Medida de não lineariedade

Kelly Pereira de Lima

2019

Algumas propriedades assintóticas dos estimadores de mínimos quadrados

- Quanto mais próximo do linear for o comportamento de um modelo, mais precisos serão os resultados assintóticos.
- Avalia-se a extensão do comportamento não linear através das bem conhecidas medidas de nãolinearidade.

Vantagens

As vantagens em se buscar modelos com comportamento próximo do comportamento linear:

- as estimativas de mínimos quadrados podem ser facilmente obtidas;
- os estimadores são aproximadamente não viciados, normalmente distribuídos com variância mínima, mesmo em pequenas amostras;
- os valores de predições são mais precisos;
- os métodos iterativos convergem mais rapidamente;
- os estimadores têm propriedades similares às propriedades ótimas de modelos lineares.

Medidas de não lineriedade

- Os métodos para estudar o comportamento do ajuste de modelos de regressão não lineares são:
- Medidas de Curvatura de Bates e Watts
- Medida de Vício de Box
- Medida de assimetria de Hougaard

E como surgiu?

Medidas de não-linearidade

- Box (1971) propôs uma fórmula para estimar os vícios dos estimadores de mínimos quadrados de um modelo de regressão univariado.
- Gillis e Ratkowsky (1978) concluíram que a medida de vício de Box fornece uma boa indicação da extensão do comportamento não linear do modelo.
- Bates e Watts (1980) apresentaram nova medidas de não linearidade baseada no conceito geométrico de curvatura.
    - a não linearidade intrínseca (curvatura do espaço de estimação).
    - a não linearidade devida ao efeito de parâmetros.

Medidas de curvatura de Bates e Watts

- Bates e Watts provaram que a não linearidade de um modelo pode ser decomposta em duas componentes:
  - não linearidade intrínseca (IN)
  - não-linearidade devida ao efeito de parâmetros (PE).
- Os estimadores têm propriedades similares às propriedades ótimas de modelos lineares.
- A medida (PE) é uma quantidade escalar que representa o máximo valor do efeito da parametrização e aumenta à medida que o seu comportamento se afasta do comportamento linear. Além disso, mede a extensão do comportamento não-linear causado pela parametrização.

Medida de Vício de Box

Box (1971) propôs uma estatística para avaliar o vício dos estimadores de mínimos quadrados dos parâmetros de um modelo de regressão não-linear univariado, dada por:

$$Vicio(\hat{\theta}) = -\frac{\sigma^2}{2}\left[\sum_{i=1}^{n} F(\theta)F^{\top}(\theta)\right]^{-1} \sum_{i=1}^{n} F(\theta)tr$$

em que $F(\theta)_{p\times 1}$ é o vetor de primeiras derivadas de $f(X_i, \theta)$ e $H(\theta)_{p\times p}$ é a segunda derivadas com relação a cada elementos de θ.

Medida de Vício de Box

É comum expressar o valor da estimativa do vício em porcentagem: `

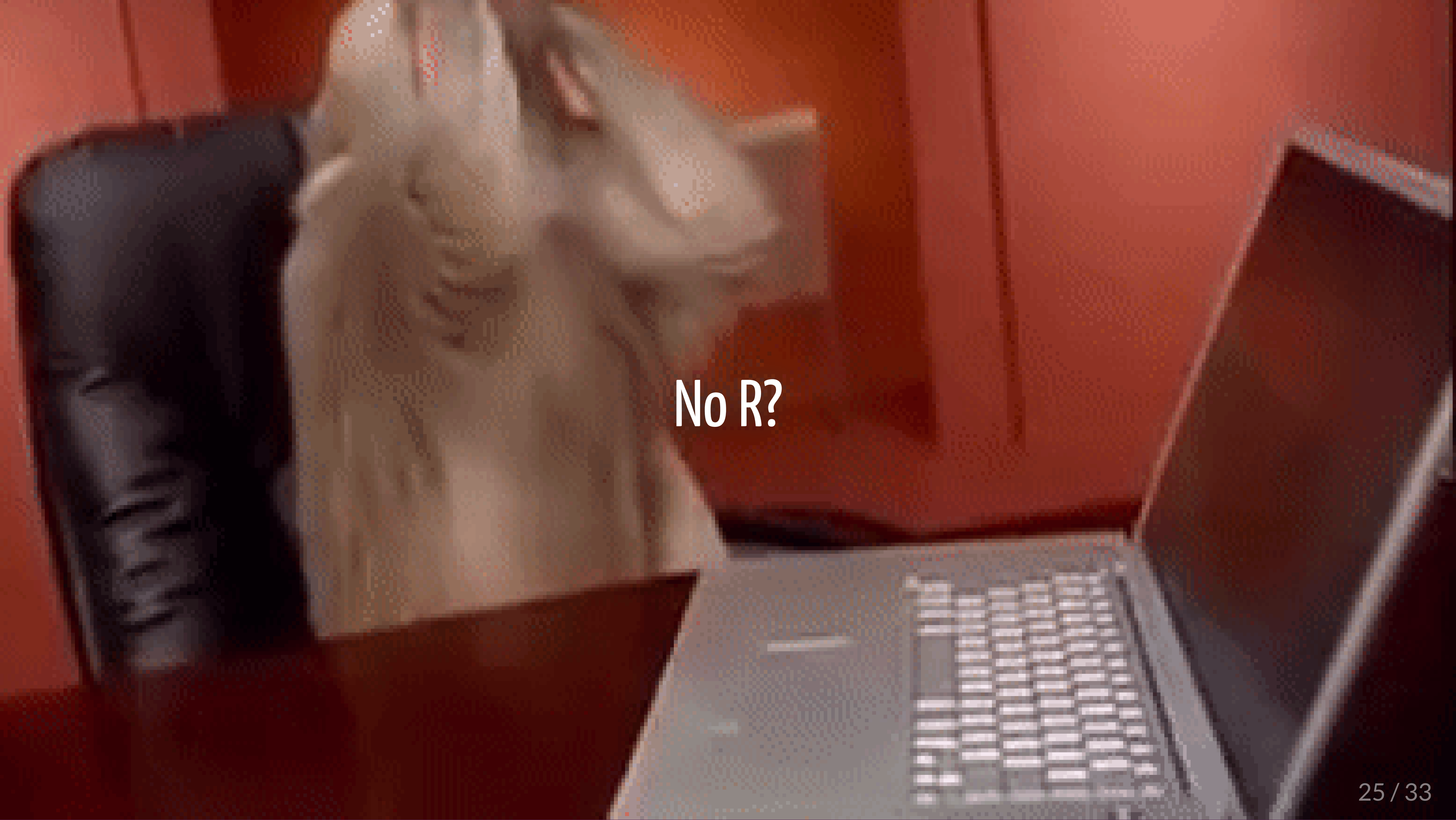
$$\%Vicio(\hat{\theta}) = \frac{100 * Vicio(\hat{\theta})}{\hat{\theta}}$$

` Considera-se vícios acima de 1%, em valor absoluto, como um indicador do comportamento não-linear do modelo.

Medida de assimetria de Hougaard

Para um estimador de parâmetro $i \ \ (g_i)$ fornece uma ideia da extensão do comportamento não linear desse estimador. Conforme Ratkowsky (1990), pode-se estabelecer uma regra de decisão:

- Para $g_i < 0,1$, o estimador do i-ésimo parâmetro tem comportamento muito próximo ao linear;
- Para $0,1 < g_i < 0,25$, o estimador é razoavelmente próximo ao linear;
- Para $g_i \geq 0,25$, a assimetria é muito aparente;
- Para $g_i > 1$, indica considerável comportamento não linear.



No R?

Medidas de curvatura

```r
rm(list=ls())
# Massa fresca do fruto
massa<- c(0.55,2.94,9.99,19.71,39.20,49.94,64.46,79.22,82.09,93.12,96.02,97.65)
daa <- c(40,60,80,100,120,140,160,180,200,220,240,260)

logistico <- deriv3(~ alpha/(1+ exp(k*(gamma-daa))),c("alpha", "gamma","k"), function(alpha,gamma,k,daa){NULL})
Log <- nls(massa~logistico(alpha,gamma,k,daa), star=list(alpha = 100,gamma=59, k=0.04))
Log
```

```
## Nonlinear regression model
##   model: massa ~ logistico(alpha, gamma, k, daa)
##    data: parent.frame()
##     alpha     gamma         k
##  97.84300 138.73454   0.03394
##  residual sum-of-squares: 82.42
##
## Number of iterations to convergence: 9
## Achieved convergence tolerance: 4.604e-06
```

Medidas de curvatura

```r
summary(Log)
```

```
##
## Formula: massa ~ logistico(alpha, gamma, k, daa)
##
## Parameters:
##        Estimate Std. Error t value Pr(>|t|)
## alpha 9.784e+01  2.349e+00   41.65 1.32e-11 ***
## gamma 1.387e+02  2.771e+00   50.07 2.54e-12 ***
## k     3.393e-02  2.551e-03   13.30 3.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.026 on 9 degrees of freedom
##
## Number of iterations to convergence: 9
## Achieved convergence tolerance: 4.604e-06
```

Medidas de curvatura

```r
require(MASS)
rms.curv(Log)
```

```
## Parameter effects: c^theta x sqrt(F) = 0.3944
##         Intrinsic: c^iota  x sqrt(F) = 0.1173
```

Vício de box

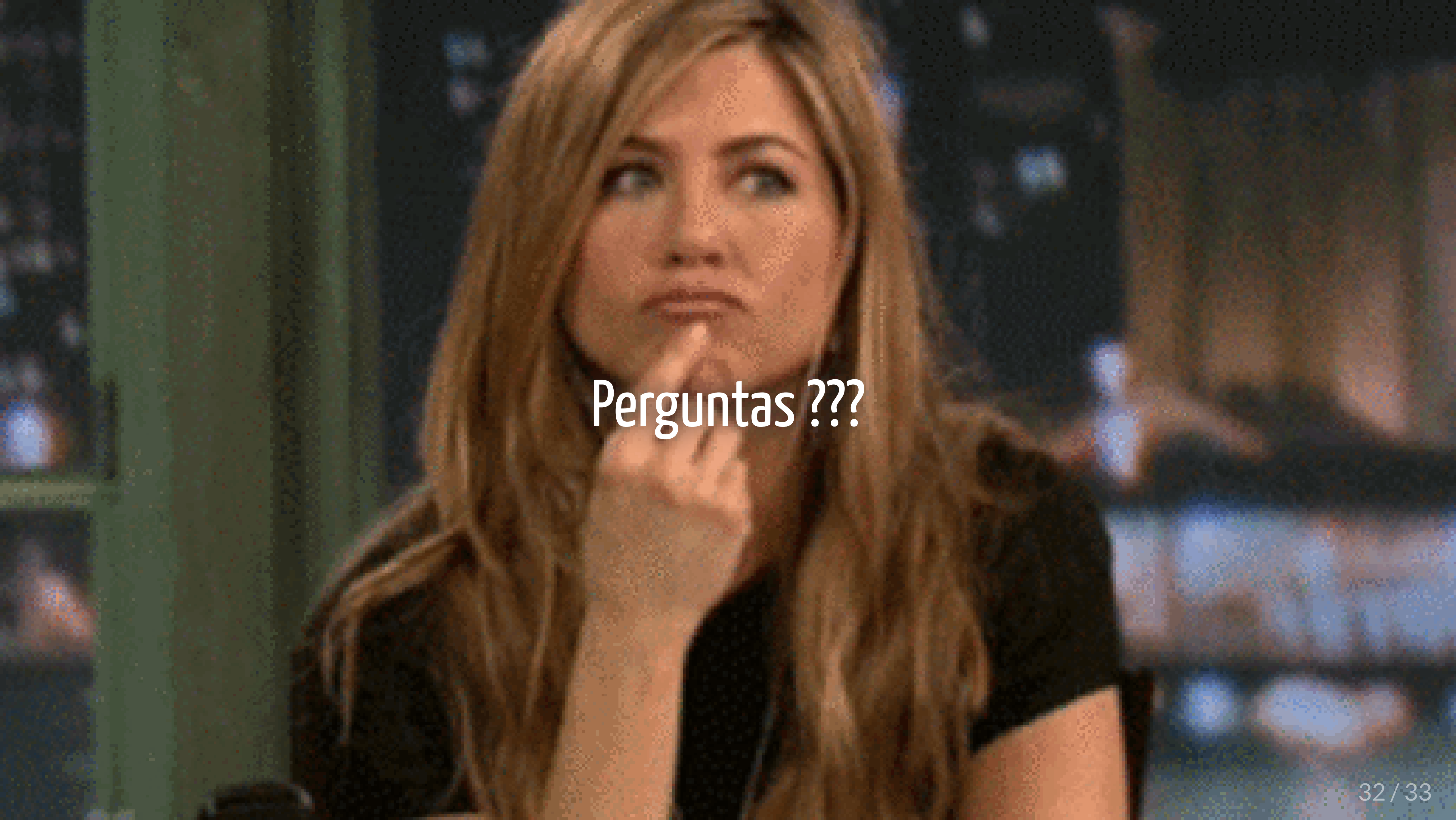
```
biasbox <- function(nls.obj){
theta <- summary(nls.obj)$coef[,1]
F <- attr(nls.obj$m$fitted(), "gradient")
H <- attr(nls.obj$m$fitted(), "hessian")
sig <- summary(nls.obj)$sigma
n <- dim(F)[1]
FlFi <- t(F)%*%F
d <- -(sig^2/2)*sapply(1:n, function(x){
sum(diag(solve(FlFi)%*%H[x, , ]))})
bias <- as.vector(solve(FlFi)%*%t(F)%*%d)
names(bias) <- names(coef(nls.obj))
bias.th <- 100*bias/theta
return(list("viés bruto"=bias,
"%viés(theta)"=bias.th))
}
```

Vício de box

```
biasbox(Log)
```

```
## $`viés bruto`
##        alpha        gamma            k
## 0.1131021070 0.1016682001 0.0001542466
##
## $`%viés(theta)`
##      alpha      gamma          k
## 0.11559550 0.07328254 0.45453416
```

Vamos para o R ???


Perguntas ???

Obrigada