ACA228 - Modelos de Regressão e Previsão

Regressão Linear Multipla: Medidas de Influência

Prof. Carlos Trucíos
carlos.trucios@facc.ufrj.br
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis,
Universidade Federal do Rio de Janeiro

Aula 15

Universidade Federal do Rio de Janeiro

Introdução

Leverage

Distância de Cook

DFFITS

DFBETAS

COVRATIO

Exemplo

Introdução

Introdução

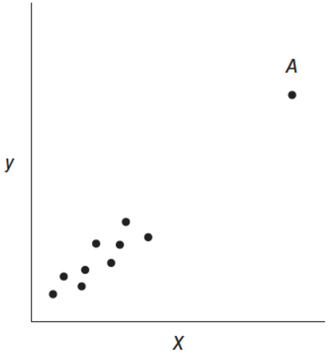

Figure 6.1 An example of a leverage point.

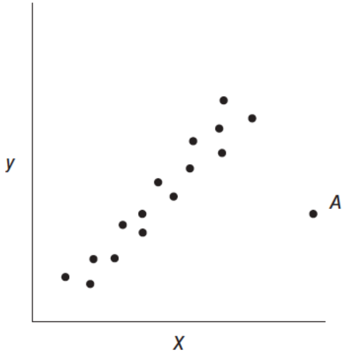

Figure 6.2 An example of an influential observation.

Introdução

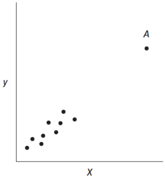

Figure 6.1 An example of a leverage point.

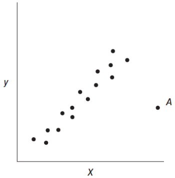

Figure 6.2 An example of an influential observation.

Leverage

- O ponto A não afetará os $\hat{\beta}$'s mas afetara o R^2 e $\widehat{V}(\hat{\beta}|X)$.

Influência

- O ponto A terá um forte impacto nos $\hat{\beta}$'s

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Além disso:

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Além disso:

- Algumas vezes podemos encontrar um coeficiente que tem um sinal ($\pm$) que não faz sentido.

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Além disso:

- Algumas vezes podemos encontrar um coeficiente que tem um sinal ($\pm$) que não faz sentido.
- Uma variavel que sabemos que é importante no modelo aparece como estatísticamente não significativa.

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Além disso:

- Algumas vezes podemos encontrar um coeficiente que tem um sinal ($\pm$) que não faz sentido.
- Uma variavel que sabemos que é importante no modelo aparece como estatísticamente não significativa.
- Um modelo que se ajusta bem aos dados e que faz total sentido no problema de negócio, apresenta predições *pobres*

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Além disso:

- Algumas vezes podemos encontrar um coeficiente que tem um sinal ($\pm$) que não faz sentido.
- Uma variavel que sabemos que é importante no modelo aparece como estatísticamente não significativa.
- Um modelo que se ajusta bem aos dados e que faz total sentido no problema de negócio, apresenta predições *pobres*

Introdução

Para evitar essas consequências nos $\hat{\beta}$'s, $\widehat{V}(\hat{\beta}|X)$ e R^2, é importante detetar esses pontos e analisar sua possível eliminação.

Além disso:

- Algumas vezes podemos encontrar um coeficiente que tem um sinal ($\pm$) que não faz sentido.
- Uma variavel que sabemos que é importante no modelo aparece como estatísticamente não significativa.
- Um modelo que se ajusta bem aos dados e que faz total sentido no problema de negócio, apresenta predições *pobres*

Todas essas situações podem ser uma consequência da presença de observações influentes/leverage.

Leverage

Leverage

A matriz H (leia-se matriz *hat*) tem um papel importante para identificar pontos leverage.

$$H = X(X'X)^{-1}X'$$

Leverage

A matriz H (leia-se matriz *hat*) tem um papel importante para identificar pontos leverage.

$$H = X(X'X)^{-1}X'$$

- vamos utilizar os elementos da diagonal da matriz H, $h_{ii} = x_i(X'X)^{-1}x_i'$

Leverage

A matriz H (leia-se matriz *hat*) tem um papel importante para identificar pontos leverage.

$$H = X(X'X)^{-1}X'$$

- vamos utilizar os elementos da diagonal da matriz H, $h_{ii} = x_i(X'X)^{-1}x_i'$
- é uma medida padronizada da distância da i-ésima observação ao centro (do espaço x)

Leverage

A matriz H (leia-se matriz *hat*) tem um papel importante para identificar pontos leverage.

$$H = X(X'X)^{-1}X'$$

- vamos utilizar os elementos da diagonal da matriz H, $h_{ii} = x_i(X'X)^{-1}x_i'$
- é uma medida padronizada da distância da i-ésima observação ao centro (do espaço x)
- Valores grandes indicam potenciais pontos de leverage.

Leverage

A matriz H (leia-se matriz *hat*) tem um papel importante para identificar pontos leverage.

$$H = X(X'X)^{-1}X'$$

- vamos utilizar os elementos da diagonal da matriz H, $h_{ii} = x_i(X'X)^{-1}x_i'$
- é uma medida padronizada da distância da i-ésima observação ao centro (do espaço x)
- Valores grandes indicam potenciais pontos de leverage.
- Se $h_{ii} > 2(k+1)/n$ a i-ésima observação é um possível ponto de leverage.

Leverage

A matriz H (leia-se matriz *hat*) tem um papel importante para identificar pontos leverage.

$$H = X(X'X)^{-1}X'$$

- vamos utilizar os elementos da diagonal da matriz H, $h_{ii} = x_i(X'X)^{-1}x_i'$
- é uma medida padronizada da distância da i-ésima observação ao centro (do espaço x)
- Valores grandes indicam potenciais pontos de leverage.
- Se $h_{ii} > 2(k+1)/n$ a i-ésima observação é um possível ponto de leverage.
- No R, utilizamos a função `hatvalues()`

Leverage

O *dataset* `delivery` disponivel no pacote `robustbase`, contém 25 observações referentes a maquinas automáticas de vendas de bebidas. Estamos interessados em predizer o tempo necessário para o motorista da rota fazer a manutenção dessas maquinas (o que inclui repor o estoque de bebidas na máquina e fazer pequenas manutenções). O modelo a ser utilizado é da forma:

$$delTime = \beta_0 + \beta_1 \text{ n.prod} + \beta_2 \text{ distance} + u,$$

em que:

- *delTime*: é o tempo utilizado (em minutos) pelo motorista da rota para fazer a manutenção da máquina,
- *n.prod*: é o número de caixas do produto estocado e
- *distance*: é a distância (em pés) percorrida pelo motorista da rota.

Leverage

```
library(robustbase)
head(delivery,4)
```

```
##   n.prod distance delTime
## 1      7      560   16.68
## 2      3      220   11.50
## 3      3      340   12.03
## 4      4       80   14.88
```

```
modelo <- lm(delTime ~ n.prod + distance, data = delivery)
2*(2+1)/nrow(delivery)  # Ponto de corte hii
```

```
## [1] 0.24
```

```
hatvalues(modelo)  # Valores hii
```

Leverage

	n.prod	distance	delTime	hii
1	7	560	16.68	0.1018
2	3	220	11.50	0.0707
3	3	340	12.03	0.0987
4	4	80	14.88	0.0854
5	6	150	13.75	0.0750
6	7	330	18.11	0.0429
7	2	110	8.00	0.0818
8	7	210	17.83	0.0637
9	30	1460	79.24	0.4983
10	5	605	21.50	0.1963
11	16	688	40.33	0.0861
12	10	215	21.00	0.1137
13	4	255	13.50	0.0611

Leverage

	n.prod	distance	delTime	hii
14	6	462	19.75	0.0782
15	9	448	24.00	0.0411
16	10	776	29.00	0.1659
17	6	200	15.35	0.0594
18	7	132	19.00	0.0963
19	3	36	9.50	0.0964
20	17	770	35.10	0.1017
21	10	140	17.90	0.1653
22	26	810	52.32	0.3916
23	9	450	18.75	0.0413
24	8	635	19.83	0.1206
25	4	150	10.75	0.0666

Distância de Cook

Distância de Cook

Mede como a i-ésima observação influencia na estimação dos β's.

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{(k+1)\hat{\sigma}^2} \equiv \frac{(\hat{y}_{(i)} - \hat{y})' (\hat{y}_{(i)} - \hat{y})}{(k+1)\hat{\sigma}^2} \equiv \frac{r_i^2}{k+1} \frac{h_{ii}}{1 - h_{ii}},$$

em que (i) significa *sem a i-ésima observação*

Distância de Cook

Mede como a i-ésima observação influencia na estimação dos β's.

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{(k+1)\hat{\sigma}^2} \equiv \frac{(\hat{y}_{(i)} - \hat{y})' (\hat{y}_{(i)} - \hat{y})}{(k+1)\hat{\sigma}^2} \equiv \frac{r_i^2}{k+1} \frac{h_{ii}}{1-h_{ii}},$$

em que (i) significa *sem a i-ésima observação*

- Valores grandes de D_i indicam que a observação pode ser influente.

Distância de Cook

Mede como a *i*-ésima observação influencia na estimação dos β's.

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'X'X(\hat{\beta}_{(i)} - \hat{\beta})}{(k+1)\hat{\sigma}^2} \equiv \frac{(\hat{y}_{(i)} - \hat{y})'(\hat{y}_{(i)} - \hat{y})}{(k+1)\hat{\sigma}^2} \equiv \frac{r_i^2}{k+1}\frac{h_{ii}}{1-h_{ii}},$$

em que (i) significa *sem a i-ésima observação*

- Valores grandes de D_i indicam que a observação pode ser influente.
- Geralmente, se $D_i > F_{0.5,k+1,n-(k+1)}$ é influente.

Distância de Cook

Mede como a i-ésima observação influencia na estimação dos β's.

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{(k+1)\hat{\sigma}^2} \equiv \frac{(\hat{y}_{(i)} - \hat{y})' (\hat{y}_{(i)} - \hat{y})}{(k+1)\hat{\sigma}^2} \equiv \frac{r_i^2}{k+1} \frac{h_{ii}}{1 - h_{ii}},$$

em que (i) significa *sem a i-ésima observação*

- Valores grandes de D_i indicam que a observação pode ser influente.
- Geralmente, se $D_i > F_{0.5,k+1,n-(k+1)}$ é influente.
- Como $F_{0.5,k+1,n-(k+1)} \approx 1$, alguns livros consideram pontos influentes se $D_i > 1$

Distância de Cook

Mede como a *i*-ésima observação influencia na estimação dos β's.

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{(k+1)\hat{\sigma}^2} \equiv \frac{(\hat{y}_{(i)} - \hat{y})' (\hat{y}_{(i)} - \hat{y})}{(k+1)\hat{\sigma}^2} \equiv \frac{r_i^2}{k+1} \frac{h_{ii}}{1 - h_{ii}},$$

em que (i) significa *sem a i-ésima observação*

- Valores grandes de D_i indicam que a observação pode ser influente.
- Geralmente, se $D_i > F_{0.5,k+1,n-(k+1)}$ é influente.
- Como $F_{0.5,k+1,n-(k+1)} \approx 1$, alguns livros consideram pontos influentes se $D_i > 1$
- No R, utilizamos a função `cooks.distance()`

Distância de Cook

```r
k = 2
n = 25
qf(0.5,k+1,n-(k+1))  # Ponto de corte Di
```

```
## [1] 0.813655
```

```r
cooks.distance(modelo)  # Distância de Cook
```

Distância de Cook

	n.prod	distance	delTime	hii	Di
1	7	560	16.68	0.1018	0.1001
2	3	220	11.50	0.0707	0.0034
3	3	340	12.03	0.0987	0.0000
4	4	80	14.88	0.0854	0.0776
5	6	150	13.75	0.0750	0.0005
6	7	330	18.11	0.0429	0.0001
7	2	110	8.00	0.0818	0.0022
8	7	210	17.83	0.0637	0.0031
9	30	1460	79.24	0.4983	3.4193
10	5	605	21.50	0.1963	0.0538
11	16	688	40.33	0.0861	0.0162
12	10	215	21.00	0.1137	0.0016
13	4	255	13.50	0.0611	0.0023

Distância de Cook

	n.prod	distance	delTime	hii	Di
14	6	462	19.75	0.0782	0.0033
15	9	448	24.00	0.0411	0.0006
16	10	776	29.00	0.1659	0.0033
17	6	200	15.35	0.0594	0.0004
18	7	132	19.00	0.0963	0.0440
19	3	36	9.50	0.0964	0.0119
20	17	770	35.10	0.1017	0.1324
21	10	140	17.90	0.1653	0.0509
22	26	810	52.32	0.3916	0.4510
23	9	450	18.75	0.0413	0.0299
24	8	635	19.83	0.1206	0.1023
25	4	150	10.75	0.0666	0.0001

DFFITS

DFFITS

Mede a influência da i-ésima observação sobre o valor ajustado (fitted value, $\hat{y}_i$).

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i,(i)}}{\sqrt{s^2_{(i)} h_{ii}}} \equiv= \left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} t_i$$

DFFITS

Mede a influência da i-ésima observação sobre o valor ajustado (fitted value, $\hat{y}_i$).

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i,(i)}}{\sqrt{s^2_{(i)} h_{ii}}} \equiv= \Big(\frac{h_{ii}}{1 - h_{ii}}\Big)^{1/2} t_i$$

- Valors grandes de DDFITS são possíveis pontos influentes/leverage

DFFITS

Mede a influência da i-ésima observação sobre o valor ajustado (fitted value, $\hat{y}_i$).

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i,(i)}}{\sqrt{s^2_{(i)} h_{ii}}} \equiv= \Big(\frac{h_{ii}}{1-h_{ii}}\Big)^{1/2} t_i$$

- Valors grandes de DDFITS são possíveis pontos influentes/leverage
- $\big|DFFITS_i\big| > 2\sqrt{(k+1)/n}$

DFFITS

Mede a influência da i-ésima observação sobre o valor ajustado (fitted value, $\hat{y}_i$).

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i,(i)}}{\sqrt{s^2_{(i)} h_{ii}}} \equiv= \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} t_i$$

- Valors grandes de DDFITS são possíveis pontos influentes/leverage
- $|DFFITS_i| > 2\sqrt{(k+1)/n}$
- No R, utilizamos a função `dffits()`

DFFITS

Mede a influência da i-ésima observação sobre o valor ajustado (fitted value, $\hat{y}_i$).

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i,(i)}}{\sqrt{s_{(i)}^2 h_{ii}}} \equiv= \left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} t_i$$

- Valors grandes de DDFITS são possíveis pontos influentes/leverage
- $|DFFITS_i| > 2\sqrt{(k+1)/n}$
- No R, utilizamos a função `dffits()`

DFFITS

Mede a influência da i-ésima observação sobre o valor ajustado (fitted value, $\hat{y}_i$).

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i,(i)}}{\sqrt{s^2_{(i)} h_{ii}}} \equiv= \left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} t_i$$

- Valors grandes de DDFITS são possíveis pontos influentes/leverage
- $\left|DFFITS_i\right| > 2\sqrt{(k+1)/n}$
- No R, utilizamos a função `dffits()`

```
2*sqrt((k+1)/n)    # Ponto de corte DFFITS
```

```
## [1] 0.6928203
```

```
dffits(modelo)
```

DFFITS

	n.prod	distance	delTime	hii	Di	DFFITS
1	7	560	16.68	0.1018	0.1001	-0.5709
2	3	220	11.50	0.0707	0.0034	0.0986
3	3	340	12.03	0.0987	0.0000	-0.0052
4	4	80	14.88	0.0854	0.0776	0.5008
5	6	150	13.75	0.0750	0.0005	-0.0395
6	7	330	18.11	0.0429	0.0001	-0.0188
7	2	110	8.00	0.0818	0.0022	0.0790
8	7	210	17.83	0.0637	0.0031	0.0938
9	30	1460	79.24	0.4983	3.4193	4.2961
10	5	605	21.50	0.1963	0.0538	0.3987
11	16	688	40.33	0.0861	0.0162	0.2180
12	10	215	21.00	0.1137	0.0016	-0.0677
13	4	255	13.50	0.0611	0.0023	0.0813

DFFITS

	n.prod	distance	delTime	hii	Di	DFFITS
14	6	462	19.75	0.0782	0.0033	0.0974
15	9	448	24.00	0.0411	0.0006	0.0426
16	10	776	29.00	0.1659	0.0033	-0.0972
17	6	200	15.35	0.0594	0.0004	0.0339
18	7	132	19.00	0.0963	0.0440	0.3653
19	3	36	9.50	0.0964	0.0119	0.1862
20	17	770	35.10	0.1017	0.1324	-0.6718
21	10	140	17.90	0.1653	0.0509	-0.3885
22	26	810	52.32	0.3916	0.4510	-1.1950
23	9	450	18.75	0.0413	0.0299	-0.3075
24	8	635	19.83	0.1206	0.1023	-0.5711
25	4	150	10.75	0.0666	0.0001	-0.0176

DFBETAS

DFBETAS

Mede a influência da i-ésima observação na estimação dos β_j.

$$DFBETAS_i = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\sqrt{S^2_{(i)} C_{jj}}},$$

em que C_{jj} é o j-ésimo elemento da diagonal da matriz $(X'X)^{-1}$

DFBETAS

Mede a influência da i-ésima observação na estimação dos β_j.

$$DFBETAS_i = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\sqrt{S^2_{(i)} C_{jj}}},$$

em que C_{jj} é o j-ésimo elemento da diagonal da matriz $(X'X)^{-1}$

- Se $|DFBETAS_i| > 2/\sqrt{n}$ o valor é um possível ponto influente.

DFBETAS

Mede a influência da i-ésima observação na estimação dos β_j.

$$DFBETAS_i = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\sqrt{S^2_{(i)} C_{jj}}},$$

em que C_{jj} é o j-ésimo elemento da diagonal da matriz $(X'X)^{-1}$

- Se $|DFBETAS_i| > 2/\sqrt{n}$ o valor é um possível ponto influente.
- No R, utilizamos a função `dfbetas()`

DFBETAS

Mede a influência da i-ésima observação na estimação dos β_j.

$$DFBETAS_i = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\sqrt{S^2_{(i)} C_{jj}}},$$

em que C_{jj} é o j-ésimo elemento da diagonal da matriz $(X'X)^{-1}$

- Se $|DFBETAS_i| > 2/\sqrt{n}$ o valor é um possível ponto influente.
- No R, utilizamos a função `dfbetas()`
- **Cuidado:** existe uma função `dfbeta()` mas é diferente do que desejamos calcular.

DFBETAS

Mede a influência da i-ésima observação na estimação dos β_j.

$$DFBETAS_i = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\sqrt{S^2_{(i)} C_{jj}}},$$

em que C_{jj} é o j-ésimo elemento da diagonal da matriz $(X'X)^{-1}$

- Se $|DFBETAS_i| > 2/\sqrt{n}$ o valor é um possível ponto influente.
- No R, utilizamos a função `dfbetas()`
- **Cuidado:** existe uma função `dfbeta()` mas é diferente do que desejamos calcular.

DFBETAS

Mede a influência da i-ésima observação na estimação dos β_j.

$$DFBETAS_i = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\sqrt{S^2_{(i)} C_{jj}}},$$

em que C_{jj} é o j-ésimo elemento da diagonal da matriz $(X'X)^{-1}$

- Se $|DFBETAS_i| > 2/\sqrt{n}$ o valor é um possível ponto influente.
- No R, utilizamos a função `dfbetas()`
- **Cuidado:** existe uma função `dfbeta()` mas é diferente do que desejamos calcular.

```
2/sqrt(n)  # Ponto de corte DIFBETAS
```

```
## [1] 0.4
```

```
dfbetas(modelo)
```

DFBETAS

	(Intercept)	n.prod	distance
1	-0.1873	0.4113	-0.4349
2	0.0898	-0.0478	0.0144
3	-0.0035	0.0039	-0.0028
4	0.4520	0.0883	-0.2734
5	-0.0317	-0.0133	0.0242
6	-0.0147	0.0018	0.0011
7	0.0781	-0.0223	-0.0110
8	0.0712	0.0334	-0.0538
9	-2.5757	0.9287	1.5076
10	0.1079	-0.3382	0.3413
11	-0.0343	0.0925	-0.0027
12	-0.0303	-0.0487	0.0540
13	0.0724	-0.0356	0.0113

DFBETAS

	(Intercept)	n.prod	distance
14	0.0495	-0.0671	0.0618
15	0.0223	-0.0048	0.0068
16	-0.0027	0.0644	-0.0842
17	0.0289	0.0065	-0.0157
18	0.2486	0.1897	-0.2724
19	0.1726	0.0236	-0.0990
20	0.1680	-0.2150	-0.0929
21	-0.1619	-0.2972	0.3364
22	0.3986	-1.0254	0.5731
23	-0.1599	0.0373	-0.0527
24	-0.1197	0.4046	-0.4654
25	-0.0168	0.0008	0.0056

COVRATIO

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.
- Para mensurar o papel que a i-ésima observação exerçe sobre a precisão da estimação utilizaremos o COVRATIO.

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.
- Para mensurar o papel que a i-ésima observação exerçe sobre a precisão da estimação utilizaremos o COVRATIO.

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.
- Para mensurar o papel que a i-ésima observação exerçe sobre a precisão da estimação utilizaremos o COVRATIO.

$$COVRATIO_i = \frac{\left|\left(X'_{(i)}X_{(i)}\right)^{-1} s^2_{(i)}\right|}{\left|\left(X'X\right)^{-1}\hat{\sigma}^2\right|} \equiv \frac{s^{k+1}_{(i)}}{\hat{\sigma}^{k+1}}\left(\frac{1}{1-h_{ii}}\right)$$

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.
- Para mensurar o papel que a i-ésima observação exerçe sobre a precisão da estimação utilizaremos o COVRATIO.

$$COVRATIO_i = \frac{\left|\left(X'_{(i)}X_{(i)}\right)^{-1} s^2_{(i)}\right|}{\left|\left(X'X\right)^{-1}\hat{\sigma}^2\right|} \equiv \frac{s_{(i)}^{k+1}}{\hat{\sigma}^{k+1}}\left(\frac{1}{1-h_{ii}}\right)$$

- Se $COVRATIO_i > 1$, a i-ésima observação ajuda a melhorar a precisão da estimação

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.
- Para mensurar o papel que a i-ésima observação exerçe sobre a precisão da estimação utilizaremos o COVRATIO.

$$COVRATIO_i = \frac{\left|\left(X'_{(i)}X_{(i)}\right)^{-1} s^2_{(i)}\right|}{\left|\left(X'X\right)^{-1}\hat{\sigma}^2\right|} \equiv \frac{s_{(i)}^{k+1}}{\hat{\sigma}^{k+1}}\left(\frac{1}{1-h_{ii}}\right)$$

- Se $COVRATIO_i > 1$, a i-ésima observação ajuda a melhorar a precisão da estimação
- Se $COVRATIO_i < 1$, a i-ésima observação piora a precisão da estimação

COVRATIO

- A distância de Cook, DFFITS e DFBETAS fornecem informações úteis para conhecer o efeito da i-ésima observação nos $\hat{\beta}_j$ e $\hat{y}_i$
- Contudo, eles não fornecem nenhuma informação sobre a precisão geral da estimativa.
- Para mensurar o papel que a i-ésima observação exerçe sobre a precisão da estimação utilizaremos o COVRATIO.

$$COVRATIO_i = \frac{\left|\left(X'_{(i)}X_{(i)}\right)^{-1}s^2_{(i)}\right|}{\left|\left(X'X\right)^{-1}\hat{\sigma}^2\right|} \equiv \frac{s^{k+1}_{(i)}}{\hat{\sigma}^{k+1}}\left(\frac{1}{1-h_{ii}}\right)$$

- Se $COVRATIO_i > 1$, a i-ésima observação ajuda a melhorar a precisão da estimação
- Se $COVRATIO_i < 1$, a i-ésima observação piora a precisão da estimação
- No R, utilizamos a função `covratio()`

COVRATIO

	n.prod	distance	delTime	hii	Di	DFFITS	COVRATIO
1	7	560	16.68	0.1018	0.1001	-0.5709	0.8711
2	3	220	11.50	0.0707	0.0034	0.0986	1.2149
3	3	340	12.03	0.0987	0.0000	-0.0052	1.2757
4	4	80	14.88	0.0854	0.0776	0.5008	0.8760
5	6	150	13.75	0.0750	0.0005	-0.0395	1.2396
6	7	330	18.11	0.0429	0.0001	-0.0188	1.1999
7	2	110	8.00	0.0818	0.0022	0.0790	1.2398
8	7	210	17.83	0.0637	0.0031	0.0938	1.2056
9	30	1460	79.24	0.4983	3.4193	4.2961	0.3422
10	5	605	21.50	0.1963	0.0538	0.3987	1.3054
11	16	688	40.33	0.0861	0.0162	0.2180	1.1717
12	10	215	21.00	0.1137	0.0016	-0.0677	1.2906
13	4	255	13.50	0.0611	0.0023	0.0813	1.2070

COVRATIO

	n.prod	distance	delTime	hii	Di	DFFITS	COVRATIO
14	6	462	19.75	0.0782	0.0033	0.0974	1.2277
15	9	448	24.00	0.0411	0.0006	0.0426	1.1918
16	10	776	29.00	0.1659	0.0033	-0.0972	1.3692
17	6	200	15.35	0.0594	0.0004	0.0339	1.2192
18	7	132	19.00	0.0963	0.0440	0.3653	1.0692
19	3	36	9.50	0.0964	0.0119	0.1862	1.2153
20	17	770	35.10	0.1017	0.1324	-0.6718	0.7598
21	10	140	17.90	0.1653	0.0509	-0.3885	1.2377
22	26	810	52.32	0.3916	0.4510	-1.1950	1.3981
23	9	450	18.75	0.0413	0.0299	-0.3075	0.8897
24	8	635	19.83	0.1206	0.1023	-0.5711	0.9476
25	4	150	10.75	0.0666	0.0001	-0.0176	1.2311

Exemplo

Exemplo

```r
modelo1 <- lm(delTime ~ n.prod + distance,
              data = delivery)
modelo2 <- lm(delTime ~ n.prod + distance,
              data = delivery[-9,])
modelo3 <- lm(delTime ~ n.prod + distance,
              data = delivery[-22,])
modelo4 <- lm(delTime ~ n.prod + distance,
              data = delivery[-c(9,22),])
```

Exemplo

	Beta0	Beta1	Beta2	sigma	R2
1	2.341	1.616	0.014	3.259	0.956
2	4.447	1.498	0.010	2.430	0.944
3	1.916	1.786	0.012	3.173	0.952
4	4.643	1.456	0.011	2.483	0.898

O modelo2 (sem a observação 9) produz uma mudança considerável na estimação dos β's, reduz o $\hat{\sigma}$ e o R^2 é pouco alterado. O modelo3 não apresenta muitas mudanças em comparação com o modelo1. O modelo3, não presenta muita diferença com o modelo2. Talvez devemos retirar apenas a observação 9.

Leituras Recomendadas

- Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. Introduction to linear regression analysis. John Wiley & Sons, 2021. Chapter “Diagnostic for leverage and influence”