



REGRESSÃO LOGÍSTICA EM AÇÕES JUDICIAIS

Consultores Responsáveis:

Alfredo Rossi

Ana Gabriela Vasconcelos

Gabriel Ravi

Lucas Gongora

Renata Dias

Requerente

Débora Costa Ferreira

Brasília, Agosto de 2017

Conteúdo

1	Introdução	3
2	Metodologia	4
2.1	Média	4
2.2	Mediana	4
2.3	Desvio Padrão Amostral	4
2.4	Quartil	5
2.5	Coeficiente de Variação	5
2.6	Boxplot	5
2.7	Teste de Hipóteses	6
2.7.1	Estatística do Teste	6
2.7.2	Nível de Significância	7
2.7.3	P-valor	7
2.8	Teste de Qui-Quadrado de Independência	7
2.9	Teste Exato de Fisher	8
2.10	Teste de Wilcoxon	9
2.11	Teste de Kruskall Wallis	10
2.12	Regressão Logística	11
2.13	Teste de Wald	11
2.14	Teste da Razão de Verossimilhança	12
2.15	Critério de Informação	12
2.16	Teste de Hosmer e Lemeshow	13
2.17	Deviance	13
2.18	Dfbeta	14
2.19	Leverage	14
2.20	Curva ROC	15
2.21	Métodos Automáticos de Seleção de Variáveis	15
3	Frequência de Amicus Curiae por relator	17
4	Análise entre número de AC favoráveis e Relator	18

5	Procedimentos para a Regressão Logística	19
5.1	Multicolinearidade	19
5.2	Modelo para Cita no Voto	20
5.3	Modelo para Resultado	27
5.3.1	Modelo sem interação	32
5.3.2	Modelo com interação	35
6	Conclusão	37

1 Introdução

O objetivo do relatório é criar modelos de regressão voltados para a área jurídica, de modo que verifique a relação do envolvimento de Amicus Curiae na resultado da ação judicial. Além disso, será verificado a relação de Amicus Curiae por relator para os processos apresentados.

O banco de dados foi coletado pelo cliente, e contém as duas variáveis respostas de resultado do processo e de citação no voto. Como o objetivo é verificar a relação com Amicus Curiae, foram criadas algumas variáveis para expor diversos perfis. Os resultados obtidos foram feitos nos softwares R e SAS.

2 Metodologia

2.1 Média

A média é a soma das observações dividida pelo tamanho da amostra:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

sendo X_i o resultado da i -ésima observação e n o tamanho da amostra.

2.2 Mediana

A mediana é definida como a medida que separa os dados em duas partes iguais. Em outras palavras, é a realização que ocupa a posição central da série de observações quando estão ordenadas em ordem crescente. Assim, das n observações da amostra, metade será inferior ao valor da mediana e a outra metade será superior à mediana. Sua forma de cálculo é condicionada ao fato do tamanho da amostra (n) ser par ou ímpar e é dada por:

$$\begin{cases} X_{(\frac{n+1}{2})}, & \text{se } n \text{ ímpar} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ par} \end{cases}$$

2.3 Desvio Padrão Amostral

O desvio padrão amostral é uma medida que indica o quanto os dados variam em relação à média e é dado por:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

sendo X_i o resultado da i -ésima observação, \bar{X} a média da amostra e n o tamanho da amostra.

2.4 Quartil

Os quartis são pontos que dividem a amostra em quatro partes iguais, ou seja, $(\frac{i \cdot 100}{4})\%$ dos dados possuem valores menores que o i -ésimo quartil. Dessa forma, o segundo quartil, também chamado mediana, divide os dados ordenados em duas partes iguais: os 50% menores valores e os 50% maiores valores.

Para $i = 1, 2, 3$, o i -ésimo quartil será dado por:

$$Q_i = X_{(\frac{i \cdot (n+1)}{4})}$$

sendo n o tamanho da amostra e $X_{(j)}$ a j -ésima observação da variável ordenada.

Se $\frac{i \cdot (n+1)}{4}$ não for um valor inteiro, o i -ésimo quartil será dado pela média das observações contidas nas posições imediatamente antes e imediatamente depois do valor obtido.

2.5 Coeficiente de Variação

O coeficiente de variação fornece a variação dos dados em relação à média, eliminando a magnitude deles e permitindo a comparação entre dados que possuem medições diferentes. Quanto menor for o seu valor, mais homogêneos são os dados, e quanto maior, mais heterogêneos. A fórmula para cálculo é dada por:

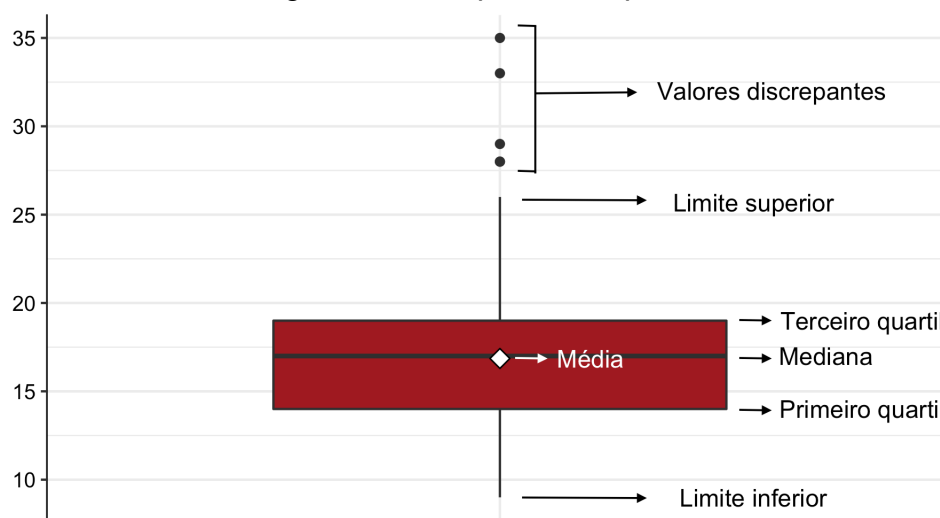
$$CV = \frac{S}{\bar{X}}$$

sendo S e \bar{X} o desvio padrão e a média amostral, respectivamente.

2.6 Boxplot

O boxplot é um tipo de gráfico no qual se pode perceber de forma mais clara como os dados estão distribuídos. A figura abaixo ilustra um exemplo de boxplot.

Figura 1: Exemplo de boxplot



A porção inferior do retângulo diz respeito ao segundo quartil, enquanto a superior do retângulo indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada pelo losango branco e os pontos são *outliers*. Os *outliers* são valores discrepantes da série de dados, ou seja, valores que não demonstram a realidade de um conjunto de dados.

2.7 Teste de Hipóteses

O teste de hipótese é um método utilizado para verificar se os dados são compatíveis com alguma hipótese, podendo concordar ou rejeitar com a hipótese testada. Ele é composto por duas hipóteses:

$$\begin{cases} H_0 : \text{Hipótese a ser testada} \\ H_1 : \text{Hipótese aceitável caso a nula seja rejeitada} \end{cases}$$

2.7.1 Estatística do Teste

A Estatística do teste é um valor obtido na amostra relacionado à distribuição do parâmetro testado. Dessa forma, cada teste possui sua estatística que apresenta uma relação com uma determinada distribuição teórica de probabilidade.

2.7.2 Nível de Significância

O nível de significância é a probabilidade com a qual se sujeita a correr o risco de um Erro do Tipo I (rejeitar uma hipótese nula dada que ela é verdadeira). O nível de significância deve ser determinado antes da extração da amostra. Um nível de significância de $\alpha = 0,05$ significa que se tem a probabilidade de errar 5 em 100 vezes, ou seja, em 100 vezes, a probabilidade de se tomar a decisão correta é de 95%.

2.7.3 P-valor

Nível descritivo ou p-valor é uma medida utilizada para sintetizar o resultado de um teste de hipóteses. Ele é formalmente definido como a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra, assumindo verdadeira a hipótese nula. Se $p\text{-valor} < \alpha$ concluí-se que há evidências estatísticas suficientes para se rejeitar a hipótese nula.

2.8 Teste de Qui-Quadrado de Independência

Qui-Quadrado, simbolizado por χ^2 , é um teste de hipótese que se destina a encontrar um valor da dispersão para duas variáveis nominais, avaliando a associação existente entre variáveis. O princípio básico deste método é comparar proporções, isto é, as possíveis divergências entre as frequências observadas e esperadas para um certo evento. Para esse teste, foram formuladas as seguintes hipóteses:

$$\left\{ \begin{array}{l} H_0 : \text{As variáveis são independentes} \\ H_1 : \text{As variáveis não são independentes} \end{array} \right.$$

Este teste é baseado no cálculo dos valores esperados. Os Valores Esperados são os valores que seriam observados, caso a hipótese nula fosse verdadeira:

$$E_{ij} = \frac{(\text{Total da linha } i) \times (\text{total da coluna } j)}{\text{Total geral}}$$

E a estatística do teste é:

$$\chi^2 = \sum \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx \chi^2[(r - 1) \times (s - 1)]$$

em que

- E_{ij} é valor esperado na i-ésima linha e na j-ésima coluna;
- O_{ij} é valor observado na i-ésima linha e na j-ésima coluna;
- r é o número total de linhas;
- s é o número total de colunas.

Para que a aproximação Qui-quadrado seja satisfatória, é preciso que a amostra seja relativamente grande, ou seja, pode ser considerado um tamanho de amostra suficiente se todos os valores esperados, forem no mínimo iguais a 5.

2.9 Teste Exato de Fisher

O teste de Fisher é utilizado para comparar variáveis qualitativas, quando os pressupostos do teste Qui-Quadrado não forem obedecidos, analisando o grau de dependência entre as variáveis. Assim, as hipóteses desse teste são:

$$\begin{cases} H_0 : \text{As variáveis estudadas não dependem uma da outra} \\ H_1 : \text{As variáveis estudadas dependem uma da outra} \end{cases}$$

Após estipuladas as hipóteses, a função hipergeométrica é utilizada. Ela é dada pela seguinte fórmula:

$$P = \frac{(A + B)!(C + D)!(A + C)!(B + D)!}{N!A!B!C!D!}$$

em que

- A = número de observações na casela A;
- B = número de observações na casela B;
- C = número de observações na casela C;
- D = número de observações na casela D;
- N = número de observações na amostra.

Assim, se um p-valor menor que o nível de significância de 5% for obtido, rejeita-se a hipótese nula de que as variáveis não são dependentes, com nível de confiança de 95%.

2.10 Teste de Wilcoxon

O teste de Wilcoxon é um teste não paramétrico que tem o objetivo de verificar se duas amostras aleatórias X e Y possuem a mesma função de distribuição. Em função disso, esse teste também pode ser utilizado para a comparação de médias entre amostras dependentes. Seja $D_i = x_i - y_i$ para $i=1, \dots, n$. Então, a amostra D_1, \dots, D_n é obtida a partir das diferenças entre os valores de cada par. As hipóteses desse teste são:

$$\begin{cases} H_0 : \mu_x - \mu_y = \mu_d = 0 \\ H_1 : \mu_x \neq \mu_0 \neq 0 \text{ ou } \mu_x > \mu_y \text{ ou } \mu_x < \mu_y \end{cases}$$

em que:

μ_x = média populacional de X;

μ_y = média populacional de Y.

O teste é realizado a partir da ordenação da variável D_i e postos (ou ranks) são atribuídos a cada observação. Algumas observações podem receber a mesma posição na ordenação. Esse fenômeno é denominado empate. A estatística do teste é calculada da seguinte maneira:

Considere:

$$\begin{cases} H_0 : R(x_i, y_i) \text{ se } D > 0 \\ H_1 : -R(x_i, y_i) \text{ se } D < 0 \end{cases}$$

em que $R(x_i, y_i)$ é o posto associado a (x_i, y_i) . Então:

- Se não houver empates, a estatística a ser usada é:

$$T^+ = \sum (R_i \text{ com } D_i > 0)$$

ou seja, a soma dos postos positivos.

- Se houver empates, a estatística a ser usada tem distribuição Normal com média zero e variância 1 e é dada por:

$$V = \frac{\sum_{i=1}^n R_i}{\sum_{i=1}^n R_i^2}$$

Caso o p-valor seja menor que o nível de significância de 5%, a hipótese nula é rejeitada.

2.11 Teste de Kruskal Wallis

O teste de Kruskal-Wallis é utilizado para comparar dois ou mais grupos independentes, onde a distribuição de probabilidade não necessita ser normal. É um método baseado em postos da forma que para cada valor de uma variável quantitativa será atribuído um posto. O teste de Kruskal-Wallis consiste exatamente em comparar estes postos.

Desta maneira as hipóteses do teste de Kruskal-Wallis são formuladas da seguinte maneira:

$$\begin{cases} H_0 : \text{Não existe diferença entre os grupos} \\ H_1 : \text{Pelo menos um grupo difere dos demais} \end{cases}$$

A estatística do teste de Kruskal-Wallis é definida da seguinte maneira:

$$H_{Kruskal-Wallis} = \frac{\left[\frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right] - 3(N+1)}{1 - \left[\frac{\sum_j (t_j^3 - t_j)}{N^3 - N} \right]} \approx \chi^2_{(k-1)}$$

Rejeita-se H_0 se o p-valor < nível de significância. Tem-se que:

- k é o número de grupos;
- R_i é a soma dos postos do grupo i ;
- n_i é número de elementos do grupo i ;
- N é o tamanho total da amostra;
- t_j número de elementos no j -ésimo empate (se houver).

2.12 Regressão Logística

A análise de regressão logística binária é um instrumento eficaz para verificar a relação entre duas ou mais variáveis no caso específico em que a resposta é dicotomizada em sucesso e fracasso. A regressão logística multinomial estende a regressão logística binária a um modelo com múltiplas categorias na variável resposta. Porém, há uma limitação: não se assume uma ordenação dessas categorias. Sua modelagem é feita a partir da equação:

$$P(Y_i = 1 | X_{1i}, \dots, X_{pi}) = \pi(X_i) = \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}}}$$

em que a probabilidade da variável resposta ser igual a 1 está em função das variáveis explicativas X_i , $i = 1, 2, \dots, p$.

Essa equação pode ser escrita de maneira linear pela transformação logito:

$$\pi^*(X_i) = \ln \left(\frac{\pi(X_i)}{1 - \pi(X_i)} \right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$$

O parâmetro β_j corresponde ao efeito do aumento de uma unidade de X_j sobre o logaritmo da odds de $Y = 1$, mantendo as outras variáveis constantes. Dessa forma, a $\exp(\beta_j)$ tem como efeito a multiplicação na odds de $Y = 1$ para o aumento de uma unidade de X_j , mantendo as demais variáveis constantes.

2.13 Teste de Wald

O Teste de Wald é utilizado para verificar se o efeito da variável explicativa é significativo. As hipóteses desse teste são dadas por:

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

A estatística do teste é obtida por comparação entre a estimativa de máxima verossimilhança do parâmetro ($\hat{\beta}_i$) e a estimativa de seu desvio padrão:

$$Z^2 = \left(\frac{\hat{\beta}_i}{\widehat{DP}(\hat{\beta}_i)} \right)^2 \sim X^2$$

2.14 Teste da Razão de Verossimilhança

O teste de Razão de Verossimilhança testa a significância conjunta das variáveis explicativas. Usualmente é utilizado para verificar se o modelo reduzido é mais adequado em relação ao completo, conforme as hipóteses:

$$\begin{cases} H_0 : \ln \left(\frac{\pi(X_i)}{1-\pi(X_i)} \right) = \beta_0 + \beta_{m+1}X_{i(m+1)} + \dots + \beta_p X_{pi} \\ H_1 : \ln \left(\frac{\pi(X_i)}{1-\pi(X_i)} \right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_m X_{im} + \beta_{m+1}X_{i(m+1)} + \dots + \beta_p X_{pi} \end{cases}$$

Com estatística de teste sob H_0 :

$$G^2 = -2(L_0 - L_1) \sim X^2_{(p+1)-(p+1-m) = m \text{ g.l.}}$$

em que

- L_0 é o máximo da log verossimilhança do modelo sob H_0 ;
- L_1 é o máximo da log verossimilhança do modelo sob H_1 .

2.15 Critério de Informação

Os critérios de informação são medidas que auxiliam na seleção de modelos. Um dos critérios existentes é o de Akaike (AIC).

$$AIC = -2(\text{Log Verossimilhança} - p)$$

em que p é o número de parâmetros do modelo.

Porém, além dele não ser consistente, ele não é o ideal para comparar modelos com número de parâmetros diferentes. Por isso, utiliza-se o critério de Akaike Corrigido (AIC_c).

$$AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2}$$

em que n é o tamanho da amostra.

Outro critério é o Schwarz (SC), baseado, em parte, na função de verossimilhança e é dado por:

$$SC = -2\log L + p \log \sum_j f_j n_j$$

em que p é o número de parâmetros no modelo, n_j é o número de ensaios quando a sintaxe dos eventos/ensaios são especificados e $n_j = 1$ a sintaxe de um único ensaio. Valores menores para o SC implica em menos variáveis explicativas, ou melhor ajustamento, ou os dois.

2.16 Teste de Hosmer e Lemeshow

O teste de Hosmer e Lemeshow é utilizado para testar a adequabilidade do modelo, verificando se o modelo proposto explica bem o que é observado. Para esse teste, assume-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{O modelo de regressão logística ajusta-se bem aos dados} \\ H_a : \text{O modelo de regressão logística não se ajusta bem aos dados} \end{cases}$$

Para realizar o teste é necessário dividir os dados em grupos com base nos percentis das probabilidades estimadas. A estatística do teste é o Qui-quadrado de Pearson:

$$X^2 = \sum_{i=1}^c \sum_{j=0}^1 \frac{(f_{ji} - f_{eji})^2}{f_{eji}}$$

2.17 Deviance

A Deviance resulta em um teste com hipótese nula de que todos os parâmetros que estão no modelo saturado mas não estão no modelo testado são iguais a zero. A Deviance tem distribuição aproximadamente Qui-Quadrado (X^2) com a diferença entre os parâmetros do modelo como graus de liberdade.

$$Deviance = -2[L_M - L_S]$$

sendo que L_M denota o log da função de verossimilhança do modelo M de interesse e L_S denota o log da função de verossimilhança do modelo completo, também conhecido como modelo saturado (ele contém todas as variáveis explicativas).

Para se comparar mais de um modelo simultaneamente basta utilizar a deviance parcial, dada pela seguinte expressão:

$$Deviance\ Parcial = -2[L_0 - L_S] - [-2(L_1 - L_S)]$$

no qual L_0 é o log da função de verossimilhança de um modelo (M_0) que é um caso especial de um modelo (M_1) cujo logaritmo da função de verossimilhança é L_1 . Para amostras grandes a expressão acima tem distribuição aproximadamente Qui-Quadrado (X^2) com graus de liberdade igual à diferença de graus de liberdade entre cada expressão $[L_i - L_S]$ que por sua vez tem seus graus de liberdade como a diferença entre o número de parâmetros de L_i e de L_S .

2.18 Dfbeta

O Dfbeta, ou diferença nos valores beta, é uma medida que permite verificar a influência de cada observação nos parâmetros estimados. Seu cálculo é feito excluindo cada observação i individualmente e verificando a diferença no parâmetro j , dividido pelo seu desvio padrão.

$$(DFbeta)_{j(i)} = \frac{b_j - b_{j(i)}}{\sigma(\bar{b}_j)}$$

Para grandes amostras observação é considerada influente se:

- $|DFbeta| > \frac{2}{\sqrt{n}}$.

Para amostras pequenas e médias é considerado influente se :

- $|DFbeta| > 1$

2.19 Leverage

O leverage é uma medida que permite a identificação de observações discrepantes por meio da distância entre o valor da variável X para a observação i e a média de X.

Ele é calculado a partir dos elementos h_{ii} correspondentes à diagonal da matriz chapéu calculada como:

$$H = X' \cdot (X' \cdot X)^{-1} \cdot X$$

Nota-se que os valores de h_{ii} podem variar entre 0 e 1 e a soma de todos resulta no número de parâmetros. Considera-se que valores superiores a 0,5 indicam um leverage alto e entre 0,2 e 0,5 moderado.

2.20 Curva ROC

A curva ROC tem como função verificar se o modelo está bem ajustado aos dados, verificando os níveis de especificidade e sensibilidade para cada possível ponto de corte. Essas medidas fornecem informação a respeito da classificação do modelo, entre a categoria real da observação e a que foi predita (a matriz de confusão explicita cada uma das categorias). Para a curva ROC ser perfeita, é necessário que tenha sensibilidade ($P(\hat{Y} = 1|Y = 1)$) igual à 1 e que o complementar da especificidade ($1 - P(\hat{Y} = 0|Y = 0)$) seja igual a 0. É esperado que o modelo se aproxime ao máximo da curva perfeita, aumentando sua área abaixo da curva, indicando um bom modelo. Além de verificar a qualidade de ajuste, a curva ROC também é utilizada para determinar o ponto de corte que melhor se ajusta aos dados, uma vez que o ponto intuitivo de 0,5 não é o ideal em todos os casos.

2.21 Métodos Automáticos de Seleção de Variáveis

Qualquer procedimento para seleção ou exclusão de variáveis de um modelo é baseado em um algoritmo que checa a importância das variáveis, incluindo ou excluindo-as do modelo se baseando em uma regra de decisão. A importância da variável é definida em termos de uma medida de significância estatística do coeficiente associado à variável para o modelo.

O método *forward* se inicia com nenhuma variável no modelo, apenas o intercepto. Esse método se resume em adicionar uma variável a cada passo. Assim, calcula-se a estatística "F" para cada variável e seu respectivo p-valor. Dentre as variáveis com p-valor menor que o nível de significância α de 5%, adiciona-se aquela com menor

p-valor para compor o modelo. Repete-se esse procedimento até que mais nenhuma variável apresente $p\text{-valor} < \alpha$.

Já o método *backward* de seleção de variáveis se inicia com todas as variáveis explicativas no modelo. Para cada uma dessas variáveis, calcula-se a estatística "F" e seu p-valor. Dentre todas as variáveis que apresentarem $p\text{-valor} > \alpha$, seleciona-se aquela com maior p-valor para sair do modelo de regressão. Repete-se esse procedimento até que nenhuma variável apresente $p\text{-valor} > \alpha$.

Por fim, o método *stepwise* pode começar tanto com o modelo sem nenhuma variável como com todas. Caso comece com nenhuma o algoritmo realiza testes para verificar a inclusão da variável e logo após testa para retirar outra. Se iniciar com o modelo completo, o procedimento acontece de maneira invertida. Assim como os anteriores, os testes são feitos respeitando o nível de significância para verificar a inclusão ou exclusão das variáveis no modelo.

3 Frequência de Amicus Curiae por relator

A fim de se ter uma melhor ideia da quantidade de Amicus Curiae (AC) de cada um dos relatores em que a data de julgamento foi no período contemplado nesse estudo, formulou-se a tabela abaixo.

Tabela 1: Frequência de AC por Relator

Relator	AC			
	Cooperativo	Servidor Público	Órgao Público	Rep. Sociedade Civil
2	12	0	0	24
3	18	0	2	13
4	11	0	20	3
5	1	0	0	0
6	20	2	2	3
7	2	1	5	9
8	2	0	0	0
9	1	0	0	4
10	9	0	6	0
11	17	0	0	2
13	0	0	1	11
14	10	0	28	11
15	17	0	3	12
16	1	0	0	1
17	13	0	3	20
18	0	0	0	1
19	10	9	13	2
20	28	22	23	1
Total	172	34	106	137

Pela tabela acima, pode-se ver que os julgamentos de responsabilidade dos relatores César Peluso e Roberto Barroso tiveram apenas um Amicus Curiae no período analisado. O relator com maior número de AC foi Teori Zavascki, com 74 Amicus Curiae, 25 a mais que o segundo colocado, Luis Fux. É interessante notar que Amicus Curiae servidores públicos estiveram presentes apenas em **25 julgamentos**, sendo 22 deles (88%) do relator Teori Zavascki.

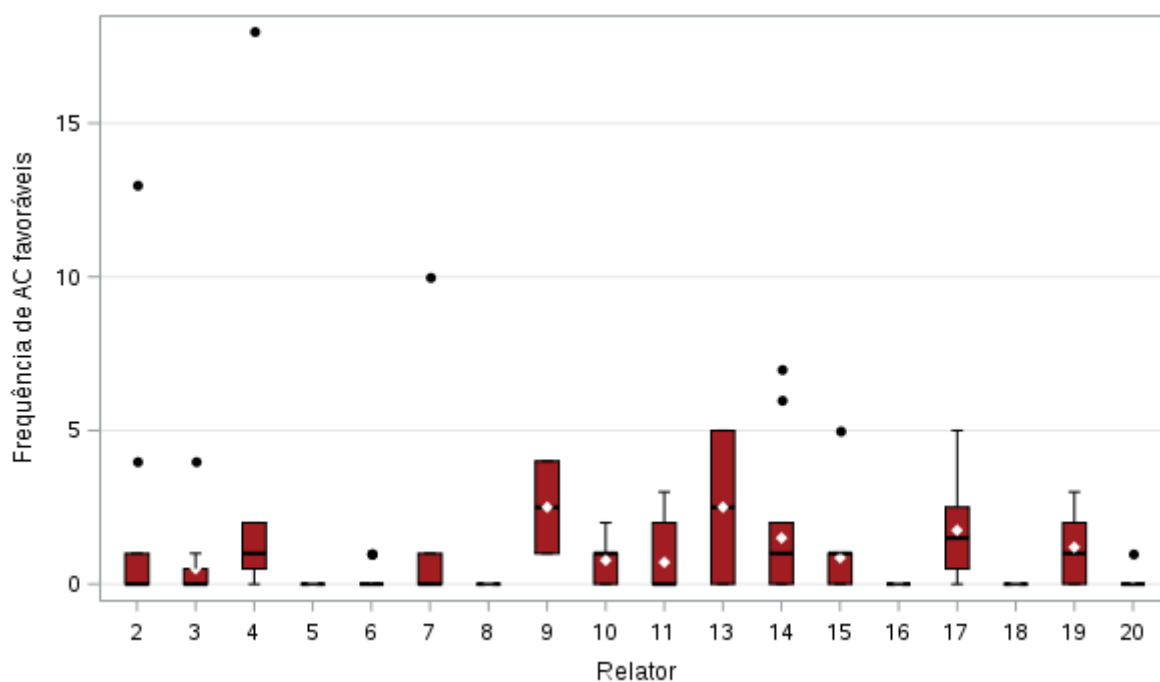
de onde tirou?

Além disso, todos os AC servidores públicos foram favoráveis, não havendo nenhum desfavorável. Ainda, AC cooperativo foi a categoria mais frequente, com 172 AC's no total, 55 a mais que o segundo colocado, representante da sociedade civil.

4 Análise entre número de AC favoráveis e Relator

O estudo abaixo corresponde a verificar a relação entre as variáveis *AC favoráveis* e *relator*. Assim, observe o box-plot da frequência de *AC favorável* por cada *relator*.

Figura 2: Box-plots da frequência de AC's favoráveis por relator



Ao observar os box-plots acima, nota-se que os *relatores* que possuem maior frequência de AC's positivos são, respectivamente, os relatores 13 e 9. Além disso, nota-se que os relatores 4, 17 e 19 possuem também um destaque por sua frequência maior.

Por fim, nota-se que os relatores 5, 6, 8, 16, 18 e 20 não possuem resultado positivo com relação ao número de *Amicus Curiae*

5 Procedimentos para a Regressão Logística

Para a decisão do melhor modelo de regressão logística será verificado a multicolinearidade das variáveis respostas. Em seguida, serão criados alguns modelos que atendem às suposições do cliente e eles serão comparados por meio de critérios de informação de Akaike (AIC), Bayesiano (SC) e por meio do logaritmo da máxima verossimilhança. Então, será realizada a análise de diagnóstico dos modelos para a confirmação dos seus ajustes. Assim, será criado um modelo que foi considerado o mais adequado pelos responsáveis.

Os modelos testados, de forma geral, têm como variáveis controle tempo de tramitação, composição da corte, posição PGR, tipo de requerente e relator. O primeiro modelo possui, além dessas variáveis, Amicus Curiae (AC) favoráveis e Amicus Curiae desfavoráveis. Já o segundo utiliza AC cooperativo, AC servidor público, AC órgão/agente público e AC representante da sociedade civil, todas separadas entre favoráveis e desfavoráveis. Para o último modelo foram criadas quatro novas variáveis, sendo cada uma delas a diferença entre Amicus Curiae favoráveis e desfavoráveis das categorias do segundo modelo (cooperativo, servidor público, órgão/agente público e representante da sociedade civil).

5.1 Multicolinearidade

Ao analisar as correlações e associações entre as variáveis explicativas do modelo dois a dois, pelos testes de Kruskal-Wallis, Qui-Quadrado e Fisher verifica-se a independência das variáveis. Para o qual, todas não indicam nenhuma relação forte com exceção das apresentadas no quadro a seguir:

Quadro 1: Variáveis correlacionadas

Variáveis	Teste	P-valor
PGR com o tipo de requerente	Qui-quadrado	0,01566
Relator com AC rep. da sociedade civil desfavoráveis	Kruskall-Wallis	0,0015
Tipo de requerente com AC órgão público favoráveis	Kruskall-Wallis	0,0068
Relator com Diferença de AC de rep. da sociedade civil	Kruskall-Wallis	0,0473

Foram apresentados as associações com p-valor inferior a 0,05 de acordo com os

testes utilizados, entretanto, não será considerado multicolinearidade, já que acredita-se que há interferência devido ao pequeno tamanho do banco de dados.

5.2 Modelo para Cita no Voto

Para a variável resposta cita no voto foram testados os modelos com e sem interação utilizando os métodos de seleção de variáveis *forward*, *backward* e *stepwise*. Não foi registrada nenhuma diferença ao realizar as interações e, por motivos de simplificação, a tabela abaixo cita apenas os casos em que os métodos de seleção convergiram e que não apresentaram apenas o intercepto.

Tabela 2: Modelos reduzidos para a variável cita no voto

Modelo	Método de Seleção	Modelo reduzido	AIC	SC	-2 log L
Modelo 1	<i>forward</i>	x_{10}	153,78	173,58	139,78
Modelo 2	<i>forward</i>	$x_{10} + x_{20}$	146,96	169,59	130,96
Modelo 3	<i>backward</i>	x_{20}	150,43	156,08	146,43

Como mencionado anteriormente, a tabela acima fornece os modelos reduzidos com as devidas variáveis explicativas para a variável *cita no voto*. Note que os modelos reduzidos no quadro não necessariamente correspondem a cada um dos modelos sugeridos.

Para o primeiro modelo proposto o método *forward* de seleção de variáveis selecionou apenas a variável *tipo de requerente* (x_{10}). Já os métodos *backward* e *stepwise* resultaram em um modelo apenas com o intercepto, ou seja, sem nenhuma variável explicativa significativa. Para o segundo modelo os métodos de seleção também não convergiram. Dessa forma, o método *forward* incluiu as variáveis *tipo de requerente* (x_{10}) e *Amicus Curiae representante da sociedade civil desfavoráveis* (x_{20}) e para os métodos *backward* e *stepwise* obteve-se um modelo com apenas essa última variável. Por fim, no último modelo os métodos não convergiram novamente. Para o *forward* o modelo reduzido incluiu a variável *tipo de requerente* e os métodos *backward* e *stepwise* mantiveram apenas o intercepto.

Sabe-se que quanto menor o valor do critério AIC, SC e da log verossimilhança, melhor a adequabilidade do modelo selecionado. Dessa forma, por meio da tabela acima, pode-se observar que o modelo 2 apresenta o menor AIC (146,90) e log da

verossimilhança (130,96). Já para o critério SC, o modelo 3 apresenta menor valor.

A fim de testar a adequabilidade dos modelos, realizou-se o teste de Hosmer e Lemeshow e calculou-se a curva ROC de cada um deles. Para esse teste, assume-se como hipótese nula o modelo de regressão se ajustar bem aos dados e, conseqüentemente, a hipótese alternativa consiste em o modelo não se ajustar bem. Para esse teste e curva ROC obteve-se os seguintes resultados:

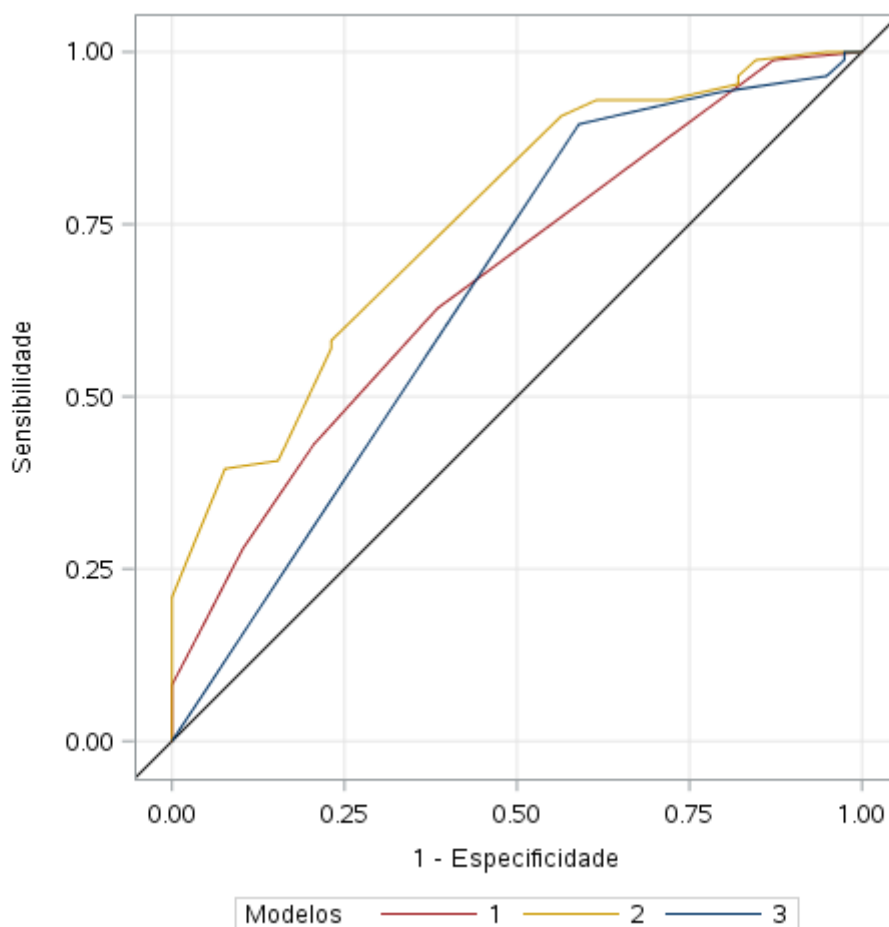
Quadro 2: Área do modelo da Curva Roc

Modelo	Curva ROC	p-valor
x_{10}	0,6710	>0,9999
$x_{10} + x_{20}$	0,7572	0,1553
x_{20}	0,6498	0,0428

O p-valor correspondente no quadro acima está relacionado ao teste de Hosmer e Lemeshow. Dessa forma, não há evidências para se rejeitar a hipótese nula para os modelos 1 (que possui apenas a variável explicativa *tipo de requerente*) e 2 (com as variáveis *tipo de requerente* e *Amicus Curiae representante da sociedade civil desfavoráveis*), com nível de significância de 5%. Por outro lado, para o modelo com a variável explicativa *AC representante da sociedade civil desfavoráveis*, há evidências para se rejeitar a hipótese de o modelo se ajustar bem aos dados.

Já em relação à curva ROC, quanto maior a área da curva gerada pelo modelo, melhor. Dessa forma, pode-se observar que o segundo modelo é melhor em relação a esse critério, seguido do primeiro e terceiro modelo, respectivamente. Para efeito de comparação e uma melhor visualização da curva ROC referente a cada um desses modelos, gerou-se o gráfico abaixo com as três curvas.

Figura 3: Curva Roc dos modelos de cita no voto



A curva ROC mostra o poder de predição de cada modelo e é mais informativa que uma tabela de classificação pelo fato de resumir o poder de predição de todos os possíveis π_0 . Para uma dada especificidade, um poder de predição melhor corresponde a uma sensibilidade maior. Assim, quanto melhor o poder de predição, maior a curva ROC.

Um valor de $c=0,5$ significa que as predições não foram nada melhores que um chute. Isso corresponde a um modelo com apenas o intercepto e sua curva ROC é uma linha reta conectando os pontos (0,0) e (1,1), como se pode observar pelo gráfico acima (linha preta). Como pode-se ver, o modelo com as variáveis *tipo de requerente* e *AC representante da sociedade civil desfavoráveis* é o que possui maior área acima da linha preta de referência, sendo maior que os demais modelos em todos os pontos do gráfico, mostrando ser o com maior poder de predição.

As últimas ferramentas utilizadas para verificar qual o modelo mais se ajusta para explicar a variável cita no voto são as medidas DFbeta e Leverage. A primeira detecta

valores influentes em relação a cada parâmetro do modelo e a segunda *outliers* (valores discrepantes). Nota-se que nos modelos 1 e 2 há apenas um valor influente, visto que o DF_{beta} , em módulo, é maior que 1, sendo tal valor referente à ação ADC19. Além disso em ambos os modelos foram detectados três valores discrepantes com leverage moderado (maior que 0,3, mas menor que 0,4). Já o terceiro modelo não apresentou nenhum valor influente nem discrepante, de acordo com seus DF_{betas} e Leverage.

Com base em todas as análises realizadas, optou-se pelo modelo que tem como variáveis explicativas *tipo de requerente* e *AC representante da sociedade civil desfavoráveis*. Para ele, obteve-se o maior valor para a curva ROC e não se rejeitou a hipótese nula no teste de Hosmer e Lemeshow, mostrando que o modelo ajusta-se bem aos dados.

Além disso, realizou-se o teste da razão de verossimilhança para comparar os modelos 1 (com a variável *tipo de requerente*) e 2 (com *AC representante da sociedade civil desfavoráveis* e *tipo de requerente*), testando assim se o efeito da variável *AC representante da sociedade civil desfavoráveis* é nulo. Para esse teste utilizou-se a deviance e houve evidências suficientes para se rejeitar a hipótese nula com nível de significância de 5%. Com isso, o efeito da variável *AC representante da sociedade civil desfavoráveis* não é nulo e, portanto, **o modelo 2 se mostrou ser o mais adequado para explicar a variável *cita no voto*.**

Para o modelo final, obteve-se as seguintes estimativas de seus parâmetros, erro padrão e estatística e p-valor para o teste de Wald:

Tabela 3: Análise das estimativas de máxima verossimilhança

Variável	Estimativa	Erro Padrão	Teste de Wald	
			Estatística	P-valor
Intercepto	-1,4627	1,1051	1,7519	0,1856
Tipo de Requerente 4	14,8676	407,2	0,0013	0,9709
Tipo de Requerente 5	3,0288	1,2396	5,9706	0,0145
Tipo de Requerente 6	2,6555	1,2004	4,8936	0,0270
Tipo de Requerente 7	14,8676	470,2	0,0010	0,9748
Tipo de Requerente 8	3,3261	1,3023	6,5229	0,0106
Tipo de Requerente 9	2,1921	1,1462	3,6579	0,0558
AC rep. da sociedade civil desfavoráveis	-0,6469	0,2472	6,8492	0,0089

Pode-se ver que a variável **AC representantes da sociedade civil desfavoráveis e o intercepto** possuem um efeito negativo sobre a variável resposta *cita no voto*. Além disso, outro fato importante é o p-valor para o teste de Wald (utilizado para verificar se o efeito da variável explicativa é significativo) das categorias 4, 7 e 9 da variável *tipo de requerente*. Assim, ao obter-se um p-valor maior que o nível de significância de 5%, não há evidências para se rejeitar a hipótese desses efeitos não serem significativos, ou seja, não influenciam na variável resposta.

Isso pode se dever ao fato de nos tipos de requerente 4 e 7, apenas a categoria "não menciona" da variável *cita no voto* fazer parte da amostra. Por fim, é importante observar que a categoria de referência da variável *tipo de requerente* é 1 (presidente da república) e, por isso, ela não está mencionada na tabela acima.

Com base nos valores da razão de chance (ou *odds ratio*) apresentadas na tabela abaixo é possível obter interpretações a cerca dos parâmetros da regressão logística.

Tabela 4: Estimativas da razão de chances

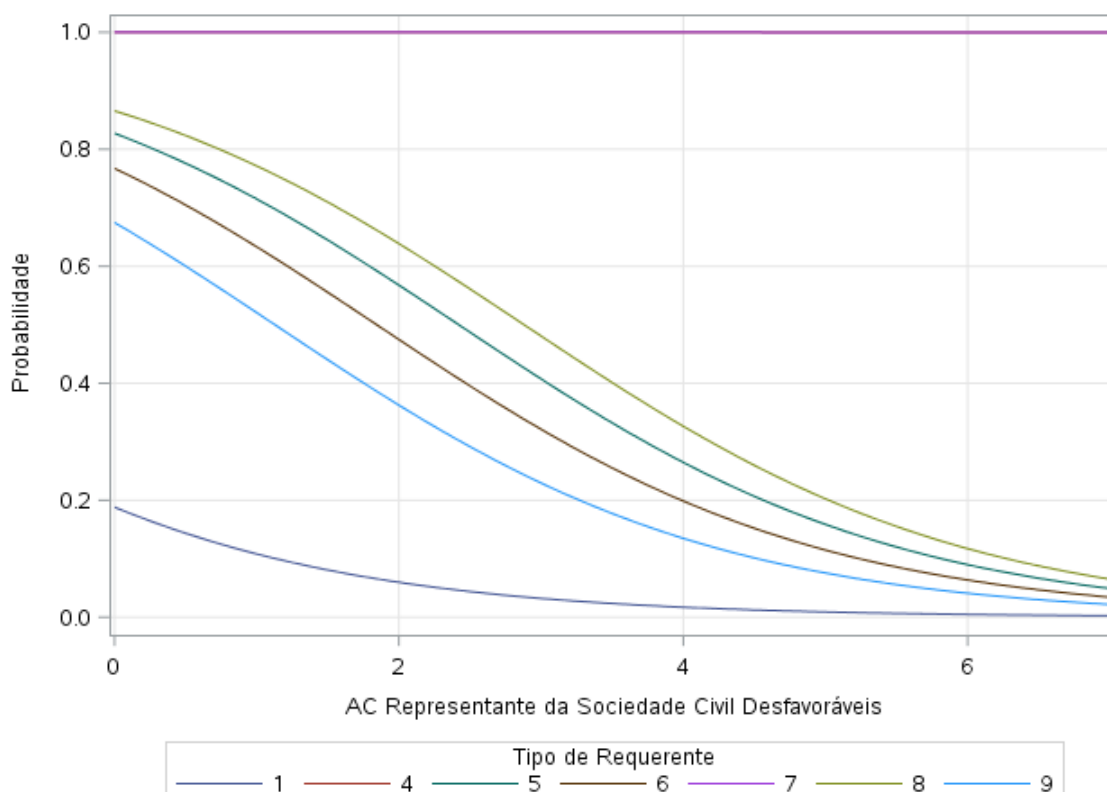
Efeito	Estimativa Pontual	Intervalo de 95% de Confiança de Wald	
Tipo de Requerente 4 vs 1	>999,999	<0,001	>999,999
Tipo de Requerente 5 vs 1	20,673	1,821	234,705
Tipo de Requerente 6 vs 1	14,232	1,354	149,640
Tipo de Requerente 7 vs 1	>999,999	<0,001	>999,999
Tipo de Requerente 8 vs 1	27,831	2,168	357,330
Tipo de Requerente 9 vs 1	8,954	0,947	84,660
AC rep. da sociedade civil desfavoráveis	0,524	0,323	0,850

Nota-se que as estimativas das categorias 4 e 7 da variável *tipo de requerente* não são passíveis de se calcular e interpretar. Isso se deve ao fato de que em ambas apenas a categoria "não menciona" da variável *cita no voto* fazer parte da amostra. Ou seja, nela, 100% dos tipos de requerente 4 e 7 tem resposta *cita no voto = não menciona*, não apresentando nenhuma observação para a categoria complementar "considera argumentos", como citado anteriormente. Já na categoria 9 observa-se que o intervalo de confiança contém o 1, o que novamente indica que tal tipo de requerente não influencia na variável *cita no voto*.

Além disso, mantendo-se o nível das demais variáveis constantes, tem-se que a chance de **não ser mencionado no voto aumenta entre 82,1% e 23.379,5% quando o tipo de requerente é o 5 em relação ao do tipo 1.** Quando o requerente é do tipo 6 a chance de não ser citado no voto aumenta entre 35,4% e 14.864,0% ao comparar com os relatores do tipo 1, mantendo as outras variáveis constantes. Já ao passar de um requerente do tipo 1 para um do tipo 8 a chance de não ser mencionado no voto é multiplicada por um número entre 2,168 e 357,330. **Por fim para cada aumento de 1 AC representante da sociedade civil desfavorável a chance de sucesso (cita no voto = não menciona) diminui entre 15,0% e 67,7%, novamente mantendo as demais variáveis constantes.**

Com o objetivo de se ter uma melhor visualização do efeito de cada variável explicativa do modelo selecionado, optou-se por apresentar o gráfico abaixo:

Figura 4: Gráfico de probabilidade predita para *cita no voto* = não menciona



A figura acima mostra as probabilidades preditas para a ocorrência de sucesso (*cita no voto* = não menciona) para cada *tipo de requerente* em função da quantidade de *AC representante da sociedade civil desfavoráveis*. É possível notar que essas probabilidades decrescem à medida que o número de Amicus Curiae representante da sociedade civil desfavoráveis cresce. Além disso, a probabilidade de sucesso do *tipo de requerente 8* é sempre maior que a do 5, que por sua vez é maior que a do 6. Essa última ainda é maior que essa mesma probabilidade para o *tipo de requerente 9*, a qual é maior que aquela para o *tipo de requerente 1*. Dessa forma, essas probabilidades nunca se cruzam. É importante destacar que as curvas de probabilidade predita para os *tipos de requerente 4 e 7* não decresce em nenhum momento, sempre permanecendo na probabilidade 1, pois apenas a categoria "não menciona" está presente na amostra para esses *tipos de requerente*.

5.3 Modelo para Resultado

O objetivo agora é encontrar o melhor modelo que explica a variável *resultado*. Assim como realizado para a formulação do modelo de regressão para a variável *cita no voto*, serão utilizados alguns critérios de informação, como o de Akaike (AIC) e Bayesiano (SC), além do logaritmo da máxima verossimilhança. Por fim, serão realizados testes de adequabilidade e análise de diagnóstico para cada um dos modelos.

Os modelos testados, de forma geral, possuem as mesmas variáveis controle explicitadas na seção 5.2. Porém, além dessas, incluem também a variável *cita no voto*. Para a variável resposta *resultado* agregou-se as categorias 1 e 2 (resultado procedente e parcialmente procedente, respectivamente). Isso foi realizado pelo fato da amostra ser pequena e, portanto, a fim de melhorar a qualidade dos modelos.

O primeiro modelo possui, além dessas variáveis de controle, *Amicus Curiae favoráveis* e *Amicus Curiae desfavoráveis* como variáveis explicativas. Já o segundo modelo tem *AC corporativo favorável*, *AC corporativo desfavorável*, *AC servidor público favorável*, *AC servidor público desfavorável*, *AC órgão/agente público favorável*, *AC órgão/agente público desfavorável* e *AC representante da sociedade civil favorável* e *desfavorável*. Por fim, o terceiro modelo possui quatro variáveis explicativas além das de controle, sendo cada uma delas a diferença entre *Amicus Curiae favoráveis* e *desfavoráveis* para cada uma das categorias do segundo modelo (*corporativo*, *servidor público*, *órgão/agente público* e *representante da sociedade civil*).

Para a variável *resultado* também foram testados os modelos com e sem interação entre as variáveis qualitativas e quantitativas utilizando os métodos de seleção automática de variáveis *forward*, *backward* e *stepwise*. Diferentemente do ocorrido com a variável resposta *cita no voto*, alguns modelos com interações foram escolhidos pelos métodos automáticos de seleção, sendo elas, dessa forma, significativas.

A fim de simplificar, para os modelos em que os métodos de seleção convergiram, a tabela abaixo menciona apenas um desses métodos. Nela estão explicitados os modelos selecionados e os métodos utilizados para obtê-los, além de seus respectivos critérios de informação e logaritmo da máxima verossimilhança.

Tabela 5: Modelos reduzidos para a variável *resultado*

Modelo	Método Seleção	Modelo Reduzido	AIC	SC	-2 log L
Modelo 1	<i>backward</i>	$x_5 + x_6 + x_8 + x_9 + x_{12} + x_8 * x_{12} + x_8 * x_6 + x_9 * x_{12}$	131,80	165,74	107,80
Modelo 2	<i>forward</i>	$x_5 + x_{10} + x_{12}$	135,05	163,34	115,05
Modelo 3	<i>backward</i>	$x_5 + x_{12}$	143,09	154,40	135,09
Modelo 4	<i>backward</i>	$x_5 + x_{12} + x_{14} + x_{14} * x_{12}$	136,32	156,11	122,32
Modelo 5	<i>backward</i>	$x_5 + x_{12} + x_{23} + x_{25} + x_{23} * x_{12}$	126,67	149,30	110,67

É importante ressaltar que os modelos apresentados na tabela acima não necessariamente correspondem a cada um dos modelos sugeridos. Além disso, a variável representada por x_5 é o tempo de tramitação, x_6 cita no voto, x_8 é AC favoráveis e x_9 AC desfavoráveis. Tipo de requerente é representada por x_{10} e PGR por x_{12} . Ainda, x_{14} é a quantidade de AC corporativo favoráveis, x_{23} é a diferença entre AC corporativo favoráveis e desfavoráveis e, por último, x_{25} é a diferença entre AC órgão/agente público favoráveis e desfavoráveis.

Assim, para o primeiro modelo proposto, ou seja, com as variáveis AC favoráveis e AC desfavoráveis além das variáveis de controle e interações, o modelo 1 acima foi selecionado pelo método *backward* de seleção. Esse modelo apresentou o menor log da verossimilhança dentre todos os acima mencionados. Além disso, o modelo 2 também foi selecionado para o primeiro modelo proposto, mas por meio do método *forward* e *stepwise* de seleção de variáveis.

Já para o segundo modelo proposto, ou seja, aquele que possui, além das variáveis de controle e interações, AC corporativo, AC servidor público, AC órgão/agente público e AC representante da sociedade civil, todas separadas em favoráveis e desfavoráveis, os modelos 2, 3 e 4 foram obtidos. O modelo 2 foi selecionado por meio do método *forward*, 3 pelo *stepwise* e o 4 pelo *backward*.

Para o terceiro e último modelo proposto, que possui como variáveis explicativas, além das de controle, a diferença de AC favoráveis e desfavoráveis para cada categoria de AC (corporativo, servidor público, órgão/agente público e representante da sociedade civil), o modelo 5 foi obtido. Ele foi selecionado pelo método *backward* de seleção e possui, dentre todos os modelos mencionados na tabela acima, o menor valor para os critérios AIC e SC.

Com o objetivo de testar a adequabilidade dos modelos, foi realizado o teste de

Hosmer e Lemeshow, além de calculada a curva ROC de cada um desses modelos. Assim, obteve-se os seguintes resultados para eles:

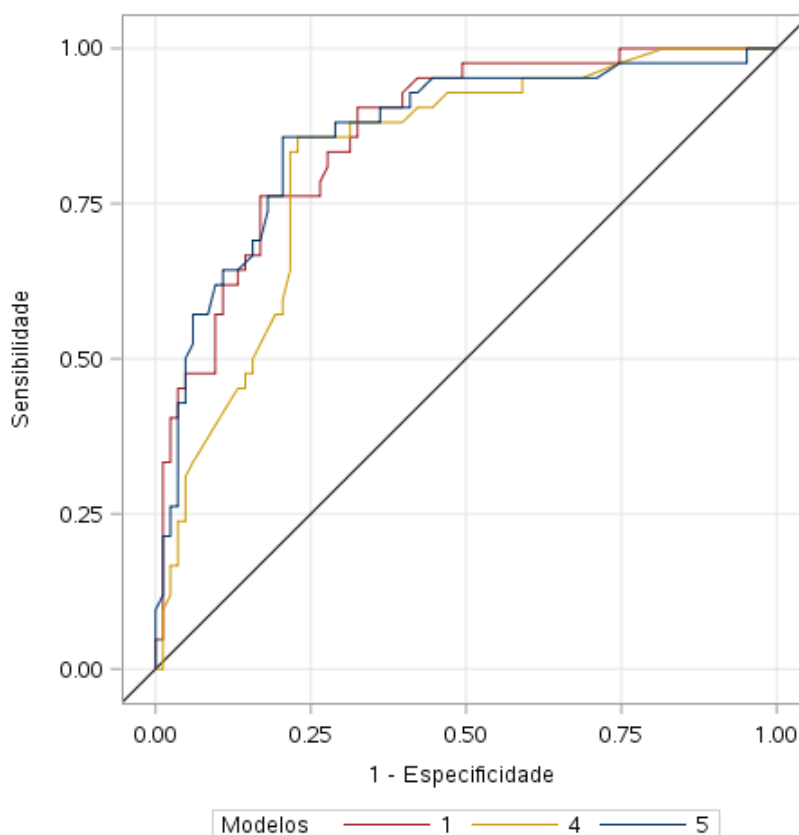
Quadro 3: Área do modelo da Curva ROC

Modelo	Curva ROC	P-valor
$x_5 + x_6 + x_8 + x_9 + x_{12} + x_8 * x_{12} + x_8 * x_6 + x_9 * x_{12}$	0,8639	0,4607
$x_5 + x_{10} + x_{12}$	0,8246	0,1205
$x_5 + x_{12}$	0,7423	0,0902
$x_5 + x_{12} + x_{14} + x_{14} * x_{12}$	0,8193	0,1366
$x_5 + x_{12} + x_{23} + x_{25} + x_{23} * x_{12}$	0,8622	0,2719

Ao observar o p-valor do teste de Hosmer e Lemeshow, nota-se que para todos os modelos selecionados não existem evidências suficientes para se rejeitar a hipótese nula (o modelo se ajusta bem aos dados) com nível de significância de 5%. Além disso, em relação à curva ROC, sabe-se que quanto maior o seu valor, maior o poder de predição do modelo. No caso do modelo 1, observa-se o maior valor para essa curva, sendo ele de 0,8639. Dessa forma, esse modelo abrange 86,39% da área total possível, sendo ela 1. É importante ressaltar que a curva do modelo 5 possui valor muito próximo ao do modelo 1, sendo ele de 0,8622.

A fim de se ter uma melhor visualização da curva ROC de cada modelo com interação e perceber suas diferenças, gerou-se o gráfico abaixo.

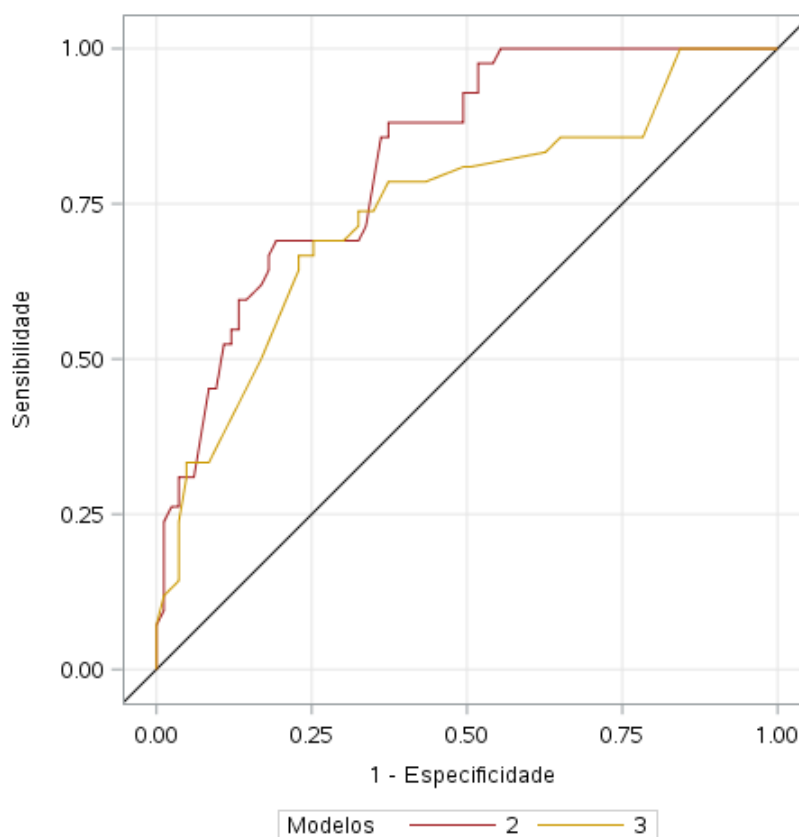
Figura 5: Curva Roc dos modelos de resultado para as interações



Como já mencionado, a curva ROC mostra o poder de predição de cada modelo, de modo que, quanto maior esse poder de predição, maior a curva ROC. Dessa forma, pode-se perceber pela figura acima que nenhum dos três modelos apresenta maior curva ROC em todos os pontos do gráfico. Além disso, pode-se ver mais explicitamente que os modelos 1 e 5 (curvas vermelha e azul, respectivamente) apresentam curva ROC muito próximas, sendo eles então os modelos com maior poder de predição. Já para o modelo 4 a visualização não é tão clara, mas pode-se ver que sua curva não está tão próxima das dos outros modelos.

Para os modelos sem interação também foi gerado o gráfico de suas respectivas curvas ROC. Por meio da figura abaixo é evidente que a curva ROC do modelo 2 abrange uma área maior que a do modelo 3. Assim, esse modelo possui maior curva ROC e, conseqüentemente, maior poder de predição.

Figura 6: Curva Roc dos modelos de resultado sem interação



Para a medida Leverage, valores acima de 0,5 indicam um índice alto de valores discrepantes e valores entre 0,2 e 0,5 indicam um índice moderado. Assim, no modelo 1 quatro pontos apresentaram Leverage acima de 0,5. Já em relação ao DFbeta, esse modelo apresentou 1 valor influente no coeficiente x_9 e 1 outro influente nos coeficientes das interações e no x_{12} . Para o modelo 2 apenas um valor apresentou Leverage acima de 0,5 e nenhuma observação mostrou-se ser influente. Já para o modelo 3, nenhum valor influente ou discrepante foi detectado, com todas as observações apresentando Leverage abaixo de 0,2. O modelo 4 apresentou duas observações acima de 0,2 para a medida Leverage (nenhum deles acima de 0,5) e, em relação à medida DFbeta, uma observação apresentou DFbeta de 0,99 para o coeficiente x_{12} e próximo de -2 para a interação $x_{12} * x_{14}$. Por fim, o modelo 5 apresentou quatro valores com Leverage acima de 0,2, sendo três deles valores acima de 0,5. Além disso, detectou-se a presença de dois valores influentes para esse modelo, um relacionado ao coeficiente de x_{12} e outro à interação $x_{12} * x_{23}$.

5.3.1 Modelo sem interação

Após todas as análises realizadas, optou-se por escolher, dentre os modelos apresentados na Tabela 5, o melhor modelo com e sem interação. Dessa forma, os modelos 5 (com interação) e 2 (sem interação) mostraram-se ser os mais adequados para explicar a variável *resultado*. Para o modelo 2, obteve-se as seguintes estimativas:

Tabela 6: Análise das estimativas de máxima verossimilhança do modelo 2

Variável	Estimativa	Erro Padrão	Teste de Wald	
			Estatística	P-valor
Intercepto	0,7885	1,1811	0,4456	0,5044
Tempo de tramitação	0,2006	0,0703	8,1426	0,0043
Tipo de requerente 4	-2,0199	1,6450	1,5077	0,2195
Tipo de requerente 5	-1,7416	1,3459	1,6744	0,1957
Tipo de requerente 6	2,0431	1,2642	2,6116	0,1061
Tipo de requerente 7	-15,0299	650,6	0,0005	0,9816
Tipo de requerente 8	-1,5459	1,3109	1,3907	0,2383
Tipo de requerente 9	-0,4278	1,1873	0,1298	0,7186
PGR 1	-3,9127	0,9296	17,7166	<,0001
PGR 2	-1,6818	0,6747	6,2131	0,0127

Ao observar a tabela acima, pode-se ver que, ao nível de significância de 5%, não se rejeita a hipótese do efeito de todos os tipos de requerente não serem significativos, ou seja, **nenhum tipo influencia no resultado**. Porém mesmo que não sejam significativos é importante mantê-los no modelo como variáveis de controle, uma vez que os critérios de informação se mostraram melhores quando *tipo de requerente* está no modelo.

Outro ponto para se observar refere-se ao **sinal das estimativas dos parâmetros**. Assim, pode-se ver que a variável *PGR* e todas as categorias de *tipo de requerente*, menos a categoria 6 (Procurador-Geral da República) possuem um efeito negativo sobre a variável resposta *resultado*. Por fim, um fato importante refere-se ao alto erro padrão e p-valor para o teste de Wald para o *tipo de requerente* 7. Isso pode se dever ao fato de que nesse nível da variável *tipo de requerente* apenas uma categoria de resposta da variável *resultado* fazer parte da amostra.

Como dito na seção anterior, a interpretação dos parâmetros deve ser feita em cima de suas razões de chances, ou *odds ratio*. Assim, abaixo encontram-se as estimativas para a razão de chances de cada variável e, no caso de variáveis dummies, a esti-

mativa para cada categoria, além de seus respectivos intervalos de Wald de 95% de confiança.

Tabela 7: Estimativa da razão de chances do modelo 2

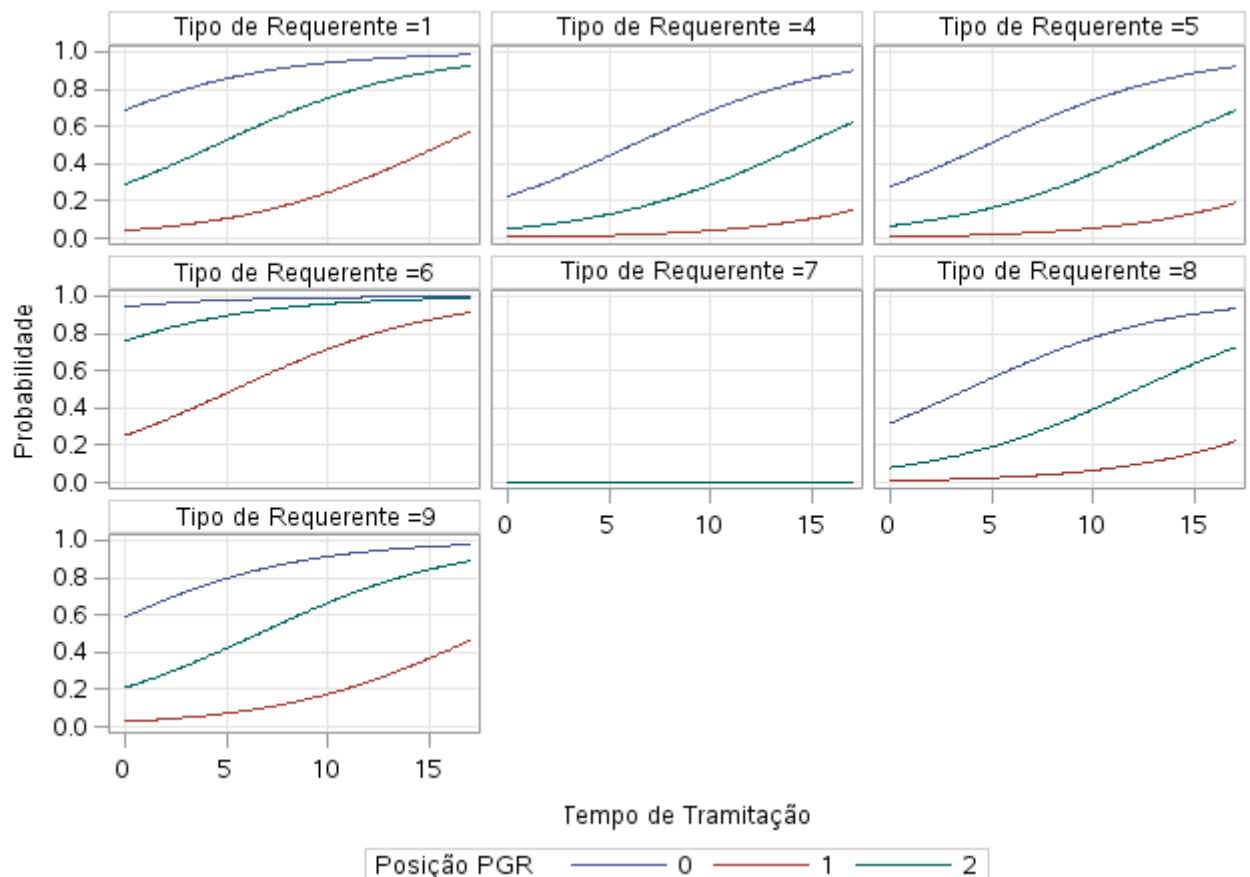
Efeito	Estimativa Pontual	Intervalo de 95% de Confiança de Wald	
Tempo de tramitação	1,222	1,065	1,403
Tipo de requerente 4	0,133	0,005	3,334
Tipo de requerente 5	0,175	0,013	2,451
Tipo de requerente 6	7,714	0,647	91,923
Tipo de requerente 7	<,001	<,001	>999,999
Tipo de requerente 8	0,213	0,016	2,783
Tipo de requerente 9	0,652	0,064	6,681
PGR 1	0,020	0,003	0,124
PGR 2	0,186	0,050	0,698

Antes da interpretação da *odds ratio* é importante lembrar que a categoria de referência da variável resposta *resultado* é 0 (improcedente) e, assim, as interpretações são feitas em cima dela. Nota-se que o intervalo de confiança para a razão de chance de todos os tipos de requerente possuem o número 1, o que indica novamente que não influenciam no resultado. As estimativas da categoria 7 da variável *tipo de requerente* não são passíveis de se calcular e interpretar. Isso se deve ao fato de que nela apenas a categoria "procedente/parcialmente procedente" da variável *resultado* fazer parte da amostra. Ou seja, nela, 100% do *tipo de requerente 7* tem resposta *resultado* = *procedente/parcialmente procedente*, não apresentando nenhuma observação para a categoria complementar "improcedente".

Além disso, a chance de *resultado* improcedente aumenta entre 1,065 e 1,403 vezes, ou entre 6,5% e 40,3%, a cada aumento de um ano de *tempo de tramitação*, mantidas constantes as demais variáveis. Já para a posição do PGR a chance de *resultado* improcedente diminui entre 87,6% e 99,7% quando a posição passa de Improcedente(0) para Procedente(1), e diminui entre 30,2% e 95,0% quando passa de Improcedente(0) para Parcialmente Procedente (2).

Com o objetivo de se analisar visualmente o efeito de cada variável do modelo selecionado sem interação, gerou-se o gráfico abaixo. Ele mostra a probabilidade predita de sucesso (*resultado* = improcedente) em função do *tempo de tramitação* para cada *posição PGR* quando o *tipo de requerente* é o 9.

Figura 7: Gráfico de probabilidade predita para *resultado* = improcedente



A figura acima apresenta a probabilidade do *resultado* ser "improcedente" em relação ao *Tempo de Tramitação* para cada *Tipo de Requerente*. Nota-se que o comportamento dos gráficos não é exatamente o mesmo para cada tipo, com isso é possível ver a importância da variável x_{10} no modelo mesmo com nenhum nível significativo. Nota-se que a probabilidade de sucesso quando a *posição PGR* é "improcedente" (0) é sempre maior que "procedente" (1), que por sua vez é maior que "parcialmente procedente" (2) para todos os requerentes. O comportamento diferenciado do gráfico para o tipo igual a 7 se dá pelo fato de se ter apenas 1 observação para cada posição de PGR na amostra.

5.3.2 Modelo com interação

Como mencionado, optou-se por escolher o melhor modelo com e sem interação. Dessa forma, os modelos 5 (com interação) e 2 (sem interação) mostraram-se ser os mais adequados. Para o modelo com interação obteve-se as seguintes estimativas:

Tabela 8: Análise das estimativas de máxima verossimilhança do modelo 5

Variável	Estimativa	Erro Padrão	Teste de Wald	
			Estatística	P-valor
Intercepto	-0,0494	0,3911	0,0159	0,8996
Tempo de tramitação	0,1899	0,0681	7,7798	0,0053
PGR 1	-4,0004	0,8880	20,2968	<,0001
PGR 2	-1,8810	0,6647	8,0081	0,0047
Diferença 1	0,1566	0,1544	1,0285	0,3105
Diferença 3	-0,4547	0,1303	12,1697	0,0005
Diferença 1 * PGR 1	-1,7228	0,4365	15,5766	<,0001
Diferença 1 * PGR 2	0,2637	0,2388	1,2195	0,2695

Primeiramente, é importante ressaltar que a variável *diferença 1* na tabela acima refere-se à diferença de *AC corporativo* favoráveis e desfavoráveis e *diferença 3* refere-se à diferença de *AC órgão/agente público* favoráveis e desfavoráveis. Pode-se observar pela tabela acima que essa primeira variável possui um efeito positivo sobre a variável *resultado*, enquanto a segunda um efeito negativo. Nota-se que para a variável *diferença 1* o teste não apresentou evidências para rejeitar a hipótese de efeito nulo, porém devido à interação com *PGR 1* significativa ela deve permanecer no modelo. As demais variáveis sem interação se mostraram significativas pois apresentaram p-valor para o teste de Wald inferior ao nível de significância de 5%.

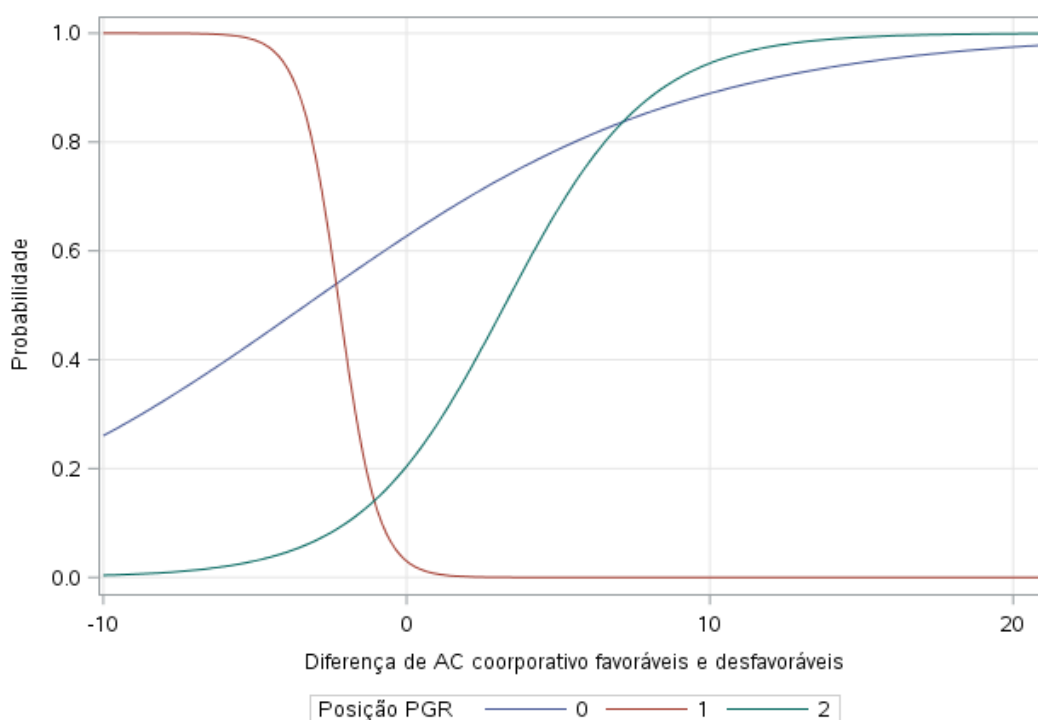
Para uma melhor interpretação dos parâmetros, a tabela abaixo fornece os valores da razão de chances das variáveis sem interação, além de seus respectivos intervalos de confiança. As variáveis com interação possuem interpretação da razão de chance mais complexa, sendo assim elas serão interpretadas pela figura 8. Vale ressaltar que uma vez que uma variável faz parte de uma interação sua interpretação não faz mais sentido individualmente.

Tabela 9: Estimativa da razão de chances do modelo 5

Efeito	Estimativa Pontual	Intervalo de 95% de Confiança de Wald	
Tempo de tramitação	1,209	1,058	1,382
Diferença 3	0,635	0,492	0,819

Mantendo-se as demais variáveis constantes, tem-se que a chance de *resultado improcedente* aumenta entre 5,8% e 38,2% vezes a cada aumento de um ano de *tempo de tramitação*. Além disso, a cada aumento de uma unidade na diferença entre *Amicus Curiae órgão/agente público* favoráveis e desfavoráveis a chance de *resultado improcedente* diminui entre 18,1% e 50,8%, mantendo-se constantes as demais variáveis.

Figura 8: Gráfico de probabilidade predita para *resultado = improcedente*



Dif. AC órgão/agente pub = 0,128 ; Tempo de Tramitação = 3,312

Para permitir a realização do gráfico acima as variáveis diferença 3 e tempo de tramitação foram fixadas em um valor médio. Pode-se observar que a curva correspondente a PGR procedente possui um comportamento discrepante, em que quando o número de *AC órgão/agente público* desfavoráveis é superior ao de favoráveis a probabilidade de *resultado improcedente* é um, porém quando o número de favoráveis é maior tal probabilidade é próxima de zero. Isso provavelmente se deve pelo falta de

observações suficientes para estimar os coeficientes, com a presença de apenas uma das respostas em cada grupo.

Além disso nota-se que a probabilidade de *resultado* improcedente quando a *PGR* também é improcedente cresce lentamente a medida que a diferença 1 aumenta. Já quando a posição da *PGR* é parcialmente procedente a probabilidade de *resultado* improcedente é próxima de zero quando o número de *AC órgão/agente público* desfavoráveis é superior (diferença negativa). A medida que tal diferença passa a ficar positiva e crescente, mais *AC's* em questão favoráveis a probabilidade de resultado improcedente cresce, ultrapassando a curva referente à posição da *PGR* improcedente quando a diferença está em torno de 7 ou 8.

6 Conclusão

Todos os três modelos propostos possuem as mesmas variáveis de controle tanto para variável resposta *cita no voto* quanto para *resultado*. Porém, os modelos para a variável dependente *resultado* possuem também *cita no voto* como variável de controle. Vale lembrar que as categorias de referência para as variáveis resposta *cita no voto* e *resultado* são, respectivamente, "não menciona" e "improcedente".

Após todas as análises necessárias, nenhuma interação foi significativa para explicar a variável *cita no voto*. Assim, o modelo que se mostrou ser o mais adequado para essa variável resposta é aquele que possui como variáveis explicativas *tipo de requerente* e *Amicus Curie representante da sociedade civil desfavoráveis*.

Já para a variável resposta *resultado*, algumas interações foram significativas. Assim, optou-se por escolher o melhor modelo com e sem interação. Dessa forma, o primeiro modelo obtido tem como variáveis explicativas *tempo de tramitação*, *posição PGR* e *tipo de requerente*. Já o segundo modelo contém *tempo de tramitação*, *posição PGR*, *diferença 1*, *diferença 3* e a interação entre as variáveis *diferença 1* e *PGR*. É importante lembrar que *diferença 1* refere-se à diferença de *Amicus Curiae* *coorporativo* favoráveis e desfavoráveis e *diferença 3* à diferença entre *AC órgão/agente público* favoráveis e desfavoráveis.