Lab 6 Solutions

Name: Kaushik Raj V Nadar

Roll No.: 208160499

Date: 10/09/2024

P 1. Consider the **Delivery Time** data set. Calculate the following measures of outliers:

(a) Leverage point,

(b) COOK'S D,

(c) DFBETAS,

(d) DFFITS,

(e) COVRATIO.

Based on the above measures, find the outliers present in the data.

Code:

```
library(car)
library(readxl)

TimeDeliveryData <- read_excel("TimeDeliveryData.xlsx") %>% as.data.frame()

model <- lm(Y ~ X1 + X2, data = TimeDeliveryData)

leverage_points <- hatvalues(model)
cooks_d <- cooks.distance(model)
dfbetas_values <- dfbetas(model)
dffits_values <- dffits(model)
covratio_values <- covratio(model)

# Print the Results
leverage_points
cooks_d
dfbetas_values
dffits_values
covratio_values

influence_measures <- influence.measures(model)
summary(influence_measures)
```

Output:

```
> leverage_points
         1          2          3          4          5          6 
0.10180178 0.07070164 0.09873476 0.08537479 0.07501050 0.04286693 
         7          8          9         10         11         12 
0.08179867 0.06372559 0.49829216 0.19629595 0.08613260 0.11365570 
        13         14         15         16         17         18 
```

```
0.06112463 0.07824332 0.04111077 0.16594043 0.05943202 0.09626046
        19         20         21         22         23         24
0.09644857 0.10168486 0.16527689 0.39157522 0.04126005 0.12060826
        25
0.06664345

> cooks_d
           1            2            3            4            5
1.000921e-01 3.375704e-03 9.455785e-06 7.764718e-02 5.432217e-04
           6            7            8            9           10
1.231067e-04 2.171604e-03 3.051135e-03 3.419318e+00 5.384516e-02
          11           12           13           14           15
1.619975e-02 1.596392e-03 2.294737e-03 3.292786e-03 6.319880e-04
          16           17           18           19           20
3.289086e-03 4.013419e-04 4.397807e-02 1.191868e-02 1.324449e-01
          21           22           23           24           25
5.086063e-02 4.510455e-01 2.989892e-02 1.023224e-01 1.084694e-04

> dfbetas_values
    (Intercept)            X1           X2
1  -0.187267279  0.4113118750 -0.434862094
2   0.089793299 -0.0477642427  0.014414155
3  -0.003515177  0.0039483525 -0.002846468
4   0.451964743  0.0882802920 -0.273373097
5  -0.031674102 -0.0133001129  0.024240457
6  -0.014681480  0.0017921068  0.001078986
7   0.078071285 -0.0222783194 -0.011018802
8   0.071202807  0.0333823324 -0.053823961
9  -2.575739806  0.9287433421  1.507550618
10  0.107919369 -0.3381628707  0.341326746
11 -0.034274535  0.0925271540 -0.002686252
12 -0.030268935 -0.0486664488  0.053973390
13  0.072366473 -0.0356212226  0.011335105
14  0.049516699 -0.0670868604  0.061816778
15  0.022279094 -0.0047895025  0.006838236
16 -0.002693186  0.0644208340 -0.084187552
17  0.028855555  0.0064876499 -0.015696507
18  0.248558020  0.1897331043 -0.272430555
19  0.172558506  0.0235737344 -0.098968842
20  0.168036548 -0.2149950233 -0.092915080
21 -0.161928685 -0.2971750929  0.336406248
22  0.398566309 -1.0254140704  0.573140240
23 -0.159852248  0.0372930389 -0.052651959
24 -0.119720216  0.4046225960 -0.465446949
25 -0.016816024  0.0008498979  0.005592192

> dffits_values
           1            2            3            4            5
-0.570850478  0.098618619 -0.005203676  0.500801817 -0.039458989
```

```
      6           7           8           9          10
-0.018779374  0.078990030  0.093760764  4.296080927  0.398713071
       11          12          13          14          15
 0.217953207 -0.067670223  0.081259033  0.097362643  0.042584374
       16          17          18          19          20
-0.097159801  0.033915978  0.365309285  0.186167873 -0.671771402
       21          22          23          24          25
-0.388501185 -1.195036104 -0.307538544 -0.571139627 -0.017626149

> covratio_values
        1         2         3         4         5         6         7
0.8710782 1.2149209 1.2756813 0.8759964 1.2396032 1.1999120 1.2397501
        8         9        10        11        12        13        14
1.2056413 0.3422132 1.3054035 1.1717266 1.2906069 1.2070490 1.2276758
       15        16        17        18        19        20        21
1.1918460 1.3692181 1.2192451 1.0692145 1.2152541 0.7598217 1.2376914
       22        23        24        25
1.3980787 0.8896761 0.9476321 1.2310981

> summary(influence_measures)
Potentially influential observations of
	 lm(formula = Y ~ X1 + X2, data = TimeDeliveryData) :

   dfb.1_  dfb.X1  dfb.X2  dffit   cov.r   cook.d  hat   
9  -2.58_*  0.93    1.51_*  4.30_*  0.34_*  3.42_*  0.50_*
22  0.40   -1.03_*  0.57   -1.20_*  1.40    0.45    0.39_*
```

P 2. Consider the **Acetylene** data set.

(a) Normalise the data set by subtracting the associated mean and dividing by the square root of $(n-1)\times$ the sample variance of each variables.

(b) Consider all the regressors including the interactions and second order terms. Find the correlation matrix. Observe the correlation.

(c) Find the VIF for the regression model where all the regressors including the interactions and second order terms are considered.

Code:

```
library(car)
library(dplyr)
library(readxl)

Acetylene_Data <- read_excel("Acetylene_Data.xlsx") %>% as.data.frame()

# Normalize the dataset
acetylene_norm <- lapply(Acetylene_Data, function(x) ((x - mean(x)) / sqrt((length(x) - 1) * var(x)))) %>%
as.data.frame()
```

```
# Create interaction and second-order terms
acetylene_final <- cbind(acetylene_norm,
                    X1X2 = acetylene_norm$X1 * acetylene_norm$X2,
                    X1X3 = acetylene_norm$X1 * acetylene_norm$X3,
                    X2X3 = acetylene_norm$X2 * acetylene_norm$X3,
                    X1_sq = acetylene_norm$X1^2,
                    X2_sq = acetylene_norm$X2^2,
                    X3_sq = acetylene_norm$X3^2)

# Calculate correlation matrix
corr <- cor(acetylene_final)

# Print correlation matrix
cat("Correlation Matrix: ")
corr

# Calculate VIF
model <- lm(Y ~ ., data = acetylene_final)
vif_values <- vif(model)

# Print VIF values
cat("VIF Values: ", vif_values)
```

Output:

```
> acetylene_norm
             Y          X1          X2           X3
1   0.27978969  0.28022427 -0.22544387 -0.231055577
2   0.30582925  0.28022427 -0.15704118 -0.231055577
3   0.31233914  0.28022427 -0.06583759 -0.235136028
4   0.26893987  0.28022427  0.04816689 -0.222894673
5   0.24724023  0.28022427  0.20777317 -0.218814222
6   0.18214132  0.28022427  0.48138393 -0.231055577
7  -0.17590268 -0.04003204 -0.32576781 -0.002550282
8  -0.09995395 -0.04003204 -0.22544387 -0.018872089
9  -0.03485504 -0.04003204 -0.06583759 -0.067837509
10 -0.02400522 -0.04003204  0.04816689 -0.116802929
11  0.04109369 -0.04003204  0.20777317 -0.051515703
12  0.05194350 -0.04003204  0.48138393  0.005610621
13 -0.45799795 -0.36028835 -0.32576781  0.356529466
14 -0.41459868 -0.36028835 -0.22544387  0.470782113
15 -0.33864995 -0.36028835 -0.06583759  0.421816693
16 -0.14335322 -0.36028835  0.20777317  0.372851273

> acetylene_final
             Y          X1          X2           X3         X1X2
1   0.27978969  0.28022427 -0.22544387 -0.231055577 -0.063174843
2   0.30582925  0.28022427 -0.15704118 -0.231055577 -0.044006749
3   0.31233914  0.28022427 -0.06583759 -0.235136028 -0.018449290
4   0.26893987  0.28022427  0.04816689 -0.222894673  0.013497533
```

```
5   0.24724023  0.28022427  0.20777317 -0.218814222  0.058223085
6   0.18214132  0.28022427  0.48138393 -0.231055577  0.134895461
7  -0.17590268 -0.04003204 -0.32576781 -0.002550282  0.013041150
8  -0.09995395 -0.04003204 -0.22544387 -0.018872089  0.009024978
9  -0.03485504 -0.04003204 -0.06583759 -0.067837509  0.002635613
10 -0.02400522 -0.04003204  0.04816689 -0.116802929 -0.001928219
11  0.04109369 -0.04003204  0.20777317 -0.051515703 -0.008317584
12  0.05194350 -0.04003204  0.48138393  0.005610621 -0.019270780
13 -0.45799795 -0.36028835 -0.32576781  0.356529466  0.117370347
14 -0.41459868 -0.36028835 -0.22544387  0.470782113  0.081224798
15 -0.33864995 -0.36028835 -0.06583759  0.421816693  0.023720516
16 -0.14335322 -0.36028835  0.20777317  0.372851273 -0.074858253
           X1X3         X2X3       X1_sq       X2_sq        X3_sq
1  -0.0647473801  0.0520900627 0.078525641 0.050824937 5.338668e-02
2  -0.0647473801  0.0362852397 0.078525641 0.024661931 5.338668e-02
3  -0.0658908217  0.0154807893 0.078525641 0.004334588 5.528895e-02
4  -0.0624604969 -0.0107361442 0.078525641 0.002320050 4.968204e-02
5  -0.0613170553 -0.0454637249 0.078525641 0.043169691 4.787966e-02
6  -0.0647473801 -0.1112264424 0.078525641 0.231730492 5.338668e-02
7   0.0001020930  0.0008307999 0.001602564 0.106124668 6.503940e-06
8   0.0007554882  0.0042545967 0.001602564 0.050824937 3.561557e-04
9   0.0027156738  0.0044662581 0.001602564 0.004334588 4.601928e-03
10  0.0046758594 -0.0056260344 0.001602564 0.002320050 1.364292e-02
11  0.0020622786 -0.0107035809 0.001602564 0.043169691 2.653868e-03
12 -0.0002246046  0.0027008628 0.001602564 0.231730492 3.147907e-05
13 -0.1284534116 -0.1161458245 0.129807692 0.106124668 1.271133e-01
14 -0.1696173089 -0.1061349402 0.129807692 0.050824937 2.216358e-01
15 -0.1519756387 -0.0277713943 0.129807692 0.004334588 1.779293e-01
16 -0.1343339684  0.0774684917 0.129807692 0.043169691 1.390181e-01

> cat("Correlation Matrix: ")
Correlation Matrix:
> corr
                Y         X1          X2          X3        X1X2
Y      1.00000000  0.9450377  0.37003503 -0.91397769 -0.36615167
X1     0.94503771  1.0000000  0.22362776 -0.95820405 -0.13241739
X2     0.37003503  0.2236278  1.00000000 -0.24023098  0.03868762
X3    -0.91397769 -0.9582041 -0.24023098  1.00000000  0.19498531
X1X2  -0.36615167 -0.1324174  0.03868762  0.19498531  1.00000000
X1X3   0.42055075  0.4428236  0.19226267 -0.66052652 -0.26485039
X2X3   0.41877316  0.2055387 -0.02306559 -0.27411884 -0.97448134
X1_sq -0.24850590 -0.2707456 -0.14771083  0.50096224  0.24631222
X2_sq -0.03803937  0.0309599  0.49754636 -0.01751058  0.39789760
X3_sq -0.55517440 -0.5767868 -0.22390584  0.76515318  0.27476802
           X1X3        X2X3      X1_sq       X2_sq      X3_sq
Y      0.4205507  0.41877316 -0.2485059 -0.03803937 -0.5551744
X1     0.4428236  0.20553866 -0.2707456  0.03095990 -0.5767868
X2     0.1922627 -0.02306559 -0.1477108  0.49754636 -0.2239058
X3    -0.6605265 -0.27411884  0.5009622 -0.01751058  0.7651532
```

```
X1X2  -0.2648504 -0.97448134  0.2463122  0.39789760  0.2747680
X1X3   1.0000000  0.32351596 -0.9722442  0.12583104 -0.9721670
X2X3   0.3235160  1.00000000 -0.2792725 -0.37460454 -0.3585293
X1_sq -0.9722442 -0.27927248  1.0000000 -0.12359068  0.8936644
X2_sq  0.1258310 -0.37460454 -0.1235907  1.00000000 -0.1579789
X3_sq -0.9721670 -0.35852931  0.8936644 -0.15797895  1.0000000

> cat("VIF Values: ", vif_values)
VIF Values:  375.2478 1.740631 680.28 31.03706 6563.345 35.61129 1762.575 3.164318 1156.766
```