Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Condition-Number-Based Criterion: Critério para Seleção da Matriz de Correlação de Trabalho

Marcelo dos Santos[1]
IFBA, Paulo Afonso, BA
Fernanda de Bastiani[2]
UFPE, Recife, PE
Miguel A. Uribe-Opazo[3]
UNIOESTE, Cascavel, PR
Manuel Galea[4]
UC, Santiago, CHL

Resumo. O método das Equações de Estimação Generalizada proporciona ótimos resultados para estimativas em modelo regressão. No entanto, o desafio na aplicação deste método está em uma boa especificação para matriz de correlação de trabalho. Particularmente, em análise de dados espaciais, a matriz de correlação espacial. Nesta conjuntura, a escolha de uma matriz de correlação espacial adequada é imprescindível, pois uma matriz inadequada poderá gerar estimativas pouco confiáveis haja vista o aumento do viés de estimação. Diante disso, neste estudo foi proposto um novo método de seleção da matriz de correlação de trabalho, baseado no condicionamento da matriz de variância-covariância *naive*. A avaliação da performance do método foi realizada por um estudo de simulações, utilizando as distribuições marginais da Normal, Poisson e Gama para dados espacialmente correlacionados.

Palavras-chave. Equações de Estimação Generalizadas, Condicionamento, Variância-Covariância.

1 Introdução

As Equações de Estimação Generalizada (EEG's) permitem modelar de forma explícita a estrutura da correlação de dados, sendo assim possível ajustar a variabilidade entre as observações além de fornecer estimativas consistentes dos parâmetros de regressão. Contudo, apesar de ser amplamente utilizada, a abordagem das EEG's ainda tem sido pouco explorada no contexto da análise de dados espacialmente correlacionados. Sua aplicação requer, a *priori*, que seja especificada uma matriz de correlação, comumente nomeada de correlação de trabalho. Esta matriz é geralmente desconhecida e representada por $\mathbf{R}(\boldsymbol{\varphi})$, em que $\boldsymbol{\varphi}$, também desconhecido, é um vetor paramétrico de correlação que caracteriza completamente a matriz $\mathbf{R}$. De modo geral, as EEG's tem como ponto forte gerar estimativas consistentes para os parâmetros de regressão, mesmo que a matriz de correlação de trabalho não seja especificada corretamente, ver [11]. No entanto, uma especificação da matriz de correlação de trabalho que se aproxima da verdadeira matriz correlação que gerou os dados, poderá melhorar a eficiência dos estimadores dos parâmetros, sobretudo em pequenas amostras.

[1]marcelo.santos@ifba.edu.br
[2]fernanda.bastiani@ufpe.br
[3]mopazo@uol.com.br
[4]mgalea@mat.uc.cl

Os critérios os mais populares para escolha da melhor matriz de correlação de trabalho são *Rotnizky and Jewell's criterion* (RJC), *quasi-likelihood under the independence model criterion* (QIC), e *Critério de Informação de Correlação* (CIC), propostos por [1, 8], [5] e [2], respectivamente.

Estes critérios, bem como alguns outros não tão populares, foram estabelecidos no contexto da análise de dados longitudinais. Ademais, pouco se conhece sobre o desempenho relativo, de cada critério, com relação a seleção de uma estrutura da matriz de correlação de trabalho ótima para melhorar a eficiência dos estimadores de regressão. Quando os dados são georreferenciados com uma estrutura de correlação gerada a partir de algum modelo de semivariograma, nada se sabe sobre o desempenho destes critérios.

Desta forma, neste estudo, propomos um novo critério, ver [9], para auxiliar na seleção da matriz de correlação de trabalho espacial em EEG, com base no número de condição da matriz de variâncias-covariâncias *naive*. Particularmente, esta proposta é direcionada a análise de dados espacialmente correlacionados. Para avaliar o desempenho do critério proposto, realizamos um extenso estudo de simulação e comparamos os resultados com os critérios RJC, QIC e CIC.

2 Equações de Estimação Generalizada no Contexto de Dados Espacialmente Correlacionados

A análise de dados espacialmente correlacionados, a abordagem por EEG inicia-se definindo um vetor $\mathbf{Z} = (Z(\mathbf{s}_1) = z_1, \ldots, Z(\mathbf{s}_n) = z_n)^\top$ de variáveis respostas observadas em diferentes localizações $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathbf{S} \subset \mathbb{R}^2$. Então, a partir da proposta de [3] a matriz de covariância espacial é definida por $\mathbf{V} = \mathbf{A}^{1/2}\mathbf{R}(\boldsymbol{\varphi})\mathbf{A}^{1/2}$, desde que se assuma um estimador consistente para os parâmetros de correlação espacial $\boldsymbol{\varphi}$, e $\mathbf{A} = diag(c(\sigma_1^2, \ldots, \sigma_n^2))$ é uma matriz de variâncias dos dados com c definindo a dispersão. Aqui, considera-se $c = 1$. Desta forma, as estimativas de quasi-verossimilhança, para os parâmetros do modelo de regressão, são obtidas pela solução da função escore

$$\mathbf{U} = \mathbf{D}^\top \mathbf{V}^{-1}(\mathbf{Z} - \boldsymbol{\mu}), \tag{1}$$

em que o vetor de médias tem elementos $\mu_i = g(\mathbf{x}_i^\top \boldsymbol{\beta})$, para alguma função $g(\cdot)$ monótona, diferenciável e invertível. E $\mathbf{x}$ é um vetor de ordem $p \times 1$ que representam as entradas das linhas da matriz de covariáveis e, $\boldsymbol{\beta}$ de ordem $p \times 1$ é um vetor de parâmetros do modelo de regressão. $\mathbf{D}$ é uma matriz de derivadas da média com relação a $\boldsymbol{\beta}$.

Devido a forma não fechada da função escore, sua solução é realizada por algum método iterativo. Então, dado um estimador consistente para os parâmetros da estrutura de correlação, um estimador consistente para os parâmetros de regressão, disponível na literatura como por exemplo nos trabalhos de [6], [10], [4] e [7] é dado por:

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} + \left[\left(\hat{\mathbf{D}}^{(m)}\right)^\top \left(\hat{\mathbf{V}}^{(m)}\right)^{-1} \left(\hat{\mathbf{D}}^{(m)}\right)\right]^{-1} \left(\hat{\mathbf{D}}^{(m)}\right)^\top \left(\hat{\mathbf{V}}^{(m)}\right)^{-1} \left(\mathbf{Z} - \hat{\boldsymbol{\mu}}^{(m)}\right). \tag{2}$$

Com isso, um estimador robusto (ou *sanduich*) e consistente para variância é:

$$\hat{\mathbf{V}}_{rob} = \left(\hat{\mathbf{D}}^\top \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}}\right)^{-1} \left(\hat{\mathbf{D}}^\top \hat{\mathbf{V}}^{-1} Cov(\mathbf{Z}) \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}}\right) \left(\hat{\mathbf{D}}^\top \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}}\right)^{-1}, \tag{3}$$

em que $Cov(\mathbf{Z}) = (\mathbf{Z} - \hat{\boldsymbol{\mu}})(\mathbf{Z} - \hat{\boldsymbol{\mu}})^\top$. Note que, se a matriz de trabalho coincide ou é próxima da verdadeira matriz de correlação de trabalho, então, $\hat{\mathbf{V}}$ aproxima-se ou coincide com $Cov(\mathbf{Z})$. Neste caso, o estimador de variância-covariâncias robusto reduz-se a um estimador ótimo, comumente nomeado de *naive*, e dado por:

$$\hat{\mathbf{V}}_N = \left(\hat{\mathbf{D}}^\top \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}}\right)^{-1}. \tag{4}$$

Portanto, qualquer que seja a matriz de correlação de trabalho, dado um vetor não-nulo $\mathbf{x}$, temos que

$$\mathbf{x}^\top(\hat{\mathbf{V}}_{rob} - \hat{\mathbf{V}}_N)\mathbf{x} \geq 0, \tag{5}$$

isto é, a diferença entre os estimadores robusto e *naive* resulta em uma matriz definida não-negativa. Logo, uma matriz de correlação de trabalho mal especificada poderá resultar em uma significativa perda de eficiência. Mas, este problema poderá ser controlado pelo condicionamento da matriz $\hat{\mathbf{V}}_N$, haja vista sua influência no controle da propagação dos erros de arredondamento na Equação iterativa (2).

3 Condition-Number-Based Criterion

O condicionamento ou número de condição matricial está fortemente relacionado ao conceito de eficiência numérica, sobretudo em problemas que envolvem inversão matricial. Para este estudo, nos motivamos no fato de que o estimador robusto depende da inversão de termos equivalentes ao estimador $\hat{\mathbf{V}}_N$. Neste caso, se $\hat{\mathbf{V}}_N$ é mal condicionada, então, $\hat{\mathbf{V}}_{rob}$ tende a ser mal condicionado. Ainda, com relação a eficiência numérica do estimador descrito na Equação (2), o processo iterativo depende da inversão matricial repetidas vezes até a convergência. Então, o mal condicionamento da matriz $\hat{\mathbf{V}}_N$ poderá prejudicar a eficiência numérica do estimador $\hat{\boldsymbol{\beta}}$, podendo ainda ser dilatada pelos erros de arredondamento e aproximação numérica.

Dentre os critérios mais populares estão o *quasi-likelihood under the independence model criterion* (QIC), *Critério de Informação de Correlação* (CIC) e *Rotnizky and Jewell's criterion* (RJC). No entanto, estes critérios levam em consideração apenas os elementos da diagonal principal do produto das matrizes de variâncias-covariâncias, obtidas pelo estimador robusto e sob independência ou robusto e *naive*. Ademais, de acordo com [1] quando se utiliza o critério RJC, sob a especificação de uma correta matriz de correlação, o produto $\hat{\mathbf{V}}_{Naive}^{-1} \times \hat{\mathbf{V}}_{rob}$ se aproxima de uma matriz identidade. Contudo, os elementos fora da diagonal se aproximam de zero, mas não necessariamente são iguais a zero. Da mesma forma, os elementos da diagonal não são necessariamente iguais a um. Diante disso, sente-se a necessidade de um critério que carregue toda a informação presente nas matrizes variâncias-covariâncias bem como da matriz de covariância espacial.

Nota-se que os critérios para seleção da matriz correlação de trabalho dependem do estimador de variâncias robusto, que por sua vez depende do processo de inversão matricial. Portanto, é importante considerar como os erros de arredondamento se propagam, ou seja, avaliar como pequenas variações nos elementos de $\mathbf{V}$ e $\mathbf{V}^{-1}$ influenciam nos resultados da estatística de seleção. Diante disso, a proposta aqui, é estabelecer um critério baseado em uma modificação nas expressões do critério RJC, introduzindo a norma e número de condição das matrizes $\hat{\mathbf{V}}_{rob}$ e $\hat{\mathbf{V}}_{Naive}$, respectivamente. Desta maneira, espera-se capturar toda informação trazida da matriz de correlação espacial para estimativas das matrizes de variâncias-covariâncias.

A introdução do número de condição de uma matriz tem como objetivo avaliar a sensibilidade da matriz de variâncias-covariâncias. O mal condicionamento de uma matriz está relacionado a quase singularidade da matriz e, sua definição é facilmente encontrada nos principais textos sobre análise numérica como sendo:

$$Cond(\mathbf{B}) = \|\mathbf{B}\| \left\|\mathbf{B}^{-1}\right\|, \tag{6}$$

em que $\|\cdot\| : \mathbb{R}^n \longrightarrow \mathbb{R}$ define uma norma matricial induzida por uma norma vetorial, aqui a norma euclidiana. Portanto, a grandeza do número de condição da matriz $\mathbf{B}$ revela a sensibilidade da solução do problema as pertubações na matriz de correlação de trabalho.

Se o modelo marginal de regressão e a matriz de correlação de trabalho estão corretamente especificados, então, estimador robusto reduz-se ao estimador *naive*. Com isso, aplicando a norma em ambos os lados da equação iterativa para estimativas dos parâmetros de regressão, temos que

DOI: 10.5540/03.2025.011.01.0368

$$\left\|\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)}\right\| = \left\|\left[\left(\hat{\mathbf{D}}^{(m)}\right)^\top \left(\hat{\mathbf{V}}^{(m)}\right)^{-1} \left(\hat{\mathbf{D}}^{(m)}\right)\right]^{-1} \left(\hat{\mathbf{D}}^{(m)}\right)^\top \left(\hat{\mathbf{V}}^{(m)}\right)^{-1} \left(\mathbf{Z} - \hat{\boldsymbol{\mu}}^{(m)}\right)\right\| \quad (7)$$

$$\leq \left\|\hat{\mathbf{V}}_{Naive}\right\| \left\|\left(\hat{\mathbf{D}}^{(m)}\right)^\top\right\| \left\|\left(\hat{\mathbf{V}}^{(m)}\right)^{-1}\right\| \left\|\mathbf{Z} - \hat{\boldsymbol{\mu}}^{(m)}\right\|. \quad (8)$$

Então, como $\hat{\boldsymbol{\beta}}^{(m+1)}$ é solução do sistema de EEG, podemos afirmar que

$$\hat{\boldsymbol{\beta}}^{(m+1)} > \left(\hat{\mathbf{D}}^{(m)}\right)^\top \left(\hat{\mathbf{V}}^{(m)}\right)^{-1} \left(\mathbf{Z} - \hat{\boldsymbol{\mu}}^{(m)}\right). \quad (9)$$

Portanto, utilizando uma manipulação algébrica temos que

$$\frac{\left\|\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)}\right\|}{\left\|\hat{\boldsymbol{\beta}}^{(m+1)}\right\|} \leq \frac{\left\|\hat{\mathbf{V}}_{Naive}\right\| \left\|\hat{\mathbf{V}}^{-1}_{Naive}\right\| \left\|\left(\hat{\mathbf{D}}^{(m)}\right)^\top\right\| \left\|\left(\hat{\mathbf{V}}^{(m)}\right)^{-1}\right\| \left\|\mathbf{Z} - \hat{\boldsymbol{\mu}}^{(m)}\right\|}{\left\|\hat{\mathbf{V}}^{-1}_{Naive}\right\| \left\|\left(\hat{\mathbf{D}}^{(m)}\right)^\top\right\| \left\|\left(\hat{\mathbf{V}}^{(m)}\right)^{-1}\right\| \left\|\mathbf{Z} - \hat{\boldsymbol{\mu}}^{(m)}\right\|} \quad (10)$$

$$= Cond(\hat{\mathbf{V}}_{Naive}) \frac{1}{\left\|\hat{\mathbf{V}}^{-1}_{Naive}\right\|}. \quad (11)$$

Desta forma, temos um majorante para o erro relativo das estimativas dos parâmetros do modelo marginal de regressão sob EEG. Portanto, o erro relativo é proporcional ao inverso da norma da matriz estimada para variâncias-covariâncias, sob o estimador *naive*. Neste caso, o número de condição é um fator de proporcionalidade, isto é, quantifica a sensibilidade para estimativas dos parâmetros de regressão. Portanto, se a norma da matriz de variâncias-covariâncias estimada for muito grande, então, uma pequena perturbação na matriz de correlação de trabalho ou na resposta, poderá significar em um grande erro relativo para estimativas dos β's.

Com isso, a proposta aqui é selecionar a matriz de correlação de trabalho espacial que minimiza Condition-Number-Based Criterion (CNBC) (12).

$$RJ_1M = \frac{\left\|\hat{\mathbf{V}}_{rob}\right\| Cond(\hat{\mathbf{V}}_{Naive})}{p},$$
$$RJ_2M = \frac{\left(\left\|\hat{\mathbf{V}}_{rob}\right\| Cond(\hat{\mathbf{V}}_{Naive})\right)^2}{p},$$
$$CNBC = \sqrt{(RJ_1M - 1)^2 + (RJ_2M - 1)^2}, \quad (12)$$

em que p é o número de covariáveis envolvidas no modelo marginal de regressão . Para avaliar o desempenho do critério $CNBC$ é realizado um estudo de simulação, no qual os resultados são comparados com os critérios QIC, CIC e RJC, utilizando modelos de correlação gerados a partir das famílias Wendland, Matérn, e o modelo Esférico.

4 Validação do Critério de Seleção de Correlação Espacial por Estudo de Simulação

Com objetivo de avaliar o desempenho do método proposto, realiza-se um estudo de simulação em termos da proporção de seleção de uma estrutura de correlação espacial verdadeira em comparação com os critérios QIC, CIC e RJC. A ideia é avaliar a taxa de acerto, ou seja, o percentual em que os critérios identificou a verdadeira matriz de correlação que gerou os dados. Para gerar os dados utilizamos três cenários, a saber: considerando a variância da Normal, Poisson e Gama para dados espacialmente correlacionados. Contudo, neste trabalho serão apresentados apenas os resultados para o caso da variância da Gama.

Para definir o modelo de regressão consideramos a função de ligação inversa $g(\mu_i) = \dfrac{1}{\mu_i}$. O parâmetro de regressão ($\boldsymbol{\beta}$) é considerado fixo e igual a $\beta_0 = 0,5$. As respostas são geradas sob estrutura de correlação espacial, considerando os modelos das famílias Wendland, Matérn e modelo Esférico. Consideramos uma suavização $\kappa = 1, 2$ e 3, e $\nu = 0,5, 1,5$ e $3,5$ para covariâncias da família Wendland e Matérn, respectivamente. O vetor de parâmetros que caracteriza a estrutura espacial é considerado fixo para os três casos, igual a $\boldsymbol{\varphi} = (\phi_1, \phi_2, \phi_3)^\top = (0,2; 0,6; 0,9)^\top$. São consideradas dois tipos de grades. Primeiramente consideramos uma grade de tamanhos 10×10 e 20×20, ambas com espaçamento de 1 (uma) unidade entre os pontos.

Nas Tabelas 1 e 2 estão os resultados para taxa de acertos da verdadeira matriz de trabalho, considerando resposta correlacionada da Gama. Notamos que os critérios CIC e QIC obtiveram desempenho razoáveis tanto para estrutura da Wendland quanto da Matérn e modelos Esférico, embora o QIC tenha obtido um desempenho muito ruim para o modelo Esférico considerando uma grade de tamanho 10×10. Os critério RJC e CNBC obtiveram excelentes resultados, com taxas de acertos superiores a 83% e 93%, respectivamente, em todas as configurações.

Tabela 1: Proporções (%) de acertos na seleção da estrutura de correlação espacial para respostas correlacionada da Gama. Os pontos são distribuídos em grades regulares de tamanho 10×10 e 20×20. Os dados são gerados alternando os modelos da família Wendland, Matérn e Esférico.

Estrutura Verdadeira	Critério	Especificação com $n = 100$			Especificação com $n = 400$		
		W - κ	M - $4,5$	Esf	W - κ	M - $4,5$	Esf
W - 1	CIC	25,9	35,8	38,3	32,0	31,2	36,8
	QIC	42,1	49,5	8,4	37,4	54,0	8,6
	RJC	**89,4**	2,7	7,9	**84,5**	8,5	7,0
	$CNBC$	**95,7**	1,8	2,5	**93,4**	4,0	2,6
W - 2	CIC	**34,1**	32,8	33,1	**35,5**	29,7	34,8
	QIC	41,5	51,3	7,2	36,7	55,4	7,9
	RJC	**89,2**	2,4	8,4	**85,3**	8,8	5,9
	$CNBC$	**95,6**	2,1	2,3	**93,7**	3,9	2,4
W - 3	CIC	**39,6**	30,0	30,4	32,5	33,1	34,4
	QIC	43,7	52,7	3,6	35,8	57,0	7,2
	RJC	**89,5**	3,1	7,4	**88,4**	6,6	5,5
	$CNBC$	**95,5**	2,0	2,5	**95,1**	3,7	1,2

CNBC: Condition-Number-Based Criterion. W - κ : Wendland com suavização ν.

Tabela 2: Proporções (%) de acertos na seleção da estrutura de correlação espacial para respostas correlacionada da Gama. Os pontos são distribuídos em grades regulares de tamanho 10×10 e 20×20. Os dados são gerados alternando os modelos da família Wendland, Matérn e Esférico.

Estrutura Verdadeira	Critério	Especificação com $n = 100$			Especificação com $n = 400$		
		W - κ	M - 4, 5	Esf	W - κ	M - 4, 5	Esf
		M - ν	W - 3	Esf	M - ν	W - 3	Esf
M - 0, 5	CIC	34,4	38,3	27,3	**40, 2**	36,6	23,2
	QIC	41,0	54,0	5,0	35,8	47,3	16,9
	RJC	**83, 6**	8,7	7,7	**84, 6**	9,0	6,4
	$CNBC$	**94, 2**	3,1	2,7	**94, 9**	3,3	1,8
M - 1, 5	CIC	23,7	34,7	41,6	32,7	31,1	36,2
	QIC	43,0	50,7	6,3	35,4	37,5	27,1
	RJC	**87, 6**	7,5	4,9	**84, 2**	9,4	6,4
	$CNBC$	**95, 1**	3,2	1,7	**94, 0**	3,9	2,1
M - 3, 5	CIC	30,2	37,7	32,1	32,1	28,9	39,0
	QIC	44,0	49,5	6,5	40,4	41,2	18,4
	RJC	**85, 7**	6,6	7,7	**89, 2**	5,5	5,3
	$CNBC$	**94, 7**	2,7	2,6	**96, 1**	1,8	2,1
		Esf	W - 3	M - 4, 5	Esf	W - 3	M - 4, 5
Esférico	CIC	31,0	43,3	25,7	35,9	40,8	23,3
	QIC	3,3	49,7	47,0	39,9	9,1	51,0
	RJC	**90, 4**	7,6	2,0	**89, 3**	5,1	5,6
	$CNBC$	**96, 0**	2,7	1,3	**95, 4**	1,6	3,0

CNBC: Condition-Number-Based Criterion. M - ν : Matérn com suavização κ.

Portanto, tanto o critério aqui proposto, quanto o RJC podem ser considerados para auxiliar na escolha da matriz de correlação espacial. Ademais, convém ressaltar que o critério CNBC leva em consideração os efeitos dos erros de arredondamento envolvidos no processo de inversão matricial, que estão presentes nos estimadores robusto e *naive*.

5 Considerações Finais

Neste trabalho, um novo critério, nomeado Condition-Number-Based Criterion (CNBC), foi desenvolvido com base no número de condição das matrizes de variâncias-covariância *naive*, para selecionar uma estrutura de correlação de trabalho espacial apropriada em GEE. Comparamos a nossa proposta com os critérios RJC, QIC e CIC, disponíveis na literatura, por meio de um extenso estudo de simulação.

Destacamos que o critério proposto permite que as informações, presentes na matriz de correlação espacial, contribuam para estimativa de todas as partes do modelo de regressão sob EEG, ao contrário dos outros critérios QIC, CIC e RJC. Ademais, neste trabalho focamos na seleção da matriz de correlação de trabalho sob a suposição de que o modelo de média está especificado corretamente. A especificação da estrutura de correlação é baseada em modelos de semivariogramas, utilizando as famílias Wendland, Matérn e modelo Esférico.

Referências

[1] L. Y. Hin, V. J. Carey e Y. G. Wang. "Criteria for working–correlation–structure selection in GEE: Assessment via simulation". Em: **The American Statistician** 61.4 (2007), pp. 360–364.

[2] L. Y. Hin e Y. G. Wang. "Working-correlation-structure identification in generalized estimating equations". Em: **Statistics in medicine** 28.4 (2009), pp. 642–658.

[3] K. Y. Liang e S. L. Zeger. "Longitudinal data analysis using generalized linear models". Em: **Biometrika** 73.1 (1986), pp. 13–22.

[4] D. T. Nava et al. "Local Influence for Spatially Correlated Binomial Data: An Application to the Spodoptera frugiperda Infestation in Corn". Em: **Journal of Agricultural, Biological and Environmental Statistics** 22.4 (2017), pp. 540–561.

[5] W. Pan. "Akaike's information criterion in generalized estimating equations". Em: **Biometrics** 57.1 (2001), pp. 120–125.

[6] A. S. Paul e L. M. McShane. "A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data". Em: **Biometrics** (1995), pp. 627–638.

[7] R. L. Prentice. "Correlated binary regression with covariates specific to each binary observation". Em: **Biometrics** (1988), pp. 1033–1048.

[8] A. Rotnitzky e N. P. Jewell. "Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data". Em: **Biometrika** 77.3 (1990), pp. 485–497.

[9] M. Santos et al. "Selection criterion of working correlation structure for spatially correlated data". Em: **The American Statistician** 77.3 (2023), pp. 283–291.

[10] M. K. Venezuela, D. B. Aparecida e M. C. Sandoval. "Diagnostic techniques in generalized estimating equations". Em: **Journal of Statistical Computation and Simulation** 77.10 (2007), pp. 879–888.

[11] M. K. Venezuela, M. C. Sandoval e D. A. Botter. "Local influence in estimating equations". Em: **Computational statistics & data analysis** 55.4 (2011), pp. 1867–1883.

DOI: 10.5540/03.2025.011.01.0368