RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Assignment 2.R | Assignment 2 - Daniel Yim.Rmd | Assignment 3.R | Assignment 5 - Daniel Yim.R | Assignment 5 - Daniel Yim.Rmd | Assignment 3 - Daniel Yim.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline

```
---
title: "Assignment 5 - Daniel Yim"
author: "Daniel Yim"
date: "2023-04-10"
output:
  pdf_document: default
  word_document: default
---

#Assignment 5

#1. Perform two separate regression analysis models on this dataset using as target variable (dependent variable) y1
# and then using y2. (20 marks)

#Linear Regression: Y1 Heating Load, X1 Relative Compactness

```{r}
energy_eff_bdg <-lm(Y1 ~ X1, data = ENB2012_data)
summary(energy_eff_bdg)
```
```

```
Call:
lm(formula = Y1 ~ X1, data = ENB2012_data)

Residuals:
    Min      1Q  Median      3Q     Max
-19.569  -6.332  -1.028   3.393  19.259

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -23.053      2.081  -11.08   <2e-16 ***
X1            59.359      2.698   22.00   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.904 on 766 degrees of freedom
Multiple R-squared:  0.3872,    Adjusted R-squared:  0.3864
F-statistic:   484 on 1 and 766 DF,  p-value: < 2.2e-16
```

```
#Linear Regression: Y1 Heating Load, X2 Surface Area

```{r}
energy_eff_bdg <-lm(Y1 ~ X2, data = ENB2012_data)
summary(energy_eff_bdg)
```
```

```
Call:
lm(formula = Y1 ~ X2, data = ENB2012_data)

Residuals:
    Min      1Q  Median      3Q     Max
-18.609  -5.524  -1.300   3.529  18.176

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 72.945395   2.111064   34.55   <2e-16 ***
X2          -0.075387   0.003116  -24.19   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.602 on 766 degrees of freedom
Multiple R-squared:  0.4331,    Adjusted R-squared:  0.4324
F-statistic: 585.3 on 1 and 766 DF,  p-value: < 2.2e-16
```

```
#Linear Regression: Y1 Heating Load, X3 Wall Area
```{r}
energy_eff_bdg <-lm(Y1 ~ X3, data = ENB2012_data)
summary(energy_eff_bdg)
```
```

```
Call:
lm(formula = Y1 ~ X3, data = ENB2012_data)

Residuals:
     Min       1Q   Median       3Q      Max
-19.0213  -7.3937  -0.4882   7.5728  18.2107
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.259681   2.391323  -4.709 2.96e-06 ***
X3            0.105391   0.007439  14.168  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.988 on 766 degrees of freedom
Multiple R-squared:  0.2076,    Adjusted R-squared:  0.2066
F-statistic: 200.7 on 1 and 766 DF,  p-value: < 2.2e-16
```

#Linear Regression: Y1 Heating Load, X4 Roof Area

```{r}
energy_eff_bdg <-lm(Y1 ~ X4, data = ENB2012_data)
summary(energy_eff_bdg)
```

```
Call:
lm(formula = Y1 ~ X4, data = ENB2012_data)

Residuals:
     Min       1Q   Median       3Q      Max
-19.5327  -2.6392  -0.3191   2.4997  15.0930

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 56.309657   0.746269   75.45   <2e-16 ***
X4          -0.192535   0.004094  -47.03   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.121 on 766 degrees of freedom
Multiple R-squared:  0.7427,    Adjusted R-squared:  0.7424
F-statistic:  2212 on 1 and 766 DF,  p-value: < 2.2e-16
```

#Linear Regression: Y1 Heating Load, X5 Overall Height

```{r}
energy_eff_bdg <-lm(Y1 ~ X5, data = ENB2012_data)
summary(energy_eff_bdg)
```

```
Call:
lm(formula = Y1 ~ X5, data = ENB2012_data)

Residuals:
     Min       1Q   Median       3Q      Max
-15.7259  -2.5929  -0.3085   2.0015  11.8241

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.59887    0.52661  -8.733   <2e-16 ***
X5           5.12497    0.09516  53.857   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.615 on 766 degrees of freedom
Multiple R-squared:  0.7911,    Adjusted R-squared:  0.7908
F-statistic:  2901 on 1 and 766 DF,  p-value: < 2.2e-16
```

#Linear Regression: Y1 Heating Load, X6 Orientation

```{r}
energy_eff_bdg <-lm(Y1 ~ X6, data = ENB2012_data)
summary(energy_eff_bdg)
```

```
Call:
lm(formula = Y1 ~ X6, data = ENB2012_data)

Residuals:
    Min      1Q  Median      3Q     Max
-16.285  -9.320  -3.345   9.375  20.781

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.38885    1.19733  18.699   <2e-16 ***
X6          -0.02333    0.32587  -0.072    0.943
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.1 on 766 degrees of freedom
Multiple R-squared:  6.691e-06, Adjusted R-squared:  -0.001299
F-statistic: 0.005126 on 1 and 766 DF,  p-value: 0.9429
```

#Linear Regression: Y1 Heating Load, X7 Glazing Area

```{r}
energy_eff_bdg <-lm(Y1 ~ X7, data = ENB2012_data)
summary(energy_eff_bdg)
```

```
Call:
lm(formula = Y1 ~ X7, data = ENB2012_data)

Residuals:
    Min      1Q  Median      3Q     Max
-13.272  -9.193  -3.054   7.253  17.699

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.5170     0.7103  24.662  < 2e-16 ***
X7           20.4380     2.6351   7.756  2.8e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.722 on 766 degrees of freedom
Multiple R-squared:  0.07281,   Adjusted R-squared:  0.0716
F-statistic: 60.16 on 1 and 766 DF,  p-value: 2.796e-14
```

#Linear Regression: Y1 Heating Load, X8 Glazing Area Distribution

```{r}
energy_eff_bdg <-lm(Y1 ~ X8, data = ENB2012_data)
summary(energy_eff_bdg)
```

```
Call:
lm(formula = Y1 ~ X8, data = ENB2012_data)

Residuals:
    Min      1Q  Median      3Q     Max
-14.699  -8.968  -3.132   8.796  21.343

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.7086     0.7520  27.539   <2e-16 ***
X8            0.5684     0.2342   2.427   0.0154 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.06 on 766 degrees of freedom
Multiple R-squared:  0.007633,  Adjusted R-squared:  0.006338
F-statistic: 5.892 on 1 and 766 DF,  p-value: 0.01544
```

# Before performing the multivariable regression analysis run some tests to see if the regression model assumptions
# (i.e., multicollinearity, normal errors, etc) are met. (20 marks)

#Multiple Regression Analysis Y2 Cooling Load:

```{r}
energy_eff_bdg <-lm(Y2 ~ ., data = ENB2012_data)
summary(energy_eff_bdg)
```

```
Call:
lm(formula = Y2 ~ ., data = ENB2012_data)

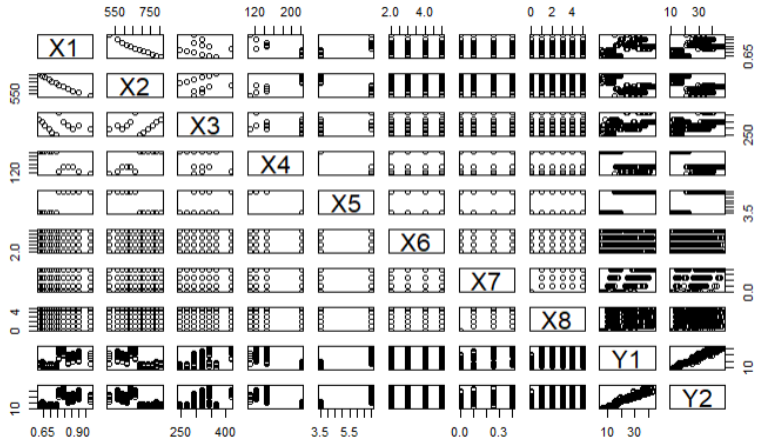
Residuals:
    Min      1Q  Median      3Q     Max
-4.9848 -1.1019 -0.1268  0.9315  7.0398

Coefficients: (1 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.435514  12.781437   1.912  0.05628 .
X1          -14.651807   6.998282  -2.094  0.03662 *
X2           -0.012596   0.011515  -1.094  0.27438
X3           -0.008022   0.004644  -1.727  0.08454 .
X4                 NA         NA      NA       NA
X5            0.669953   0.245532   2.729  0.00651 **
X6            0.141730   0.062799   2.257  0.02430 *
X7           -2.557639   0.721915  -3.543  0.00042 ***
X8           -0.135906   0.046619  -2.915  0.00366 **
Y1            0.866650   0.024052  36.032  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.946 on 759 degrees of freedom
Multiple R-squared:  0.9586,    Adjusted R-squared:  0.9582
F-statistic:  2197 on 8 and 759 DF,  p-value: < 2.2e-16
```

```
#Check for multicollinearity:

#visually:
```{r}
pairs(ENB2012_data)
```
```




```
#in terms of correlation numbers:
```{r}
cor(ENB2012_data)
```
```

```
              X1            X2          X3            X4          X5           X6           X7          X8           Y1
X1  1.000000e+00 -9.919015e-01 -0.2037817 -8.688234e-01  0.8277473  0.000000000 7.617400e-20 0.00000000  0.622271936
X2 -9.919015e-01  1.000000e+00  0.1955016  8.807195e-01 -0.8581477  0.000000000 4.664140e-20 0.00000000 -0.658119917
X3 -2.037817e-01  1.955016e-01  1.0000000 -2.923165e-01  0.2809757  0.000000000 0.000000e+00 0.00000000  0.455671365
X4 -8.688234e-01  8.807195e-01 -0.2923165  1.000000e+00 -0.9725122  0.000000000 -1.197187e-19 0.00000000 -0.861828052
X5  8.277473e-01 -8.581477e-01  0.2809757 -9.725122e-01  1.0000000  0.000000000 0.000000e+00 0.00000000  0.889430464
X6  0.000000e+00  0.000000e+00  0.0000000  0.000000e+00  0.0000000  1.000000000 0.000000e+00 0.00000000 -0.002586763
X7  7.617400e-20  4.664140e-20  0.0000000 -1.197187e-19  0.0000000  0.000000000 1.000000e+00 0.21296422  0.269841685
X8  0.000000e+00  0.000000e+00  0.0000000  0.000000e+00  0.0000000  0.000000000 2.129642e-01 1.00000000  0.087368460
Y1  6.222719e-01 -6.581199e-01  0.4556714 -8.618281e-01  0.8894305 -0.002586763 2.698417e-01 0.08736846  1.000000000
Y2  6.343391e-01 -6.729989e-01  0.4271170 -8.625466e-01  0.8957852  0.014289598 2.075050e-01 0.05052512  0.975861739
           Y2
X1  0.63433907
X2 -0.67299893
X3  0.42711700
X4 -0.86254660
X5  0.89578517
X6  0.01428960
X7  0.20750499
X8  0.05052512
Y1  0.97586174
Y2  1.00000000
```

```
#vif scores: not working for me
#install.packages("caTools")
#library(caTools)
#install.packages('car')
#library('car')
#energy_eff_bdg <-lm(Y2 ~ ., data = ENB2012_data)
#vif(mod)

#2. Which variables are significant in both models (with y1 and y2)? (20 marks)

#For Y1: significance found with x1, x2, x3, x4, x5, x7, x8

#For Y2: significance found with x1, x5, x6, x7, x8

#In both models: X1, X5, X7, and X8 are significant

# Remove least significant variables
```{r}
energy_eff_bdg2 <- lm(Y2 ~ . - X2 - X3 - X4 - X6 - X7,data = ENB2012_data)
summary(energy_eff_bdg2)
```
```

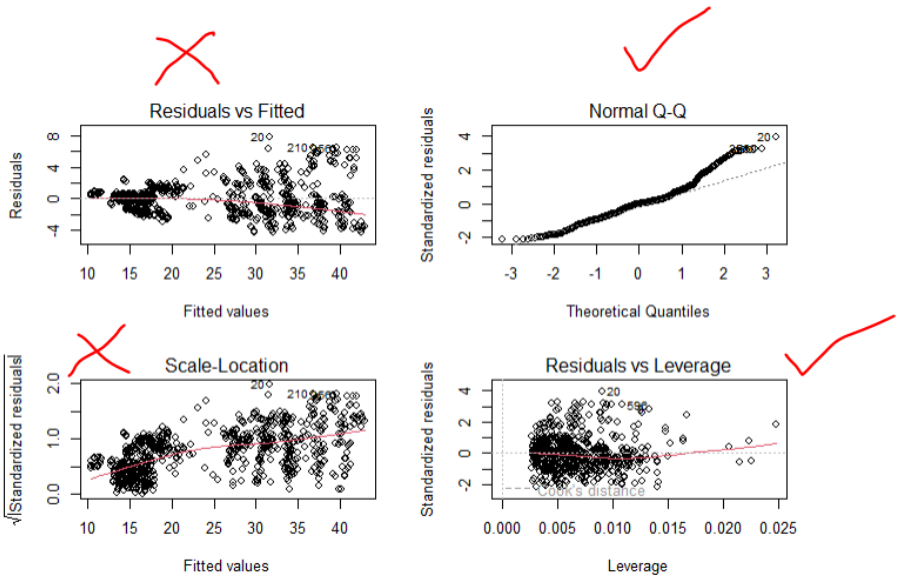
```
Call:
lm(formula = Y2 ~ . - X2 - X3 - X4 - X6 - X7, data = ENB2012_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2465 -1.2952 -0.0042  0.7360  7.7676

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.61456    0.68697   6.717 3.62e-11 ***
X1          -2.95905    1.35103  -2.190  0.02881 *
X5           0.91035    0.14150   6.433 2.20e-10 ***
X8          -0.14526    0.04719  -3.078  0.00216 **
Y1           0.80080    0.01771  45.209  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.98 on 763 degrees of freedom
Multiple R-squared:  0.9569,    Adjusted R-squared:  0.9567
F-statistic:  4235 on 4 and 763 DF,  p-value: < 2.2e-16
```

```{r}
par(mfrow=c(2,2))
plot(energy_eff_bdg2)
```




```
#3. Interpret some of the coefficients of the variables in the models you run (20 marks)
# For Y2 Cooling Load, the variable is reduced by X1, X2, X3, X7, and X8.

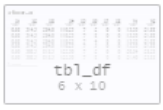
#4. Is your model overall significant or well specified (hint: Fisher) (20 marks)


#perform Fisher's LSD
```{r}
install.packages('agricolae')
library(agricolae)
head(ENB2012_data)
energy_eff_bdg3 <- aov(X1 ~ X8,data = ENB2012_data)
summary(energy_eff_bdg3)
print(LSD.test(energy_eff_bdg3,"X1"))
```
```

```r
#3. Interpret some of the coefficients of the variables in the models you run (20 marks)
# For Y2 Cooling Load, the variable is reduced by X1, X2, X3, X7, and X8.

#4. Is your model overall significant or well specified (hint: Fisher) (20 marks)


#perform Fisher's LSD
```{r}
install.packages('agricolae')
library(agricolae)
head(ENB2012_data)
energy_eff_bdg3 <- aov(X1 ~ X8,data = ENB2012_data)
summary(energy_eff_bdg3)
print(LSD.test(energy_eff_bdg3,"X1"))
```
```

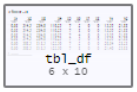



```
0.71 0.71    0 64 0.684027 0.735973 0.71 0.71 0.71 0.71 0.71
0.74 0.74    0 64 0.714027 0.765973 0.74 0.74 0.74 0.74 0.74
0.76 0.76    0 64 0.734027 0.785973 0.76 0.76 0.76 0.76 0.76
0.79 0.79    0 64 0.764027 0.815973 0.79 0.79 0.79 0.79 0.79
0.82 0.82    0 64 0.794027 0.845973 0.82 0.82 0.82 0.82 0.82
0.86 0.86    0 64 0.834027 0.885973 0.86 0.86 0.86 0.86 0.86
0.9  0.90    0 64 0.874027 0.925973 0.90 0.90 0.90 0.90 0.90
0.98 0.98    0 64 0.954027 1.005973 0.98 0.98 0.98 0.98 0.98

$comparison
NULL

$groups
       X1 groups
0.98 0.98      a
0.9  0.90      b
0.86 0.86      c
0.82 0.82      d
0.79 0.79     de
0.76 0.76     ef
0.74 0.74     fg
0.71 0.71     gh
0.69 0.69     hi
0.66 0.66     ij
0.64 0.64     jk
0.62 0.62      k

attr(,"class")
[1] "group"
```




A tibble: 6 × 10

| X1 <dbl> | X2 <dbl> | X3 <dbl> | X4 <dbl> | X5 <dbl> | X6 <dbl> | X7 <dbl> | X8 <dbl> | Y1 <dbl> | Y2 <dbl> |
|---|---|---|---|---|---|---|---|---|---|
| 0.98 | 514.5 | 294.0 | 110.25 | 7 | 2 | 0 | 0 | 15.55 | 21.33 |
| 0.98 | 514.5 | 294.0 | 110.25 | 7 | 3 | 0 | 0 | 15.55 | 21.33 |
| 0.98 | 514.5 | 294.0 | 110.25 | 7 | 4 | 0 | 0 | 15.55 | 21.33 |
| 0.98 | 514.5 | 294.0 | 110.25 | 7 | 5 | 0 | 0 | 15.55 | 21.33 |
| 0.90 | 563.5 | 318.5 | 122.50 | 7 | 2 | 0 | 0 | 20.84 | 28.28 |
| 0.90 | 563.5 | 318.5 | 122.50 | 7 | 3 | 0 | 0 | 21.46 | 25.38 |

6 rows