Regression Analysis HW-02

Likhit Garimella
lg836

Q(3.7)

```
# Likhit Garimella
# Regression Analysis HW-2

# libraries
#install.packages("faraway")
#install.packages("olsrr")
#install.packages("psych")
#install.packages("readr")
library(faraway)
library(olsrr)
library(psych)
library(readr)

# import dataset
data1 <- read.csv("/Users/likhitgarimella/Desktop/
SummerSemester/B4.csv", header = T, sep = ",")
data1
```

(a) Fitting the multiple regression model:-

```
# (a) fitting the multiple regression model
fitMod <- lm(y ~ ., data = data1)
sum1 <- summary(fitMod)
fitMod
```

Text output:-

```
> # import dataset
> data1 <- read.csv("/Users/likhitgarimella/Desktop/
SummerSemester/B4.csv", header = T, sep = ",")
> data1
     y     x1  x2     x3    x4  x5 x6 x7 x8 x9
1  29.5 5.0208 1.0 3.5310 1.500 2.0  7  4 62  0
2  27.9 4.5429 1.0 2.2750 1.175 1.0  6  3 40  0
3  25.9 4.5573 1.0 4.0500 1.232 1.0  6  3 54  0
4  29.9 5.0597 1.0 4.4550 1.121 1.0  6  3 42  0
5  29.9 3.8910 1.0 4.4550 0.988 1.0  6  3 56  0
6  30.9 5.8980 1.0 5.8500 1.240 1.0  7  3 51  1
7  28.9 5.6039 1.0 9.5200 1.501 0.0  6  3 32  0
8  35.9 5.8282 1.0 6.4350 1.225 2.0  6  3 32  0
9  31.5 5.3003 1.0 4.9883 1.552 1.0  6  3 30  0
10 31.0 6.2712 1.0 5.5200 0.975 1.0  5  2 30  0
11 30.9 5.9592 1.0 6.6660 1.121 2.0  6  3 32  0
12 30.0 5.0500 1.0 5.0000 1.020 0.0  5  2 46  1
13 36.9 8.2464 1.5 5.1500 1.664 2.0  8  4 50  0
14 41.9 6.6969 1.5 6.9020 1.488 1.5  7  3 22  1
15 40.5 7.7841 1.5 7.1020 1.376 1.0  6  3 17  0
16 43.9 9.0384 1.0 7.8000 1.500 1.5  7  3 23  0
17 37.5 5.9894 1.0 5.5200 1.256 2.0  6  3 40  1
18 37.9 7.5422 1.5 5.0000 1.690 1.0  6  3 22  0
19 44.5 8.7951 1.5 9.8900 1.820 2.0  8  4 50  1
20 37.9 6.0831 1.5 6.7265 1.652 1.0  6  3 44  0
21 38.9 8.3607 1.5 9.1500 1.777 2.0  8  4 48  1
22 36.9 8.1400 1.0 8.0000 1.504 2.0  7  3  3  0
23 45.8 9.1416 1.5 7.3262 1.831 1.5  8  4 31  0
24 25.9 4.9176 1.0 3.4720 0.998 1.0  7  4 42  0
>
> # (a) fitting the multiple regression model
> fitMod <- lm(y ~ ., data = data1)
> sum1 <- summary(fitMod)
> fitMod

Call:
lm(formula = y ~ ., data = data1)

Coefficients:
(Intercept)          x1          x2          x3          x4
x5          x6          x7          x8          x9
   14.92765     1.92472     7.00053     0.14918
2.72281     2.00668    -0.41012    -1.40324
-0.03715     1.55945
```

The linear regression model between sale price of the house/1000 and all regressors is :-
(taking the obtained values as coefficients)

$$\boxed{\hat{y} = 14.92765 + 1.92472x_1 + 7.00053x_2 + 0.14918x_3 + 2.72281x_4 + 2.00668x_5 - 0.41012x_6 - 1.40324x_7 - 0.03715x_8 + 1.55945x_9}$$

Coeff of $x_1, x_2, x_3, x_4, x_5, x_9$ <u>are > 0</u>
↳ indicates positive relationships among them

Coeff of x_6, x_7, x_8 <u>are < 0</u>
↳ indicating negative relationships among them

<u>Example</u>:-
No. of baths x_2 ↑ 1 unit ⇒ selling price ↑ 7 units.
&
No. of bed x_7 ↑ 1 unit ⇒ selling price ↓ 1.4 units.

(b) Test for significance of regression :-

```
>
> # (b) test for significance of regression -> output
from console
> sum1

Call:
lm(formula = y ~ ., data = data1)

Residuals:
   Min     1Q Median     3Q    Max 
-3.720 -1.956 -0.045  1.627  4.253 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 14.92765    5.91285   2.525   0.0243 *
x1           1.92472    1.02990   1.869   0.0827 .
x2           7.00053    4.30037   1.628   0.1258  
x3           0.14918    0.49039   0.304   0.7654  
x4           2.72281    4.35955   0.625   0.5423  
x5           2.00668    1.37351   1.461   0.1661  
x6          -0.41012    2.37854  -0.172   0.8656  
x7          -1.40324    3.39554  -0.413   0.6857  
x8          -0.03715    0.06672  -0.557   0.5865  
x9           1.55945    1.93750   0.805   0.4343  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.949 on 14 degrees of
freedom
Multiple R-squared:  0.8531, Adjusted R-squared:
0.7587
F-statistic: 9.037 on 9 and 14 DF,  p-value: 0.000185
```

F-stat is 9.037 & p-value is 2×10^{-4}
↳ which is less than significance level 0.05

⇒ so, we would reject $H_0: \beta_1 = \cdots = \beta_9 = 0$ & conclude there is a linear relationship b/w selling price y and any of the regressors $x_1, x_2, \ldots, x_9$

(c)

→ From results in (b), we can see that p-values for t-stat of $\beta_1, \beta_2, \dots \beta_9$ are all greater than the significance level 0.05.

→ i.e., none of the regressors are significant in the linear model.

→ ∴ we may face with multicolinearity problem in this linear reg. model.

(d)

```
>
> # (d)
> newMod <- lm(y ~ x1 + x2 + x5 + x6 + x7 + x8 + x9,
data = data1)
> sum2 <- summary(newMod)
>
> anov1 <- anova(fitMod)
> anov2 <- anova(newMod)
>
> F_34 <- round((sum(anov1$`Sum Sq`[1:9])-
sum(anov2$`Sum Sq`[1:7]))/2/anov1$`Mean Sq`[10], 4)
> p_34 <- round(1-pf(F_34, 2, 14), 4)
>
```

The partial F test measures the contribution of the regressors in x_3 & x_4 given that the other regressor in the model, where F statistics value is 0.3225 & correspod. p-value is 0.7296.

The p-value is greater than significance level 0.05 ⇒ there is no contribution of lot size & living space given that all the other regressors are in the model.

(e) > # (e)

```
> vif(fitMod)
       x1        x2        x3        x4        x5        x6        x7        x8        x9 
 7.021036  2.835413  2.454907  3.836477  1.823605 11.710101  9.722663  2.320887  1.942494 
```

→ From parts b,c,d , we find insignificant coeff. of all regressors , while the regression is significant based on F statistics.

→ And VIF for x_6 is > 10.

→ This situation indicates multicollinearity in this model.