

Análise de Dados - UFPE/2019 - Lista 10

Maria Eduarda R. N. Lessa

25 de junho de 2019

Definir diretório:

```
setwd("C:/Users/Duda/Desktop/PPGCP/Análise de Dados/lista_10")
```

Questão 1:

```
# Instalar e requerer pacotes necessários:
# install.packages("MASS")
require(MASS)
# install.packages("fields")
require(fields)
require(car)

# Carregar bases de dados e atribuí-las a objetos:
banco <-
  read.csv("C:/Users/Duda/Desktop/PPGCP/Análise de Dados/lista_10/simula.correto.csv",
           header=T)

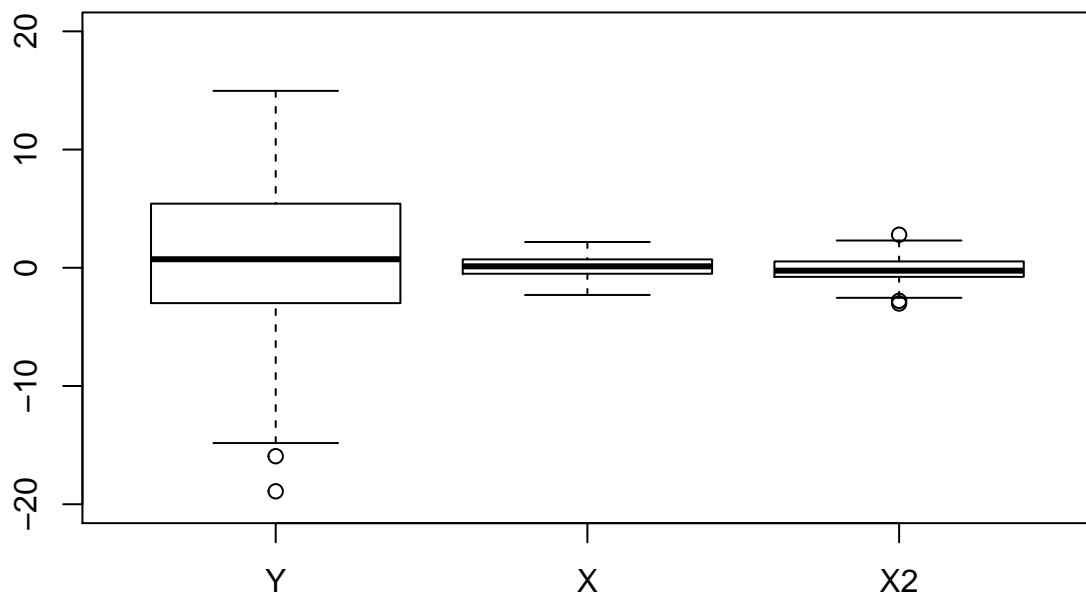
banco2 <-
  read.csv("C:/Users/Duda/Desktop/PPGCP/Análise de Dados/lista_10/simula.errado.csv",
           header=T)

attach(banco)

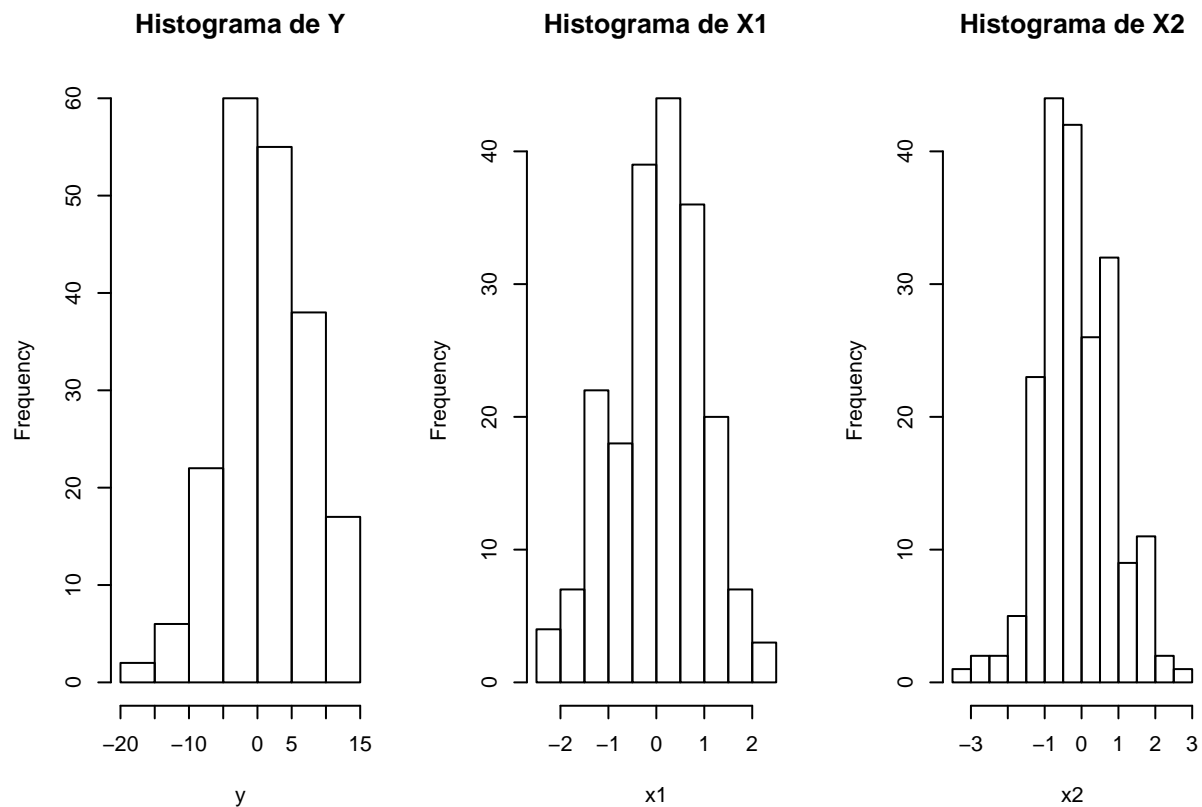
# Gerar tabela com análise descritiva das variáveis:
variav.descrit <- data.frame(y, x1, x2, x3)
t(stats(variav.descrit))

##      N      mean Std.Dev.      min      Q1      median      Q3
## y  200  0.86397597 6.2272734 -18.905095 -2.9469167  0.7242765  5.4142191
## x1 200  0.05653829 0.9165942  -2.299111 -0.5029032  0.1245492  0.7024212
## x2 200 -0.13319370 0.9774098  -3.020530 -0.7653276 -0.2416209  0.5365086
## x3 200  0.52500000 0.5006277   0.000000  0.0000000  1.0000000  1.0000000
##      max missing values
## y  14.965617          0
## x1  2.181491          0
## x2  2.798559          0
## x3  1.000000          0

# Gerar box-plots para analisar distribuição de y, x1 e x2:
par(mfrow=c(1,1))
boxplot(y, x1, x2, names=c('Y','X','X2'), ylim=c(-20,20))
```



```
# Gerar histogramas para analisar distribuição de y, x1 e x2:  
par(mfrow=c(1,3))  
hist(y, main="Histograma de Y")  
hist(x1, main="Histograma de X1")  
hist(x2, main="Histograma de X2")
```



```
# Gerar modelos de regressão de mínimos quadrados:
modelo.1 <- lm(y ~ x1)
modelo.2 <- lm(y ~ x1 + x2)
modelo.3 <- lm(y ~ x2 + x3)
modelo.4 <- lm(y ~ x1 + x3)
modelo.f <- lm(y ~ x1 + x2 + x3)

# Gerar tabela da análise de variância (ANOVA) para comparar modelo estimado
# com o modelo nulo:
tab.ANOVA <- function(modelo, gln, gld) {
  SSreg <- sum((y - mean(y))^2)
  print(paste('SSreg:'))
  print(SSreg)
  RSS <- sum(modelo$res^2)
  print(paste('RSS:'))
  print(RSS)
  F.teste <- ((SSreg - RSS) / (gln) / (RSS / gld))
  print(paste('F:'))
  print(F.teste)
  print(paste('P-valor:'))
  1 - pf(F.teste, 3, 196)
}

tab.ANOVA(modelo.f, 3, 196)
```

```
## [1] "SSreg:"
```

```
## [1] 7717.008
## [1] "RSS:"
## [1] 5005.588
## [1] "F:"
## [1] 35.38967
## [1] "P-valor:"
```

```
## [1] 0
```

```
# Analisar resultados das regressões:
summary(modelo.1)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.117  -3.330   0.240   3.814  15.066
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6488     0.3664   1.771  0.0781 .
## x1            3.8063     0.3999   9.517 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.171 on 198 degrees of freedom
## Multiple R-squared:  0.3139, Adjusted R-squared:  0.3104
## F-statistic: 90.58 on 1 and 198 DF,  p-value: < 2.2e-16
```

```
summary(modelo.2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.2107  -3.3899   0.2805   3.4299  14.9529
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.7266     0.3695   1.966  0.0507 .
## x1            3.7007     0.4058   9.120 <2e-16 ***
## x2            0.5396     0.3806   1.418  0.1578
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.158 on 197 degrees of freedom
## Multiple R-squared:  0.3208, Adjusted R-squared:  0.3139
## F-statistic: 46.53 on 2 and 197 DF,  p-value: < 2.2e-16
```

```
summary(modelo.3)
```

```
##
## Call:
## lm(formula = y ~ x2 + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.4409  -3.9256  -0.0877   4.5997  15.3132
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.6333     0.6323   2.583  0.01052 *
## x2            1.1930     0.4443   2.685  0.00787 **
## x3           -1.1626     0.8674  -1.340  0.18167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.123 on 197 degrees of freedom
## Multiple R-squared:  0.04282, Adjusted R-squared:  0.0331
## F-statistic: 4.406 on 2 and 197 DF, p-value: 0.01343
```

```
summary(modelo.4)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.2918  -3.4481  -0.1911   3.4822  16.1835
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.7928     0.5211   3.440 0.000709 ***
## x1            3.9875     0.3965  10.058 < 2e-16 ***
## x3           -2.1987     0.7259  -3.029 0.002782 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.068 on 197 degrees of freedom
## Multiple R-squared:  0.3444, Adjusted R-squared:  0.3378
## F-statistic: 51.75 on 2 and 197 DF, p-value: < 2.2e-16
```

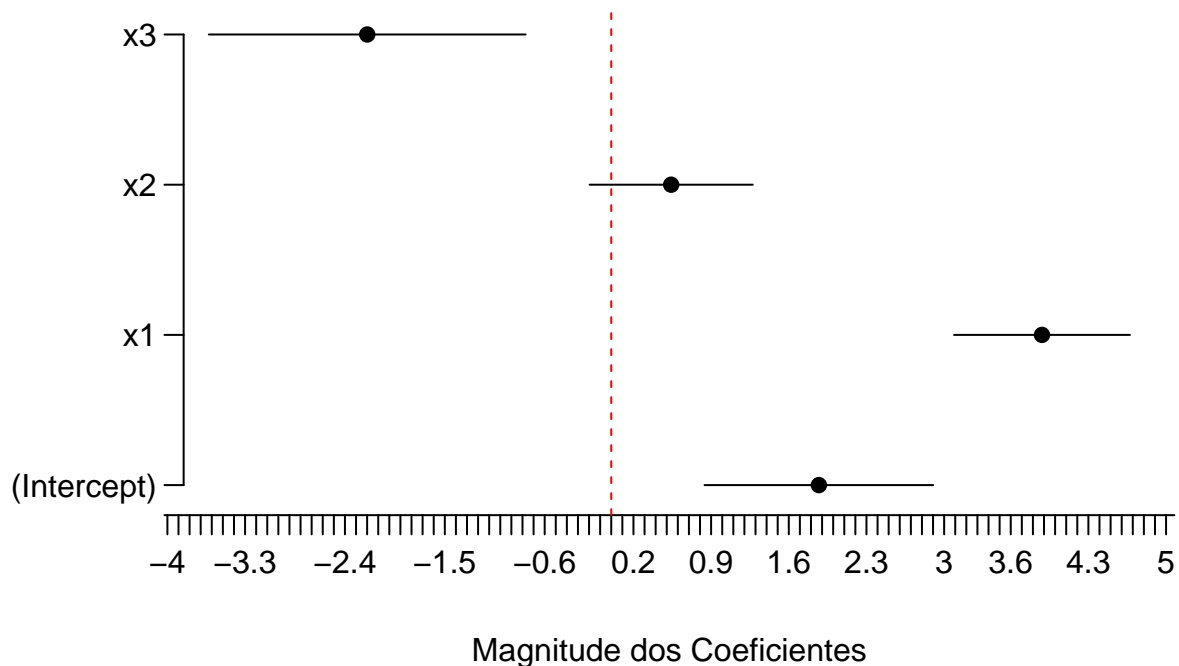
```
summary(modelo.f)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -17.3857 -3.3964 -0.0111 3.5749 16.0700
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.8710      0.5225   3.581 0.000432 ***
## x1           3.8819      0.4020   9.656 < 2e-16 ***
## x2           0.5400      0.3728   1.448 0.149100
## x3          -2.1991      0.7239  -3.038 0.002706 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.054 on 196 degrees of freedom
## Multiple R-squared:  0.3514, Adjusted R-squared:  0.3414
## F-statistic: 35.39 on 3 and 196 DF, p-value: < 2.2e-16
```

```
# Analisar a magnitude dos coeficientes da regressão a partir dos IC:
par(mfrow=c(1,1))
betas <- coefficients(modelo.f)
IC <- confint(modelo.f, level=0.95)

y.axis <- seq(from=1, to=length(betas))
plot(betas, y.axis, type="p", pch=19, xlab="Magnitude dos Coeficientes",
      ylab="", axes=F, xlim=c(min(IC-.4), max(IC+.4)), ylim=c(min(y.axis-.2),
                                                                max(y.axis+.2)),
      cex=1, yaxs="i", xaxs="i")
segments(IC[,1], y.axis, IC[,2], y.axis)
axis(1, at=seq(round(min(IC-.9)), round(max(IC+.9)), by=0.1),
      labels=seq(round(min(IC-.9)), round(max(IC+.9)), by=0.1),
      tick=T, cex.axis=1, mgp=c(2,.7,0))
axis(2, at=y.axis, label=names(betas), las=1, tick=T,
      line=-.5, cex.axis=1, mgp=c(2,.7,0))
abline(v=0, lty=2, col="red")
```



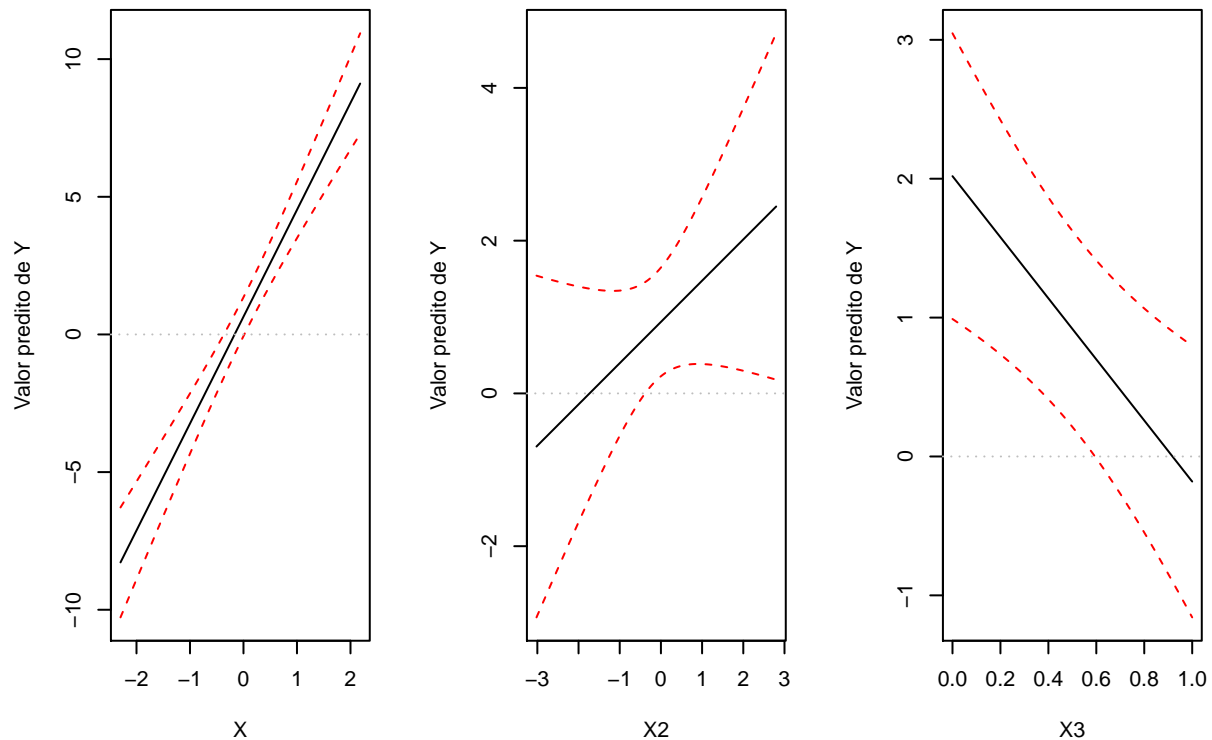
```
# Analisar coeficientes a partir de gráfico dos valores estimados versus observados:
novo.x <- data.frame(x1=seq(min(x1), max(x1), length.out=dim(banco)[1]),
                     x2=mean(x2), x3=mean(x3))
novo.x2 <- data.frame(x1=mean(x1), x2=seq(min(x2), max(x2), length.out=dim(banco)[1]),
                     x3=mean(x3))
novo.x3 <- data.frame(x1=mean(x1), x2=mean(x2), x3=seq(min(x3), max(x3),
                                                         length.out=dim(banco)[1]))

par(mfrow=c(1,3))

y.predito <- predict(modelo.f, novo.x, interval="confidence")
matplot(novo.x$x1, y.predito, lty=c(1,2,2), col=c('black', 'red', 'red'), type="l",
        ylab="Valor predito de Y", xlab="X")
abline(h=0, col='gray', lty=3)

y.predito <- predict(modelo.f, novo.x2, interval="confidence")
matplot(novo.x2$x2, y.predito, lty=c(1,2,2), col=c('black', 'red', 'red'), type="l",
        ylab="Valor predito de Y", xlab="X2")
abline(h=0, col='gray', lty=3)

y.predito <- predict(modelo.f, novo.x3, interval="confidence")
matplot(novo.x3$x3, y.predito, lty=c(1,2,2), col=c('black', 'red', 'red'), type="l",
        ylab="Valor predito de Y", xlab="X3")
abline(h=0, col='gray', lty=3)
```



```
detach(banco)
```

Diagnosticar problemas nos modelos de regressão:

```
# Gerar modelos de regressão:
modelo1 <- lm(y ~ x1 + x2 + x3, data = banco)
modelo2 <- lm(y ~ x1 + x2 + x3, data = banco2)
```

Os dados serão analisados a fim de verificar se cumprem os pressupostos necessários requeridos pelo modelo de mínimos quadrados. Os gráficos gerados a partir da base “banco” atendem a estes pressupostos, enquanto os gerados a partir da base “banco2”, não atendem. Os autores buscam apresentar estes dois exemplos para comparar dados que se adequam ao modelo desenvolvido com dados que não se adequam.

```
# Configurar painel para exibir gráficos, lado a lado:
par(mfrow=c(1,2))

# Analisar influência dos outliers em cada modelo:
outlierTest(modelo1)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
```



```
##      rstudent unadjusted p-value Bonferroni p
## 91 -3.572163      0.00044604      0.089208
```

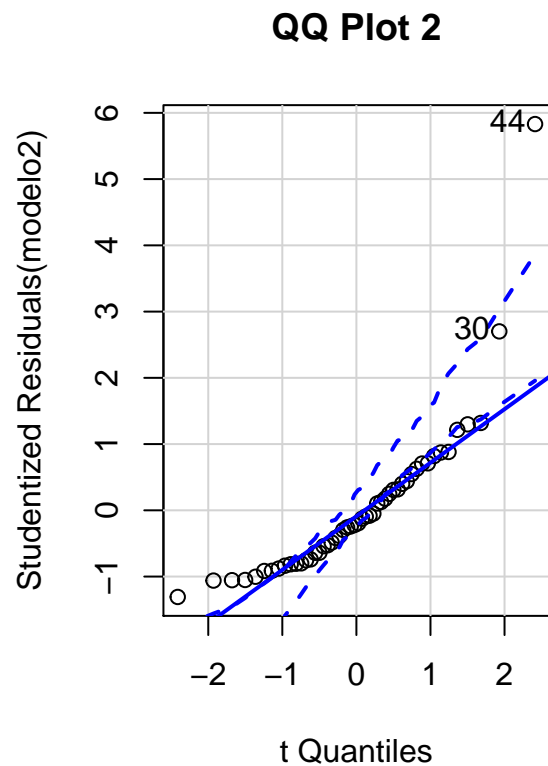
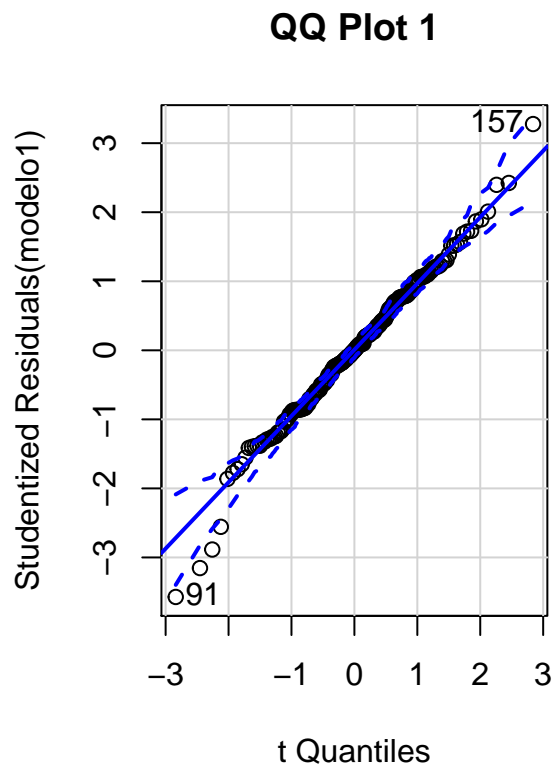
```
outlierTest(modelo2)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 44 5.830979      5.5607e-07      2.7803e-05
```

```
# Analisar distribuição dos resíduos com qqPlot:
qqPlot(modelo1, main="QQ Plot 1")
```

```
## [1] 91 157
```

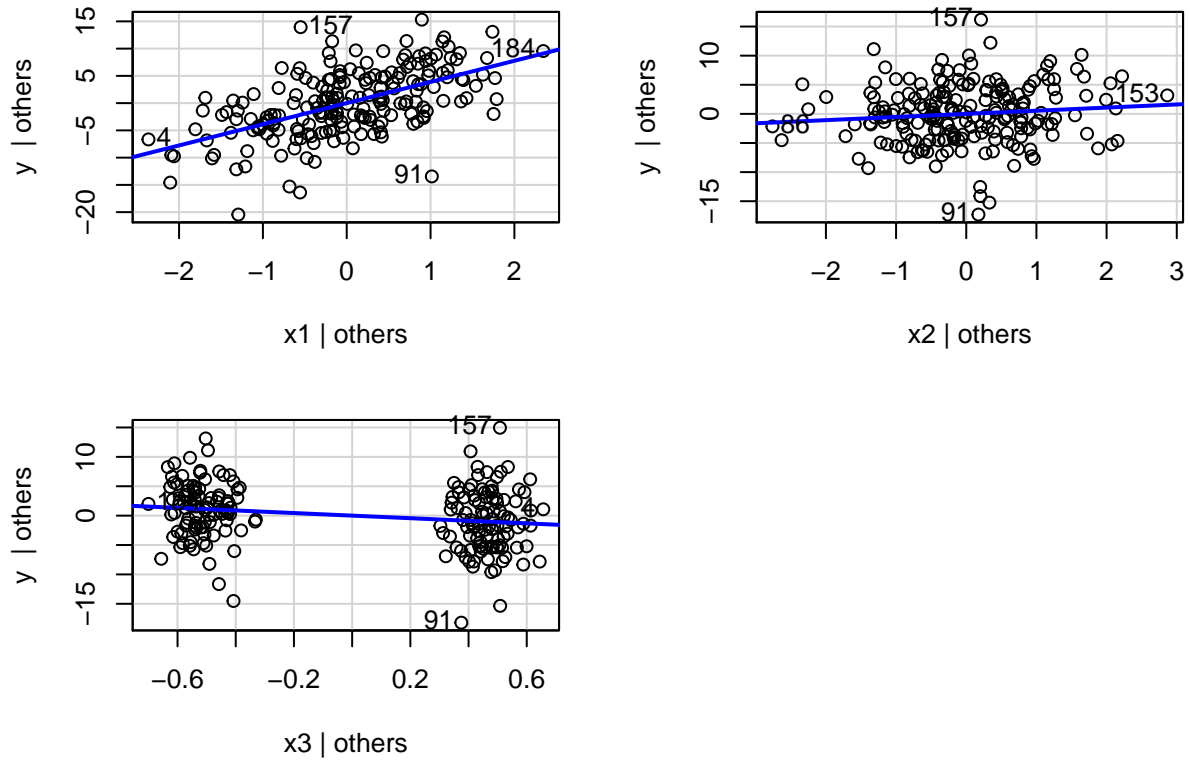
```
qqPlot(modelo2, main="QQ Plot 2")
```



```
## [1] 30 44
```

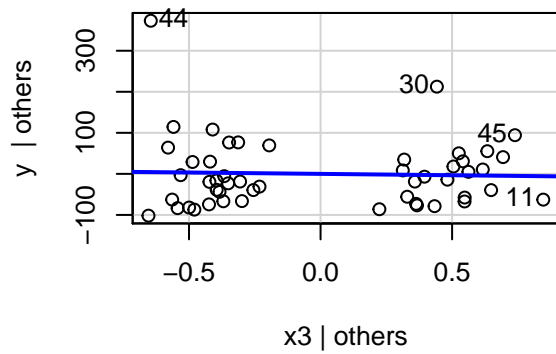
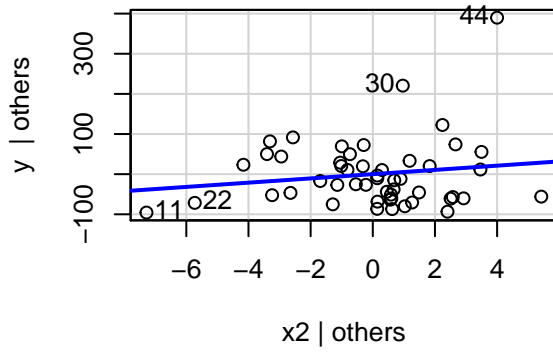
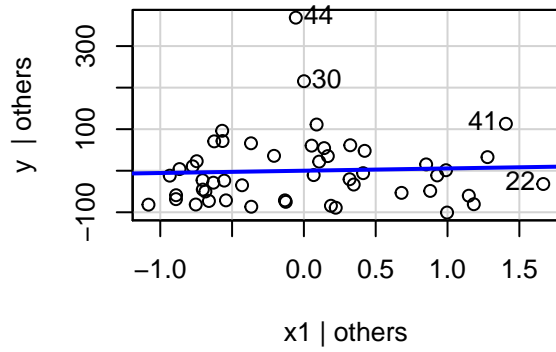
```
# Analisar observações influentes com partial plots:
avPlots(modelo1)
```

Added-Variable Plots

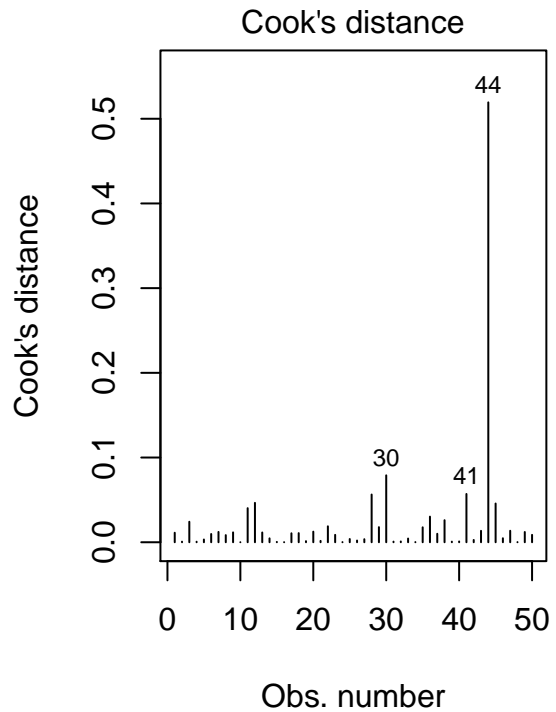
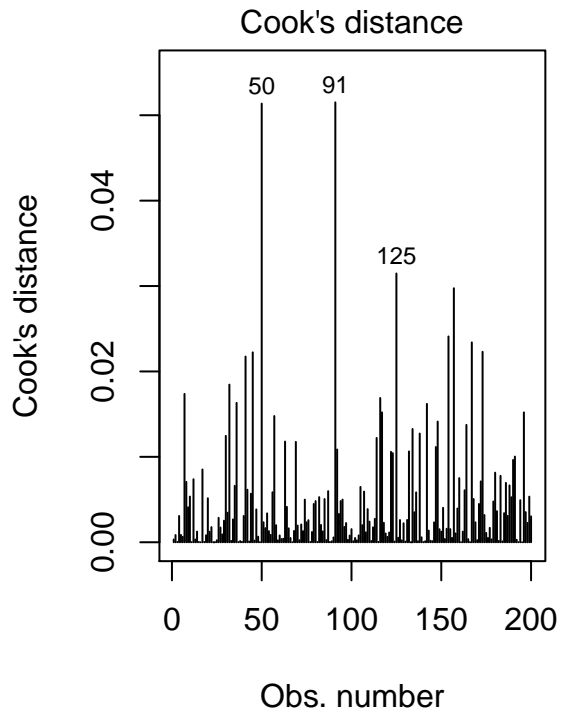


```
avPlots(modelo2)
```

Added-Variable Plots



```
# Analisar observações influentes com o plot do teste das distâncias de Cook:
limite <- 4/((length(banco$x)-length(modelo1$coefficients)-1))
plot(modelo1, which=4, cook.levels=limite)
limite <- 4/((length(banco2$x)-length(modelo2$coefficients)-1))
plot(modelo2, which=4, cook.levels=limite)
```



Gerar gráfico de Influência:

```
influencePlot(modelo1, id.method="identify", main="Gráfico de Influência 1",
              ylim=c(-6,6), xlim=c(0,0.25))
```

##	StudRes	Hat	CookD
## 50	-3.1539967	0.02114137	0.0513675223
## 86	-0.1334193	0.05471987	0.0002589068
## 91	-3.5721633	0.01682806	0.0515110817
## 153	0.3306770	0.05446451	0.0015818367
## 157	3.2768532	0.01149709	0.0297444490

```
influencePlot(modelo2, id.method="identify", main="Gráfico de Influência 2",
              ylim=c(-6,6 ), xlim=c(0,0.25))
```

Gráfico de Influência 1

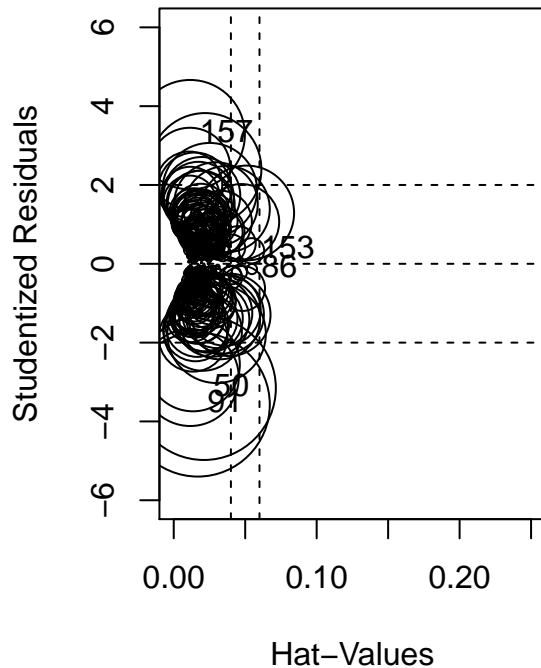
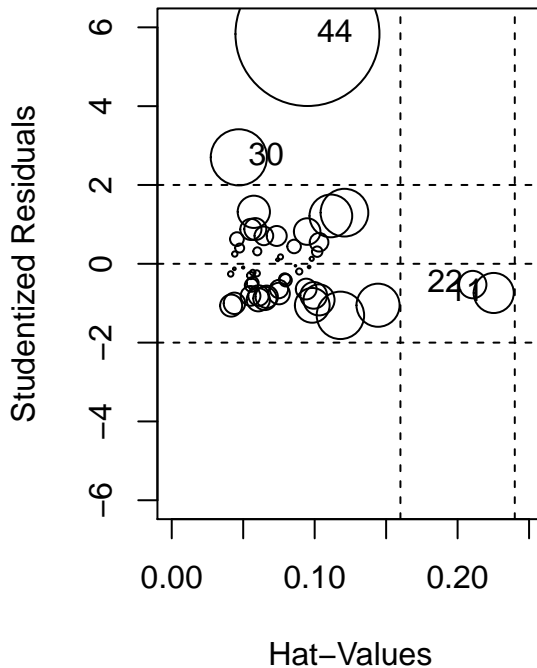


Gráfico de Influência 2



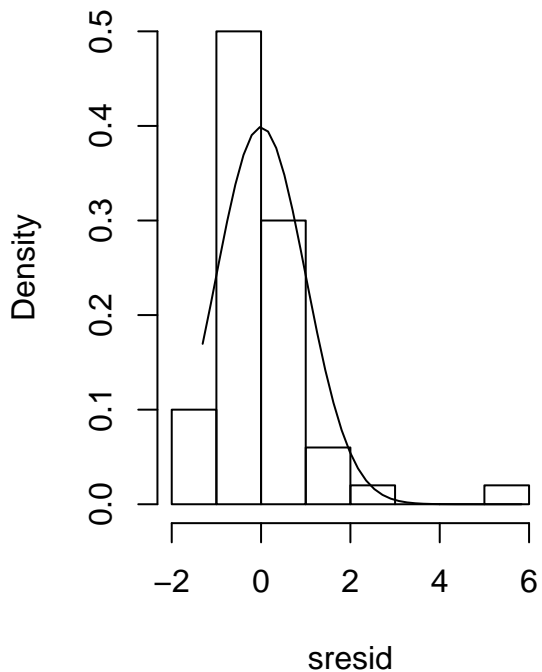
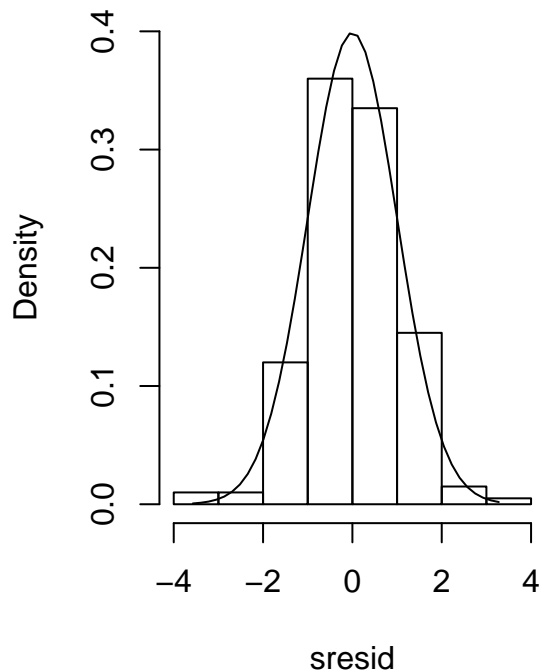
```
##      StudRes      Hat      CookD
## 11 -0.7405834 0.22526537 0.04026368
## 22 -0.5276173 0.21056136 0.01885842
## 30  2.7012080 0.04693638 0.07901850
## 44  5.8309791 0.09496963 0.51936607
```

Checar se resíduos apresentam distribuição normal:

```
sresid <- studres(modelo1)
hist(sresid, freq=FALSE, ylim=c(0,0.4), main="Distribuição dos Resíduos Padronizados 1")
xfit <- seq(min(sresid),max(sresid),length=40)
yfit <- dnorm(xfit)
lines(xfit, yfit)

sresid <- studres(modelo2)
hist(sresid, freq=FALSE, main="Distribuição dos Resíduos Padronizados 2")
xfit <- seq(min(sresid),max(sresid),length=40)
yfit <- dnorm(xfit)
lines(xfit, yfit)
```

istribuição dos Resíduos Padronizaistribuição dos Resíduos Padroniza



```
# Analisar homocedasticidade:
```

```
ncvTest(modelo1)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.0260565, Df = 1, p = 0.87176
```

```
ncvTest(modelo2)
```

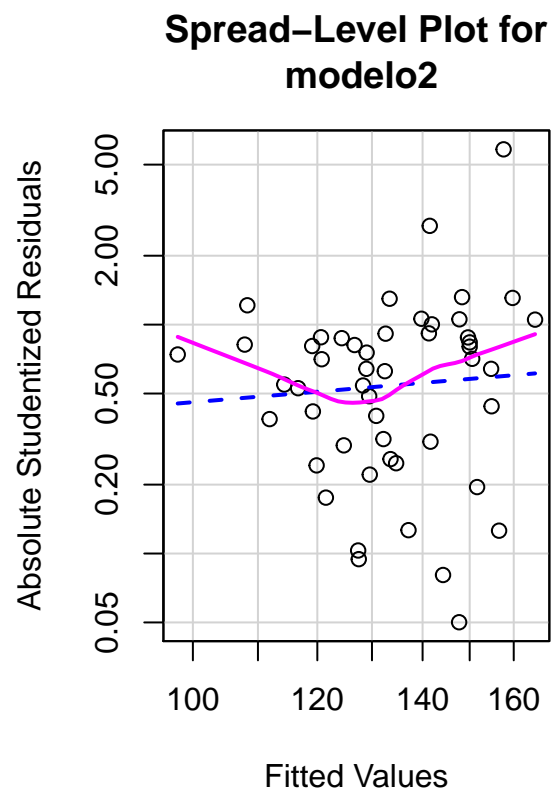
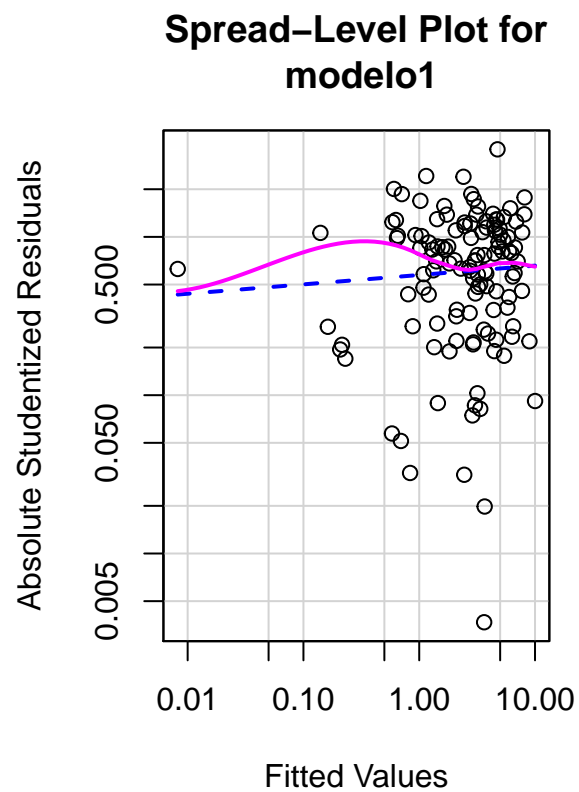
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 13.18187, Df = 1, p = 0.00028267
```

```
# Plotar resíduos padronizados versus valores preditos:
```

```
spreadLevelPlot(modelo1)
```

```
##
## Suggested power transformation: 0.9408374
```

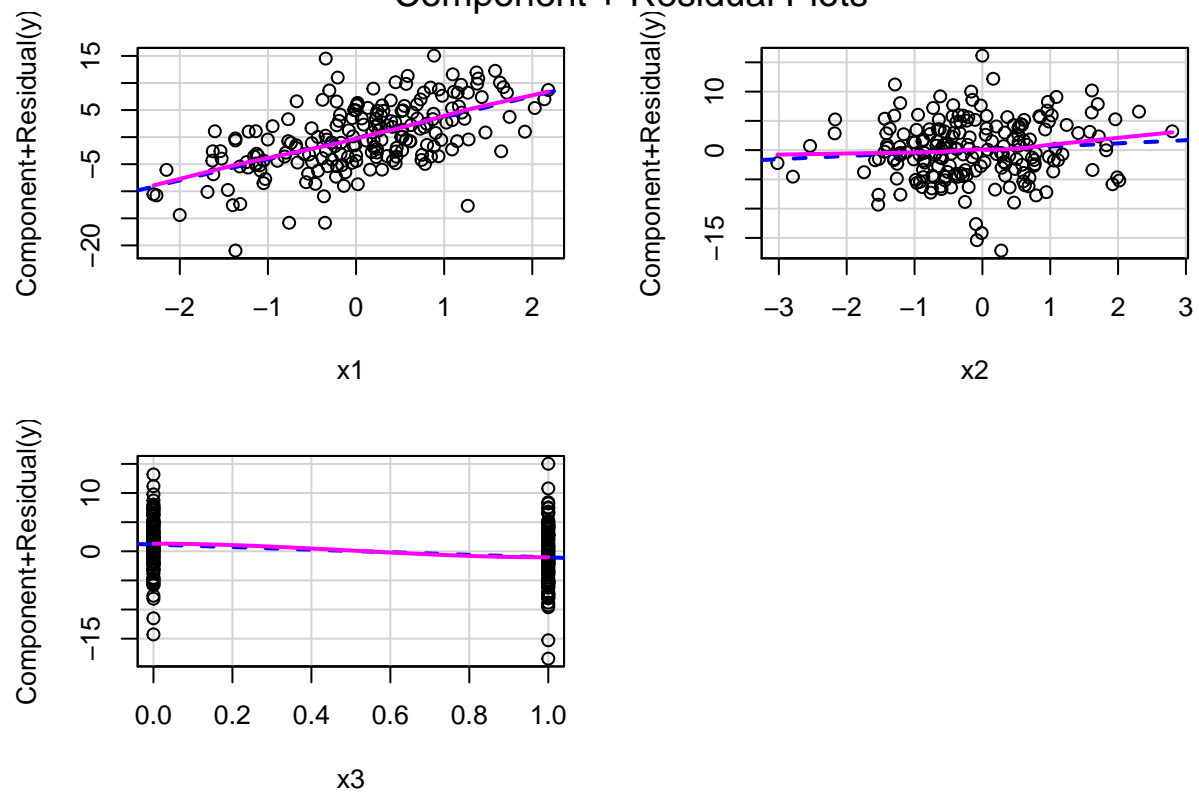
```
spreadLevelPlot(modelo2)
```



```
##
## Suggested power transformation: 0.4227659
```

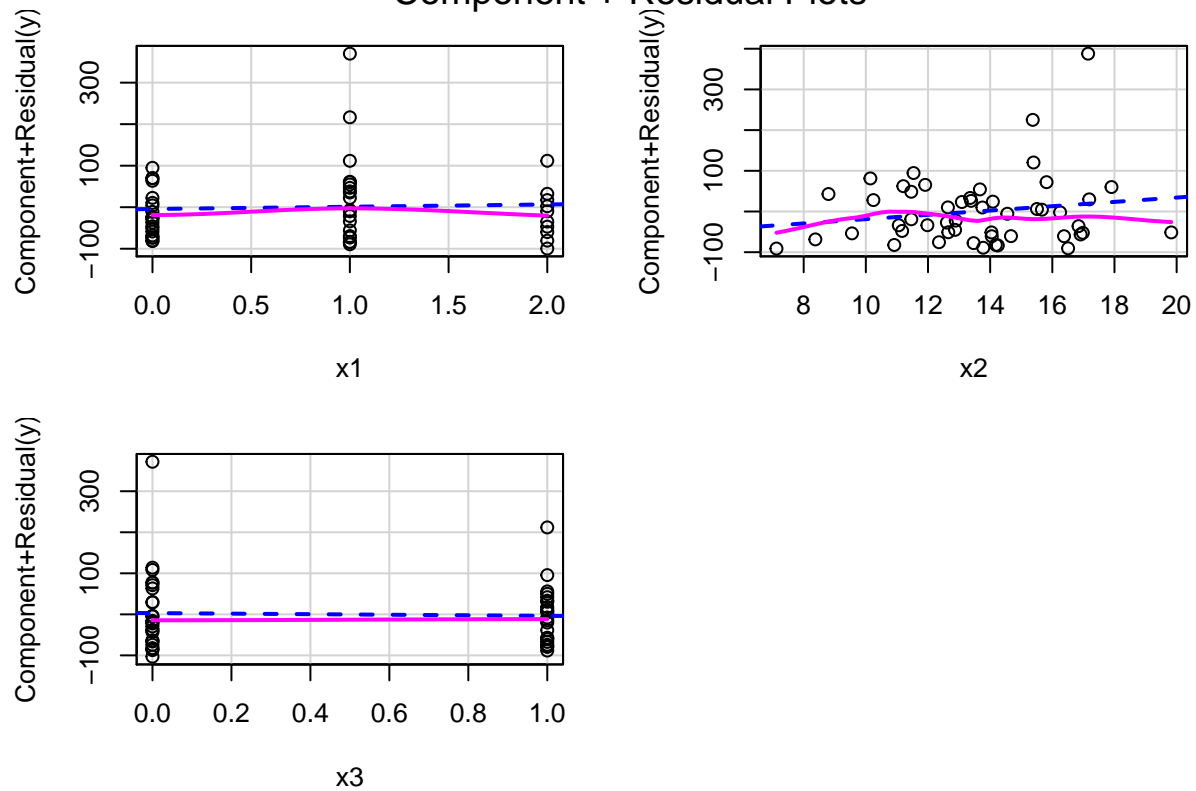
```
# Plotar component + residual plot para analisar se existe não-linearidade:
crPlots(modelo1)
```

Component + Residual Plots



```
crPlots(modelo2)
```


Component + Residual Plots



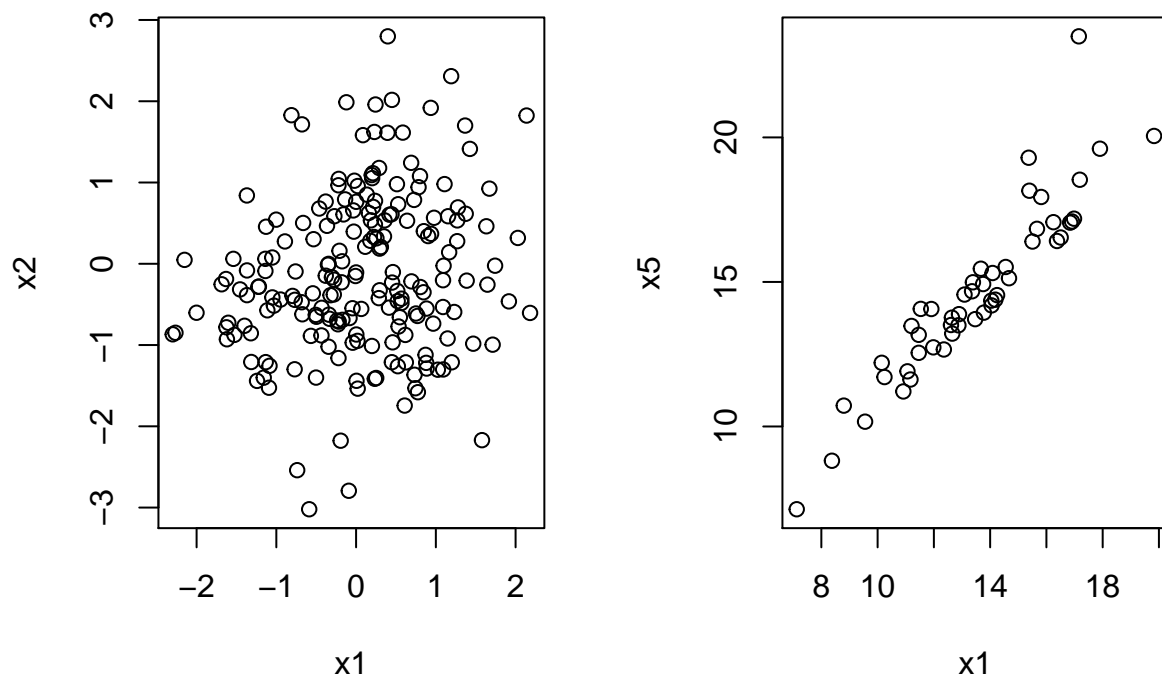
```
# Verificar se existe colinearidade:
vif(modelo1)
```

```
##      x1      x2      x3
## 1.058094 1.034828 1.023288
```

```
sqrt(vif(modelo1)) > 2
```

```
##      x1      x2      x3
## FALSE FALSE FALSE
```

```
par(mfrow=c(1,2))
plot(banco$x1, banco$x2, xlab='x1', ylab='x2')
plot(banco2$x2, banco2$x5, xlab='x1', ylab='x5')
```



```
# Verificar se existe autocorrelação dos erros:
durbinWatsonTest(modelo1)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 -0.09088988 2.177764 0.244
## Alternative hypothesis: rho != 0
```

Questão 2:

letra a)

1. É recomendável analisar graficamente os dados trabalhados antes de interpretar o p-valor.
2. Não faz sentido estimar o p-valor para amostras não aleatórias.
3. O p-valor é afetado pelo tamanho da amostra.
4. Não faz sentido estimar o p-valor quando analisamos dados populacionais (censo).

letra b)

1. Algumas vezes, a análise gráfica dos dados não é útil. É necessário que o pesquisador saiba o efeito que procura identificar a partir da análise dos dados. Gráficos são úteis para relações lineares sem a presença de covariáveis, mas este nem sempre é o caso. Também, os autores que a teoria deve prever o modelo adequado a ser utilizado.
2. O p-valor é uma medida de ajuste da hipótese nula aos dados analisados. Ainda que as amostras não sejam aleatórias, o p-valor pode trazer valiosas informações.
3. Não necessariamente uma amostra maior resultará em um p-valor significativo. Caso a hipótese nula seja verdadeira, uma amostra grande não apresentará um p-valor significativo. É necessário distinguir o PODER de um teste e a sua SIGNIFICÂNCIA. Também, os autores afirmam que efeitos que apresentam um p-valor significativo em uma amostra pequena é tão relevante quanto aqueles que apresentam um p-valor significativo em uma amostra grande.
4. Mesmo para dados populacionais o p-valor é útil, já que em um censo um dos objetivos é rejeitar uma hipótese nula. Ademais, o p-valor nestes casos, pode apontar alguma possível falha na coleta dos dados.

letra c)

1. Por vezes, os gráficos podem fornecer, de fato, informações que não são úteis para uma pesquisa; mas ignorá-los é um cenário ainda pior. Os autores afirmam que gráficos são poderosas ferramentas não só para relações lineares, como também para identificar relações exponenciais, cúbicas e quadráticas. Por fim, os autores afirmam que os dados podem ter um comportamento diferente do previsto na teoria e, por isso, a análise gráfica torna-se essencial.
2. Os autores apontam que o p-valor não é uma medida de ajuste, mas sim de probabilidade de encontrar o valor observado no teste estatístico quando a hipótese nula é verdadeira. Para amostras não aleatórias existe um problema de super ou subestimação do teste t, o que afeta o p-valor. O teorema do limite central se aplica apenas às amostras aleatórias, portanto, nas amostras não aleatórias, o p-valor, assim como os intervalos de confiança, apresentarão problemas nas suas medidas.
3. Os autores reiteram que quanto maior a amostra, menor o p-valor e apontam que as medidas das amostras menores são mais instáveis; quando uma amostra é pequena, apenas efeitos fortes são capazes de alcançar significância estatística. Um dos pressupostos para a análise do p-valor é que a amostra apresenta uma distribuição normal e em amostras pequenas este pressuposto é difícil de ser testado com precisão. Em amostras grandes, mesmo efeitos fracos podem ser estatisticamente significantes. Outro ponto é que em amostras pequenas os outliers afetam mais as estimativas, quando comparadas às amostras maiores.
4. Os autores acreditam que falhas na coleta dos dados não é uma razão suficiente para justificar a estimação do p-valor quando são utilizados dados populacionais. Eles afirmam que o p-valor não é uma medida capaz de apresentar a qualidade da mensuração das variáveis. Quando os dados utilizados fornecem informações a respeito de toda a população, não há incerteza e não é necessária a utilização desta estimativa.

Questão 3:

letra a)

Samuels conclui que as eleições legislativas no Brasil são “*state-centered*”, ou seja, para os candidatos que disputam uma vaga no Congresso, as conexões a nível estadual são mais importantes do que aquelas a nível nacional. Eles acreditam que os candidatos à presidência detêm poucas ferramentas capazes de influenciar o resultado das eleições para o Congresso (enquanto governadores exercem grande influência); este processo torna menos provável a formação de uma coalizão que apoie o governo do presidente incumbente.

O modelo: O autor postula a hipótese de que será encontrada uma correlação, condicionada ao período das eleições para governador, entre o número efetivo de candidatos a governador e o número efetivo de candidatos nas listas abertas que disputam uma vaga no Congresso. O autor utilizou uma série de dados “cross-sectional” sobre as eleições no Brasil nos períodos democráticos, entre 1945 e 1964 e 1989 e 1998. A variável dependente, ENEL, é a medida do número de listas formadas de candidatos que disputam uma vaga na Câmara dos Deputados em um estado “s” e tempo “t”. São utilizadas 5 variáveis independentes principais:

PROXGOV: a proximidade entre a eleição governamental e a eleição para o Congresso, em um estado “s” em um tempo “t”. Recebe valor 1 caso aconteçam no mesmo período e para proximidade e 0 para quando as eleições para o Congresso acontecem no meio do mandato do governador.

ENGOV: número efetivo de candidatos a governador em um estado “s” em um período “t”. O autor postula que o número de candidatos afetará indiretamente a VD a depender do período da eleição, como apontará a variável PROXGOV.

Interação PROXGOV x ENGOV.

PROXPRES: a proximidade entre a eleição presidencial e a eleição para o Congresso, em um estado “s” em um tempo “t”. Medida da mesma forma que PROXGOV, o autor postula que não haverá efeito desta variável na variável dependente.

ENPRES: número efetivo de candidatos à presidência em um estado “s” em um período “t”.

Interação PROXPRES x ENPRES.

logM: é o log da magnitude do distrito, ou seja, o número de assentos a serem preenchidos na Câmara dos Deputados em um estado “s” e período “t”.

Foram adicionadas as variáveis dummies YEARx e STATEx.

Nos anos de 1950, 1994 e 1998 as eleições ocorreram simultaneamente, por isso, foram rodadas três regressões, cada uma excluindo um dos anos em questão.

O autor analisa o efeito das variáveis LogM, PROXGOV, PROXGOV x ENGOV, PROXPRES e PROXPRES x ENPRES sobre a variável dependente e encontra que o efeito das variáveis associadas com as eleições para governador são estatisticamente significantes, enquanto àquelas associadas às eleições presidenciais não apresentaram efeitos significantes.

letra b)

Os autores afirmam que nos modelos interativos desenvolvidos por Samuels (2000) os termos constitutivos foram omitidos. Quando são incluídos, nenhuma das variáveis anteriormente analisadas apresentam efeitos estatisticamente significantes. Quando o efeito marginal das eleições para governador no número de listas é plotado, para os 3 modelos apresentados, as eleições para governador não apresentam o “coattails effect” (quando um político popular atrai votos para outros candidatos). Os autores afirmam que se existe algum “coattails effect” nas eleições para o Congresso Nacional no Brasil, este efeito, na verdade, será percebido a nível presidencial e não estadual (governadores).

letra c)

As quatro principais recomendações são: utilizar modelos interativos quando a hipótese a ser testada é condicional por natureza; incluir todos os termos constitutivos nas especificações do modelo interativo; não interpretar termos constitutivos como se fossem efeitos marginais incondicionais; calcular efeitos marginais substancialmente significantes e os erros padrão.

Questão 4:

```
# Instalar e requerer pacotes necessários:
# install.packages("rio")
require(rio)
require(dplyr)
```

```
# Carregar base de dados:
base_samuels <- import(file = "samuels.dta")
```

```
# Checar variáveis:
names(base_samuels)
```

```
## [1] "state"          "yeardeputy"     "enlists"
## [4] "magnitude"      "enpres"         "yearpres"
## [7] "proxpres"       "engov"          "yeargov"
## [10] "proxgov"        "logmag"         "engov_proxgov"
## [13] "enpres_proxpres" "ac"             "al"
## [16] "am"             "ap"             "ba"
## [19] "ce"             "df"             "es"
```

```
## [22] "gb"          "go"          "ma"
## [25] "mg"          "ms"          "mt"
## [28] "pa"          "pb"          "pe"
## [31] "pi"          "pr"          "rj"
## [34] "rn"          "ro"          "rr"
## [37] "rs"          "sc"          "se"
## [40] "sp"          "toc"         "north"
## [43] "neast"       "cwest"       "seast"
## [46] "south"       "year1945"    "year1947"
## [49] "year1950"    "year1954"    "year1958"
## [52] "year1962"    "year1990"    "year1994"
## [55] "year1998"
```

Replicar o primeiro modelo (sem o ano de 1950):

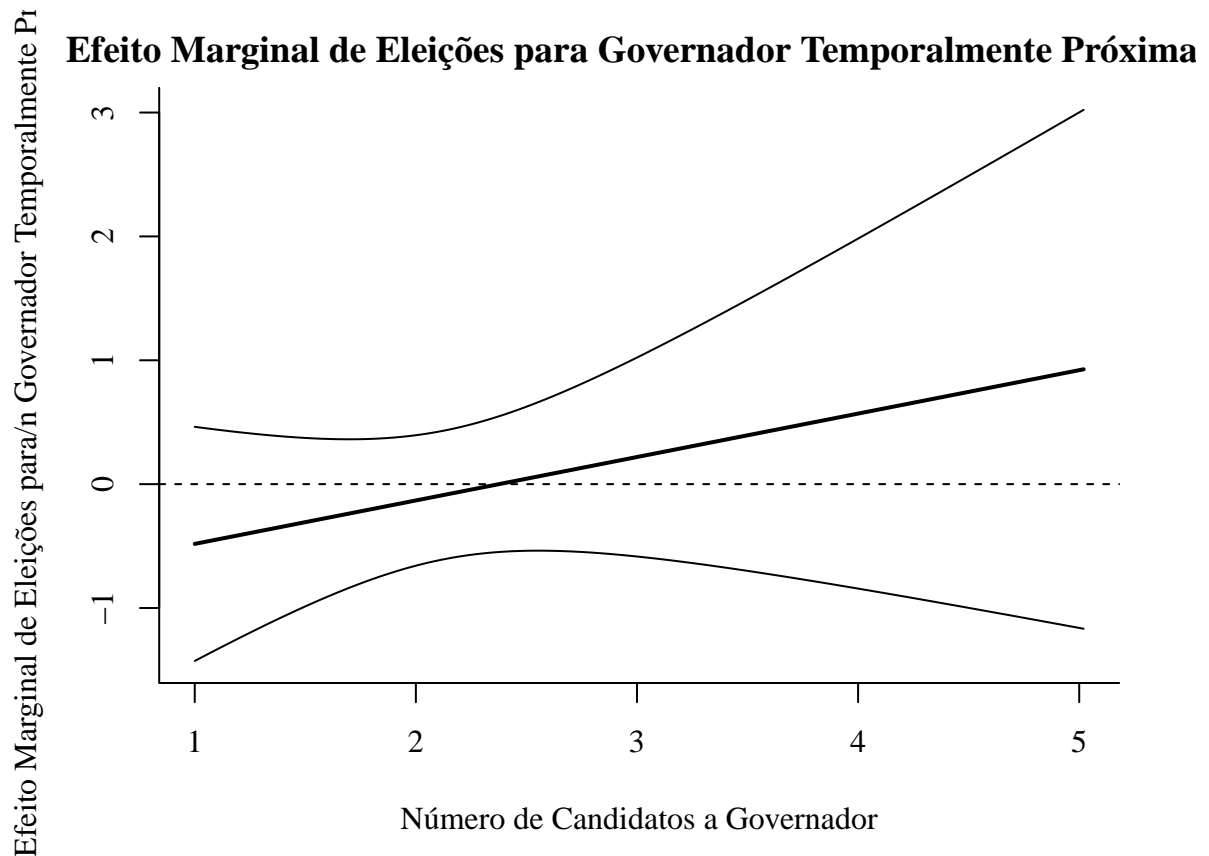
```
reg1 <- lm(enlists ~ proxgov + engov + engov_proxgov + proxpres + enpres +
  enpres_proxpres + logmag + year1945 + year1947 + year1954 +
  year1958 + year1962 + year1994 + year1998 + al + am + ac + ap +
  ba + ce + df + es + go + ma + mg + ms + mt + pa + pb + pe + pi +
  pr + rj + rn + ro + rr + rs + sc + se + sp + toc, data = base_samuels)
```

```
summary(reg1)
```

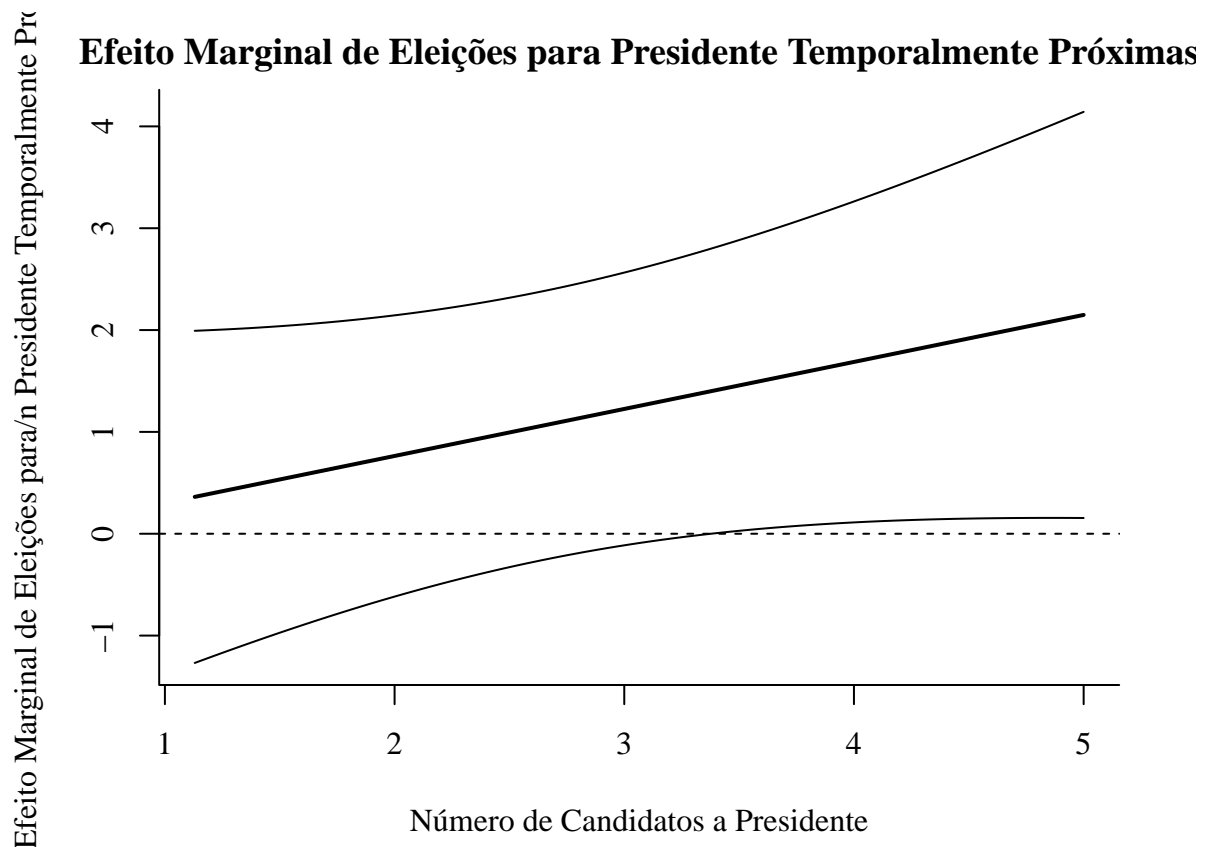
```
##
## Call:
## lm(formula = enlists ~ proxgov + engov + engov_proxgov + proxpres +
##   enpres + enpres_proxpres + logmag + year1945 + year1947 +
##   year1954 + year1958 + year1962 + year1994 + year1998 + al +
##   am + ac + ap + ba + ce + df + es + go + ma + mg + ms + mt +
##   pa + pb + pe + pi + pr + rj + rn + ro + rr + rs + sc + se +
##   sp + toc, data = base_samuels)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26649 -0.27902 -0.01931  0.22650  2.23748
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.264490   1.712917   0.738   0.4617
## proxgov       -0.832856   0.805439  -1.034   0.3029
## engov          0.141182   0.303267   0.466   0.6423
## engov_proxgov  0.350609   0.357752   0.980   0.3288
## proxpres      -0.159684   1.087746  -0.147   0.8835
## enpres        -0.218299   0.230210  -0.948   0.3447
## enpres_proxpres 0.461669   0.323420   1.427   0.1557
## logmag         0.911734   0.526582   1.731   0.0856
## year1945      -0.086789   0.198392  -0.437   0.6625
## year1947       0.751871   0.800881   0.939   0.3495
## year1954       0.146064   0.254675   0.574   0.5672
## year1958       0.462212   0.464606   0.995   0.3216
## year1962       0.183359   0.472451   0.388   0.6985
## year1994      -0.406197   0.208596  -1.947   0.0535
## year1998      -0.203562   0.203279  -1.001   0.3184
## al             0.032568   0.682143   0.048   0.9620
## am             0.286282   0.709456   0.404   0.6872
```

```
## ac          0.212578    0.780464    0.272    0.7857
## ap         -0.217654    0.770398   -0.283    0.7780
## ba          0.214144    0.678333    0.316    0.7527
## ce         -0.363694    0.681235   -0.534    0.5943
## df          0.075602    0.814915    0.093    0.9262
## es          0.231562    0.712872    0.325    0.7458
## go         -0.009147    0.696944   -0.013    0.9895
## ma         -0.688633    0.679249   -1.014    0.3125
## mg          0.772835    0.689782    1.120    0.2645
## ms         -0.191411    0.809232   -0.237    0.8134
## mt         -0.134494    0.697001   -0.193    0.8473
## pa          0.576608    0.678994    0.849    0.3972
## pb         -0.406140    0.690832   -0.588    0.5576
## pe         -0.414013    0.681957   -0.607    0.5448
## pi         -0.378995    0.710753   -0.533    0.5947
## pr          0.854990    0.655272    1.305    0.1942
## rj          1.073520    0.687036    1.563    0.1205
## rn         -0.120071    0.710816   -0.169    0.8661
## ro          0.137871    0.766239    0.180    0.8575
## rr          0.310703    0.771565    0.403    0.6878
## rs          0.215481    0.693917    0.311    0.7566
## sc         -0.026683    0.689345   -0.039    0.9692
## se         -0.081348    0.717027   -0.113    0.9098
## sp          0.876809    0.715900    1.225    0.2228
## toc        -0.533252    0.803236   -0.664    0.5079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5744 on 137 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.595
## F-statistic: 7.379 on 41 and 137 DF,  p-value: < 2.2e-16
```

```
# Plotar gráfico dos efeitos marginais de proxgov sobre enlists, condicionado a engov:
beta.hat <- coef(reg1)
cov <- vcov(reg1)
z0 <- seq(min(base_samuels$engov), max(base_samuels$engov), length.out = 1000)
dy.dx <- beta.hat["proxgov"] + beta.hat["engov_proxgov"]*z0
se.dy.dx <- sqrt(cov["proxgov", "proxgov"] + z0^2*cov["engov_proxgov", "engov_proxgov"] +
  2*z0*cov["proxgov", "engov_proxgov"])
upr <- dy.dx + 1.96*se.dy.dx
lwr <- dy.dx - 1.96*se.dy.dx
par(family="serif",bty="l",mar=c(5,5.5,2,2))
plot(x=z0, y=dy.dx,type="n",xlim=c(min(z0),max(z0)),
     ylim=c(min(lwr),max(upr)),
     xlab = "Número de Candidatos a Governador",
     ylab = "Efeito Marginal de Eleições para/n Governador Temporalmente Próximas",
     main= "Efeito Marginal de Eleições para Governador Temporalmente Próximas")
lines(z0, dy.dx, lwd = 2)
lines(z0, lwr)
lines(z0, upr)
abline(h=0,lty=2)
```



```
# Plotar gráfico dos efeitos marginais de proxpres sobre enlists, condicionado a enpres:
beta.hat2 <- coef(reg1)
cov2 <- vcov(reg1)
z2 <- seq(min(base_samuels$enpres), max(base_samuels$enpres), length.out = 1000)
dy.dx2 <- beta.hat2["proxpres"] + beta.hat2["enpres_proxpres"]*z2
se.dy.dx2 <- sqrt(cov2["proxpres", "proxpres"] + z2^2*cov2["enpres_proxpres",
"enpres_proxpres"] +
2*z2*cov2["proxpres", "enpres_proxpres"])
upr2 <- dy.dx2 + 1.96*se.dy.dx2
lwr2 <- dy.dx2 - 1.96*se.dy.dx2
par(family="serif",bty="l",mar=c(5,5.5,2,2))
plot(x=z2, y=dy.dx2,type="n",xlim=c(min(z2),max(z2)),
ylim=c(min(lwr2),max(upr2)),
xlab = "Número de Candidatos a Presidente",
ylab = "Efeito Marginal de Eleições para/n Presidente Temporalmente Próximas",
main= "Efeito Marginal de Eleições para Presidente Temporalmente Próximas")
lines(z2, dy.dx2, lwd = 2)
lines(z2, lwr2)
lines(z2, upr2)
abline(h=0,lty=2)
```

```
# Replicar o segundo modelo (sem o ano de 1994):
reg2 <- lm(enlists ~ proxgov + engov + engov_proxgov + proxpres + enpres +
           enpres_proxpres + logmag + year1945 + year1947 + year1950 +
           year1954 + year1958 + year1962 + year1998 + al + am + ac + ap +
           ba + ce + df + es + go + ma + mg + ms + mt + pa + pb + pe + pi +
           pr + rj + rn + ro + rr + rs + sc + se + sp + toc, data = base_samuels)

summary(reg2)
```

```
##
## Call:
## lm(formula = enlists ~ proxgov + engov + engov_proxgov + proxpres +
##     enpres + enpres_proxpres + logmag + year1945 + year1947 +
##     year1950 + year1954 + year1958 + year1962 + year1998 + al +
##     am + ac + ap + ba + ce + df + es + go + ma + mg + ms + mt +
##     pa + pb + pe + pi + pr + rj + rn + ro + rr + rs + sc + se +
##     sp + toc, data = base_samuels)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26649 -0.27902 -0.01931  0.22650  2.23748
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.873785    1.650741    1.135   0.2583
```

```

## proxgov      -0.832856   0.805439  -1.034   0.3029
## engov        0.141182   0.303267   0.466   0.6423
## engov_proxgov 0.350609   0.357752   0.980   0.3288
## proxpres     -1.175176   1.010084  -1.163   0.2467
## enpres       -0.218299   0.230210  -0.948   0.3447
## enpres_proxpres 0.461669   0.323420   1.427   0.1557
## logmag       0.911734   0.526582   1.731   0.0856 .
## year1945     0.319408   0.220413   1.449   0.1496
## year1947     0.345674   0.735878   0.470   0.6393
## year1950     0.406197   0.208596   1.947   0.0535 .
## year1954     0.146064   0.254675   0.574   0.5672
## year1958     0.056015   0.441364   0.127   0.8992
## year1962     -0.222838   0.455747  -0.489   0.6257
## year1998     0.202635   0.159746   1.268   0.2068
## al           0.032568   0.682143   0.048   0.9620
## am           0.286282   0.709456   0.404   0.6872
## ac           0.212578   0.780464   0.272   0.7857
## ap           -0.217654   0.770398  -0.283   0.7780
## ba           0.214144   0.678333   0.316   0.7527
## ce           -0.363694   0.681235  -0.534   0.5943
## df           0.075602   0.814915   0.093   0.9262
## es           0.231562   0.712872   0.325   0.7458
## go           -0.009147   0.696944  -0.013   0.9895
## ma           -0.688633   0.679249  -1.014   0.3125
## mg           0.772835   0.689782   1.120   0.2645
## ms           -0.191411   0.809232  -0.237   0.8134
## mt           -0.134494   0.697001  -0.193   0.8473
## pa           0.576608   0.678994   0.849   0.3972
## pb           -0.406140   0.690832  -0.588   0.5576
## pe           -0.414013   0.681957  -0.607   0.5448
## pi           -0.378995   0.710753  -0.533   0.5947
## pr           0.854990   0.655272   1.305   0.1942
## rj           1.073520   0.687036   1.563   0.1205
## rn           -0.120071   0.710816  -0.169   0.8661
## ro           0.137871   0.766239   0.180   0.8575
## rr           0.310703   0.771565   0.403   0.6878
## rs           0.215481   0.693917   0.311   0.7566
## sc           -0.026683   0.689345  -0.039   0.9692
## se           -0.081348   0.717027  -0.113   0.9098
## sp           0.876809   0.715900   1.225   0.2228
## toc         -0.533252   0.803236  -0.664   0.5079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5744 on 137 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.595
## F-statistic: 7.379 on 41 and 137 DF,  p-value: < 2.2e-16

```

```

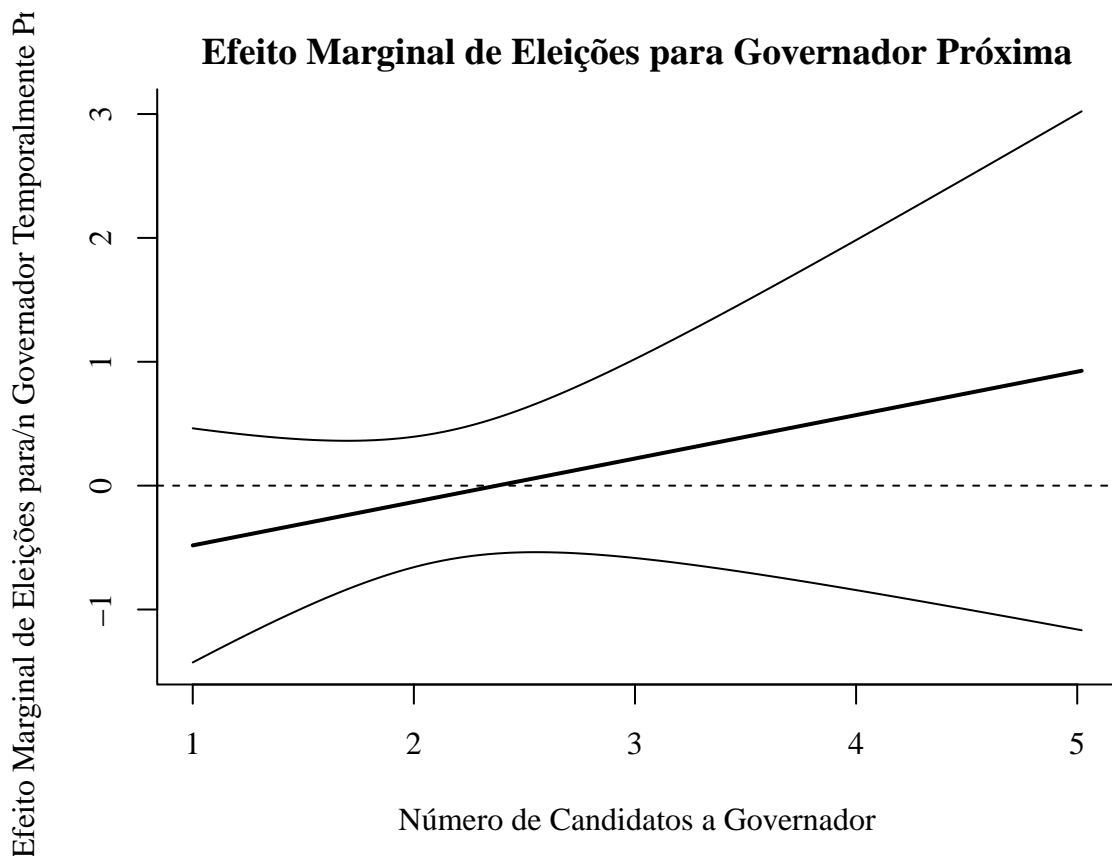
# Plotar gráfico dos efeitos marginais de proxgov sobre enlists, condicionado a engov:
beta.hat3 <- coef(reg2)
cov3 <- vcov(reg2)
z3 <- seq(min(base_samuels$engov), max(base_samuels$engov), length.out = 1000)
dy.dx3 <- beta.hat3["proxgov"] + beta.hat3["engov_proxgov"]*z3
se.dy.dx3 <- sqrt(cov3["proxgov", "proxgov"] + z3^2*cov3["engov_proxgov",

```

```

                                "engov_proxgov"] +
                                2*z3*cov3["proxgov", "engov_proxgov"])
upr3 <- dy.dx3 + 1.96*se.dy.dx3
lwr3 <- dy.dx3 - 1.96*se.dy.dx3
par(family="serif",bty="l",mar=c(5,5.5,2,2))
plot(x=z3, y=dy.dx3,type="n",xlim=c(min(z3),max(z3)),
     ylim=c(min(lwr3),max(upr3)),
     xlab = "Número de Candidatos a Governador",
     ylab = "Efeito Marginal de Eleições para/n Governador Temporalmente Próximas",
     main = "Efeito Marginal de Eleições para Governador Próxima")
lines(z3, dy.dx3, lwd = 2)
lines(z3, lwr3)
lines(z3, upr3)
abline(h=0,lty=2)

```



```

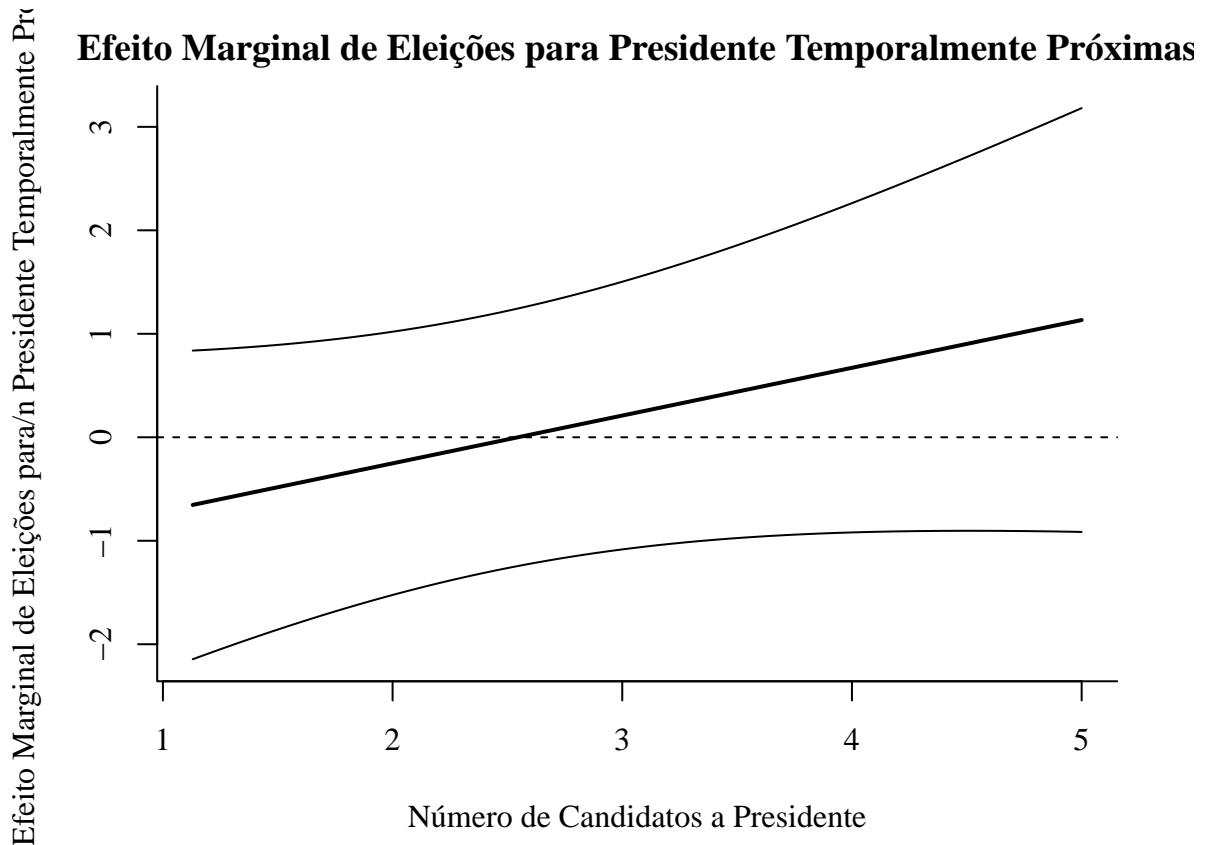
# Plotar gráfico dos efeitos marginais de proxpres sobre enlists, condicionado a enpres:
beta.hat4 <- coef(reg2)
cov4 <- vcov(reg2)
z4 <- seq(min(base_samuels$enpres), max(base_samuels$enpres), length.out = 1000)
dy.dx4 <- beta.hat4["proxpres"] + beta.hat4["enpres_proxpres"]*z4
se.dy.dx4 <- sqrt(cov4["proxpres", "proxpres"] + z4^2*cov4["enpres_proxpres",
                                                             "enpres_proxpres"] +
                  2*z4*cov4["proxpres", "enpres_proxpres"])
upr4 <- dy.dx4 + 1.96*se.dy.dx4
lwr4 <- dy.dx4 - 1.96*se.dy.dx4

```

```

par(family="serif",bty="l",mar=c(5,5.5,2,2))
plot(x=z4, y=dy.dx4,type="n",xlim=c(min(z4),max(z4)),
     ylim=c(min(lwr4),max(upr4)),
     xlab = "Número de Candidatos a Presidente",
     ylab = "Efeito Marginal de Eleições para/n Presidente Temporalmente Próximas",
     main= "Efeito Marginal de Eleições para Presidente Temporalmente Próximas")
lines(z4, dy.dx4, lwd = 2)
lines(z4, lwr4)
lines(z4, upr4)
abline(h=0,lty=2)

```



```

# Replicar o terceiro modelo (sem o ano 1998):
reg3 <- lm(enlists ~ proxgov + engov + engov_proxgov + proxpres + enpres +
           enpres_proxpres + logmag + year1945 + year1947 + year1950 +
           year1954 + year1958 + year1962 + year1994 + al + am + ac + ap +
           ba + ce + df + es + go + ma + mg + ms + mt + pa + pb + pe + pi +
           pr + rj + rn + ro + rr + rs + sc + se + sp + toc, data = base_samuels)

summary(reg3)

```

```

##
## Call:
## lm(formula = enlists ~ proxgov + engov + engov_proxgov + proxpres +
##     enpres + enpres_proxpres + logmag + year1945 + year1947 +
##     year1950 + year1954 + year1958 + year1962 + year1994 + al +

```

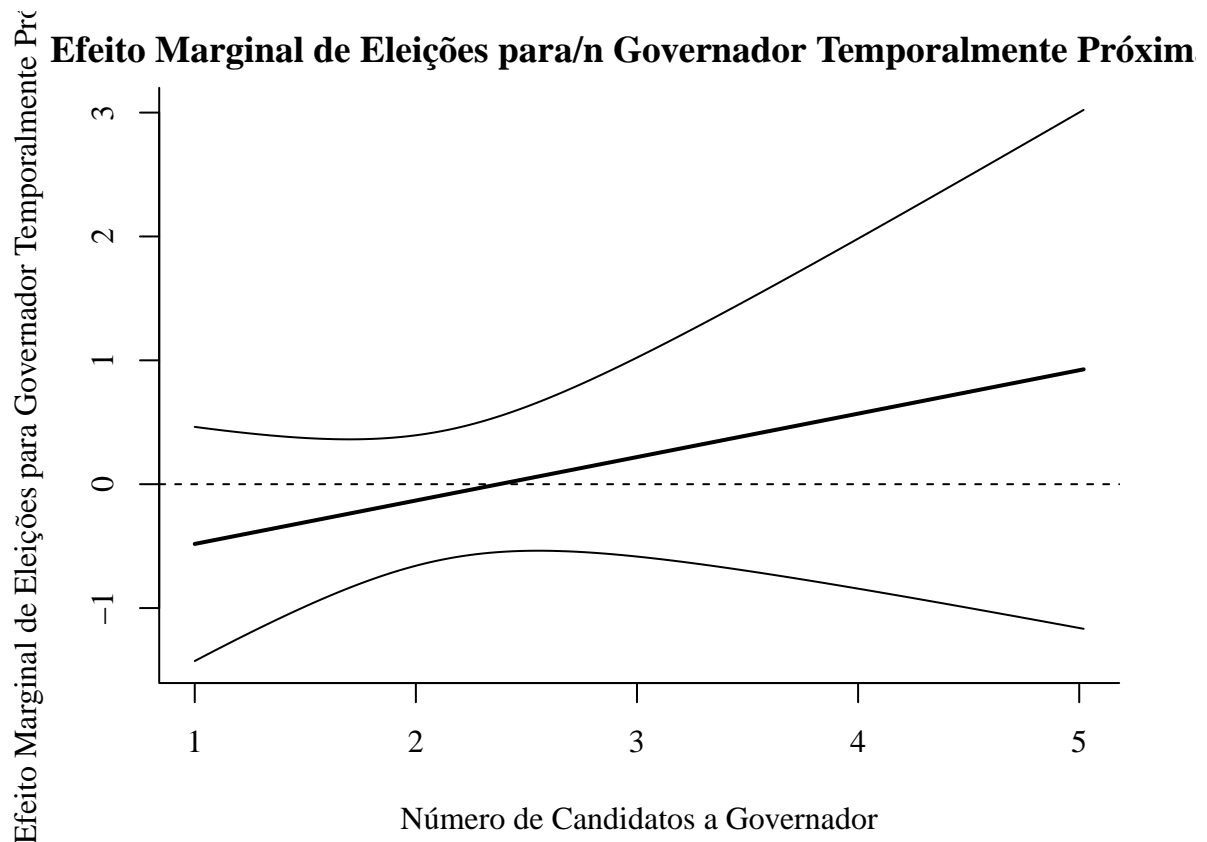
```

##      am + ac + ap + ba + ce + df + es + go + ma + mg + ms + mt +
##      pa + pb + pe + pi + pr + rj + rn + ro + rr + rs + sc + se +
##      sp + toc, data = base_samuels)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -1.26649 -0.27902 -0.01931  0.22650  2.23748
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.569833   1.644845   0.954   0.3416
## proxgov      -0.832856   0.805439  -1.034   0.3029
## engov         0.141182   0.303267   0.466   0.6423
## engov_proxgov 0.350609   0.357752   0.980   0.3288
## proxpres     -0.668589   1.018828  -0.656   0.5128
## enpres       -0.218299   0.230210  -0.948   0.3447
## enpres_proxpres 0.461669   0.323420   1.427   0.1557
## logmag        0.911734   0.526582   1.731   0.0856
## year1945      0.116773   0.219802   0.531   0.5961
## year1947      0.548308   0.729409   0.752   0.4535
## year1950      0.203562   0.203279   1.001   0.3184
## year1954      0.146064   0.254675   0.574   0.5672
## year1958      0.258649   0.435453   0.594   0.5535
## year1962     -0.020203   0.450870  -0.045   0.9643
## year1994     -0.202635   0.159746  -1.268   0.2068
## al            0.032568   0.682143   0.048   0.9620
## am            0.286282   0.709456   0.404   0.6872
## ac            0.212578   0.780464   0.272   0.7857
## ap           -0.217654   0.770398  -0.283   0.7780
## ba            0.214144   0.678333   0.316   0.7527
## ce           -0.363694   0.681235  -0.534   0.5943
## df            0.075602   0.814915   0.093   0.9262
## es            0.231562   0.712872   0.325   0.7458
## go           -0.009147   0.696944  -0.013   0.9895
## ma           -0.688633   0.679249  -1.014   0.3125
## mg            0.772835   0.689782   1.120   0.2645
## ms           -0.191411   0.809232  -0.237   0.8134
## mt           -0.134494   0.697001  -0.193   0.8473
## pa            0.576608   0.678994   0.849   0.3972
## pb           -0.406140   0.690832  -0.588   0.5576
## pe           -0.414013   0.681957  -0.607   0.5448
## pi           -0.378995   0.710753  -0.533   0.5947
## pr            0.854990   0.655272   1.305   0.1942
## rj            1.073520   0.687036   1.563   0.1205
## rn           -0.120071   0.710816  -0.169   0.8661
## ro            0.137871   0.766239   0.180   0.8575
## rr            0.310703   0.771565   0.403   0.6878
## rs            0.215481   0.693917   0.311   0.7566
## sc           -0.026683   0.689345  -0.039   0.9692
## se           -0.081348   0.717027  -0.113   0.9098
## sp            0.876809   0.715900   1.225   0.2228
## toc          -0.533252   0.803236  -0.664   0.5079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##
## Residual standard error: 0.5744 on 137 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.595
## F-statistic: 7.379 on 41 and 137 DF,  p-value: < 2.2e-16

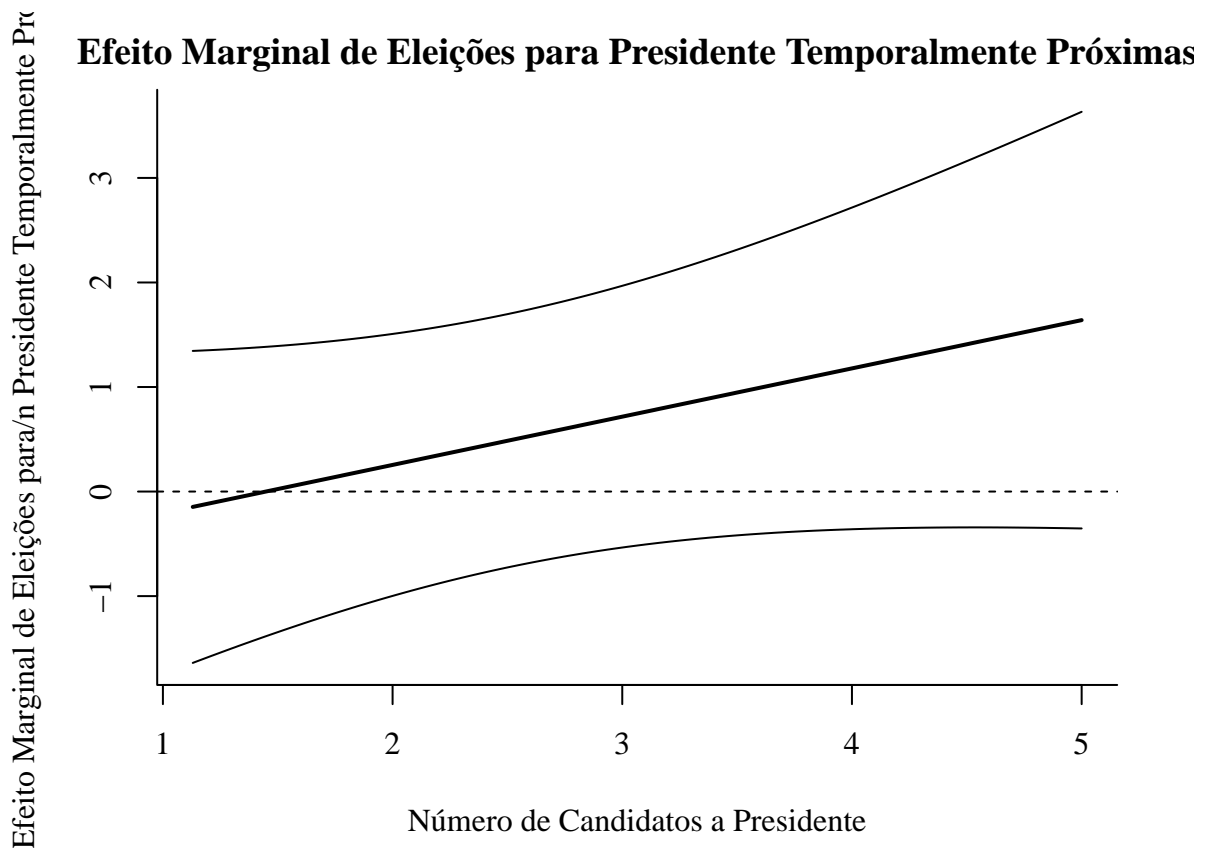
# Plotar gráfico dos efeitos marginais de proxgov sobre enlists, condicionado a engov:
beta.hat5 <- coef(reg3)
cov5 <- vcov(reg3)
z5 <- seq(min(base_samuels$engov), max(base_samuels$engov), length.out = 1000)
dy.dx5 <- beta.hat5["proxgov"] + beta.hat5["engov_proxgov"]*z5
se.dy.dx5 <- sqrt(cov5["proxgov", "proxgov"] + z5^2*cov5["engov_proxgov",
                                                         "engov_proxgov"] +
                  2*z5*cov5["proxgov", "engov_proxgov"])
upr5 <- dy.dx5 + 1.96*se.dy.dx5
lwr5 <- dy.dx5 - 1.96*se.dy.dx5
par(family="serif",bty="l",mar=c(5,5.5,2,2))
plot(x=z5, y=dy.dx5,type="n",xlim=c(min(z5),max(z5)),
     ylim=c(min(lwr5),max(upr5)),
     xlab = "Número de Candidatos a Governador",
     ylab = "Efeito Marginal de Eleições para Governador Temporalmente Próximas",
     main= "Efeito Marginal de Eleições para/n Governador Temporalmente Próximas")
lines(z5, dy.dx5, lwd = 2)
lines(z5, lwr5)
lines(z5, upr5)
abline(h=0,lty=2)
```



```

# Plotar gráfico dos efeitos marginais de proxpres sobre enlists, condicionado a empres:
beta.hat6 <- coef(reg3)
cov6 <- vcov(reg3)
z6 <- seq(min(base_samuels$enpres), max(base_samuels$enpres), length.out = 1000)
dy.dx6 <- beta.hat6["proxpres"] + beta.hat6["enpres_proxpres"]*z6
se.dy.dx6 <- sqrt(cov6["proxpres", "proxpres"] + z6^2*cov6["enpres_proxpres",
                                                             "enpres_proxpres"] +
                  2*z6*cov6["proxpres", "enpres_proxpres"])
upr6 <- dy.dx6 + 1.96*se.dy.dx6
lwr6 <- dy.dx6 - 1.96*se.dy.dx6
par(family="serif",bty="l",mar=c(5,5.5,2,2))
plot(x=z6, y=dy.dx6,type="n",xlim=c(min(z6),max(z6)),
     ylim=c(min(lwr6),max(upr6)),
     xlab = "Número de Candidatos a Presidente",
     ylab = "Efeito Marginal de Eleições para/n Presidente Temporalmente Próximas",
     main= "Efeito Marginal de Eleições para Presidente Temporalmente Próximas")
lines(z6, dy.dx6, lwd = 2)
lines(z6, lwr6)
lines(z6, upr6)
abline(h=0,lty=2)

```



Questão 5:

```
# Requerer pacotes:
```

```
require(rio)
require(ggplot2)
```

```
# Carregar base de dados:
```

```
base_hillary <- import(file = "hillary.dta")
```

```
# Checar variáveis:
```

```
names(base_hillary)
```

```
## [1] "gender"          "v960067"          "v960070"          "v960071"
## [5] "v960073"         "v960115"          "v960119"          "v960272"
## [9] "hillary_thermo"  "v960284"          "v960292"          "v960293"
## [13] "v960365"         "v960385"          "v960420"          "v960568"
## [17] "v960571"         "v960605"          "v960606"          "v960610"
## [21] "income"          "wmnmvt"           "v961300"          "religion"
## [25] "female"          "male"             "hat_hillary"      "protestant"
## [29] "catholic"        "jewish"           "other_rel"        "not_relig"
## [33] "_est_model_3"     "_est_model_4"      "_est_model_5"      "cathprot"
## [37] "wmnmvt_fem"      "_est_model_1"      "_est_model_2"
```

```
# Gerar regressão:
```

```
reghillary <- lm(hillary_thermo ~ income + jewish + income:jewish, data = base_hillary)
```

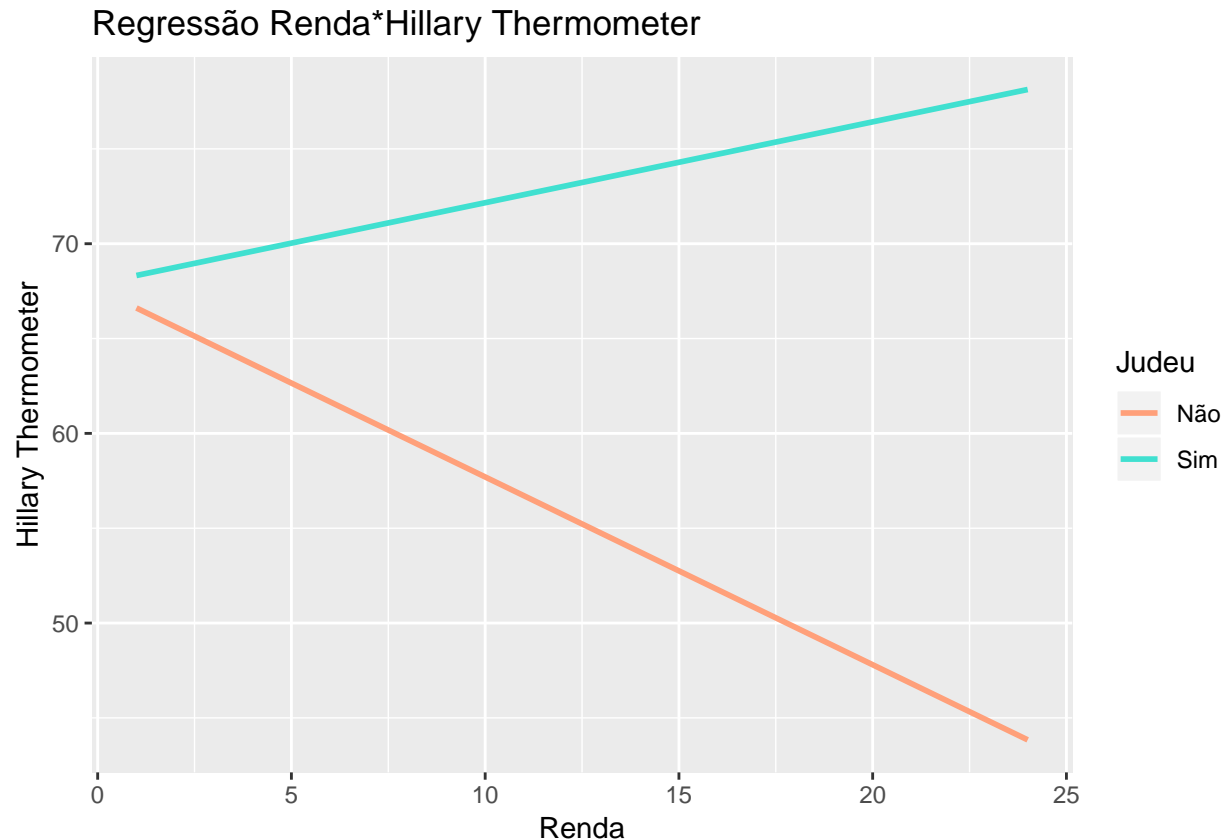
```
# Analisar resultados:
```

```
summary(reghillary)
```

```
##
## Call:
## lm(formula = hillary_thermo ~ income + jewish + income:jewish,
##     data = base_hillary)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -66.613 -18.791   4.278  21.357  56.160
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   67.6031     1.9224  35.167  <2e-16 ***
## income        -0.9901     0.1175  -8.426  <2e-16 ***
## jewish         0.2948    14.1528   0.021   0.983
## income:jewish  1.4165     0.8297   1.707   0.088 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.84 on 1538 degrees of freedom
## (172 observations deleted due to missingness)
## Multiple R-squared:  0.05053,    Adjusted R-squared:  0.04868
## F-statistic: 27.29 on 3 and 1538 DF,  p-value: < 2.2e-16
```



```
# Gerar plot:
ggplot(base_hillary,aes(y = hillary_thermo, x= income, color = factor(jewish)))+
  xlab("Renda") + ylab("Hillary Thermometer") +
  labs(title = "Regressão Renda*Hillary Thermometer",size = 10) +
  scale_colour_manual(name = "Judeu", labels = c("Não", "Sim"),
    values = c("lightsalmon", "turquoise")) +
  stat_smooth(method="lm", se = F)
```



A hipótese condicional, de que especificamente para os americanos judeus é observado um efeito da renda sobre a avaliação de Hillary Clinton, não é estatisticamente significativa (observando os resultados da regressão para a interação entre income e jewish); a hipótese nula, portanto, não pode ser rejeitada. Ao analisar o coeficiente de income, controlado para jewish, é possível observar que o resultado é estatisticamente significativo, ou seja, a renda do eleitor que não é judeu tem efeito negativo sobre a variável dependente, que é a avaliação dada por estes eleitores à Hillary Clinton (quanto maior a renda, pior a avaliação). Ademais, é possível perceber que o efeito marginal da interação entre income e jewish é representado pelo beta de income (-0.99) mais o beta da interação (1.42), que é igual a 0.43; ou seja, caso fosse significativo, o aumento de uma unidade na renda dos eleitores que são judeus seria responsável por uma variação de +0.43 unidade no termômetro de avaliação de Hillary Clinton.