

Modelo de Regressão Multinomial Ordinal com Chances Proporcionais: Uma Aplicação a Avaliação do Presidente

André F. B. Menezes

Universidade Estadual de Maringá, Departamento de Estatística, Brasil

4 de Julho de 2018

1 Introdução

2 O modelo

3 Resultados

4 Referências

Objetivos

- ▶ Apresentar uma revisão do modelo de regressão multinomial ordinal;
- ▶ Determinar quais os fatores que influenciam a opinião do eleitor em relação a avaliação do presidente Michel Temer no seu 1 ano e 4 meses de governo.

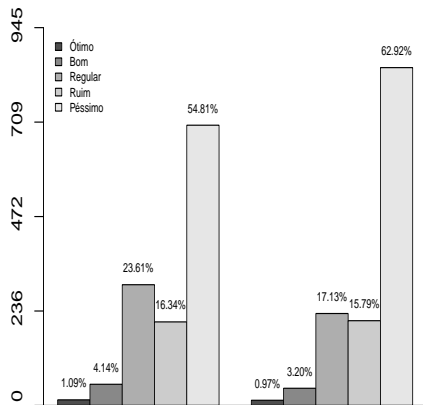
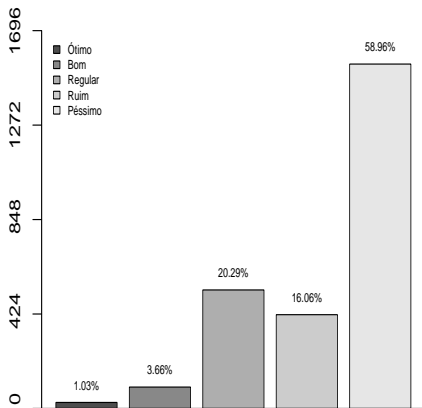
Os dados

- ▶ Fonte: pesquisa de opinião do Datafolha
- ▶ Ano: 2018
- ▶ Número de observações: 2772
- ▶ Número de variáveis: 124
- ▶ Estudo transversal

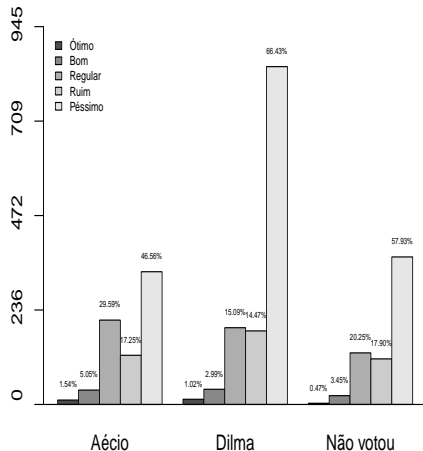
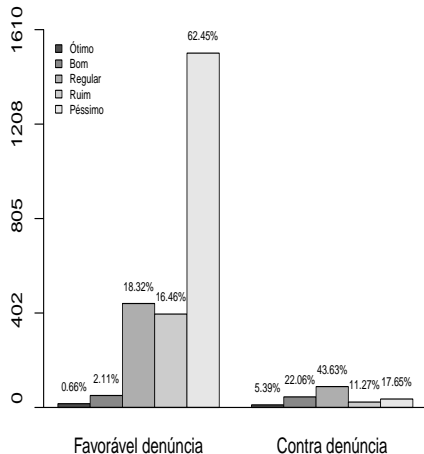
Variáveis selecionadas

- ▶ *Y*: “Na sua opinião o presidente Michel Temer está fazendo um governo: (1) Ótimo, (2) Bom, (3) Regular, (4) Ruim, (5) Péssimo”
- ▶ SEXO: Sexo do indivíduo: (1) Masculino, (2) Feminino
- ▶ DENUNCIA: “Na sua opinião, os deputados federais deveriam ou não autorizar a segunda denúncia do Ministério Público contra o presidente Michel Temer? (1) Sim, (2) Não”
- ▶ VOTOU2014: “Em quem você votou no segundo turno da eleição para presidente em 2014? (1) Aécio, (2) Dilma, (3) Branco/Nulo/Não votou”

Análise exploratória



Análise exploratória



O modelo

- ▶ Considere uma variável aleatória Y politômica ordinal com J categorias.
- ▶ Assume-se que Y tem distribuição multinomial com probabilidades $\{p_1, \dots, p_J\}$, satisfazendo $\sum_{j=1}^J p_j = 1$.
- ▶ Para introduzir o modelo de regressão multinomial basta associar os parâmetros $\{p_1, \dots, p_J\}$ as variáveis explicativas $\mathbf{x} = (x_1, \dots, x_p)^\top$.
- ▶ Denotaremos por $p_j(\mathbf{x})$ a probabilidade de ocorrência da categoria j com $j = 1, \dots, J$, para um dado vetor \mathbf{x} de variáveis explicativas tal que $\sum_{j=1}^J p_j(\mathbf{x}) = 1$.

- Além disso, considere as seguintes probabilidades acumuladas, definidas por

$$\theta_1(\mathbf{x}) = p_1(\mathbf{x}) = \Pr(Y \leq 1 \mid \mathbf{x})$$

$$\theta_2(\mathbf{x}) = p_1(\mathbf{x}) + p_2(\mathbf{x}) = \Pr(Y \leq 2 \mid \mathbf{x})$$

$$\vdots$$

$$\theta_J(\mathbf{x}) = p_1(\mathbf{x}) + p_2(\mathbf{x}) + \dots + p_J(\mathbf{x}) = \Pr(Y \leq J \mid \mathbf{x})$$

tal que $\theta_1(\mathbf{x}) \leq \theta_2(\mathbf{x}) \leq \dots \leq \theta_J = 1$.

- Desse modo, o modelo de regressão multinomial ordinal ou mais conhecido como modelo logitos cumulativos é expresso por

$$\log \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \log \left[\frac{\Pr(Y \leq j \mid \mathbf{x})}{\Pr(Y > j \mid \mathbf{x})} \right] \quad (1)$$

com $j = 1, \dots, J - 1$.

- ▶ A partir de (1) três principais modelos logitos cumulativos foram propostos na literatura.
- ▶ Modelo com chances não proporcionais proposto por Williams e Grizzle (1972);
- ▶ Modelo com chances proporcionais proposto por McCullagh (1980);
- ▶ Modelo com chances proporcionais parciais proposto por Peterson e Harrel (1990).

- Assume-se que os efeitos das covariáveis \mathbf{x} diferem entre os $J - 1$ logitos cumulativos. O modelo é definido por

$$\log \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \log \left[\frac{\Pr(Y \leq j \mid \mathbf{x})}{\Pr(Y > j \mid \mathbf{x})} \right] = \beta_{0j} + \beta_j \mathbf{x}, \quad (2)$$

em que $j = 1, \dots, J - 1$ e $\beta_j = (\beta_{1j}, \dots, \beta_{pj})$ são os coeficientes da regressão tal que β_{kj} descreve, para o logito j , o efeito da covariável k .

- Assume que $\beta_j = \beta$ para todo j . O modelo é expresso por

$$\log \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \log \left[\frac{\Pr(Y \leq j \mid \mathbf{x})}{\Pr(Y > j \mid \mathbf{x})} \right] = \beta_{0j} + \beta \mathbf{x}, \quad (3)$$

em que $j = 1, \dots, J - 1$ e $\beta = (\beta_1, \dots, \beta_p)$ são os coeficientes da regressão.

Modelo com chances proporcionais

- Em termos das probabilidades acumuladas o modelo pode ser definido por

$$\theta_j(\mathbf{x}) = \frac{\exp(\beta_{0j} + \boldsymbol{\beta} \mathbf{x})}{1 + \exp(\beta_{0j} + \boldsymbol{\beta} \mathbf{x})}, \quad j = 1, \dots, J - 1. \quad (4)$$

- As probabilidades em cada classe podem ser obtidas a partir da seguinte relação

$$p_j(\mathbf{x}) = \theta_j(\mathbf{x}) - \theta_{j-1}(\mathbf{x}), \quad j = 1, \dots, J, \quad (5)$$

em que $\theta_0(\mathbf{x}) = 0$.

Estimação via máxima verossimilhança

- ▶ Considerando que $\mathbf{y} = (y_{1j}, \dots, y_{nj}), j = 1, \dots, J$ seja uma amostra aleatória de tamanho n de uma distribuição Multinomial com J categorias ordinais.
- ▶ Assumindo o modelo de chances proporcionais (3), a função de verossimilhança é dada por

$$\mathcal{L}(\boldsymbol{\nu} \mid \mathbf{y}, \mathbf{x}) = \prod_{i=1}^n \left\{ \prod_{j=1}^J [p_j(\mathbf{x}_i)]^{y_{ij}} \right\} = \prod_{i=1}^n \left\{ \prod_{j=1}^J [\theta_j(\mathbf{x}_i) - \theta_{j-1}(\mathbf{x}_i)]^{y_{ij}} \right\}$$

em que $\boldsymbol{\nu} = (\beta_{0j}, \boldsymbol{\beta})$ e $y_{ij} = 1$ se a resposta da observação i esta na categoria j e $y_{ij} = 0$, caso contrário, com $\sum_{j=1}^J y_{ij} = 1$.

- ▶ A função log-verossimilhança pode ser escrita por

$$\ell(\boldsymbol{\nu} \mid \mathbf{y}, \mathbf{x}) = \sum_{i=1}^n \left\{ \sum_{j=1}^J y_{ij} \log [\theta_j(\mathbf{x}_i) - \theta_{j-1}(\mathbf{x}_i)] \right\}. \quad (6)$$

- ▶ Os estimadores de máxima verossimilhança são obtidos pela maximização da função (6).
- ▶ No presente trabalho a *procedure* LOGISTIC do software **SAS** foi empregada para estimação dos parâmetros.

Suposição de chances proporcionais

- ▶ Para avaliar se a suposição de chances proporcionais do modelo esta condizente com os dados é necessário testar a hipótese $\mathcal{H}_0 : \beta = \beta_j$ para todo j .
- ▶ A estatística mais utilizada para esta finalidade é a estatística escore definida por

$$S_R = \mathbf{U}(\beta_0)^\top \mathbf{I}^{-1}(\beta_0) \mathbf{U}(\beta_0) \sim \chi_{p-q}^2 \quad (7)$$

em que p e q são o número de parâmetros dos modelos sem e com chances proporcionais, respectivamente.

- ▶ Quando a hipótese nula for rejeitada para uma parte das covariáveis então o modelo de chances proporcionais parciais pode ser uma alternativa viável.

- Seja \mathbf{x}_i a i -ésima combinação de categorias das covariáveis, com $i = 1, \dots, s$ para o qual se tem n_{1+} observações, em que n_{ij} denota a frequência observada para a categoria j , $j = 1, \dots, J$, tal que

$$\sum_{j=1}^J n_{ij} = n_{i+}.$$

- As estatísticas de Pearson e Deviance são definidas por

$$Q_P = \sum_{i=1}^s \sum_{j=1}^J \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad \text{e} \quad Q_L = 2 \sum_{i=1}^s \sum_{j=1}^J n_{ij} \log \left(\frac{n_{ij}}{e_{ij}} \right),$$

sendo $e_{ij} = n_{i+} \times \hat{p}_j(\mathbf{x}_i)$ é a frequência esperada sob a hipótese nula de que o modelo está bem ajustado.

- ▶ Quando o modelo é corretamente especificado ambas as estatísticas Q_P e Q_L tem distribuição assintótica qui-quadrado com $g.l. = (J - 1)(s - 1) - q$, sendo J o número de categorias da variável resposta Y , s o número de subpopulações e q o número de parâmetros do modelo associado as covariáveis.
- ▶ Na presença de variáveis contínuas as estatísticas Q_P e Q_L não possuem um bom comportamento Agresti (2007). Nesse contexto, Lipsitz et al. (1996) introduziram uma generalização da estatística de Hosmer-Lemeshow.

Resultados

- Considerando que a variável resposta Y (avaliação do governo Temer) é composta por $r = 5$ categorias, os quatro logitos cumulativos ficam definidos por

$$\log \left[\frac{\theta_1(\mathbf{x})}{1 - \theta_1(\mathbf{x})} \right] = \log \left[\frac{p_1(\mathbf{x})}{p_2(\mathbf{x}) + p_3(\mathbf{x}) + p_4(\mathbf{x}) + p_5(\mathbf{x})} \right],$$

$$\log \left[\frac{\theta_2(\mathbf{x})}{1 - \theta_2(\mathbf{x})} \right] = \log \left[\frac{p_1(\mathbf{x}) + p_2(\mathbf{x})}{p_3(\mathbf{x}) + p_4(\mathbf{x}) + p_5(\mathbf{x})} \right],$$

$$\log \left[\frac{\theta_3(\mathbf{x})}{1 - \theta_3(\mathbf{x})} \right] = \log \left[\frac{p_1(\mathbf{x}) + p_2(\mathbf{x}) + p_3(\mathbf{x})}{p_4(\mathbf{x}) + p_5(\mathbf{x})} \right],$$

$$\log \left[\frac{\theta_4(\mathbf{x})}{1 - \theta_4(\mathbf{x})} \right] = \log \left[\frac{p_1(\mathbf{x}) + p_2(\mathbf{x}) + p_3(\mathbf{x}) + p_4(\mathbf{x})}{p_5(\mathbf{x})} \right].$$

- Sendo que $\mathbf{x} = (x_1, x_2, x_3, x_4)^\top$ é um dado vetor de covariáveis X_1 (1 se sexo masculino e 0 se feminino), X_2 (1 se votou no Aécio, 0 se votou na Dilma ou não votou), X_3 (1 se não votou, 0 se votou na Dilma ou Aécio) e X_4 (1 se é favorável a denuncia e 0 se é contra).
- É importante destacar que no modelo as probabilidades acumuladas são acumuladas sobre os valores inferiores, sendo o menor valor inferior o nível de avaliação ótimo.

Inicialmente, foram considerados os seguintes modelos

$$m_1 : \log \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} + \beta_{1j} x_1 + \beta_{2j} x_2 + \beta_{3j} x_3 + \beta_{4j} x_4,$$

e

$$m_2 : \log \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4,$$

com $j = 1, 2, 3, 4$.

Tabela: Comparação entre os modelos m_1 e m_2 .

Modelo	AIC	BIC	S_R	valor-p
m_1	5445.0751	5562.5091	20.5415	0.0575
m_2	5442.2803	5489.2539		

Tabela: Seleção das covariáveis para o modelo m_2 .

	G.L.	TRV	valor- p	AIC	BIC
Nulo	—	—	—	6028.4309	6052.0615
X_1	1	17.1873	<0.0001	6013.2436	6042.7818
$X_2, X_3 \mid X_1$	1	521.5128	<0.0001	5493.7308	5528.9609
$X_4 \mid X_1, X_2, X_3$	2	55.4504	<0.0001	5442.2803	5489.2539
$X_1 * X_2 * X_3 * X_4 \mid X_1, X_2, X_3, X_4$	7	9.4453	0.2222	5446.8350	5534.9104

Tabela: Resumos inferências do modelo m_2 .

Parâmetro	Estimativa	E.P.	Wald χ^2	valor- p
β_{01} (ótimo)	-3.3746	0.2295	216.2090	<0.0001
β_{02} (bom)	-1.7414	0.1580	121.4214	<0.0001
β_{03} (regular)	0.4190	0.1534	7.4612	0.0063
β_{04} (ruim)	1.2381	0.1552	63.6629	<0.0001
β_1 (masculino)	0.2850	0.0787	13.1224	0.0003
β_2 (aécio)	0.6893	0.0929	55.0789	<0.0001
β_3 (não votou)	0.3479	0.0982	12.5496	0.0004
β_4 (favorável)	-2.1629	0.1443	224.6461	<0.0001

- Os valores obtidos para as estatísticas de Pearson e Deviance foram $Q_P = 50.6142$ e $Q_L = 50.6458$ com graus de liberdade iguais a 20 e o valor- p para ambas as estatística foi aproximadamente 0.12.

Desse modo tem-se que o modelo estimado pode ser expresso em termos dos logitos cumulativos por

$$\log \left[\frac{\hat{\theta}_j(\mathbf{x})}{1 - \hat{\theta}_j(\mathbf{x})} \right] = \hat{\beta}_{0j} + 0.2650 x_1 + 0.6893 x_2 + 0.3479 x_3 - 2.1629 x_4,$$

ou ainda, em termos das probabilidades acumuladas, por

$$\hat{\theta}_j(\mathbf{x}) = \frac{\exp \left(\hat{\beta}_{0j} + 0.2650 x_1 + 0.6893 x_2 + 0.3479 x_3 - 2.1629 x_4 \right)}{1 + \exp \left(\hat{\beta}_{0j} + 0.2650 x_1 + 0.6893 x_2 + 0.3479 x_3 - 2.1629 x_4 \right)},$$

para $j = 1, 2, 3, 4$.

As probabilidades empíricas e previstas pelo modelo em cada nível da variável resposta foram comparadas calculando a raiz quadrada do erro quadrático médio, definida por

$$\text{REQM} = \sqrt{\sum_{i=1}^{12} [\tilde{p}_j(\mathbf{x}_i) - \hat{p}_j(\mathbf{x}_i)]^2}, \quad j = 1, 2, \dots, 5$$

em que $\tilde{p}_j(\mathbf{x}_i)$ denota a probabilidade empírica e $\hat{p}_j(\mathbf{x}_i)$ representa a probabilidade prevista pelo modelo para o i -ésimo nível das covariáveis.

Resultados

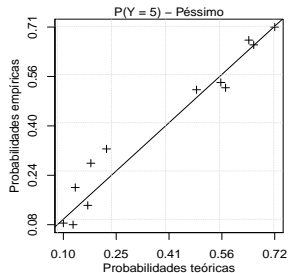
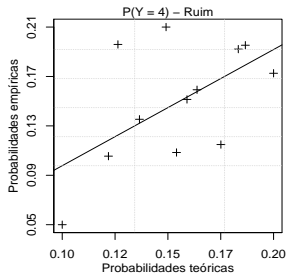
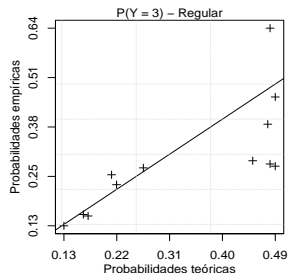
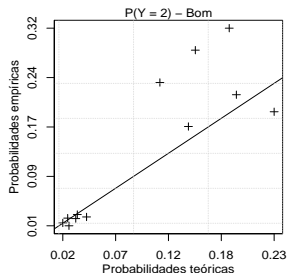
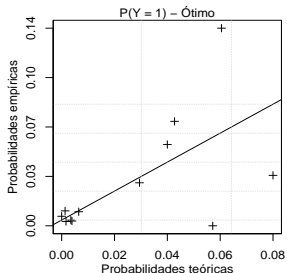


Tabela: Probabilidades empíricas $\tilde{p}_j(\mathbf{x})$ e preditas $\hat{p}_j(\mathbf{x})$ para dado valores das covariáveis.

X_1	X_2	X_3	$\tilde{p}_1(\mathbf{x})$	$\hat{p}_1(\mathbf{x})$	$\tilde{p}_2(\mathbf{x})$	$\hat{p}_2(\mathbf{x})$	$\tilde{p}_3(\mathbf{x})$	$\hat{p}_3(\mathbf{x})$
1	1	1	0.0096	0.0103	0.0256	0.0404	0.2756	0.2659
1	1	2	0.0102	0.0052	0.0238	0.0209	0.1548	0.1625
1	1	3	0.0038	0.0074	0.0231	0.0292	0.2577	0.2111
1	2	1	0.0345	0.0831	0.1897	0.2340	0.6379	0.4840
1	2	2	0.0556	0.0435	0.1667	0.1455	0.3889	0.4801
1	2	3	0.0000	0.0606	0.3200	0.1876	0.2800	0.4930
2	1	1	0.0033	0.0078	0.0294	0.0308	0.2320	0.2197
2	1	2	0.0065	0.0039	0.0163	0.0158	0.1254	0.1291
2	1	3	0.0030	0.0055	0.0118	0.0222	0.1509	0.1707
2	2	1	0.1351	0.0638	0.2162	0.1950	0.4595	0.4930
2	2	2	0.0294	0.0331	0.2353	0.1160	0.2941	0.4541
2	2	3	0.0714	0.0462	0.2857	0.1526	0.2857	0.4840
REQM			0.1095		0.2296		0.3825	

X_1 : 1 masculino e 2 feminino; X_2 : 1 favorável denuncia e 2 não favorável denúncia; X_3 : 1 votou Aécio, 2 votou na Dilma e 3 não votou.

Tabela: continuação...

X_1	X_2	X_3	$\tilde{p}_4(\mathbf{x})$	$\hat{p}_4(\mathbf{x})$	$\tilde{p}_5(\mathbf{x})$	$\hat{p}_5(\mathbf{x})$
1	1	1	0.1763	0.1958	0.5128	0.4876
1	1	2	0.1548	0.1567	0.6565	0.6547
1	1	3	0.1962	0.1799	0.5192	0.5725
1	2	1	0.0517	0.1002	0.0862	0.0986
1	2	2	0.1111	0.1519	0.2778	0.1790
1	2	3	0.2000	0.1254	0.2000	0.1334
2	1	1	0.1993	0.1831	0.5359	0.5586
2	1	2	0.1384	0.1352	0.7134	0.7160
2	1	3	0.1627	0.1612	0.6716	0.6403
2	2	1	0.1081	0.1211	0.0811	0.1270
2	2	2	0.1176	0.1720	0.3235	0.2248
2	2	3	0.2143	0.1472	0.1429	0.1699
REQM			0.1347		0.1786	

Tabela: Estimativas pontuais e intervalares da razão de chances.

Efeito	$\exp(\hat{\beta})$	I.C. 95%
Masculino vs Feminino	1.3297	(1.1397, 1.5514)
Aécio vs Dilma	1.9923	(1.6607, 2.3900)
Não votou vs Dilma	1.4161	(1.1682, 1.7168)
Favorável vs Não favorável	0.1150	(0.0867, 0.1526)

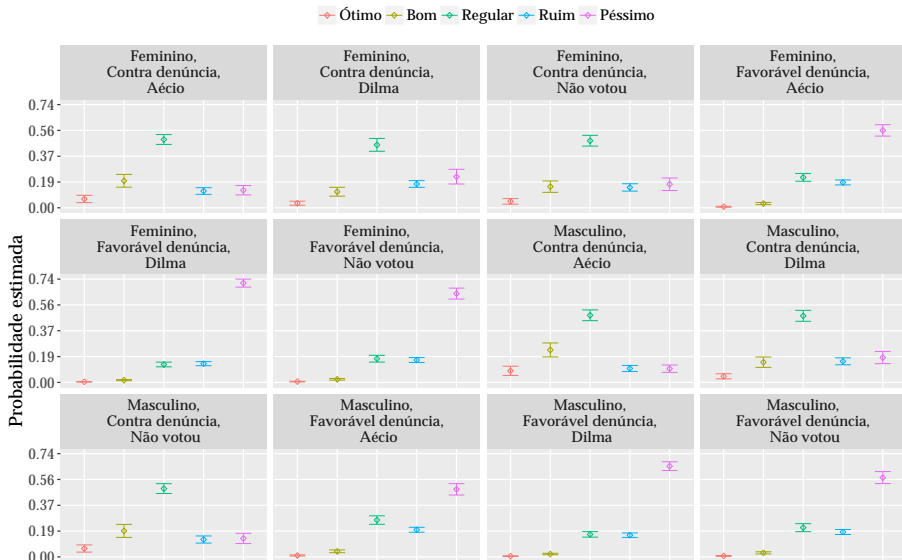
Interpretação $\exp(\hat{\beta}_2)$

- Para um nível fixo j da avaliação do governo Temer, a chance estimada de uma pessoa que votou no Aécio em 2014 avaliar o governo Temer na direção ótima (i.e., $Y \leq j$ ao invés de $Y > j$) é aproximadamente 2 vezes a chance de uma pessoa que votou na Dilma em 2014. Portanto, existe uma associação substancial das pessoas que votaram no Aécio tenderem a acharem o governo Temer mais ótimo do que as pessoas que votaram na Dilma;

Interpretação $\exp(\hat{\beta}_4)$

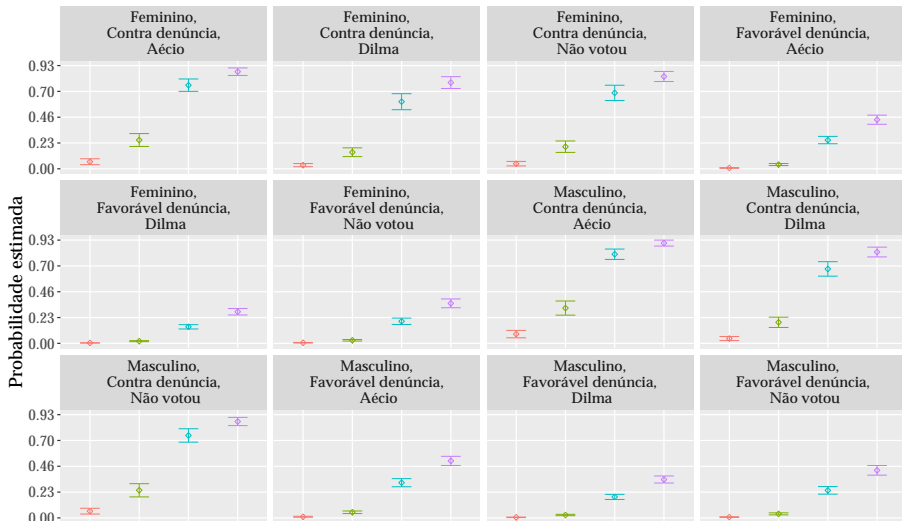
- Para um nível fixo j da avaliação do governo Temer, a chance estimada de uma pessoa a qual acredita que os deputados deveriam autorizar a denúncia contra o Temer avaliar o governo na direção ótima (i.e., $Y \leq j$ ao invés de $Y > j$) é aproximadamente 0.11 vezes a chance de uma pessoa a qual acredita que os deputados não deveriam autorizar a denúncia. Isso significa que as pessoas que são favoráveis a denúncia contra o Temer tender a avaliar o governo na direção Péssimo comparada com as pessoas que não são favoráveis a denúncia.

Resultados



Resultados

◆ Ótimo
 ◆ Bom\Ótimo
 ◆ Bom\Ótimo\Regular
 ◆ Bom\Ótimo\Regular\Ruim



- [1] Agresti, A., 1990. **Categorical Data Analysis**. Wiley: New York.
- [2] Agresti, A., 2007. **An Introduction to Categorical Data Analysis**. John Wiley & Sons, Inc: New Jersey.
- [3] Derr, B., 2013. **Ordinal response modeling with the LOGISTIC procedure**. In: Proceedings of the SAS Global Forum 2013 Conference. Paper 446–2013.
- [4] Giolo, S. R., 2017. **Introdução à análise de dados categóricos com aplicações**. Edgard Blucher Ltda.

- [5] High, R., 2013. **Models for ordinal response data.** In: Proceedings of the SAS Global Forum 2013 Conference. Paper 445–2013.
- [6] McCullagh, P., 1980. **Regression models for ordinal data.** Journal of the Royal Statistical Society. Series B (Methodological) 42 (2), 109–142.
- [7] Peterson, B., Harrell, F. E., 1990. **Partial proportional odds models for ordinal response variables.** Journal of the Royal Statistical Society. Series C (Applied Statistics) 39 (2), 205–217.
- [8] Williams, O. D., Grizzle, J. E., 1972. **Analysis of contingency tables having ordered response categories.** Journal of the American Statistical Association 67 (337), 55–63.

Muito obrigado!