Q5

Zhen Zhang

4/24/2020

```
library(car)
```

```
## Loading required package: carData
```

```
commericial = read.table("../data/commercial.txt", header=T)
names(commericial)
```

```
## [1] "y"  "x1" "x2" "x3" "x4"
```

```
head(commericial)
```

```
##      y x1    x2   x3     x4
## 1 13.5  1  5.02 0.14 123000
## 2 12.0 14  8.19 0.27 104079
## 3 10.5 16  3.00 0.00  39998
## 4 15.0  4 10.70 0.05  57112
## 5 14.0 11  8.97 0.07  60000
## 6 10.5 15  9.45 0.24 101385
```

(a) Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X4; with X1 given X4; with X2 given X1 and X4; and with X3 given X1, X2 and X4. Hint use the lm and anova functions.

```
fit = lm(y ~ x4 + x1 + x2 + x3, commericial)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x4         1 67.775  67.775 52.4369 3.073e-10 ***
## x1         1 42.275  42.275 32.7074 2.004e-07 ***
## x2         1 27.857  27.857 21.5531 1.412e-05 ***
## x3         1  0.420   0.420  0.3248    0.5704
## Residuals 76 98.231   1.293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b) Test whether X3 can be dropped from the regression model given that X1, X2 and X4 are retained. Use the F test statistic and level of significance of .01. State the null and alternative hypotheses, test statistic, P-value and decision.

```
drop1(fit, test = "F")
```

```
## Single term deletions
##
## Model:
## y ~ x4 + x1 + x2 + x3
##        Df Sum of Sq     RSS    AIC F value    Pr(>F)
## <none>               98.231 25.622
## x4      1    42.325 140.556 52.643 32.7464 1.976e-07 ***
## x1      1    57.243 155.473 60.814 44.2881 3.894e-09 ***
## x2      1    25.759 123.990 42.486 19.9294 2.747e-05 ***
## x3      1     0.420  98.650 23.968  0.3248    0.5704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c) Test whether both X2 and X3 can be dropped from the regression model given that X1 and X4 are retained; use alpha = 0.01. State the null and alternative hypotheses, test statistic, P -value and decision. Hint: use the pf function to find P values.

```
fit2 = lm(y ~ x1 + x4, commericial)
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x4, data = commericial)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2032 -0.4593  0.0641  0.7730  2.5083
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.436e+01  2.771e-01  51.831  < 2e-16 ***
## x1          -1.145e-01  2.242e-02  -5.105 2.27e-06 ***
## x4           1.045e-05  1.363e-06   7.663 4.23e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.274 on 78 degrees of freedom
## Multiple R-squared:  0.4652, Adjusted R-squared:  0.4515
## F-statistic: 33.93 on 2 and 78 DF,  p-value: 2.506e-11
```

```
"???"
```

```
## [1] "???"
```

(d) Find the variance inflation factors for the full model with all four predictors in the model. What do they tell you?

```
vif(fit)
```

```
##       x4       x1       x2       x3
## 1.412722 1.240348 1.648225 1.323552
```