Solution to HW 7

Guanhua FANG

November 27, 2017

8.6

a. R^2 is 0.81. Plot is shown in graph

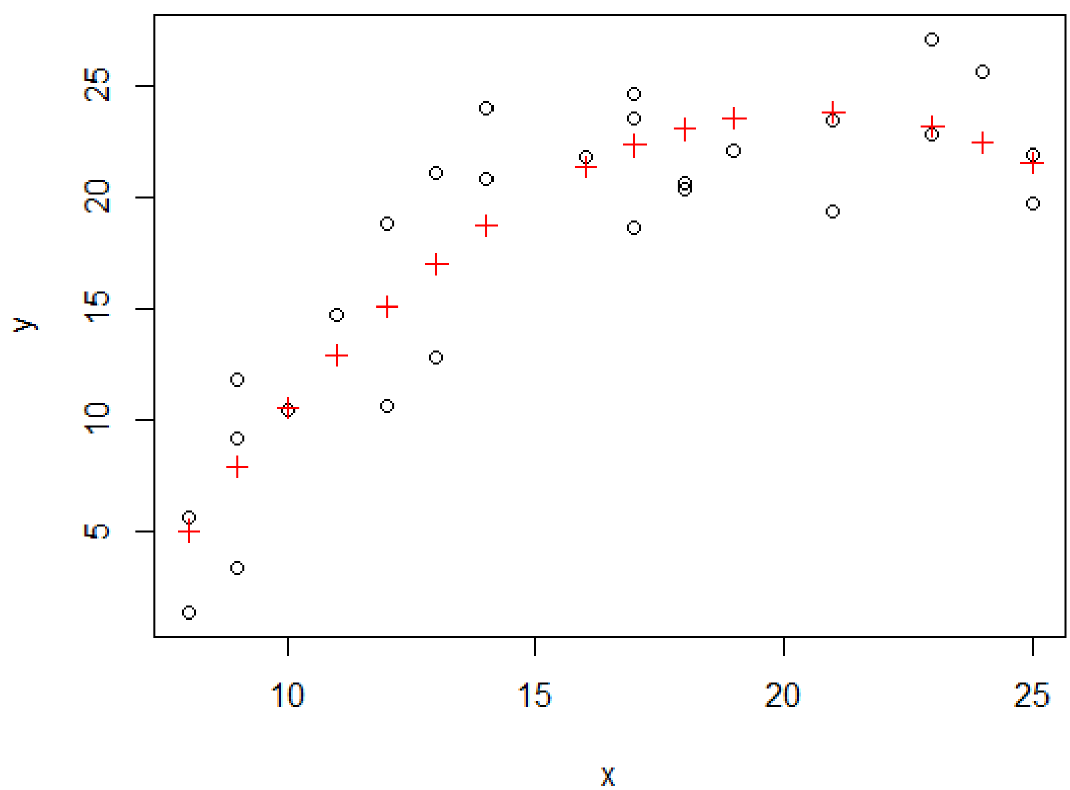

Figure 1: Plot for part(a) of 8.6

b. $H_0 : Y = \beta_0 + \beta_1 X + \beta_2 X^2$. $H_a : Y \neq \beta_0 + \beta_1 X + \beta_2 X^2$. F statistic is 52.63. Cut off is

$f_{0.99}(2,24) = 5.61$. The null can not be rejected. The model is not lack of fit.

c. Mean at $X = 10$ is 10.6. Mean at $X = 15$ is 20.1. Mean at $X = 20$ is 23.8. MSE is 9.94. The covariance matrix of $\hat{Y}_{10}, \hat{Y}_{15}, \hat{Y}_{20}$ is $(x_{10}, x_{15}, x_{20})^T (X^T X)^{-1} (x_{10}, x_{15}, x_{20})$ which is equal to We use

	1	2	3
1	0.0859	0.0263	-0.0035
2	0.0263	0.0802	0.0575
3	-0.0035	0.0575	0.0740

Bonferroni method, then get intervals are $(7.56, 13.58), (17.2, 23.05)$ and $(20.99, 26.58)$ respectively.

d. Mean at $X = 15$ is 20.1, sd at $X = 15$ is $([(1,15,15^2)(X^T X)^{-1}(1,15,15^2)^T] + 1)^{1/2} * \sqrt{9.94}$ $= 3.27$. So 99% interval is $(20.1 - 2.79 * 3.28, 20.1 + 2.79 * 3.28) = (10.97, 29.130)$. Female at age 15 will have steroid level between $(10.97, 29.30)$ with prob 0.99.

e. $H_0 : \beta_2 = 0$. $H_a : \beta_2 \neq 0$. Use statistic $\hat{\beta}_2 / sd(\beta_2) = -5.0$, it follows $t(1, 24)$. The cutoff is 2.79. Hence, the null is rejected. There exists relation between X^2 and Y.

f. The regression function is $Y = -26.32 + 4.87X - 0.12X^2$. So as age goes up towards 20.5, the steroid level will increase correspondingly, then the level drops as age increases.

8.42

a. The residual tends to become large when fitted value is large. It may violate the constant variance assumption.

b. $H_0 : \beta_{x_1^2} = 0, \beta_{x_2^2} = 0, \beta_{x_1:x_2} = 0$, $H_a :$ at least one of $\beta_{x_1^2}, \beta_{x_2^2}, \beta_{x_1:x_2}$ not zero. We use F test, the statistic value equals 0.37. The cut off is $F_{0.95}(3, 25) = 2.99$. The Null cannot be rejected. Hence, all second order terms can be dropped.

c. $H_0 : \beta_2 = 0, \beta_{99} = \beta_{01} = \beta_{02} = 0$, $H_a :$ at least one of $\beta_2, \beta_{99}, \beta_{01}, \beta_{02}$ not zero. We use F test, the statistic value equals 0.68. The cut off is $F_{0.95}(4, 28) = 2.71$. The Null cannot be rejected. Hence, "advertise index" and "year" can be dropped.

8.43

We first fit first order model and set "year" as the categorical variable. We can see that only X_1 and X_2 are significant. The score is not so much related with variable "year". Hence, we only consider the second order regression against X_1, X_2. Then, we can find that X_1^2 and X_2^2 are significant variable. We'd better choose $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$. After we check the residual plot, we could see it still exists the phenomenon of non-constant variance. We may further consider weighted linear regression.

10.5

a.

b. The fitted function in 6.5 (b) is $Y = 37.65 + 4.42X_1 + 4.38X_2$. Coefficients β_1 and β_2 match the slope in two add-ed variable plots. Hence, the model is appropriate.

c. The function is $e(Y|X_1) = 4.38e(X_2|X_1)$, it is consistent with previous regression function.

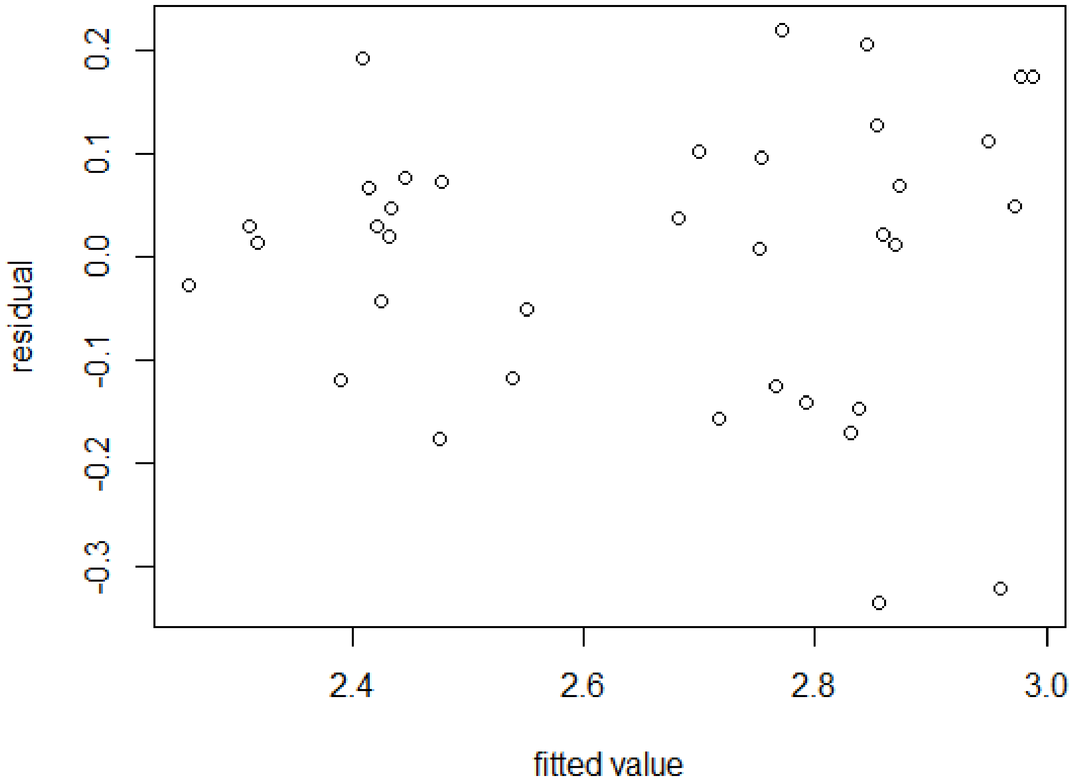

Figure 2: Plot for part (a) of 8.42

10.9

a. SSE is 94.3. The studentized deleted residual is in Table 1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-0.04	0.06	-1.41	1.44	-0.38	-0.69	-0.80	0.53	0.48	-0.63	1.90	1.02	-1.19	-2.19	1.55	0.26

The cutoff for boneronni test with level $\alpha = 0.10$ is 2.84. So there is no detected outlier.

b. The h_{ii} is in the Table 2. All leverage value is quite small.

c. $2p/n = 6/16 = 0.375$. There is no observation according to the rule.

d. From, graph, the new point is within range. $h_{new,new}$ is 0.175, it is within range. Hence, two degree agree with each other.

g. The case 14 is a influential case.

6 Let $\hat{\beta}$ be the LSE for regression of all observation We know that $\hat{\beta}^1$ be the LSE for regression of all observation. If, $\hat{\beta} = \hat{\beta}^1$, then $\hat{\beta}$ satisfies $\sum_{i=1}^{n}(y_1 - x_1^T\beta)x_1 = \mathbf{0}$ and $\sum_{i=2}^{n}(y_1 - x_1^T\beta)x_1 = \mathbf{0}$.

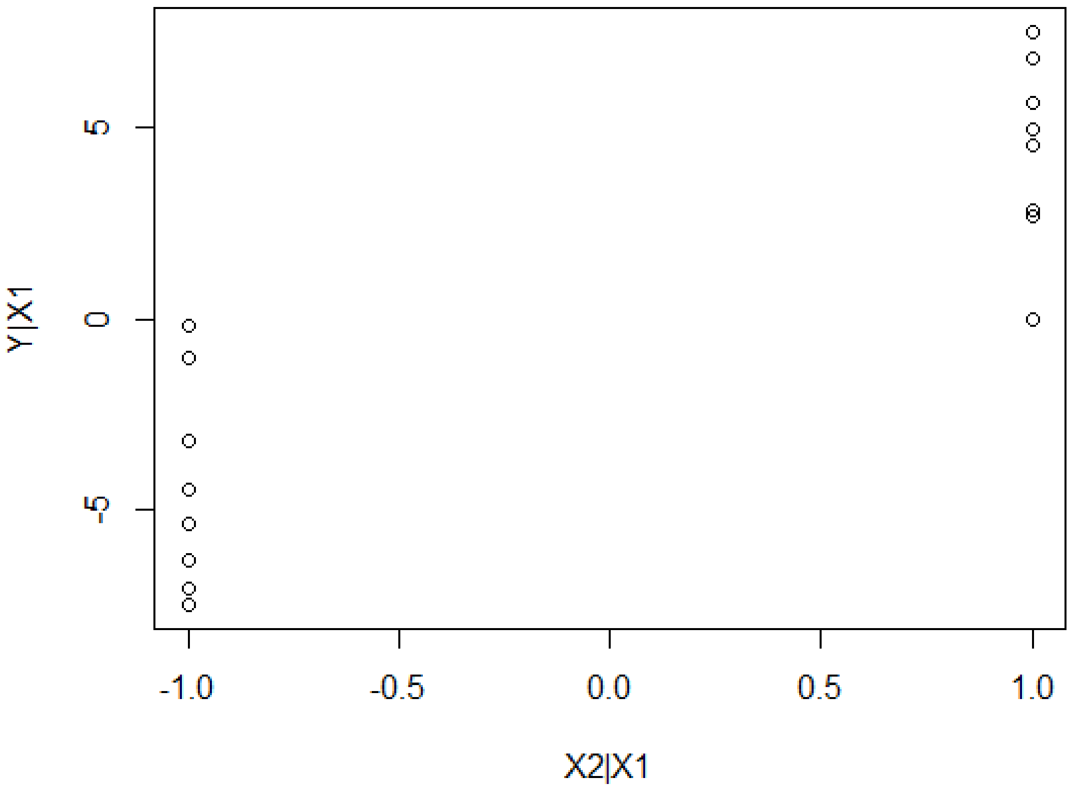

Figure 3: Added variable plot for part (a) of 10.5

Then, we have $(y_1 - x_1^T\hat{\beta})x_1 = 0$. Since $x_1 = (1, x_{11}, \ldots, x_{1p})^T \neq 0$. Then, $y_1 - x_1^T\hat{\beta}$ must be zero. In other words,it lies on the regression line. The other way is easy to see.

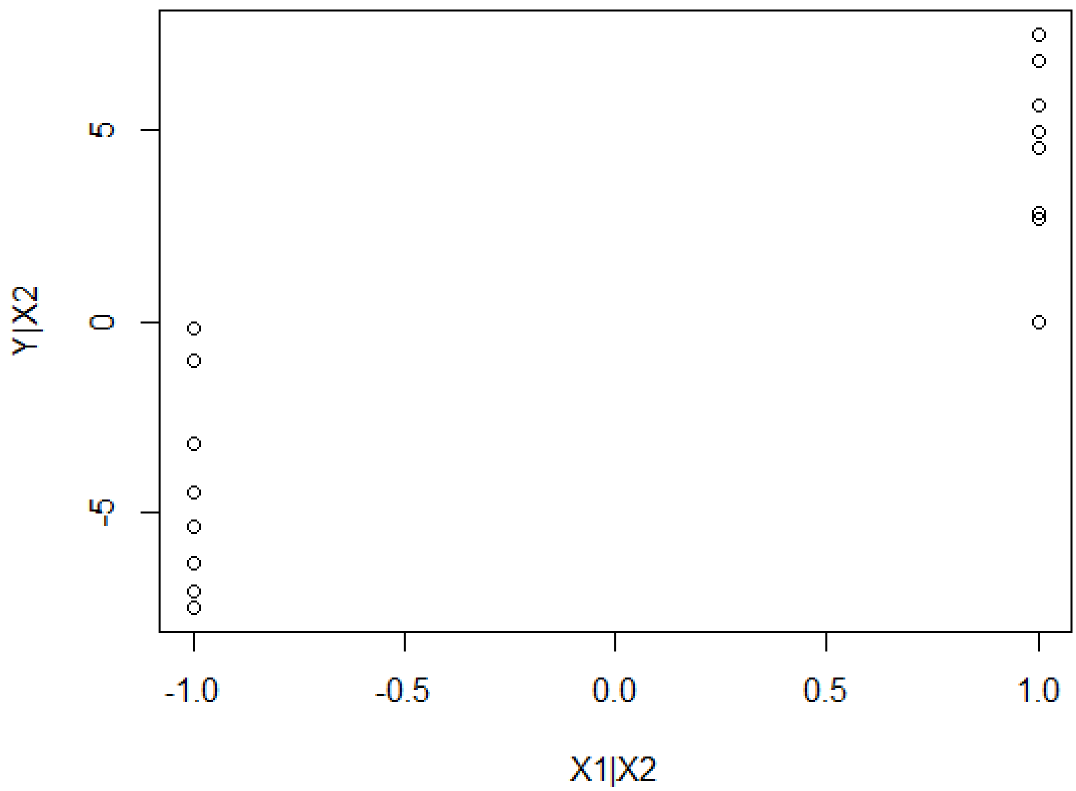

Figure 4: Added variable plot (a) of 8.42

1	2	3	4	5	6	7	8
0.238	0.238	0.238	0.238	0.138	0.138	0.138	0.138
9	10	11	12	13	14	15	16
0.138	0.138	0.138	0.138	0.237	0.238	0.237	0.238

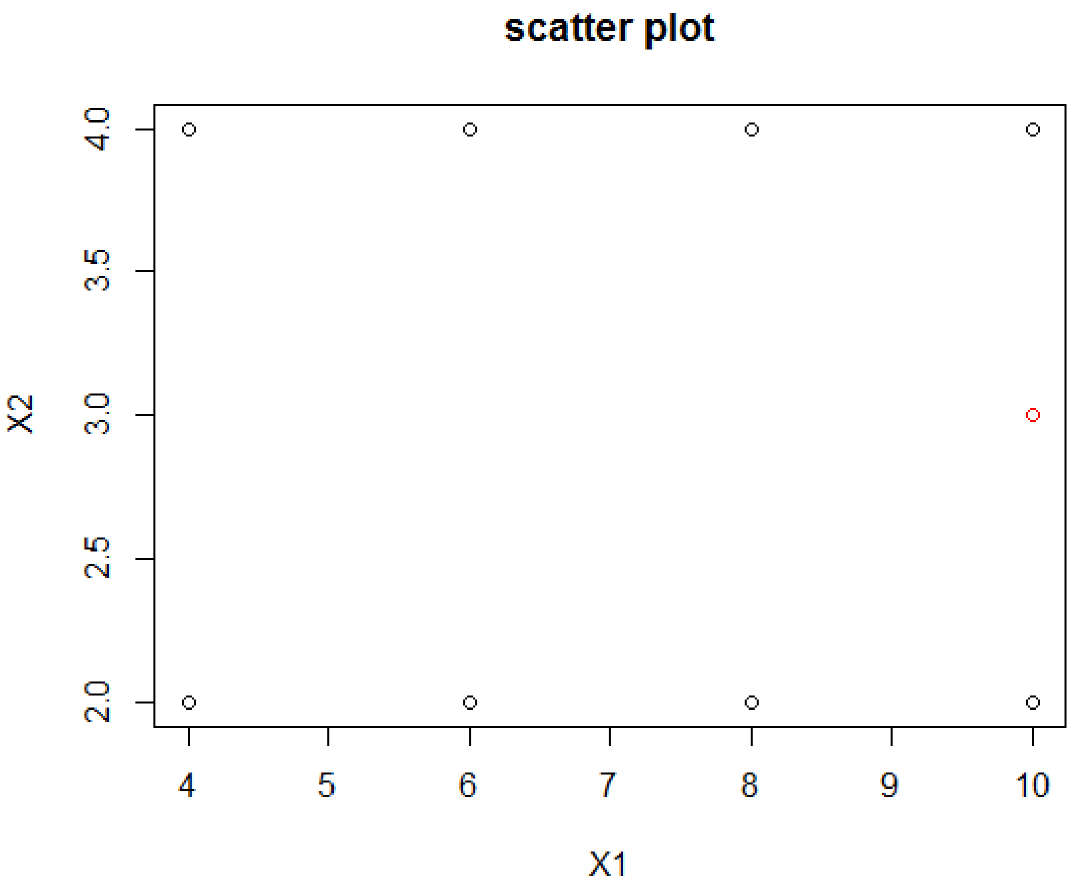

Figure 5: Added variable plot (a) of 8.42

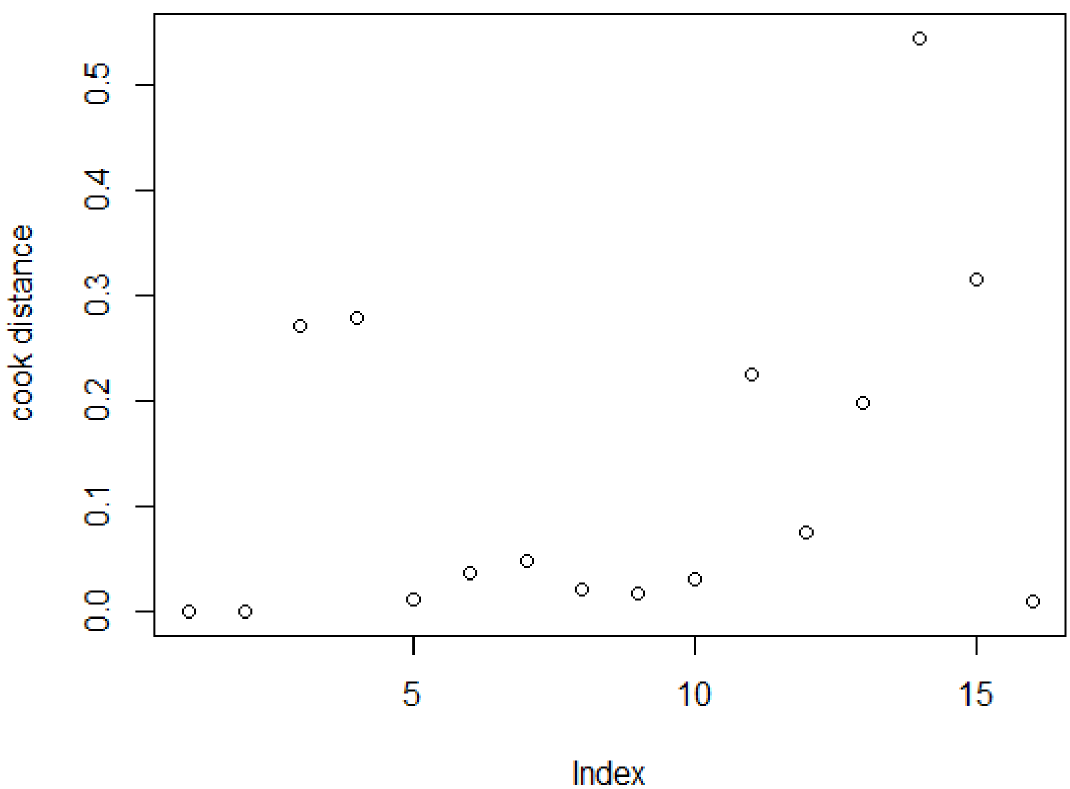

Figure 6: Index plot