LECTURE - APPLIED Regression - MLR Models

Example of Second Degree MLR Model

Problem - 4.88

Problem 4.88. Cost of modifying a naval air base. A naval base is considering modifying or adding to its fleet of 48 standard aircraft. The final decision regarding the type and number of aircraft to be added depends on a comparison of cost versus effectiveness of the modified fleet. Consequently, the naval base would like to model the projected percentage increase y in fleet effectiveness by the end of the decade as a function of the cost x of modifying the fleet. A first proposal is the quadratic model $E(y) = \beta_0 + \beta_1.x + \beta_2.x^2$ The data provided in the table were collected on 10 naval bases of similar size that recently expanded their fleets.

(a) Fit a simple linear regression model to predict the percentage improvement using cost.

(b) Fit a quadratic model to the data.

(c) Is there sufficient evidence to conclude that the percentage improvement y increases more quickly for more costly fleet modifications than for less costly fleet modifications? Test with $\alpha = 0.05$.

(d) Now consider the complete second-order model

$$E(y) = \beta_0 + \beta_1.X_1 + \beta_2.X_1^2 + \beta_3.X_2 + \beta_4.X_1.X_2$$

where x =Cost to modifying the fleet $X_2 = \begin{cases} 1, & \text{if if U.S base} \\ 0, & \text{if foreign base} \end{cases}$

Data Table

	PERCENT	COST	BASE
1	18.00	125.00	US
2	32.00	160.00	US
3	9.00	80.00	US
4	37.00	162.00	US
5	6.00	110.00	US
6	3.00	90.00	FOR
7	30.00	140.00	FOR
8	10.00	85.00	FOR
9	25.00	150.00	FOR
10	2.00	50.00	FOR

Solution:

(a) Output from SLR Model $PERCENT \ = \beta_0 \ + \ \beta_1 \times COST +$ error

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
COST	1	1274.82	1274.82	42.71	0.0002
Residuals	8	238.78	29.85		

Parameter Estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-18.6228	5.7472	-3.24	0.0119
COST	0.3110	0.0476	6.54	0.0002

Residual standard error: 5.463 on 8 degrees of freedom

Multiple R-squared: 0.8422, Adjusted R-squared: 0.8225

F-statistic: 42.71 on 1 and 8 DF, p-value: 0.0002

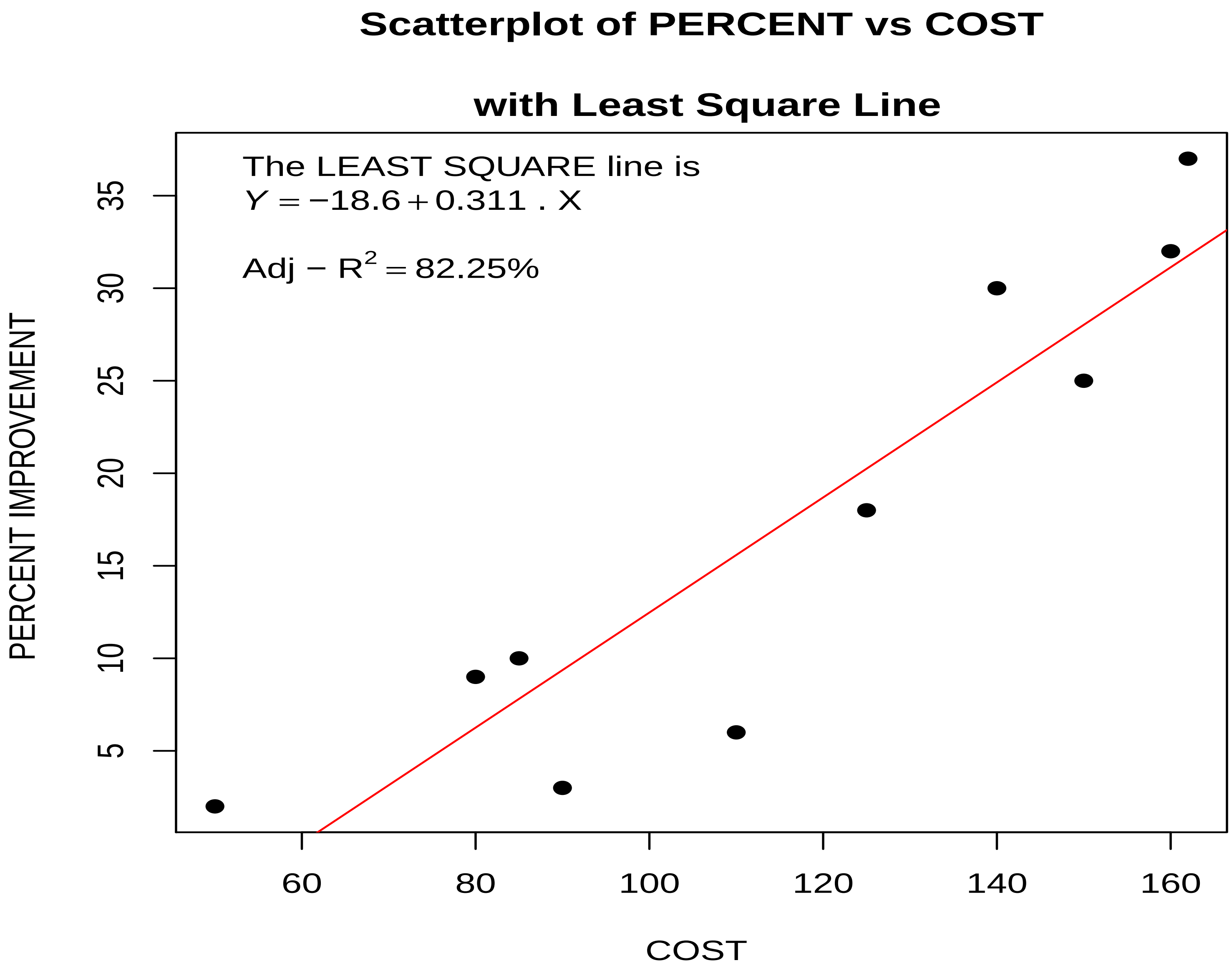

Figure 1: Scatter Plot of the Data and SLR Fitted Line

(b) Output from Quadratic Model

$$PERCENT = \beta_0 + \beta_1 \times COST + \beta_2 \times COST^2 + \text{error}$$

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
COST	1	1274.82	1274.82	61.62	0.0001
COST^2	1	93.95	93.95	4.54	0.0706
Residuals	7	144.82	20.69		

Parameter Estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.6590	14.5501	0.73	0.4876
COST	-0.2816	0.2809	-1.00	0.3494
$COST^2$	0.0027	0.0013	2.13	0.0706

Residual standard error: 4.549 on 7 degrees of freedom

Multiple R-squared: 0.9043, Adjusted R-squared: 0.877

F-statistic: 33.08 on 2 and 7 DF, p-value: 0.000271

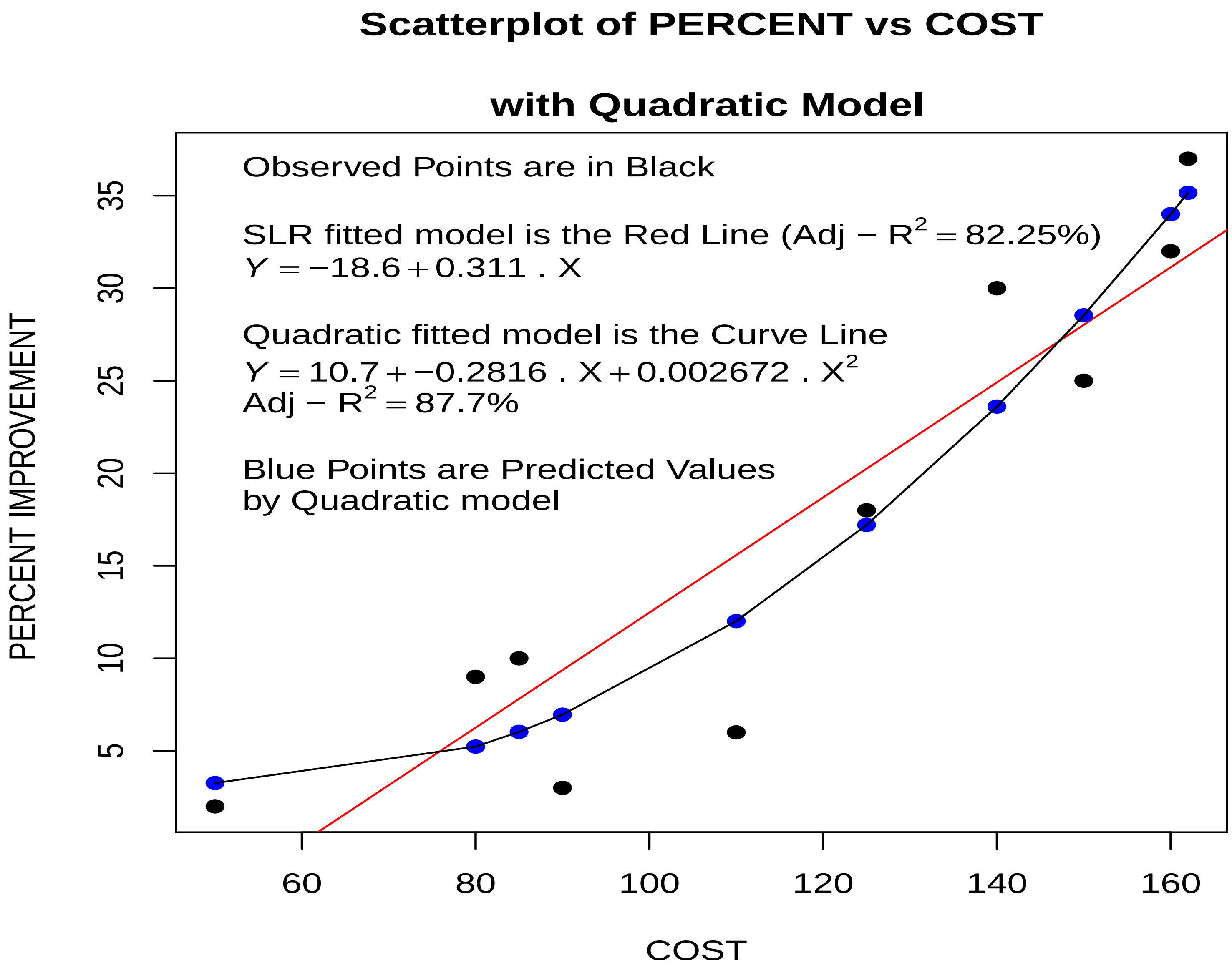

Figure 2: Scatter Plot and Other Two Fitted Models

(c) To check whether the percentage improvement y increases more quickly for more costly fleet, we test

$$H_0 : \beta_2 \leq 0 \quad vs \quad H_1 : \beta_2 > 0$$

t-stat = 2.131 and one sided p-value = $0.5 \times 0.0706 = 0.0353$.

As p-value = $0.0353 < 0.05$, we reject the null hypothesis.

(d)

Model 3: PERCENT = COST + COST^2 + BASE + COST $\times$ BASE + error)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
COST	1	1274.82	1274.82	45.86	0.0011
COST^2	1	93.95	93.95	3.38	0.1254
BASE	1	1.19	1.19	0.04	0.8444
COST:BASE	1	4.65	4.65	0.17	0.6994
Residuals	5	138.98	27.80		

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.2537	18.3929	0.72	0.5034
COST	-0.3546	0.3860	-0.92	0.4004
COST^2	0.0031	0.0019	1.68	0.1539
BASEUS	5.4030	15.4189	0.35	0.7403
COST:BASEUS	-0.0514	0.1256	-0.41	0.6994

Residual standard error: 5.272 on 5 degrees of freedom Multiple R-squared: 0.9082, Adjusted R-squared: 0.8347 F-statistic: 12.36 on 4 and 5 DF, p-value: 0.008356

COMPARISON OF THREE MODELS

1: PERCENT = COST

2: PERCENT = COST + COST^2

3: PERCENT = COST + COST^2 + BASE + COST $\times$ BASE

	Res.Df	RSS	Diff	Sum of Sq	F	Pr(>F)
1	8	238.78				
2	7	144.82	1	93.95	3.38	0.1254
3	5	138.98	2	5.84	0.11	0.9022