Statistical Linear Model Assignment 5

12111603 谭致恒

2023-12-07

Question 1

(a)

We derive the model as matrix form as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_{p-1}(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_{p-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & \cdots & P_{p-1}(x_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

and denote it as

$$Y = Pa + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I)$.

Now we perform the least square on it and we can yield that

$$\hat{a} = (P^T P)^{-1} P^T Y.$$

Since $\sum\limits_{i=1}^{n} P_l(x_i) P_m(x_i) = 0$ for all l and m, and by the definition of matrix multiplication, we can obtain

$$P^T P = \begin{bmatrix} \sum\limits_{i=1}^{n} (P_0(x_i))^2 & & & \\ & \sum\limits_{i=1}^{n} (P_1(x_i))^2 & & \\ & & \ddots & \\ & & & \sum\limits_{i=1}^{n} (P_{p-1}(x_i))^2 \end{bmatrix}.$$

Notice that we reach a consensus: $n >> p$, which is a usual case in the practical problem we focus on in this course, then $\sum\limits_{i=1}^{n}(P_j(x_i))^2 \neq 0$ since the polynomial P_j has j real roots at most for $j = 0, 1, 2, \cdots, p-1$. Consequently, the matrix P^TP is invertible with its inverse as follows:

$$(P^TP)^{-1} = \begin{bmatrix} \frac{1}{\sum\limits_{i=1}^{n}(P_0(x_i))^2} & & & \\ & \frac{1}{\sum\limits_{i=1}^{n}(P_1(x_i))^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sum\limits_{i=1}^{n}(P_{p-1}(x_i))^2} \end{bmatrix}.$$

Additionally, we can derive P^TY by matrix multiplication:

$$P^TY = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i P_0(x_i) \\ \sum\limits_{i=1}^{n} y_i P_1(x_i) \\ \vdots \\ \sum\limits_{i=1}^{n} y_i P_{p-1}(x_i) \end{bmatrix}.$$

Now we obtain the least square estimator of a is

$$\hat{a} = (P^TP)^{-1}P^TY = \begin{bmatrix} \frac{\sum\limits_{i=1}^{n} y_i P_0(x_i)}{\sum\limits_{i=1}^{n}(P_0(x_i))^2} \\ \vdots \\ \frac{\sum\limits_{i=1}^{n} y_i P_{p-1}(x_i)}{\sum\limits_{i=1}^{n}(P_{p-1}(x_i))^2} \end{bmatrix}.$$

Therefore, the least square estimator of a_j is

$$\hat{a}_j = \frac{\sum\limits_{i=1}^{n} y_i P_j(x_i)}{\sum\limits_{i=1}^{n} (P_j(x_i))^2}.$$

To verify the uncorrelation between $\hat{a}_j$'s, we compute the covariance matrix of $\hat{a}$:

$$Cov(\hat{a}) = Cov((P^TP)^{-1}P^TY) = (P^TP)^{-1}P^TCov(Y)P(P^TP)^{-1} = \sigma^2(P^TP)^{-1}$$

$$= \begin{bmatrix} \frac{\sigma^2}{\sum\limits_{i=1}^{n}(P_0(x_i))^2} & & & \\ & \frac{\sigma^2}{\sum\limits_{i=1}^{n}(P_1(x_i))^2} & & \\ & & \ddots & \\ & & & \frac{\sigma^2}{\sum\limits_{i=1}^{n}(P_{p-1}(x_i))^2} \end{bmatrix}.$$

Since $Cov(\hat{a})$ is a diagonal matrix, then we can maintain that $Cov(\hat{a}_i, \hat{a}_j) = 0$ for $i \neq j$, implying $\hat{a}_j$'s are uncorrelated.

(b)

Since $\mathbb{E}\hat{a} = \mathbb{E}[(P^TP)^{-1}P^TY] = (P^TP)^{-1}P^T\mathbb{E}Y = (P^TP)^{-1}P^TPa = a$, $Cov(\hat{a}) = \sigma^2(P^TP)^{-1}$ and $\hat{a}$ is a linear transformation of Y where Y obeys a Gaussian distribution, then $\hat{a}$ also obeys a Gaussian distribution:

$$\hat{a} \sim \mathcal{N}_p(a, \sigma^2(P^TP)^{-1}).$$

Thus, $\hat{a}_j \sim \mathcal{N}_p(a_j, \frac{\sigma^2}{\sum\limits_{i=1}^{n}(P_j(x_i))^2})$.

Moreover,

$$SSE = (Y - P\hat{a})^T(Y - P\hat{a}) = Y^T(I - H_P)Y,$$

where $H_P = P(P^TP)^{-1}P^T$ is a projection matrix.

Since $I - H_P$ is idempotent with $rank(I - H_P) = n - p$, then we have

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-p).$$

Consequently,

$$\frac{(\hat{a}_j - a_j)\sqrt{\sum_{i=1}^{n}(P_j(x_i))^2}}{\hat{\sigma}} \sim t_{n-p}$$

where $\hat{\sigma}^2 = \frac{SSE}{n-p}$.

Therefore, we can reject $H_0 : a_j = 0$ at α level of confidence if

$$\left|\frac{\hat{a}_j\sqrt{\sum_{i=1}^{n}(P_j(x_i))^2}}{\hat{\sigma}}\right| > t_{(\alpha/2, n-p)}.$$

(c)

Denote

$$[\,P_0(x^*) \quad P_1(x^*) \quad \cdots \quad P_{p-1}(x^*)\,] \triangleq p^*,$$

then $y^* = (p^*)^T a + \varepsilon^*$, $\mathbb{E}(y^*) = (p^*)^T a$ and $\widehat{\mathbb{E}(y^*)} = (p^*)^T \hat{a}$.

Thus we have

$$\widehat{\mathbb{E}(y^*)} - \mathbb{E}(y^*) \sim \mathcal{N}(0, \sigma^2 (p^*)^T (P^T P)^{-1} p^*).$$

Hence we yield that

$$\frac{\widehat{\mathbb{E}(y^*)} - \mathbb{E}(y^*)}{\hat{\sigma}\sqrt{(p^*)^T (P^T P)^{-1} p^*}} \sim t_{n-p}$$

where $\hat{\sigma}^2 = \frac{SSE}{n-p}$.

Therefore, $100(1-\alpha)\%$ confidence interval is

$$[\widehat{\mathbb{E}(y^*)} - t_{(\alpha/2,n-p)}\sqrt{(p^*)^T (P^T P)^{-1} p^*}\hat{\sigma},\ \widehat{\mathbb{E}(y^*)} + t_{(\alpha/2,n-p)}\sqrt{(p^*)^T (P^T P)^{-1} p^*}\hat{\sigma}]$$

i.e.,

$$[(p^*)^T\hat{a} - t_{(\alpha/2,n-p)}\sqrt{(p^*)^T(P^TP)^{-1}p^*}\hat{\sigma},\ (p^*)^T\hat{a} + t_{(\alpha/2,n-p)}\sqrt{(p^*)^T(P^TP)^{-1}p^*}\hat{\sigma}].$$

Question 2

(a)

```
data <- read.csv("C:/Users/Lenovo/Desktop/6data(1).csv")
fit <- lm(Y ~ X1 + X2 + X3 + X4, data = data)
summary(fit)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.17355 -0.55425 -0.00316  0.61569  2.02727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.22211    0.71119  17.185  < 2e-16 ***
## X1          -0.18698    0.02497  -7.489 9.04e-09 ***
## X2           0.29510    0.07349   4.016 0.000298 ***
## X3          -1.21196    1.40668  -0.862 0.394786
## X4           0.07479    0.01637   4.569 5.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9353 on 35 degrees of freedom
## Multiple R-squared:  0.7541, Adjusted R-squared:  0.726
## F-statistic: 26.84 on 4 and 35 DF,  p-value: 3.088e-10
```

We establish a hypothesis test for the utility of the model
$H_0: \beta_1 = \beta_1 = \beta_3 = \beta_4 = 0$, and H_1 : at least one of them is not 0.

Since the p-value of F-statistic is 3.088e-10, which is extremely small. Thus, we can say that the model is significantly effective.

(b)

```
r1 <- rstandard(fit)
r2 <- rstudent(fit)
leverage <- hatvalues(fit)
dffits <- dffits(fit)
cooks_distance <- cooks.distance(fit)

statistics <- data.frame(r1,r2,leverage,dffits,cooks_distance)
names(statistics) <- c("Studentized residuals","Studentized deleted residuals",
                       "Leverage values", "Dffits","Cook's distance")
statistics
```

```
##    Studentized residuals Studentized deleted residuals Leverage values
## 1           -0.881313929                  -0.878434210      0.13469748
## 2           -0.547027035                  -0.541475420      0.14719530
## 3            0.114526120                   0.112899333      0.36914451
## 4            0.001844146                   0.001817610      0.08380724
## 5            0.908126276                   0.905794106      0.05964149
## 6           -2.520284183                  -2.745620746      0.14979177
## 7           -0.239868320                  -0.236611361      0.10028484
## 8            2.137663634                   2.259566038      0.52194545
## 9            2.429263793                   2.625895934      0.20392176
## 10           0.503208583                   0.497771708      0.05951632
## 11           0.730853920                   0.725897877      0.16180864
## 12          -0.303250984                  -0.299280866      0.07711837
## 13           0.775534861                   0.771029052      0.07742031
## 14           0.271122006                   0.267501818      0.15000359
## 15          -0.131615597                  -0.129753863      0.10232066
## 16          -0.941069237                  -0.939490166      0.14157544
## 17           0.400112460                   0.395260142      0.06830628
## 18          -1.151851331                  -1.157426559      0.14920288
## 19          -0.586051309                  -0.580473602      0.18019839
## 20          -0.608227275                  -0.602668830      0.05929007
## 21           0.067067672                   0.066106867      0.09300209
## 22           0.338608468                   0.334284136      0.09243832
## 23          -0.008970208                  -0.008841144      0.09775913
## 24           0.270208563                   0.266598685      0.08846972
## 25          -0.431894646                  -0.426818898      0.12784686
## 26          -1.776371827                  -1.835507218      0.17601682
## 27           1.247561875                   1.257897088      0.08369103
## 28           0.174568468                   0.172131513      0.08252251
## 29          -0.733720368                  -0.728789274      0.07953229
## 30           0.730635239                   0.725677315      0.10296404
## 31          -1.249953962                  -1.260421576      0.08660188
## 32          -0.108416507                  -0.106874423      0.07103079
## 33          -1.043192193                  -1.044548682      0.07718264
## 34           0.688281672                   0.683015946      0.09531188
## 35           1.002134178                   1.002197155      0.08919865
## 36           1.600626622                   1.638711484      0.05806995
## 37           1.721521774                   1.773496669      0.13203018
## 38          -1.693003760                  -1.741473039      0.17171092
## 39          -0.109289486                  -0.107735278      0.07536270
## 40          -0.640939528                  -0.635457160      0.12206680
##           Dffits Cook's distance
## 1  -0.3465811939    2.418147e-02
## 2  -0.2249577359    1.032980e-02
## 3   0.0863623876    1.534990e-03
## 4   0.0005497282    6.221786e-08
## 5   0.2281166588    1.046110e-02
## 6  -1.1524494177    2.238163e-01
## 7  -0.0789952047    1.282644e-03
## 8   2.3610156478    9.978296e-01
## 9   1.3290197065    3.023341e-01
## 10  0.1252196844    3.204873e-03
## 11  0.3189369120    2.062290e-02
## 12 -0.0865136968    1.536902e-03
## 13  0.2233553000    1.009447e-02
## 14  0.1123748224    2.594443e-03
## 15 -0.0438067804    3.948997e-04
## 16 -0.3815356611    2.921184e-02
## 17  0.1070229134    2.347370e-03
## 18 -0.4846955654    4.653439e-02
## 19 -0.2721469992    1.509883e-02
## 20 -0.1513010173    4.663243e-03
## 21  0.0211684851    9.224501e-05
## 22  0.1066850849    2.335616e-03
## 23 -0.0029102204    1.743692e-06
## 24  0.0830557620    1.417267e-03
## 25 -0.1634151456    5.468686e-03
```

```
## 26 -0.8483479432     1.348136e-01
## 27  0.3801575295     2.843094e-02
## 28  0.0516236723     5.481995e-04
## 29 -0.2142246442     9.303065e-03
## 30  0.2458563427     1.225482e-02
## 31 -0.3881051666     2.962683e-02
## 32 -0.0295526507     1.797489e-04
## 33 -0.3020860161     1.820382e-02
## 34  0.2216944949     9.981838e-03
## 35  0.3136320696     1.967054e-02
## 36  0.4068823945     3.158951e-02
## 37  0.6916950670     9.016202e-02
## 38 -0.7929115605     1.188398e-01
## 39 -0.0307574597     1.947026e-04
## 40 -0.2369487002     1.142353e-02
```

(c)

```
absstudentized <- abs(r1)

which(absstudentized > 2)
```

```
## 6 8 9
## 6 8 9
```

Thus, the 6^{th}, 8^{th} and 9^{th} observations are outlying Y observations.

(d)

```
which(leverage > 2 *(4+1)/40)
```

```
## 3 8
## 3 8
```

We consider leverage values. If the leverage value of the points is larger than $\frac{2(p+1)}{n}$, then we say it is a high-leverage points, i.e., an outlier of X.

From this criterion we can say that the 3^{rd} and 8^{th} observations are outlying X observations.

(e)

```
absdffits <- abs(dffits)
which(absdffits > 2*sqrt((4+1)/40))
```

```
##  6  8  9 26 38
##  6  8  9 26 38
```

We maintain that an observation is a high influential point if its dffits is larger than $2\sqrt{\frac{p+1}{n}}$. According to this criterion, then we can say the 6^{th}, 8^{th}, 9^{th}, 26^{th} and 38^{th} observations are high influential points.

```
which(cooks_distance > qf(0.5, (4+1), (40-4-1)))
```

```
## 8
## 8
```

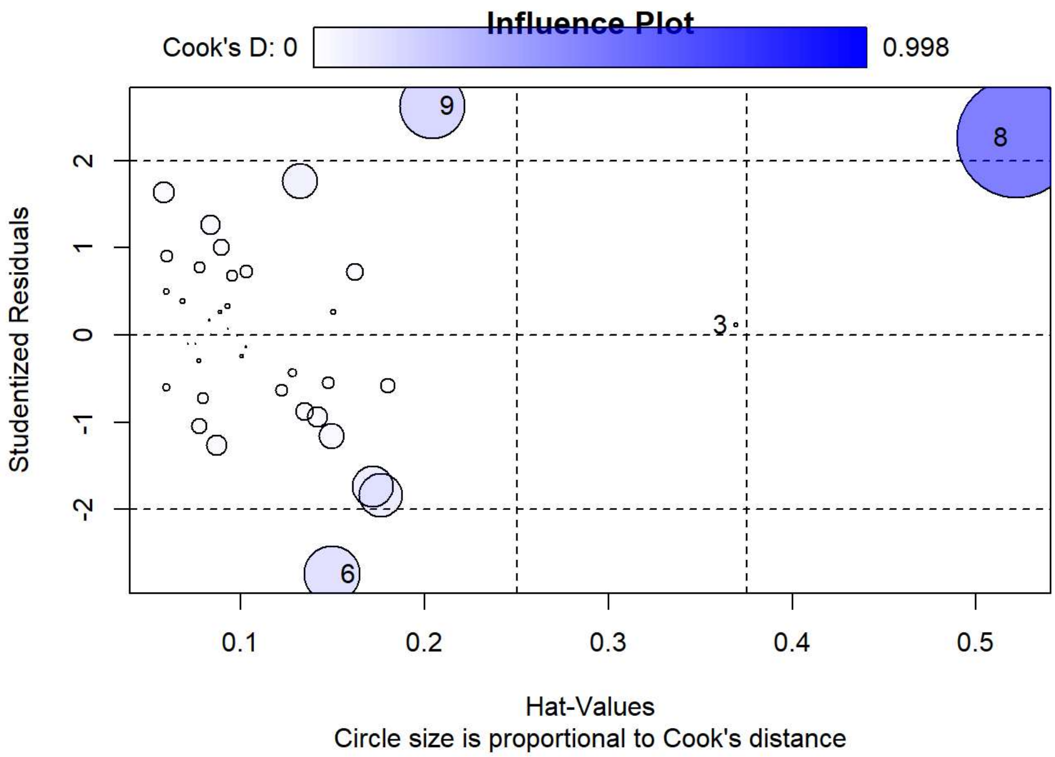
We can also maintain that an observation is a high influential point if its Cook's distance is larger than $F_{(0.5,p+1,n-p-1)}$. According to this criterion, then we can say the 8^{th} observation is the high influential point.

(Appendix) Influential plot

```
library(car)
```

```
## 载入需要的程辑包：carData
```

```
influencePlot(fit, id.method="identify", main="Influence Plot",
              sub="Circle size is proportional to Cook's distance")
```




```
##       StudRes       Hat       CookD
## 3   0.1128993 0.3691445 0.00153499
## 6  -2.7456207 0.1497918 0.22381628
## 8   2.2595660 0.5219454 0.99782961
## 9   2.6258959 0.2039218 0.30233413
```