Econ 3790: Business and Economics Statistics

Instructor: Yogesh Uppal
yuppal@ysu.edu

Chapter 15: Multiple Regression Model

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

where:

$\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are the parameters, and
ε is a random variable called the error term

Estimated Multiple Regression Equation

A simple random sample is used to compute sample statistics $b_0, b_1, b_2, \ldots, b_p$ that are used as the point estimators of the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$.

- The <u>estimated multiple regression equation</u> is:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p$$

Interpreting the Coefficients

In multiple regression analysis, we interpret each regression coefficient as follows:

> b_i represents an estimate of the change in y corresponding to a 1-unit increase in x_i when all other independent variables are held constant.

Multiple Regression Model

Example: Car Sales

Suppose we believe that number of cars sold (y) is not only related to the number of ads (x_1), but also to the minimum down payment required at the (x_2). The regression model can be given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where

y = number of cars sold

x_1 = number of ads

x_2 = minimum down payment required ('000)

Estimated Regression Equation

$$y = 14.4 + 3.7 * x_1 - 25 * x_2$$

- Interpretation?
- Estimated values of y?
- Error?
- Prediction?

Multiple Coefficient of Determination

- Relationship Among SST, SSR, SSE

SST = SSR + SSE

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

where:

SST = total sum of squares

SSR = sum of squares due to regression

SSE = sum of squares due to error

Multiple Coefficient of Determination

$$R^2 = SSR/SST$$

$$R^2 = 84.63/89.2 = .949$$

Adjusted Multiple Coefficient of Determination

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Standard Error of Estimate

$$s = \sqrt{MSE} = \sqrt{SSE/(n-p-1)}$$

Testing for Significance: t Test

Hypotheses

$$H_0: \beta_i = 0$$
$$H_a: \beta_i \neq 0$$

Test Statistics

$$t = \frac{b_i}{SE(b_i)}$$

Rejection Rule

Reject H_0 if p-value $\leq \alpha$ or
if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$ where $t_{\alpha/2}$
is based on a t distribution
with $n - p - 1$ degrees of freedom.

Example: Testing for significance of coefficients

Hypotheses

$$H_0: \beta_i = 0$$

$$H_a: \beta_i \neq 0$$

Rejection Rule

For α = .05 and d.f. = ?, $t_{.025}$ =

Test Statistics

$$t = \frac{b_i}{SE(b_i)}$$

Testing for Significance of Regression: F Test

- **Hypotheses**

  H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
  H_a: One or more of the parameters is not equal to zero.

- **Test Statistics**

  $F = \text{MSR/MSE}$

- **Rejection Rule**

  Reject H_0 if p-value $\leq \alpha$ or if $F > F_\alpha$, where F_α is based on an F distribution with p d.f. in the numerator and $n - p - 1$ d.f. in the denominator.

Multiple Regression Model

- Example 2: Programmer Salary
  - Survey

A software firm collected data for a sample of 20 computer programmers. A suggestion was made that regression analysis could be used to determine if salary was related to the years of experience and the score on the firm's programmer aptitude test.

The years of experience, score on the aptitude test, and corresponding annual salary ($1000s) for a sample of 20 programmers is shown on the next slide.

Multiple Regression Model

Exper.	Score	Salary	Exper.	Score	Salary
4	78	24	9	88	38
7	100	43	2	73	26.6
1	86	23.7	10	75	36.2
5	82	34.3	5	81	31.6
8	86	35.8	6	74	29
10	84	38	8	87	34
0	75	22.2	4	79	30.1
1	80	23.1	6	94	33.9
6	83	30	3	70	28.2
6	91	33	3	89	30

Multiple Regression Model

Suppose we believe that salary (y) is related to the years of experience (x_1) and the score on the programmer aptitude test (x_2) by the following regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where

y = annual salary ($1000)

x_1 = years of experience

x_2 = score on programmer aptitude test

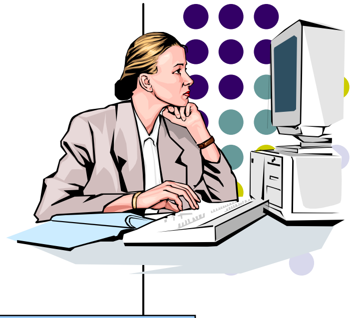
Solving for β_0, β_1 and β_2:

	A	B	C
38			
39		*Coeffic.*	*Std. Err.*
40	Intercept	3.17394	6.15607
41	Experience	1.4039	0.19857
42	Test Score	0.25089	0.07735

Anova Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-statistic
Regression	500.34	……	……..	……….
Error	……..	…….	…….	
Total	599.8	……..		

Estimated Regression Equation



SALARY = 3.174 + 1.404(EXPER) + 0.251(SCORE)

b_1 = 1.404 implies that salary is expected to increase by $1,404 for each additional year of experience (when the variable *score on programmer attitude test* is held constant).

b2 = 0.251 implies that salary is expected to increase by $251 for each additional point scored on the programmer aptitude test (when the variable *years of experience* is held constant).

Prediction

- Suppose Bob had an experience of 4 years and had a score of 78 on the aptitude test. What would you estimate (or expect) his score to be?

$$\hat{y} = 3.174 + 1.404*(4) + 0.251(78)$$

$$= 28.358$$

- Bob’s estimated salary is $28,358.

Error

- Bob's actual salary is $24000. How much error we made in estimating his salary based on his experience and score?

$$error = y - \hat{y} = 24000 - 28358 = -4358$$

- So, we shall overestimate Bob's salary.

Multiple Coefficient of Determination

- Relationship Among SST, SSR, SSE

SST = SSR + SSE

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

where:

SST = total sum of squares

SSR = sum of squares due to regression

SSE = sum of squares due to error

Multiple Coefficient of Determination

$$R^2 = \text{SSR}/\text{SST}$$

- $R^2 = 500.3285/599.7855 = .83418$

Adjusted Multiple Coefficient of Determination

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

$$R_a^2 = 1 - (1 - .834179)\frac{20-1}{20-2-1} = .814671$$

Testing for Significance: t Test

Hypotheses

$$H_0 : \beta_i = 0$$
$$H_a : \beta_i \neq 0$$

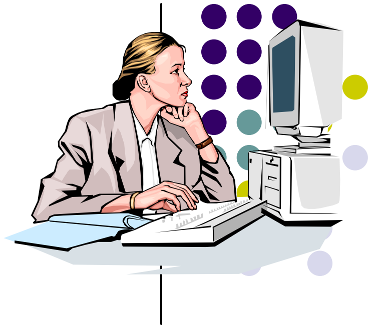
Test Statistics

$$t = \frac{b_i}{SE(b_i)}$$

Rejection Rule

Reject H_0 if p-value $\leq \alpha$ or if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$ where $t_{\alpha/2}$ is based on a t distribution with $n - p - 1$ degrees of freedom.

Example



Hypotheses

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

Rejection Rule

For $\alpha = .05$ and d.f. = 17, $t_{.025} = 2.11$

Reject H_0 if p-value $\leq .05$ or if $t \geq 2.11$

Test Statistics

$$t = \frac{b_1}{SE(b_1)} = \frac{1.404}{0.199} = 7.07$$

Since t=7.07 > $t_{0.025}$ =2.11, we reject H_0.

Testing for Significance of Regression: F Test

- **Hypotheses**

  H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$

  H_a: One or more of the parameters is not equal to zero.

- **Test Statistics**

  $F = \text{MSR}/\text{MSE}$

- **Rejection Rule**

  Reject H_0 if p-value $\leq \alpha$ or if $F > F_\alpha$, where F_α is based on an F distribution with p d.f. in the numerator and $n - p - 1$ d.f. in the denominator.

Example

Hypotheses

H_0: $\beta_1 = \beta_2 = 0$

H_a: One or both of the parameters is not equal to zero.

Rejection Rule

For $\alpha = .05$ and d.f. = 2, 17; $F_{.05} = 3.59$

Reject H_0 if p-value $\leq .05$ or $F \geq 3.59$

Test Statistics

F = MSR/MSE

= 250.17/5.86 = 42.8

$F = 42.8 \geq F_{0.05} = 3.59$, so we can reject H_0.