STAT 43000/STAT 53001 Applied Statistics Spring 2023

Lab 4

Q.N. 1) A hospital administrator wished to study the relationship between patients satisfaction (y) and patient's age (x_1, in years), severity of illness (x_2, an index) and anxiety level (x_3, an index). 46 patients are randomly selected and data is collected. The data is as below

ID	y	x_1	x_2	x_3
1	48	50	51	2.3
2	70	41	44	1.8
3	46	42	50	2.2
4	77	29	50	2.1
5	47	38	55	2.2
6	66	36	49	2.0
7	60	33	49	2.1
8	52	44	58	2.9
9	43	47	53	2.5
10	72	32	46	2.6
11	59	33	42	2.0
12	47	40	48	2.2
13	82	29	48	2.5
14	42	47	50	2.6
15	37	44	51	2.6
16	92	28	46	1.8
17	57	36	46	2.3
18	89	28	43	1.8
19	54	45	48	2.4
20	89	29	48	2.4
21	51	34	51	2.3
22	79	33	56	2.5
23	49	55	51	2.4
24	60	43	50	2.3
25	34	55	54	2.5
26	57	32	52	2.4
27	83	36	49	1.8
28	36	53	57	2.8
29	64	30	51	2.4
30	66	43	53	2.3
31	68	45	51	2.2
32	66	40	48	2.2
33	36	49	54	2.9
34	26	52	62	2.9
35	67	43	53	2.4
36	57	53	54	2.2
37	88	29	46	1.9
38	77	29	52	2.3
39	86	23	41	1.8
40	63	25	49	2.0
41	55	42	51	2.7
42	76	31	47	2.0
43	80	34	49	2.2
44	37	47	60	2.4
45	83	22	51	2.0
46	59	37	53	2.1

a. Fit a multiple linear regression model to the data and state the estimated regression line. How is b_2 interpreted here?

```
> model=lm(y~x1+x2+x3)
> model

Call:
lm(formula = y ~ x1 + x2 + x3)

Coefficients:
(Intercept)           x1           x2           x3
    158.491       -1.142       -0.442      -13.470
```

Y=-1.142x1-0.442x2-13.47x3+158.491

B2 is the coefficient of x2, b2=-0.442 here, which means b2 represents the change in the response variable (y) associated with a one-unit increase in the value of the predictor variable (x2)

b. Calculate the coefficient of multiple determination. What does it indicate?

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
x1           -1.1416     0.2148  -5.315 3.81e-06 ***
x2           -0.4420     0.4920  -0.898   0.3741
x3          -13.4702     7.0997  -1.897   0.0647 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.06 on 42 degrees of freedom
Multiple R-squared:  0.6822,    Adjusted R-squared:  0.6595
F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

a multiple R^2 value of 0.6822 means that the predictor variables in the model explain 68.22% of the variance in the response variable.

c. Predict the patient satisfaction score of a 47 years old patient whose severity of illness index is 53 and anxiety level index is 2.5?

```
> predict(model,data.frame(x1=47,x2=53,x3=2.5))
       1
47.73386
```

d. Please provide the 95% confidence interval and prediction interval for the value calculated in part (c).

```
> predict(model,data.frame(x1=47,x2=53,x3=2.5),interval="pred")
       fit      lwr      upr
1 47.73386 27.00058 68.46714
> predict(model,data.frame(x1=47,x2=53,x3=2.5),interval="conf")
       fit      lwr      upr
1 47.73386 43.50688 51.96085
```

Q.N. 2) The dataset *mtcars* in R was extracted from the 1974 Motor Trend US magazine and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

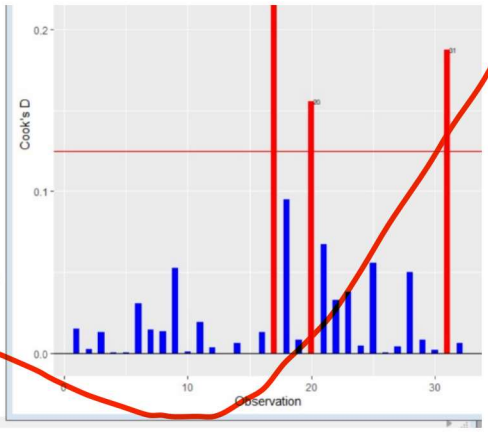
a. Import the data in R and extract the variables included in the dataset.

```
> data(mtcars)
> head(mtcars,4)
                mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710     22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
> names(mtcars)
 [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear" "carb"
>
```

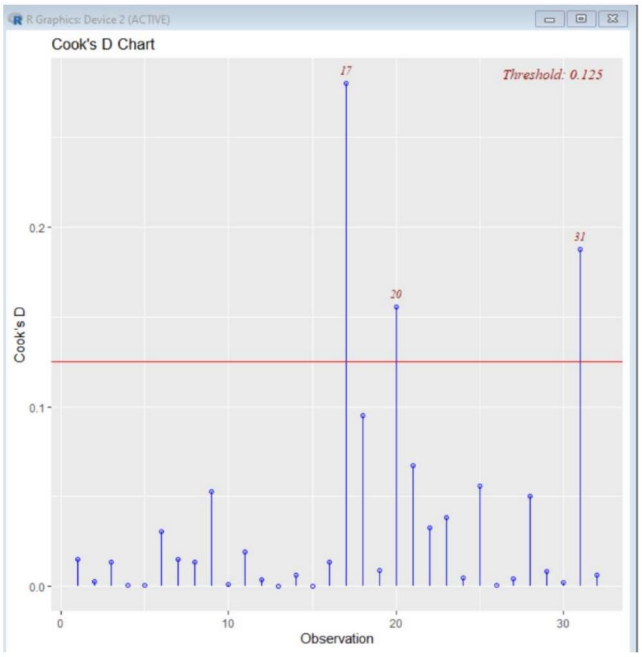
b. Install the `olsrr` package in R and use R code below to draw the Cook's D bar plot

```
model <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)

ols_plot_cooksd_bar(model)
```

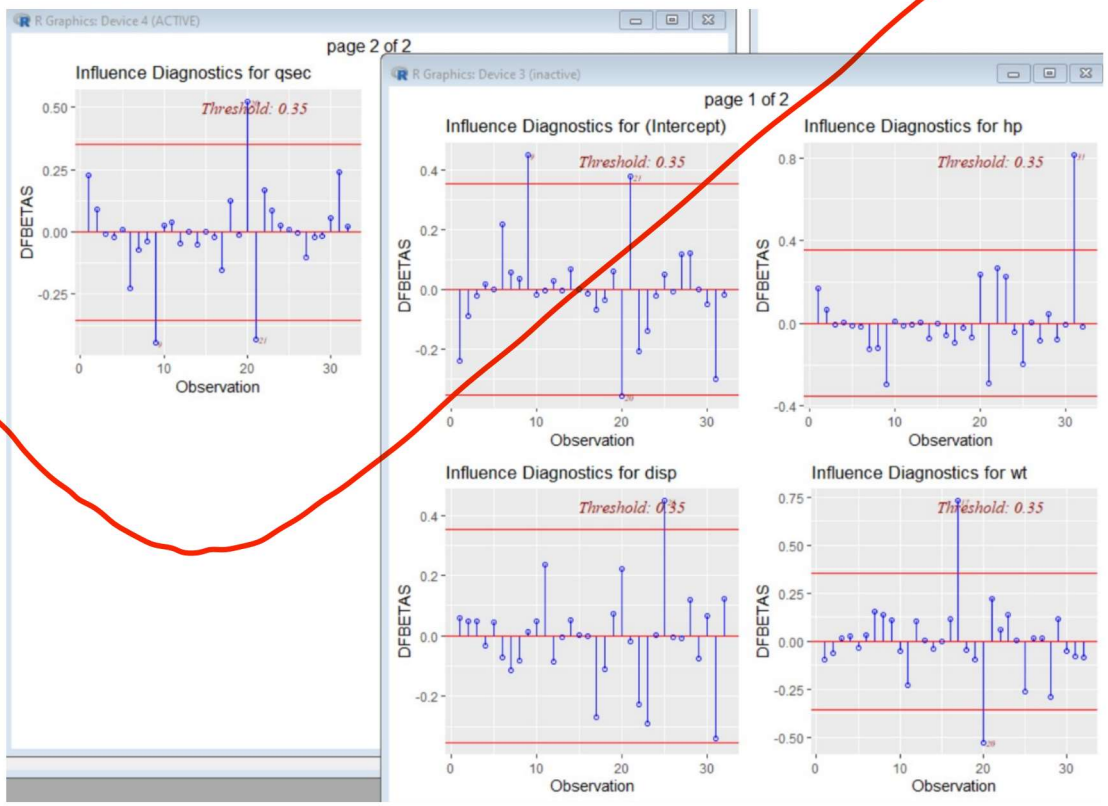
```
package 'data.table' successfully unpacked and MD5 sums checked
package 'ggplot2' successfully unpacked and MD5 sums checked
package 'goftest' successfully unpacked and MD5 sums checked
package 'gridExtra' successfully unpacked and MD5 sums checked
package 'nortest' successfully unpacked and MD5 sums checked
package 'Rcpp' successfully unpacked and MD5 sums checked
package 'olsrr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Zhang\AppData\Local\Temp\Rtmpw9vsvg\downloaded_packages
> model <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)
>
> ols_plot_cooksd_bar(model)
Error in ols_plot_cooksd_bar(model) :
  could not find function "ols_plot_cooksd_bar"
> model <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)
> ols_plot_cooksd_bar(model)
Error in ols_plot_cooksd_bar(model) :
  could not find function "ols_plot_cooksd_bar"
> library(olsrr)
Need help getting started with regression models? Visit: https://www.rsquaredacademy.com

Attaching package: 'olsrr'

The following object is masked from 'package:datasets':

    rivers

> ols_plot_cooksd_bar(model)
>
```



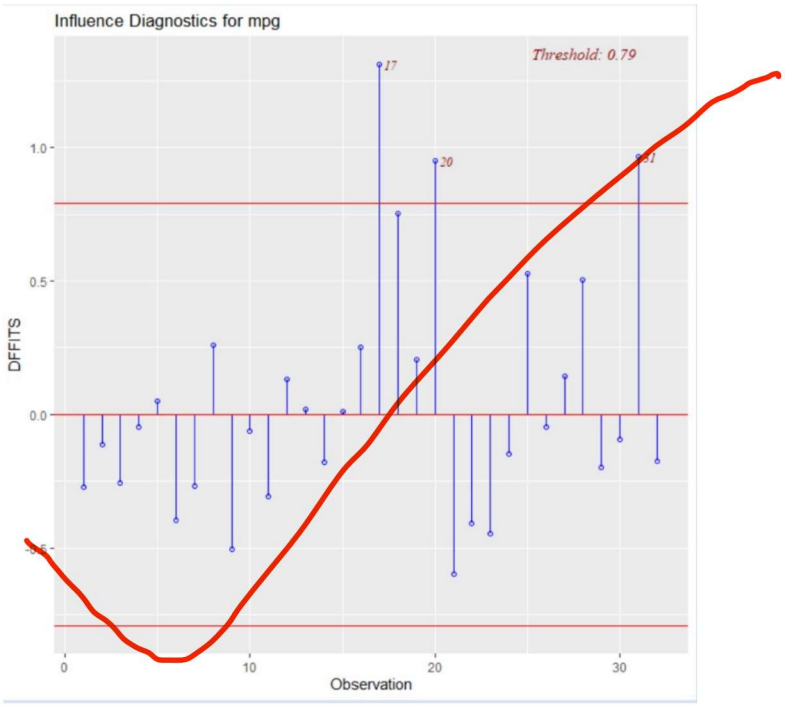
c. Use `ols_plot_cooksd_chart(model)` to display the Cook's D chart



d. Identify the influence diagnostics for each variable using *ols_plot_dfbetas(model)*

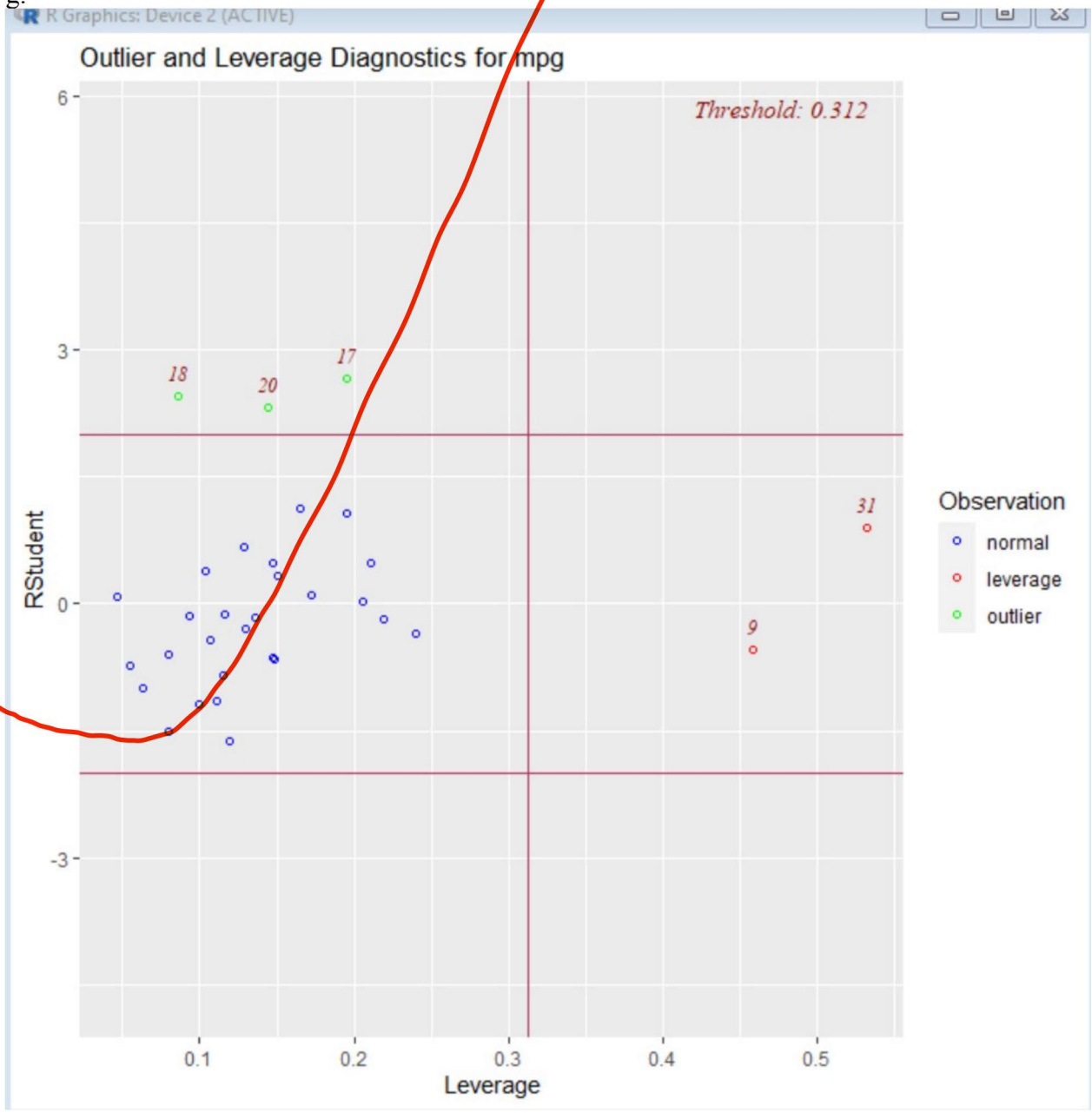


e. Use `ols_plot_dffits(model)` to check the difference in fits.



f. Use `ols_plot_resid_lev(model)` to detect and display the influential observations.

g.


R Graphics: Device 2 (ACTIVE)
Outlier and Leverage Diagnostics for mpg
Threshold: 0.312
RStudent
Leverage
6
3
0
-3
0.1
0.2
0.3
0.4
0.5
18
20
17
31
9
Observation
normal
leverage
outlier