



Aspectos da biologia reprodutiva da espécie *Atya scabra*

Fernando Henrique Ferraz P. da Rosa
Matheus Moreira Costa
Vagner Aparecido Pedro Junior

26 de setembro de 2004

Resumo

Projeto de MAE0328 - Análise de Regressão baseado no Relatório de Análise Estatística do Centro de Estatística Aplicada (CEA-USP), RAE 9607[1].

[Home Page](#)

[Title Page](#)

[Contents](#)



[Page 1 of 24](#)

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

[Home Page](#)[Title Page](#)[Contents](#)[Page 2 of 24](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

1. Descrição do estudo

O estudo envolveu 537 camarões da espécie *Atya scabra*, colhidos no período de fevereiro de 95 a fevereiro de 96 no Rio Guaecá, em São Sebastião (SP).

1.1. Variáveis

- CT - comprimento total (mm), obtido com paquímetro
- CC - comprimento da carapaça (mm), obtido com paquímetro
- PESO - peso úmido (g), obtido com balança, sendo que os ovos das fêmeas ovíferas foram retirados antes da pesagem total
- SEXO - sexo do animal.

1.2. Relações de Interesse

- Peso em função de CT e Sexo
- CC em função de CT e Sexo



2. Lendo o conjunto de Dados

Para ler o conjunto de dados no data.frame cam no R [2] fazemos:

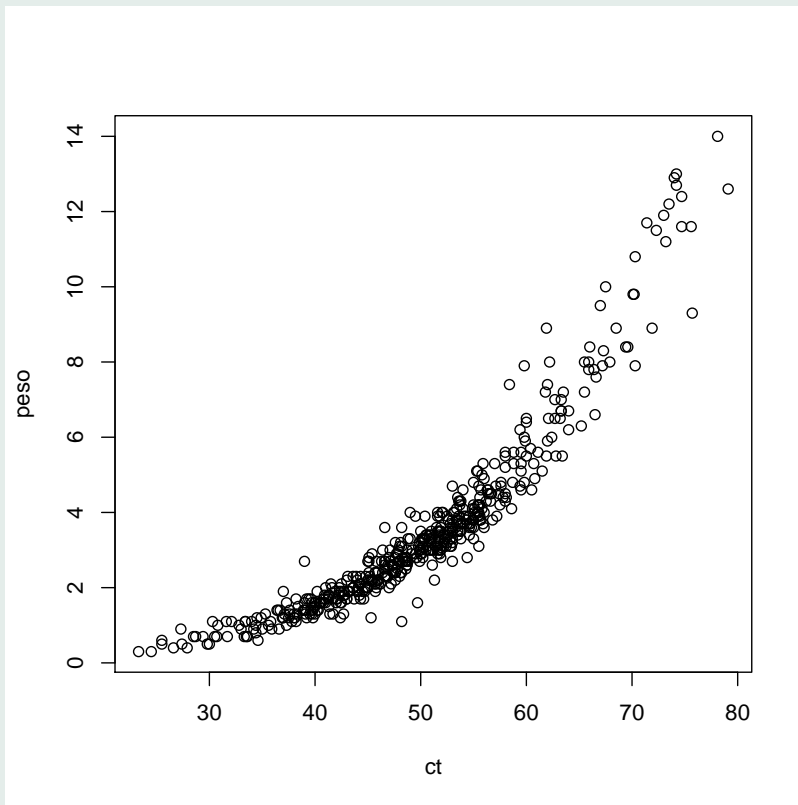
```
> cam <- read.delim("Dados.dat", header = FALSE, row.names = 1,  
+ strip.white = TRUE, col.names = c("camarao", "mes", "sexo",  
+ "peso", "ct", "cc", "cm", "am", "tam.ovario", "ovos",  
+ "olhos", "cor.ovario"), quote = "", as.is = 10:12)  
> cam$cor.ovario <- factor(cam$cor.ovario, exclude = "")  
> cam$ovos <- factor(cam$ovos, exclude = "")  
> cam$olhos <- factor(cam$olhos, exclude = "")  
> cam$ct[414] <- 55.4  
> cam$cm[403] <- 8.6
```

[Home Page](#)[Title Page](#)[Contents](#)[Page 3 of 24](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)



3. Peso em função de CT e Sexo

Começemos observando o diagrama de dispersão do peso pelo comprimento total (ct):



[Home Page](#)

[Title Page](#)

[Contents](#)



[Page 4 of 24](#)

[Go Back](#)

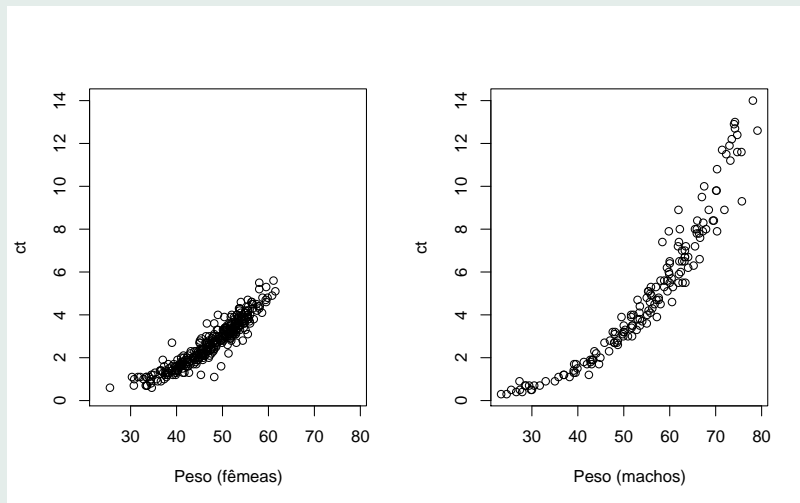
[Full Screen](#)

[Close](#)

[Quit](#)



O mesmo gráfico, por sexo:



Os gráficos anteriores sugerem uma relação exponencial para ambos os sexos. Tomamos a transformação logarítmica $Y = \log(\text{peso})$:

[Home Page](#)

[Title Page](#)

[Contents](#)



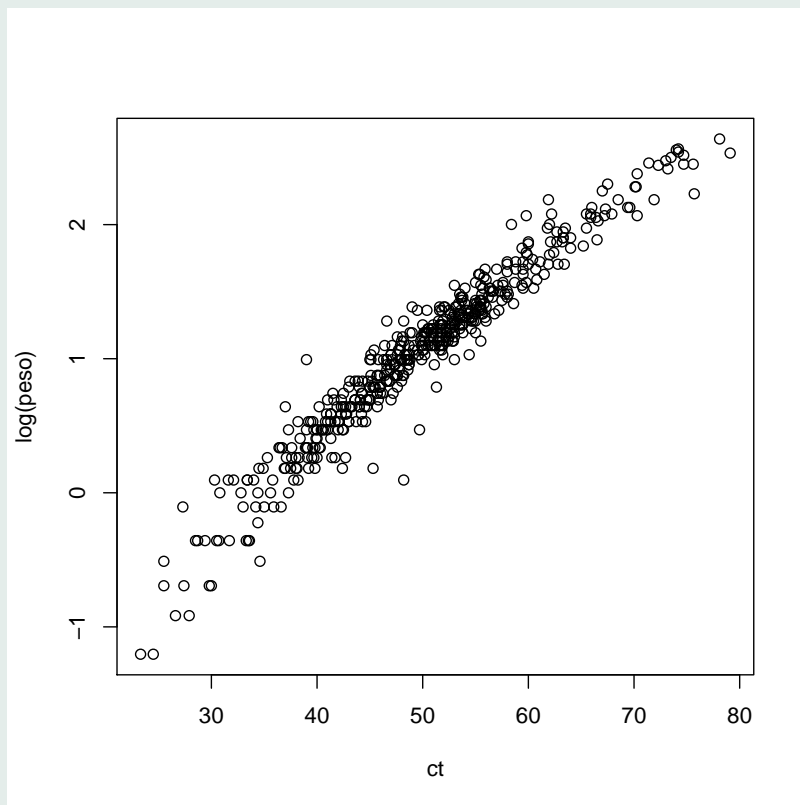
Page 5 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



O gráfico anterior sugere uma curvatura na relação entre a variável transformada e o peso. Por hora, entretanto, ajustemos o modelo linear de primeira ordem:

[Home Page](#)

[Title Page](#)

[Contents](#)



Page 6 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



$$Y = \beta_0 + \beta_1 ct + \beta_2 sexo + \beta_3 ct * sexo + \epsilon \quad (1)$$

```
> z <- lm(log(peso) ~ ct + sexo + ct:sexo)
> summary(z)
```

Call:

```
lm(formula = log(peso) ~ ct + sexo + ct:sexo)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.854446	-0.071988	0.005979	0.089148	0.622172

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.082004	0.061737	-33.724	<2e-16 ***
ct	0.062900	0.001290	48.754	<2e-16 ***
sexoM	-0.162194	0.082298	-1.971	0.0493 *
ct:sexoM	0.002787	0.001616	1.725	0.0852 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1607 on 493 degrees of freedom

Multiple R-Squared: 0.94, Adjusted R-squared: 0.9396

F-statistic: 2575 on 3 and 493 DF, p-value: < 2.2e-16

[Home Page](#)

[Title Page](#)

[Contents](#)



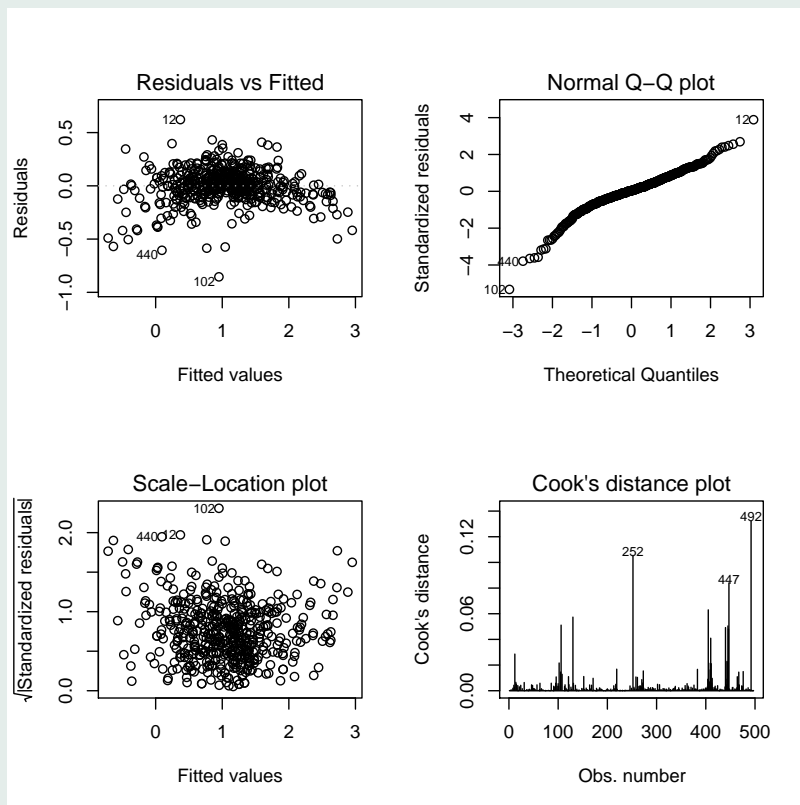
Page 7 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



O padrão mostrado pelo gráfico de resíduos versus ajustados indica que o efeito da curvatura em ct é realmente necessário. Colocamos então no modelo uma variável quadrática:

[Home Page](#)

[Title Page](#)

[Contents](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

[Page 8 of 24](#)

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



$$Y = \beta_0 + \beta_1 ct + \beta_2 sexo + \beta_3 ct * sexo + \beta_4 ct^2 + \epsilon \quad (2)$$

```
> z2 <- lm(log(peso) ~ ct * sexo + I(ct^2))
> summary(z2)
```

Call:

```
lm(formula = log(peso) ~ ct * sexo + I(ct^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.8793461	-0.0712487	-0.0006775	0.0857149	0.6217571

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.266e+00	1.231e-01	-26.539	< 2e-16 ***
ct	1.157e-01	5.032e-03	22.993	< 2e-16 ***
sexoM	-4.003e-01	7.731e-02	-5.178	3.27e-07 ***
I(ct^2)	-5.749e-04	5.331e-05	-10.784	< 2e-16 ***
ct:sexoM	9.180e-03	1.571e-03	5.843	9.33e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1446 on 492 degrees of freedom

Multiple R-Squared: 0.9515, Adjusted R-squared: 0.9511

F-statistic: 2412 on 4 and 492 DF, p-value: < 2.2e-16

[Home Page](#)

[Title Page](#)

[Contents](#)



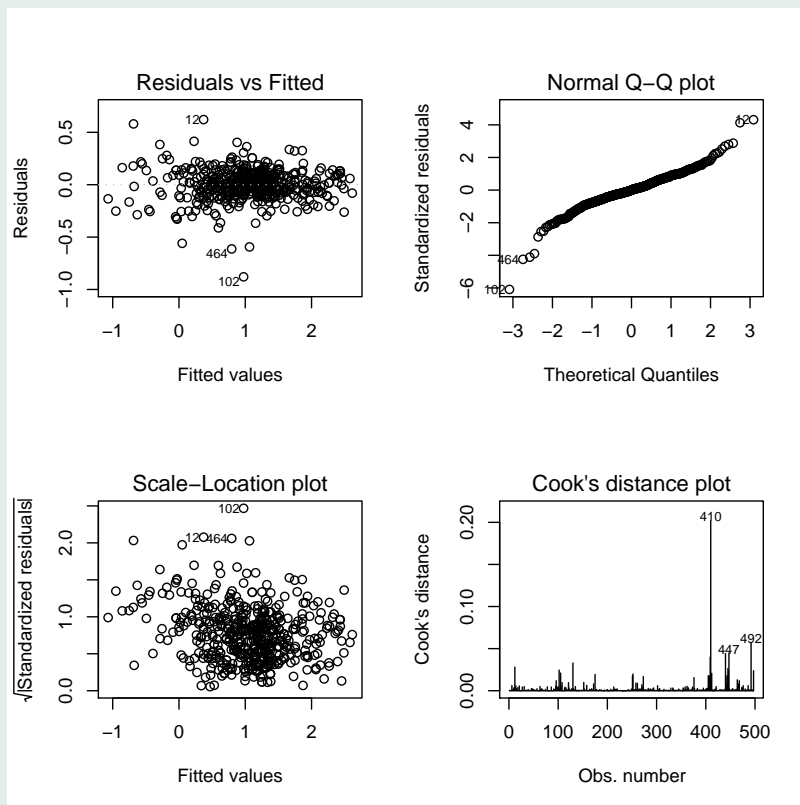
Page 9 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



O padrão nos resíduos desapareceu, mostrando que o efeito quadrático era adequado. O gráfico QQ-Plot e o Scale-Location plot indicam entretanto que há pontos extremos que podem estar influenciando o modelo. A seguir temos o gráfico do modelo ajustado

[Home Page](#)

[Title Page](#)

[Contents](#)



[Page 10 of 24](#)

[Go Back](#)

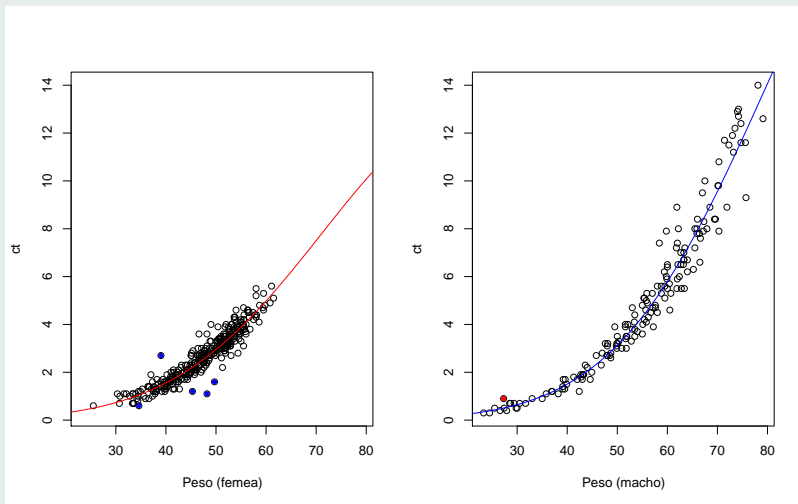
[Full Screen](#)

[Close](#)

[Quit](#)



e os pontos com resíduos padronizados maiores em módulo a 3:



Removendo esses 6 pontos do modelo, temos:

```
> influentes <- abs(rstudent(z2)) > 3  
> z3 <- update(z2, subset = !influentes)  
> summary(z3)
```

Call:

```
lm(formula = log(peso) ~ ct * sexo + I(ct^2), subset = !influentes)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

[Home Page](#)

[Title Page](#)

[Contents](#)



Page 11 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

[Home Page](#)[Title Page](#)[Contents](#)[Page 12 of 24](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

-0.404887 -0.071089 -0.001402 0.084869 0.408650

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.339e+00	1.087e-01	-30.731	< 2e-16 ***
ct	1.193e-01	4.442e-03	26.858	< 2e-16 ***
sexoM	-4.640e-01	6.850e-02	-6.773	3.66e-11 ***
I(ct^2)	-6.151e-04	4.705e-05	-13.075	< 2e-16 ***
ct:sexoM	1.032e-02	1.390e-03	7.424	5.12e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.126 on 486 degrees of freedom

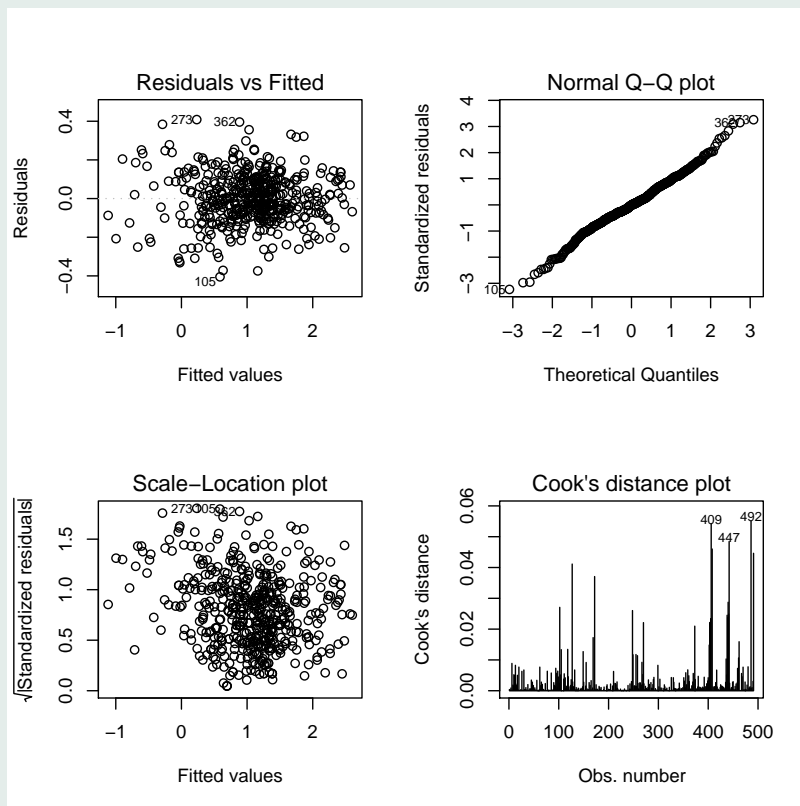
Multiple R-Squared: 0.9626, Adjusted R-squared: 0.9623

F-statistic: 3127 on 4 and 486 DF, p-value: < 2.2e-16

Comparando os coeficientes dos modelos 2 e 3, temos:

	(Intercept)	ct	sexoM	I(ct ²)	ct:sexoM
z2	-3.27	0.12	-0.40	-0.00	0.01
z3	-3.34	0.12	-0.46	-0.00	0.01

Donde concluímos que aqueles pontos tinham pouca influência sobre os valores dos parâmetros então não perdemos muito em descartá-los. Para o modelo 3 temos os gráficos de resíduos:



Onde não se observa nenhuma quebra das suposições para modelagem, concluído que o modelo está bem ajustado. Temos ainda $R_{adj}^2 = 0.9623$.

[Home Page](#)

[Title Page](#)

[Contents](#)



[Page 13 of 24](#)

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

[Home Page](#)[Title Page](#)[Contents](#)[Page 14 of 24](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Podemos ainda testar se o efeito de interação de primeira ordem CT:Sexo pode ser descartado do modelo, através do modelo linear geral [3]. Nosso modelo completo é z3, e o reduzido o que não tem o efeito de interação:

```
> zcompleto <- z3
> zreduzido <- update(zcompleto, ~. - ct:sexo)
> anova(zreduzido, zcompleto)
```

Analysis of Variance Table

Model 1: log(peso) ~ ct + sexo + I(ct^2)

Model 2: log(peso) ~ ct * sexo + I(ct^2)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	487	8.5949				
2	486	7.7194	1	0.8755	55.121	5.115e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Onde rejeitamos a hipótese nula de que o modelo reduzido é tão adequado quanto o completo, concluindo que o efeito de interação de primeira ordem ct:sexo é significativo.

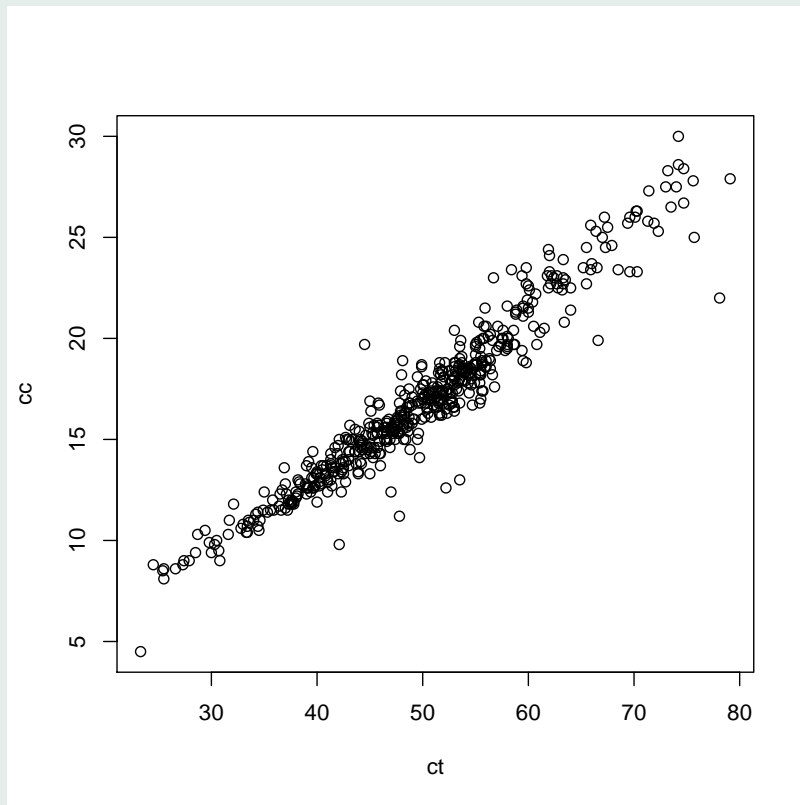
Voltando a escala original temos que o modelo final é dado então por:

$$\begin{aligned} \text{machos} & \text{ Peso} = \exp \{-3.803 + 0.1296CT - 0.0006151CT^2\} \\ \text{fêmeas} & \text{ Peso} = \exp \{-3.339 + 0.1193CT - 0.0006151CT^2\} \end{aligned} \quad (3)$$



4. CC em função de CT e Sexo

Começemos observando o diagrama de dispersão do peso pelo comprimento total (ct):



[Home Page](#)

[Title Page](#)

[Contents](#)



Page 15 of 24

[Go Back](#)

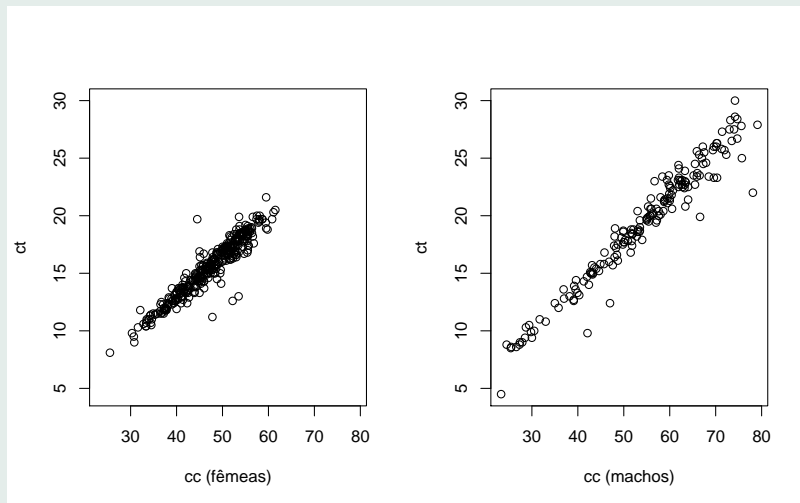
[Full Screen](#)

[Close](#)

[Quit](#)



O mesmo gráfico, por sexo:



Os gráficos anteriores sugerem uma relação linear de primeira ordem simples entre o comprimento da carapaça (cc) e o comprimento total (ct). Ajustemos então o modelo:

$$cc = \beta_0 + \beta_1 ct + \beta_2 \text{sexo} + \beta_3 ct * \text{sexo} + \epsilon \quad (4)$$

```
> z <- lm(cc ~ ct * sexo)
> summary(z)
```

Call:
`lm(formula = cc ~ ct * sexo)`

[Home Page](#)

[Title Page](#)

[Contents](#)



Page 16 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



Residuals:

	Min	1Q	Median	3Q	Max
	-6.63481	-0.34563	0.05155	0.44738	5.12590

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.978160	0.353662	-2.766	0.00588	**
ct	0.349489	0.007406	47.192	< 2e-16	***
sexoM	-0.770804	0.468015	-1.647	0.10016	
ct:sexoM	0.039548	0.009223	4.288	2.15e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9386 on 524 degrees of freedom

Multiple R-Squared: 0.946, Adjusted R-squared: 0.9456

F-statistic: 3057 on 3 and 524 DF, p-value: < 2.2e-16

[Home Page](#)

[Title Page](#)

[Contents](#)



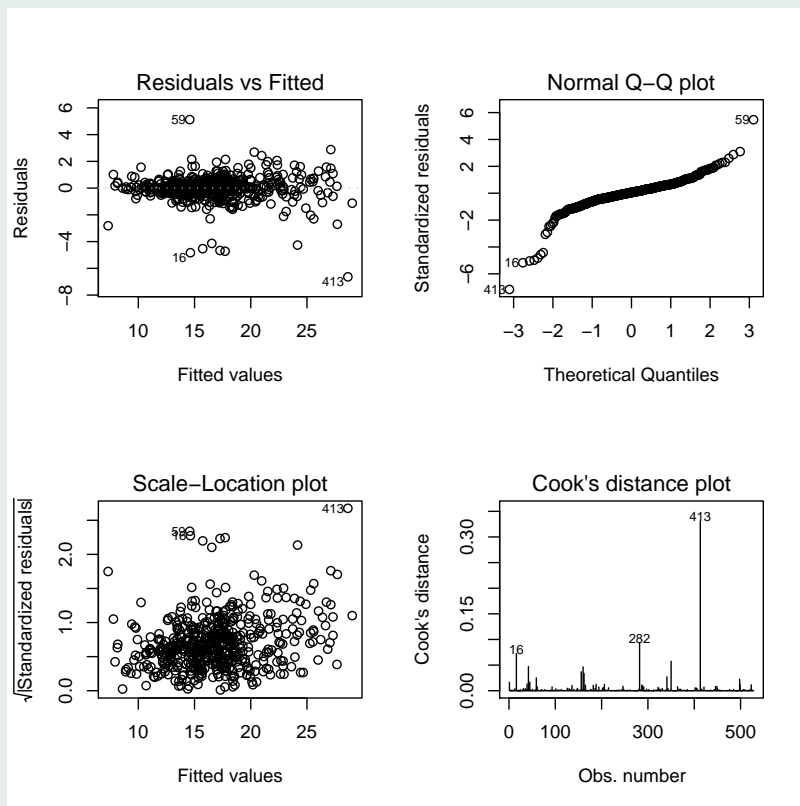
Page 17 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



Os gráficos de resíduos indicam a presença de possíveis valores extremos (outliers), abaixo temos destacados os 10 pontos com resíduo padronizado maior que 3:

[Home Page](#)

[Title Page](#)

[Contents](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

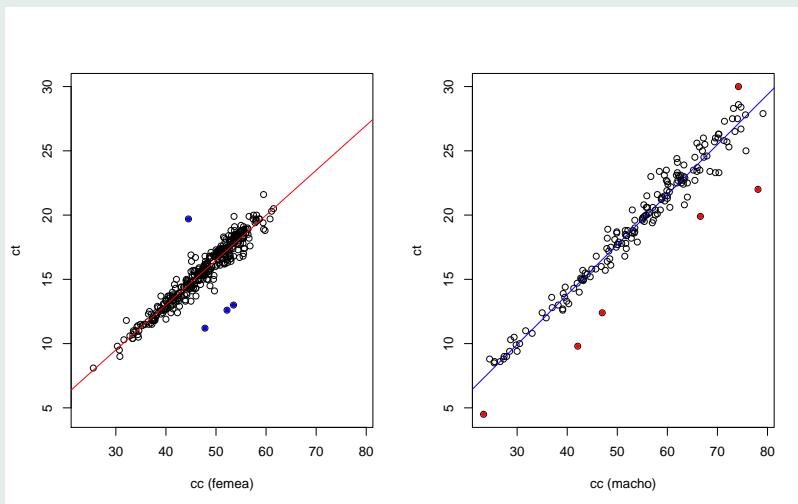
[Page 18 of 24](#)

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



Removendo esses 10 pontos do modelo, temos:

```
> influentes <- abs(rstudent(z)) > 3  
> z2 <- update(z, subset = !influentes)  
> summary(z2)
```

Call:

```
lm(formula = cc ~ ct * sexo, subset = !influentes)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.80011	-0.35589	-0.01071	0.38585	2.57616

[Home Page](#)

[Title Page](#)

[Contents](#)



Page 19 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



Coefficients:

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.157288    0.262799  -4.404 1.30e-05 ***
ct           0.353814    0.005506  64.265 < 2e-16 ***
sexoM       -0.431198    0.353471  -1.220  0.223
ct:sexoM     0.034410    0.006953   4.949 1.01e-06 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6963 on 514 degrees of freedom

Multiple R-Squared: 0.969, Adjusted R-squared: 0.9688

F-statistic: 5356 on 3 and 514 DF, p-value: < 2.2e-16

Comparando os coeficientes dos modelos 1 e 2, temos:

	(Intercept)	ct	sexoM	ct:sexoM
z	-0.98	0.35	-0.77	0.04
z2	-1.16	0.35	-0.43	0.03

Donde concluimos que aqueles pontos tinham pouca influência sobre os valores dos parâmetros então não perdemos muito em descartá-los. Para o modelo 3 temos os gráficos de resíduos:

[Home Page](#)

[Title Page](#)

[Contents](#)

◀▶

◀▶

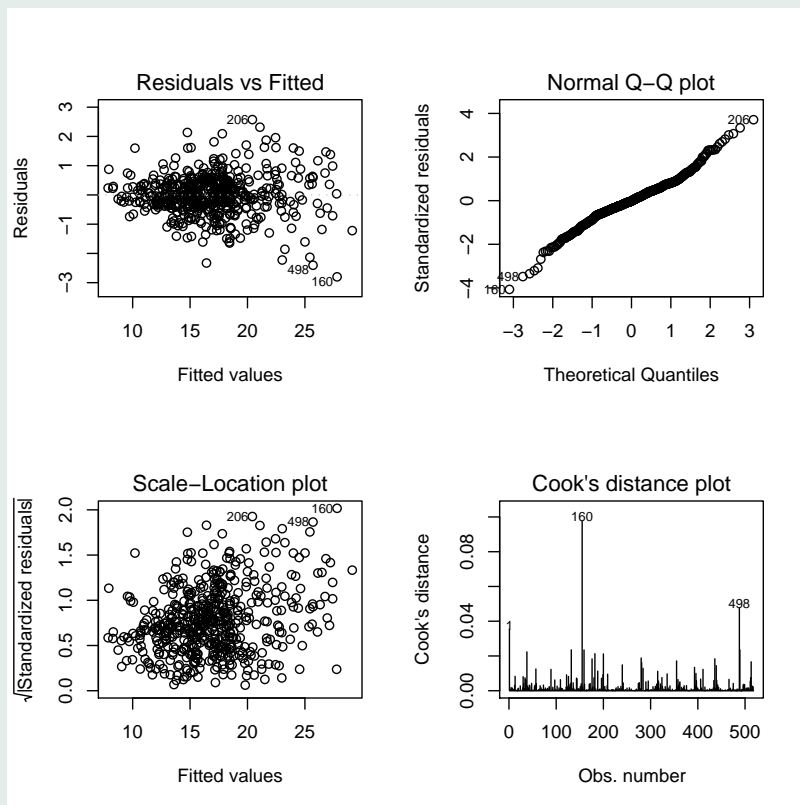
Page 20 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



Onde não se observa nenhuma quebra das suposições para modelagem, concluído que o modelo está bem ajustado. Temos ainda $R_{adj}^2 = 0.9688$.

[Home Page](#)

[Title Page](#)

[Contents](#)



[Page 21 of 24](#)

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

[Home Page](#)[Title Page](#)[Contents](#)[Page 22 of 24](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Podemos ainda testar se o efeito de interação de primeira ordem CT:Sexo pode ser descartado do modelo, através do modelo linear geral [3]. Nosso modelo completo é z2, e o reduzido o que não tem o efeito de interação:

```
> zcompleto <- z2
> zreduzido <- update(z2, ~. - ct:sexo)
> anova(zreduzido, zcompleto)
```

Analysis of Variance Table

Model 1: cc ~ ct + sexo

Model 2: cc ~ ct * sexo

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	515	261.045				
2	514	249.173	1	11.872	24.489	1.014e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Onde rejeitamos a hipótese nula de que o modelo reduzido é tão adequado quanto o completo, concluindo que o efeito de interação de primeira ordem ct:sexo é significativo.

Por fim, o modelo é dado por:

$$\begin{aligned} \text{machos} \quad CC &= -1.588 + 0.3882CT \\ \text{fêmeas} \quad CC &= -1.157 + 0.3538CT \end{aligned} \quad (5)$$



Referências

- [1] CORDANI, L. K.; LIMA, L. P. de. *Relatório de Análise Estatística Sobre o Projeto: Aspectos da biologia reprodutiva da espécie Atya scabra*. [S.l.], 1996.
- [2] R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2004. ISBN 3-900051-00-3. Disponível em: <<http://www.R-project.org>>.
- [3] NETER, J. et al. *Applied Linear Statistical Models*. Chicago: Irwin, 1996.

[Home Page](#)

[Title Page](#)

[Contents](#)



Page 23 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)



Sobre

A versão eletrônica desse arquivo pode ser obtida em <http://www.feferraz.net>

Copyright (c) 1999-2005 Fernando Henrique Ferraz Pereira da Rosa.
É dada permissão para copiar, distribuir e/ou modificar este documento
sob os termos da Licença de Documentação Livre GNU (GFDL), versão 1.2,
publicada pela Free Software Foundation;
Uma cópia da licença em está inclusa na seção intitulada
"Sobre / Licença de Uso".

[Home Page](#)

[Title Page](#)

[Contents](#)



Page 24 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)