Prova 2 Prática - Estatística Não-Paramétrica

Augusto Cesar Ribeiro Nunes

6 de julho de 2015

Introdução

O presente trabalho tem como produto final uma minuta do relatório de administração semestral do Instituto de Resseguros do Brasil (IRB), a partir de um estudo de caso reduzido disponibilizado pelo Professor Doutor Raul Yukihiro Matsushita. Esta tarefa faz parte da 2a Prova Prática da disciplina Métodos Estatísticos 2 (Estatística Não-Paramétrica), ministrada no 1o semestre de 2015 aos alunos de Graduação em Estatística da Universidade de Brasília.

Em particular, apresentaremos uma estimativa para a contabilização de provisão para contingências judiciais passivas do IRB, a partir de uma amostra de dados de encerramento referentes a processos cíveis, criminais, fiscais e trabalhistas onde o IRB foi réu, até o dia 31/12/2003.

Esta análise de risco é inerente à natureza de uma empresa de resseguros, que, convenientemente, "adquire" parte da carteira de riscos de seguradoras menores, usualmente quando da ocorrência de apólices de grande valor. Considera-se que os casos estudados são aqueles em que há desacordo entre a parte segurada e a IRB após um dado sinistro, desacordo que usualmente deve-se a interpretação de cobertura contratual.

Análise

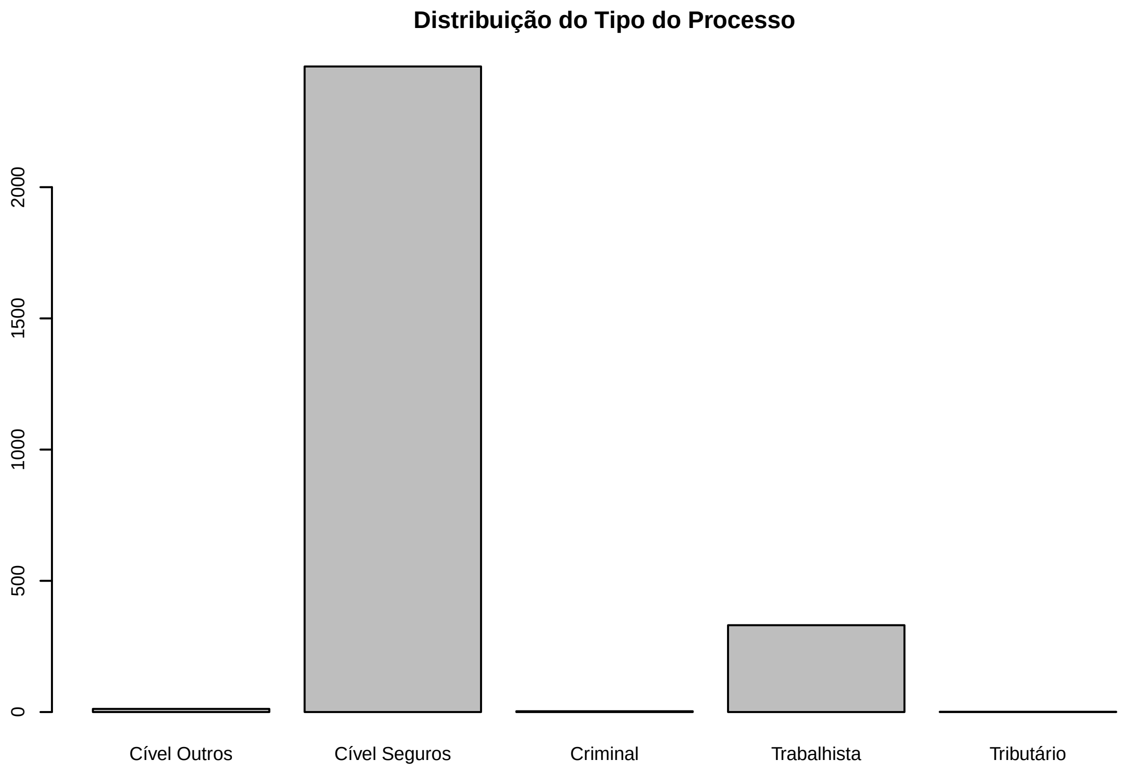
Estudo Exploratório

É de praxe iniciar um trabalho desta natureza com uma descrição exploratória do problema a ser estudado e as variáveis observadas. Temos as seguintes variáveis no conjunto de dados disponibilizado:

- **Posição** indica o estado litigioso do IRB no processo. No conjunto de dados disponibilizado, o IRB é réu para todas as observações.
- **Tipo** tipo do processo em questão: Cível, Trabalhista, Tributário ou Outros.
- **Estado** Unidade da Federação onde foi impetrado o litígio judicial.
- **VrCausa** Valor da Causa, em Reais.
- **VrPago** Valor Pago, quando aplicável. Em caso de decisão favorável ao IRB, o valor desta variável é nulo na observação.
- **Procedência** Resumo do resultado do processo, se resultou em acordo, outras hipóteses ou julgado procedente/improcedente em decisão monocrática.
- **Tempo** Tempo, em meses, decorrido da autuação do processo até a decisão monocrática.

Variável Tipo do Processo

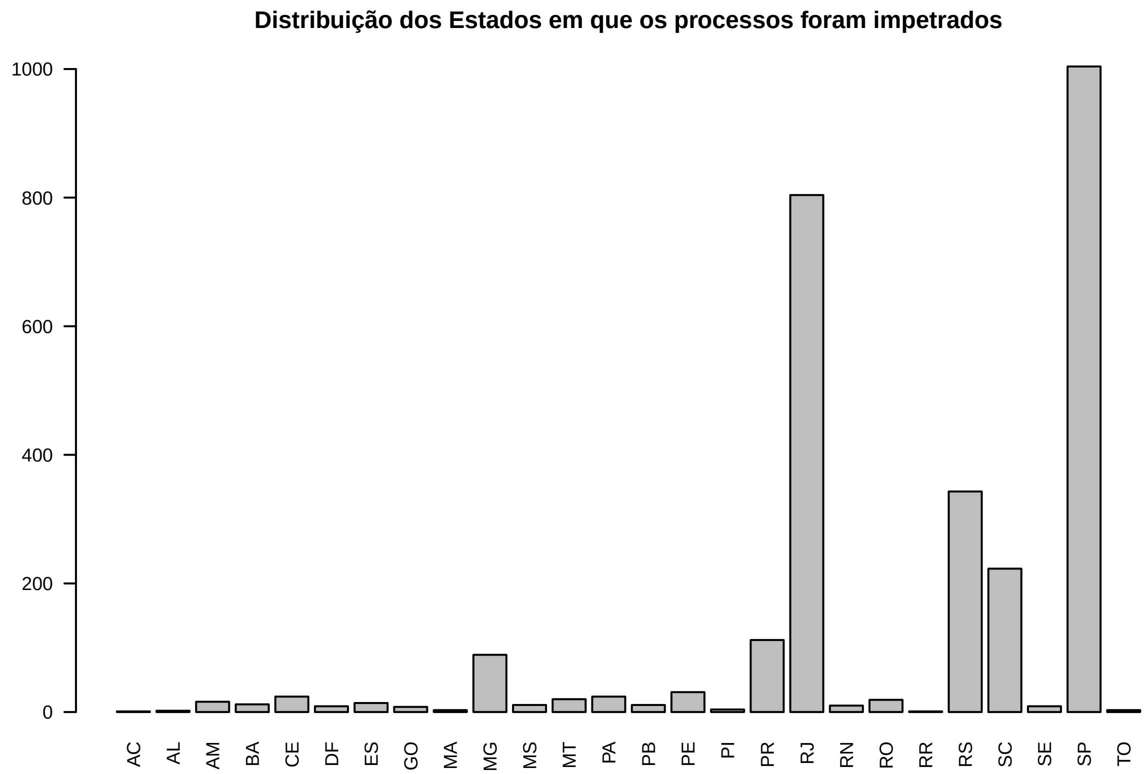
```
##  Cível Outros Cível Seguros      Criminal   Trabalhista    Tributário
##            12          2460             3           331             1
```



Nota-se uma maioria esmagadora de processos cíveis referentes a seguros.

Variável Estado

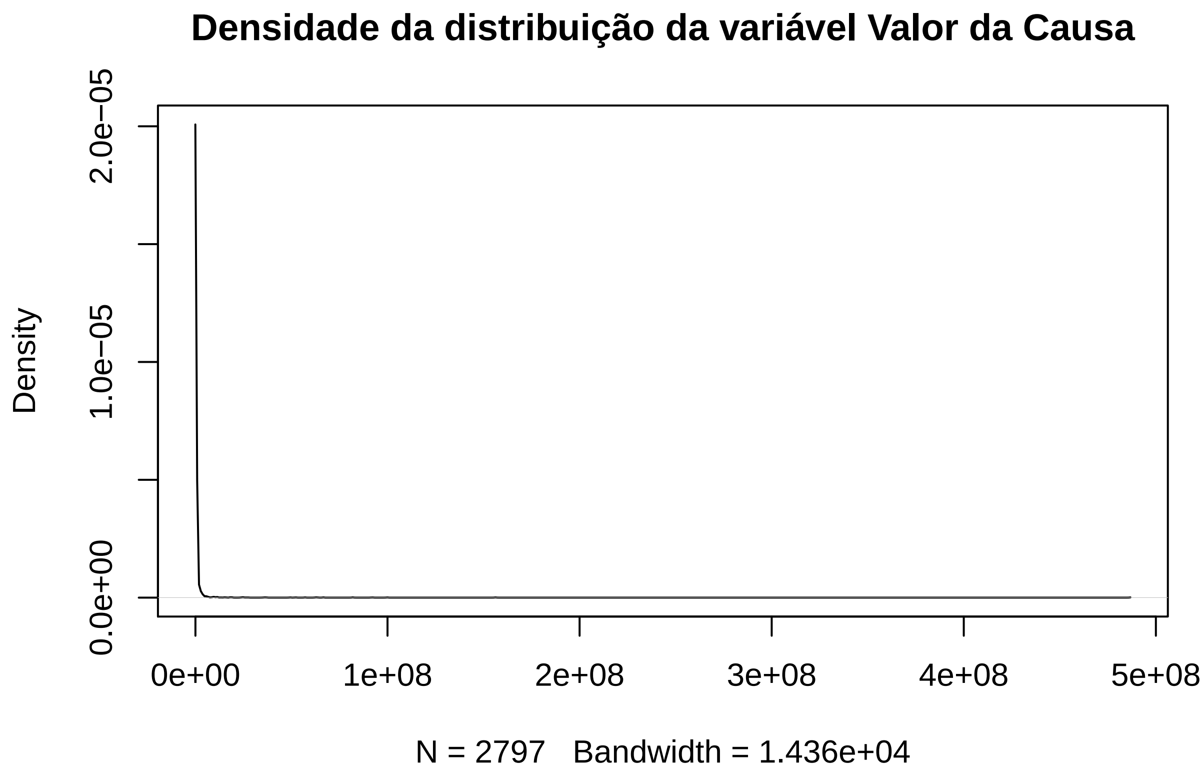
```
##   AC   AL   AM   BA   CE   DF   ES   GO   MA   MG   MS   MT   PA   PB   PE
##    1    2   16   12   24    9   14    8    3   89   11   20   24   11   31
##   PI   PR   RJ   RN   RO   RR   RS   SC   SE   SP   TO
##    4  112  804   10   19    1  343  223    9 1004    3
```

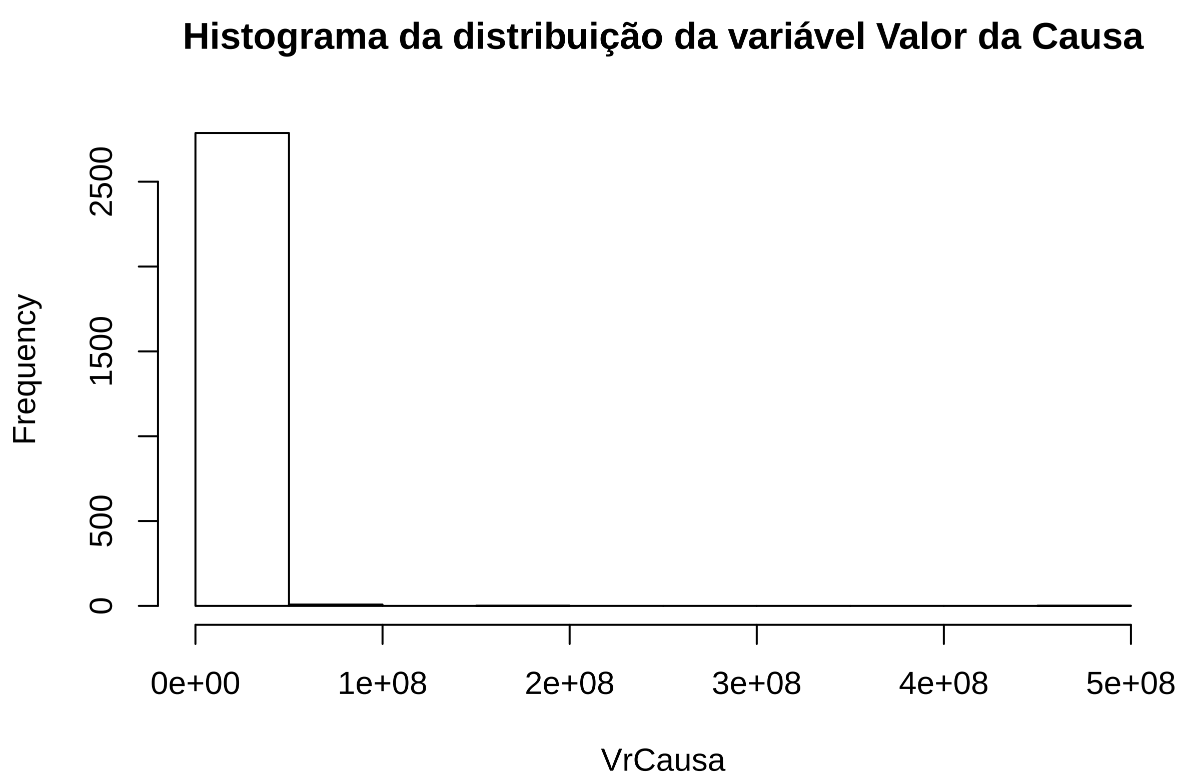


Como era de se esperar, temos ampla concentração de litígios em Estados onde há maior atividade econômica, como SP, RJ e MG.

Variável Valor da Causa

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##         0      1520     15960    790600    106100 486700000        10
```


Densidade da distribuição da variável Valor da Causa
Density
0.0e+00
1.0e−05
2.0e−05
0e+00
1e+08
2e+08
3e+08
4e+08
5e+08
N = 2797 Bandwidth = 1.436e+04


Histograma da distribuição da variável Valor da Causa
Frequency
0
500
1500
2500
0e+00
1e+08
2e+08
3e+08
4e+08
5e+08
VrCausa

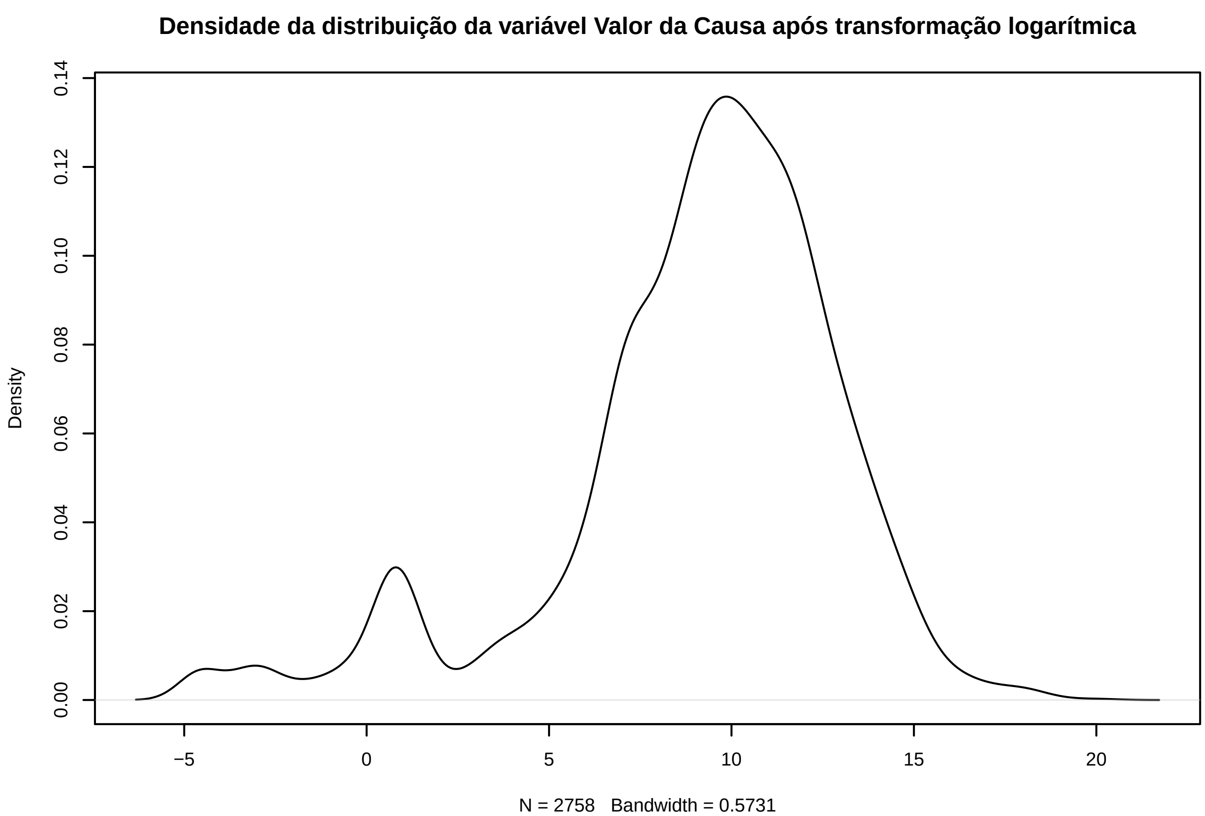
Note que tanto as estatísticas de ordem quanto os gráficos de Densidade e Histograma da variável sugerem que sua distribuição segue um modelo de cauda pesada, uma *Power Law* ou Lei de Potências. Basta atentar para o máximo (486700000), que é cerca de 4587 vezes maior que a estatística do 3o quartil (106100).

Podemos então aplicar uma transformação (monótona) logarítmica nesta variável, chamada de **logVrCausa** no conjunto de dados, cuja distribuição é a seguinte.

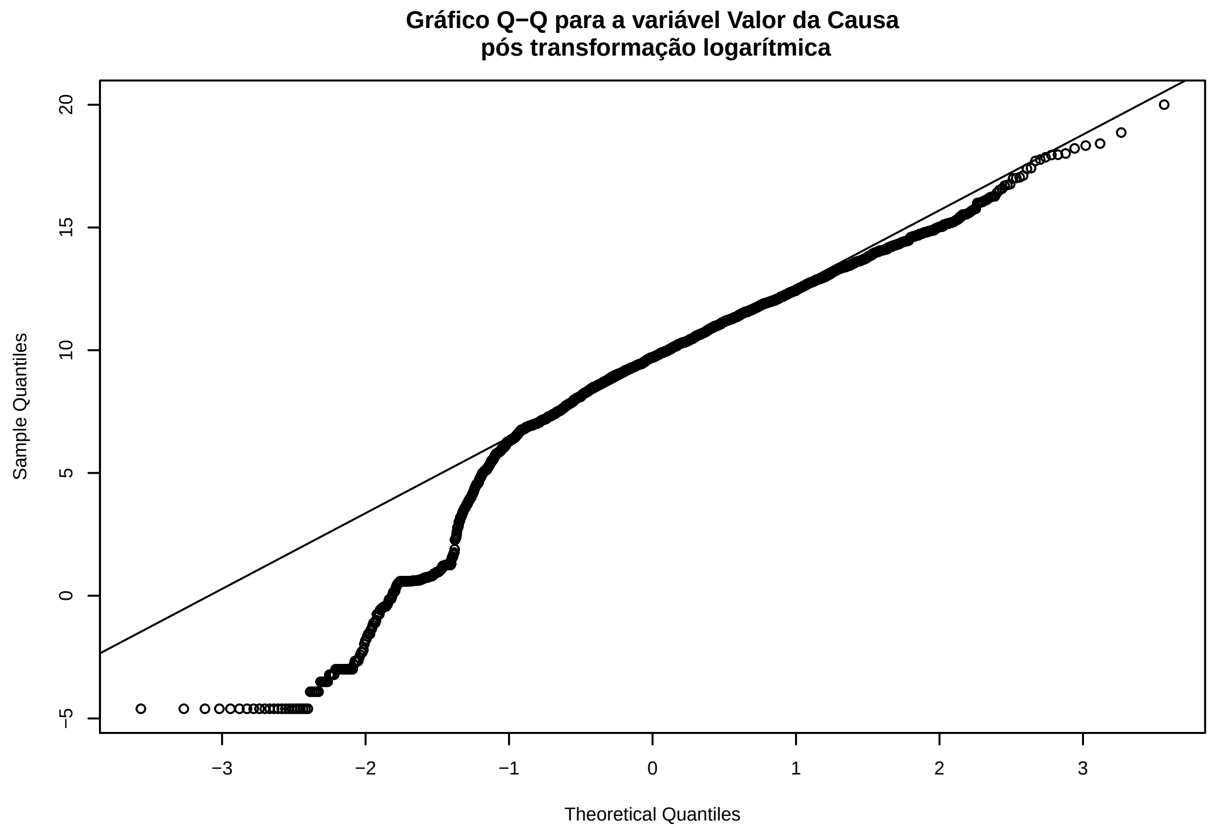
```
logVrCausa = log(na.exclude(VrCausa))
logVrCausa[mapply(is.infinite, logVrCausa)] <- NA
summary(logVrCausa)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  -4.605   7.452   9.710   9.082  11.610  20.000      39
```

```
par(cex=0.6)
plot(density(na.exclude(logVrCausa)), main="Densidade da distribuição da variável Valor da Causa após t
```




```
qqnorm(logVrCausa, main="Gráfico Q-Q para a variável Valor da Causa\n pós transformação logarítmica")
qqline(logVrCausa)
```




```r
fitdistr(na.exclude(logVrCausa), "normal")
```

```
##       mean           sd
##   9.08203565   3.94234679
##  (0.07506849) (0.05308144)
```

```r
ks.test(logVrCausa, "pnorm")
```

```
## Warning in ks.test(logVrCausa, "pnorm"): ties should not be present for the
## Kolmogorov-Smirnov test

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  logVrCausa
## D = 0.91056, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```r
shapiro.test(logVrCausa)
```

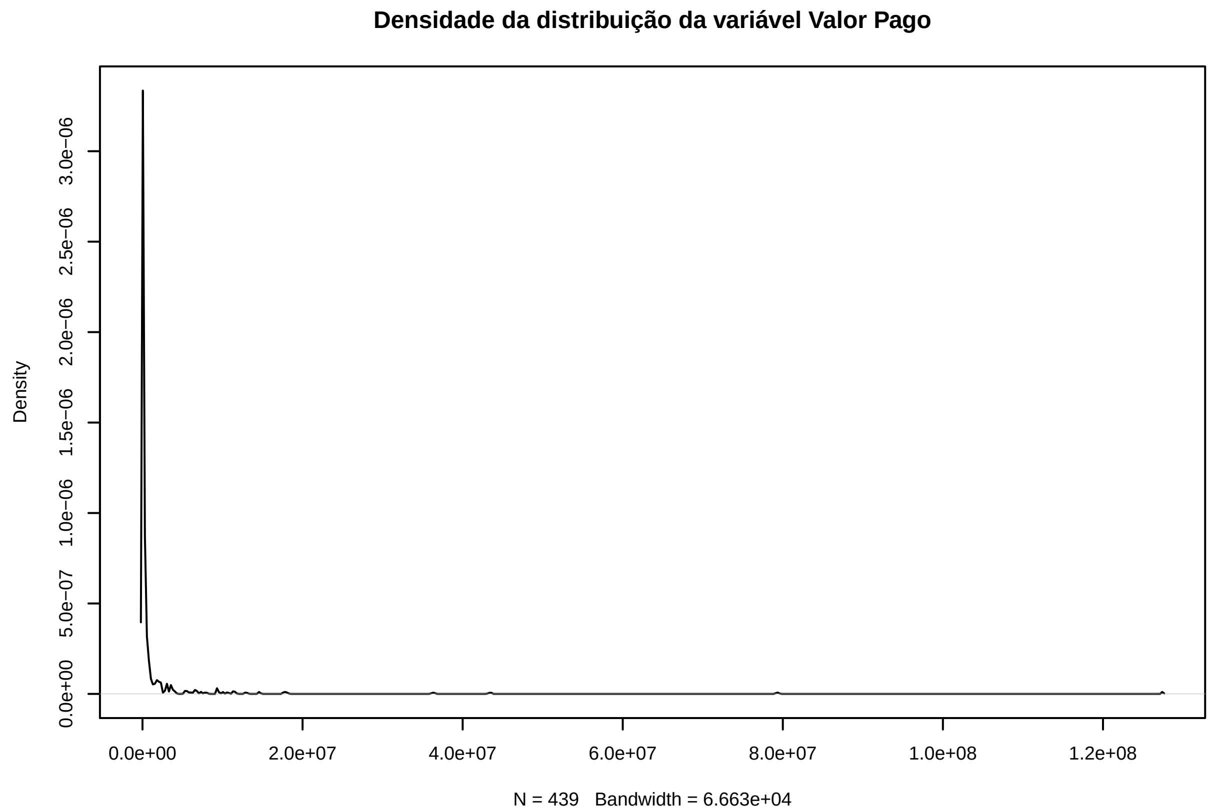
```
##
##  Shapiro-Wilk normality test
##
## data:  logVrCausa
## W = 0.92249, p-value < 2.2e-16
```

Note que após a transformação a variável deixa de assumir comportamento de distribuição com cauda pesada. O gráfico Q-Q sugere que ainda assim a distribuição da variável após a transformação é não-Normal. Na verdade nem seria necessário o gráfico QQ, pois a própria densidade da variável nos mostra uma distribuição bimodal.

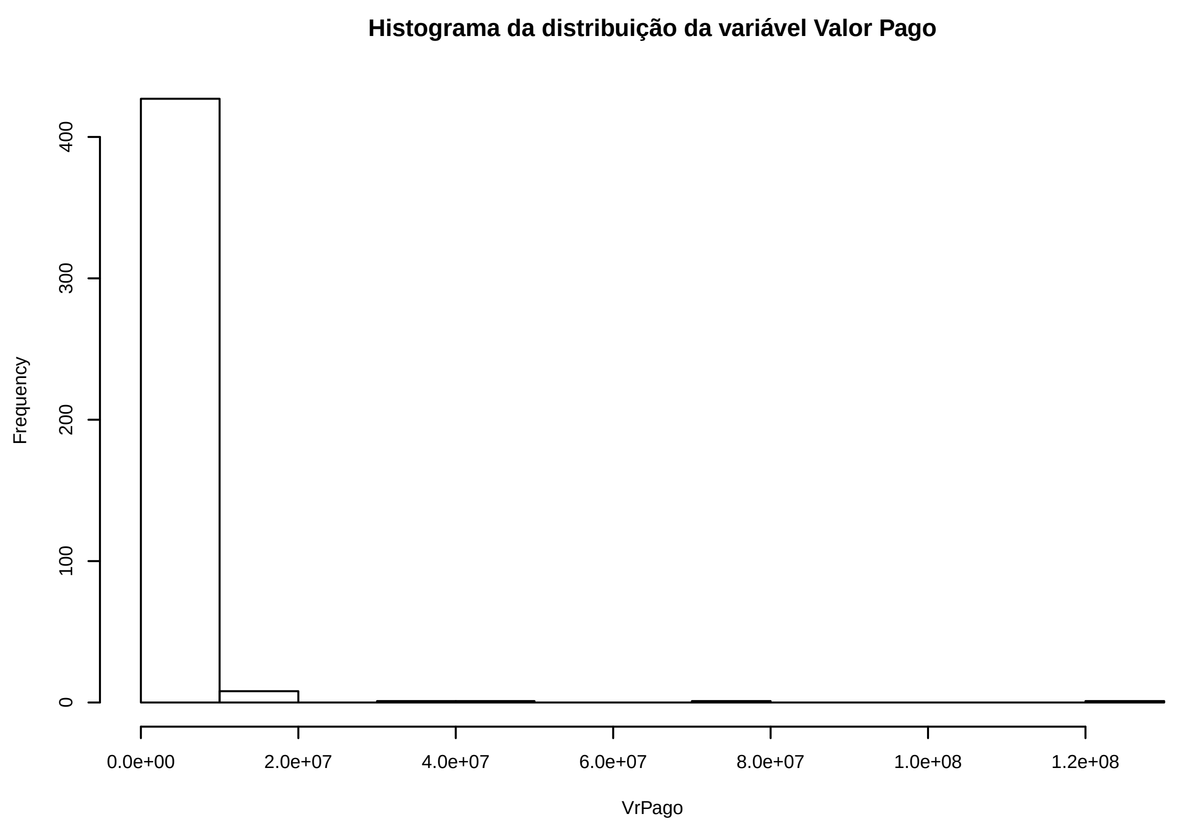
Forçando a barra e ajustando a partir da máxima verossimilhança uma distribuição Normal a partir da variável transformada, a função *fitdistr* dá uma estimativa com $\mu = 9.08203565 \pm 0.07506849$ e $\sigma^2 = 3.94234679 \pm 0.05308144$. Entretanto, os testes de Kolmogorov-Smirnov e Shapiro-Wilk **REJEITAM** a hipótese de Normalidade desta variável.

Variável Valor Pago

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##         5     18440     78280   1414000    353400 127400000      2368
```



Densidade da distribuição da variável Valor Pago
Density
0.0e+00
5.0e−07
1.0e−06
1.5e−06
2.0e−06
2.5e−06
3.0e−06
0.0e+00
2.0e+07
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08
N = 439   Bandwidth = 6.663e+04



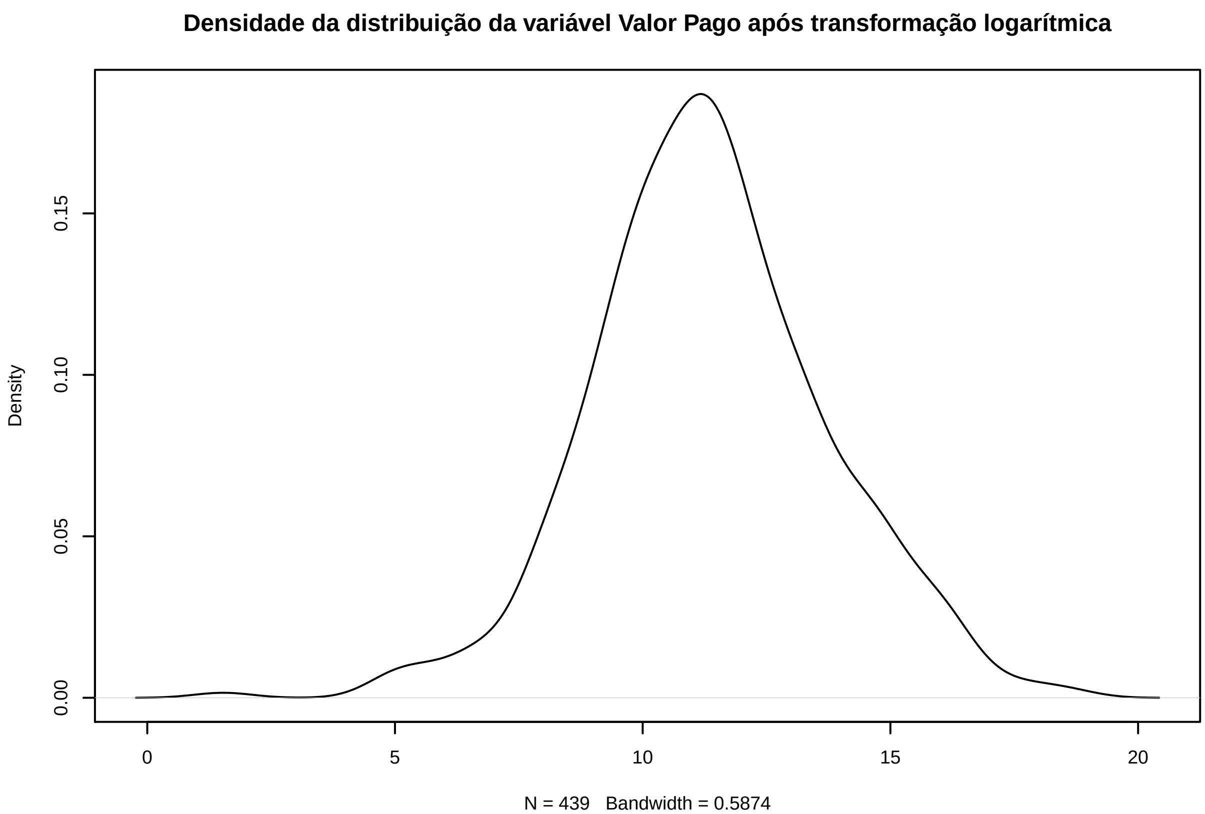
Histograma da distribuição da variável Valor Pago
Frequency
0
100
200
300
400
0.0e+00
2.0e+07
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08
VrPago

Similarmente ao observado no Valor da Causa, o Valor Pago no processo também segue uma distribuição de cauda longa, ou uma Lei de Potência. Aplicando a transformação logarítmica e atribuindo os valores à variável **logVrPago**, temos:

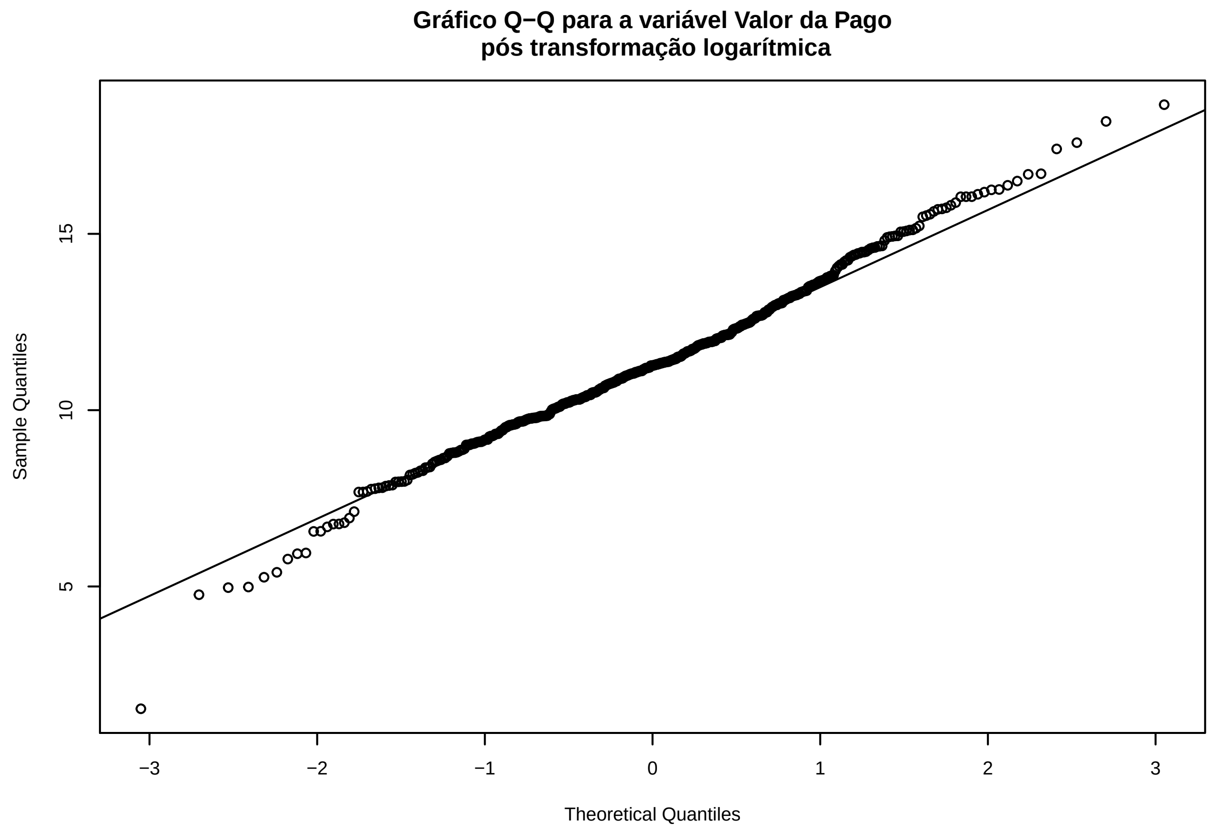
```
logVrPago = log(na.exclude(VrPago))
logVrPago[mapply(is.infinite, logVrPago)] <- NA
summary(logVrPago)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.533   9.822  11.270  11.340  12.780  18.660
```

```
par(cex=0.6)
plot(density(na.exclude(logVrPago)), main="Densidade da distribuição da variável Valor Pago após transf
```




```
qqnorm(logVrPago, main="Gráfico Q-Q para a variável Valor da Pago\n pós transformação logarítmica")
qqline(logVrPago)
```




```r
fitdistr(na.exclude(logVrPago), "normal")
```

```
##       mean            sd
##   11.34039334    2.37292628
##  ( 0.11325359) ( 0.08008238)
```

```r
ks.test(logVrPago, "pnorm")
```

```
## Warning in ks.test(logVrPago, "pnorm"): ties should not be present for the
## Kolmogorov-Smirnov test

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  logVrPago
## D = 0.99772, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

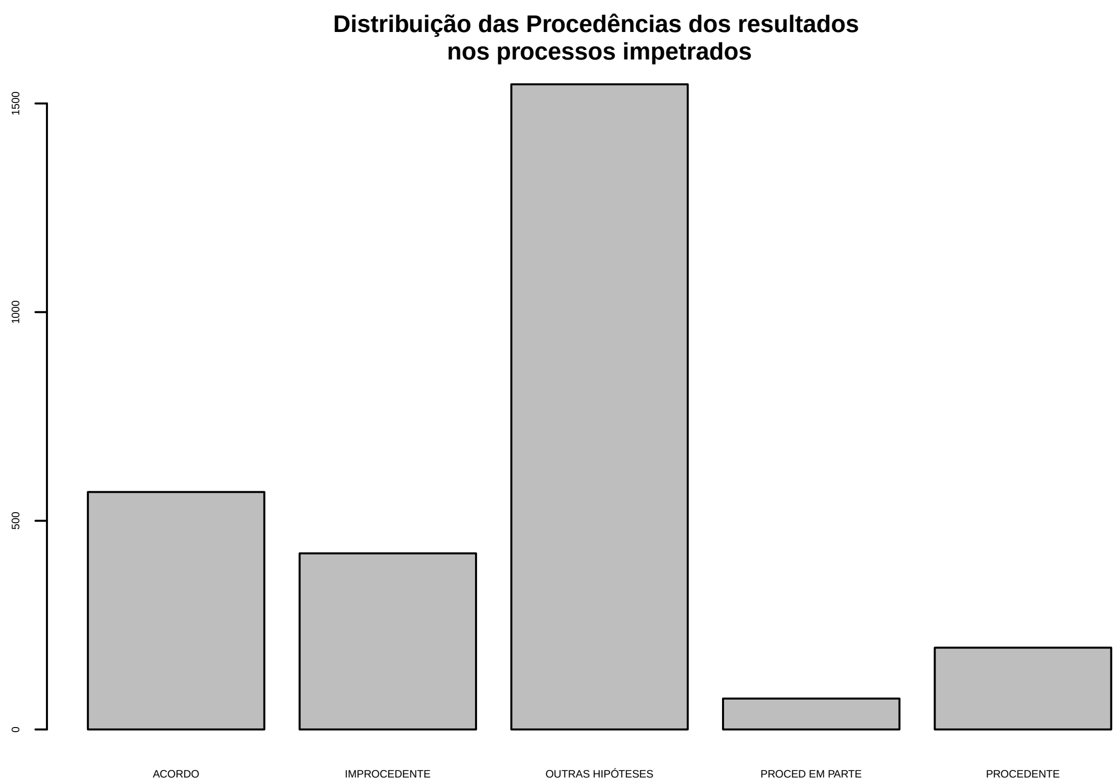
```r
shapiro.test(logVrPago)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  logVrPago
## W = 0.99215, p-value = 0.02066
```

Agora sim temos um melhor ajuste à distribuição Normal da variável transformada, justificado pelo gráfico de densidade e histograma, bem como pelo gráfico Q-Q. O ajuste de Máxima-Verossimilhança a partir de uma distribuição Normal hipótetica nos dá a seguinte estimativa para os parâmetros: $\mu = 11.34039334 \pm 0.11325359$ e $\sigma^2 = 2.37292628 \pm 0.08008238$. Entretanto, os testes de Kolmogorov-Smirnov e Shapiro-Wilk **REJEITAM** a hipótese de Normalidade desta variável.

Variável Procedência

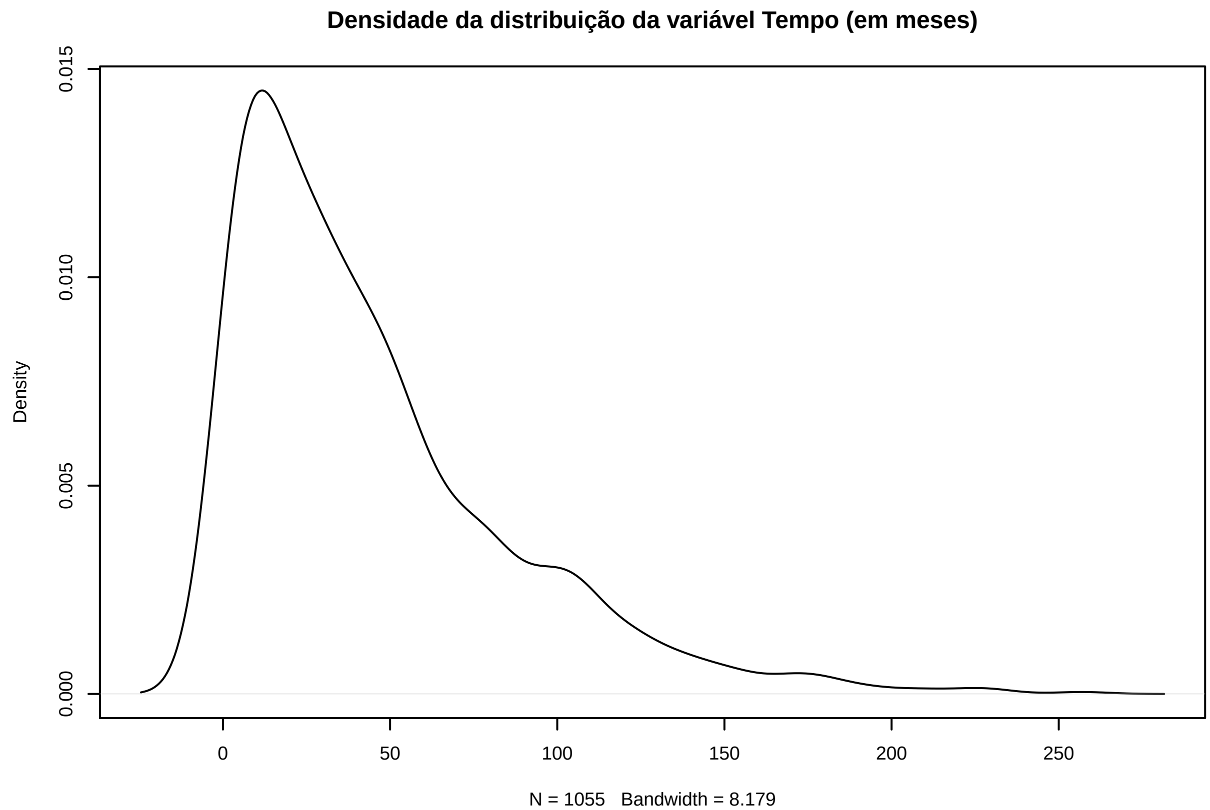
```
##           ACORDO     IMPROCEDENTE OUTRAS HIPÓTESES  PROCED EM PARTE
##              569              422             1546               74
##       PROCEDENTE
##              196
```

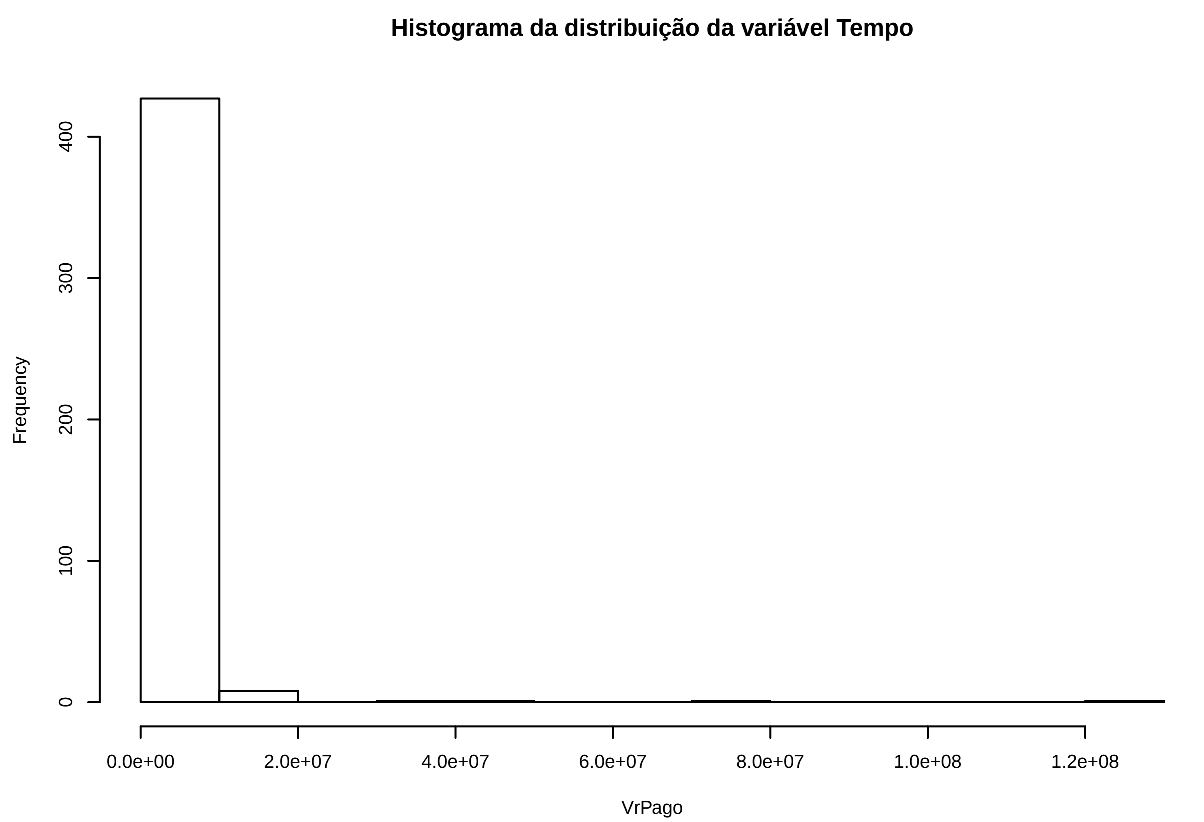


Nota-se prevalência de processos que resultaram em "Outras Hipóteses" (que seriam quais?), seguida de acordos, processos considerados improcedentes, e finalmente, procedentes em sua totalidade ou parte.

Variável Tempo

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   13.00   33.00   44.67   62.00  257.00    1752
```


Densidade da distribuição da variável Tempo (em meses)
Density
0.000
0.005
0.010
0.015
0
50
100
150
200
250
N = 1055   Bandwidth = 8.179


Histograma da distribuição da variável Tempo
Frequency
0
100
200
300
400
0.0e+00
2.0e+07
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08
VrPago

```
##        rate
##   0.0223848928
##  (0.0006891738)
```

Aqui vemos um retrato da morosidade do Judiciário no Brasil, processos que demoraram até 250 meses (mais de 20 anos) para chegarem a uma decisão. O gráfico de densidade sugere uma distribuição assimétrica, mas não necessariamente com cauda longa, de valores não nulos, então podemos ajustar a distribuição do tempo a uma distribuição Exponencial. O ajuste de máxima verossimilhança nos dá a estimativa $\lambda = 0.0223848928 \pm 0.0006891738$ para o parâmetro de taxa da distribuição Exponencial empírica. Note que $Exp(\lambda) = \Gamma(1, \lambda)$. O teste de Kolmogorov-Smirnov para uma amostra no entanto rejeita a hipótese nula com alto nível de significância.

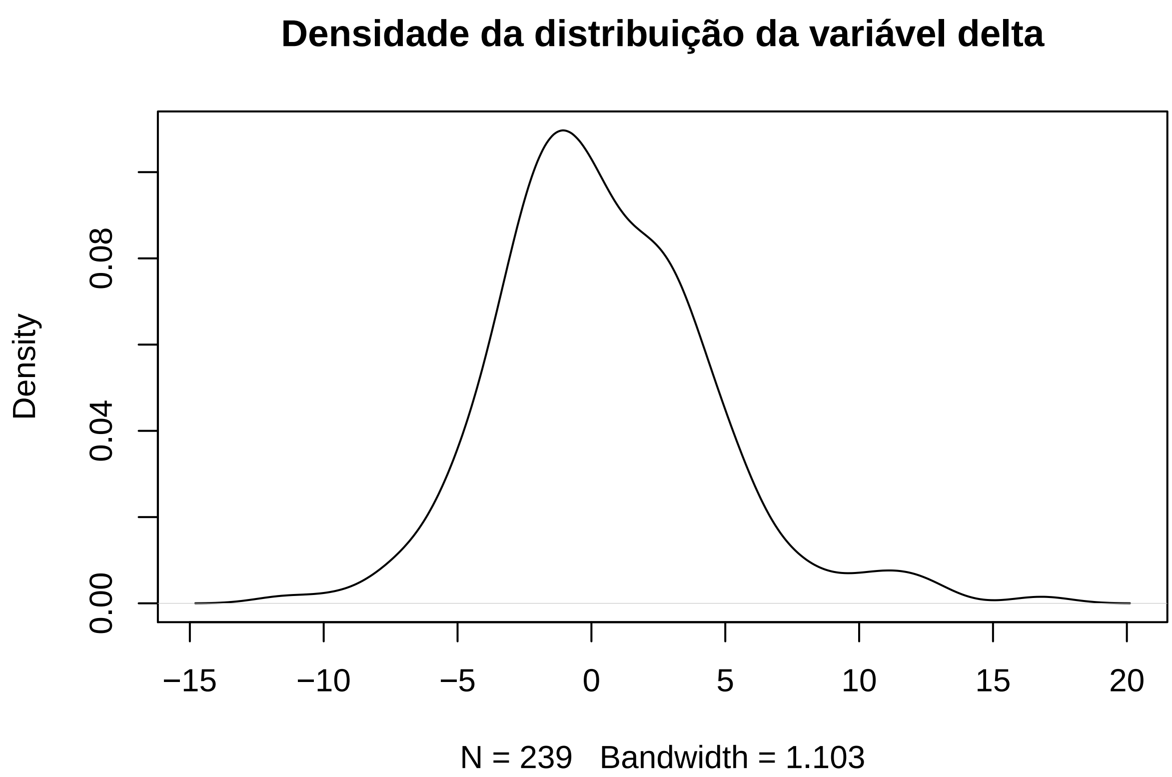
Distribuição da variável diferença entre o valor pago e o valor da causa

Criaremos a variável auxiliar delta = logVrPago - logVrCausa.

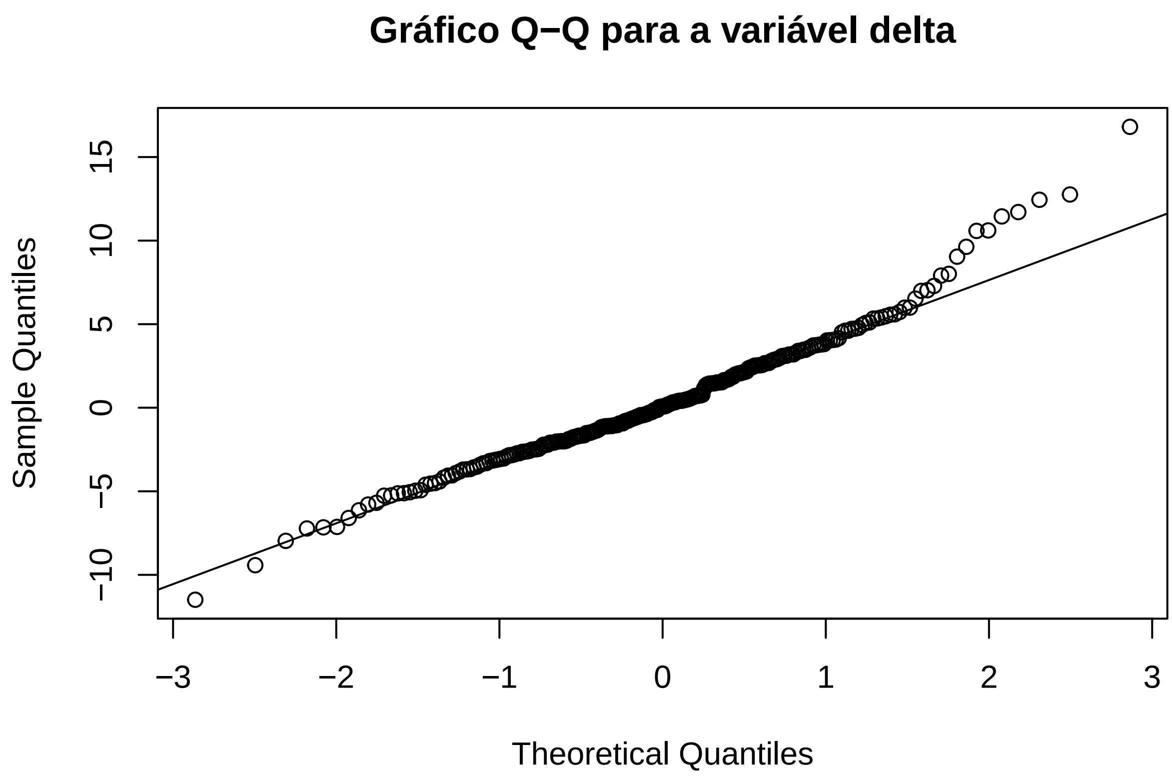
```
delta = logVrPago[VrPago != 0] - logVrCausa[VrPago != 0]
delta[mapply(is.infinite, delta)] <- NA
summary(delta)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## -11.4900  -2.0900   0.0805   0.4534   2.8190  16.8100     2568
```

```
plot(density(na.exclude(delta)), main="Densidade da distribuição da variável delta")
```



```r
qqnorm(delta, main="Gráfico Q-Q para a variável delta")
qqline(delta)
```




```r
fitdistr(na.exclude(delta), "normal")
```

```
##      mean          sd    
##   0.4534014   3.9994362 
##  (0.2587020) (0.1829300)
```

```r
shapiro.test(delta)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  delta
## W = 0.97272, p-value = 0.0001471
```

Note que o gráfico nos dá indícios de que a distribuição desta variável é normal, dada sua simetria. O ajuste usando máxima verossimilhança supondo a distribuição normal estima $\mu = 0.8538448 \pm 0.2597719$ e $\sigma^2 = 4.0159756 \pm 0.1836864$. Entretato, os testes de Kolmogorov-Smirnov e Shapiro-Wilk rejeitam a hipótese de normalidade com considerável significância.

Estimativa da Probabilidade de um processo durar mais de 260 meses

Utilizaremos um argumento de máxima verossimilhança. Supondo que a duração do processo é uma realização de uma variável Bernoulli com sucesso quando o processo dura mais de 260 meses, e sabendo pelo Princípio da verossimilhança que a mesma resume toda a informação sobre o parâmetro dada pela amostra, criamos uma variável indicadora auxiliar **nProcLongos** que é igual a um quando o processo durou mais de 260 meses (não estrito), e nula caso contrário.

```
nProcLongos = sum(na.exclude(tempo[tempo>260]))
nProcLongos
```

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

Note que, por este argumento, um processo pode durar mais de 260 meses com probabilidade nula. Utilizando a distribuição empírica ajustada pela verossimilhança acima, podemos estimar essa probabilidade calculando $P(X > 260)$, com $X \sim Exp(\lambda = 0.0223848928)$, que nos dá 6.6424732×10^{-5}, muito próximo de zero.

Estimativa para a probabilidade de o valor a ser pago em um processo superar 30 mil vezes o valor da causa

Como supomos que a distribuição da variável delta é normal, podemos simplesmete obter esta estimativa calculando $P(\frac{logVrPago}{logVrCausa} > 10.30895)$. Como supomos que logVrPago e logVrCausa seguem distribuições Normais, a distribuição de sua razão é Cauchy. Então $P(\frac{logVrPago}{logVrCausa} > 10.30895) = 0.0617349$.

Estimativa da probabilidade de o valor a ser pago em um processo superar a quantia de R$ 130.000.000,00

Por raciocínio similar ao feito para a estimativa do processo durar mais de 260 meses, podemos usar um argumento de verossimilhança e obter quantas vezes, na amostra, o valor a ser pago ultrapassa o estipulado.

```
nValoresAltos = sum(na.exclude(VrPago[VrPago>130000000]))
nValoresAltos
```

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

O que poderia nos levar à conclusão de que o evento é impossível. No entanto, como trata-se de distribuição de cauda longa, o comportamento nos extremos é inesperado, e então deveríamos estudar o comportamento da distribuição testada nos valores extremos. como $log 130000000 = 18.683045$, e supomos que $logVrPago \sim Normal(\mu = 11.34039334, \sigma^2 = 2.37292628)$, podemos calcular $P(logVrPago > 18,68305) = 0.001401$.

Estudo acerca das probabilidades de ocorrências dos eventos [Y>X] e [Y<X]

Vamos criar duas novas variáveis: CausamqPago, e PagomqCausa.

```
CausamqPago = sum(na.exclude(logVrCausa > logVrPago))
```

```
## Warning in logVrCausa > logVrPago: comprimento do objeto maior não é
## múltiplo do comprimento do objeto menor
```

```r
CausamqPago
```

```
## [1] 873
```

```r
PagomqCausa = sum(na.exclude(logVrCausa < logVrPago))
```

```
## Warning in logVrCausa < logVrPago: comprimento do objeto maior não é
## múltiplo do comprimento do objeto menor
```

```r
PagomqCausa
```

```
## [1] 1885
```

Ou seja, em 31.1008194% dos casos o valor da Causa é maior que o valor Pago, e em 67.1535447% dos casos o Valor Pago é maior que o valor da Causa.

Teste de dependência entre o Tipo do Processo e as variáveis tempo, Valor Pago, Valor da Causa e delta

```r
chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" | TIPO == "Trabalhista"], dados$t
```

```
## Warning in chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível
## Outros" | : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"] and dados$tempo[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"]
## X-squared = 349.95, df = 328, p-value = 0.1936
```

```r
chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" | TIPO == "Trabalhista"], logVrCa
```

```
## Warning in chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível
## Outros" | : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"] and logVrCausa[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"]
## X-squared = 5449.8, df = 4974, p-value = 1.801e-06
```

```r
chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" | TIPO == "Trabalhista"], logVrPa
```

```
## Warning in chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível
## Outros" | : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"] and logVrPago[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"]
## X-squared = 439, df = 433, p-value = 0.4108
```

```r
chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" | TIPO == "Trabalhista"], delta[T
```

```
## Warning in chisq.test(dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível
## Outros" | : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dados$TIPO[TIPO == "Cível Seguros" | TIPO == "Cível Outros" |
##     TIPO == "Trabalhista"] and delta[TIPO == "Cível Seguros" | TIPO == "Cível Outros" | TIPO ==
##     "Trabalhista"]
## X-squared = 239, df = 238, p-value = 0.4696
```

Não há dependência entre o tipo de processo e a variável tempo Não há dependência entre o tipo de processo e a variável log-Valor da Causa *Não há dependência entre o tipo de processo e a variável log-Valor Pago* Não há dependência entre o tipo de processo e a variável delta

Testar se as distribuições em SP diferem das no RJ

```r
ks.test(tempo[Estado == "RJ"], tempo[Estado == "SP"])
```

```
## Warning in ks.test(tempo[Estado == "RJ"], tempo[Estado == "SP"]): p-value
## will be approximate in the presence of ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  tempo[Estado == "RJ"] and tempo[Estado == "SP"]
## D = 0.17635, p-value = 0.0001254
## alternative hypothesis: two-sided
```

```r
ks.test(VrCausa[Estado == "RJ"], VrCausa[Estado == "SP"])
```

```
## Warning in ks.test(VrCausa[Estado == "RJ"], VrCausa[Estado == "SP"]): p-
## value will be approximate in the presence of ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  VrCausa[Estado == "RJ"] and VrCausa[Estado == "SP"]
## D = 0.27827, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
ks.test(VrPago[Estado == "RJ"], VrPago[Estado == "SP"])
```

```
## Warning in ks.test(VrPago[Estado == "RJ"], VrPago[Estado == "SP"]): p-value
## will be approximate in the presence of ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  VrPago[Estado == "RJ"] and VrPago[Estado == "SP"]
## D = 0.067932, p-value = 0.9618
## alternative hypothesis: two-sided
```

```
ks.test(delta[Estado == "RJ"], delta[Estado == "SP"])
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  delta[Estado == "RJ"] and delta[Estado == "SP"]
## D = 0.23932, p-value = 0.1235
## alternative hypothesis: two-sided
```

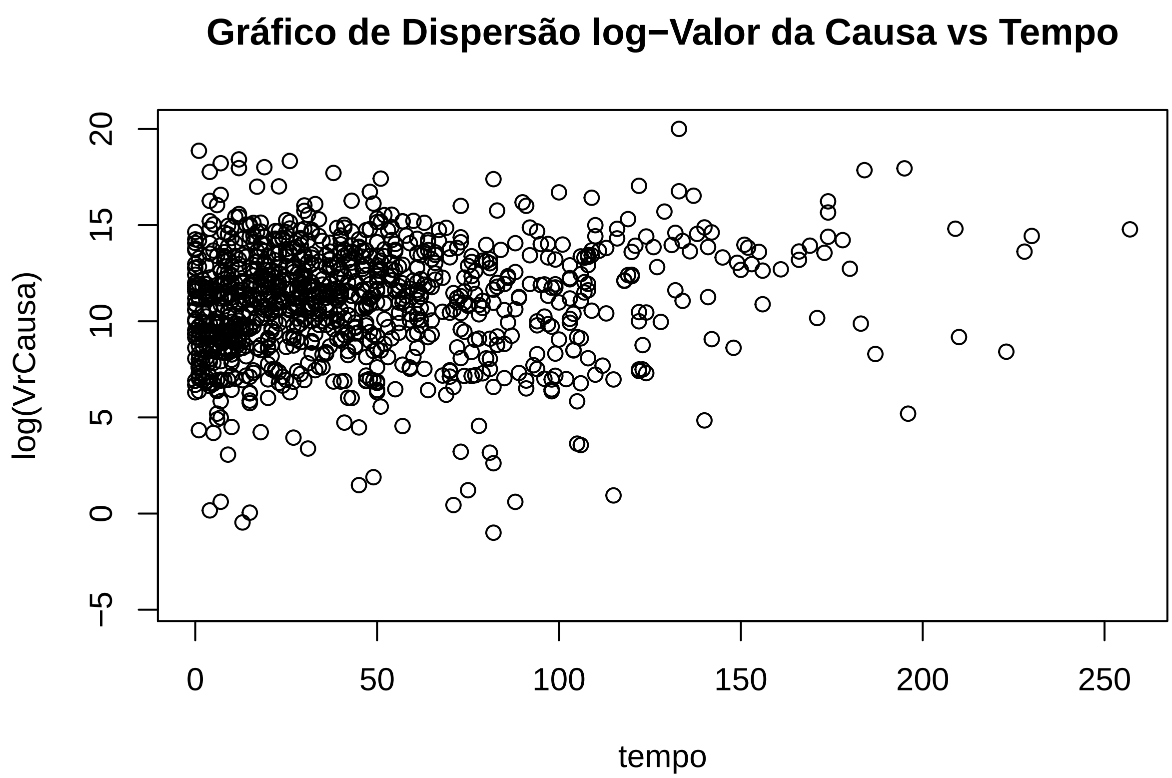
As seguintes distribuições são diferentes em SP e RJ:

- Tempo: p-valor = 0.0001254
- Valor da Causa: p-valor ~ 0

Por outro lado, o Valor Pago (p-valor = 0.9618), e o delta (p-valor = 0.7147) têm mesma distribuição em RJ e SP com alta significância estatística.

Estruturas de dependência

```
plot(tempo,log(VrCausa), main = "Gráfico de Dispersão log-Valor da Causa vs Tempo")
```




```r
cor.test(tempo, log(VrCausa))
```

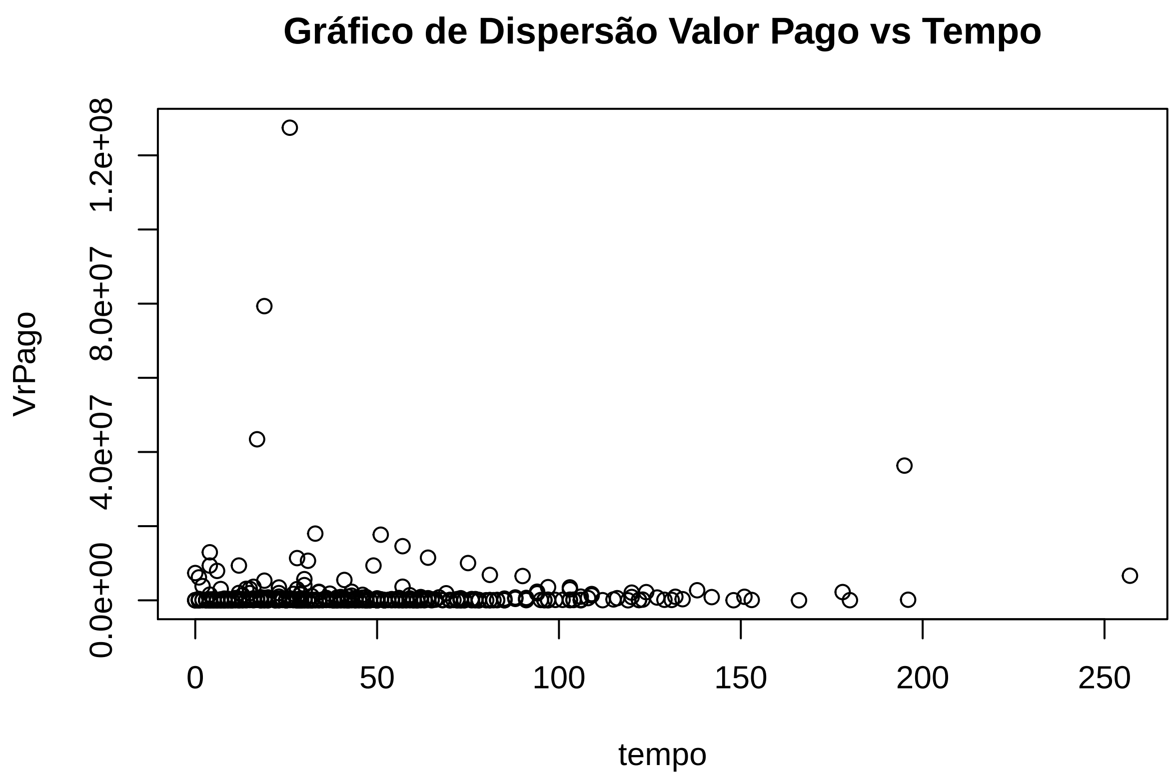
```
##
##  Pearson's product-moment correlation
##
## data:  tempo and log(VrCausa)
## t = 3.8922, df = 1051, p-value = 0.0001056
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.0592175 0.1783311
## sample estimates:
##       cor
## 0.1192032
```

```r
hoeffd(tempo, VrCausa)
```

```
## D
##      x    y
## x 1.00 0.01
## y 0.01 1.00
##
## avg|F(x,y)-G(x)H(y)|
##        x      y
## x 0.0000 0.0134
```

```
## y 0.0134 0.0000
##
## max|F(x,y)-G(x)H(y)|
##        x      y
## x 0.0000 0.0424
## y 0.0424 0.0000
##
## n
##      x    y
## x 1055 1053
## y 1053 2797
##
## P
##   x  y
## x    0
## y  0
```

```
plot(tempo,VrPago, main = "Gráfico de Dispersão Valor Pago vs Tempo")
```




```
cor.test(tempo, log(VrPago))
```

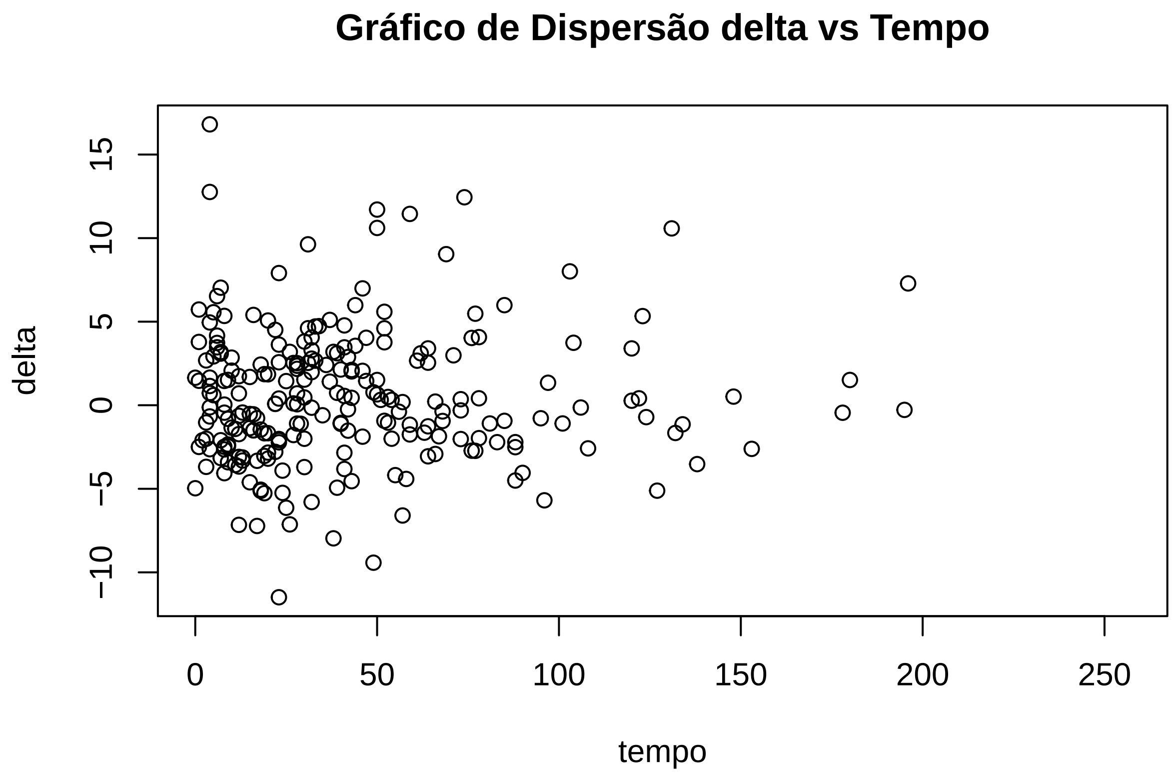
```
##
##  Pearson's product-moment correlation
##
## data:  tempo and log(VrPago)
```

```
## t = 3.5348, df = 437, p-value = 0.0004517
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.07429365 0.25631595
## sample estimates:
##       cor
## 0.1667249
```

```r
hoeffd(tempo, VrPago)
```

```
## D
##   x y
## x 1 0
## y 0 1
##
## avg|F(x,y)-G(x)H(y)|
##       x     y
## x 0.000 0.013
## y 0.013 0.000
##
## max|F(x,y)-G(x)H(y)|
##       x     y
## x 0.000 0.035
## y 0.035 0.000
##
## n
##      x   y
## x 1055 439
## y  439 439
##
## P
##   x      y
## x        0.0049
## y 0.0049
```

```r
plot(tempo,delta, main = "Gráfico de Dispersão delta vs Tempo")
```




```r
cor.test(tempo, delta)
```

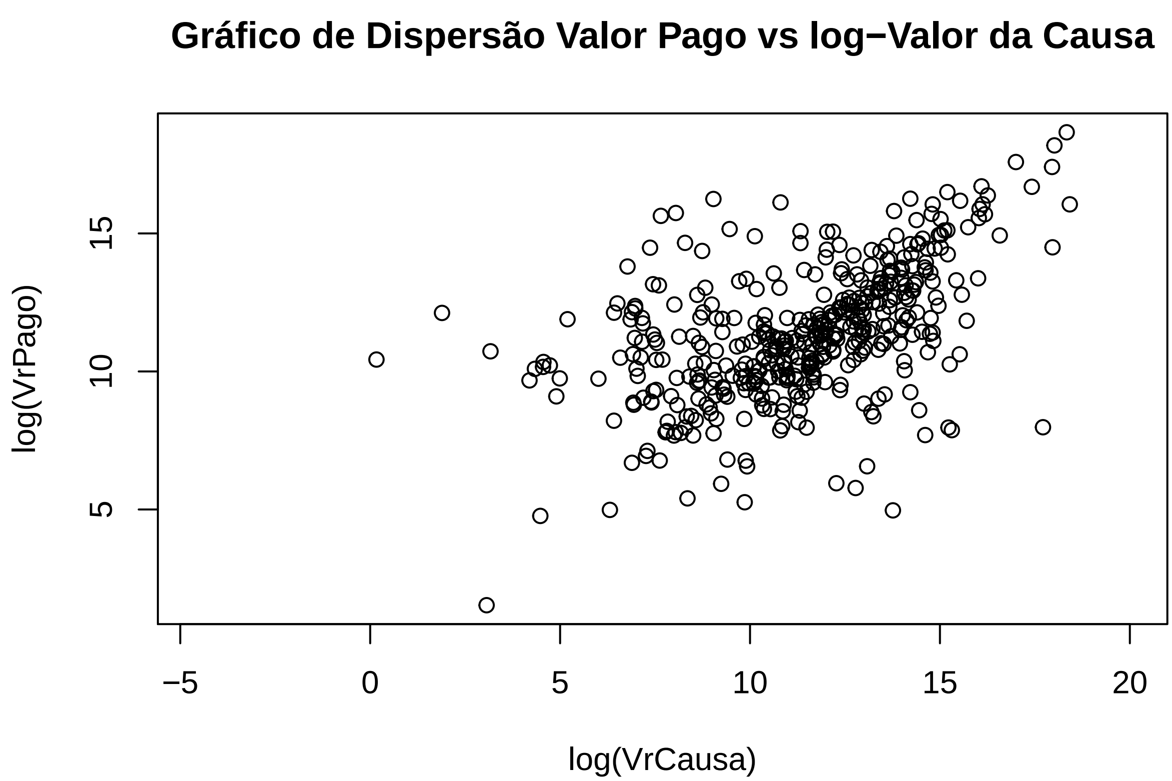
```
##
##  Pearson's product-moment correlation
##
## data:  tempo and delta
## t = 0.60635, df = 237, p-value = 0.5449
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.08797812  0.16542515
## sample estimates:
##        cor
## 0.03935628
```

```r
hoeffd(tempo, delta)
```

```
## D
##   x y
## x 1 0
## y 0 1
##
## avg|F(x,y)-G(x)H(y)|
##        x      y
## x 0.0000 0.0095
```

```
## y 0.0095 0.0000
##
## max|F(x,y)-G(x)H(y)|
##        x      y
## x 0.0000 0.0375
## y 0.0375 0.0000
##
## n
##      x   y
## x 1055 239
## y  239 239
##
## P
##   x      y
## x        0.1656
## y 0.1656
```

```r
plot(log(VrCausa), log(VrPago), main = "Gráfico de Dispersão Valor Pago vs log-Valor da Causa" )
```

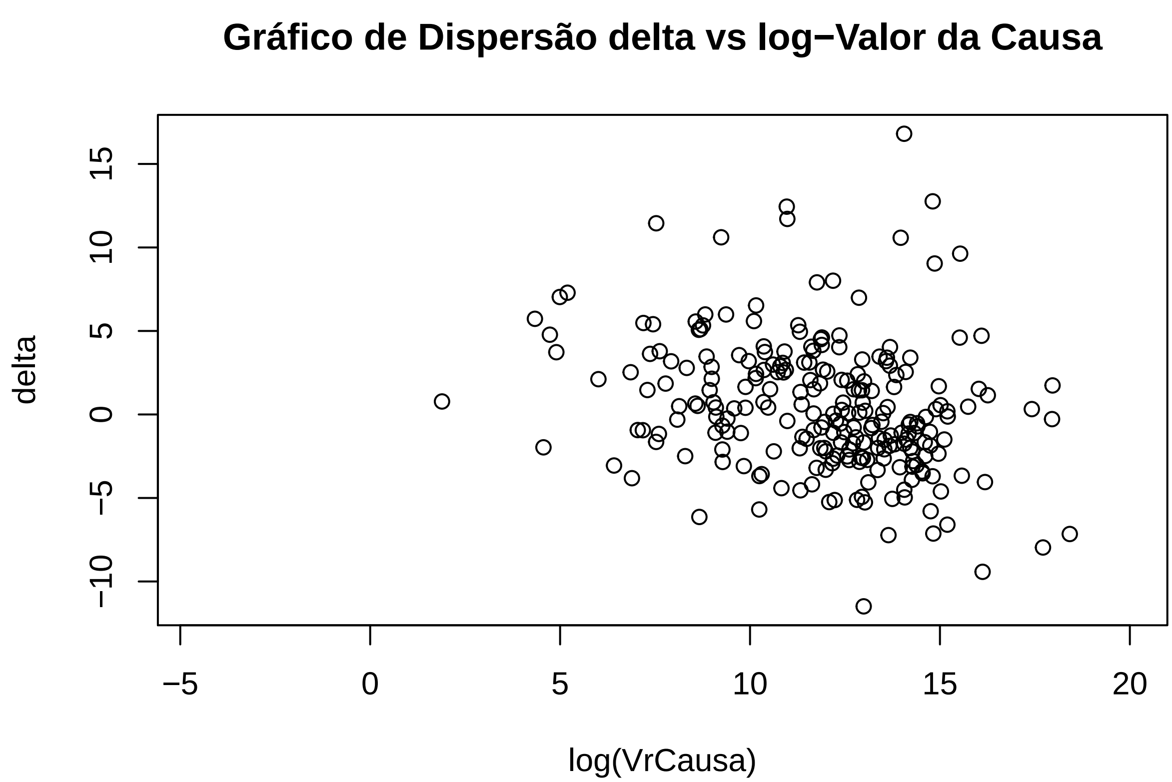



```r
cor.test(log(VrCausa), log(VrPago))
```

```
##
##  Pearson's product-moment correlation
##
## data:  log(VrCausa) and log(VrPago)
```

```
## t = 12.328, df = 437, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4350762 0.5742706
## sample estimates:
##       cor
## 0.5079822
```

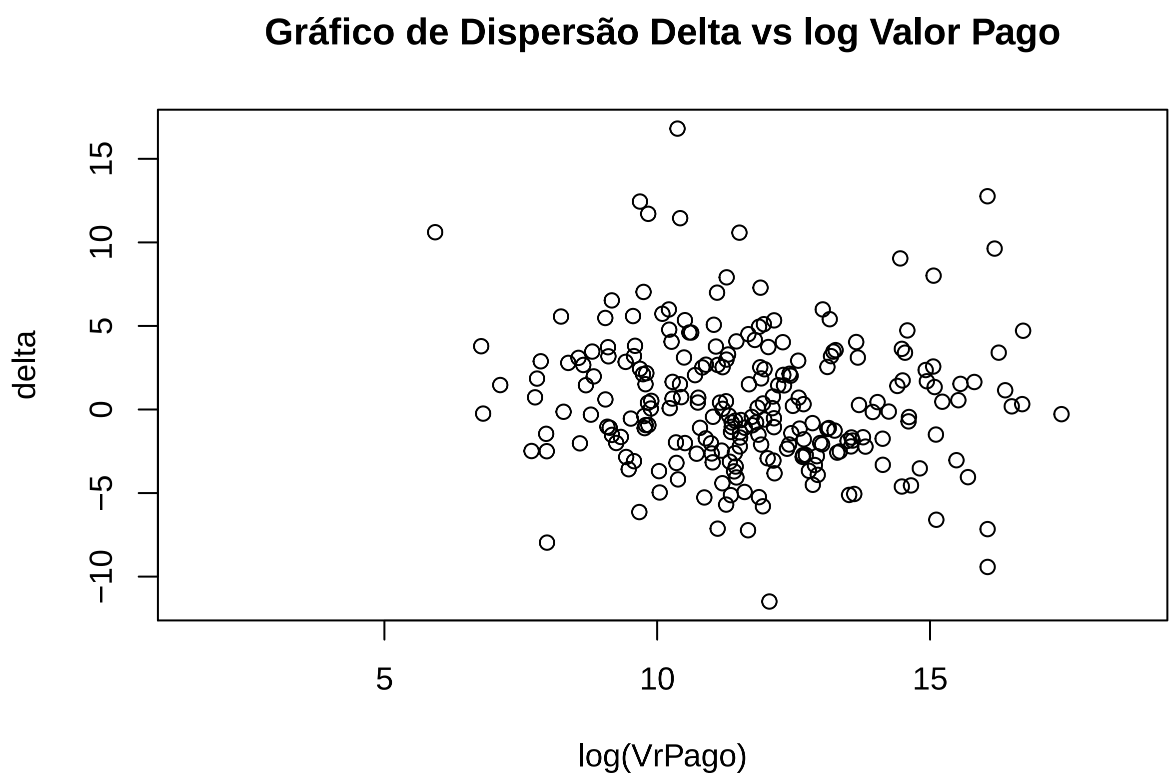
```r
plot(log(VrCausa), delta, main = "Gráfico de Dispersão delta vs log-Valor da Causa")
```




```r
cor.test(log(VrCausa), delta)
```

```
##
##  Pearson's product-moment correlation
##
## data:  log(VrCausa) and delta
## t = -4.3834, df = 237, p-value = 1.757e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3872854 -0.1522441
## sample estimates:
##        cor
## -0.2738486
```

```r
plot(log(VrPago), delta, main = "Gráfico de Dispersão Delta vs log Valor Pago")
```




```r
cor.test(log(VrPago), delta)
```

```
##
##  Pearson's product-moment correlation
##
## data:  log(VrPago) and delta
## t = -1.5946, df = 237, p-value = 0.1121
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.22695414  0.02418502
## sample estimates:
##        cor
## -0.1030262
```

Não há correlação linear entre o tempo e o Valor Pago, nem entre o tempo e o delta, tampouco entre o tempo e o logaritmo de delta. Há correlação linear entre o tempo e o Valor da Causa (e também portanto o logaritmo do Valor da Causa), bem como entre o tempo e o logaritmo do Valor Pago.

Há correlação linear entre o logaritmo do valor da causa e o logaritmo do valor pago . Não há correlação linear entre o logaritmo do Valor da Causa e o delta, e nem entre log do Valor Pago e o delta.

n é muito grande para calcular-se o teste de Hoeffding no computador que foi utilizado para este trabalho. Ou seja, não pude estimar dependência não-linear entre as variáveis.

Conclusão

Esta minuta de relatório de administração nos possibilitou prever eventos extremos quanto à variável Tempo e Valor Pago, bem como a proporação de eventos em que a Variável Valor da Causa é maior que Valor Pago, e também a ocorrência ou não de correlações - lineares apenas em razão da limitação computacional - entre as variáveis de interesse.