(Lecture No:- 18) (01/11/23)

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

Dependent Variable (Quantitative)

Independent Variable. (Quantitative)

(One of Qualitative/Categorical Independent Variables)

(i) (Qualitative 'X' Variable Binary Categories)

Wage ⟶ Dependent Variable.

Gender ⟶ Categorical Variable

Wage(Rs) = 20,000 − 3500 female

Gender

male → 1

female → 0 ⟶ Dummy Variable.

Wage(Rs) = Rs. 20,000.

↓

Intercept Value

This is actually the average wage of a male.

→ 300 means that the female will earn on average Rs. 300 less than the average male Income.

(ii) (Qualitative 'X' Variable with 'K' Categories)

South ⟶ Reference Category.

Price = 200 + 50 East - 75 West

200 = Intercept = The average Price in South Zone is $200,000.

50 = Home Price in East Zone is $50,000 higher than the home price in South Zone

-75 = Home Price in West Zone is $75,000 lower

than the house price in the South Zone

P# 756, Anderson

(Example)

Dummy Variable

or

Indicator Variable.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

$$\hat{y} = 0.93 + 0.3876 x_1 + 1.263 x_2$$

When $x_2 = 0$ (Mechanical)

$$\hat{y} = 0.93 + 0.3876 x_1 \checkmark$$

When $x_2 = 1$ (Electrical)

$$\hat{y} = 0.93 + 0.3876 x_1 + (1.263 - 1) \checkmark$$

$$\hat{y} = 2.193 + 0.3876 x_1 \checkmark$$

1.263 = The electrical repair will require 1.263 hours longer than

The mechanical Repair.

P# 761

Q 34

$$\hat{y} = 10.1 - 4.2x_1 + 6.8x_2 + 15.3x_3.$$

(a) $$\hat{y} = 10.1 - 4.2x_1 + 6.8x_2 + 15.3(1).$$

15.3 = A restaurant

having a drive-up facility will earn $15300 more than the restaurant without having the drive-up facility.

(b) $x_1 = 2$ ✓

$x_2 = 8,000$

$x_3 = 0$

$\hat{y} = 10.1 - 4.2x_1 + 1.8x_3 + 15.3x_2$

$\hat{y} = 10.1 - 4.2(2)$

$+ 6.8(8) + 15.3(0)$

$\hat{y} = \$56{,}100$

(c) $x_1 = 1, \quad x_2 = \frac{300}{100} = 3$

$x_3 = 1$

$\hat{y} = 10.1 - 4.2(1)$

$+ 6.8(3) + 15.3(1)$

$\hat{y} = \$41{,}600$

$$e_i = y_i - \hat{y}_i$$

original values of "y"

Estimated values of "y"

(Residual Analysis and Regression Diagnostics)

(1) There should not be auto-correlation between the errors. The errors

Should not be Correlated

$e_1, e_2, e_3, \ldots, e_k$

(2) The Variance of errors should be constant.

$Var(\varepsilon_i) = \sigma^2$

(3) The errors should be normally distributed.

(4) There should be no correlation between

The Independent Variable

The Independent Variable should not be Correlated