



# ME613 - Análise de Regressão

Parte 12

# Gráficos de Regressão Parcial

# Introdução

Vimos anteriormente:

- Gráfico dos resíduos versus variável preditora: podemos usar para checar presença de curvatura.
- Gráfico dos resíduos versus variável preditora não incluída no modelo: decidir se deve ser incluída.

Problema: estes gráficos não mostram o efeito marginal de uma variável, dado que as demais já estão no modelo.

# Gráfico de regressão parcial

*Added-variable plots* ou *adjusted variable plots* fornecem informação sobre a importância marginal de  $X_k$ , considerando as demais variáveis já incluídas no modelo.

Para o efeito marginal de  $X_k$ , consideramos os resíduos da regressão de  $Y$  nas demais variáveis e os resíduos da regressão de  $X_k$  nas demais variáveis.

O gráfico destes dois resíduos mostra a importância marginal de  $X_k$  na redução da variabilidade do resíduo. E também pode fornecer informação sobre a natureza da relação marginal de  $X_k$  com  $Y$ .

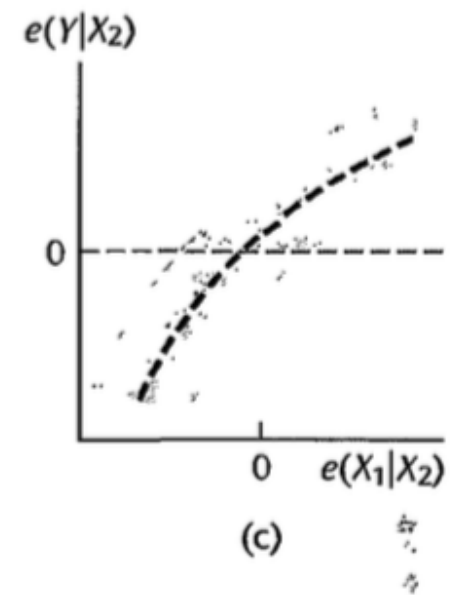
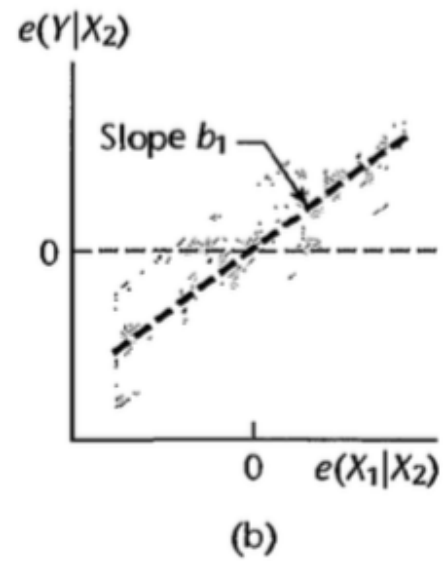
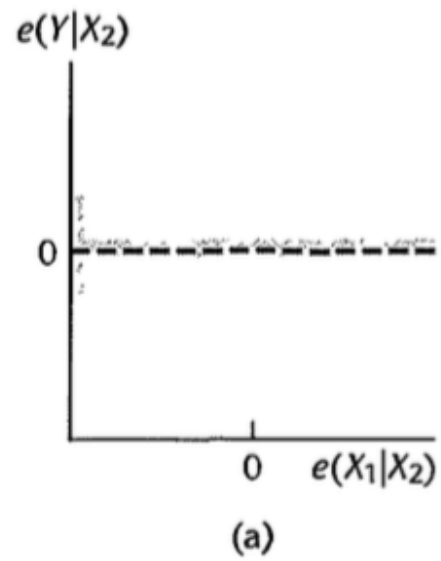
# Exemplo

Considere uma regressão múltipla de primeira ordem com duas variáveis preditoras:  $X_1$  e  $X_2$ .

Queremos estudar o efeito de  $X_1$ , dado que  $X_2$  já está no modelo.

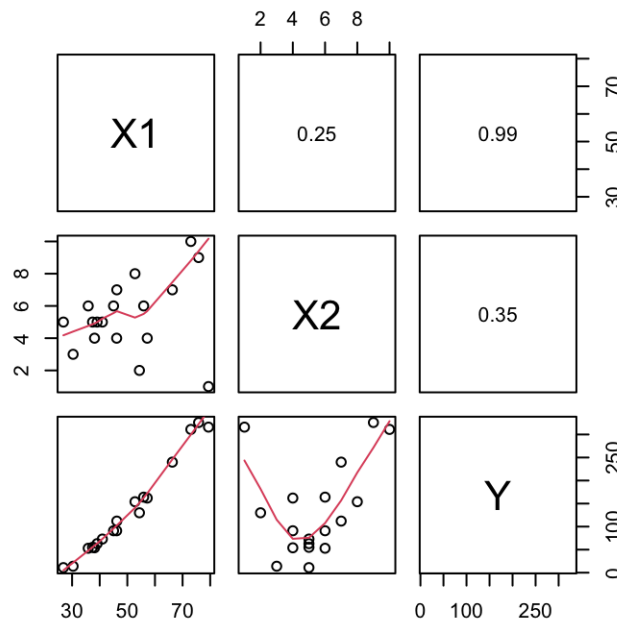
- Fazemos a regressão  $Y$  em  $X_2$  e obtemos os resíduos:  $e(Y \mid X_2)$ .
- Fazemos a regressão de  $X_1$  em  $X_2$  e obtemos os resíduos:  $e(X_1 \mid X_2)$ .
- Fazemos o gráfico de  $e(Y \mid X_2)$  versus  $e(X_1 \mid X_2)$ .

# Exemplo



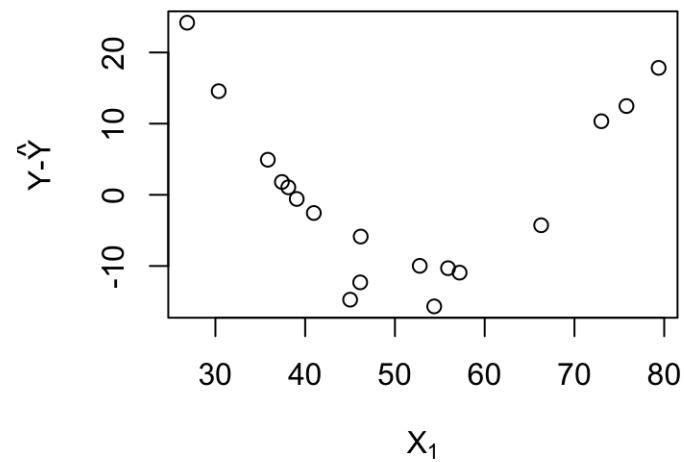
# Exemplo: Salário de gerentes

Para cada gerente: média salarial anual nos últimos 2 anos ( $X_1$ ), medida de aversão a risco ( $X_2$ ) e valor do seguro de vida ( $Y$ ).



# Exemplo: Salário de gerentes

	Estimativa	Erro-Padrão	t	valor de p
(Intercept)	-205.718659	11.3926829	-18.057086	0.0000000
X1	6.288029	0.2041495	30.801102	0.0000000
X2	4.737602	1.3780794	3.437829	0.0036622

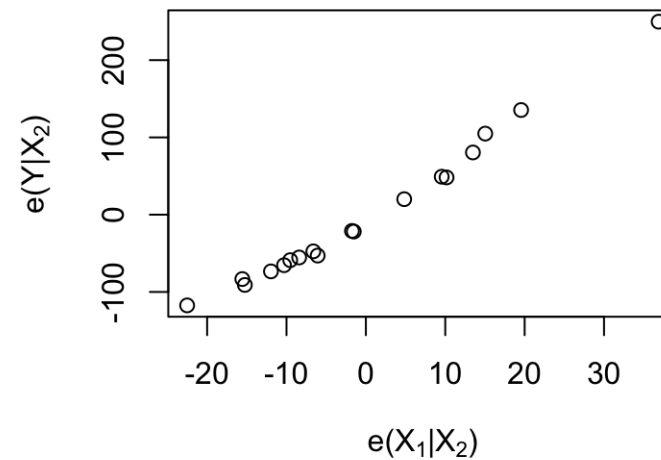




# Exemplo: Salário de gerentes

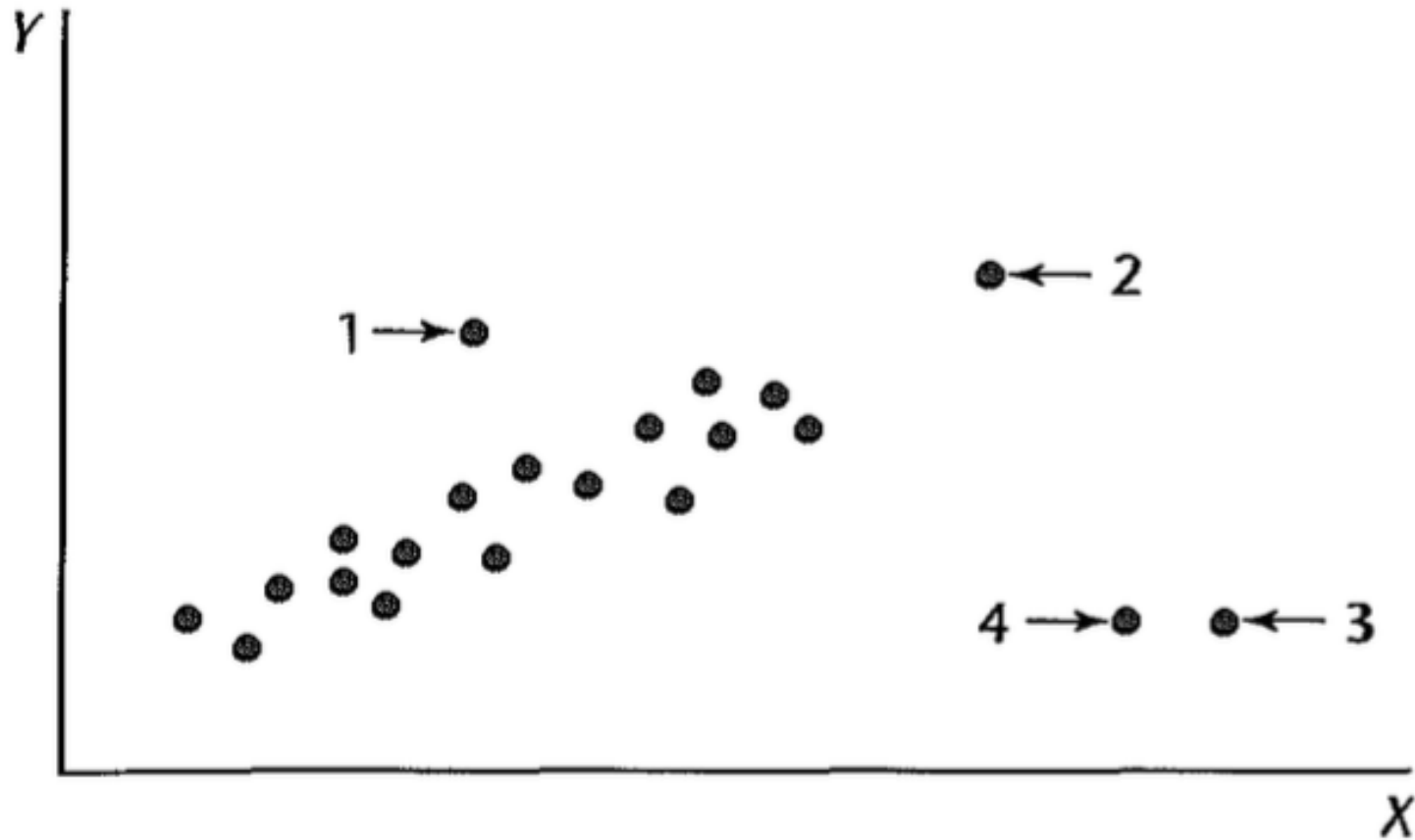
```
modelo1 <- lm(Y ~ X2, data=dados)
y_x2 <- resid(modelo1)
modelo2 <- lm(X1 ~ X2, data=dados)
x1_x2 <- resid(modelo2)
modelo3r <- lm(y_x2 ~ x1_x2)
```

	Estimativa	Erro-Padrão	t	valor de p
x1_x2	6.288029	0.1976669	31.81124	0



*Outliers*

# Introdução



*Outliers* em  $Y$

# Resíduo Semi-studentizado

$$e_i^* = \frac{Y_i - \hat{Y}_i}{\sqrt{QME}}$$

Para  $n$  grande, quando  $|e_i^*| > 4$  considera-se a  $i$ -ésima observação como *outlier*.

# Exemplo - Gordura corporal

```
dat = read.table('./dados/fat.txt')
colnames(dat) <- c("X1", "X2", "X3", "Y")
X1 = dat[,1]
X2 = dat[,2]
X3 = dat[,3]
Y = dat[,4]
modelo1 <- lm(Y ~ X1 + X2, data=dat)
rstandard(modelo1)
```

```
##           1           2           3           4           5
## -0.7402261049  1.4765806027 -1.5757913418 -1.3171518783 -0.0001308889
##           6           7           8           9          10
## -0.1519867578  0.3064529233  1.6606069555  1.1095479783 -1.0316491806
##          11          12          13          14          15
##  0.1407848595  0.9272163634 -1.7121512617  1.4686083237  0.2747599004
##          16          17          18          19          20
##  0.2655244112 -0.3538020409 -0.3435016352 -1.1631291289  0.4197625143
```

# Matriz “chapéu”

$$\mathbf{H}_{n \times n} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$$

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\text{Var}(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$$

$$\text{Var}(e_i) = \sigma^2(1 - h_{ii})$$

$$h_{ii} = \mathbf{X}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_i$$

$$\mathbf{X}_{i \times p} = \begin{pmatrix} 1 \\ X_{i,1} \\ \vdots \\ X_{i,p-1} \end{pmatrix}$$

$$\text{Cov}(e_i, e_j) = -h_{ij}\sigma^2 \quad i \neq j$$

# Resíduo Studentizado

$$\widehat{Var}(e_i) = QME(1 - h_{ii})$$

$$\widehat{Cov}(e_i, e_j) = -h_{ij}QME \quad i \neq j$$

Resíduo studentizado:

$$r_i = \frac{e_i}{\sqrt{QME(1 - h_{ii})}}$$



# Resíduo Studentizado com observação excluída

Se uma observação é muito discrepante, ela pode influenciar no ajuste.

Procedimento:

- excluir a  $i$ -ésima observação
- ajustar o modelo com as  $n - 1$  observações restantes
- obter  $\hat{Y}_{i(i)}$ : o valor predito para a  $i$ -ésima observação quando esta foi excluída no ajuste do modelo.

$$d_i = Y_i - \hat{Y}_{i(i)} = \frac{e_i}{1 - h_{ii}}$$

Obs: quanto maior  $h_{ii}$ , maior será  $d_i$ , em comparação com  $e_i$ .

# Resíduo Studentizado com observação excluída

$$\widehat{Var}(d_i) = QME_{(i)}(1 + \mathbf{X}_i^T (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_i) = \frac{QME_{(i)}}{1 - h_{ii}}$$

Resíduo Studentizado com observação excluída

$$t_i = \frac{d_i}{\sqrt{\widehat{Var}(d_i)}} \sim t_{n-p-1}$$

podemos calcular  $t_i$  sem de fato fazer ajustes separados para cada observação excluída:

$$t_i = e_i \left[ \frac{n - p - 1}{SQE(1 - h_{ii}) - e_i^2} \right]^{1/2}$$

# Resíduo Studentizado com observação excluída

A observação  $i$  é um *outlier* se  $|t_i| > t_{n-p-1}(1 - \alpha/2n)$ , utilizando Bonferroni.

# Exemplo - Gordura corporal

```
e <- resid(modelo1)
h <- hatvalues(modelo1)
t <- rstudent(modelo1)
round(data.frame("e"=e, "h"=h, "t"=t), 3)
```

##		e	h	t
## 1		-1.683	0.201	-0.730
## 2		3.643	0.059	1.534
## 3		-3.176	0.372	-1.654
## 4		-3.158	0.111	-1.348
## 5		0.000	0.248	0.000
## 6		-0.361	0.129	-0.148
## 7		0.716	0.156	0.298
## 8		4.015	0.096	1.760
## 9		2.655	0.115	1.118
## 10		-2.475	0.110	-1.034
## 11		0.336	0.120	0.137
## 12		2.226	0.109	0.923
## 13		-3.947	0.178	-1.826
## 14		3.447	0.148	1.525
## 15		0.571	0.333	0.267
## 16		0.642	0.095	0.258
## 17		-0.851	0.106	-0.345
## 18		-0.783	0.197	-0.334
## 19		-2.857	0.067	-1.176
## 20		1.040	0.050	0.409

# Exemplo - Gordura corporal

```
alpha=0.10  
n = dim(dat)[1]  
p = length(coefficients(modelo1))  
t_c <- qt(1-alpha/(2*n),df=n-p-1)
```

Se  $|t_i| > 3.25$ , então  $i$  é observação *outlier*.

*Outliers* em  $X$

# Matriz “chapéu”

- $0 \leq h_{ii} \leq 1$
- $\sum_{i=1}^n h_{ii} = p$ , lembrando que  $p$  é o número de parâmetros no modelo, incluindo o intercepto.
- Alavanca (*leverage*):  $h_{ii}$  mede a distância entre os valores de  $X$  da  $i$ -ésima observação e os valores médios de  $X$  para as  $n$  observações.

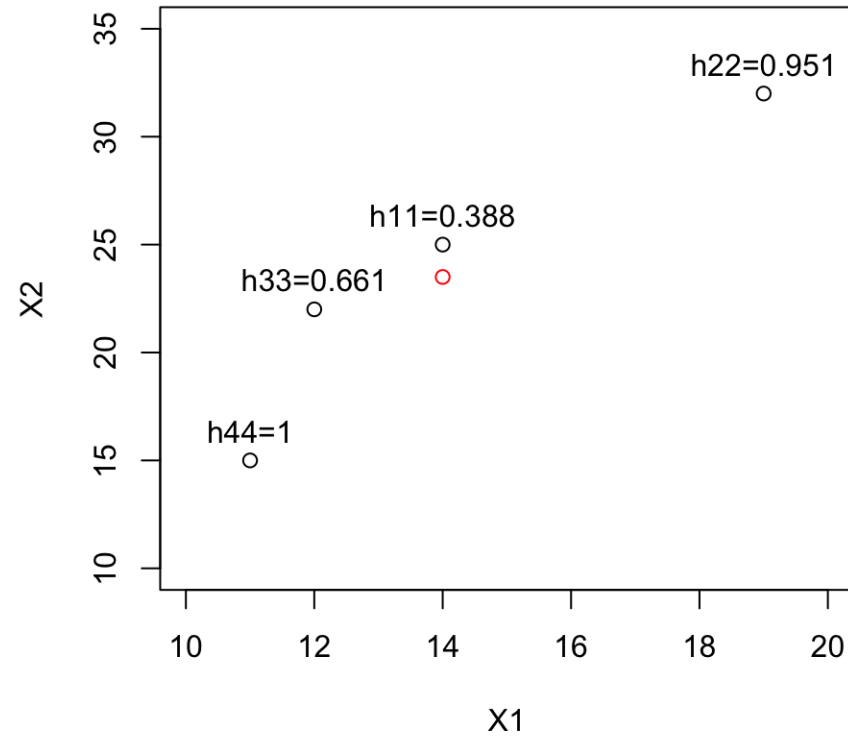
# Exemplo

```
X1 <- c(14,19,12,11)
X2 <- c(25,32,22,15)
Y <- c(301,327,246,187)
Xmatriz <- matrix(cbind(rep(1,length(X1)),X1,X2),ncol=3)
H <- Xmatriz %*% solve(t(Xmatriz)%*%Xmatriz) %*% t(Xmatriz)
hii <- diag(H)
cbind(X1,X2,hii)
```

```
##      X1 X2      hii
## [1,] 14 25 0.3876812
## [2,] 19 32 0.9512882
## [3,] 12 22 0.6614332
## [4,] 11 15 0.9995974
```



# Exemplo



# Matriz “chapéu”

$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$ , portanto cada  $\hat{Y}_i$  é uma combinação linear de todos os valores de  $Y$  e o peso de cada  $Y_i$  para o valor ajustado  $\hat{Y}_i$  depende de  $h_{ii}$ .

Quanto maior  $h_{ii}$ , maior o peso de  $Y_i$  em  $\hat{Y}_i$ .

$h_{ii}$  é uma função que depende apenas dos valores de  $X$ , portanto, mede o papel dos valores de  $X$  na determinação da importância de cada  $Y_i$  no valor ajustado  $\hat{Y}_i$ .

Quanto maior  $h_{ii}$ , menor a variância de  $e_i$ . Desta forma, quanto maior  $h_{ii}$ , mais próximo  $\hat{Y}_i$  tenderá a estar de  $Y_i$ .

# Ponto de alavanca

Um valor alavanca  $h_{ii}$  é considerado alto se é duas vezes maior que o valor de alavanca médio, denotado por  $\hat{h}$ :

$$\bar{h} = \frac{\sum_{i=1}^n h_{ii}}{n} = \frac{p}{n}$$

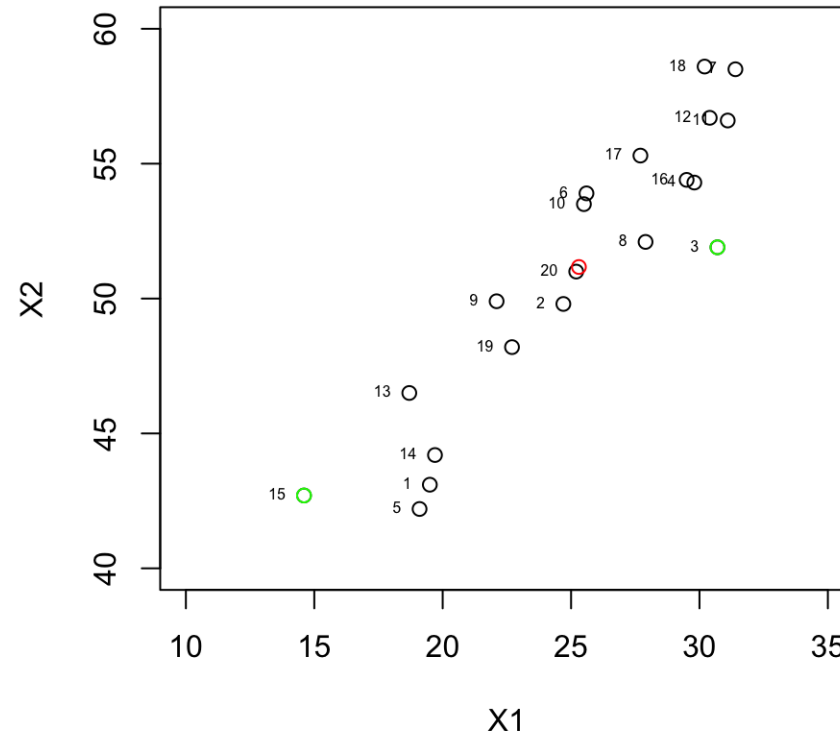
Desta maneira, observações em que  $h_{ii} > 2p/n$  são consideradas *outliers* com respeito aos valores de  $X$ .

# Exemplo - Gordura corporal

```
e <- resid(modelo1)
h <- hatvalues(modelo1)
t <- rstudent(modelo1)
round(data.frame("e"=e, "h"=h, "t"=t), 3)
```

##		e	h	t
## 1		-1.683	0.201	-0.730
## 2		3.643	0.059	1.534
## 3		-3.176	0.372	-1.654
## 4		-3.158	0.111	-1.348
## 5		0.000	0.248	0.000
## 6		-0.361	0.129	-0.148
## 7		0.716	0.156	0.298
## 8		4.015	0.096	1.760
## 9		2.655	0.115	1.118
## 10		-2.475	0.110	-1.034
## 11		0.336	0.120	0.137
## 12		2.226	0.109	0.923
## 13		-3.947	0.178	-1.826
## 14		3.447	0.148	1.525
## 15		0.571	0.333	0.267
## 16		0.642	0.095	0.258
## 17		-0.851	0.106	-0.345
## 18		-0.783	0.197	-0.334
## 19		-2.857	0.067	-1.176
## 20		1.040	0.050	0.409

# Exemplo - Gordura corporal



$$2p/n = 0.3$$

Observações influentes

# Introdução

Após identificar casos *outliers* com respeito a  $Y$  e/ou  $X$ , o próximo passo é verificar se esses casos são *influentes*.

Uma observação é considerada influente se sua exclusão causa grandes mudanças na regressão ajustada.

# *DFFITS* - Influência em um único valor ajustado

A influência que a  $i$ -ésima observação tem no valor ajustado  $\hat{Y}_i$  é medida por:

$$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{QME_{(i)}h_{ii}}}$$

O denominador é o desvio-padrão estimado de  $\hat{Y}_i$ .

$$Var(\hat{\mathbf{Y}}) = \mathbf{H}Var(\mathbf{Y})\mathbf{H}^T = \mathbf{H}(\sigma^2\mathbf{I})\mathbf{H}^T$$

Como  $\mathbf{H} = \mathbf{H}^T$  (simétrica) e  $\mathbf{H}\mathbf{H} = \mathbf{H}$  (idempotente), temos que:

$$Var(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H}$$



# *DFFITS* - Influência em um único valor ajustado

$$DFFITS_i = e_i \left[ \frac{n - p - 1}{SQE(1 - h_{ii} - e_i^2)} \right]^{1/2} \left( \frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} = t_i \left( \frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}$$

Se uma observação é um *outlier* em  $X$  e tem alto valor de alavanca, *DFFITS* tenderá a ter um alto valor.

Para pequenos conjuntos de dados, se  $|DFFITS_i| > 1$ , a observação é considerada influente.

Para grandes conjuntos de dados, se  $|DFFITS_i| > 2\sqrt{p/n}$ , a observação é considerada influente.

# Exemplo - Gordura corporal

```
format(dffits(modelol),scientific = FALSE)
```

```
##           1           2           3           4
## "-0.36614723450" " 0.38381029454" "-1.27306744543" "-0.47634829704"
##           5           6           7           8
## "-0.00007292347" "-0.05668650028" " 0.12793708219" " 0.57452120091"
##           9          10          11          12
## " 0.40216488535" "-0.36387250695" " 0.05054582680" " 0.32333658364"
##          13          14          15          16
## "-0.85078122680" " 0.63551411143" " 0.18885207780" " 0.08376828665"
##          17          18          19          20
## "-0.11837349597" "-0.16552651714" "-0.31507065348" " 0.09399705521"
```

# *Cook's Distance* - Influência em todos os valores ajustados

Medida para avaliar a influência de uma observação  $i$  no ajuste das  $n$  observações:

$$D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(i)})^2}{pQME}$$

Em forma matricial:

$$D_i = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})^T (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})}{pQME}$$

# *Cook's Distance* - Influência em todos os valores ajustados

$$D_i = \frac{e_i^2}{pQME} \left[ \frac{h_{ii}}{(1 - h_{ii})^2} \right]$$

Comparamos  $D_i$  com percentis de  $F(p, n - p)$ : se for maior que o percentil 50, a  $i$ -ésima observação deve ser investigada como possível ponto influente.

# Exemplo - Gordura corporal

```
format(cooks.distance(modelo1),scientific = FALSE)
```

```
##           1           2           3           4
## "0.045950548956025" "0.045481177306003" "0.490156678050177" "0.072161900262186"
##           5           6           7           8
## "0.000000001883399" "0.001136518329757" "0.005764939254433" "0.097938531802951"
##           9          10          11          12
## "0.053133515087085" "0.043957035227563" "0.000903798575500" "0.035154363872331"
##          13          14          15          16
## "0.212150240723778" "0.124892510084211" "0.012575299122032" "0.002474925197901"
##          17          18          19          20
## "0.004926142442252" "0.009636470211422" "0.032360064480030" "0.003096787039886"
```

# Exemplo - Gordura corporal

```
caso3 <- cooks.distance(modelo1)[3]  
p=length(coefficients(modelo1))  
n=dim(dat)[1]  
perc=pf(caso3,df1=p,df2=n-p)
```

Caso 3 tem o maior valor:  $D_3 = 0.49$ . Este valor corresponde ao percentil 30.6 da distribuição  $F(p, n - p)$ .

# *DFBETAS* - influência nos coeficientes da regressão

Medida de influência da  $i$ -ésima observação no coeficiente  $\hat{\beta}_k$ :

$$DFBETAS_{k(i)} = \frac{\hat{\beta}_k - \hat{\beta}_{k(i)}}{\sqrt{QME_{(i)} c_{kk}}}$$

em que  $c_{kk}$  é o  $k$ -ésimo elemento da diagonal de  $(\mathbf{X}^T \mathbf{X})^{-1}$ .

Para pequenos conjuntos de dados, se  $|DFBETAS_{k(i)}| > 1$ , a observação é considerada influente.

Para grandes conjuntos de dados, se  $|DFBETAS_{k(i)}| > 2/\sqrt{n}$ , a observação é considerada influente.

# Exemplo - Gordura corporal

```
format(dfbetas(modelo1),scientific = FALSE)
```

```
##      (Intercept)      X1      X2
## 1  "-0.30518208063" "-0.13148559192" " 0.23203185250"
## 2  " 0.17257315757" " 0.11502507830" "-0.14261289523"
## 3  "-0.84710125907" "-1.18252488189" " 1.06690317579"
## 4  "-0.10161195986" "-0.29351950126" " 0.19607192232"
## 5  "-0.00006372122" "-0.00003052747" " 0.00005023715"
## 6  " 0.03967715415" " 0.04008114113" "-0.04426759013"
## 7  "-0.07752748175" "-0.01561293306" " 0.05431633626"
## 8  " 0.26143123587" " 0.39112622713" "-0.33245331815"
## 9  "-0.15135207505" "-0.29465556652" " 0.24690908623"
## 10 " 0.23774917449" " 0.24460100708" "-0.26880860125"
## 11 "-0.00902088534" " 0.01705640110" "-0.00248451805"
## 12 "-0.13049333736" " 0.02245800361" " 0.06999608166"
## 13 " 0.11941465394" " 0.59242024965" "-0.38949127544"
## 14 " 0.45174371217" " 0.11317216394" "-0.29770422361"
## 15 "-0.00300427628" "-0.12475670942" " 0.06876928638"
## 16 " 0.00930846262" " 0.04311346974" "-0.02512498857"
## 17 " 0.07951207831" " 0.05504356655" "-0.07609007641"
## 18 " 0.13205215004" " 0.07532874247" "-0.11610031509"
## 19 "-0.12960322961" "-0.00407202961" " 0.06442930786"
## 20 " 0.01019045469" " 0.00229079680" "-0.00331414570"
```



# Exemplo - Gordura corporal

```
influence.measures(modelo1)
```

```
## Influence measures of
##   lm(formula = Y ~ X1 + X2, data = dat) :
##
##      dfb.1_    dfb.X1    dfb.X2    dffit cov.r   cook.d    hat inf
## 1  -3.05e-01 -1.31e-01  2.32e-01 -3.66e-01 1.361 4.60e-02 0.2010
## 2   1.73e-01  1.15e-01 -1.43e-01  3.84e-01 0.844 4.55e-02 0.0589
## 3  -8.47e-01 -1.18e+00  1.07e+00 -1.27e+00 1.189 4.90e-01 0.3719  *
## 4  -1.02e-01 -2.94e-01  1.96e-01 -4.76e-01 0.977 7.22e-02 0.1109
## 5  -6.37e-05 -3.05e-05  5.02e-05 -7.29e-05 1.595 1.88e-09 0.2480  *
## 6   3.97e-02  4.01e-02 -4.43e-02 -5.67e-02 1.371 1.14e-03 0.1286
## 7  -7.75e-02 -1.56e-02  5.43e-02  1.28e-01 1.397 5.76e-03 0.1555
## 8   2.61e-01  3.91e-01 -3.32e-01  5.75e-01 0.780 9.79e-02 0.0963
## 9  -1.51e-01 -2.95e-01  2.47e-01  4.02e-01 1.081 5.31e-02 0.1146
## 10  2.38e-01  2.45e-01 -2.69e-01 -3.64e-01 1.110 4.40e-02 0.1102
## 11 -9.02e-03  1.71e-02 -2.48e-03  5.05e-02 1.359 9.04e-04 0.1203
## 12 -1.30e-01  2.25e-02  7.00e-02  3.23e-01 1.152 3.52e-02 0.1093
## 13  1.19e-01  5.92e-01 -3.89e-01 -8.51e-01 0.827 2.12e-01 0.1784
## 14  4.52e-01  1.13e-01 -2.98e-01  6.36e-01 0.937 1.25e-01 0.1480
## 15 -3.00e-03 -1.25e-01  6.88e-02  1.89e-01 1.775 1.26e-02 0.3332  *
## 16  9.31e-03  4.31e-02 -2.51e-02  8.38e-02 1.309 2.47e-03 0.0953
## 17  7.95e-02  5.50e-02 -7.61e-02 -1.18e-01 1.312 4.93e-03 0.1056
## 18  1.32e-01  7.53e-02 -1.16e-01 -1.66e-01 1.462 9.64e-03 0.1968
## 19 -1.30e-01 -4.07e-03  6.44e-02 -3.15e-01 1.002 3.24e-02 0.0670
## 20  1.02e-02  2.29e-03 -3.31e-03  9.40e-02 1.224 3.10e-03 0.0501
```

Inflação da variância

# Introdução

Problemas quando variáveis preditoras apresentam correlação alta entre si:

- Incluir ou excluir uma variável preditora altera os coeficientes da regressão.
- Os erros padrão dos coeficientes estimados ficam muito grandes.

A presença de multicolinearidade pode ser investigada através dos seguintes diagnósticos informais:

- Grandes mudanças nas estimativas dos parâmetros de regressão quando uma variável é incluída ou excluída do modelo.
- Estimativa dos parâmetros de regressão com sinal oposto do que seria esperado, segundo informações/conhecimento prévio.

# Exemplo - Gordura corporal

	Estimativa	Erro-Padrão	t	valor de p
(Intercept)	-19.1742456	8.3606407	-2.2933943	0.0348433
X1	0.2223526	0.3034389	0.7327755	0.4736790
X2	0.6594218	0.2911873	2.2645969	0.0368987

---

# Exemplo - Gordura corporal

	Estimativa	Erro-Padrão	t	valor de p
(Intercept)	117.084695	99.782403	1.173400	0.2578078
X1	4.334092	3.015511	1.437266	0.1699111
X2	-2.856848	2.582015	-1.106441	0.2848944
X3	-2.186060	1.595499	-1.370142	0.1895628

# VIF - fator de inflação da variância

Lembrando:

$$Var(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

Para medir o impacto da multicolinearidade, é mais útil trabalhar com as variáveis com transformação de correlação, vistas anteriormente.

$$Var(\hat{\beta}^*) = (\sigma^*)^2 \mathbf{r}_{XX}^{-1}$$

Definindo  $VIF_k$  (fator de inflação da variância para  $\hat{\beta}_k^*$ ) como o  $k$ -ésimo elemento de  $\mathbf{r}_{XX}^{-1}$ , temos:

$$Var(\hat{\beta}_k^*) = (\sigma^*)^2 VIF_k$$

# *VIF* - fator de inflação da variância

Pode-se reescrever:

$$VIF_k = (1 - R_k^2)^{-1}$$

em que  $R_k^2$  é o coeficiente de determinação da regressão de  $X_k$  nas demais variáveis preditoras.

Quando  $R_k^2 = 0$ ,  $VIF_k = 1$ , caso contrário,  $VIF_k > 1$ .

# Exemplo - Gordura corporal

Variáveis escala original

```
modelo2 <- lm(Y ~ X1 + X2 + X3,data=dat)
kable(summary(modelo2)$coef,col.names = c("Estimativa","Erro-Padrão","t","valor de p"))
```

	Estimativa	Erro-Padrão	t	valor de p
(Intercept)	117.084695	99.782403	1.173400	0.2578078
X1	4.334092	3.015511	1.437266	0.1699111
X2	-2.856848	2.582015	-1.106441	0.2848944
X3	-2.186060	1.595499	-1.370142	0.1895628



# Exemplo - Gordura corporal

Variáveis padronizadas

	Estimativa	Erro-Padrão	t	valor de p
X1	4.263705	2.877965	1.481500	0.1567712
X2	-2.928701	2.567924	-1.140493	0.2698942
X3	-1.561417	1.105576	-1.412310	0.1759032

---

# Exemplo - Gordura corporal

```
library(car)
```

```
## Loading required package: carData
```

```
vif(modelo2)
```

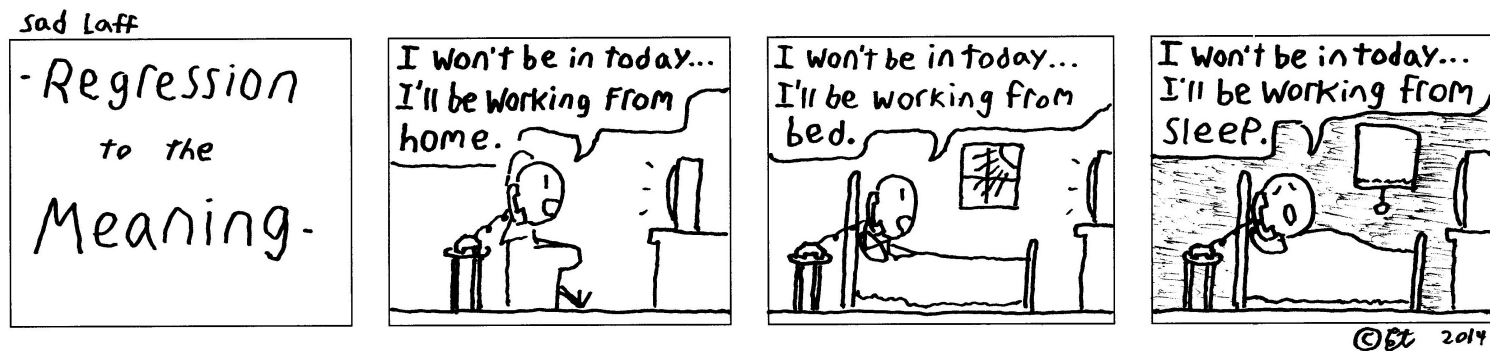
```
##           x1           x2           x3  
## 708.8429 564.3434 104.6060
```

# Agradecimento

- Slides criados por Samara F Kiihl / IMECC / UNICAMP

# Leitura

- Applied Linear Statistical Models: Capítulo 10.
- Faraway - [Linear Models with R](#): Capítulo 6, Seção 7.3
- Draper & Smith - [Applied Regression Analysis](#): Capítulo 8.
- Caffo - [Regression Models for Data Science in R](#): Residuals, variation, diagnostics.



# Agradecimentos

- Samara F Kiihl
- Benilton S Carvalho
- Rafael P Maia