Foundations of Linear Models, Assignment 2

Tanriver Ayder Ezgi (1541821), Oana Petrof (1541809),
Olusoji Oluwafemi Daniel (1541893), Van Baelen Wessel (1234318),
Owokotomo Olajumoke Evangelina (1539654)

October 2, 2016

PROBLEM 1

Refer to Patient satisfaction Problem 6.15. Test whether both X_2 and X_3 can be dropped from the regression model given that X_1 is retained. Use $\alpha = 0.025$. State the alternatives, decision rule and conclusion. What is the P-value of the test?

PROBLEM 2

Refer to Patient satisfaction Problem 6.15. Test whether $\beta_1 = -1$ and $\beta_2 = 0$; use α= 0.025. State the alternatives, full and reduced models, decision rule, and conclusion.

PROBLEM 3

Fit first order linear regression model for relating patient's satisfaction (Y) to patient's age (X_1) and severity of illness (X_2). State the fitted regression function.

The regression function is $\hat{Y} = 156.67186 - 1.26765X_1 - 0.92079X_2$

Compare the estimated regression coefficients for patients' age and severity of illness obtained in the previous question with the corresponding coefficients obtained by fitting a full model.

Model	X_1	X_2
Full	-1.14161(0.21480)	-0.44200(0.43489)
Reduced	-1.26765(0.21035)	-0.92079(0.49197)

Table 1: Coefficients(standard error) of Age and Severity of Illness

From the table above, it can be observed that the standard errors of X_1 (Age) remains approximately the same in both models, while the coefficient estimate on the otherhand exhibited a slight change (a difference of 0.13 and a ratio of 0.9), hence the inclusion or withdrawal of X_3 (anxiety level) has a very mild effect on the coefficient of age. On the other hand, a difference of about 0.49 (a ratio of about 2.1) is observed between the coefficent estimate of X_2 (severity of illness) in the full model and in the reduced model. A change is also noticed in its standard error but not as huge as that observed in its coefficient estimate, hence the inclusion or removal of X_3 has an effect on the coefficient estimate of X_2

Does $SSR(X_1) = SSR(X_1|X_3)$ here? Does $SSR(X_2) = SSR(X_2|X_3)$?

SS	Value
$SSR(X_1)$	8275.38885
$SSR(X_2)$	4860.26000
$SSR(X_1, X_3)$	9038.80461
$SSR(X_2, X_3)$	6262.91029

Table 2: Sum of Squares

$$SSR(X_1|X_3) = SSR(X_1, X_3) - SSR(X_3) = 3483.89147 \neq SSR(X_1)$$

$$SSR(X_2|X_3) = SSR(X_2, X_3) - SSR(X_3) = 707.99714 \neq SSR(X_2)$$

Since $SSR(X_1|X_3) \neq SSR(X_1)$ and $SSR(X_2|X_3) \neq SSR(X_2)$, it implies that adding X_1 and X_2 improves the regression function and it does contain information that is not contained in X_3.

Refer to the correlation matrix of the variables in the full model, what bearing does it have on your findings from the two previous questions?

	Y	X_1	X_2	X_3
Y	1.00000	−0.78676	−0.60294	−0.64459
		(< .0001)	(< .0001)	(< .0001)
X_1	−0.78676	1.00000	0.56795	0.56968
	(< .0001)		(< .0001)	(< .0001)
X_2	−0.60294	0.56795	1.00000	0.67053
	(< .0001)	(< .0001)		(< .0001)
X_3	−0.64459	0.56968	0.67053	1.00000
	(0.0001)	(< .0001)	(< .0001)	

Although the correlation between X_1 and X_3 is significant and moderate (0.570), X_3 still explains some part of the response (Y) that was not explained by the variable X_1 as indicated by the extra sum of squares computation in the previous question ($SSR(X_1|X_3) \neq SSR(X_1)$). In other words, although the inclusion or exclusion of X_3 has a very mild effect on the coefficient estimate and standard

error of X_1, it contains some information that explains variation in Y that is not contained in X_1 despite the significant and moderate correlation between X_1 and X_3.
Also the correlation between X_2 and X_3 is significant and ≈ 0.7, and the inclusion or exclusion of X_3 does have an effect on both the coefficient estimate and standard error of X_2, both variables do not have the same information about the response Y, since $SSR(X_2|X_3) \neq SSR(X_2)$.

In summary, conclusions from the first two problems answered, i.e comparing parameter estimates and computing extra sum of squares gained by adding X_2 and X_1 to models containing X_3 led to the assertion that X_1 and X_3 are potentially not correlated (since the inclusion of X_3 had little effect ont the estimates of X_1), an assertion proved wrong by the results obtained from the correlation matrix. On the other hand, the change observed in the coefficient estimates of X_2 when X_3 was inserted into the model might lead one to think X_2 is possibly highly correlated with X_3 and possibly not much would be gained by adding X_2 to a model already containing X_3. This was somehow true as X_3 and X_2 has a significant correlation of ≈ 0.7 and only a gain of 707.99714 is observed when X_2 is added to a model containing X_3.

Appendix (SAS Codes and Graphs)

SAS Codes

```
data assignment;
infile 'C:\Users\OODOOE\Downloads\Video\Second Year\First Semester\
Foundation\foundation_assignment2\Data\CH06PR15.txt' firstobs=2;
input Y X1 X2 X3;
label Y='Patients Satisfaction'
  X1='Patients Age'
  X2='severity of illness'
  X3='anxiety level';
run;

*Question 1;
proc reg data=assignment;
model Y = X1 X2;
run;

*Question 2;
proc reg data=assignment;
model Y = X1 X2 X3 / ss1 ss2;
model Y = X1;
model Y = X2;
model Y = X3;
model Y = X1 X3 /ss1;
model Y = X2 X3 /ss1;
model Y = X3 X2 /ss1;
model Y = X3 X1 /ss1;
```

```
run;
*Question 3;
proc corr data=assignment plots=matrix(histogram);
var Y X1 X2 X3;
run;
```