CH 15: Multiple Regression: Part 1 The Model

1. Review of the simple linear regression model

   (A) The population

   (B) The prediction equation (regression equation)

   In this chapter, we are interested in developing a model with more than one independent variable (multiple regression).

2. Multiple regression model: describes the relationship between one dependent variable (y) and two or more independent variables ($x_1, x_2, \ldots, x_p$) in a linear function. Note: p is the number of independent variables.

   (A) The population model:

   $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

   (B) The multiple regression equation:

   $$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

   (C) The prediction equation (estimated multiple linear regression equation)

   $$\text{eq15.3:}\ \hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_p x_p$$

   Where $b_0, b_1, \ldots, b_p$ are the regression coefficients: (b_0 is the y intercept and $b_1, \ldots, b_p$ are the slopes.)

   (1) $b_0, b_1, \ldots, b_p$ are the estimates of $\beta_0, \beta_1, \ldots . \beta_p$.

   (2) The least squares method is used to minimize $\sum(y_i - \hat{y}_i)^2$ (for the ith observation) to provide the values of $b_0, b_1, \ldots, b_p$.

   (D) The interpretation of the regression coefficients:

   (1) The y intercept (b_0): The estimated average value of y when all the independent variables satisfy $x_1 = x_2 = \cdots = x_p = 0$.

   (2) The slope (b_i and $i = 1, 2, \ldots, p$): Estimate the average of y changes by b_i for each one-unit increase in x_i holding constant the effect of all other independent variables.

EX1 To study the relationship amount the number of Omni-Powerbars sold in a month (y), the price of the Omni-Powerbar (x_1, in cents), and the monthly budget of promotion (x_2, in \$), thirty-four stores were selected and resulting in the following computer output of the multiple regressions model:

a). What is the value of p (the number of independent variable)?

b). What is the prediction equation?

c). Interpret of meaning of b_1.

d). Interpret the meaning of b_2.

e). Predict the average number of bars sold for a store that has a sales price of \$.79 and the promotion expenditures of \$400.

3. Multiple Coefficient of Determination:

   Multiple coefficient of determination measures the proportion of total variation in y explained by all independent variables $x_1, x_2, \ldots, x_p$.

$$\text{eq15.8: Multiple Coefficient of Determination: } R^2 = \frac{SSR}{SST}$$

$$\text{eq15.9: Adjusted Multiple Coefficient of Determination: } R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

EX 1 (cont). From a computer output we find out that $R^2 = 0.7577$, interpret this result.

CH 15: Multiple Regression: Part 2 Hypotheses Test and Confidence Intervals

1. The multiple linear regression model and equation (Population):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the population parameters. Moreover, β_0 is the y-intercept; $\beta_j, j = 1, \ldots, k$ is the slope; and ε is the random error in y (assumed to be normally distributed with $E(\varepsilon) = 0$ and $var(\varepsilon) = \sigma^2$).

2. t-test for the slope β_i:

To determine the existence of a significant linear relationship between the x and y variables. In this case, a hypothesis test of whether β_j is equal to zero or not.

Step 1: State H_0 vs. H_1.

Step 2: Compute the test statistic

The test statistic:

$$\text{Eq 13.15:} \quad t_{cal} = \frac{b_i}{S_{b_i}}$$

with $(n - p - 1)$ degrees of freedom.

Note:

p is the number of independent variables;

b_i is the slope of variable x_i, holding constant the effects of all other independent variables;

S_{b_i} is the standard error of the slope b_i.

Step 3: Make a decision using p-value approach or the critical value approach.

p-value approach: Reject H_0 if p-value $\leq \alpha$

Critical value approach ($CV = \pm t_{\alpha/2}$): Reject H_0 if $t_{cal} \leq -t_{\alpha/2}$ or if $t_{cal} \geq t_{\alpha/2}$

Note: If we reject H_0, the corresponding independent is significant in explaining y, and should be included in the model. Otherwise, it should not be included in the model.

3. The $100(1 - \alpha)\%$ confidence interval for the true slope β_i

$$b_i \pm t_{\alpha/2} S_{b_i}$$

with $n - p - 1$ degrees of freedom

EX 2 The firm wants predict the sales (y, in \$1,000's) using the market value (x_1, in \$1,000's), the total assets (x_2, in \$1,000's), and the number of employees (x_3). To do so, thirty-four firms were selected and the following Excel Output was obtained:

(a) If the firm wants to test whether the coefficient on Market value is significant, what is the relevant test statistic? What decision should be made? (Use the critical value approach with $\alpha = 0.05$).

(b) If the firm wants to test whether the coefficient on total assets is significant, what is the relevant p-value? What decision should be made? (Use the p-value approach with $\alpha = 0.05$).

(c) Find the 95% confidence interval for the true slope of the number of employees (β_3).

4. Which multiple regression model to choose?

   (a) The multiple of coefficient determination

   (b) The standard error of estimate

EX 2 (cont.)