Linear 6

Tonghao Zhang

2016/11/2

6.25

The analyst can modify the Y_i by deducing the X_{i2} component: $Y_i^* = Y_i - \beta_2 X_{i2}$. Then fit Y_i^* with covariates $X_1,\ X_3$ will yield desiring result.

6.26

Simple regression $\hat{Y}$ on Y will yield the exact set of coefficient H, for there exist only one unique least square solution. To be specific, we are looking for $c_0,\ c_1$ such that $\sum(c_0 + c_1\hat{Y}_i - Y_i)^2$ is minimized. As $\hat{Y}_i = b_0 + b_1 X_{1i} + ... + b_p X_{pi}$, the criterion is to minimize $\sum(c_0 + c_1 b_0 + c_1 b_1 X_{1i} + ... + c_1 b_p X_{pi})^2$. This is still a least square solution of X_i on Y, the uniqueness of least square ensures $c_0 = 0,\ c_1 = 1$, so the regression fit is identical to that of X_i and Y. Now that the fitted values are the same, SSE is the same, so will be R^2.

6.27

- b: [b_0,b_1,b_2] = [33.932, 2.784, -0.264]
- e: [-2.700, -1.230, -1.637, -1.330, -0.090, 6.987]
- H:

```
##             [,1]        [,2]        [,3]       [,4]        [,5]       [,6]
## [1,]  0.23143293  0.25167585  0.21178735  0.1488684 -0.05475543 0.21099091
## [2,]  0.25167585  0.31240459  0.09437844  0.2662773 -0.14787283 0.22313666
## [3,]  0.21178735  0.09437844  0.70442026 -0.3191744  0.10446672 0.20412159
## [4,]  0.14886839  0.26627729 -0.31917435  0.6142563  0.14143492 0.14833743
## [5,] -0.05475543 -0.14787283  0.10446672  0.1414349  0.94039955 0.01632707
## [6,]  0.21099091  0.22313666  0.20412159  0.1483374  0.01632707 0.19708635
```

- SSR: 3009.92
- $s^2\{b\}$:

```
##                       X1          X2
##    715.43035 -34.1569695 -13.5941622
## X1 -34.15697   1.6615716   0.6440307
## X2 -13.59416   0.6440307   0.2624528
```

- $\hat{Y}_h = 53.847$
- $s^2\{\hat{Y}_h\} = 5.424$

7.1

- $SSR(X_1|X_2)$ df = 1
- $SSR(X_2|X_1, X_3)$ df = 1
- $SSR(X_1, X_2|X_3, X_4)$ df = 2
- $SSR(X_1, X_2, X_3|X_4, X_5)$ df = 3

7.3

```r
data7.3<-read.table("CH06PR05.txt",header=FALSE)
colnames(data7.3)<-c("Y","X1","X2")
lm7.3<-lm(Y~X1+X2,data = data7.3)
anova(lm7.3)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X1         1 1566.45 1566.45 215.947 1.778e-09 ***
## X2         1  306.25  306.25  42.219 2.011e-05 ***
## Residuals 13   94.30    7.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a)

$SSR(X_1) = 1566.45$, $SSR(X_2|X_1) = 306.25$.

b)

$H_0 : \beta_2 = 0$, $H_A : \beta_2 \neq 0$

decision rule: if $F^* > F(0.99; 1, 13)$, reject H_0.

$F^* = \frac{306.25}{7.25}$, while threshold $= 9.0738057$, reject H_0. The p-value is 2.0056483×10^{-5}.

7.6

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X1         1 8275.4  8275.4 81.8026 2.059e-11 ***
## X2         1  480.9   480.9  4.7539   0.03489 *
## X3         1  364.2   364.2  3.5997   0.06468 .
## Residuals 42 4248.8   101.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a)

$SSR(X_2, X_3|X_1) = 480.9 + 364.2$

$H_0 : \beta_2 = \beta_3 = 0$, H_A: at least one $\beta_i \neq 0$.

decision rule: if $F^* > F(0.975 : 2, 42)$, reject H_0.

$F^* = \frac{\frac{480.9+364.2}{2}}{101.2} = 4.175$, while the threshold $F(0.975; 2, 42) = 4.0327099$, reject H_0. p-value is 0.0221872.

7.11

a)

- $R^2_{Y1} = \frac{SSR(X_1)}{SSTO} = 0.550$, it measures the usefulness of fitting X_1 to Y in linear regression model. Here 0.550 is far from 1, the fit is not very useful.
- $R^2_{Y2} = \frac{SSR(X_2)}{SSTO} = 0.409$, it measures the usefulness of fitting X_2 to Y in linear regression model. Here 0.409 means this model is not very useful.
- $R^2_{12} = \frac{SSR(X_1)}{SSE(X_2)} = 0$, it measures the correlation of X_1 and X_2. 0 basicly says there is no correlation between X_1 and X_2.
- $R^2_{Y1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = 0.930$, it measures the effectiveness of adding X_1 into linear regression model given X_2 is already in the model. 0.930 indicates it's very effective.
- $R^2_{Y2|1} = \frac{SSR(X_2|X_1)}{SSE(X_1)} = 0.907$, it measures the effectiveness of adding X_2 into linear regression model given X_1 is already in the model. 0.907 indicates it's very effective.

b)

Because the two variable are uncorrelated. $R^2_{12} = 0$. Furthermore the extra sum of variance explain by adding new variable is the same regardless of $X_1,\ X_2$ sequence.

Problem 9

part A

Just plug $X = X^* C^t$ into $X^t Y = X^t X \hat{\beta}$, $CX^{*t} Y = CX^{*t} X^* C^t \hat{\beta}$, cancel C on both side yield the desire result $X^{*t} Y = X^{*t} X^* \hat{\beta}^*$.

part B

```
Xt%*%X
```

```
##      x1 x2  x3
##    6  0  0   0
## x1 0 70  0   0
## x2 0  0 84   0
## x3 0  0  0 180
```

```
round(solve(Xt%*%X),3)
```

```
##             x1    x2    x3
##    0.167 0.000 0.000 0.000
## x1 0.000 0.014 0.000 0.000
## x2 0.000 0.000 0.012 0.000
## x3 0.000 0.000 0.000 0.006
```

```
round(solve(Xt%*%X)%*%Xt%*%Y,3)
```

```
##     [,1]
##    4.400
## x1 0.583
## x2 0.239
## x3 0.074
```

```r
round(t(C)%*%(solve(Xt%*%X)%*%Xt%*%Y),3)
```

```
##        [,1]
## [1,] 3.353
## [2,] 1.360
## [3,] 2.243
## [4,] 1.924
```

part C