Economics Honours (Sixth Semester)

Basic Econometrics

Application of Multiple Regression Analysis

*Problem

Table 1 stated below gives the bushels of corn per acre, Y, resulting from use of various amounts of fertilizer X_1 and insecticides X_2 both in pounds per acre, from 1971 to 1980. Run a multiple regression analysis in three variables and test the significance of the partial regression coefficients. Also determine the overall significance of the regression.

Table 1: Data on Corn produced with Fertilizer and Insecticide

Year	Y	X_1	X_2
1971	40	6	4
1972	44	10	4
1973	46	12	5
1974	48	14	7
1975	52	16	9
1976	58	18	12
1977	60	22	14
1978	68	24	20
1979	74	26	21
1980	80	32	24

Source: [The problem and the Table has been taken from Schaum's Outlines on Statistics and Econometrics by Dominick Salvatore and Derrick Reagle]

*Solution:

Suppose the sample linear regression model in three variables for running multiple regression is of the form,

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{u}$$

where Y = bushels of corn per acre, X_1= use of various amounts of fertilizer and X_1 and X_2 = insecticides (both in pounds per acre), $\hat{\beta}_0$ = intercept coefficient, $\hat{\beta}_1$ measures the change in the mean value of Y, per unit change in X_1, holding the value of X_2 constant. Similarly, $\hat{\beta}_2$ measures the change in the mean value of Y per unit change in X_2, holding the value of X_1 constant. $\hat{\beta}_1$ and $\hat{\beta}_2$ are known as partial regression coefficients; $\hat{u}$ = residual term.

Table 1: Showing statistical working corn-fertilizer-pesticide problem

Year	Obs	(Y) (Corn)	(X_1) (Fertilizer)	(X_2) (Pesticide)	y	x_1	x_2
1971	1	40	6	4	-17	-12	-8
1972	2	44	10	4	-13	-8	-8
1973	3	46	12	5	-11	-6	-7
1974	4	48	14	7	-9	-4	-5
1975	5	52	16	9	-5	-2	-3
1976	6	58	18	12	1	0	0
1977	7	60	22	14	3	4	2
1978	8	68	24	20	11	6	8
1979	9	74	26	21	17	8	9
1980	10	80	32	24	23	14	12
	n =10	$\Sigma Y = 570$ $\bar{Y} = 57$	$\Sigma X_1 = 180$ $\bar{X}_1 = 18$	$\Sigma X_2 = 120$ $\bar{X}_2 = 12$	$\Sigma y_i = 0$	$\Sigma x_1 = 0$	$\Sigma x_2 = 0$

Table 1: Showing statistical working corn-fertilizer-pesticide problem contd…

Obs	(Y) (Corn)	(X_1) (Fertilizer)	(X_2) (Pesticide)	x_1y	x_2y	x_1x_2	x_1^2	x_2^2
1	40	6	4	204	136	96	144	64
2	44	10	4	104	104	64	64	64
3	46	12	5	66	77	42	36	49
4	48	14	7	36	45	20	16	25
5	52	16	9	10	15	6	4	9
6	58	18	12	0	0	0	0	0
7	60	22	14	12	6	8	16	4
8	68	24	20	66	88	48	36	64
9	74	26	21	136	153	72	64	81
10	80	32	24	322	276	168	196	144
n =10	$\Sigma Y = 570$ $\bar{Y} = 57$	$\Sigma X_1 = 180$ $\bar{X}_1 = 18$	$\Sigma X_2 = 120$ $\bar{X}_2 = 12$	$\Sigma x_1y = 956$	$\Sigma x_2y = 900$	$\Sigma x_1x_2 = 524$	$\Sigma x_1^2 = 576$	$\Sigma x_2^2 = 504$

[Note: $y = (Y - \bar{Y}_1)$, $x_1 = (X_1 - \bar{X}_1)$ and $x_2 = (X_2 - \bar{X}_2)$]

*1. Determining the regression coefficients:

[Discussed in Multiple regression analysis]

$$\hat{\beta}_1 = \frac{(\Sigma x_1 y)(\Sigma x_2^2) - (\Sigma x_2 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2} = \frac{(956)(504) - (900)(524)}{(576)(504) - (524)^2} = 0.65$$

$$\hat{\beta}_2 = \frac{(\Sigma x_2 y)(\Sigma x_1^2) - (\Sigma x_1 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2} = \frac{(900)(576) - (956)(524)}{(576)(504) - (524)^2} = 1.11$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 = 57 - (0.65)(18) - (1.11)(12) = 31.98$$

The estimated multiple regression equation is **= 31.98 + 0.65 X_1 + 1.11X_2**

*2. Testing the significance of partial regression coefficients:

To test for the significance of partial regression coefficients, we need first to determine se ($\hat{\beta}_1$) and se ($\hat{\beta}_2$).

Table 2: Calculations for testing the significance of ($\hat{\beta}_1$) and ($\hat{\beta}_2$)

Obs	(Y) (Corn)	(X_1) (Fertilizer)	(X_2) (Pesticide)	$\hat{Y}$	u^	$u\hat{}^2$	y^2
1	40	6	4	40.32	-0.32	0.1024	289
2	44	10	4	42.92	1.08	1.1664	169
3	46	12	5	45.33	0.67	0.4489	121
4	48	14	7	48.85	-0.85	0.7225	81
5	52	16	9	52.37	-0.37	0.1369	25
6	58	18	12	57.00	1.00	1.000	1
7	60	22	14	61.82	-1.82	3.3124	9
8	68	24	20	69.78	-1.78	3.1684	121
9	74	26	21	72.19	1.81	3.2761	289
10	80	32	24	79.42	0.58	0.3364	529
n =10	$\Sigma Y = 570$ $\bar{Y} = 57$	$\Sigma X_1 = 180$ $\bar{X}_1 = 18$	$\Sigma X_2 = 120$ $\bar{X}_2 = 12$		Σu^ = 0	$\Sigma u\hat{}^2$ = 13.6704	Σy^2 = 1634

We get from the formulae determined in previous lesson

$$\sigma^2 = \hat{\sigma}^2 = \frac{\Sigma \hat{u}^2}{(n-k)} = 13.6704/(10-3) = 1.9529$$

$$Var(\hat{\beta}_1) = \sigma^2 . \frac{\Sigma x_2^2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} = \frac{1.9529 \times (504)}{(576)(504) - (524)^2} = 0.06$$

$se(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{0.06} = 0.24$

$$Var(\hat{\beta}_2) = \sigma^2 . \frac{\Sigma x_1^2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} = \frac{1.9529 \text{ x } (576)}{(576)(504) - (524)^2} = 0.07$$

$se(\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_2)} = \sqrt{0.07} = 0.27$

To test the significance of the partial regression coefficient β_1 we set the null hypothesis against the alternative hypothesis as

$H_0 : \beta_1 = 0,$

against $H_1 : \beta_1 \neq 0.$

- Here the null hypothesis states that with X_2 held constant, X_1 has no influence on Y. To test the null hypothesis we compute the value of *t*.

$$t_1 = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$$

Therefore $t_1 = (\hat{\beta}_1 - \beta_1)/ se(\hat{\beta}_1) = (0.65 - 0)/0.24 = 2.7$

- Again to test the significance of the partial regression coefficient β_2 we set the null hypothesis against the alternative hypothesis as

$H_0 : \beta_2 = 0,$

against $H_1 : \beta_2 \neq 0.$

Here the null hypothesis states that with X_1 held constant, X_2 has no influence on Y. To test the null hypothesis we compute the value of *t*.

$$t_2 = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)}$$

Therefore $t_2 = (\hat{\beta}_2 - \beta_2)/ se(\hat{\beta}_2) = (1.11 - 0)/0.27 = 4.11.$

Since both t_1 and t_2 exceed t=2.365 with 7 degrees of freedom (df) at 5% level of significance, we reject the null hypotheses in both the cases and conclude that both $\hat{\beta}_1$ and $\hat{\beta}_2$ (the partial regression coefficients) are statistically significant at the 5% level.

*3. Overall significance of the regression

To test the overall significance of the regression we must first determine the Coefficient of Multiple Determination, R^2.

The formula for $R^2 = 1 - (\Sigma \hat{u}^2)/ \Sigma y^2$

In our example, we get R^2 = 1- (13.6704/1634) = 1 - 0.0084 = 0.9916 = 99.16%.

We use the F test to test the overall significance of the multiple regression. We set the null hypothesis against the alternative hypothesis as

$H_0 : \beta_1 = \beta_2 = 0,$

(all slope coefficients are simultaneously zero)

against H_1 : Not all slope coefficients are simultaneously zero

To test the null hypothesis we compute the value of F.

$$F = \frac{R^2 / (k-1)}{(1-R^2)/(n-k)}$$

where n = number of observations, k = total number of parameters in the multiple regression model and R^2 = Coefficient of determination of the multiple regression model.

$$\text{Therefore } F_{2,7} = \frac{0.9916/2}{(1-0.9916)/7} = 413.17$$

The calculated value of F exceeds the tabular value of F = 4.74 at the 5% level of significance and with numerator df = 2 and denominator df = 7, the null hypothesis is rejected and we conclude that R^2 is significantly different from zero.

***Exercise**

1. For the given data solve the questions that follow:

n	1	2	3	4	5	6	7	8	9	10
Y	20	28	40	45	37	52	54	43	65	56
X_1	2	3	5	4	3	5	7	6	7	8
X_2	5	6	6	5	5	7	6	6	7	7

(a) Determine the regression coefficients

(b) Determine the regression equation

(c) Test the significance of the partial regression coefficients

(d) Determine the coefficient of multiple regression

(e) Test the overall significance of the regression

(f) Determine the partial correlation coefficients

*N.B. The problem, explanations of all the theories and solution to the problem has been accessed from the reference mentioned.

References:

Salvatore, D., & Reagle, D. (2011). *Schaum's outline of statistics and econometrics* (2nd ed.). McGraw-Hill Education.