



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA**

AMAURI DOS SANTOS LIMA NETO

**MODELOS LINEARES GENERALIZADOS: UMA APLICAÇÃO A
DADOS VOCAIS**

**SOLÂNEA
2024**

1 Introdução

A voz humana é uma ferramenta poderosa de comunicação, expressão e identidade pessoal(4). No entanto, a produção vocal pode ser afetada por diversas condições que resultam em irregularidades na voz. Essas irregularidades podem se manifestar de várias formas, como rouquidão, aspereza, tremor, variação no tom ou volume, e podem ser causadas por fatores temporários ou permanentes. As causas incluem desde abusos vocais, infecções respiratórias, até mesmo problemas mais sérios como lesões nas cordas vocais, paralisia laríngea, ou condições neurológicas. Compreender a origem e o impacto dessas irregularidades é crucial para o diagnóstico e tratamento adequado, visando a recuperação ou melhoria da qualidade vocal e, conseqüentemente, da comunicação e qualidade de vida do indivíduo.

2 Metodologia

Modelos Lineares Generalizados (MLG) são uma extensão dos modelos lineares que permitem uma maior flexibilidade na modelagem de variáveis dependentes que não seguem uma distribuição normal (Cordeiro e Demétrio 2008). Eles são especialmente úteis quando a relação entre a variável resposta e as variáveis preditoras não é linear. Os MLGs foram introduzidos por John Nelder e Robert Wedderburn em 1972. Eles generalizaram os modelos lineares tradicionais para acomodar distribuições da família exponencial e diferentes funções de ligação, permitindo a modelagem de uma variedade maior de dados.

2.1 Função de Ligação

A função de ligação $g(\mu)$ relaciona a média da variável resposta μ com a combinação linear dos preditores η . No caso de uma distribuição gama com função de ligação log, temos:

$$g(\mu) = \log(\mu) \quad (1)$$

$$\eta = X\beta \quad (2)$$

onde X é a matriz de preditores e β é o vetor de coeficientes.

A função de ligação é crucial porque determina como as variáveis preditoras influenciam a variável resposta. Ela permite a transformação de uma relação não linear em uma linear.

2.2 Distribuição da Família Exponencial

Os MLGs são baseados em distribuições da família exponencial, que incluem distribuições comuns como a normal, binomial, Poisson, e gama. A forma geral da função de

densidade da família exponencial é:

$$f(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right) \quad (3)$$

onde θ é o parâmetro canônico, ϕ é o parâmetro de dispersão, e $b(\theta)$ e $c(y, \phi)$ são funções específicas da distribuição.

2.3 Estrutura Linear e Componente de Variância

Nos MLGs, a estrutura linear é representada por $\eta = X\beta$. A variância da variável resposta é uma função da média μ e do parâmetro de dispersão ϕ :

$$\text{Var}(Y) = \phi V(\mu) \quad (4)$$

onde $V(\mu)$ é a função de variância. No caso da distribuição gama, temos $V(\mu) = \mu^2$.

2.4 Distribuição Gama

A distribuição gama é apropriada para modelar variáveis positivas contínuas com distribuição assimétrica. Com a função de ligação log, temos:

$$g(\mu) = \log(\mu) \quad (5)$$

2.5 Métodos de Estimação de Parâmetros

Os parâmetros dos MLGs são tipicamente estimados usando o método da máxima verossimilhança. Comparado aos mínimos quadrados, a máxima verossimilhança é mais flexível para distribuições não normais.

2.6 Função Desvio (Deviance)

A função desvio (deviance) mede a qualidade do ajuste do modelo comparando um modelo ajustado com um modelo saturado. O desvio é definido como:

$$D = 2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i) \right) \quad (6)$$

onde y_i são os valores observados e $\hat{\mu}_i$ são os valores ajustados pelo modelo. Para a distribuição gama, esta fórmula se adapta à forma específica da distribuição.

2.7 Avaliação de Ajuste do Modelo

Para avaliar a adequação do modelo, usamos critérios como o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion), além de testes de bondade de ajuste

e a análise do desvio.

2.8 Overfitting e Underfitting

Overfitting ocorre quando o modelo se ajusta excessivamente aos dados de treinamento, enquanto underfitting ocorre quando o modelo é muito simples. Estratégias como validação cruzada e regularização ajudam a mitigar esses problemas.

2.9 Interpretabilidade dos Resultados

Tornar os resultados dos MLGs interpretáveis envolve a escolha apropriada de funções de ligação e transformações, além de técnicas de visualização de dados.

2.10 vantagens

- **Modelo Linear Generalizado e Modelo Linear Comum:**

MLGs permitem a modelagem de variáveis resposta que seguem distribuições não normais e utilizam funções de ligação para transformar relações não lineares em lineares.

- **Modelagem:**

Flexibilidade em modelar diferentes tipos de dados e variáveis resposta.

- **Modelo de Poisson para prever eventos raros:**

Sim, modelos de Poisson são adequados para contagem de eventos raros.

- **Variedade de funções de ligação**

A escolha da função de ligação depende da natureza da variável resposta e da distribuição dos dados.

- **robustez contra outliers:**

A robustez contra outliers depende da escolha da distribuição e da função de ligação. Algumas distribuições são mais sensíveis a outliers do que outras.

2.11 Modelo Linear Generalizado com Distribuição Gama e Função de Ligação Log

O modelo linear generalizado (MLG) com distribuição Gama e função de ligação log é definido por:

1. **Distribuição:** A variável resposta Y segue uma distribuição Gama com função de densidade de probabilidade dada por:

$$f_Y(y; \mu, \phi) = \frac{y^{\phi-1} e^{-y/(\mu\phi)}}{\Gamma(\phi)(\mu\phi)^\phi}$$

onde μ é o parâmetro da média e ϕ é o parâmetro de dispersão.

2. **Modelo Completo:** Combinando as relações acima, obtemos:

$$\log(\mu) = X\beta$$

ou equivalentemente:

$$\mu = e^{X\beta}$$

3 Descrição dos dados

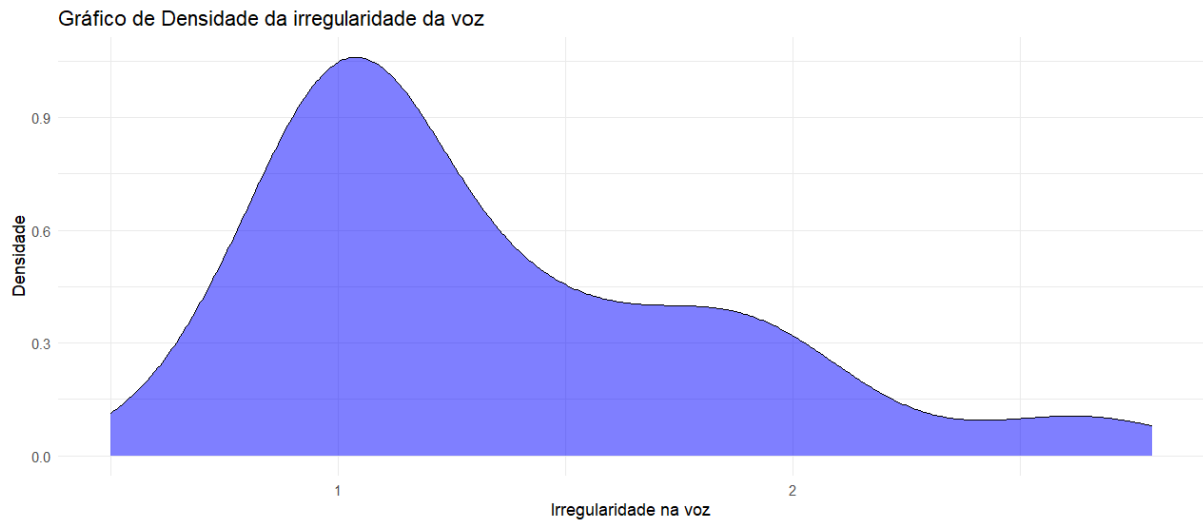
Desta forma, O banco de dados utilizado trás informações sobre a avaliação na irregularidade na voz de 131 pacientes por meio de 5 variáveis e está disponível no kaggle (1).

- **jitter:** Nervosismo do paciente
- **shimmer:** mede a oscilação de atingir tons afinados
- **GNE:** impacto genético na voz
- **Irregularity:** variável resposta que mede a irregularidade da voz do paciente
- **Noise:** Barulho medido do paciente
- **OverallSeverity:** Gravidade geral da voz do paciente

Resumo das variáveis

Variável	Tipo	Mínimo	Média	Máximo
jitter	contínua	0.03	1.961	31.58
shimmer	contínua	2.7	13.821	62.42
GNE	contínua	0.15	0.5905	0.96
Irregularity	contínua	0.5	1.34	2.79
Noise	contínua	0.16	0.9224	2.66
OverallSeverity	contínua	0.47	1.206	2.6

E a variável Irregularity segue a seguinte densidade



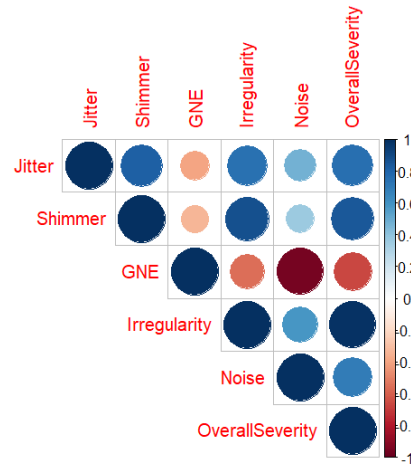
3.1 Teste de Normalidade

O teste de Shapiro-Wilk é um teste estatístico utilizado para avaliar a normalidade de uma amostra de dados. Ele verifica se a distribuição dos dados se desvia de uma distribuição normal.

- **Hipótese nula (H_0):** Os dados seguem uma distribuição normal.
- **Hipótese alternativa (H_1):** Os dados não seguem uma distribuição normal.

Aplicando o teste à Variável Irregularity, temos um p-valor de $1.71e-07$ que não nos dá uma probabilidade suficiente para aceitar a H_0 , então aceitamos H_1 e seguimos com a análise.

3.2 Correlações entre as variáveis



O gráfico acima mostra as correlações entre as variáveis que serão utilizadas no modelo, e observa-se correlações positivamente e negativamente fortes entre as variáveis.

4 Resultados

Ao ser ajustado o modelo linear generalizado com a distribuição Gama e função de ligação Log, apenas 3 variáveis se mostraram ser estatisticamente significativas para o modelo.

Modelo ajustado com as variáveis significativas

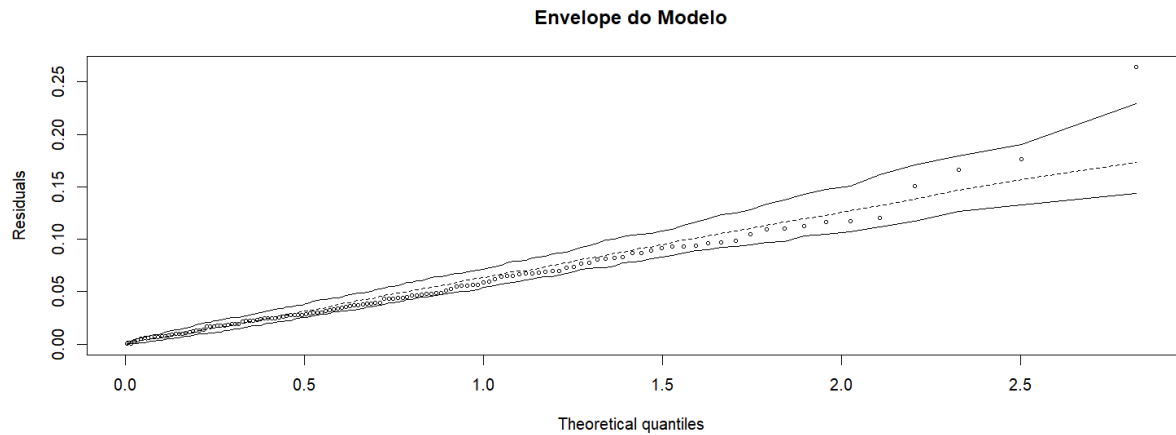
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.79376	0.020838	-38.098	< 2e-16 ***
Jitter	-0.016900	0.001789	-9.446	2.30e-16 ***
Noise	-0.118851	0.012839	-9.257	6.62e-16 ***
OverallSeverity	0.968587	0.022800	42.481	< 2e-16 ***

Com um AIC de -276.76 e Deviance de 0.54245 com 127 Graus de Liberdade.

Valores dos fatores de inflação

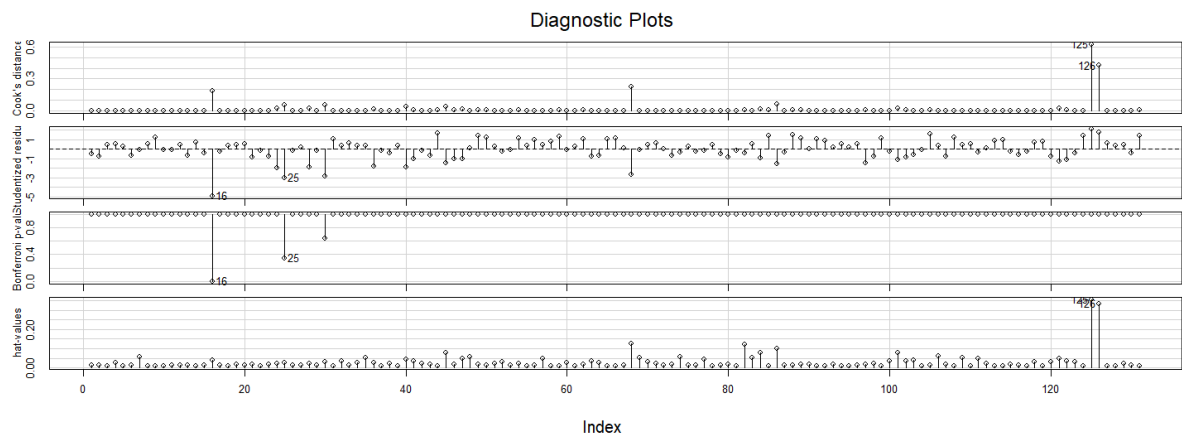
	Jitter	Noise	OverallSeverity
VIF	2.354515	1.976416	3.573493

Como nenhuma das variáveis tem um VIF maior que 10, não há evidência de multicolinearidade.

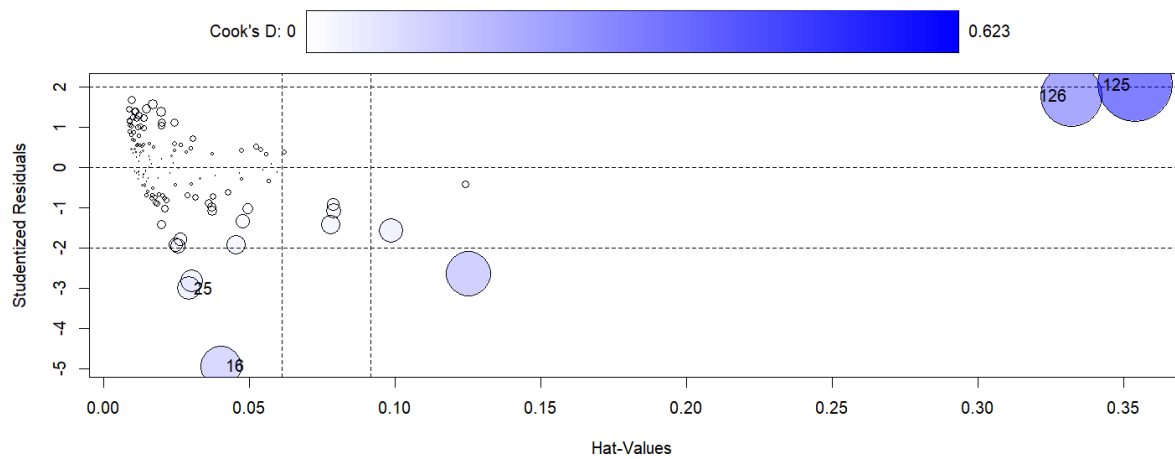


O gráfico de envelope nos dá a ideia de um bom ajuste no modelo, não contendo apenas 1 valor que representa 0,76% dos valores que é inferior ao limite de 5%.

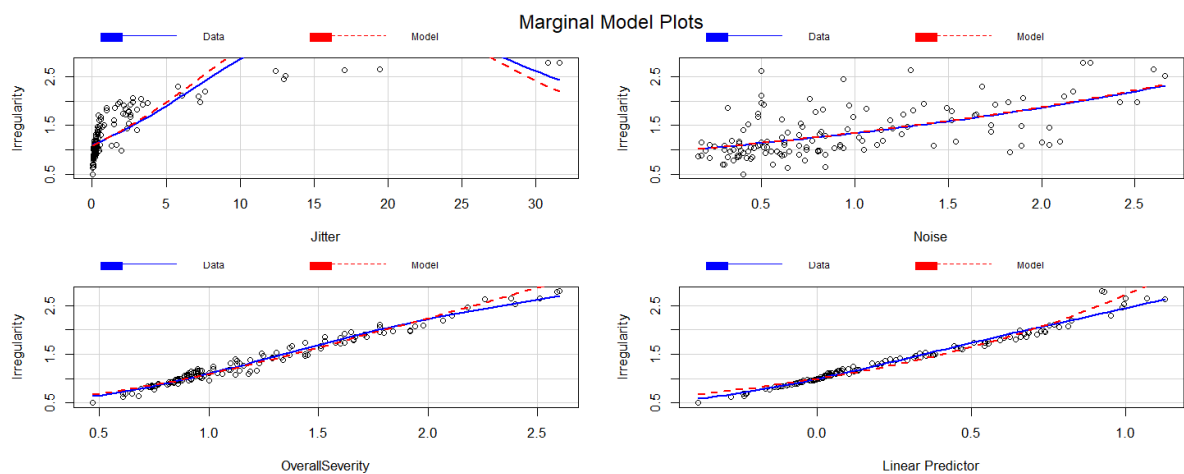
4.1 Diagnóstico do modelo



Nos gráficos de diagnóstico observa-se que o ponto 16 é um possível outlier, e os pontos 125 e 126 são os com maiores valores de distância de Cook e hat-values, ou seja, apresenta os maiores valores nas variáveis independentes.



Observando o gráfico de influência acima vemos que como já esperado, os pontos 125 e 126 são os que mais influenciam no modelo.



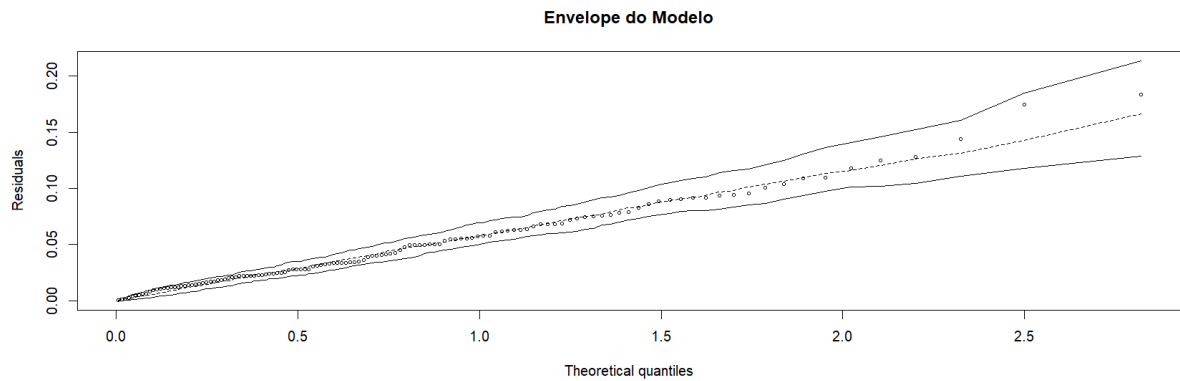
Os gráficos acima destacam a relação entre a variável dependente e cada uma das independentes assim como o preditor linear em que podemos observar uma boa relação entre os dados reais e as estimativas do modelo.

4.2 Modelo sem a observação 16

observando o modelo acima foi constatado que o ponto 16 é um possível outlier e foi criado um novo modelo retirando o ponto.

Modelo ajustado com as variáveis significativas

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.775630	0.019726	-39.321	<2e-16
Jitter	-0.016038	0.001671	-9.597	<2e-16
Noise	-0.116283	0.011929	-9.748	<2e-16
OverallSeverity	0.951828	0.021458	44.359	<2e-16



Com um AIC de -294.76 e Deviance de 0.45442 com 126 Graus de Liberdade.

O novo modelo se mostrou melhor, com AIC e Deviance menores, assim como menores devios nos parâmetros e seu envelope conseguiu conter 100% dos pontos.

4.3 interpretação dos resultados

Para podermos interpretar os parâmetros do modelo, é necessário retorna-los para sua escala de origem, ou seja, aplicar a função inversa da função de ligação.

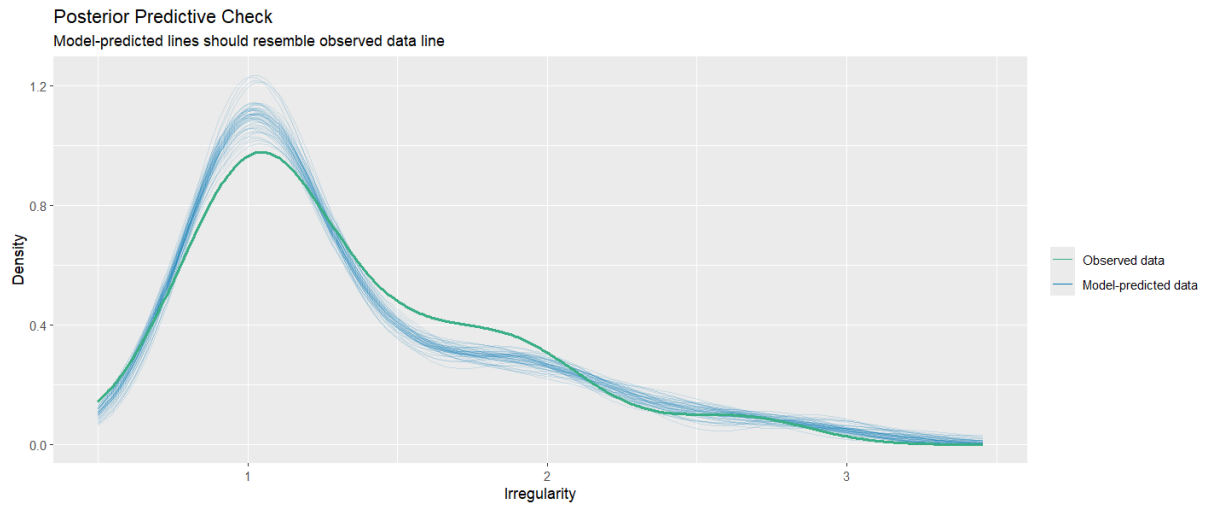
Valores dos parâmetros

	(Intercept)	Jitter	Noise	OverallSeverity
$exp(\beta)$	0,4604	0,984	0,89	2,5904

Neste sentido, a cada unidade de nervosismo do paciente(jitter) sua irregularidade na voz aumenta cerca de 0.984, para o barulho(Noise) aumenta 0.89 e para cada unidade de gravidade na voz, sua irregularidade vocal aumenta 2.59. Caso todas as variáveis fossem 0 sua irregularidade vocal seria de 0.4604(intercept). Todas as análises e gráficos foram elaborados utilizando o Software R. (R Core Team 2023)

5 Conclusão

De modo Geral, o modelo mostrou uma boa adequação aos dados, modelando de forma aceitável a variável resposta(Irregularity) e apresentando bons resultados no diagnóstico.



O gráfico acima mostra 100 simulações (em azul) utilizando o modelo criado e a densidade dos dados(em verde) para uma melhor visualização do comportamento preditivo e adequação do modelo.

REFERÊNCIAS

. [S.l.]. Disponível em: <<https://www.kaggle.com/daniilkrasnoproshin/healthy-vs-laryngeal-disorder-classification/data>>.

CORDEIRO, G. M.; DEMÉTRIO, C. G. Modelos lineares generalizados e extensões. *Piracicaba: USP*, p. 31, 2008.

R Core Team. *The R Project for Statistical Computing*. [S.l.], 2023. Disponível em: <<https://cran.r-project.org/manuals.html>>.

TEIXEIRA, J. P.; FERREIRA, D.; CARNEIRO, S. M. Análise acústica vocal-determinação do jitter e shimmer para diagnóstico de patologias da fala. In: INEGI. *6º Congresso Luso-Moçambicano de Engenharia, 3º Congresso de Engenharia de Moçambique*. [S.l.], 2011.

1 Anexos

```
> #carregar os dados
> dados = read.csv('C:/Users/Amauri/Downloads/data/healthy_vs_laryngeal_disorder_cl
> #teste de normalidade
> shapiro.test(dados$Irregularity)
```

Shapiro-Wilk normality test

```
data: dados$Irregularity
W = 0.90707, p-value = 1.71e-07
#carregar pacotes
pacman::p_load('glm','MASS','hnp','car','ggplot2','tidyverse')

> #criação do modelo
> modelo1 = glm(data = dados[,],formula = Irregularity ~ . -idx -SD_F0 -Max_F0 -mea
> summary(modelo1)
```

Call:

```
glm(formula = Irregularity ~ . - idx - SD_F0 - Max_F0 - mean_F0 -
      Min_F0 - GNE - Diagnosis - Shimmer, family = Gamma(link = "log"),
      data = dados[, ])
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.793876	0.020838	-38.098	< 2e-16 ***
Jitter	-0.016900	0.001789	-9.446	2.30e-16 ***
Noise	-0.118851	0.012839	-9.257	6.62e-16 ***
OverallSeverity	0.968587	0.022800	42.481	< 2e-16 ***

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

(Dispersion parameter for Gamma family taken to be 0.004081012)

```
Null deviance: 16.54175 on 130 degrees of freedom
Residual deviance: 0.54245 on 127 degrees of freedom
AIC: -276.76
```

Number of Fisher Scoring iterations: 4

```

> #correlações
> r = cor(dados[2:7])
> #Grafico de correlação
> corrplot::corrplot(r,type = 'upper')
> #envelope
> hnp(modelo1$residuals,sim = 99,resid.type = 'deviance',how.many.out = T,conf = 0.
Half-normal plot with simulated envelope generated assuming the residuals are
normally distributed under the null hypothesis.
Estimated mean: -5.533661e-12
Estimated variance: 0.003986835
Total points: 131
Points out of envelope: 1 ( 0.76 %)
#graficos de diagnóstico
> influenceIndexPlot(modelo1)
> marginalModelPlots(modelo1)
> influencePlot(modelo1)
      StudRes      Hat      CookD
16  -4.940662 0.04015159 0.18600256
25  -3.007302 0.02918068 0.05908444
125  2.063642 0.35390718 0.62273610
126  1.787131 0.33205257 0.42567522
> #Grafico de predições
> library(performance)
> check_model(modelo1, theme = "theme_minimal")[]
$VIF
# Check for Multicollinearity

$QQ
Simulated residuals from a model of class 'glm' based on 250 simulations. Use
'check_residuals()' to check uniformity of residuals or 'residuals()' to
extract simulated residuals. It is recommended to refer to
'?DHARMA::simulateResiduals' and 'vignette("DHARMA")' for more information
about different settings in particular situations or for particular models.

$HOMOGENEITY
      x      y
1  0.9106835 0.72932834
2  0.8936725 0.87832576

```

3	1.1453879	0.68959549
4	1.1282761	0.76592641
5	1.0231874	0.50863749
6	0.9096017	0.82903998
7	1.1079857	0.34110727
8	1.0227745	0.75698532
9	1.0708112	1.14468999
10	0.9765903	0.32601940
11	1.0058619	0.30314975
12	1.0798173	0.66373037
13	0.9203486	0.83145434
14	1.0419122	0.85212144
15	0.9356621	0.65754949
16	0.6793034	2.05361823
17	0.9455614	0.50933759
18	1.0261078	0.60530764
19	1.0363963	0.71549647
20	1.0309615	0.77257019
21	0.8689637	0.94352347
22	0.9815704	0.43088243
23	0.8829387	0.87713855
24	0.7957304	1.38116627
25	0.7893125	1.67453428
26	0.9814175	0.42823673
27	0.9676873	0.44809226
28	0.7929191	1.36291858
29	0.9906043	0.41080807
30	0.7558233	1.62664606
31	1.2911564	1.04042221
32	1.4576415	0.59513409
33	1.0399316	0.77968789
34	1.4646070	0.61947748
35	1.1252414	0.59532792
36	0.7864594	1.32060103
37	0.9996080	0.38902288
38	1.7178202	0.66427945
39	1.0239818	0.63268253
40	2.5912743	1.36482591
41	2.0362787	0.99457290

42	1.8074522	0.39102866
43	0.8883718	0.82620331
44	1.1239523	1.32820696
45	2.6828890	1.18950316
46	0.8446450	1.01167504
47	2.1904575	1.01461740
48	0.9545597	0.30315512
49	1.3924339	1.20350740
50	1.0815705	1.13521533
51	1.7371685	0.54711614
52	1.7405247	0.52835174
53	1.0171172	0.33217308
54	1.5562692	1.07618848
55	1.0442427	0.62312643
56	1.0611789	1.01103878
57	1.4209621	0.66376847
58	1.3654719	0.92650835
59	1.4174288	1.16699007
60	1.8334094	0.17174028
61	1.0110201	0.54381090
62	1.0749583	1.05056184
63	2.0654937	0.85891800
64	1.9337997	0.82961661
65	1.3761441	1.03653153
66	1.1050197	1.09966545
67	1.7208143	0.29028634
68	3.0837302	1.58649856
69	1.8673308	0.25140932
70	1.3377397	0.70858864
71	1.5690159	0.78484003
72	1.0004115	0.08056920
73	0.9314823	0.83921243
74	2.1555130	0.58332419
75	1.4540710	0.53045372
76	0.9869085	0.51988833
77	1.9774900	0.37656410
78	1.2234764	0.68520900
79	0.9278887	0.68837412
80	0.8795093	0.94294544

81	0.9777602	0.35354040
82	2.7175207	0.64462369
83	1.0449863	0.73400021
84	2.0866336	0.95474612
85	1.2441820	1.21039300
86	2.9120913	1.24118576
87	0.9513444	0.59486282
88	1.0836068	1.24435012
89	1.0340730	1.07756976
90	1.6005984	0.07687305
91	1.2338711	1.04771733
92	1.2174227	0.96817546
93	1.5891393	0.45562410
94	1.0219218	0.76598906
95	1.0197068	0.39874245
96	1.2752926	0.74326015
97	0.8560897	1.18546377
98	0.9064987	0.90097717
99	1.3083064	1.10553147
100	2.1183052	0.46185366
101	2.7012098	1.04603376
102	1.9701398	0.94412447
103	2.2797650	0.79369903
104	1.0072724	0.33711010
105	1.2674142	1.28516823
106	1.1422192	0.62700310
107	0.8929098	0.87121768
108	1.0640942	1.12738261
109	1.9244662	0.68150395
110	1.1392014	0.75110114
111	1.9568884	0.54984905
112	1.7184834	0.32586965
113	1.0481396	0.96365348
114	1.0405148	1.02547806
115	0.9975436	0.52722020
116	0.9367964	0.78715649
117	0.9881340	0.53761094
118	1.6622644	0.86229908
119	1.0357288	0.90839821

```

120 1.9007102 0.87126003
121 2.2547923 1.14657349
122 2.0913433 1.03841339
123 1.9823801 0.64808185
124 1.1951044 1.22951966
125 2.5146451 1.46030104
126 2.5350495 1.36040572
127 1.0630506 0.83367794
128 1.3952222 0.62647069
129 1.0611202 0.65674355
130 0.9471690 0.67233784
131 1.0697225 1.21468740

```

\$OUTLIERS

OK: No outliers detected.

- Based on the following method and threshold: cook (0.945).
- For variable: (Whole model)

\$INFLUENTIAL

	Hat	Cooks_Distance	Predicted	Residuals	Std_Residuals	Index	Influent
1	0.016864594	1.213375e-03	0.9106835	-0.0340787930	-0.538014022	1	
2	0.016738221	2.532810e-03	0.8936725	-0.0496881248	-0.784393681	2	
3	0.010523445	6.012708e-04	1.1453879	0.0299195480	0.470834560	3	
4	0.026303473	2.324222e-03	1.1282761	0.0365339907	0.579563947	4	
5	0.011545511	1.954474e-04	1.0231874	0.0163424223	0.257308508	5	
6	0.016767766	2.014007e-03	0.9096017	-0.0441858310	-0.697543074	6	
7	0.059792215	2.152409e-04	1.1079857	-0.0072247753	-0.116634890	7	
8	0.011436144	9.496526e-04	1.0227745	0.0359642070	0.566218667	8	
9	0.009909191	4.295906e-03	1.0708112	0.0810846165	1.275607342	9	
10	0.012265983	3.507326e-05	0.9765903	-0.0067634845	-0.106528681	10	
11	0.014618499	3.132328e-05	1.0058619	-0.0058391050	-0.092078904	11	
12	0.013537375	6.658258e-04	1.0798173	0.0276953963	0.436499072	12	
13	0.014559530	1.765268e-03	0.9203486	-0.0444981107	-0.701685389	13	
14	0.010004706	1.332040e-03	1.0419122	0.0454618572	0.715231554	14	
15	0.014026685	6.648806e-04	0.9356621	-0.0276815216	-0.436388640	15	
16	0.040151586	1.860026e-01	0.6793034	-0.2915746458	-4.658697257	16	
17	0.013890159	2.369981e-04	0.9455614	-0.0165484372	-0.260861747	17	
18	0.010409420	3.530333e-04	1.0261078	0.0231060402	0.363591570	18	

19	0.017068118	1.137712e-03	1.0363963	0.0320796094	0.506504629	19
20	0.013760549	1.242640e-03	1.0309615	0.0373984778	0.589493198	20
21	0.018329616	3.699462e-03	0.8689637	-0.0574417274	-0.907529404	21
22	0.012236704	1.067546e-04	0.9815704	-0.0118343069	-0.186394221	22
23	0.020971404	3.169895e-03	0.8829387	-0.0494430321	-0.782210183	23
24	0.025422328	2.373136e-02	0.7957304	-0.1254993105	-1.989982096	24
25	0.029180679	5.908444e-02	0.7893125	-0.1881033042	-2.988431303	25
26	0.013875356	1.183009e-04	0.9814175	-0.0116791180	-0.184102717	26
27	0.015994084	1.638217e-04	0.9676873	0.0126702565	0.199941346	27
28	0.024775429	2.191473e-02	0.7929191	-0.1221045348	-1.935510509	28
29	0.014090262	1.017602e-04	0.9906043	-0.0107432851	-0.169369257	29
30	0.030067718	5.425894e-02	0.7558233	-0.1767255045	-2.808953703	30
31	0.009879556	2.922994e-03	1.2911564	0.0672915351	1.058601439	31
32	0.037392298	1.218238e-03	1.4576415	0.0220370739	0.351597128	32
33	0.015663285	1.470152e-03	1.0399316	0.0380458368	0.600276520	33
34	0.028321900	1.073105e-03	1.4646070	0.0239735637	0.380704007	34
35	0.055576155	1.847934e-03	1.1252414	0.0218435639	0.351848807	35
36	0.026317565	2.055204e-02	0.7864594	-0.1142434404	-1.812335776	36
37	0.011582422	6.709644e-05	0.9996080	-0.0096427633	-0.151826414	37
38	0.024793002	1.237587e-03	1.7178202	-0.0281003316	-0.445429625	38
39	0.012678118	5.143745e-04	1.0239818	0.0251967812	0.396946285	39
40	0.045402256	4.125780e-02	2.5912743	-0.1211034534	-1.940270961	40
41	0.036899972	9.372191e-03	2.0362787	-0.0633448514	-1.010396149	41
42	0.022844816	1.366465e-04	1.8074522	-0.0096869340	-0.153398329	42
43	0.018888802	2.242714e-03	0.8883718	-0.0438313975	-0.692695329	43
44	0.009754490	7.664162e-03	1.1239523	0.1082090427	1.702190538	44
45	0.077721302	4.217745e-02	2.6828890	-0.0894524225	-1.458065639	45
46	0.020914085	5.593997e-03	0.8446450	-0.0661461767	-1.046430534	46
47	0.049301331	1.373931e-02	2.1904575	-0.0655466646	-1.052313670	47
48	0.057636355	1.291450e-04	0.9545597	0.0056885282	0.091729011	48
49	0.019701685	1.054096e-02	1.3924339	0.0889566527	1.406421065	49
50	0.013712075	5.772348e-03	1.0815705	0.0796326036	1.255177413	50
51	0.023187072	5.317318e-04	1.7371685	0.0187816626	0.297470835	51
52	0.032876870	6.622794e-04	1.7405247	-0.0176412229	-0.280804369	52
53	0.014501813	4.478826e-05	1.0171172	-0.0070138534	-0.110597377	53
54	0.024413553	8.391867e-03	1.5562692	0.0713711998	1.131113476	54
55	0.011165300	4.255900e-04	1.0442427	0.0244660319	0.385139220	55
56	0.013658573	3.617340e-03	1.0611789	0.0635022376	1.000901752	56
57	0.047259789	2.407264e-03	1.4209621	0.0272252645	0.436617278	57

58	0.009375026	1.743414e-03	1.3654719	0.0536179221	0.843279225	58
59	0.012043559	5.652306e-03	1.4174288	0.0841003590	1.324478794	59
60	0.025943563	5.792609e-06	1.8334094	-0.0018607574	-0.029513027	60
61	0.012566698	2.782563e-04	1.0110201	0.0186567935	0.293899723	61
62	0.019679031	6.113106e-03	1.0749583	0.0682476551	1.078995530	62
63	0.037663405	5.325243e-03	2.0654937	-0.0469652176	-0.749426243	63
64	0.028630730	3.490582e-03	1.9337997	-0.0439764888	-0.698464647	64
65	0.012590896	3.679846e-03	1.3761441	0.0667104645	1.050900162	65
66	0.008971220	3.309387e-03	1.1050197	0.0750164456	1.179585451	66
67	0.016705644	3.015964e-05	1.7208143	0.0053285358	0.084116689	67
68	0.125315441	2.269091e-01	3.0837302	-0.1586570157	-2.655520199	68
69	0.054701158	5.779535e-05	1.8673308	-0.0039309770	-0.063289549	69
70	0.029988766	1.948492e-03	1.3377397	0.0312641438	0.496905972	70
71	0.024366824	2.369065e-03	1.5690159	0.0383753103	0.608169559	71
72	0.015912816	1.703451e-07	1.0004115	-0.0004114323	-0.006492287	72
73	0.020239214	2.561541e-03	0.9314823	-0.0452124026	-0.715012472	73
74	0.056457608	1.731973e-03	2.1555130	-0.0212651535	-0.342691925	74
75	0.015890915	3.196213e-04	1.4540710	0.0177271159	0.279725798	75
76	0.015427313	2.861691e-04	0.9869085	-0.0172316190	-0.271843059	76
77	0.046707698	2.462964e-04	1.9774900	-0.0088707417	-0.142220745	77
78	0.009408198	5.234123e-04	1.2234764	0.0295602894	0.464919085	78
79	0.014174364	8.071266e-04	0.9278887	-0.0303626468	-0.478691407	79
80	0.017846485	3.591365e-03	0.8795093	-0.0573843767	-0.906400298	80
81	0.011998560	4.743160e-05	0.9777602	-0.0079578106	-0.125323025	81
82	0.123937635	6.107075e-03	2.7175207	-0.0250552460	-0.419032027	82
83	0.052231403	3.999025e-03	1.0449863	0.0331392995	0.532853993	83
84	0.078627212	1.772675e-02	2.0866336	-0.0569723498	-0.929100083	84
85	0.010739921	5.825540e-03	1.2441820	0.0903465599	1.421911074	85
86	0.098627644	6.492046e-02	2.9120913	-0.0965147427	-1.591320280	86
87	0.014953283	4.752116e-04	0.9513444	-0.0226060521	-0.356543372	87
88	0.014751930	8.974570e-03	1.0836068	0.0951432389	1.500448845	88
89	0.020192388	6.946539e-03	1.0340730	0.0717013630	1.133895586	89
90	0.019108971	1.700799e-07	1.6005984	-0.0003739355	-0.005910203	90
91	0.009137245	2.777911e-03	1.2338711	0.0682426790	1.073162215	91
92	0.008860390	1.963696e-03	1.2174227	0.0584704354	0.919358656	92
93	0.020200586	2.221230e-04	1.5891393	0.0130700093	0.206691857	93
94	0.011765419	1.024655e-03	1.0219218	0.0368083962	0.579606076	94
95	0.012334927	7.892903e-05	1.0197068	0.0100604764	0.158463687	95
96	0.013246579	1.024232e-03	1.2752926	0.0346551368	0.546109028	96

97	0.019851743	1.000000e-02	0.8560897	-0.0916593040	-1.449261406	97
98	0.021583900	3.634139e-03	0.9064987	-0.0521990659	-0.826070274	98
99	0.008902815	3.354560e-03	1.3083064	0.0758019274	1.191895514	99
100	0.038457244	4.549536e-04	2.1183052	-0.0134221552	-0.214266371	100
101	0.078923169	2.564670e-02	2.7012098	-0.0686463747	-1.119658897	101
102	0.036149392	7.449863e-03	1.9701398	-0.0569816960	-0.908545260	102
103	0.042711668	4.426568e-03	2.2797650	-0.0399036718	-0.638421545	103
104	0.010985962	3.586433e-05	1.0072724	-0.0072373078	-0.113917874	104
105	0.017010650	1.180190e-02	1.2674142	0.1011709944	1.597341115	105
106	0.062125950	2.559454e-03	1.1422192	0.0241273605	0.389990126	106
107	0.017745207	2.601978e-03	0.8929098	-0.0488483218	-0.771531439	107
108	0.011375405	4.646873e-03	1.0640942	0.0786557489	1.238314115	108
109	0.054083111	3.083346e-03	1.9244662	0.0285837453	0.460053883	109
110	0.012521842	1.008961e-03	1.1392014	0.0353944995	0.557555475	110
111	0.047409243	1.137288e-03	1.9568884	-0.0189703201	-0.304254868	111
112	0.024075573	6.954664e-05	1.7184834	0.0066867088	0.105954736	112
113	0.010225350	2.227232e-03	1.0481396	0.0578965900	0.910963298	113
114	0.011869601	3.320988e-03	1.0405148	0.0653484436	1.029068244	114
115	0.019065750	3.754241e-04	0.9975436	-0.0176910162	-0.279607547	115
116	0.015293038	1.490634e-03	0.9367964	-0.0398053191	-0.627919056	116
117	0.012094438	2.556717e-04	0.9881340	-0.0184652452	-0.290812736	117
118	0.030740653	4.383740e-03	1.6622644	0.0460551529	0.732275135	118
119	0.011962654	2.061101e-03	1.0357288	0.0515107948	0.811199416	119
120	0.031452029	4.677984e-03	1.9007102	-0.0485054255	-0.771517548	120
121	0.047688573	2.163630e-02	2.2547923	-0.0843077686	-1.352365725	121
122	0.037388448	1.129035e-02	2.0913433	-0.0691705316	-1.103599860	122
123	0.030225423	1.374552e-03	1.9823801	-0.0266591889	-0.423767448	123
124	0.008973229	5.173034e-03	1.1951044	0.0932196946	1.465821342	124
125	0.353907180	6.227361e-01	2.5146451	0.1057410805	2.059265503	125
126	0.332052573	4.256752e-01	2.5350495	0.0936778203	1.794244792	126
127	0.010555180	1.288271e-03	1.0630506	0.0435308839	0.685042876	127
128	0.011567902	4.506619e-04	1.3952222	0.0247222135	0.389251227	128
129	0.024318520	1.159184e-03	1.0611202	0.0269732700	0.427460143	129
130	0.013321997	6.897380e-04	0.9471690	-0.0289634157	-0.456434131	130
131	0.010902733	5.999216e-03	1.0697225	0.0909628920	1.431728983	131

\$PP_CHECK

Warning: Minimum value of original data is not included in the replicated data.

Model may not capture the variation of the data.

```
attr(,"panel")
[1] TRUE
attr(,"dot_size")
[1] 2
attr(,"line_size")
[1] 0.8
attr(,"base_size")
[1] 10
attr(,"axis_title_size")
[1] 10
attr(,"title_size")
[1] 12
attr(,"check")
[1] "all"
attr(,"alpha")
[1] 0.2
attr(,"dot_alpha")
[1] 0.8
attr(,"show_dots")
[1] TRUE
attr(,"detrend")
[1] FALSE
attr(,"colors")
[1] "#3aaf85" "#1b6ca8" "#cd201f"
attr(,"theme")
[1] "theme_minimal"
attr(,"model_info")
attr(,"model_info")$is_binomial
[1] FALSE

attr(,"model_info")$is_bernoulli
[1] FALSE

attr(,"model_info")$is_count
[1] FALSE

attr(,"model_info")$is_poisson
[1] FALSE
```

```
attr(,"model_info")$is_negbin  
[1] FALSE
```

```
attr(,"model_info")$is_beta  
[1] FALSE
```

```
attr(,"model_info")$is_betabinomial  
[1] FALSE
```

```
attr(,"model_info")$is_orderedbeta  
[1] FALSE
```

```
attr(,"model_info")$is_dirichlet  
[1] FALSE
```

```
attr(,"model_info")$is_exponential  
[1] TRUE
```

```
attr(,"model_info")$is_logit  
[1] FALSE
```

```
attr(,"model_info")$is_probit  
[1] FALSE
```

```
attr(,"model_info")$is_censored  
[1] FALSE
```

```
attr(,"model_info")$is_truncated  
[1] FALSE
```

```
attr(,"model_info")$is_survival  
[1] FALSE
```

```
attr(,"model_info")$is_linear  
[1] FALSE
```

```
attr(,"model_info")$is_tweedie  
[1] FALSE
```

```
attr(,"model_info")$is_zeroinf  
[1] FALSE
```

```
attr(,"model_info")$is_zero_inflated  
[1] FALSE
```

```
attr(,"model_info")$is_dispersion  
[1] FALSE
```

```
attr(,"model_info")$is_hurdle  
[1] FALSE
```

```
attr(,"model_info")$is_ordinal  
[1] FALSE
```

```
attr(,"model_info")$is_cumulative  
[1] FALSE
```

```
attr(,"model_info")$is_multinomial  
[1] FALSE
```

```
attr(,"model_info")$is_categorical  
[1] FALSE
```

```
attr(,"model_info")$is_mixed  
[1] FALSE
```

```
attr(,"model_info")$is_multivariate  
[1] FALSE
```

```
attr(,"model_info")$is_trial  
[1] FALSE
```

```
attr(,"model_info")$is_bayesian  
[1] FALSE
```

```
attr(,"model_info")$is_gam  
[1] FALSE
```



```
attr(,"model_info")$is_anova
```

```
[1] FALSE
```

```
attr(,"model_info")$is_timeseries
```

```
[1] FALSE
```

```
attr(,"model_info")$is_ttest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_correlation
```

```
[1] FALSE
```

```
attr(,"model_info")$is_onewaytest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_chi2test
```

```
[1] FALSE
```

```
attr(,"model_info")$is_ranktest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_levenetest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_variancetest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_xtab
```

```
[1] FALSE
```

```
attr(,"model_info")$is_proptest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_binomtest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_ftest
```

```
[1] FALSE
```

```
attr(,"model_info")$is_meta
```

```
[1] FALSE
```

```
attr(,"model_info")$link_function
```

```
[1] "log"
```

```
attr(,"model_info")$family
```

```
[1] "Gamma"
```

```
attr(,"model_info")$n_obs
```

```
[1] 131
```

```
attr(,"model_info")$n_grouplevels
```

```
NULL
```

```
attr(,"bandwidth")
```

```
[1] "nrd"
```

```
attr(,"type")
```

```
[1] "density"
```

```
attr(,"model_class")
```

```
[1] "glm"
```

```
# Verificar multicolinearidade usando VIF
```

```
> valorVIF <- car::vif(modelo1)
```

```
> print(valorVIF)
```

```
      Jitter      Noise OverallSeverity
      2.354515      1.976416      3.573493
```

```
> #Gráfico de densidade
```

```
> ggplot(data = dados, aes(x = Irregularity)) +
+   geom_density(fill = "blue", alpha = 0.5) +
+   labs(title = "Gráfico de Densidade da irregularidade da voz",
+         x = "Irregularidade na voz",
+         y = "Densidade")+
+   theme_minimal()
```

```
> #resumo dos dados
```

```
> glimpse(dados[2:7])
```

```
Rows: 131
```

```
Columns: 6
```

```
$ Jitter      <dbl> 0.09, 0.07, 0.15, 0.39, 0.43, 0.09, 0.24, 0.12, 0.21, 0.21,
```

```

$ Shimmer      <dbl> 7.41, 8.35, 11.58, 7.70, 9.69, 9.14, 8.69, 10.20, 9.22, 8.5
$ GNE          <dbl> 0.83, 0.78, 0.44, 0.35, 0.51, 0.83, 0.26, 0.73, 0.68, 0.72,
$ Irregularity <dbl> 0.88, 0.85, 1.18, 1.17, 1.04, 0.87, 1.10, 1.06, 1.16, 0.97,
$ Noise        <dbl> 0.37, 0.45, 1.04, 1.54, 0.89, 0.38, 2.04, 0.53, 0.62, 0.58,
$ OverallSeverity <dbl> 0.77, 0.76, 1.09, 1.14, 0.96, 0.77, 1.18, 0.91, 0.97, 0.87,
> summary(dados[2:7])

```

Jitter	Shimmer	GNE	Irregularity	Noise
Min. : 0.030	Min. : 2.700	Min. :0.1500	Min. :0.50	Min. :0.1600
1st Qu.: 0.160	1st Qu.: 8.265	1st Qu.:0.4050	1st Qu.:0.98	1st Qu.:0.4550
Median : 0.310	Median :11.060	Median :0.6100	Median :1.16	Median :0.7300
Mean : 1.961	Mean :13.821	Mean :0.5905	Mean :1.34	Mean :0.9224
3rd Qu.: 1.700	3rd Qu.:16.320	3rd Qu.:0.7800	3rd Qu.:1.67	3rd Qu.:1.2550
Max. :31.580	Max. :62.420	Max. :0.9600	Max. :2.79	Max. :2.6600

OverallSeverity

Min. :0.470

1st Qu.:0.875

Median :1.020

Mean :1.206

3rd Qu.:1.485

Max. :2.600

```
> #valores interpretaveis
```

```
> exp(modelo1$coefficients)
```

(Intercept)	Jitter	Noise	OverallSeverity
0.4520893	0.9832416	0.8879397	2.6342190