variables. It has considered three types of models in which the focus is kept on inclusion of qualitative variables in the regression models. The first of such models is considered is a simple regression model. In this, we have considered only one dummy variable, as an independent variable, on the RHS of the regression equation. This equation is of the form: $Y_i = \beta_1 + \beta_2 D_i + u_i$. Analysis in this form is called as ANOVA. Quite often, we would be committing a specification bias if we consider the regression model in this form. This happens because the variable Y_i will be clearly related to a variable X_i which is a quantitative variable. To accommodate this, we considered the second type of model in which we included a co-variant (X_i) into the regression equation: $Y_i = \beta_1 + \beta_2 D + \beta_3 X_i + u_i$. Analysis in this form is called as ANCOVA. In both these type of models, our focus was only on observing the significance of difference in the intercepts. But in practice, we do encounter a number of situations in which not only the intercept, but the slope too could vary between categories. To allow for this kind of situation, we considered a third type of model in which we accommodated for the interactive effect of the 'dummy variable with the quantitative variable', i.e., $D_i X_i$. The regression model considered for this kind of an analysis is of the form: $Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i X_i) + u_i$. In this situation, we noted that we could come across four possibilities viz. coincidental, parallel, concurrent and dissimilar regressions. We have finally considered the case where a regression model may have to be formulated to accommodate more than one qualitative variable and a case where we might be interested in examining for the interactive effect of the two qualitative variables. For this, we considered models such as $Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 (D_{2i} D_{3i}) + \beta_5 X_i + u_i$.

9.7 ANSWERS/ HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) A qualitative variable is one which has a categorical response such as yes/no or employed/unemployed or male/female. If the response is limited to two, as in these cases, it is called as a dichotomous variable. The responses can be more than two. But they may be classified as 1, 2, 3, …….. Such responses are unambiguous or categorical. Hence, a qualitative variable is also called as dummy variable or categorical variable.

2) The model in this case can be $Y_i = \beta_1 + \beta_2 D_i + u_i$. We are considering the dependent variable Y_i as quantitative variable. The D_i's are thus fixed and hence non-stochastic. D_i is taken a dichotomous, i.e., it takes the values 0 and 1. In such cases, the factor or entity which is assigned the value 0, is called as the base category. The estimated value of the mean of Y_i, given $D_i = 0$, is given by β_1. Here, β_2 is not strictly the slope coefficient but is the 'differential intercept coefficient'. The estimated value of the mean of Y_i, given $D_i = 1$, is given by $\beta_1 + \beta_2$.

3) The mean value of Y_i for the two gender classes, the R^2 value, the absolute value of the estimated dummy variable coefficient and the standard errors will be the same. The numerical value of the intercept term and its t value will change.

4) The number of responses to the dummy variable is called as 'categories' of response. If the dummy variable refers to gender of the respondent, there are two categories of response viz. male and female. If we assign two separate dummies in such cases, we encounter a situation of perfect collinearity. Hence, we will not get unique estimates or one of the two parameters is not estimable. This situation is called as 'dummy variable trap'. To avoid this situation, the general rule is if we have m categories, we limit the number of dummies to '$m - 1$'. The models should also have a common intercept β_1.

5) If the regression model considered has only one independent variable in general, and that variable is a dummy variable as considered here in particular, then the variation or the sources of variability that is sought to be identified for the dependent variable is limited to that one variable. In such cases, the regression model considered is called as an ANOVA model. If the independent variables considered are two, with one considered as dummy variable, and the other variable considered is related to the dummy variable, then such models are called as ANCOVA model.

In other words, regression models in which some independent variables are qualitative and some others are quantitative, are called as ANCOVA models.

6) The advantage is that ANCOVA models provide a method of statistically controlling the effects of covariates. The consequence of excluding a covariant from being included in the model is that the model suffers from 'specification error'. The consequence of committing specification errors are that the ideal assumptions required for the OLS estimators to be efficient are violated. Consequently, they lose out on their efficiency properties.

7) The general form of the model is like: $Y_i = \beta_1 + \beta_2 D + \beta_3 X_i + u_i$. The slope coefficient indicates the rate of increase (or decrease) in the 'marginal propensity to consume (MPC)'. This is when the dependent variable Y relates to a consumption variable like expenditure on food and the quantitative independent variable is like disposable income as considered here.

Check Your Progress 2

1) In regression models with one intercept and one slope coefficients, our interest might be to test to know whether: (i) the intercept terms are statistically different and (ii) the slope coefficients are statistically different? For investigating the second question, we need to introduce what is called as a 'slope drifter'. The model specified with such a drifter would be like: $Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i X_i) + u_i$.The additional variable introduced here is *DiXi*. It is a multiplicative variable in the interactive form. Here β_2 and β_4 are the two slope drifters which helps us infer for the statistical difference in the intercept values and the slope values respectively.

2) We get a 'coincident regression' when there is no difference both in intercept as well as the slope. We get a 'parallel regression' when the two intercept terms are different but the two slope coefficients are the same. We get a 'concurrent regression' when the two regressions have the same intercept but different slopes. We get two 'dissimilar regressions' when both the intercept terms and the slope coefficients are different.

3) (i) $Y_i = \beta_1 + \beta_2 D_i + u_i$. (ii) $Y_i = \beta_1 + \beta_2 D + \beta_3 X_i + u_i$. (iii) $Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i X_i) + u_i$. (iv) $Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 (D_{2i} D_{3i}) + \beta_5 X_i + u_i$. The first is the ANOVA model in which we have considered only one single dummy variable as the independent variable. The second is the ANCOVA model in which we have considered one qualitative dummy variable and another quantitative exogenous variable related to the dummy variable, the omission of which would lead to a 'specification bias'. The third involves an interactive variable $(D_i X_i)$ in which we try to see whether both the slopes and the intercept coefficients differ. In this, there is a possibility of getting four different type of regressions viz. coincident, parallel, concurrent and dissimilar regressions. The fourth situation considered involves a interactive dummy variable like: $Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 (D_{2i} D_{3i}) + \beta_5 X_i + u_i$.