



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Regressão de Cox Robusta

Autor: Paula Marques Sientchkovski
Orientador: Professor Dr. Álvaro Vigo

Porto Alegre, 16 de Julho de 2013.
Universidade Federal do Rio Grande do Sul

Instituto de Matemática
Departamento de Estatística

Regressão de Cox Robusta

Autor: Paula Marques Sientchkovski

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professor Álvaro Vigo
Professor Danilo Marcondes Filho

Porto Alegre, 16 de Julho de 2013.

Dedico este trabalho à Wladimir Ullrich que esteve presente em quase todos momentos desta caminhada.

Agradecimentos

Agradeço a minha família em geral, e muito especialmente aos meus filhos, os quais foram sacrificados com a minha ausência, mas me presentearam com seus sorrisos a cada retorno meu para casa.

Agradeço a minha mãe, Mair Marques Seckler, que até hoje é meu porto seguro e que, nos meus últimos anos de curso, incontáveis vezes deixou sua casa para cuidar da minha.

Agradeço a minha sogra, Liana Brasil Ullrich, por ter tornado possível a minha ida as aulas nos primeiros anos de curso e por ser, até hoje, um grande suporte.

Agradeço a minha amiga Natalia Bordin Barbieri, pela sua paciência, dedicação e amizade, a qual foi como um bálsamo nos meus momentos mais difíceis.

Agradeço ao Professor Álvaro Vigo, por ter aceitado ser meu orientador e por ter, mesmo diante dos seus inúmeros compromissos, mostrando-se sempre solícito a sanar as minhas dúvidas. Profissional que mesmo diante de todo seu conhecimento traz consigo uma humildade natural que o torna, além de um excelente professor, um exemplo de pessoa a ser seguido.

Agradeço ao Professor Danilo Marcondes Filho, também um exemplo de profissional e pessoa, por ter aceitado compor minha banca e contribuído de forma muito positiva no meu trabalho.

Agradeço ao Isaías Prestes, que mesmo sem me conhecer, dedicou seu tempo de forma surpreendente para me ajudar na finalização da parte computacional deste estudo.

Agradeço a Aléssio Farcomeni, autor do método “Trimmed”, apresentado neste trabalho, por ter respondido aos meus inúmeros e-mails com rapidez e gentileza.

Agradeço aos meus colegas de trabalho Enara Rosana Fagundes, José Carlos Lemos da Luz e Roberto Weber de Moura e Magda Rosane Silveira que, de uma forma muito especial “abraçaram a minha causa”, me apoiando e me incentivando durante todo tempo que comigo trabalharam.

Resumo

Neste artigo apresentamos alternativas ao Modelo de Riscos Proporcionais de Cox na análise de dados de sobrevivência, por meio do emprego de métodos robustos, com o objetivo de minimizar o impacto de *outliers*. Assim, analisamos brevemente três métodos robustos disponíveis, propostos por Bednarski (1993), Sasieni (1993) e Farcomeni e Viviani (2010, 2011), bem como descrevemos seus aspectos estatísticos e computacionais, utilizando o software estatístico R.

Palavras-chave: Regressão de Cox, *outliers*, análise de sobrevivência, métodos robustos.

Sumário

1 Introdução	8
2 Métodos.....	10
3 Discussões e Conclusões	18
Tabela 1 - Descrição das características na amostra, de acordo com a presença ou ausência de inflamação.....	19
Tabela 2 - Estimativas dos coeficientes de regressão (EP) e HR (IC 95%) utilizando diferentes métodos.....	19
Tabela 3 – Comparação entre as estimativas dos coeficientes de regressão e dos erros padrão.....	20
Referências Bibliográficas	21

Este artigo será submetido à Revista HCPA

1 Introdução

O Modelo de Riscos Proporcionais de Cox é um dos métodos mais utilizados na pesquisa clínica e epidemiológica para análise de dados de sobrevivência. Nestas situações o desfecho de interesse é definido como o tempo decorrido desde um tempo inicial até a ocorrência do evento de interesse, tais como a recidiva ou cura de uma doença ou óbito do paciente. Uma das características na análise dos dados de sobrevivência é a ocorrência de censura, pois nem sempre pode ser observado o tempo completo de sobrevivência. As principais causas de censura podem ser, por exemplo: término do estudo, óbito por causas não relacionadas ao evento em estudo, perda de contato, recusa do indivíduo em continuar participando, mudança de procedimento (tratamento) ou abandono devido a efeitos adversos de tratamento.

O modelo de Cox pode ser útil para descrever a relação entre o tempo de sobrevivência e um conjunto de covariáveis, na forma:

$$\lambda(t | \mathbf{x}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i)$$

em que $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ é um vetor que representa os preditores ou covariáveis, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ é o vetor de coeficientes de regressão, $\lambda_0(t)$ é uma função de risco de base que não depende dos preditores e $i = 1, 2, \dots, n$ identifica os indivíduos da amostra. A única suposição do modelo é de que os riscos são proporcionais, o que significa que a razão de riscos (*hazard ratio*) de ocorrência do evento entre indivíduos com padrão de covariáveis \mathbf{x}_i e \mathbf{x}_j ($i \neq j$) é constante para todo tempo t . Isto pode ser expressado como:

$$\frac{\lambda(t | \mathbf{x}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\lambda(t | \mathbf{x}_j) = \lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_j)} = \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\exp(\boldsymbol{\beta}' \mathbf{x}_j)}, \forall t > 0$$

Se o pressuposto de proporcionalidade está satisfeito, o modelo permite estimar os efeitos das covariáveis sem ter que se fazer qualquer suposição a respeito da distribuição do tempo de sobrevivência.

Os parâmetros do modelo são estimados pelo método de máxima verossimilhança parcial

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[\frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{j \in R_i} \exp(\boldsymbol{\beta}' \mathbf{S}_j)} \right]^{\delta_i} \quad (1)$$

em que R_i é o conjunto de indivíduos em risco imediatamente anterior ao i -ésimo tempo de ocorrência do evento t_i , $\delta_i = 1$ se o tempo observado está associado a ocorrência do evento e zero em caso contrário e $S_i = \sum_{j \in D_i} X_j$ e D_i é o conjunto de indivíduos censurados no tempo t_i . A equação (1) é uma aproximação para a função de verossimilhança, proposta por Breslow para incorporar empates dos tempos de sobrevida observados. Sob condições de regularidade, os estimadores de máxima verossimilhança parcial derivados da equação (1) são consistentes, com distribuição assintoticamente normal (Andersen e Gill, 1982 apud Farcomeni e Viviani, 2011)

Entretanto, o estimador de máxima verossimilhança é sensível a valores extremos (outliers), os quais também podem levar à violação da suposição de riscos proporcionais, os quais nem sempre podem ser percebidos com métodos usuais de verificação. (Farcomeni e Ventura, 2012) Neste contexto, *outliers* são interpretados como tempos de falha muito pequenos ou muito grandes, comparados com o tempo mediano predito pelo modelo. (Nardi e Schemper, 1999) Alguns autores utilizam uma definição mais ampla, como indivíduos cuja contribuição para a verossimilhança parcial é pequena quando comparada aos outros sujeitos. Isto inclui tempos de falha muito curtos, muito longos ou configurações de valores dos preditores e do tempo de sobrevida que são atípicos com respeito ao modelo ajustado (Farcomeni e Viviani, 2011).

Diversos métodos – como os apresentados por Sasieni (1993), Bednarski (1993) e Bednarski; Minder (1996) e Farcomeni e Viviani (2010, 2011) – tem sido utilizados para minimizar o impacto deste tipo de situação na estimativa dos parâmetros do modelo e, portanto, na subestimação ou superestimação da razão de riscos ou na violação do pressuposto de proporcionalidade dos riscos.

O objetivo deste trabalho é fazer uma breve revisão de métodos disponíveis e descrever seus aspectos estatísticos e computacionais. Com isto, pretende-se

facilitar aos usuários a identificação da necessidade de utilizar métodos robustos na análise de dados de sobrevivência por meio do modelo de Cox.

2 Métodos

Neste trabalho, vamos explorar a utilização de três modelos robustos para dados de sobrevivência, comparando os resultados por meio do modelo de Cox tradicional, o qual não incorpora estimadores robustos.

O primeiro método foi proposto por Bednarski (1993) e baseia-se numa dupla ponderação da verossimilhança parcial que modifica a estimativa da equação (1) sem alterar fundamentalmente a sua estrutura. Essa alternativa robusta é aprimorada por Bednarski (2007) a fim de torná-la adaptável e invariante à transformação do tempo. (Hertier, S. et al, 2009)

O segundo método, proposto por Sasieni (1993), apresenta uma abordagem baseada na maximização da verossimilhança parcial ponderada. Propõe-se a utilização de dois tipos de estimadores robustos para complementar o estimador de verossimilhança parcial. No primeiro tipo, ponderam-se as falhas por uma quantidade proporcional ao número de indivíduos ainda em risco (ou uma estimativa da função de sobrevivência), de modo análogo ao estimador de Wilcoxon. No outro tipo, é introduzida uma função de ponderação, objetivando a diminuição da tendência assintótica causada pelos outliers, que diminui com o aumento de censura artificialmente introduzida (Sasieni,1993).

O terceiro método é apresentado por Farcomeni e Viviani (2010, 2011), onde a abordagem para a estimativa da regressão de Cox robusta é feita por meio de um corte/aparamento dos outliers, definidos então como indivíduos cuja contribuição para a verossimilhança parcial é pequena quando comparada com outros sujeitos, ou seja, a ideia de cortar as menores contribuições para a verossimilhança é usada para guiar a estimação.

Esses métodos foram implementados no Software R e para este trabalho foi usada a versão 2.13.2. O R é um programa livre e maiores informações podem ser obtidas em (<http://www.r-project.org/>).

Após instalar e executar o R é necessário instalar as bibliotecas. Esse processo consiste em duas etapas:

Etapa 1. Instalar pacote:

```
install.packages(" nome do pacote", repos = "http://cran-r.c3sl.ufpr.br/")
```

Etapa 2. Carregar pacote:

```
library (nome do pacote)
```

Pacotes para Modelo de Regressão de Cox

O modelo de Regressão de Cox pode ser ajustado por meio da função ***coxph***, disponível no pacote ***survival***. A função ***coxph*** ajusta o modelo de regressão de riscos proporcionais de Cox. Outra função também contida no pacote ***survival*** é a ***cox.zph***, a qual testa a hipótese de riscos proporcionais de Cox para ajuste do modelo de regressão ***coxph***.

Pacotes para Regressão de Cox Robusta

O modelo de Regressão de Cox Robusto proposto por Bednarski pode ser ajustado através da função ***coxr***, disponível no pacote ***coxrobust***. Através dessa, podemos obter as estimativas dos betas, os erros padrões e as razões de riscos. O método não apresenta os intervalos de confiança para a razão de chances.

O modelo de Cox Robusto proposto por Sasieni pode ser ajustado através da função ***coxphw***, disponível no pacote ***coxphw***, através desta função podemos obter os intervalos de confiança, os erros padrões e a razão de chances.

O modelo de Cox Robusto proposto por Farcomeni e Viviani (2011) pode ser aplicado por meio da função disponível no endereço eletrônico <http://www.afarcome.altervista.org/robcox.r>.

Exemplo

Para exemplificar passo a passo a utilização destes métodos, vamos considerar dados de um estudo de corte hipotético realizado para estimar avaliar a associação entre a presença de inflamação sistêmica na linha de base e a incidência de diabetes tipo 2. A população em estudo são indivíduos com idade acima de 37 anos residentes em uma determinada comunidade. Uma amostra aleatória de 1670 foi selecionada e os indivíduos foram acompanhados ao longo do tempo. Eles passaram por visitas clínicas, aproximadamente a cada 2 anos. Aspectos dos métodos utilizados para o diagnóstico do diabetes e para a aferição da exposição, presença de inflamação sistêmica, bem como dos demais preditores, são pouco relevantes neste contexto. Entretanto, o embasamento clínico e epidemiológico da plausibilidade da associação pode ser encontrado na literatura. (DUNCAN, B. B. et al, 2003)

Quadro 1 – Descrição das variáveis utilizadas no exemplo hipotético.

Nome da Variável	Descrição	Valores ou unidade de medida
DM	Indicador de presença de diabetes tipo 2	0=Não; 1=Sim
TEMPODM	Tempo desde a entrada na coorte até o diagnóstico de diabetes tipo 2	dias
INFLAMACAO	Indicador de presença de inflamação sistêmica na linha de base	0=Não; 1=Sim
IDADE	Idade do participante na linha de base	anos
IMC	Índice de massa corporal na linha de base	kg/m ²
RCQ	Resultado da razão entre as medidas de circunferência da cintura (em cm) e do quadril (em cm), multiplicado por 20	-
GLICOSE	Glicemia de jejum na linha de base	mmol/L
LNINS	Logaritmo natural do nível de insulina na linha de base	log _e (mmol/L)
SEXOM	Indicador de sexo masculino	0=Feminino; 1=Masculino
RACACOR	Indicador de raça/cor	0=Outra, 1=Branca

A Tabela 1 mostra resultados da análise descritiva da amostra. O tempo mediano até o diagnóstico do diabetes foi 5600 dias, com mínimo de 152 dias e máximo de 6800 dias. O programa R será utilizado para ilustrar a utilização dos

métodos robustos e interpretação dos resultados. Para tanto, será considerado o modelo com desfecho tempo até o diagnóstico do diabetes tipo 2, a variável DM identifica a ocorrência do evento ou censura e inflamação sistêmica como preditor principal, ajustando pelas demais variáveis.

O ajuste do modelo de Cox tradicional (isto é, sem métodos robustos) pode ser ajustado pela função **coxph** do programa R como segue:

```
modelo_CN<-coxph(Surv(TEMPODM,DM)~ INFLAMACAO + SEXOM + RACACOR + LNINS +
GLICOSE + IMC + RCQ + IDADE ,data=dmficticio, x=TRUE)
```

As estimativas dos parâmetros de regressão, de razão de risco, o erro padrão e intervalo de confiança são mostrados nos quadros abaixo:

	coef	exp(coef)	se(coef)	z	Pr(> z)	
INFLAMACAO	0.88881	2.43224	0.15128	5.875	4.22e-09	***
SEXOM	-0.53048	0.58832	0.16913	-3.137	0.001709	**
RACACOR	-0.52960	0.58884	0.15447	-3.428	0.000607	***
LNINS	-0.03115	0.96933	0.12922	-0.241	0.809510	
GLICOSE	0.40360	1.49720	0.09365	4.310	1.64e-05	***
IMC	0.01039	1.01044	0.01282	0.810	0.417680	
RCQ	0.36079	1.43446	0.05667	6.367	1.93e-10	***
IDADE	0.04325	1.04420	0.01157	3.737	0.000186	***

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

	exp(coef)	exp(-coef)	lower ,95	upper ,95
INFLAMACAO	2,4322	0,4111	1,8082	3,2717
SEXOM	0,5883	1,6998	0,4223	0,8195
RACACOR	0,5888	1,6983	0,4350	0,7970
LNINS	0,9693	1,0316	0,7524	1,2487
GLICOSE	1,4972	0,6679	1,2461	1,7989
IMC	1,0104	0,9897	0,9854	1,0362
RCQ	1,4345	0,6971	1,2837	1,6030
IDADE	1,0442	0,9577	1,0208	1,0682

Concordance= 0,841 (se = 0,023)
Rsquare= 0,146 (max possible= 0,805)
Likelihood ratio test= 263,7 on 8 df, p=0
Wald test 276,2 on 8 df, p=0
Score (logrank) test = 311,7 on 8 df, p=0

O modelo de Cox robusto, utilizando o método proposto por Bednarski pode ser ajustado pela função **coxr**:

```
modelo_RB <-coxr(Surv(TEMPODM,DM)~ INFLAMACAO + SEXOM + RACACOR + LNINS +
GLICOSE + IMC + RCQ + IDADE, data=dmficticio, trunc=0.95)
```

As estimativas dos parâmetros de regressão, de razão de risco e erro padrão são mostradas nos quadro abaixo:

Robust estimator				
	coef	exp(coef)	se(coef)	p
INFLAMACAO	1,1581	3,184	0,1629	1,16e-12
SEXOM	-0,7070	0,493	0,2170	1,12e-03
RACACOR	-0,8047	0,447	0,1937	3,25e-05
LNINS	-0,0571	0,945	0,1254	6,49e-01
GLICOSE	0,5412	1,718	0,0800	1,34e-11
IMC	0,0210	1,021	0,0138	1,28e-01
RCQ	0,4155	1,515	0,0679	9,24e-10
IDADE	0,0667	1,069	0,0153	1,36e-05

Extended Wald test=278 on 8 df, p=0

O modelo de Cox robusto, utilizando o método proposto por Sasieni pode ser ajustado Pela função **coxphw**:

```
modelo_RS <-coxphw(data=dmficticio,Surv(TEMPODM,DM)~INFLAMACAO + SEXOM
+ RACACOR + LNINS + GLICOSE + IMC + RCQ + IDADE , trunc = 0.95)
```

As estimativas dos parâmetros de regressão, de razão de risco, o erro padrão e intervalo de confiança são mostradas nos quadros abaixo:

Model fitted by Weighted Estimation							
	coef	se(coef)	exp(coef)	lower 0,95	upper 0,95	z	p
INFLAMACAO	0,95189996	0,19890675	2,5906,271	1,7542649	3,8257328	4,7856595	1,70e-00
SEXOM	-0,4132299	0,20647912	0,6615102	0,441348	0,9914981	-2,0013157	4,54e-04
RACACOR	-0,4927224	0,18990917	0,6109608	0,4210778	0,8864706	-2,5945163	9,47e-03
LNINS	-0,0807788	0,16971456	0,9223977	0,6613889	1,2864103	-0,4759685	6,34e-05
GLICOSE	0,36764638	0,10011816	1,4443312	1,1869873	1,7574684	3,6721249	2,41e-02
IMC	0,0278785	0,01531482	1,0282707	0,9978643	1,0596037	1,8203603	6,87e-04
RCQ	0,40732195	0,07507627	1,5027878	1,2971573	1,7410158	5,4254423	5,78e-02
IDADE	0,04848489	0,01461377	1,0496795	1,0200406	1,0801796	3,317536	9,07e-02

Wald Chi-square = 195,2421 on 8 df, p = 0

O modelo de Cox robusto, utilizando o método proposto por Farcomeni e Viviani (2011) pode ser ajustado pela função descrita no quadro abaixo, que neste trabalho foi levemente alterada para que ao invés de apenas retornar as estimativas dos parâmetros, retornassem uma lista contendo também, a razão de risco, o erro padrão e o intervalo de confiança para o objeto do modelo da sugestão ótima sugerida. Cabe lembrar que para utilização deste método é necessário primeiramente executar o modelo de Cox tradicional mostrado neste trabalho :

```
fitter=function(t,delta,x,alpha=0.1,poscaso=NULL,D=NULL,kmax=NULL,rseqmax=NULL,n.m
ultistart=NULL){
n=length(t)
if(is.null(kmax)) {kmax=10000}
if(!is.matrix(x)) {x=matrix(x)}
if(is.null(rseqmax)) {rseqmax=20}
if(is.null(D)) {D=0.1}
if(is.null(poscaso)) {poscaso=sample(n,ceiling(n*(1-alpha)))}
if(is.null(n.multistart)) {n.multistart=10}
max.lik=-Inf
for(starts in 1:n.multistart) {
  if(starts>1) {
    poscaso=sample(n,ceiling(n*(1-alpha)))
  }

t.sub=t[poscaso]
  delta.sub=delta[poscaso]
  x.sub=x[poscaso,]
lt=length(poscaso)
  pos2=poscaso
  stime=coxph(Surv(t.sub,delta.sub)~.,data=data.frame(x.sub))
  beta=stime$coefficients
  lik=stime$loglik[2]
  likold=-Inf
  k=1
  bestlik=lik
  b2=beta
  p2=pos2
  rseq=0
```

continua...

...continuação

```
while(k<kmax && rseq<rseqmax) {
  if(lik-likold!=0) {rseq=0}
  if(lik-likold==0) {rseq=rseq+1}
  likold=lik
  for(position in 1:lt) {
    Tk=log(k+1,2)/(lt*D)
    k=k+1
    candidato=sample((1:n)[-pos2],1)
    aa=coxph(Surv(t[c(pos2[-position],candidato)],delta[c(pos2[-
      position],candidato)])~x[c(pos2[-position],candidato),])
    lik.cand=aa$loglik[2]
    pii=min(exp(Tk*(lik.cand-lik)),1)
    if(rbinom(1,1,pii)==1) {pos2=c(pos2[-position],candidato)
      likold=lik
      lik=lik.cand
      beta=aa$coefficients
      rsd=aa$residuals
      objmodel=aa

      if(lik>bestlik) {
        bestlik=lik
        b2=beta
        p2=pos2
        residuos=rsd
        objetodomodelo=objmodel
      }
    }
  }
  lik=bestlik
  beta=b2
  pos2=p2
  if(lik>max.lik)
    {max.lik=lik
      max.pos2=pos2
      max.beta=beta
    }
}
return(list(lik=max.lik,outliers=(1:n)[-
max.pos2],units=max.pos2,beta=max.beta,trim=alpha,residuos=residuos,objetodomodelo
=objetodomodelo))
}
aaa <- fitter( dmficticio$TEMPODM,dmficticio$DM,
modelo_CN$x,alpha=0.1,poscaso=NULL,D=NULL,kmax=10000,rseqmax=NULL,n.multistart
=NULL)
```

As estimativas dos parâmetros de regressão, de razão de risco, o erro padrão e intervalo de confiança são mostradas nos quadros abaixo:

	coef	exp(coef)	se(coef)	z	Pr(> z)	
INFLAMACAO	0,90588	2,4741	0,17704	5,117	3,11e-07	***
SEXOM	-0,44318	0,64199	0,19112	-2,319	0,0204	*
RACACOR	-0,57799	0,56102	0,17862	-3,236	0,00121	**
LNINS	-0,16362	0,84907	0,14704	-1,113	0,26583	
GLICOSE	0,48782	1,62876	0,10097	4,832	1,35e-06	***
IMC	0,01661	1,01675	0,0148	1,122	0,26175	
RCQ	0,39243	1,48057	0,06477	6,059	1,37e-09	***
IDADE	0,04285	1,04378	0,01344	3,189	0,00143	**

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

	exp(coef)	exp(-coef)	lower .95	upper .95
INFLAMACAO	2,4741	0,4042	1,7487	3,5004
SEXOM	0,642	1,5576	0,4414	0,9337
RACACOR	0,561	1,7825	0,3953	0,7962
LNINS	0,8491	1,1778	0,6365	1,1327
GLICOSE	1,6288	0,614	1,3363	1,9852
IMC	1,0168	0,9835	0,9877	1,0467
RCQ	1,4806	0,6754	1,3041	1,681
IDADE	1,0438	0,9581	1,0166	1,0716

A Tabela 2 resume os resultados obtidos pelos diferentes modelos, com ênfase nas estimativas dos coeficientes de regressão e das razões de riscos e correspondentes intervalos de 95% de confiança.

3 Discussão e Conclusões

Apresentamos três métodos alternativos para o ajuste do Modelo de Regressão de Cox tradicional, que podem ser úteis quando há potenciais violações das suposições desse modelo, em razão de relevante sensibilidade do mesmo aos outliers. No exemplo, o tempo até o diagnóstico do diabetes (casos) varia imensamente, de 152 até 6800 dias. Estes valores extremos no tempo, bem como valores extremos nos preditores, claramente influenciam as estimativas dos coeficientes de regressão. A Tabela 2 evidencia discrepâncias substanciais nas estimativas utilizando os métodos robustos em relação ao modelo de Cox não robusto. Por exemplo, em relação ao modelo não robusto, as estimativas do coeficiente de regressão da variável inflamação tiveram um aumento de 30,3% para o modelo de Bednarski, 7,1% para o modelo de Sasieni e 1,9% para o modelo de Farcomeni & Viviani. Os coeficientes dos demais preditores também sofreram alterações com tendência de estimativas mais elevadas pelo método de Bednarski. As estimativas dos erros padrões associadas aos coeficientes de regressão foram maiores em relação àquelas obtidas pelo modelo de Cox não robusto, com exceção das variáveis LNINS e GLICOSE, que no modelo de Bednarski tiveram diminuição de 3% e 15% respectivamente. Os aumentos nos erros padrão eram esperados, pois os métodos robustos podem inflacionar a variância, e esse é o preço que se paga pela robustez, como descrito por Farcomeni & Viviane (2011) e Viviani & Farcomeni (2010). É importante salientar que as diferenças observadas nas estimativas dos coeficientes de regressão diretamente influenciam as estimativas da razão de risco (HR) e, portanto, poderiam significar potenciais vieses nas inferências.

O modelo de Cox tem sido extensamente utilizado no contexto clínico e epidemiológico. O exemplo utilizado neste trabalho ilustra a importância de avaliar potenciais efeitos de observações extremas as quais podem influenciar significativamente as estimativas dos coeficientes de regressão. Os procedimentos usuais de diagnóstico do modelo geralmente não parecem ser suficientes para revelar este tipo de discrepância, e os procedimentos robustos do modelo de Cox explorados neste trabalho podem contribuir para a verificação da adequação do modelo estimado.

Tabela 1 - Descrição das características na amostra, de acordo com a presença ou ausência de inflamação.

Variável	Com Inflamação	Sem Inflamação
	(N=512)	(N=1158)
	Média (DP) ou n(%)	Média (DP) ou n(%)
Diabetes Tipo 2		
Sim	114 (22,3)	96 (8,3)
Não	398 (77,7)	1062 (91,71)
Sexo		
Masculino	153 (29,9)	457 (27,4)
Feminino	359 (70,1)	701 (60,5)
Raça/Cor		
Branca	271 (16,2)	544 (32,6)
Não branca	241 (14,4)	614 (36,8)
Idade (em anos)	56,1(6,51)	53,5(6,01)
IMC* (kg/m ²)	29,6 (6,52)	28,1 (5,43)
RCQ [†]	19,4 (1,44)	18,4 (1,34)
Glicemia de jejum (mmol/L)	6,2 (1,03)	5,8 (0,84)
LNINS(ln mmol/L) [‡]	2,8 (0,69)	2,4 (0,84)
Tempo até o diagnóstico (em dias) [§]	3273 (3072-5600)	-

* Índice de massa corporal

† Razão entre a circunferência da cintura e a circunferência do quadril

‡ Logaritmo natural do nível de insulina

§ Mediana (Q1-Q3) estimada pelo método da Kaplan-Meier; não estimável no grupo sem inflamação.

Abaixo segue a Figura1 com os *Box Plots* das variáveis TEMPODM, GLICOSE, IDADE, IMC, LNINS e RCQ. Para que fosse possível identificar quais são as

observações que aparecem como *outliers*, foi utilizada uma função de autoria de *Tal Galili* (disponível em <http://www.r-statistics.com/wp-content/uploads/2011/01/boxplot-with-outlier-label-r.txt>).

Figura1: *Box plot* das variáveis TEMPODM, GLICOSE, IDADE, IMC, LNINS e RCQ

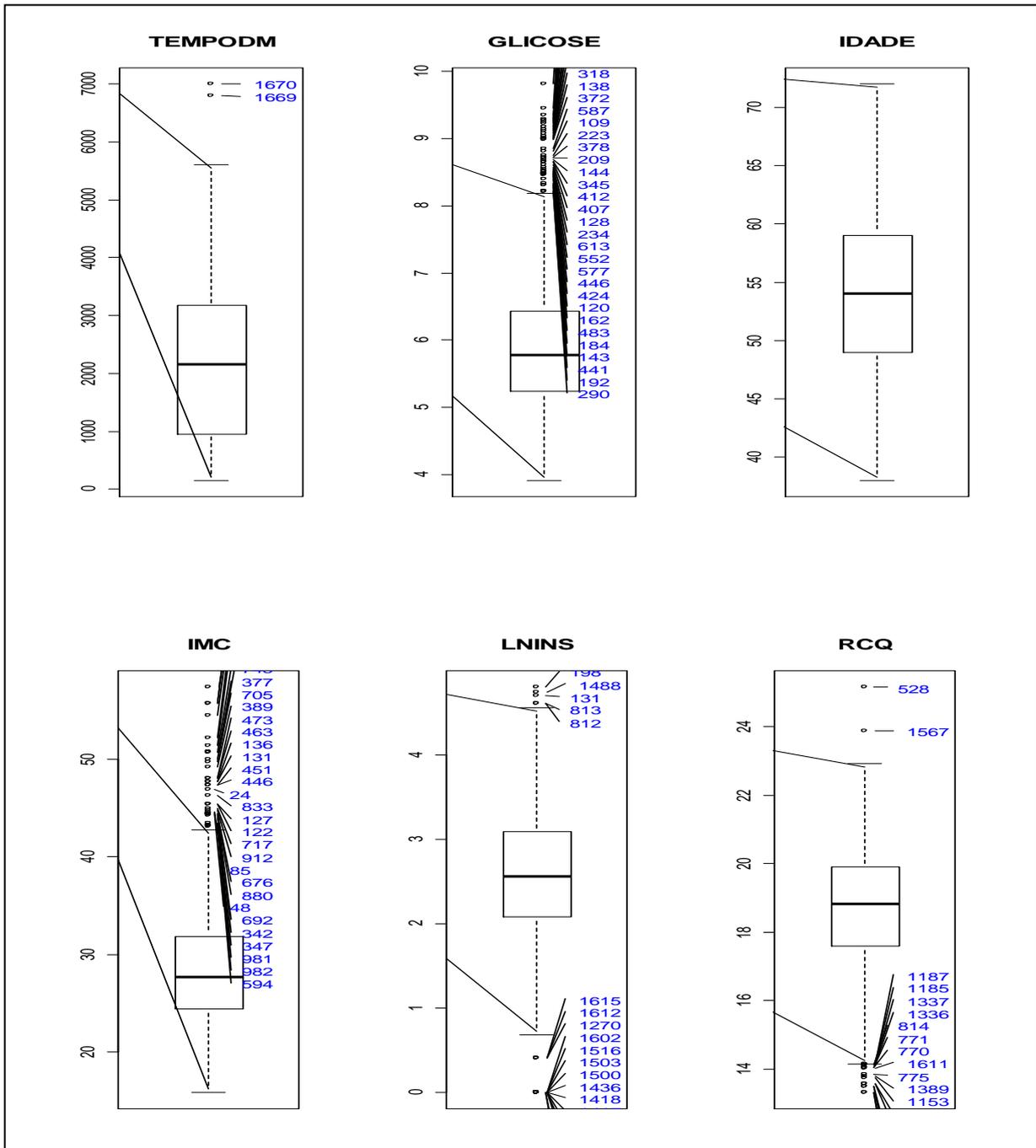


Tabela 2 - Estimativas dos coeficientes de regressão (EP) e HR (IC 95%) utilizando diferentes métodos.

Variável	<i>Modelo de Cox[†]</i>		<i>Bednarski</i>		<i>Sasieni</i>		<i>Trimmed</i>	
	<i>Coeficiente (EP)</i>	<i>HR (IC 95%)</i>	<i>Coeficiente (EP)</i>	<i>HR (IC 95%)</i>	<i>Coeficiente (EP)</i>	<i>HR (IC 95%)</i>	<i>Coeficiente(EP)</i>	<i>HR (IC 95%)</i>
Inflamação*	0,889 (0,151)	2,432 (1,808; 3,272)	1,158 (0,163)	3.184 (2,314; 4,381)	0,952 (0,199)	2,591 (1,754; 3,826)	0,906 (0,177)	2,474 (1,749; 3,500)
Sexo**	-0,530 (0,169)	0,588 (0,422; 0,820)	-0,707 (0,217)	0,493 (0,322; 0,755)	-0,413 (0,206)	0,662 (0,441; 0,991)	-0,443 (0,191)	0,642 (0,441; 0,934)
Raça/cor***	-0,530 (0,154)	0,589 (0,435; 0,797)	-0,805 (0,194)	0,447 (0,306; 0,654)	-0,493 (0,190)	0,611 (0,421; 0,886)	-0,578 (0,179)	0,561 (0,395; 0,796)
Idade (em anos)	0,043 (0,012)	1,044 (1,021; 1,068)	0,067 (0,015)	1.069 (1,037; 1,102)	0,048 (0,015)	1,050 (1,020; 1,080)	0,043 (0,013)	1,044 (1,017; 1,072)
IMC (kg/m ²)	0,010 (0,013)	1,010 (0,985; 1,036)	0,021 (0,014)	1.021 (0,994; 1,049)	0,028 (0,015)	1,028 (0,998; 1,060)	0,017 (0,015)	1,017 (0,988; 1,047)
RCQ	0,361 (0,057)	1,434 (1,284; 1,603)	0,416 (0,068)	1.515 (1,326; 1,731)	0,407 (0,075)	1,503 (1,297; 1,741)	0,392 (0,065)	1,481 (1,304; 1,681)
Glicose (mmol/L)	0,404 (0,094)	1,497 (1,246; 1,799)	0,541 (0,080)	1.718 (1,469; 2,010)	0,368 (0,100)	1,444 (1,187; 1,757)	0,488 (0,101)	1,629 (1,336; 1,985)
LNINS (log e mmol/L)	-0,031 (0,129)	0,969 (0,752; 1,249)	-0,057 (0,125)	0,945 (0,739; 1,208)	-0,081 (0,170)	0,922 (0,661; 1,286)	-0,164 (0,147)	0,849 (0,637; 1,133)

* Sim x Não

** (Masculino x Feminino)

*** Branca x Não branca

† Tradicional, sem usar métodos robustos

Tabela 3 – Comparação entre as estimativas dos coeficientes de regressão e dos erros padrão em relação às estimativas do modelo de Cox tradicional.

Preditor	Bednarski		Sasieni		Farcomeni	
	$\Delta\%$	Ef_r	$\Delta\%$	Ef_r	$\Delta\%$	Ef_r
Inflamação*	30,3	1,08	7,1	1,32	1,9	1,17
Sexo**	33,4	1,28	-22,1	1,22	-16,4	1,13
Raça/cor***	51,9	1,26	-7,0	1,23	9,1	1,16
Idade (em anos)	55,8	1,25	11,6	1,25	0,0	1,08
IMC (kg/m ²)	110,0	1,08	180,0	1,15	70,0	1,15
RCQ	15,2	1,19	12,7	1,32	8,6	1,14
Glicose (mmol/L)	33,9	0,85	-8,9	1,06	20,8	1,07
LNINS (log _e mmol/L)	83,9	0,97	161,3	1,32	429,0	1,14

$$\Delta\% = \frac{\hat{\beta}_{Modelo} - \hat{\beta}_{Cox}}{\hat{\beta}_{Cox}} \times 100$$

$$Ef_r = \text{Eficiência relativa} = \frac{\text{Erro Padrão}_{Modelo}}{\text{Erro Padrão}_{Cox}}$$

Referências Bibliográficas

ANDERSEN, P. K.; Gill, R. D. (1982): **Cox Regression Model for Counting Processes: a large sample study**. The Annals of Statistics, 10, 1100–1120.

BEDNARSKI, Tadeusz (1993): **Robust Estimation in Cox's Regression Model**. Scandinavian Journal of Statistics, Vol. 20, nº 3, 213-225

BEDNARSKI, Tadeusz (2007): **On a robust modification of Breslow's cumulated hazard estimator**. Computational Statistics & Data Analysis 52 (2007) pp. 234 – 238

BEDNARSKI, Tadeusz; MINDER, C. E. (1996): **A Robust Method for Proportional Hazards Regression**. Statistics in Medicine, Vol. 15, pp. 1033-1047

DUNCAN, B. B. et al (2003): **Low-grade systemic inflammation and the development of type 2 diabetes the atherosclerosis risk in communities study**. Diabetes.52 (7): 1799–805

FARCOMENI, Alessio; VENTURA, Laura (2012): **An overview of robust methods in medical research**. Statistical Methods in Medical Research, 21(2), 111–133.

FARCOMENI, Alessio; VIVIANI, Sara (2011): **Robust Estimation for the Cox Regression Model Based on Trimming**. Biometrical Journal, Volume 53, Issue 6, 956–973.

VIVIANI, Sara; FARCOMENI, Alessio (2010): **Trimmed Cox regression for robust estimation in survival studies**. Dipartimento di Scienze Statistiche - Sapienza Università di Roma.

HERITIER, Stephane et al (2009): **Robust Methods in Biostatistics**. John Wiley and Sons.

NARDI, Alessandra; SCHEMPER, Michael (1999): **New Residuals for Cox Regression and Their Application to Outlier Screening**. Biometrics, 55, 523-529

SASIENI, Peter (1993): **Maximum Weighted Partial Likelihood Estimators for the Cox Model**. Journal of the American Statistical Association, Vol. 88, nº 421, 144-152.