

Modelos Gaussianos

$$Y \sim N \{ \mu = f(x_i), \sigma = f(x_i) \}$$

Primeiro exemplo:

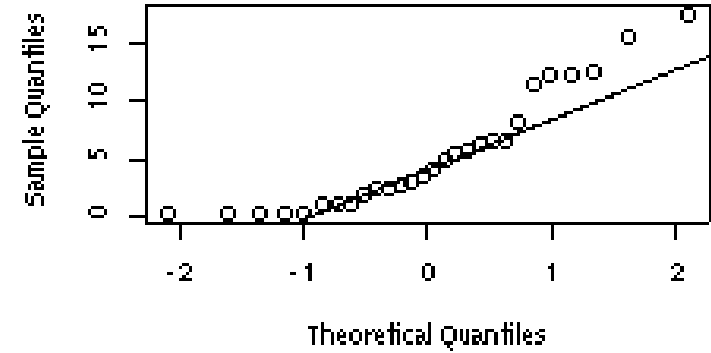
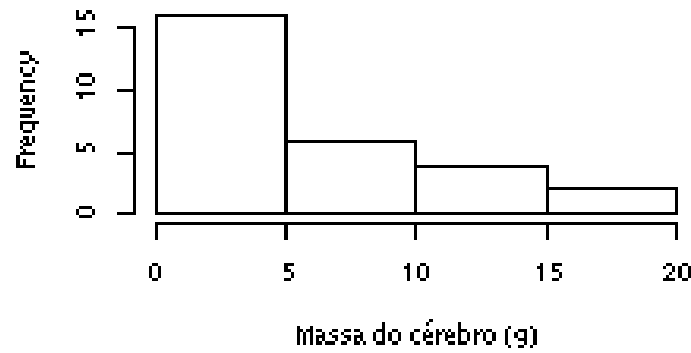
Comparação entre médias de 3 grupos

Massa (g) do cérebro de 62 espécies de mamíferos, classificados em:

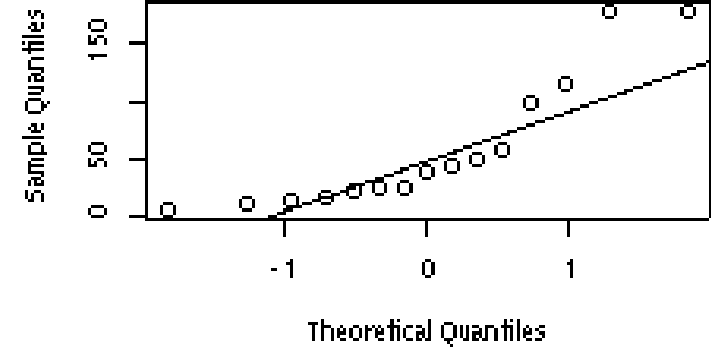
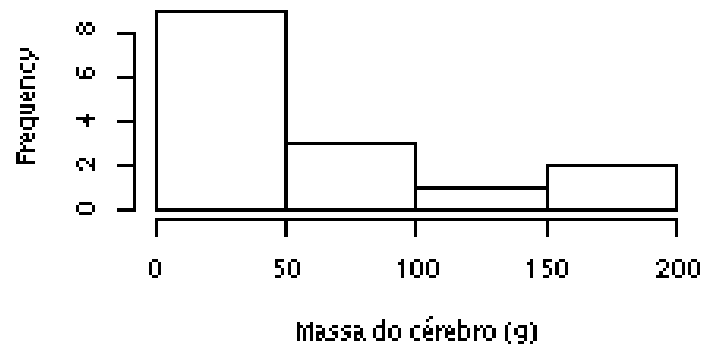
- Pequenos: até 2 kg
- Médios: entre 2 e 25 kg
- Grandes: acima de 25 kg

	tam	brain
Lesser short-tailed shrew	P	0.14
Little brown bat	P	0.25
Big brown bat	P	0.30
Mouse	P	0.40
Musk shrew	P	0.33
...		
Okapi	G	490.00
Cow	G	423.00
Horse	G	655.00
Giraffe	G	680.00
Asian elephant	G	4603.00
African elephant	G	5712.00

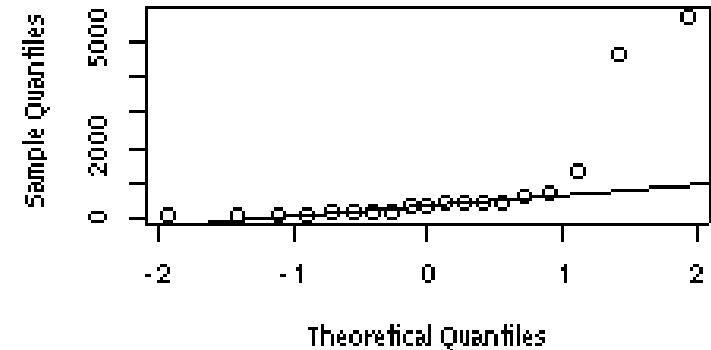
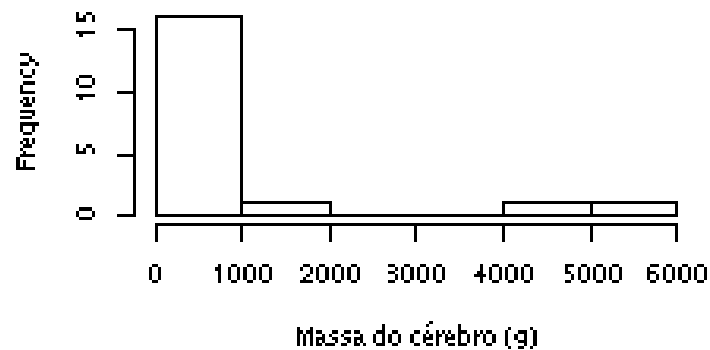
PEQUENOS



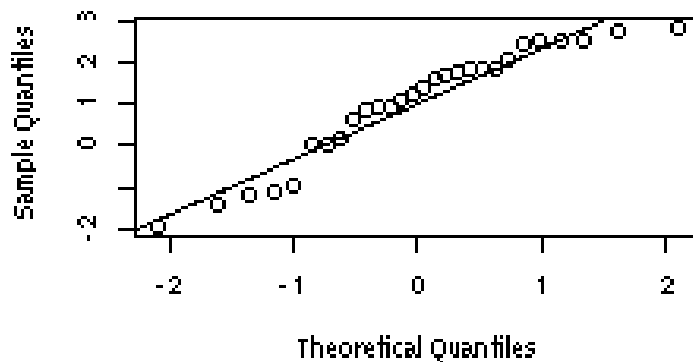
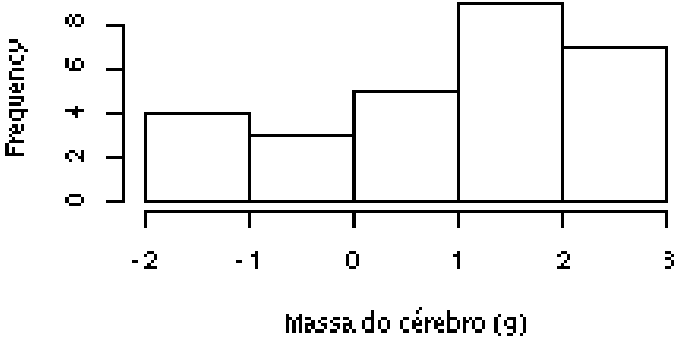
MÉDIOS



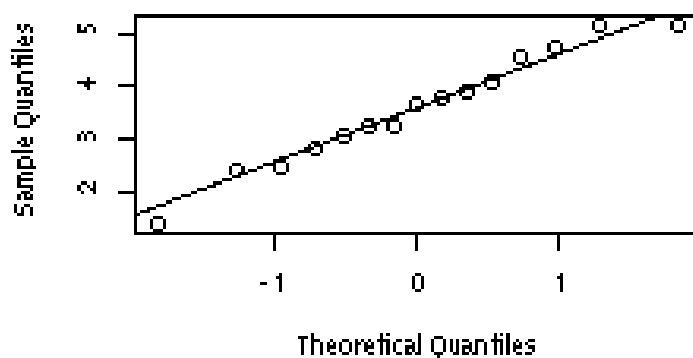
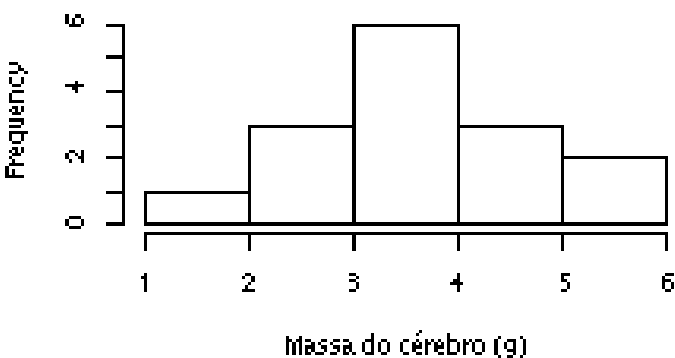
GRANDES



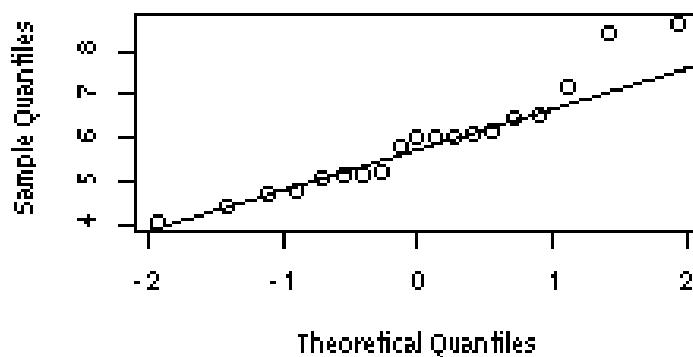
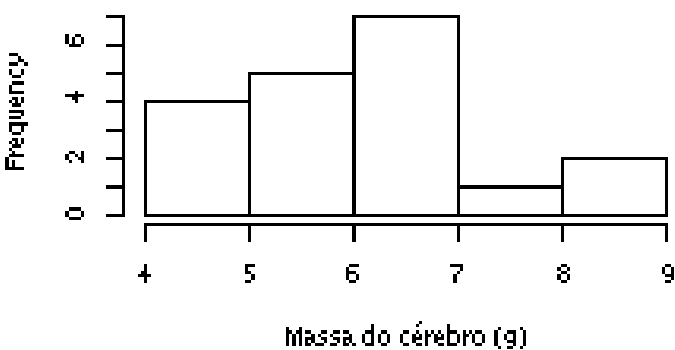
PEQUENOS



MÉDIOS



GRANDES



Transformação
logarítmica

ANOVA

```
-----  
              Df    Sum Sq Mean Sq F value    Pr(>F)  
-----  
tam              2  270.517  135.259   84.364 < 2.2e-16 ***  
Residuals       59   94.593    1.603  
-----  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testes *a posteriori*

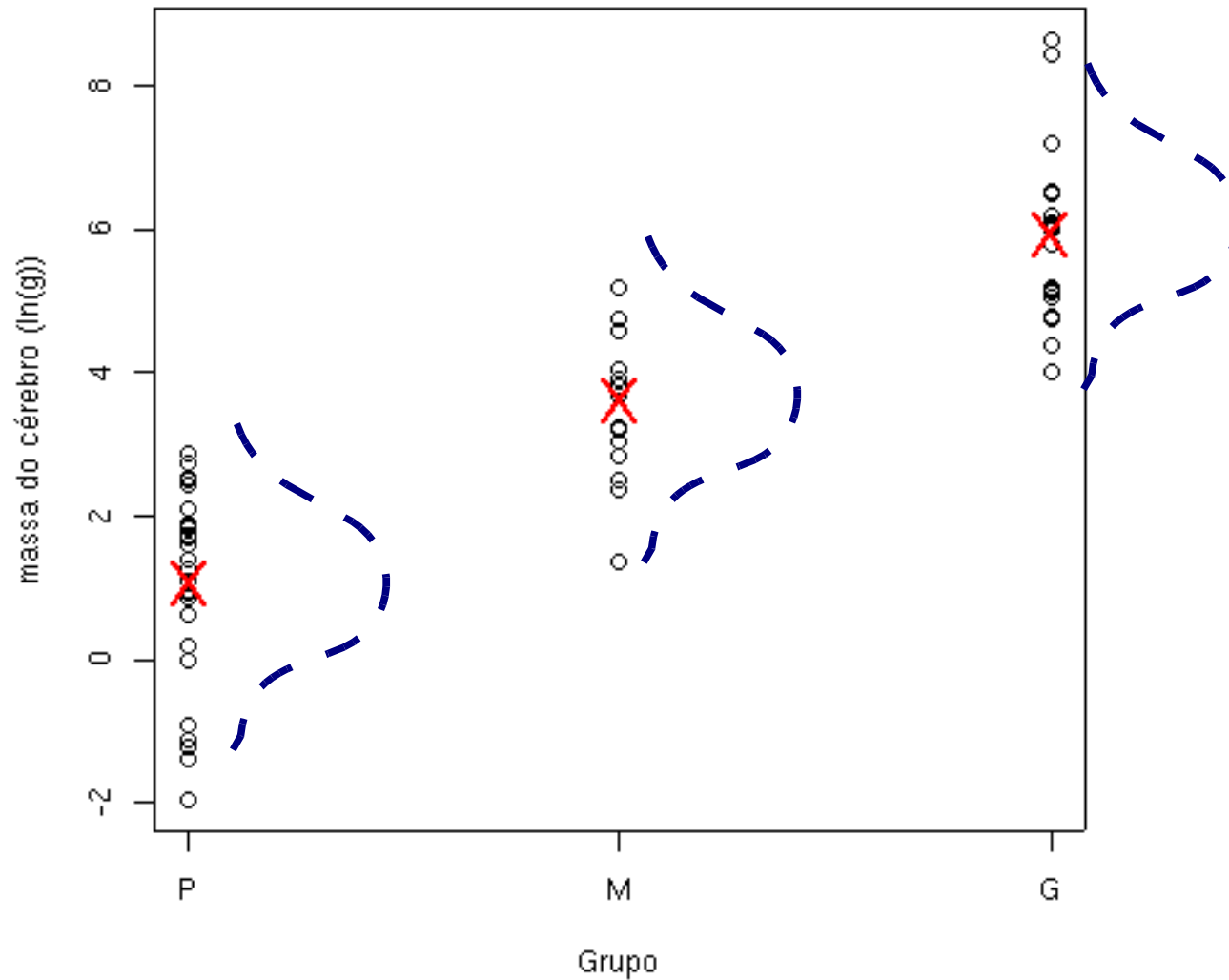
```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = log(brain) ~ tam, data = mamif)
```

```
$tam
```

	diff	lwr	upr	p adj
M-P	2.544666	1.570590	3.518742	1.0e-07
G-P	4.853677	3.948828	5.758527	0.0e+00
G-M	2.309011	1.257532	3.360489	5.8e-06

Qual o modelo implícito?



Explicitando a Hipótese Nula

$$H_0: \mu_p = \mu_m = \mu_g$$

OU: a massa do cérebro de mamíferos é uma variável aleatória com distribuição normal, cujas estimativas dos parâmetros são:

- Média da amostra = $3,14 \ln(g)$
- Desvio-padrão da amostra = $2,45 \ln(g)$

Explicitando (uma das) hipóteses alternativas

$$H1: \mu_p \neq \mu_m \neq \mu_g$$

OU: a massa do cérebro de mamíferos de cada classe de tamanho é uma variável aleatória com distribuições normais, cujos parâmetros média são diferentes, e o parâmetros desvio-padrão é o mesmo. Suas estimativas são:

- Médias de cada classe: $P = 1,04$ $M = 3,58$ $G = 5,89$
- Médias ponderadas dos desvios-padrão = 1,26

Qual a comparação implícita?

Modelo 1

$$Y \sim N\{\mu = 3.14, \sigma = 2.45\}$$

$$-LL = 142,9$$

Modelo 2

$$Y_p \sim N\{\mu = 1.04, \sigma = 1.26\}$$

$$Y_m \sim N\{\mu = 3.58, \sigma = 1.26\}$$

$$Y_g \sim N\{\mu = 5.89, \sigma = 1.26\}$$

$$-LL = 101,1$$

Um outro modelo

Modelo 3

$$Y_p \sim N\{\mu = 1.04, \sigma = 1.88\}$$

$$Y_m \sim N\{\mu = 3.58, \sigma = 1.19\}$$

$$Y_g \sim N\{\mu = 5.89, \sigma = 1.51\}$$

$$-LL = 100,5$$

Para que transformar?

Modelo 1

$$Y \sim \text{LN}\{\mu = 3.14, \sigma = 2.45\}$$

$$-LL = 337,6$$

Modelo 2

$$Y_p \sim \text{LN}\{\mu = 1.04, \sigma = 1.23\}$$

$$Y_m \sim \text{LN}\{\mu = 3.58, \sigma = 1.23\}$$

$$Y_g \sim \text{LN}\{\mu = 5.89, \sigma = 1.23\}$$

$$-LL = 295,8$$

Segundo Exemplo:

Variável explanatória contínua

	body	brain
Lesser short-tailed shrew	0.005	0.14
Little brown bat	0.010	0.25
Big brown bat	0.023	0.30
Mouse	0.023	0.40
Musk shrew	0.048	0.33
...		
Okapi	250.000	490.00
Cow	465.000	423.00
Horse	521.000	655.00
Giraffe	529.000	680.00
Asian elephant	2547.000	4603.00
African elephant	6654.000	5712.00

Regressão linear simples

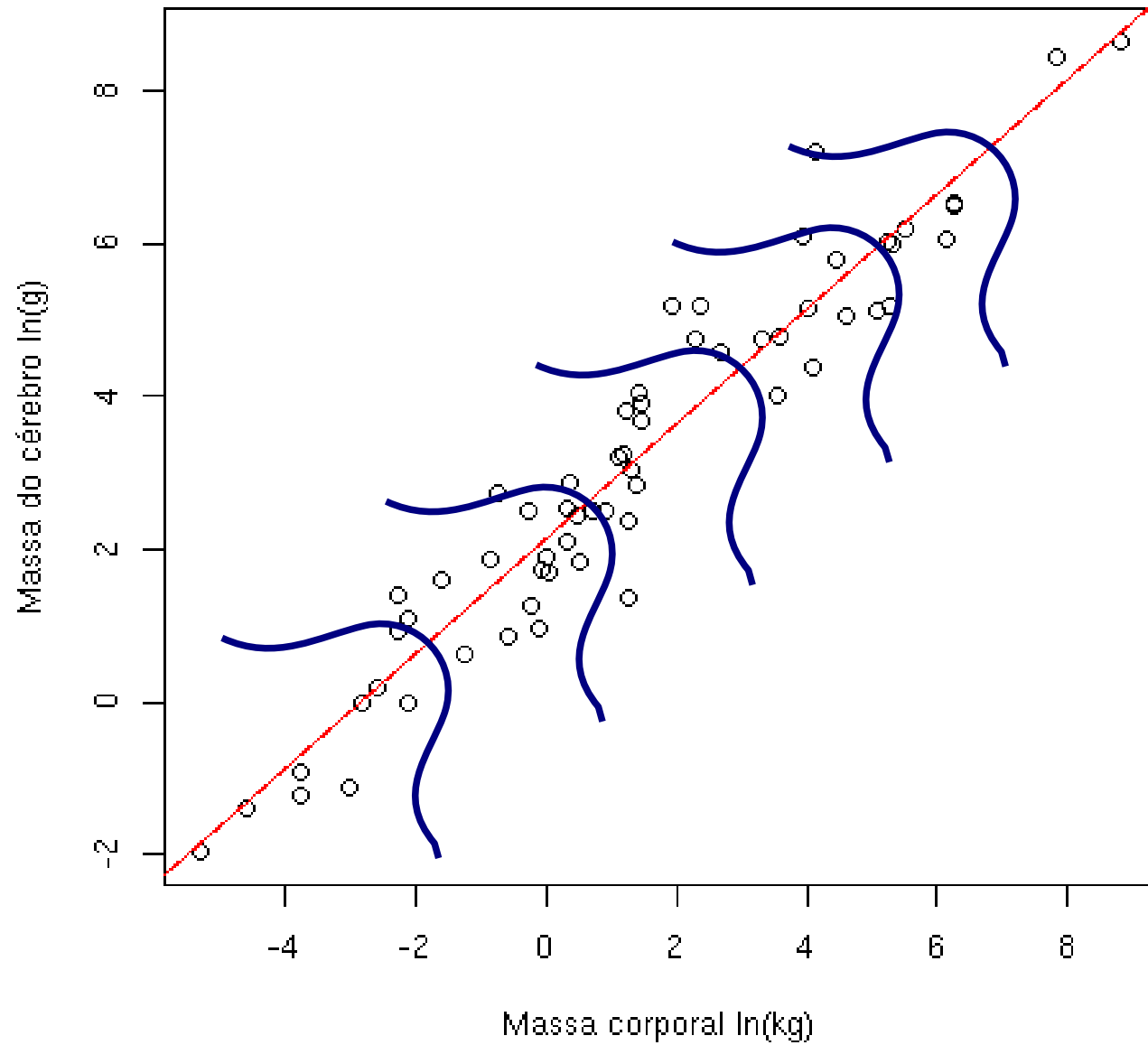
```
Call:
lm(formula = log(brain) ~ log(body), data = mamif)

Residuals:
    Min       1Q   Median       3Q      Max
-1.71550 -0.49228 -0.06162  0.43597  1.94829

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.13479    0.09604   22.23  <2e-16 ***
log(body)    0.75169    0.02846   26.41  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6943 on 60 degrees of freedom
Multiple R-Squared: 0.9208,    Adjusted R-squared: 0.9195
F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16
```

Qual o modelo implícito?



Qual a comparação implícita?

Modelo 4

$$Y \sim N\{\mu = a + b.\ln(\text{massa}), \sigma = c\}$$

Modelo 1

$$Y \sim N\{\mu = a, \sigma = b\}$$

Qual a comparação implícita?

Modelo 4

$$Y \sim N\{\mu = 2.14 + 0.75 \ln(\text{massa}), \sigma = 0.69\}$$

$$-LL = 64,3$$

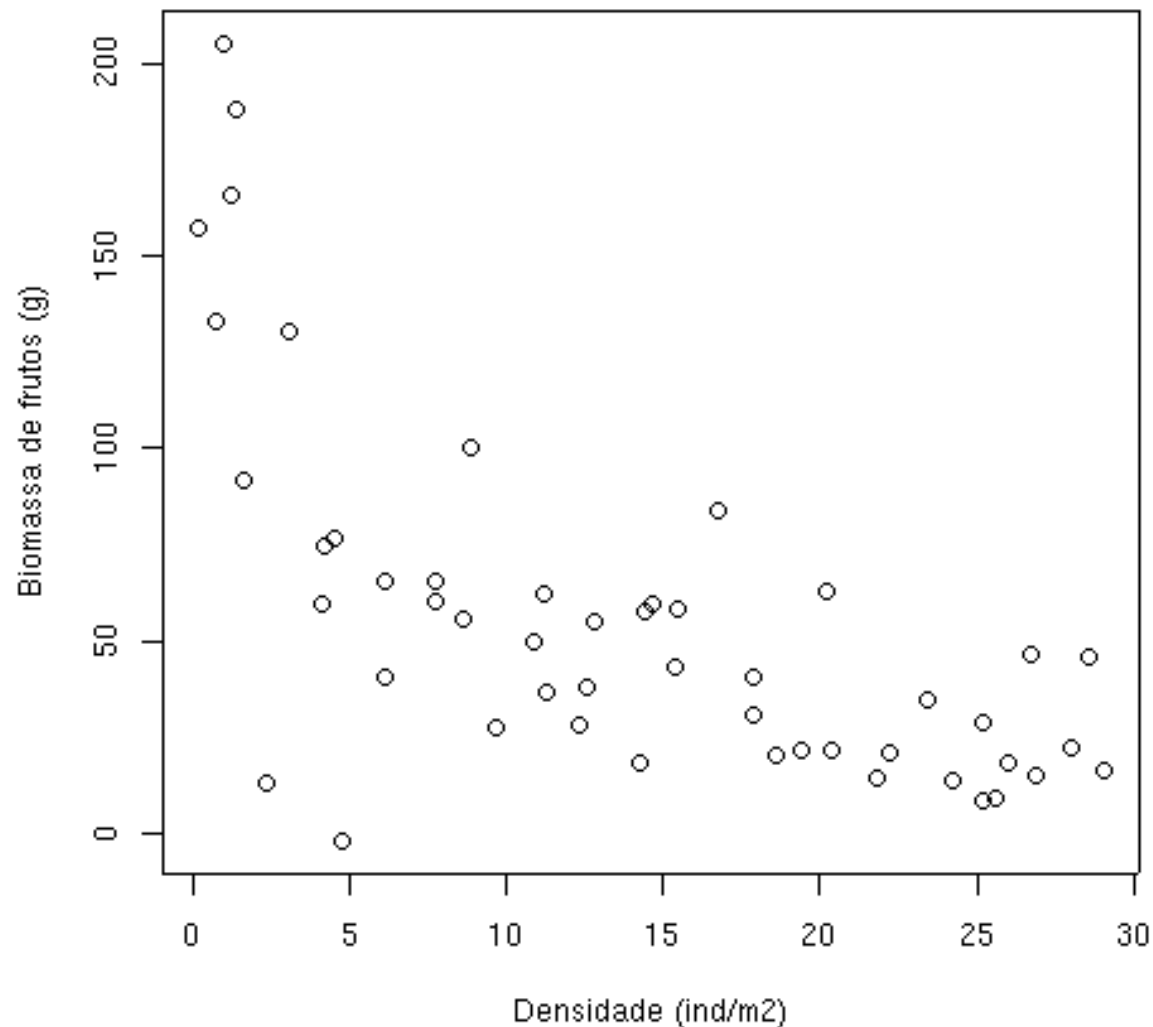
Modelo 1

$$Y \sim N\{\mu = 3.14, \sigma = 2.45\}$$

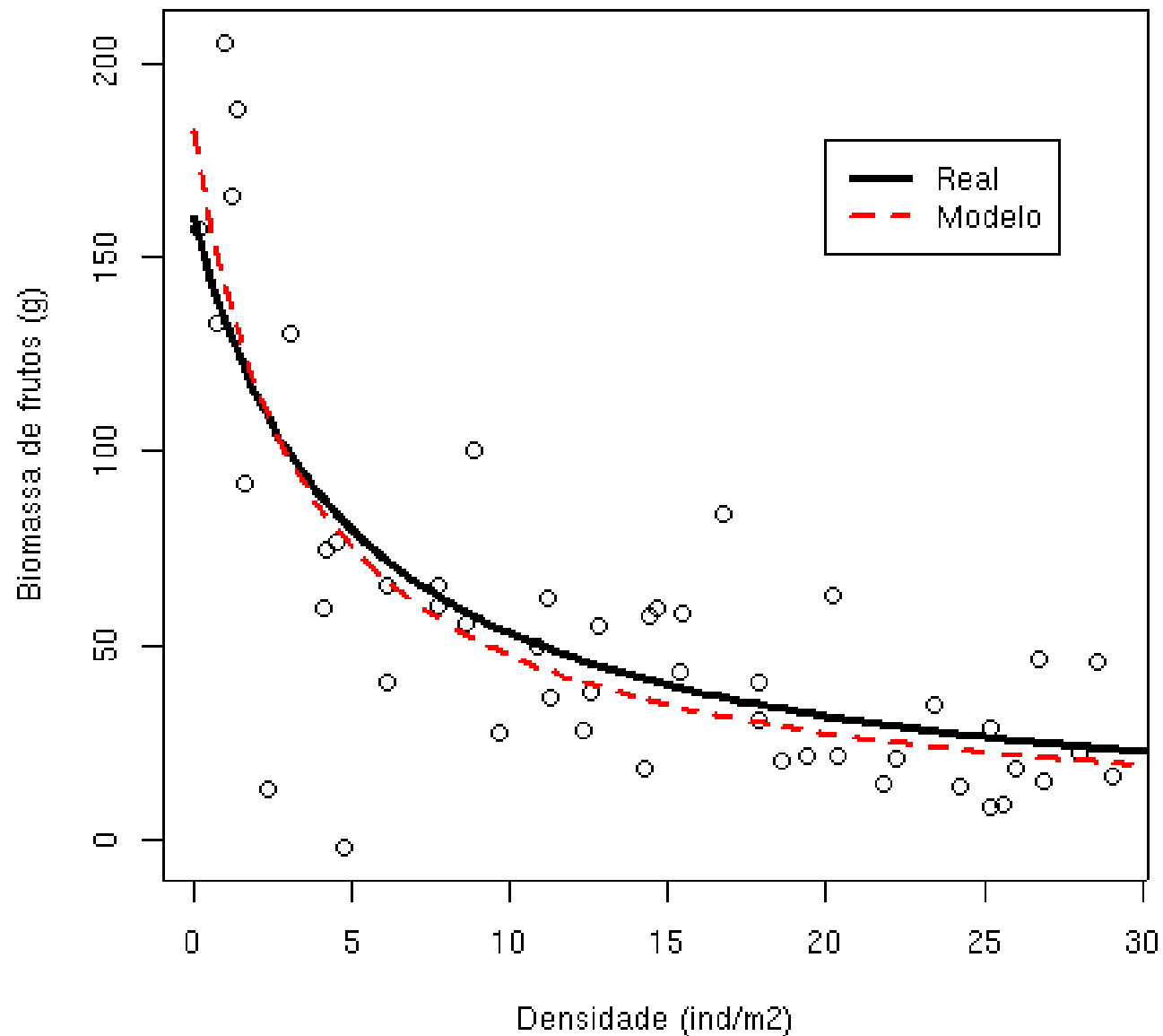
$$-LL = 142,9$$

Terceiro exemplo:

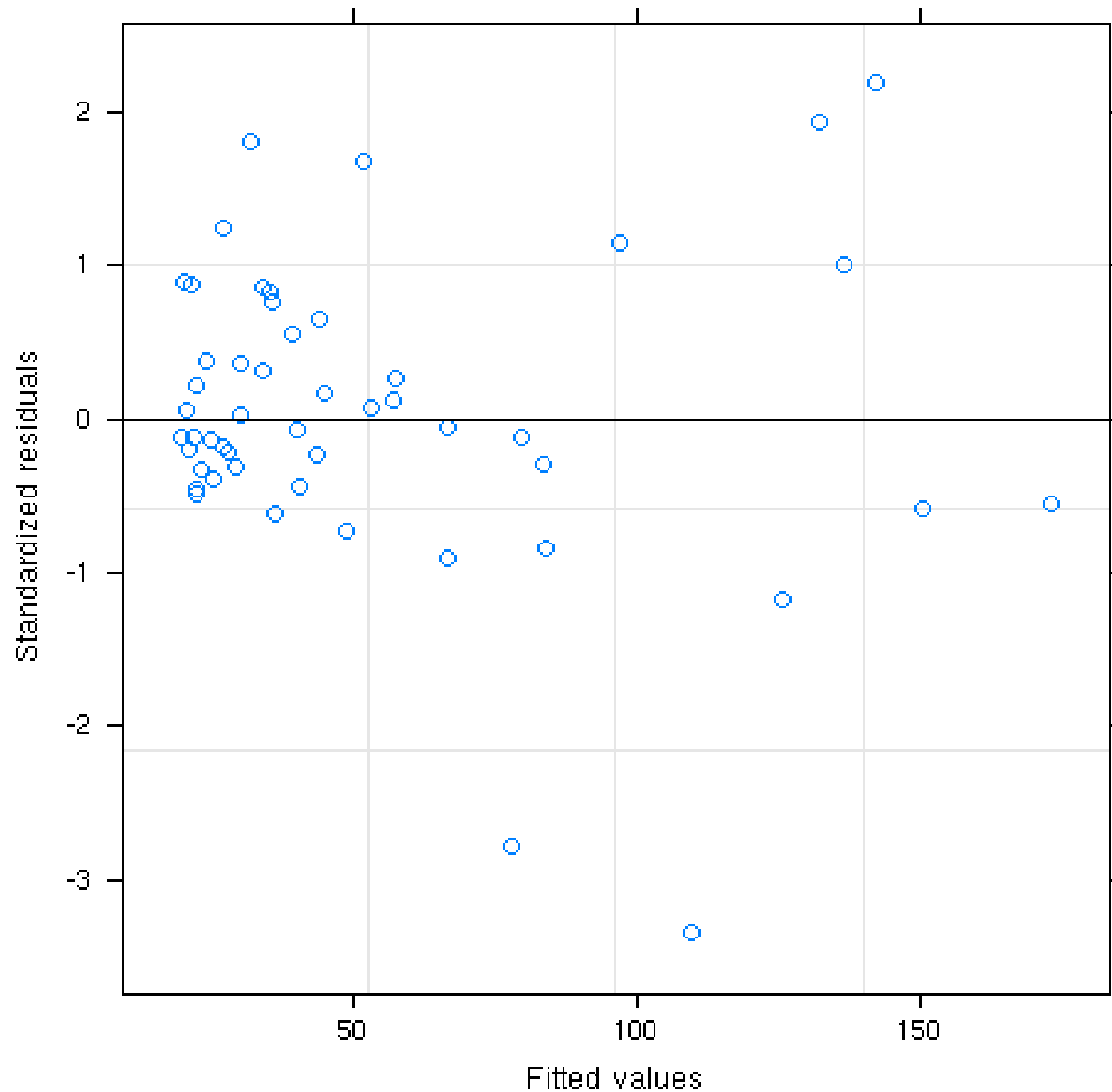
Relação não-linear, variância aumenta com a média



Regressão não-linear



Regressão não-linear: resíduos



Modelos concorrentes

Modelo 1

$$Y \sim N\{\mu = a / (\text{densidade} + b), \sigma = c\}$$

Modelo 2

$$Y \sim N\{\mu = a / (\text{densidade} + b), \sigma = \mu^c\}$$

Modelos concorrentes

Modelo 1

$$Y \sim N\{\mu = 645 / (\text{densidade} + 3.5), \sigma = 28.2\}$$

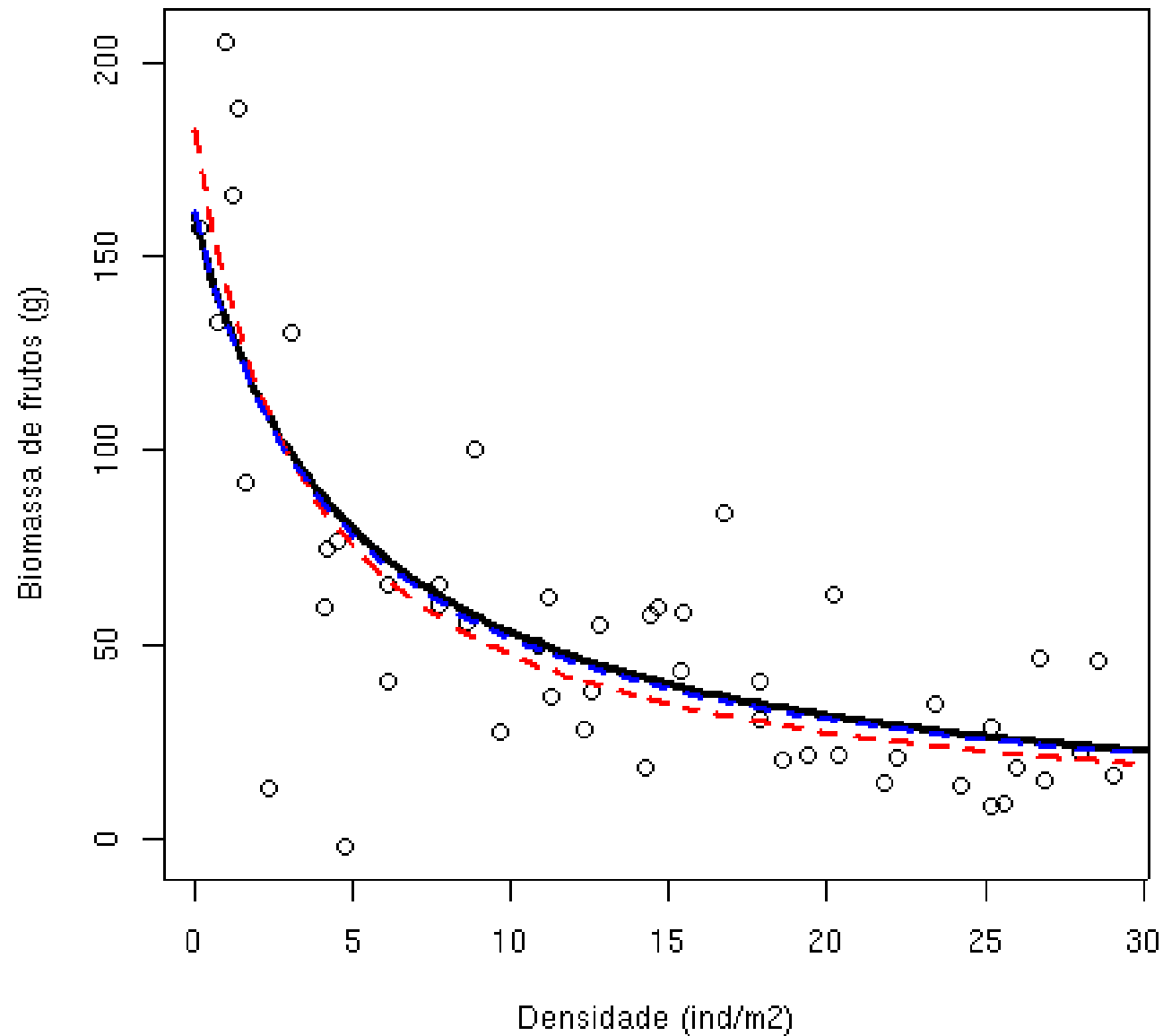
$$-LL = 238$$

Modelo 2

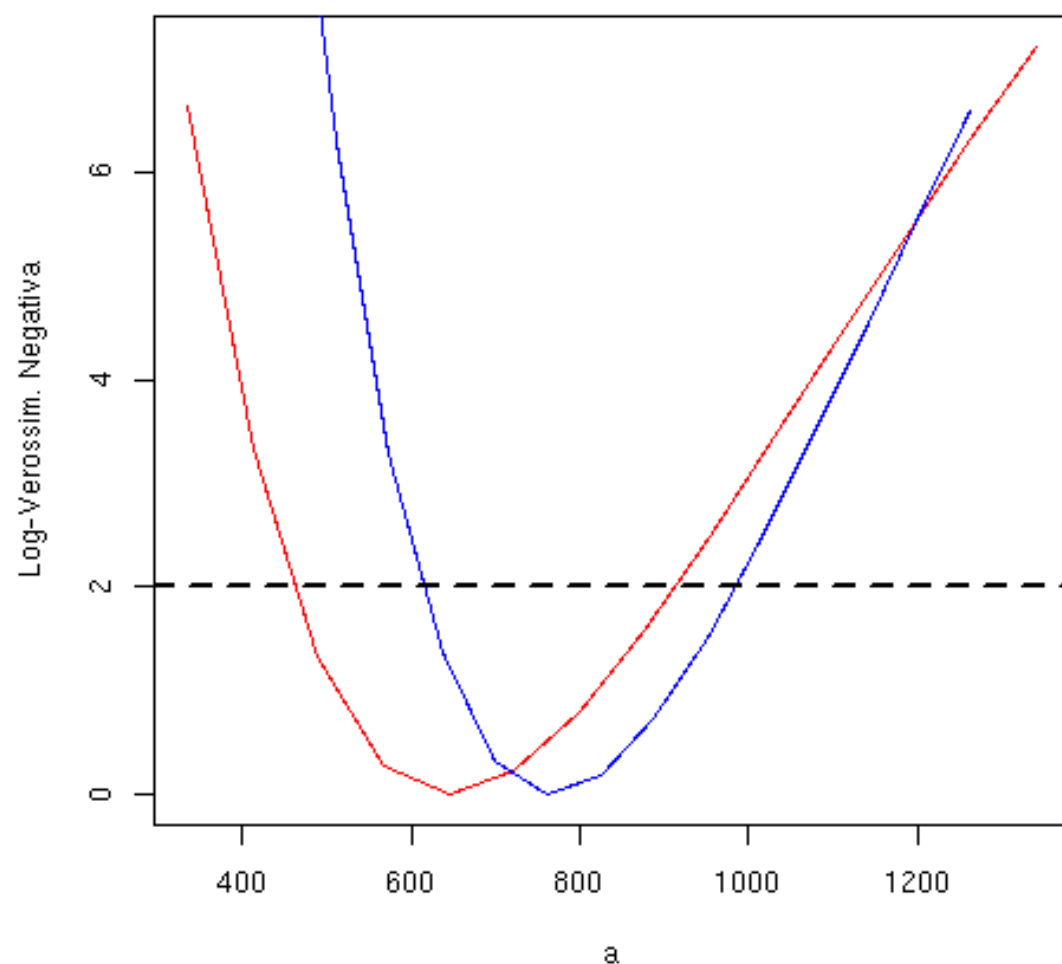
$$Y \sim N\{\mu = 762 / (\text{densidade} + 4.7), \sigma = \mu^{0.81}\}$$

$$-LL = 228$$

Esperados x “Real”



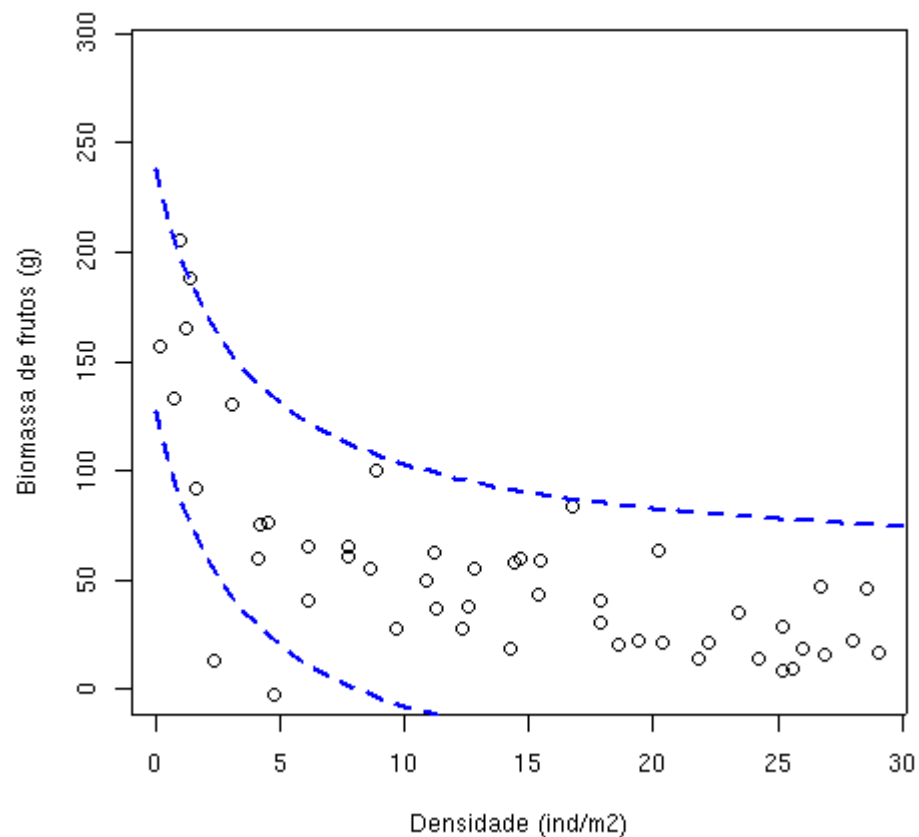
Precisão das estimativas



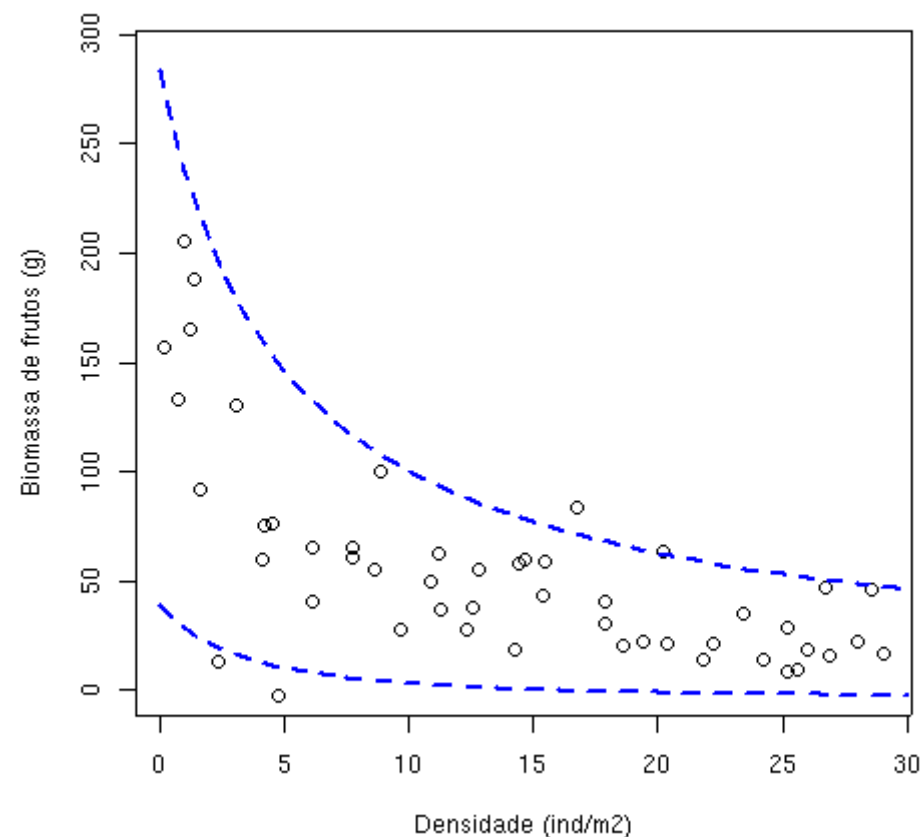
```
> confint(f2.prof)
              2.5 %      97.5 %
a      462.517044 906.581315
b        2.133277   5.963292
sigma  23.491827 34.821085
> confint(f3.prof)
              2.5 %      97.5 %
a  615.2357819 979.0660857
b    3.0791139   6.8890433
c    0.7623859   0.8765716
```


Intervalos dos valores

MODELO 1



MODELO 2



Modelos Probabilísticos: um Roteiro

- Defina a função de densidade probabilística
- Defina os parâmetros da função de densidade como funções das variáveis explanatórias
- Defina sua função de verossimilhança
- Encontre as estimativas de máxima verossimilhança de cada parâmetro
- Compare modelos concorrentes com a razão de verossimilhanças
- Faça um diagnóstico do modelo escolhido (valores esperado, resíduos, premissas, etc)

Implementação no R: Log-Verossimilhança Negativa

```
##Funcoes de max verosimilh
LLf1 <- function(a,b,sigma){
  media <- a/(densidade+b)
  -sum(dnorm(bfruto,mean=media,sd=sigma,log=T))
}

LLf2 <- function(a,b,c){
  media <- a/(densidade+b)
  -sum(dnorm(bfruto,mean=media,sd=media^c,log=T))
}
```

Implementação no R: Estimativa dos Parâmetros

Aproximação de valores iniciais

```
> frutos.1 <- nls(bfruto~a/(densidade+b),start=c(a=100,b=1))
> summary(frutos.1)
```

Formula: bfruto ~ a/(densidade + b)

Parameters:

	Estimate	Std. Error	t value	Pr(> t)	
a	644.9518	104.2357	6.187	1.30e-07	***
b	3.5402	0.8691	4.073	0.000173	***

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

Residual standard error: 28.82 on 48 degrees of freedom

Otimização numérica para encontrar os valores que maximizam a função de verossimilhança

```
> frutos.2 <- mle2(LLf1,start=list(a=645,b=3.5,sigma=28.8))
> frutos.3 <- mle2(LLf2,start=list(a=645,b=3.5,c=1))
```

Implementação no R: Diagnóstico do modelo

```
> summary(frutos.2)
Maximum likelihood estimation

Call:
mle2(minuslogl = LLf1, start = list(a = 645, b = 3.5, sigma = 28.8))

Coefficients:
              Estimate Std. Error z value      Pr(z)
a          644.99973    107.78816   5.9840 2.178e-09 ***
b           3.54062      0.90629   3.9067 9.356e-05 ***
sigma      28.23366      2.82328 10.0003 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-2 log L: 475.9473
```

```
> logLik(frutos.2)
'log Lik.' -237.9736 (df=3)
```

Implementação no R: Diagnóstico do modelo

```
> summary(frutos.3)
Maximum likelihood estimation

Call:
mle2(minuslogl = LLf2, start = list(a = 645, b = 3.5, c = 1))

Coefficients:
      Estimate Std. Error z value      Pr(z)
a 760.2656      86.9582   8.7429 < 2.2e-16 ***
b   4.7111       1.1415   4.1270 3.675e-05 ***
c   0.8136       0.0293  27.7681 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> logLik(frutos.3)
'log Lik.' -227.7474 (df=3)
```

Implementação no R: Diagnóstico do modelo

```
> f2.prof <- profile(frutos.2)
Houve 12 avisos (use warnings() para vê-los)
> f3.prof <- profile(frutos.3)
> confint(f2.prof)
              2.5 %              97.5 %
a      462.518013  906.580044
b         2.133454    5.958821
sigma  23.491819   34.821080
> confint(f3.prof)
              2.5 %              97.5 %
a 615.1296963  979.1876955
b   3.0684110    6.8719963
c   0.7624741    0.8768431
```