

# Análise Estatística

**Solicitante: Rodrigo Almeida**

**05/06/2024**

OBS: Nesse relatório os modelos nomeados significativos, refere-se a modelos com variáveis explicativas estatisticamente significativas



**Empresa Júnior d  
Estatística e Ciênci  
Atuariais**

**Consultores  
Responsáveis**

**Gabriel Moreira**

**gabriel-  
moreira@estatmg.com.br  
(mailto:gabriel-  
moreira@estatmg.com.br)**

**Tamires Balçante**

**tamires@estatmg.com.br  
(mailto:tamires@estatmg.cc)**

**Pedro Silva**

**pedro-afonso@estatmg.com  
(mailto:pedro-  
afonso@estatmg.com.br)**

## Objetivos

Esta análise tem como objetivo procurar a associação entre a carga tributária e o nível de endividamento das empresas brasileiras listadas na B3. A pesquisa pretende explorar setores específicos da economia brasileira para identificar variações na relação entre a carga tributária e a estrutura de capital dessas empresas.

Além disso, será investigado em que medida as teorias Trade-off e Pecking Order são aplicáveis no contexto brasileiro. Um dos objetivos centrais é analisar se as empresas com maiores cargas tributárias tendem a ser mais endividadas. Por fim, a pesquisa buscará verificar se o aumento da carga tributária afeta positivamente o nível de endividamento das empresas.

Essa análise proporcionará uma compreensão abrangente de como a carga tributária influencia a estrutura de capital das empresas no Brasil, destacando a importância da análise setorial e a aplicabilidade das principais teorias financeiras no cenário nacional.

## Metodologia

A base de dados continha empresas brasileiras listadas na B3 com seus respectivos indicadores trimestrais, entre eles: carga tributária (CT), distribuição da riqueza ao governo, lucratividade, retorno sobre o patrimônio líquido (ROE), grau de imobilização do ativo, liquidez corrente, giro do ativo, tamanho da firma, risco do negócio, alavancagem financeira, crescimento da receita, maior CT, aumento da CT, CT x maior CT e CT x aumento da CT.

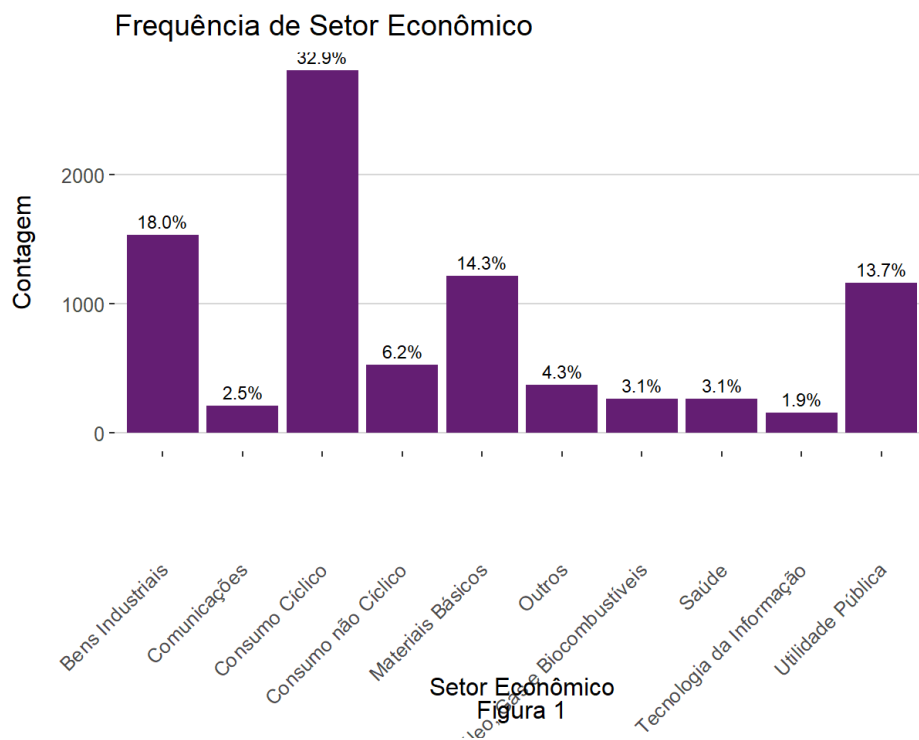
Foi realizada uma análise descritiva para entender a distribuição e a interação entre os dados. Em seguida, foi desenvolvido um modelo linear generalizado (MLG) e feita a análise de resíduos utilizando o software R.

## Resultados

### Análise Descritiva

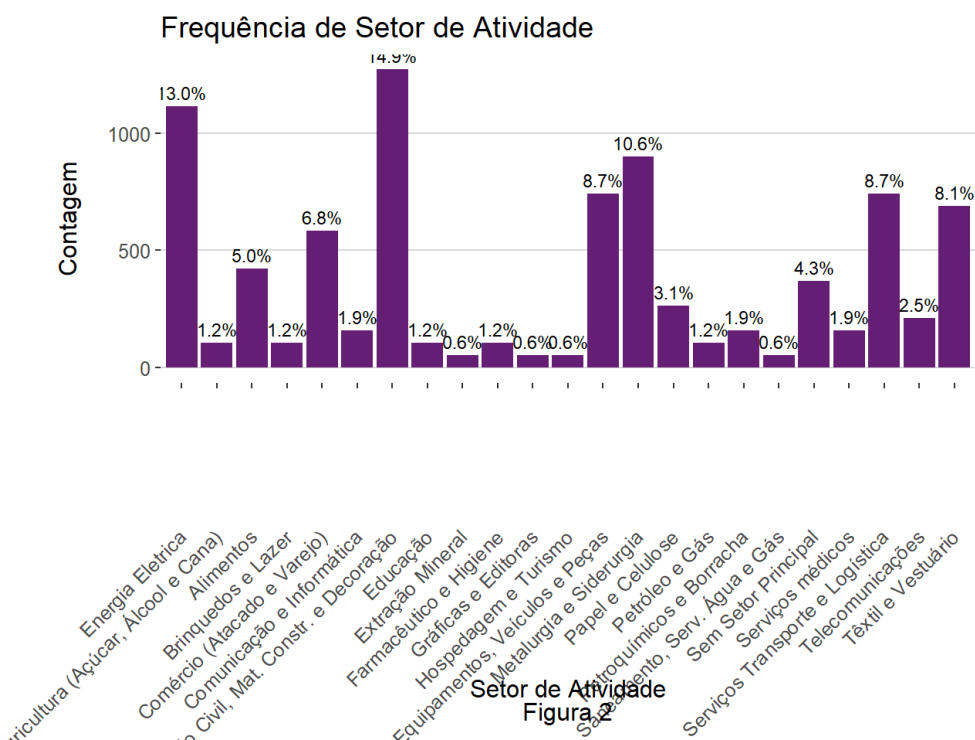
#### Gráficos de frequência para as variáveis categóricas

- Gráfico de barras para Setor\_Economico



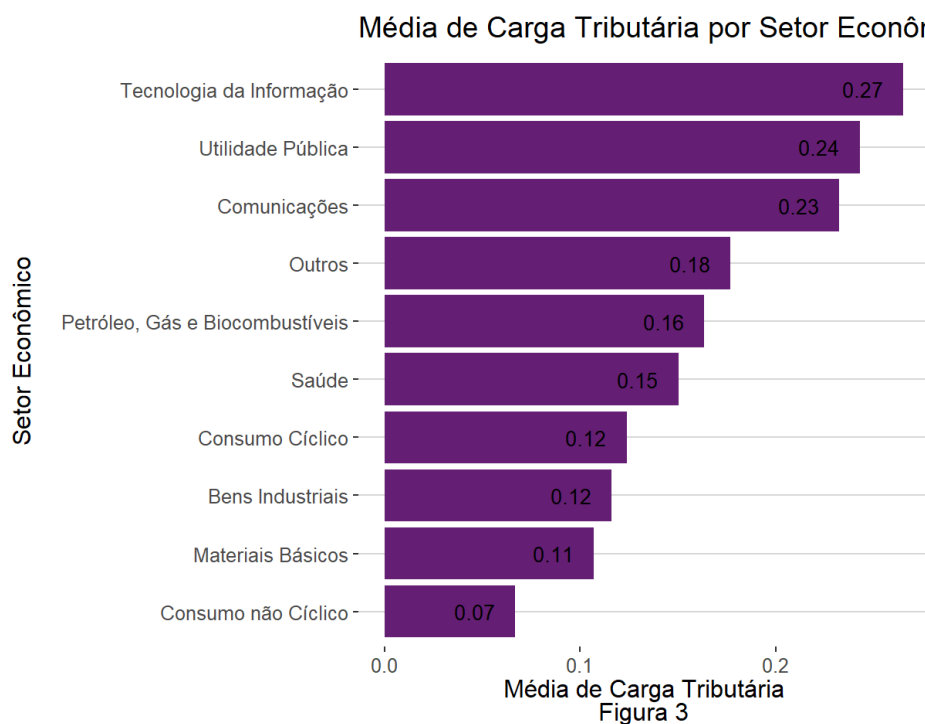
O gráfico acima (Figura 1) mostra que os setores Bens Industriais e Consumo Cíclico representam juntos a maioria dos dados, mais de 50%.

- Gráfico de barras para Setor\_Atividade



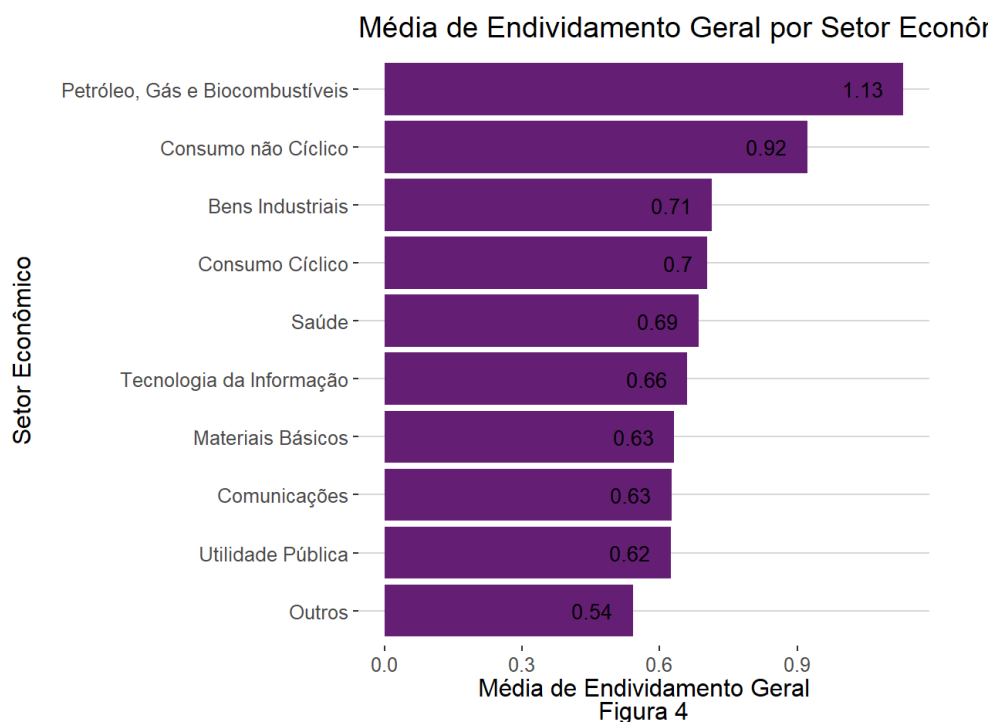
O gráfico acima (Figura 2) revela que as empresas dos setores de atividade Energia Elétrica e Construção Civil, Materiais de Construção e Decoração são os maiores setores no banco de dados, representando 14,9% e 13%, respectivamente.

- Média de Carga Tributária por Setor Econômico



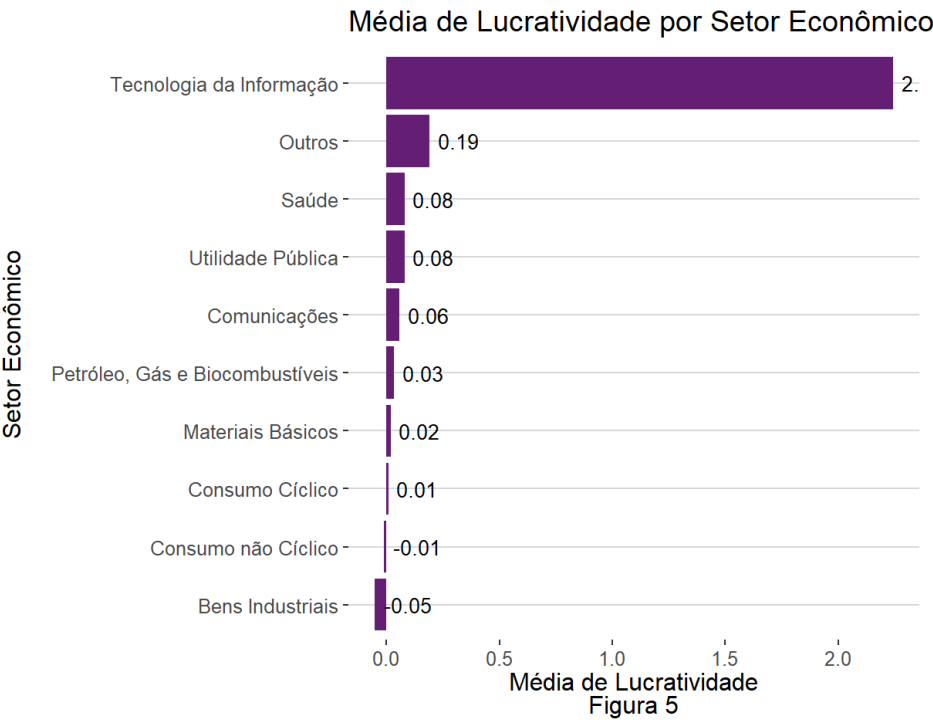
Segundo a Figura 3, as empresas do setor de Tecnologia da Informação têm a maior carga tributária média, representando 27% da receita bruta da empresa, seguidas pelas empresas do setor de Utilidade Pública, que têm em média 24% de impostos, taxas e contribuições sobre sua receita bruta. Por outro lado, as empresas do setor de Consumo não Cíclico apresentam o menor percentual de carga tributária média, representando apenas 7% da receita bruta.

- Média de Endividamento Geral por Setor Econômico



Apesar das empresas do setor de Tecnologia da Informação apresentarem as maiores cargas tributárias médias, são as empresas do setor de Petróleo, Gás e Biocombustíveis que mostram os maiores níveis médios de endividamento, conforme observado na Figura 4.

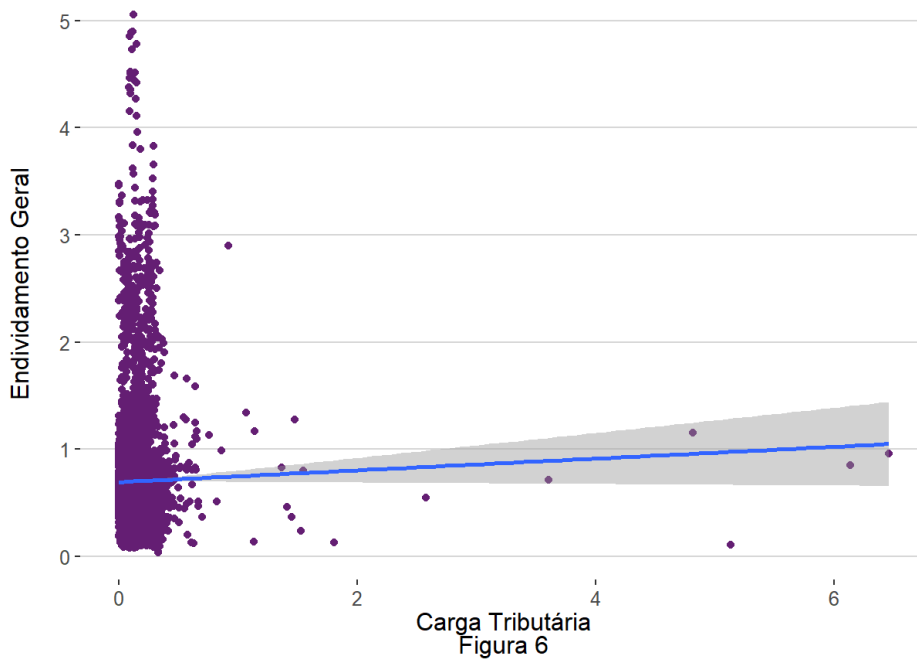
- Média de Lucratividade por Setor Econômico



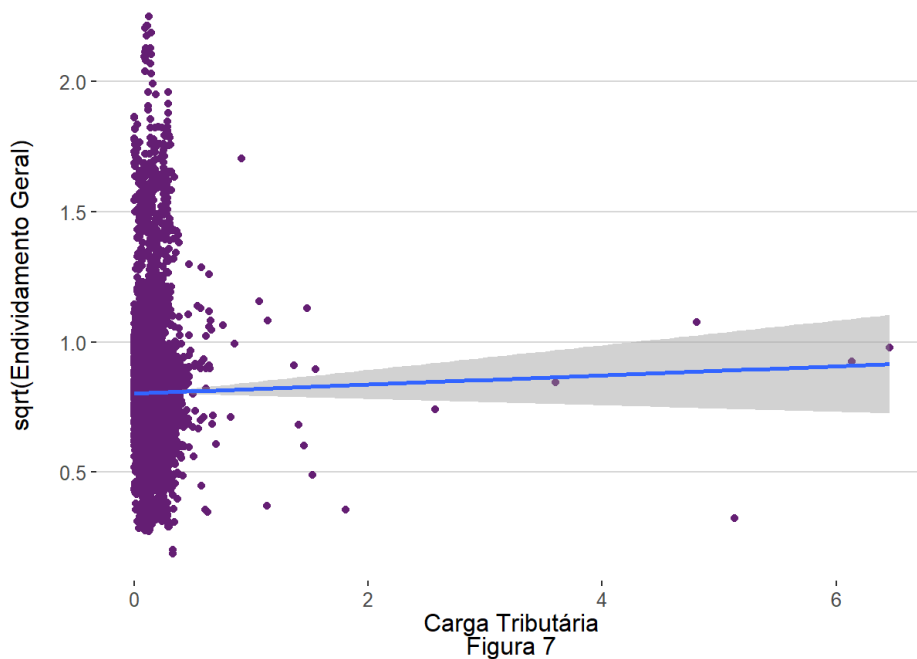
As empresas do setor de Tecnologia da Informação , apesar de terem maiores cargas tributárias médias, apresentam também maiores lucratividades médias. Por outro lado, as empresas do setor de Consumo não Cíclico , apesar de terem as menores cargas tributárias médias, registram uma lucratividade negativa.

- Relação entre Carga Tributária e Endividamento Geral

Relação entre Carga Tributária e Endividamento Geral



Relação entre Carga Tributária e  $\sqrt{\text{Endividamento Geral}}$



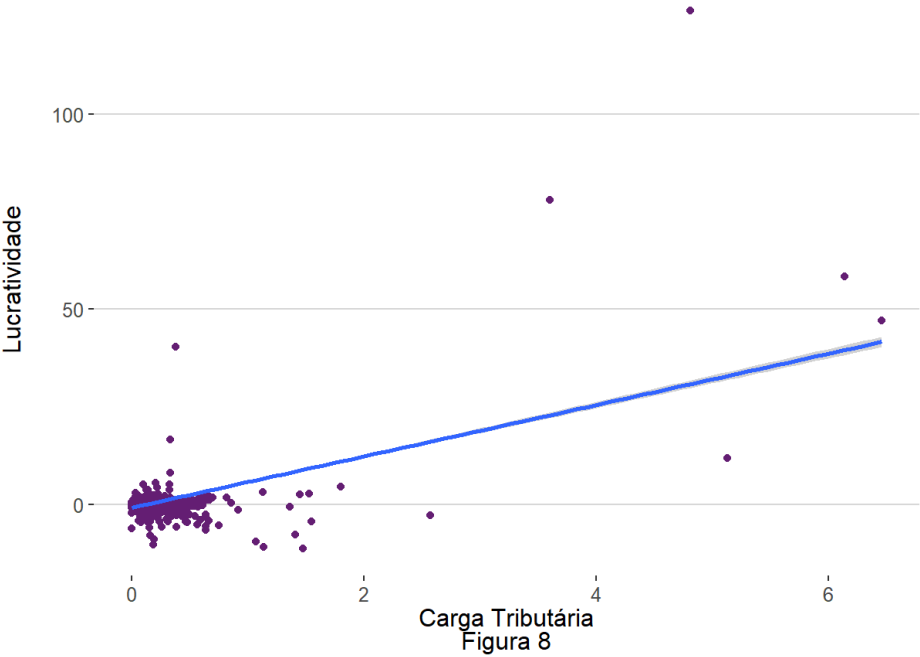
Na figura 6, aparentemente não há uma correlação linear entre Endividamento Geral e Carga Tributária. A correlação calculada é de  $0.01883047$ , indicando uma correlação positiva muito próxima de zero. Isso sugere uma tendência leve de que, em média, à medida que a carga tributária aumenta, o endividamento geral das empresas também pode aumentar ligeiramente. No entanto, essa relação é muito fraca para ser considerada significativa.

Na figura 7, foi utilizada a raiz quadrada da variável Endividamento Geral, a qual será usada posteriormente na modelagem e segue o mesmo raciocínio da variável original.

- Relação entre Carga Tributária e Lucratividade

```
## `geom_smooth()` using formula = 'y ~ x'
```

Relação entre Carga Tributária e Lucratividade



Na figura 8, observamos uma correlação positiva com um valor de 0.5568518 , o que indica uma relação linear moderadamente forte entre a carga tributária e a lucratividade das empresas nesse conjunto de dados. Isso sugere que, em média, à medida que a carga tributária aumenta, a lucratividade das empresas também tende a aumentar. No entanto, a relação não é perfeita, e pode haver uma grande variação nos dados que não está sendo capturada pela correlação linear.

Modelagem

Comparação entre Modelos Completos

- Diferença: Setor\_Economico X Setor\_Atividade

Predictors	END_GERAL			END_GERAL		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	2.15	2.05 – 2.24	<0.001	1.83	1.73 – 1.94	<0.001
CARGA TRIBUTARIA	-0.89	-1.16 – -0.62	<0.001	-1.18	-1.45 – -0.92	<0.001
DISTR RIQUEZA GOV	-0.00	-0.00 – 0.00	0.865	-0.00	-0.00 – 0.00	0.720
LUCRATIVIDADE	-0.00	-0.01 – 0.00	0.365	-0.00	-0.01 – 0.00	0.128
ROE	-0.00	-0.01 – 0.00	0.665	-0.00	-0.01 – 0.00	0.671
GRAU IMOBILZ ATIVO	-0.19	-0.24 – -0.15	<0.001	-0.22	-0.27 – -0.17	<0.001
LIQ CORRENTE	-0.14	-0.15 – -0.14	<0.001	-0.12	-0.13 – -0.12	<0.001
GIRO ATIVO	-0.03	-0.05 – -0.01	0.009	-0.04	-0.06 – -0.02	0.001
TAMANHO FIRMA	-0.16	-0.17 – -0.15	<0.001	-0.12	-0.14 – -0.11	<0.001
RISCO NEGOCIO	-0.00	-0.00 – 0.00	0.376	-0.00	-0.00 – 0.00	0.428
Cresc Receita	-0.00	-0.00 – 0.00	0.226	-0.00	-0.00 – 0.00	0.124
Maior CT	-0.10	-0.14 – -0.06	<0.001	-0.13	-0.17 – -0.09	<0.001
Aumento Carga Trib	0.02	-0.01 – 0.05	0.170	0.03	0.00 – 0.05	0.048

CT X MaioresCT	1.34	1.05 – 1.63	<0.001	1.67	1.39 – 1.95	<0.001
CT X AumentoCT	-0.18	-0.33 – -0.02	0.024	-0.22	-0.36 – -0.07	0.003
Setor Economico [Comunicações]	-0.02	-0.08 – 0.04	0.436			
Setor Economico [Consumo Cíclico]	-0.02	-0.05 – 0.00	0.068			
Setor Economico [Consumo não Cíclico]	0.18	0.14 – 0.22	<0.001			
Setor Economico [Materiais Básicos]	0.01	-0.02 – 0.04	0.646			
Setor Economico [Outros]	-0.11	-0.16 – -0.07	<0.001			
Setor Economico [Petróleo, Gás e Biocombustíveis]	0.56	0.51 – 0.61	<0.001			
Setor Economico [Saúde]	-0.03	-0.09 – 0.02	0.231			
Setor Economico [Tecnologia da Informação]	-0.15	-0.22 – -0.09	<0.001			
Setor Economico [Utilidade Pública]	-0.09	-0.12 – -0.05	<0.001			
Setor Atividade [Agricultura (Açúcar, Alcool e Cana)]				0.40	0.32 – 0.49	<0.001
Setor Atividade [Alimentos]				0.25	0.20 – 0.30	<0.001
Setor Atividade [Brinquedos e Lazer]				0.74	0.66 – 0.82	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]				0.01	-0.03 – 0.05	0.606
Setor Atividade [Comunicação e Informática]				-0.04	-0.11 – 0.03	0.225
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]				-0.06	-0.10 – -0.02	0.001
Setor Atividade [Educação]				-0.05	-0.13 – 0.03	0.240
Setor Atividade [Extração Mineral]				0.21	0.10 – 0.31	<0.001
Setor Atividade [Farmacêutico e Higiene]				0.29	0.21 – 0.37	<0.001
Setor Atividade [Gráficas e Editoras]				-0.05	-0.16 – 0.05	0.322
Setor Atividade [Hospedagem e Turismo]				0.54	0.43 – 0.64	<0.001

Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	0.03	-0.00 – 0.07	0.083
Setor Atividade [Metalurgia e Siderurgia]	0.06	0.02 – 0.10	<b>0.002</b>
Setor Atividade [Papel e Celulose]	0.17	0.11 – 0.22	<b>&lt;0.001</b>
Setor Atividade [Petróleo e Gás]	0.18	0.10 – 0.26	<b>&lt;0.001</b>
Setor Atividade [Petroquímicos e Borracha]	0.95	0.88 – 1.01	<b>&lt;0.001</b>
Setor Atividade [Saneamento, Serv. Água e Gás]	-0.07	-0.17 – 0.04	0.221
Setor Atividade [Sem Setor Principal]	-0.01	-0.06 – 0.04	0.607
Setor Atividade [Serviços médicos]	-0.06	-0.12 – 0.01	0.081
Setor Atividade [Serviços Transporte e Logística]	0.19	0.15 – 0.23	<b>&lt;0.001</b>
Setor Atividade [Telecomunicações]	0.06	0.00 – 0.12	<b>0.040</b>
Setor Atividade [Têxtil e Vestuário]	0.25	0.21 – 0.29	<b>&lt;0.001</b>
Observations	8533	8533	
R <sup>2</sup> / R <sup>2</sup> adjusted	0.302 / 0.300	0.379 / 0.376	

#### Teste AIC:

## AIC do modelo completo utilizando Setor Econômico: 8714.096

## AIC do modelo completo utilizando Setor Atividade: 7741.303

Um valor menor do Critério de Informação de Akaike (AIC) indica um modelo melhor. Portanto:

- O modelo que utiliza o **Setor\_Atividade** possui um AIC menor (7741.303) em comparação com o modelo que utiliza **Setor\_Economico** (8714.096).
- Isso sugere que o modelo que inclui o **Setor\_Atividade** é preferível com base no Critério de Informação de Akaike, pois tem um AIC mais baixo, indicando um melhor ajuste do modelo.

#### Teste ANOVA:



```
## Analysis of Variance Table
##
## Model 1: END_GERAL ~ CARGA_TRIBUTARIA + DISTR_RIQUEZA_GOV + LUCRATIVIDADE +
##   ROE + GRAU_IMOBILZ_ATIVO + LIQ_CORRENTE + GIRO_ATIVO + TAMANHO_FIRMA +
##   RISCO_NEGOCIO + Cresc_Receita + Maior_CT + Aumento_Carga_Trib +
##   CT_X_MaioresCT + CT_X_AumentoCT + factor(Setor_Economico)
## Model 2: END_GERAL ~ CARGA_TRIBUTARIA + DISTR_RIQUEZA_GOV + LUCRATIVIDADE +
##   ROE + GRAU_IMOBILZ_ATIVO + LIQ_CORRENTE + GIRO_ATIVO + TAMANHO_FIRMA +
##   RISCO_NEGOCIO + Cresc_Receita + Maior_CT + Aumento_Carga_Trib +
##   CT_X_MaioresCT + CT_X_AumentoCT + factor(Setor_Atividade)
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    8509 1379.1
## 2    8496 1226.8 13    152.34 81.154 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- O segundo modelo, que usa **Setor\_Atividade**, tem um ajuste significativamente melhor do que o primeiro modelo, que usa **Setor\_Economico**.
- A inclusão de **Setor\_Atividade** no modelo melhora significativamente o ajuste, conforme indicado pelo valor p muito baixo (< 2.2e-16).

Seguindo as análise com a variável **Setor\_Atividade** e as que testaram significativas no modelo completo

Modelo Significativo - Linear

Predictors	END_GERAL		
	Estimates	CI	p
(Intercept)	1.80	1.69 – 1.90	<0.001
CARGA TRIBUTARIA	-1.23	-1.49 – -0.97	<0.001
LUCRATIVIDADE	-0.01	-0.01 – -0.00	0.035
GRAU IMOBILZ ATIVO	-0.20	-0.25 – -0.16	<0.001
LIQ CORRENTE	-0.12	-0.13 – -0.11	<0.001
TAMANHO FIRMA	-0.12	-0.13 – -0.11	<0.001
Maior CT	-0.10	-0.14 – -0.07	<0.001
CT X MaioresCT	1.53	1.26 – 1.80	<0.001
Setor Atividade [Agricultura (Açúcar, Alcool e Cana)]	0.40	0.32 – 0.48	<0.001
Setor Atividade [Alimentos]	0.23	0.18 – 0.28	<0.001
Setor Atividade [Brinquedos e Lazer]	0.74	0.66 – 0.82	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	-0.01	-0.05 – 0.03	0.477
Setor Atividade [Comunicação e Informática]	-0.06	-0.12 – 0.01	0.097
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	-0.06	-0.10 – -0.02	0.001

Setor Atividade [Educação]	-0.06	-0.14 – 0.02	0.144
Setor Atividade [Extração Mineral]	0.19	0.08 – 0.30	<b>0.001</b>
Setor Atividade [Farmacêutico e Higiene]	0.26	0.19 – 0.34	<b>&lt;0.001</b>
Setor Atividade [Gráficas e Editoras]	-0.07	-0.18 – 0.04	0.199
Setor Atividade [Hospedagem e Turismo]	0.54	0.43 – 0.64	<b>&lt;0.001</b>
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	0.02	-0.02 – 0.06	0.279
Setor Atividade [Metalurgia e Siderurgia]	0.05	0.01 – 0.09	<b>0.013</b>
Setor Atividade [Papel e Celulose]	0.16	0.11 – 0.22	<b>&lt;0.001</b>
Setor Atividade [Petróleo e Gás]	0.17	0.09 – 0.25	<b>&lt;0.001</b>
Setor Atividade [Petroquímicos e Borracha]	0.93	0.86 – 0.99	<b>&lt;0.001</b>
Setor Atividade [Saneamento, Serv. Água e Gás]	-0.08	-0.18 – 0.03	0.147
Setor Atividade [Sem Setor Principal]	-0.02	-0.07 – 0.02	0.337
Setor Atividade [Serviços médicos]	-0.07	-0.14 – -0.01	<b>0.025</b>
Setor Atividade [Serviços Transporte e Logística]	0.18	0.14 – 0.21	<b>&lt;0.001</b>
Setor Atividade [Telecomunicações]	0.05	-0.00 – 0.11	0.072
Setor Atividade [Têxtil e Vestuário]	0.24	0.20 – 0.28	<b>&lt;0.001</b>
Observations	8533		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.377 / 0.375		

## Fator de inflação de variância (VIF)

##	GVIF	Df	GVIF^(1/(2*Df))
## CARGA_TRIBUTARIA	27.368373	1	5.231479
## LUCRATIVIDADE	1.778672	1	1.333669
## GRAU_IMOBILZ_ATIVO	1.624511	1	1.274563
## LIQ_CORRENTE	1.252707	1	1.119244
## TAMANHO_FIRMA	1.528968	1	1.236514
## Maior_CT	4.360730	1	2.088236
## CT_X_MaioresCT	37.119717	1	6.092595
## factor(Setor_Atividade)	3.733916	22	1.030395

O *Fator de inflação de variância* indica quanto a variância da estimativa de um coeficiente de regressão está aumentada devido à multicolinearidade com outras variáveis independentes no modelo. Temos que as variáveis CARGA\_TRIBUTARIA, Maior\_CT, CT\_X\_MaioresCT podem estar fortemente correlacionadas, o que pode levar a problemas de multicolinearidade.

### Modelo Significativo reduzido

Removendo as variáveis Maior\_CT e CT\_X\_MaioresCT, o vif diminui significamente.

Predictors	END_GERAL		
	Estimates	CI	p
(Intercept)	1.66	1.57 – 1.76	<0.001
CARGA TRIBUTARIA	0.24	0.19 – 0.30	<0.001
GRAU IMOBILZ ATIVO	-0.22	-0.27 – -0.18	<0.001
LIQ CORRENTE	-0.12	-0.13 – -0.12	<0.001
TAMANHO FIRMA	-0.11	-0.13 – -0.10	<0.001
Setor Atividade [Agricultura (Açúcar, Álcool e Cana)]	0.43	0.35 – 0.51	<0.001
Setor Atividade [Alimentos]	0.24	0.19 – 0.28	<0.001
Setor Atividade [Brinquedos e Lazer]	0.72	0.64 – 0.80	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	-0.02	-0.06 – 0.02	0.436
Setor Atividade [Comunicação e Informática]	-0.08	-0.15 – -0.02	0.012
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	-0.07	-0.10 – -0.04	<0.001
Setor Atividade [Educação]	-0.08	-0.16 – -0.00	0.045
Setor Atividade [Extração Mineral]	0.16	0.05 – 0.27	0.004
Setor Atividade [Farmacêutico e Higiene]	0.25	0.17 – 0.33	<0.001
Setor Atividade [Gráficas e Editoras]	-0.12	-0.23 – -0.01	0.025
Setor Atividade [Hospedagem e Turismo]	0.54	0.44 – 0.65	<0.001
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	0.01	-0.03 – 0.05	0.688
Setor Atividade [Metalurgia e Siderurgia]	0.05	0.02 – 0.09	0.005
Setor Atividade [Papel e Celulose]	0.17	0.11 – 0.22	<0.001

Setor Atividade [Petróleo e Gás]	0.18	0.10 – 0.26	<0.001
Setor Atividade [Petroquímicos e Borracha]	0.89	0.82 – 0.95	<0.001
Setor Atividade [Saneamento, Serv. Água e Gás]	-0.15	-0.25 – -0.04	0.006
Setor Atividade [Sem Setor Principal]	-0.03	-0.07 – 0.02	0.301
Setor Atividade [Serviços médicos]	-0.12	-0.18 – -0.05	<0.001
Setor Atividade [Serviços Transporte e Logística]	0.15	0.11 – 0.18	<0.001
Setor Atividade [Telecomunicações]	0.05	-0.01 – 0.11	0.076
Setor Atividade [Têxtil e Vestuário]	0.21	0.17 – 0.25	<0.001

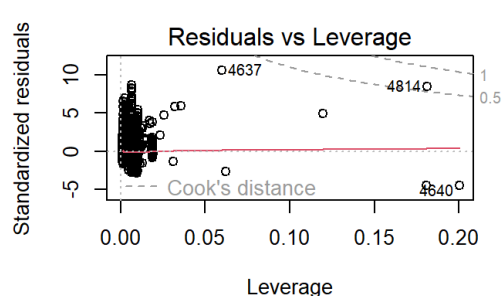
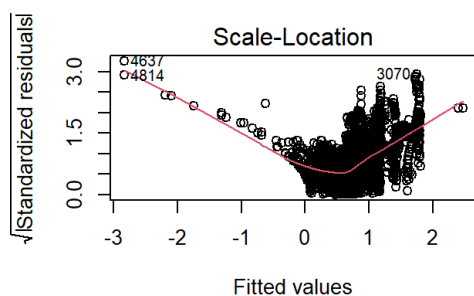
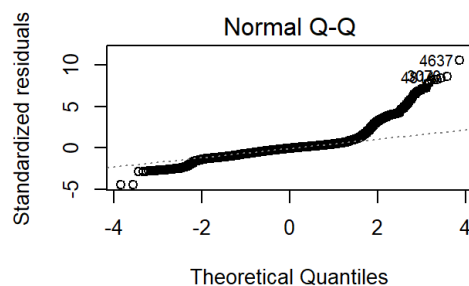
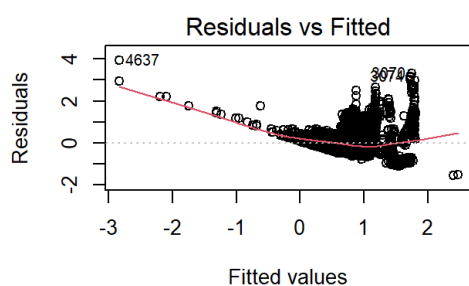
---

Observations 8533

R<sup>2</sup> / R<sup>2</sup> adjusted 0.366 / 0.364

##	GVIF	Df	GVIF^(1/(2*Df))
## CARGA_TRIBUTARIA	1.176067	1	1.084466
## GRAU_IMOBILZ_ATIVO	1.597596	1	1.263961
## LIQ_CORRENTE	1.208319	1	1.099236
## TAMANHO_FIRMA	1.467320	1	1.211330
## factor(Setor_Atividade)	2.727001	22	1.023062

## Diagnóstico de ajuste do modelo significativo reduzido



A suposição de normalidade dos resíduos não é atendida, conforme evidenciado pelos seguintes diagnósticos:

- Gráfico Normal Q-Q: Os resíduos não seguem uma distribuição normal, como indicado pela curva não linear no gráfico.
- Gráfico Residuals vs Fitted: Existe uma certa tendência nos resíduos, sugerindo que o modelo pode não capturar completamente a estrutura dos dados.
- Gráfico Scale-Location (Sqrt(standardized residuals) vs Fitted): Mostra um padrão não aleatório dos resíduos em relação aos valores ajustados, indicando heterocedasticidade (variância não constante).
- Gráfico Residuals vs Leverage: Alguns pontos mostram alavancagem, o que significa que esses pontos têm um efeito desproporcional sobre a estimativa dos coeficientes do modelo.

Esses diagnósticos indicam que o modelo pode não ser apropriado para os dados, devido à não normalidade dos resíduos, presença de tendência nos resíduos, heterocedasticidade e pontos de alavancagem. Isso sugere a necessidade de revisar o modelo, considerar transformações nos dados ou explorar modelos alternativos que possam melhor capturar a estrutura dos dados.

## Modelo completo - Box cox

Considerando uma tranformação de Box-cox

### Transformação Box-cox

A transformação de Box-Cox é uma técnica utilizada para melhorar a normalidade e a estabilização da variância dos dados. Ela é particularmente útil em situações onde os dados apresentam desvios da normalidade e/ou heterocedasticidade. A transformação de Box-Cox é definida pela seguinte fórmula:

$$\frac{y^\lambda - 1}{\lambda}, \text{ se } \lambda \neq 0$$

$$\log(y), \text{ se } \lambda = 0$$

onde *y* é a variável a ser transformada e λ é o parâmetro da transformação.

- Escolha de λ: O parâmetro λ pode assumir qualquer valor real. Um método comum é selecionar o valor de λ que maximiza a função de verossimilhança.
- Aplicação da transformação: Após selecionar o valor ótimo de λ a transformação é aplicada aos dados de acordo com a fórmula acima.

Utilizando o modelo completo de Setor\_Atividade, temos o seguinte valor de λ:

```
## Valor de Lambda:  -0.08993942
```

Predictors	y_trans		
	Estimates	CI	p
(Intercept)	0.49	0.37 - 0.60	<0.001
CARGA TRIBUTARIA	-1.72	-2.00 - -1.44	<0.001
DISTR RIQUEZA GOV	-0.00	-0.00 - 0.00	0.862
LUCRATIVIDADE	0.00	-0.00 - 0.01	0.319
ROE	-0.00	-0.01 - 0.00	0.191
GRAU IMOBILZ ATIVO	-0.44	-0.49 - -0.39	<0.001
LIQ CORRENTE	-0.23	-0.24 - -0.22	<0.001

GIRO ATIVO	-0.03	-0.05 – -0.00	<b>0.023</b>
TAMANHO FIRMA	-0.05	-0.06 – -0.04	<b>&lt;0.001</b>
RISCO NEGOCIO	0.00	-0.00 – 0.00	0.874
Cresc Receita	-0.00	-0.00 – 0.00	0.075
Maior CT	-0.23	-0.27 – -0.18	<b>&lt;0.001</b>
Aumento Carga Trib	0.03	0.00 – 0.06	<b>0.034</b>
CT X MaioresCT	2.27	1.97 – 2.56	<b>&lt;0.001</b>
CT X AumentoCT	-0.17	-0.32 – -0.01	<b>0.037</b>
Setor Atividade [Agricultura (Açúcar, Alcool e Cana)]	0.27	0.18 – 0.35	<b>&lt;0.001</b>
Setor Atividade [Alimentos]	0.20	0.15 – 0.25	<b>&lt;0.001</b>
Setor Atividade [Brinquedos e Lazer]	0.41	0.32 – 0.50	<b>&lt;0.001</b>
Setor Atividade [Comércio (Atacado e Varejo)]	-0.02	-0.06 – 0.03	0.491
Setor Atividade [Comunicação e Informática]	0.01	-0.06 – 0.09	0.701
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	-0.07	-0.11 – -0.03	<b>&lt;0.001</b>
Setor Atividade [Educação]	-0.20	-0.29 – -0.12	<b>&lt;0.001</b>
Setor Atividade [Extração Mineral]	0.11	-0.01 – 0.22	0.074
Setor Atividade [Farmacêutico e Higiene]	0.39	0.31 – 0.47	<b>&lt;0.001</b>
Setor Atividade [Gráficas e Editoras]	0.01	-0.11 – 0.12	0.924
Setor Atividade [Hospedagem e Turismo]	0.62	0.50 – 0.73	<b>&lt;0.001</b>
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	0.05	0.00 – 0.09	<b>0.029</b>
Setor Atividade [Metalurgia e Siderurgia]	0.01	-0.03 – 0.06	0.507
Setor Atividade [Papel e Celulose]	0.29	0.23 – 0.35	<b>&lt;0.001</b>
Setor Atividade [Petróleo e Gás]	0.09	0.01 – 0.18	<b>0.027</b>
Setor Atividade [Petroquímicos e Borracha]	0.70	0.63 – 0.77	<b>&lt;0.001</b>

Setor Atividade [Saneamento, Serv. Água e Gás]	-0.16	-0.28 – -0.05	<b>0.005</b>
Setor Atividade [Sem Setor Principal]	-0.09	-0.15 – -0.04	<b>&lt;0.001</b>
Setor Atividade [Serviços médicos]	-0.10	-0.17 – -0.03	<b>0.007</b>
Setor Atividade [Serviços Transporte e Logística]	0.26	0.22 – 0.31	<b>&lt;0.001</b>
Setor Atividade [Telecomunicações]	0.05	-0.01 – 0.11	0.130
Setor Atividade [Têxtil e Vestuário]	0.13	0.08 – 0.17	<b>&lt;0.001</b>
Observations	8533		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.490 / 0.488		

##	GVIF	Df	GVIF^(1/(2*Df))
## CARGA_TRIBUTARIA	28.663094	1	5.353793
## DISTR_RIQUEZA_GOV	1.004363	1	1.002179
## LUCRATIVIDADE	1.832003	1	1.353515
## ROE	1.004707	1	1.002351
## GRAU_IMOBILZ_ATIVO	1.717401	1	1.310497
## LIQ_CORRENTE	1.268600	1	1.126321
## GIRO_ATIVO	1.351876	1	1.162702
## TAMANHO_FIRMA	1.579773	1	1.256890
## RISCO_NEGOCIO	1.021900	1	1.010891
## Cresc_Receita	1.003219	1	1.001608
## Maior_CT	5.198213	1	2.279959
## Aumento_Carga_Trib	2.611623	1	1.616052
## CT_X_MaioresCT	40.676308	1	6.377798
## CT_X_AumentoCT	8.953000	1	2.992156
## factor(Setor_Atividade)	4.785576	22	1.036223

O modelo completo tem algumas variáveis não significativas e altamente correlacionada, retirando elas temos o modelo significativo abaixo

## Modelo Significativo - Box cox

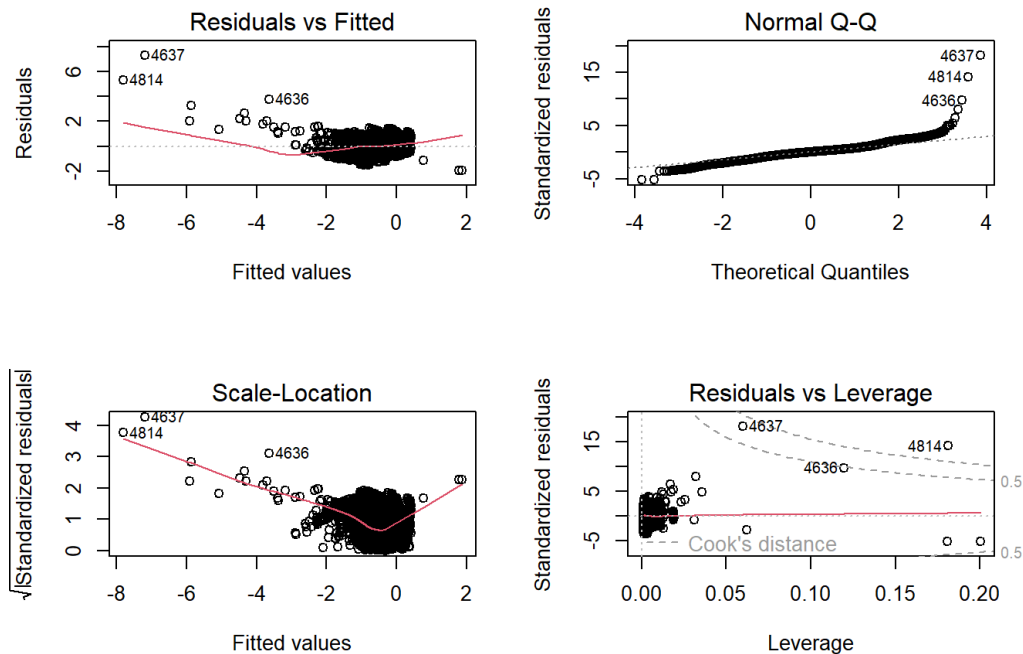
Predictors	y_trans		
	Estimates	CI	p
(Intercept)	0.19	0.08 – 0.29	<b>&lt;0.001</b>
CARGA TRIBUTARIA	0.32	0.27 – 0.38	<b>&lt;0.001</b>
GRAU IMOBILZ ATIVO	-0.47	-0.52 – -0.42	<b>&lt;0.001</b>
LIQ CORRENTE	-0.23	-0.24 – -0.22	<b>&lt;0.001</b>
TAMANHO FIRMA	-0.03	-0.04 – -0.02	<b>&lt;0.001</b>
Setor Atividade [Agricultura (Açúcar, Alcool e Cana)]	0.36	0.27 – 0.45	<b>&lt;0.001</b>
Setor Atividade [Alimentos]	0.25	0.20 – 0.29	<b>&lt;0.001</b>

Setor Atividade [Brinquedos e Lazer]	0.40	0.31 – 0.49	<b>&lt;0.001</b>
Setor Atividade [Comércio (Atacado e Varejo)]	-0.00	-0.05 – 0.04	0.890
Setor Atividade [Comunicação e Informática]	0.01	-0.06 – 0.08	0.756
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	-0.05	-0.09 – -0.02	<b>0.004</b>
Setor Atividade [Educação]	-0.19	-0.28 – -0.11	<b>&lt;0.001</b>
Setor Atividade [Extração Mineral]	0.08	-0.04 – 0.20	0.184
Setor Atividade [Farmacêutico e Higiene]	0.37	0.28 – 0.45	<b>&lt;0.001</b>
Setor Atividade [Gráficas e Editoras]	-0.04	-0.16 – 0.07	0.486
Setor Atividade [Hospedagem e Turismo]	0.64	0.53 – 0.76	<b>&lt;0.001</b>
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	0.06	0.02 – 0.10	<b>0.007</b>
Setor Atividade [Metalurgia e Siderurgia]	0.05	0.01 – 0.09	<b>0.020</b>
Setor Atividade [Papel e Celulose]	0.33	0.27 – 0.39	<b>&lt;0.001</b>
Setor Atividade [Petróleo e Gás]	0.09	0.00 – 0.17	<b>0.049</b>
Setor Atividade [Petroquímicos e Borracha]	0.68	0.61 – 0.75	<b>&lt;0.001</b>
Setor Atividade [Saneamento, Serv. Água e Gás]	-0.24	-0.35 – -0.12	<b>&lt;0.001</b>
Setor Atividade [Sem Setor Principal]	-0.08	-0.13 – -0.03	<b>0.003</b>
Setor Atividade [Serviços médicos]	-0.13	-0.20 – -0.06	<b>&lt;0.001</b>
Setor Atividade [Serviços Transporte e Logística]	0.25	0.21 – 0.29	<b>&lt;0.001</b>
Setor Atividade [Telecomunicações]	0.04	-0.02 – 0.10	0.222
Setor Atividade [Têxtil e Vestuário]	0.12	0.07 – 0.16	<b>&lt;0.001</b>
Observations	8533		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.475 / 0.474		



##		GVIF	Df	GVIF^(1/(2*Df))
##	CARGA_TRIBUTARIA	1.176067	1	1.084466
##	GRAU_IMOBILZ_ATIVO	1.597596	1	1.263961
##	LIQ_CORRENTE	1.208319	1	1.099236
##	TAMANHO_FIRMA	1.467320	1	1.211330
##	factor(Setor_Atividade)	2.727001	22	1.023062

## Diagnóstico de ajuste do modelo significativo



Embora o teste de Kolmogorov-Smirnov tenha rejeitado a hipótese de normalidade, é crucial considerar a quantidade de dados, bem como a presença de outliers e pontos de alavancagem. Na análise visual temos que os resíduos tendem a se aproximar de uma distribuição normal, no entanto, há evidências de heterocedasticidade no comportamento dos resíduos.

Observa-se pontos de alavancagem que podem está impactando no ajuste do modelo.

## Tratamento de pontos de alavanca

Pontos de alavancagem retirados foram: 4637, 4636, 4814

Representando as seguintes empresas:

##	Observação	DENOM_CIA	Setor_Economico
## 1	4637	PADTEC HOLDING S.A.	Tecnologia da Informação
## 2	4636	PADTEC HOLDING S.A.	Tecnologia da Informação
## 3	4814	CIA PARTICIPACOES ALIANCA DA BAHIA	Outros

Calculando um novo valor otimo de  $\lambda$  para o modelo sem os pontos de alavancagem, temos:

```
## Valor de Lambda: -0.120534
```

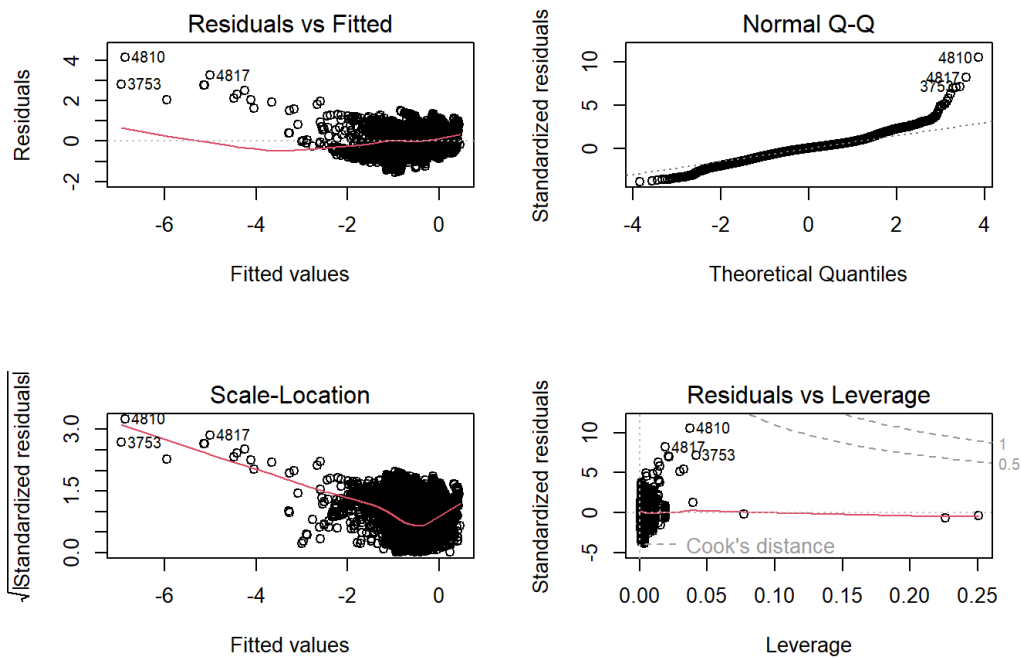
## Modelo Significativo reduzido sem pontos de alavancagem - Box cox

Predictors	y_trans		
	Estimates	CI	p
(Intercept)	0.35	0.25 - 0.46	<0.001
CARGA TRIBUTARIA	0.04	-0.02 - 0.10	0.227

GRAU IMOBILZ ATIVO	-0.52	-0.57 – -0.47	<0.001
LIQ CORRENTE	-0.27	-0.28 – -0.26	<0.001
TAMANHO FIRMA	-0.03	-0.05 – -0.02	<0.001
Setor Atividade [Agricultura (Açúcar, Álcool e Cana)]	0.29	0.21 – 0.37	<0.001
Setor Atividade [Alimentos]	0.19	0.15 – 0.24	<0.001
Setor Atividade [Brinquedos e Lazer]	0.33	0.24 – 0.41	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	-0.03	-0.07 – 0.01	0.161
Setor Atividade [Comunicação e Informática]	-0.07	-0.14 – -0.00	0.044
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	-0.07	-0.11 – -0.04	<0.001
Setor Atividade [Educação]	-0.21	-0.29 – -0.13	<0.001
Setor Atividade [Extração Mineral]	0.06	-0.05 – 0.17	0.288
Setor Atividade [Farmacêutico e Higiene]	0.33	0.25 – 0.41	<0.001
Setor Atividade [Gráficas e Editoras]	-0.06	-0.17 – 0.05	0.304
Setor Atividade [Hospedagem e Turismo]	0.58	0.47 – 0.70	<0.001
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	0.02	-0.02 – 0.06	0.317
Setor Atividade [Metalurgia e Siderurgia]	0.04	-0.00 – 0.07	0.068
Setor Atividade [Papel e Celulose]	0.33	0.27 – 0.38	<0.001
Setor Atividade [Petróleo e Gás]	0.17	0.09 – 0.25	<0.001
Setor Atividade [Petroquímicos e Borracha]	0.63	0.56 – 0.69	<0.001
Setor Atividade [Saneamento, Serv. Água e Gás]	-0.27	-0.38 – -0.16	<0.001
Setor Atividade [Sem Setor Principal]	-0.09	-0.14 – -0.04	<0.001
Setor Atividade [Serviços médicos]	-0.15	-0.22 – -0.08	<0.001

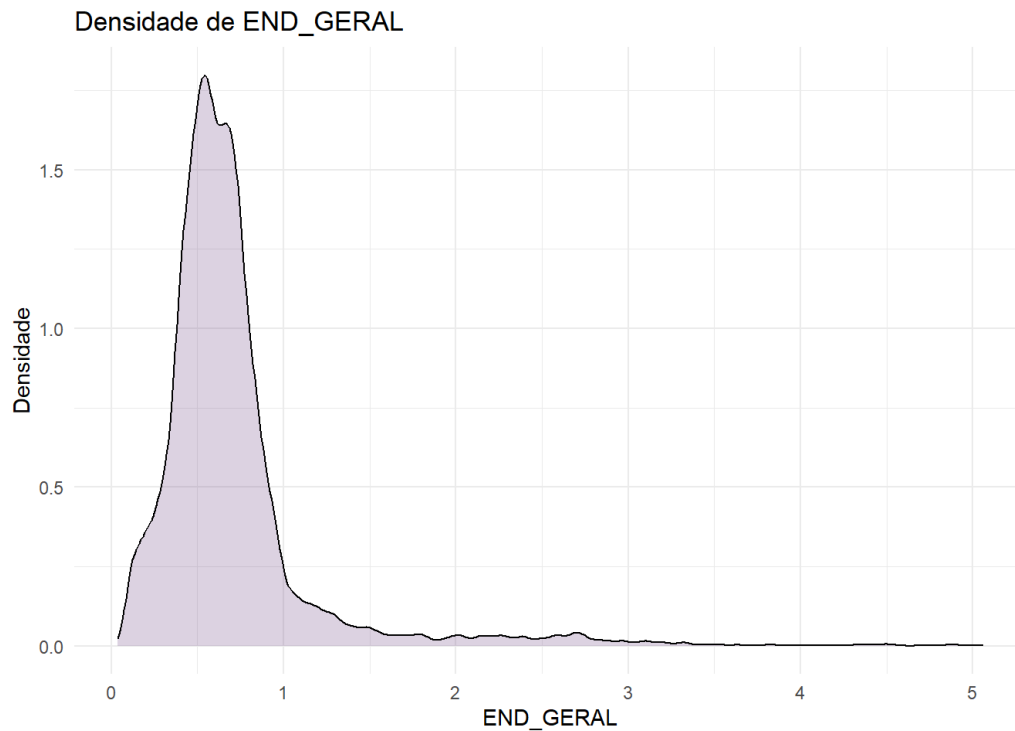
Sector Activity [Services Transport and Logistics]	0.21	0.17 - 0.25	<0.001
Sector Activity [Telecommunications]	0.04	-0.02 - 0.10	0.188
Sector Activity [Textile and Clothing]	0.08	0.04 - 0.12	<0.001
Observations	8530		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.524 / 0.522		

### Diagnóstico de ajuste do modelo significativo reduzido sem pontos de alavancagem



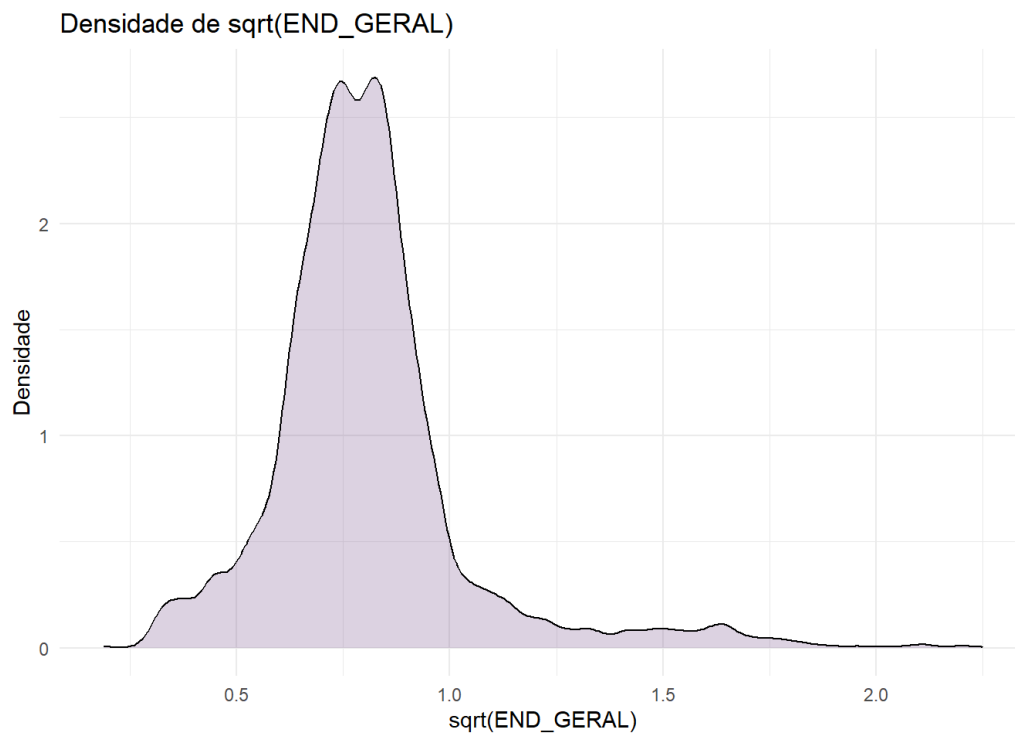
Após remover os pontos de alavancagem e aplicar uma nova transformação de Box-Cox nos dados, não foi observada uma mudança significativa no ajuste do modelo.

# Modelo Linear Generalizado



Ao observar a distribuição da variável resposta END\_GERAL, consta-se que ela apresenta valores positivos e uma assimetria à esquerda. Devido a essas características, optou-se por utilizar a distribuição Gamma para o ajuste do Modelo Linear Generalizado (MLG). A distribuição Gamma é apropriada para modelar dados positivos e assimétricos, fornecendo uma melhor adequação aos dados observados.

Devido aos problemas de convergência do modelo, causados pela presença de um número considerável de valores extremos na variável resposta, foi necessário aplicar a transformação pela raiz quadrada para corrigir a variabilidade dos dados.



# Modelo Completo

Predictors	sqrt(END_GERAL)		
	Estimates	CI	p
(Intercept)	1.36	1.28 – 1.44	<0.001
CARGA TRIBUTARIA	0.41	0.36 – 0.48	<0.001
DISTR RIQUEZA GOV	1.00	1.00 – 1.00	0.858
LUCRATIVIDADE	1.02	1.02 – 1.02	<0.001
ROE	1.00	1.00 – 1.00	0.505
GRAU IMOBILZ ATIVO	0.86	0.84 – 0.88	<0.001
LIQ CORRENTE	0.92	0.92 – 0.93	<0.001
GIRO ATIVO	0.98	0.97 – 0.99	0.006
TAMANHO FIRMA	0.96	0.95 – 0.96	<0.001
RISCO NEGOCIO	1.00	1.00 – 1.00	0.861
Cresc Receita	1.00	1.00 – 1.00	0.075
Maior CT	0.90	0.88 – 0.92	<0.001
Aumento Carga Trib	1.03	1.02 – 1.05	<0.001
CT X MaioresCT	3.30	2.82 – 3.85	<0.001
CT X AumentoCT	0.81	0.75 – 0.88	<0.001
Setor Atividade [Agricultura (Açúcar, Álcool e Cana)]	1.17	1.12 – 1.22	<0.001
Setor Atividade [Alimentos]	1.12	1.09 – 1.15	<0.001
Setor Atividade [Brinquedos e Lazer]	1.32	1.26 – 1.38	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	0.99	0.96 – 1.01	0.262
Setor Atividade [Comunicação e Informática]	1.05	1.02 – 1.09	0.005
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	0.95	0.93 – 0.97	<0.001
Setor Atividade [Educação]	0.89	0.86 – 0.93	<0.001
Setor Atividade [Extração Mineral]	1.06	1.00 – 1.13	0.054
Setor Atividade [Farmacêutico e Higiene]	1.21	1.16 – 1.27	<0.001
Setor Atividade [Gráficas e Editoras]	0.98	0.92 – 1.04	0.467
Setor Atividade [Hospedagem e Turismo]	1.35	1.27 – 1.43	<0.001

Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	1.02	1.00 – 1.05	<b>0.036</b>
Setor Atividade [Metalurgia e Siderurgia]	1.00	0.98 – 1.02	0.799
Setor Atividade [Papel e Celulose]	1.12	1.09 – 1.16	<b>&lt;0.001</b>
Setor Atividade [Petróleo e Gás]	1.00	0.95 – 1.04	0.874
Setor Atividade [Petroquímicos e Borracha]	1.50	1.45 – 1.56	<b>&lt;0.001</b>
Setor Atividade [Saneamento, Serv. Água e Gás]	0.92	0.87 – 0.97	<b>0.005</b>
Setor Atividade [Sem Setor Principal]	0.96	0.93 – 0.98	<b>0.001</b>
Setor Atividade [Serviços médicos]	0.94	0.91 – 0.98	<b>0.002</b>
Setor Atividade [Serviços Transporte e Logística]	1.14	1.12 – 1.17	<b>&lt;0.001</b>
Setor Atividade [Telecomunicações]	1.03	0.99 – 1.06	0.115
Setor Atividade [Têxtil e Vestuário]	1.09	1.07 – 1.12	<b>&lt;0.001</b>
Observations	8533		
R <sup>2</sup> Nagelkerke	0.453		

##	GVIF	Df	GVIF^(1/(2*Df))
## CARGA_TRIBUTARIA	28.663094	1	5.353793
## DISTR_RIQUEZA_GOV	1.004363	1	1.002179
## LUCRATIVIDADE	1.832003	1	1.353515
## ROE	1.004707	1	1.002351
## GRAU_IMOBILZ_ATIVO	1.717401	1	1.310497
## LIQ_CORRENTE	1.268600	1	1.126321
## GIRO_ATIVO	1.351876	1	1.162702
## TAMANHO_FIRMA	1.579773	1	1.256890
## RISCO_NEGOCIO	1.021900	1	1.010891
## Cresc_Receita	1.003219	1	1.001608
## Maior_CT	5.198213	1	2.279959
## Aumento_Carga_Trib	2.611623	1	1.616052
## CT_X_MaioresCT	40.676308	1	6.377798
## CT_X_AumentoCT	8.953000	1	2.992156
## factor(Setor_Atividade)	4.785576	22	1.036223

O modelo completo tem algumas variáveis não significativas e altamente correlacionada, retirando elas temos o modelo significativo abaixo

## Modelo Significativo

	sqrt(END_GERAL)		
Predictors	Estimates	CI	p
(Intercept)	1.19	1.12 – 1.28	<b>&lt;0.001</b>

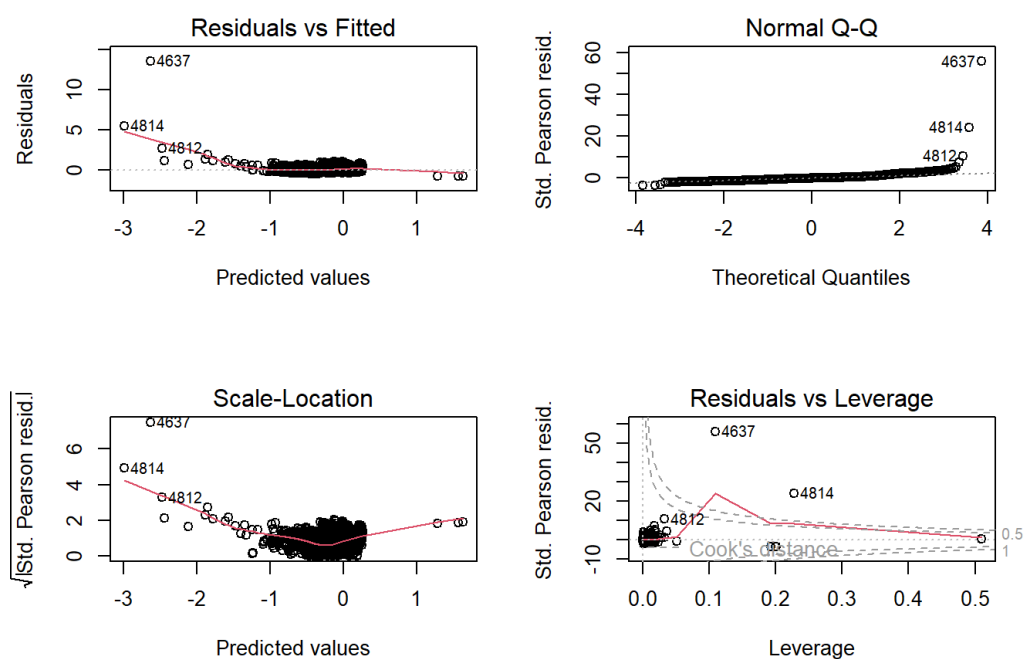
CARGA TRIBUTARIA	1.20	1.15 – 1.25	<0.001
LUCRATIVIDADE	1.01	1.01 – 1.01	<0.001
GRAU IMOBILIZ ATIVO	0.83	0.80 – 0.85	<0.001
LIQ CORRENTE	0.91	0.90 – 0.91	<0.001
TAMANHO FIRMA	0.97	0.96 – 0.98	<0.001
Setor Atividade [Agricultura (Açúcar, Álcool e Cana)]	1.23	1.17 – 1.30	<0.001
Setor Atividade [Alimentos]	1.15	1.11 – 1.18	<0.001
Setor Atividade [Brinquedos e Lazer]	1.29	1.22 – 1.36	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	1.00	0.97 – 1.02	0.712
Setor Atividade [Comunicação e Informática]	1.01	0.96 – 1.05	0.803
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	0.97	0.95 – 0.99	0.011
Setor Atividade [Educação]	0.91	0.86 – 0.96	<0.001
Setor Atividade [Extração Mineral]	1.05	0.98 – 1.13	0.201
Setor Atividade [Farmacêutico e Higiene]	1.19	1.13 – 1.25	<0.001
Setor Atividade [Gráficas e Editoras]	0.96	0.89 – 1.03	0.221
Setor Atividade [Hospedagem e Turismo]	1.35	1.26 – 1.46	<0.001
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	1.03	1.00 – 1.05	0.034
Setor Atividade [Metalurgia e Siderurgia]	1.03	1.01 – 1.06	0.007
Setor Atividade [Papel e Celulose]	1.16	1.12 – 1.20	<0.001
Setor Atividade [Petróleo e Gás]	1.04	0.98 – 1.09	0.188
Setor Atividade [Petroquímicos e Borracha]	1.46	1.40 – 1.53	<0.001
Setor Atividade [Saneamento, Serv. Água e Gás]	0.88	0.82 – 0.95	0.001
Setor Atividade [Sem Setor Principal]	0.98	0.95 – 1.01	0.215

Setor Atividade [Serviços médicos]	0.93	0.89 – 0.97	0.001
Setor Atividade [Serviços Transporte e Logística]	1.13	1.10 – 1.16	<0.001
Setor Atividade [Telecomunicações]	1.03	0.99 – 1.06	0.203
Setor Atividade [Têxtil e Vestuário]	1.09	1.06 – 1.12	<0.001

Observations 8533  
R<sup>2</sup> Nagelkerke 0.441

##	GVIF	Df	GVIF^(1/(2*Df))
## CARGA_TRIBUTARIA	1.738462	1	1.318508
## LUCRATIVIDADE	1.604764	1	1.266793
## GRAU_IMOBILZ_ATIVO	1.597766	1	1.264028
## LIQ_CORRENTE	1.246664	1	1.116541
## TAMANHO_FIRMA	1.480332	1	1.216689
## factor(Setor_Atividade)	2.902770	22	1.024515

## Diagnóstico de ajuste do modelo significativo



Na análise visual temos que os resíduos tendem a se aproximar de uma distribuição normal, no entanto, há evidências de heterocedasticidade no comportamento dos resíduos.

Observa-se pontos de alavancagem que podem está impactando no ajuste do modelo.

## Tratamento de pontos de alavanca

## Modelo Significativo sem pontos de alavacagem

- Pontos removidos: 4636, 4637, 4814, 4812

Representando as seguintes empresas:

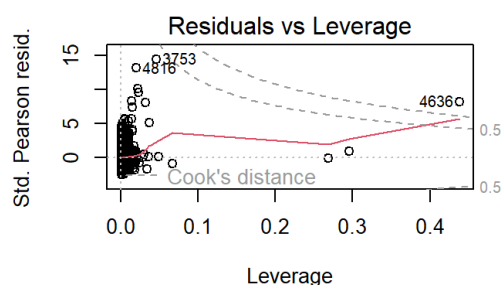
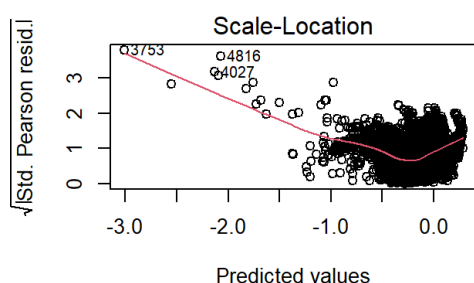
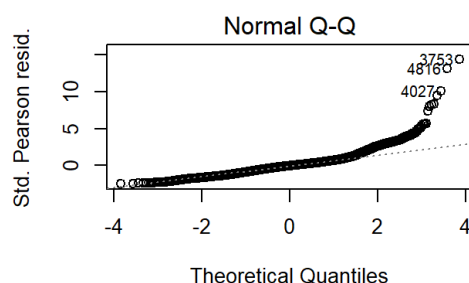
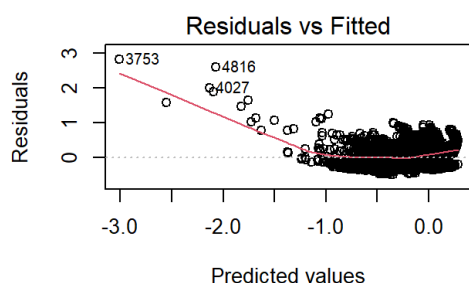


##	Observação	DENOM_CIA	Setor_Economico
## 1	4636	PADTEC HOLDING S.A.	Tecnologia da Informação
## 2	4637	PADTEC HOLDING S.A.	Tecnologia da Informação
## 3	4814	CIA PARTICIPACOES ALIANCA DA BAHIA	Outros
## 4	4812	CIA PARTICIPACOES ALIANCA DA BAHIA	Outros

Predictors	sqrt(END_GERAL)		
	Estimates	CI	p
(Intercept)	1.24	1.17 – 1.31	<0.001
CARGA TRIBUTARIA	1.14	1.10 – 1.18	<0.001
LUCRATIVIDADE	0.98	0.98 – 0.99	<0.001
GRAU IMOBILIZ ATIVO	0.80	0.78 – 0.82	<0.001
LIQ CORRENTE	0.89	0.89 – 0.90	<0.001
GIRO ATIVO	0.99	0.97 – 1.00	0.033
TAMANHO FIRMA	0.97	0.96 – 0.98	<0.001
Setor Atividade [Agricultura (Açúcar, Álcool e Cana)]	1.21	1.16 – 1.26	<0.001
Setor Atividade [Alimentos]	1.14	1.11 – 1.17	<0.001
Setor Atividade [Brinquedos e Lazer]	1.24	1.19 – 1.29	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	1.00	0.98 – 1.02	0.986
Setor Atividade [Comunicação e Informática]	0.98	0.94 – 1.01	0.201
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	0.97	0.95 – 0.99	0.001
Setor Atividade [Educação]	0.92	0.88 – 0.96	<0.001
Setor Atividade [Extração Mineral]	1.05	0.99 – 1.11	0.098
Setor Atividade [Farmacêutico e Higiene]	1.18	1.14 – 1.23	<0.001
Setor Atividade [Gráficas e Editoras]	0.96	0.91 – 1.02	0.165
Setor Atividade [Hospedagem e Turismo]	1.33	1.26 – 1.41	<0.001
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	1.02	1.00 – 1.04	0.092
Setor Atividade [Metalurgia e Siderurgia]	1.04	1.02 – 1.06	<0.001

Sector Activity [Paper and Cellulose]	1.17	1.14 – 1.20	<0.001
Sector Activity [Petroleum and Gas]	1.07	1.03 – 1.12	0.001
Sector Activity [Petrochemicals and Rubber]	1.44	1.40 – 1.50	<0.001
Sector Activity [Sanitation, Water and Gas]	0.88	0.83 – 0.93	<0.001
Sector Activity [Main Sector]	0.99	0.97 – 1.02	0.458
Sector Activity [Medical Services]	0.94	0.91 – 0.97	<0.001
Sector Activity [Transport and Logistics Services]	1.12	1.10 – 1.14	<0.001
Sector Activity [Telecommunications]	1.03	1.00 – 1.06	0.081
Sector Activity [Textile and Apparel]	1.08	1.05 – 1.10	<0.001
Observations	8529		
R <sup>2</sup> Nagelkerke	0.519		

## Diagnóstico de ajuste do modelo significativo sem pontos de alavacagem



Observa-se depois da remoção dos pontos de alavancagem o surgimento de outros pontos de alavanca, causando um leve desvio na calda superior no gráfico Normal Q-Q, a tendência de curvatura sugere a presença de não linearidade e a dispersão crescente dos resíduos indica a presença de heterocedasticidade (variância dos resíduos não constante).

# Comparando Modelo Siginificativos

Sem alavanca X Com alavanca:

Predictors	sqrt(END_GERAL)			sqrt(END_GERAL)		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	1.24	1.17 – 1.31	<0.001	1.19	1.12 – 1.28	<0.001
CARGA TRIBUTARIA	1.14	1.10 – 1.18	<0.001	1.20	1.15 – 1.25	<0.001
LUCRATIVIDADE	0.98	0.98 – 0.99	<0.001	1.01	1.01 – 1.01	<0.001
GRAU IMOBILZ ATIVO	0.80	0.78 – 0.82	<0.001	0.83	0.80 – 0.85	<0.001
LIQ CORRENTE	0.89	0.89 – 0.90	<0.001	0.91	0.90 – 0.91	<0.001
GIRO ATIVO	0.99	0.97 – 1.00	0.033			
TAMANHO FIRMA	0.97	0.96 – 0.98	<0.001	0.97	0.96 – 0.98	<0.001
Setor Atividade [Agricultura (Açúcar, Álcool e Cana)]	1.21	1.16 – 1.26	<0.001	1.23	1.17 – 1.30	<0.001
Setor Atividade [Alimentos]	1.14	1.11 – 1.17	<0.001	1.15	1.11 – 1.18	<0.001
Setor Atividade [Brinquedos e Lazer]	1.24	1.19 – 1.29	<0.001	1.29	1.22 – 1.36	<0.001
Setor Atividade [Comércio (Atacado e Varejo)]	1.00	0.98 – 1.02	0.986	1.00	0.97 – 1.02	0.712
Setor Atividade [Comunicação e Informática]	0.98	0.94 – 1.01	0.201	1.01	0.96 – 1.05	0.803
Setor Atividade [Construção Civil, Mat. Constr. e Decoração]	0.97	0.95 – 0.99	0.001	0.97	0.95 – 0.99	0.011
Setor Atividade [Educação]	0.92	0.88 – 0.96	<0.001	0.91	0.86 – 0.96	<0.001
Setor Atividade [Extração Mineral]	1.05	0.99 – 1.11	0.098	1.05	0.98 – 1.13	0.201
Setor Atividade [Farmacêutico e Higiene]	1.18	1.14 – 1.23	<0.001	1.19	1.13 – 1.25	<0.001
Setor Atividade [Gráficas e Editoras]	0.96	0.91 – 1.02	0.165	0.96	0.89 – 1.03	0.221
Setor Atividade [Hospedagem e Turismo]	1.33	1.26 – 1.41	<0.001	1.35	1.26 – 1.46	<0.001
Setor Atividade [Máquinas, Equipamentos, Veículos e Peças]	1.02	1.00 – 1.04	0.092	1.03	1.00 – 1.05	0.034
Setor Atividade [Metalurgia e Siderurgia]	1.04	1.02 – 1.06	<0.001	1.03	1.01 – 1.06	0.007
Setor Atividade [Papel e Celulose]	1.17	1.14 – 1.20	<0.001	1.16	1.12 – 1.20	<0.001
Setor Atividade [Petróleo e Gás]	1.07	1.03 – 1.12	0.001	1.04	0.98 – 1.09	0.188

Setor Atividade [Petroquímicos e Borracha]	1.44	1.40 – 1.50	<0.001	1.46	1.40 – 1.53	<0.001
Setor Atividade [Saneamento, Serv. Água e Gás]	0.88	0.83 – 0.93	<0.001	0.88	0.82 – 0.95	0.001
Setor Atividade [Sem Setor Principal]	0.99	0.97 – 1.02	0.458	0.98	0.95 – 1.01	0.215
Setor Atividade [Serviços médicos]	0.94	0.91 – 0.97	<0.001	0.93	0.89 – 0.97	0.001
Setor Atividade [Serviços Transporte e Logística]	1.12	1.10 – 1.14	<0.001	1.13	1.10 – 1.16	<0.001
Setor Atividade [Telecomunicações]	1.03	1.00 – 1.06	0.081	1.03	0.99 – 1.06	0.203
Setor Atividade [Têxtil e Vestuário]	1.08	1.05 – 1.10	<0.001	1.09	1.06 – 1.12	<0.001
Observations	8529			8533		
R <sup>2</sup> Nagelkerke	0.519			0.441		

## Teste AIC:

## AIC do modelo Siginficativo Sem alavanca: -8275.461

## AIC do modelo Siginficativo Com alavanca: -7008.732

Com base nos valores de AIC apresentados, o modelo sem pontos de alavancagem (-8275.461) é preferível ao modelo com pontos de alavancagem (-7008.732). A remoção dos pontos de alavancagem resultou em um modelo que proporciona um melhor ajuste aos dados, ao mesmo tempo em que mantém a complexidade em um nível adequado. Portanto, para obter melhores previsões e inferências, o modelo sem pontos de alavancagem deve ser considerado.

## Previsão

## RMSE do modelo com pontos de alavanca: 0.4161833

## MAE do modelo com pontos de alavanca: 0.2692005

## RMSE do modelo sem pontos de alavanca: 0.416082

## MAE do modelo sem pontos de alavanca: 0.2690156

RMSE (Root Mean Squared Error ou Raiz do Erro Quadrático Médio):

- Modelo com pontos de alavancagem: 0.4161833
- Modelo sem pontos de alavancagem: 0.416082

O RMSE é ligeiramente menor para o modelo sem pontos de alavancagem (0.416082) em comparação com o modelo com pontos de alavancagem (0.4161833). Isso sugere que o modelo sem pontos de alavancagem tem um desempenho um pouco melhor na previsão dos dados de teste, embora a diferença seja muito pequena.

MAE (Mean Absolute Error ou Erro Absoluto Médio):

- Modelo com pontos de alavancagem: 0.2692005

- Modelo sem pontos de alavancagem: 0.2690156

O MAE é também ligeiramente menor para o modelo sem pontos de alavancagem (0.2690156) em comparação com o modelo com pontos de alavancagem (0.2692005). Assim como no caso do RMSE, a diferença é muito pequena, indicando que a remoção dos pontos de alavancagem melhora marginalmente a precisão do modelo.

Os resultados indicam que a remoção dos pontos de alavancagem melhora ligeiramente o desempenho preditivo do modelo. A redução nos valores de RMSE e MAE para o modelo sem pontos de alavancagem sugere que ele fornece previsões mais precisas. No entanto, a diferença entre os dois modelos é muito pequena, o que implica que ambos os modelos têm desempenhos muito semelhantes.

Portanto, embora o modelo sem pontos de alavancagem seja ligeiramente preferível, a diferença na performance não é substancial. Isso sugere que a presença dos pontos de alavancagem não afeta significativamente a capacidade preditiva dos modelos neste caso específico.

## Considerações Finais

Dado o menor valor de AIC e o melhor desempenho preditivo, o modelo sem os pontos de alavancagem é considerado o melhor modelo. Apesar dos pressupostos de linearidade, homocedasticidade e não normalidade dos resíduos não ser atendido, a interpretação para esse conjunto de dados é válida. Isso sugere que, mesmo com desvios dos pressupostos, as conclusões derivadas do modelo são confiáveis para o conjunto de dados analisado.

## Modelo Final

$$\sqrt{\text{END\_GERAL}} = \exp \left( \begin{aligned} &\beta_0 + \beta_1 \cdot \text{CARGA\_TRIBUTARIA} + \beta_2 \cdot \text{LUCRATIVIDADE} \\ &+ \beta_3 \cdot \text{GRAU\_IMOBILZ\_ATIVO} + \beta_4 \cdot \text{LIQ\_CORRENTE} \\ &+ \beta_5 \cdot \text{GIRO\_ATIVO} + \beta_6 \cdot \text{TAMANHO\_FIRMA} \\ &+ \sum_{j=7}^{28} \beta_j \cdot \text{factor}(\text{Setor\_Atividade})_j + \epsilon \end{aligned} \right)$$

# Exponenciais dos Coeficientes ao quadrado

##	Valor
## (Intercept)	1.5364546
## CARGA_TRIBUTARIA	1.2972775
## LUCRATIVIDADE	0.9678195
## GRAU_IMOBILZ_ATIVO	0.6412223
## LIQ_CORRENTE	0.7983250
## GIRO_ATIVO	0.9734055
## TAMANHO_FIRMA	0.9422146
## factor(Setor_Atividade)Agricultura (Açúcar, Álcool e Cana)	1.4643179
## factor(Setor_Atividade)Alimentos	1.2964474
## factor(Setor_Atividade)Brinquedos e Lazer	1.5382281
## factor(Setor_Atividade)Comércio (Atacado e Varejo)	0.9996173
## factor(Setor_Atividade)Comunicação e Informática	0.9554377
## factor(Setor_Atividade)Construção Civil, Mat. Constr. e Decoração	0.9396214
## factor(Setor_Atividade)Educação	0.8449054
## factor(Setor_Atividade)Extração Mineral	1.1001805
## factor(Setor_Atividade)Farmacêutico e Higiene	1.3970402
## factor(Setor_Atividade)Gráficas e Editoras	0.9239603
## factor(Setor_Atividade)Hospedagem e Turismo	1.7626025
## factor(Setor_Atividade)Máquinas, Equipamentos, Veículos e Peças	1.0350234
## factor(Setor_Atividade)Metalurgia e Siderurgia	1.0810498
## factor(Setor_Atividade)Papel e Celulose	1.3637979
## factor(Setor_Atividade)Petróleo e Gás	1.1540462
## factor(Setor_Atividade)Petroquímicos e Borracha	2.0878715
## factor(Setor_Atividade)Saneamento, Serv. Água e Gás	0.7771501
## factor(Setor_Atividade)Sem Setor Principal	0.9810117
## factor(Setor_Atividade)Serviços médicos	0.8822149
## factor(Setor_Atividade)Serviços Transporte e Logística	1.2545968
## factor(Setor_Atividade)Telecomunicações	1.0544979
## factor(Setor_Atividade)Têxtil e Vestuário	1.1577194

## Interpretação dos coeficientes

1. Intercept (Intercepto): O valor esperado do Endividamento Geral quando todos os preditores são zero é aproximadamente 1.54.
2. CARGA\_TRIBUTARIA: Para um aumento de uma unidade na carga tributária, espera-se um aumento de aproximadamente 29,73% no Endividamento Geral (1.2972775 - 1).
3. LUCRATIVIDADE: Para um aumento de uma unidade na lucratividade, espera-se uma redução de aproximadamente 3,22% no Endividamento Geral (0.9678195-1).
4. GRAU\_IMOBILZ\_ATIVO: Para um aumento de uma unidade no grau de imobilização do ativo, espera-se uma redução de aproximadamente 35,88% no Endividamento Geral (0.6412223-1).
5. LIQ\_CORRENTE: Para um aumento de uma unidade na liquidez corrente, espera-se uma redução de aproximadamente 20,17% no Endividamento Geral (0.7983250-1).
6. GIRO\_ATIVO: Para um aumento de uma unidade no giro do ativo, espera-se uma redução de aproximadamente 2,66% no Endividamento Geral (0.9734055-1).
7. TAMANHO\_FIRMA: Para um aumento de uma unidade no tamanho da firma, espera-se uma redução de aproximadamente 5,78% no Endividamento Geral (0.9705425-1).
8. Setor de Atividade: Cada coeficiente para os diferentes setores de atividade indica o efeito relativo desse setor em comparação ao setor de referência (Energia Eletrica). Por exemplo:

- Para as empresas do setor de Agricultura (Açúcar, Alcool e Cana), espera-se um aumento no Endividamento Geral cerca de 46,43% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Alimentos, espera-se um aumento no Endividamento Geral cerca de 29,64% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Brinquedos e Lazer, espera-se um aumento no Endividamento Geral cerca de 53,82% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Comércio (Atacado e Varejo), espera-se uma redução no Endividamento Geral cerca de 0,04% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Comunicação e Informática, espera-se uma redução no Endividamento Geral cerca de 4,46% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Construção Civil, Mat. Constr. e Decoração, espera-se uma redução no Endividamento Geral cerca de 6,04% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Educação, espera-se uma redução no Endividamento Geral cerca de 15,51% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Extração Mineral, espera-se um aumento no Endividamento Geral cerca de 10,02% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Farmacêutico e Higiene, espera-se um aumento no Endividamento Geral cerca de 39,7% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Gráficas e Editoras, espera-se uma redução no Endividamento Geral cerca de 7,6% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Hospedagem e Turismo, espera-se um aumento no Endividamento Geral cerca de 76,26% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Máquinas, Equipamentos, Veículos e Peças, espera-se um aumento no Endividamento Geral cerca de 3,5% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Metalurgia e Siderurgia, espera-se um aumento no Endividamento Geral cerca de 8,1% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Papel e Celulose, espera-se um aumento no Endividamento Geral cerca de 36,38% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Petróleo e Gás, espera-se um aumento no Endividamento Geral cerca de 15,4% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Petroquímicos e Borracha, espera-se um aumento no Endividamento Geral cerca de 108,77% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Saneamento, Serv. Água e Gás, espera-se uma redução no Endividamento Geral cerca de 22,28% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Sem Setor Principal, espera-se uma redução no Endividamento Geral cerca de 1,9% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Serviços médicos, espera-se uma redução no Endividamento Geral cerca de 11,78% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Serviços Transporte e Logística, espera-se um aumento no Endividamento Geral cerca de 25,46% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Telecomunicações, espera-se um aumento no Endividamento Geral cerca de 5,45% em comparação com o Setor de Energia Eletrica;
- Para as empresas do setor de Têxtil e Vestuário, espera-se um aumento no Endividamento Geral cerca de 15,77% em comparação com o Setor de Energia Eletrica;

## Razão de chances

- A razão de chances compara as chances de um evento ocorrer entre dois grupos diferentes. No contexto do modelo gamma, podemos calcular uma forma de razão de chances utilizando os coeficientes exponenciados.

Calcular a razão de chances entre duas variáveis, por exemplo, entre "CARGA\_TRIBUTARIA" e "LUCRATIVIDADE":

$$\exp(\beta_{\text{CARGA\_TRIBUTARIA}})^2 = 1.2972775$$

$$\exp(\beta_{\text{LUCRATIVIDADE}})^2 = 0.9678195$$

$$RC_{\text{CARGA\_TRIBUTARIA vs LUCRATIVIDADE}} \approx \frac{\exp(\beta_{\text{CARGA\_TRIBUTARIA}})}{\exp(\beta_{\text{LUCRATIVIDADE}})} = \frac{1.2972775}{0.9678195} \approx 1.34$$

Para um aumento unitário na variável CARGA\_TRIBUTARIA, as chances de impacto no Endividamento Geral são aproximadamente 1.34 vezes maiores do que para um aumento unitário na variável LUCRATIVIDADE

## Referências

1. MORETIN, L.G. Estatística Básica. São Paulo: Makron Books, 7ª ed., 1999.
  2. TRIOLA, M. F. Introdução à Estatística. Tradução Vera Regina Lima de Farias e Flores. 10ª ed. Rio de Janeiro: LTC, 2008.
  3. Box, G. E. P. e Cox, D. R. (1964). An analysis of transformations (with discussion). Journal of the Royal Statistical Society B 26, 211-252.
  4. Gilberto A. Paula. MODELOS DE REGRESSÃO: com apoio computacional. São Paulo, 2024. 475 p.
- 

Site (<http://www.estatmg.com.br/>)

Instagram  
(<https://www.instagram.com/estatmg/>)

LinkedIn  
(<https://www.linkedin.com/company/estatmg>)