Solution to HW 6

Guanhua FANG

November 10, 2017

3.14

a. $H_0 : \mathrm{E}Y = \beta_0 + \beta_1 X$, $H_a : \mathrm{E}Y \neq \beta_0 + \beta_1 X$. SSE is 128.75. SSE of linear model is 146.44. Then the F statistic equals $(146.44 - 128.75)/2/128.75/12 = 0.82$. It follows distribution $F(2, 12)$. The cut off$F_{0.99}(2, 12)$ which is equal to 6.93. Thus, the null cannot be rejected.

b. The advantage is that equal number of observations under each X level can make error smaller at each level. Also, we can tell whether there is an outlier at each X level. The estimation could become more robust.

The disadvantage is that in the variance of prediction error in linear regression is not constant through all X. Hence, the error variance is bigger when X lies far away from its mean. So it may be suitable to assign more points there.

c. The linear assumption between response and predictor is appropriate from the F-test in part (a).

7.7

a. We know that $SST = \sum (y_i - \bar{y})^2 = 236.56$.

We first do linear regression between Y against $X4$, then get $SSR_4 = 67.77$.

We then do linear regression between Y against$X1, X4$ and get $SSR_{1|4} = 42.28$.

We next do linear regression between Y against$X1, X2, X4$ and get $SSR_{2|1,4} = 27.87$.

Lastly, we do linear regression between Y against $X1, X2, X3, X4$ and get $SSR_{3|1,2,4} = 0.4$.

b. $H_0 : \beta_3 = 0$ and $H_a : \beta_3 \neq 0$. F statistic is equal to $0.4/1.29 = 0.31$. The cutoff is $F_{0.99}(1, 76) = 6.98$. Hence, the null cannot be rejected. The P value is 0.58.

7.10

When $\beta_1 = 0.1$ and $\beta_2 = 0.4$, the reduced model is

$$Y = -0.1X_1 + 0.4X_2 + \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

and the full model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

we set $\tilde{Y} = Y + 0.1X_1 - 0.4X_2$. We do regression $\tilde{Y}$ against X_3 and X_4 and get SSE is 110.08. The SSE of full model is 98.25. We do F test and get statistic equal to $(110.08 - 98.25)/2/1.137^2 = 4.57$. The cutoff is $F_{0.99}(2, 76) = 4.89$. The null cannot be rejected. The reduced model is acceptable.

7.16

a. $y_i = \frac{1}{\sqrt{n-1}} \frac{y_i - \bar{y}}{s_y}$. $x_i = \frac{1}{\sqrt{n-1}} \frac{x_i - \bar{x}}{s_x}$.

b. $b_1^* = 0.89$. The correlation between standardized y and standardized x_1 is 0.89.

c. $b_1 = s_y/s_1 * b_1^* = 4.43$, $b_2 = s_y/s_2 * b_2^* = 4.38$. It is equal to original coefficients.

7.24

a. Fit $Y \sim X_1$, we get $Y = 50.78 + 4.43X1$.

b. The coefficient remains same when regress Y against both X_1 and X_2.

c. $SSR_1 = 1566.52$, $SSR_{1|2} = (0.9521 - 0.1557) * 1967 = 1566.52$. They are equal.

d. X_1 and X_2 are orthogonal. The X design is orthogonal design.

7.37

a. Fit 1 (include X_3), $R^2 = 0.90$.

Fit 2 (include X_4), $R^2 = 0.90$.

Fit 3 (include X_5), $R^2 = 0.96$.

Fit 4 (include X_6), $R^2 = 0.90$.

b. $SSY = 1406.2$. X_5 is the best predictor compared with other three variables. The extra sum is bigger than others.

c. F statistic is equal to $(0.955 - 0.8998)/1/((1 - 0.8998)/437) = 240$. The cutoff is 6.99. Hence, we should include X_5. The F statistic for other three are 12.2, 1.3, 3.0 respectively, not as large as F value when X_5 is included.