

Modelos para dados dependentes e Equações de Estimação Generalizadas

Prof. Caio Azevedo

Classe estendida de modelos de quase-verossimilhança (MQV)

- O logaritmo da (função de) quase-verossimilhança $Q(\mu; y)$ ([aqui](#)) assume que a função $V(\cdot)$ é conhecida.
- Assim, a mudança dessa função significa que um novo modelo está sendo definido.
- Para permitir comparações de diferentes funções $V(\cdot)$ para um mesmo modelo, bem como possibilitar a obtenção de uma estimativa para o erro padrão assintótico de $\hat{\sigma}^2$, Nelder e Pregibon (1987) propuseram uma (log) quase-verossimilhança estendida, definida por (próximo slide)

Cont.

- (cont.)

$$Q^+(\mu, y) = -\frac{1}{2\sigma^2} D(y; \mu) - \frac{1}{2} \ln(2\pi\sigma^2 V(y)) ,$$

em que $D(y; \mu) = 2 \int_{\mu}^y \frac{y-t}{V(t)} dt$ é o quase-desvio e $\phi = \frac{1}{\sigma^2}$, o parâmetro de dispersão.

- Similarmente a Q , Q^+ não pressupõe que a distribuição completa de Y seja conhecida, mas somente os dois primeiros momentos.

Cont.

- A estimativa de β maximizando-se $Q^+(\mathbf{y}; \boldsymbol{\mu}) = \sum_{i=1}^n Q^+(y_i; \mu_i)$, para uma amostra aleatória de tamanho n , coincide com a estimativa de quase-verossimilhança para β (exercício), uma vez que Q^+ é uma função linear de Q .
- A estimativa de ϕ maximizando Q^+ é dada por

$$\hat{\phi} = \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{n}.$$

Cont.

- Portanto, para os casos especiais em que Q^+ corresponde às distribuições normal e normal inversa, $\hat{\phi}$ corresponde à estimativa de máxima verossimilhança de ϕ .
- Maiores detalhes podem ser encontrados em [Nelder e Pregibon \(1987\)](#) e [Paula \(2024\)](#).

Respostas correlacionadas

- Seja $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir_i})'$ o vetor resposta multivariado para a i -ésima unidade experimental, $i = 1, \dots, n$, e r_i representa a quantidade de medidas feitas para a unidade i (pode ser tempo, distância, número de variáveis dentre outras).
- Assumiremos, em princípio, que apenas é conhecida a distribuição marginal de Y_{it} , $t = 1, 2, \dots, r_i$, dada por

$$f(y_{it}) = \exp \{ \phi \{ y_{it} \theta_{it} - b(\theta_{it}) \} + c(y_{it}, \phi) \} \mathbb{1}_{A_{it}}(y_{it}),$$

em que $\mathcal{E}(Y_{it}) = \mu_{it} = b'(\theta_{it})$, $\mathcal{V}(Y_{it}) = \phi^{-1} V_{it}$, $V_{it} = d\mu_{it}/d\theta_{it}$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão, em geral desconhecido.

Respostas correlacionadas

- Podemos definir um MLG para cada índice t acrescentando à estrutura anterior a parte sistemática:

$$g(\mu_{it}) = \eta_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} = \sum_{j=1}^p X_{jit}\beta_j,$$

em que η_{it} é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{X}_{it} = (X_{1it}, \dots, X_{pit})'$ representa os valores de variáveis explicativas observadas para a i -ésima unidade experimental no tempo t e $g(\cdot)$ é uma função de ligação.

Respostas correlacionadas

- O procedimento de estimação, ignorando-se a estrutura de correlação, pode ser encontrado em [Paula \(2024\)](#).
- O estimador acima é consistente e assintoticamente normal, porém muitas vezes com perda de eficiência (eventualmente apresentando estimadas deflacionadas, para os respectivos erros-padrão).
- Uma forma de considerar a correlação intra-unidades é introduzi-la na Equação de Estimação obtida sob correlações nulas.

Respostas correlacionadas

- Inicialmente, vamos considerar que a matriz de variâncias e covariâncias é dada por

$$\text{Cov}(\mathbf{Y}_i) = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i \mathbf{V}_i^{1/2},$$

em que $\mathbf{V}_i = \text{diag}(V_{i1}, \dots, V_{ir_i})'$. Note que, então a matriz de correlação de \mathbf{Y}_i é dada por (exercício, veja aqui):

$$\text{Corre}(\mathbf{Y}_i) = \mathbf{R}_i.$$

- Usualmente, assume-se que $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\rho})$, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)'$. Isso faz com se tenha sempre q parâmetros para se estimar as correlações intra-unidades, independentemente de r_i e/ou de n .

Respostas correlacionadas

- Assim, os estimadores para β , sob esta abordagem, são obtidos através da resolução do seguinte sistema de equações:

$$S_{\beta}(\hat{\beta}_G) = \mathbf{0}_{(p \times 1)},$$

chamados de equações de estimação generalizadas, em que

$$S_{\beta}(\beta) = \sum_{i=1}^n \mathbf{D}_i' \Omega_i^{-1} (\mathbf{Y}_i - \mu_i), \quad (1)$$

$$\mathbf{D}_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{1/2} \mathbf{X}_i \text{ e } \Omega_i = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i(\rho) \mathbf{V}_i^{1/2}.$$

Respostas correlacionadas

- Se assumirmos $\mathbf{R}_i \equiv \mathbf{R}_i(\boldsymbol{\rho}) = \mathbf{I}_q, \forall i$, teremos a abordagem (anteriormente mencionada) na qual se negligencia as correlações intra unidades.
- Como o sistema de equações (Equação 1), não tem solução explícita, uma forma de obtermos as respectivas estimativas é através de um processo iterativo que é uma modificação do método escore de Fisher (próximo slide):

Respostas correlacionadas

■ (cont.)

$$\begin{aligned}\beta_G^{(m+1)} &= \beta_G^{(m)} + \left\{ \sum_{i=1} \mathbf{D}_i^{(m)'} \boldsymbol{\Omega}_i^{-1(m)} \mathbf{D}_i^{(m)} \right\}^{-1} \\ &\times \left\{ \sum_{i=1}^n \mathbf{D}_i^{(m)'} \boldsymbol{\Omega}_i^{-1(m)} \left(\mathbf{y}_i - \boldsymbol{\mu}^{(m)} \right) \right\}\end{aligned}\quad (2)$$

$m=0,1,2,\dots$, (até que algum critério de convergência seja obtido).

- As estimativas $\tilde{\phi}$ e $\tilde{\rho}$ são dadas inicialmente e modificadas separadamente a cada passo do processo iterativo.

Respostas correlacionadas

- Supondo que $\hat{\phi}$ e $\hat{\rho}$ são estimadores consistentes de ϕ e ρ , respectivamente, temos que

$$\sqrt{n} \left(\hat{\beta}_G - \beta \right) \xrightarrow[n \rightarrow \infty]{D} N(\mathbf{0}, \Sigma),$$

em que,

$$\begin{aligned} \Sigma = & \lim_{n \rightarrow \infty} \left\{ n \left(\sum_{i=1}^n \mathbf{D}_i' \Omega_i^{-1} \mathbf{D}_i \right)^{-1} \left[\sum_{i=1}^n \mathbf{D}_i' \Omega_i^{-1} \text{Cov}(\mathbf{Y}_i) \Omega_i \mathbf{D}_i \right] \right. \\ & \times \left. \left(\sum_{i=1}^n \mathbf{D}_i' \Omega_i^{-1} \mathbf{D}_i \right)^{-1} \right\} \end{aligned}$$

Respostas correlacionadas

- Se a matriz de correlação $\mathbf{R}_i(\rho)$ é definida “corretamente”, então um estimador consistente para $\text{Cov}(\hat{\beta}_G)$ é dado por $\mathbf{H}_1^{-1}(\hat{\beta}_G)$, em que

$$\mathbf{H}_1(\hat{\beta}_G) = \sum_{i=1}^n \left(\hat{\mathbf{D}}_i' \hat{\Omega}_i^{-1} \hat{\mathbf{D}}_i \right),$$

em que $\hat{\mathbf{D}}_i \equiv \hat{\mathbf{D}}_i(\hat{\beta}_G)$ e $\hat{\Omega}_i \equiv \hat{\Omega}_i(\hat{\beta}_G, \hat{\rho}, \hat{\phi})$.

- Entretanto, se a matriz de correlação de trabalho $\mathbf{R}_i(\rho)$ é definida “incorretamente”, $\mathbf{H}_1^{-1}(\hat{\beta}_G)$ pode ser inconsistente.

Respostas correlacionadas

- Um estimador robusto para $\text{Cov}(\hat{\beta}_G)$, sugerido por Liang and Zeger (1986) é dada por:

$$\hat{\mathbf{V}}_G = \mathbf{H}_1^{-1} \left(\hat{\beta}_G \right) \mathbf{H}_2 \left(\hat{\beta}_G \right) \mathbf{H}_1^{-1} \left(\hat{\beta}_G \right),$$

em que

$$\mathbf{H}_2 \left(\hat{\beta}_G \right) = \sum_{i=1}^n \left\{ \hat{\mathbf{D}}_i' \hat{\Omega}_i^{-1} (\mathbf{Y}_i - \hat{\mu}_i) (\mathbf{Y}_i - \hat{\mu}_i)' \hat{\Omega}_i^{-1} \hat{\mathbf{D}}_i \right\}.$$

- O estimador $\hat{\mathbf{V}}_G$ é consistente mesmo sob a má especificação de $\mathbf{R}(\rho_i)$.

Estruturas de correlação

- Existem várias opções de escolha para $\mathbf{R}_i(\rho)$ e uma vasta literatura a respeito.
- Note que, independentemente da escolha, $R_{ijj'} = 1, \forall j = j'$. Denote por $R_{ijj'} \equiv R_{jj'}$, o $j \times j'$ -ésimo elemento de $\mathbf{R}_i(\rho)$.
- Alguns exemplos, são: não estruturada, simétrica ou permutável, Autoregressiva (AR(1)).

Estruturas de correlação

- Não estruturada: quando $R_{jj'}$ não têm uma estrutura específica e podem assumir valores diferentes entre si.
- Simétrica ou permutável: $R_{jj'} = \rho, \forall j \neq j'$.
- AR(1): $R_{jj'} = \rho^{|j-j'|}, \forall j \neq j'$.
- Nos três casos acima os respectivos estimadores de ρ , têm fórmula analítica,

Estruturas de correlação

- Não estruturada: quando $R_{ijj'}$ não têm uma estrutura específica e podem assumir valores diferentes entre si.
- Simétrica ou permutável: $R_{ijj'} = \rho, \forall j \neq j'$.
- AR(1): $R_{ijj'} = \rho^{|j-j'|}, \forall j \neq j'$.
- Nos três casos acima os respectivos estimadores de ρ , têm fórmula analítica.

Estruturas de correlação

- O parâmetro de dispersão ϕ^{-1} pode ser estimado consistentemente por $\left(N = \sum_{i=1}^n r_i\right)$
$$\hat{\phi}^{-1} = \frac{1}{N - p} \sum_{i=1}^n \sum_{j=1}^{r_i} \frac{(Y_{ij} - \hat{\mu}_{ij})^2}{\hat{V}_{ij}}.$$
- Assim, o processo iterativo (Equação (2)) deve alternar com as estimativas para ρ e ϕ até a convergência.
- Em relação à métodos de diagnósticos, sugerimos a leitura de [Hardin and Hilbe \(2013\)](#), [Venezuela et al \(2007\)](#), [Venezuela et al \(2011\)](#) e [Paula \(2024\)](#).

Análise residual

- Em termos de análise residual podemos considerar um resíduo para cada componente de \mathbf{Y}_i ou um único envolvendo todo o vetor.
- Essencialmente, utilizaremos o seguinte resíduo de Pearson (para cada resposta):

$$R_{P_{ij}} = \frac{\mathbf{e}_{ij}' \widehat{\mathbf{A}}_i^{1/2} \left(\widehat{\mathbf{V}}_i \widehat{\mathbf{W}}_i \right)^{-1} (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)}{\sqrt{1 - \widehat{h}_{ijj}}},$$

(continua no próximo slide)

Análise residual

- (Cont.) em que $i = 1, \dots, n$, $j = 1, \dots, r_i$, em que $\mathbf{A}_i = \phi \mathbf{W}_i^{1/2} \mathbf{R}_i^{-1} \mathbf{W}_i^{1/2}$, \mathbf{e}'_{ij} é um vetor de dimensão $(1 \times r_i)$ de zeros, com o valor 1, na j -ésima posição e h_{ijj} é o j -ésimo elemento da diagonal principal de

$$\mathbf{H}_i = \mathbf{A}_i^{1/2} \mathbf{X}_i (\mathbf{X}' \mathbf{A} \mathbf{X})^{-1} \mathbf{X}'_i \mathbf{A}_i^{1/2},$$

em que $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$ tem dimensão $(N \times p)$ e

$\mathbf{A} = \text{diag} \{ \mathbf{A}_1, \dots, \mathbf{A}_n \}$ / tem dimensão $N \times N$ em que $N = \sum_{i=1}^n r_i$.

Análise no R

- Pacote *gee* função *gee*. Comando básico (modelo Poisson, ligação log, com estrutura de correlação permutável)
(com intercepto)

```
gee(Y~X1 + .... +Xn, family=poisson(link='log'),  
id=indiv, corstr='exchangeable')
```

(sem intercepto)

```
gee(Y~-1+X1 + .... +Xn, family=poisson(link='log'),  
id=indiv, corstr='exchangeable')
```

Exemplo 15: Ataques epiléticos (Diggle, Liang e Zeger (2002), Seção 8.4)

- Diz respeito aos resultados de um ensaio clínico com 59 indivíduos epiléticos os quais foram aleatorizados de modo que cada um recebesse uma droga antiepilética denominada progabide ou placebo.
- Os dados de cada indivíduo consistem em um número inicial de ataques epiléticos num período de oito semanas antes do tratamento, seguido do número de ataques em cada período de duas semanas, num total de quatro períodos, após o tratamento.

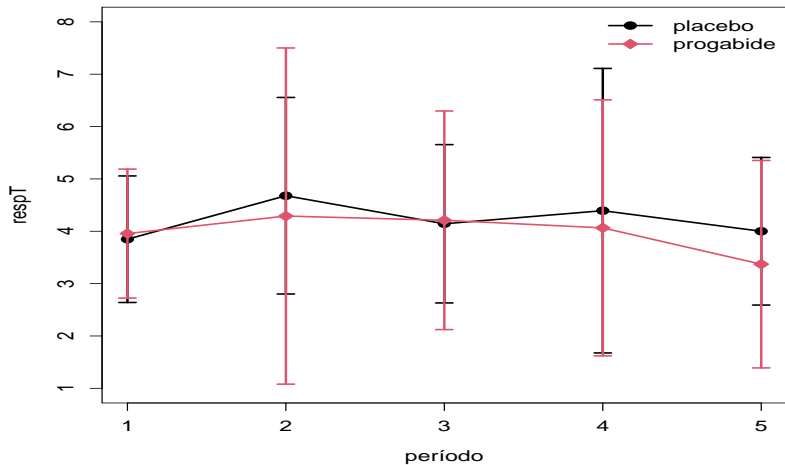
Cont.

- O interesse da pesquisa é saber se a droga reduz a taxa de ataques epilépticos.
- Estudo irregular: número de semanas varia (8 no primeiro período e 2 nos demais).
- Balanceado em relação a condição de avaliação e desbalanceado em relação ao grupo (28 - placebo e 31 - progabide).
- Completo.
- Como o número de semanas varia entre os períodos, utilizaremos, na análise descritiva: $\text{respT} = \frac{\text{número de ataques}}{\text{número de semanas}}$.

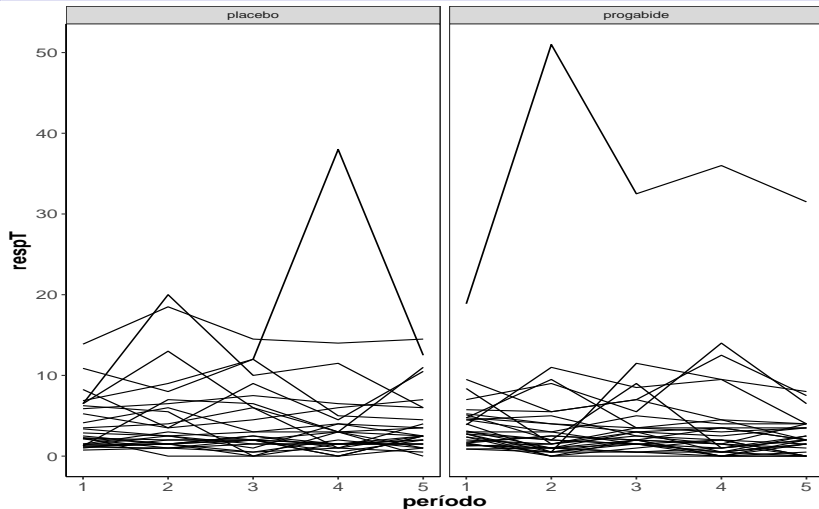
Medidas resumo

Grupo	Período	Media	DP	Var.	CV(%)	curt.	ca	Min.	Med.	Max.
pla.	1	3,85	3,26	10,65	84,79	4,64	1,44	0,75	2,38	13,88
	2	4,68	5,07	25,69	108,33	5,75	1,86	0,00	2,50	20,00
	3	4,14	4,08	16,66	98,53	3,13	1,12	0,00	2,25	14,50
	4	4,39	7,34	53,82	167,01	17,09	3,69	0,00	2,50	38,00
	5	4,00	3,81	14,48	95,14	4,12	1,47	0,00	2,50	14,50
prog.	1	3,96	3,50	12,24	88,46	11,82	2,71	0,88	3,00	18,88
	2	4,29	9,12	83,18	212,58	23,73	4,54	0,00	2,00	51,00
	3	4,21	5,93	35,16	140,86	17,99	3,71	0,00	2,50	32,50
	4	4,06	6,95	48,26	170,92	15,66	3,41	0,00	2,00	36,00
	5	3,37	5,63	31,67	166,93	21,57	4,21	0,00	2,00	31,50

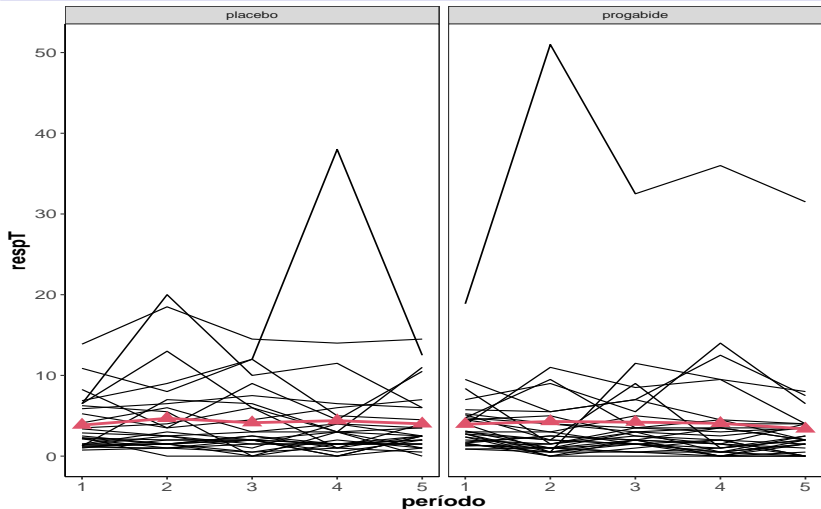
Perfis médios



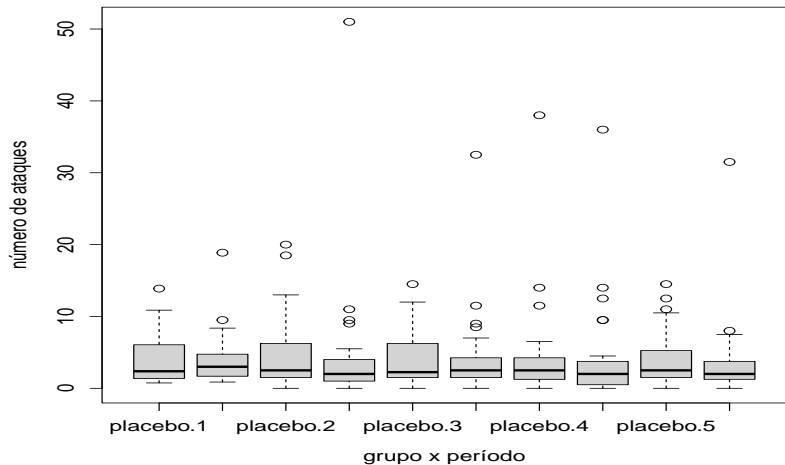
Perfis individuais



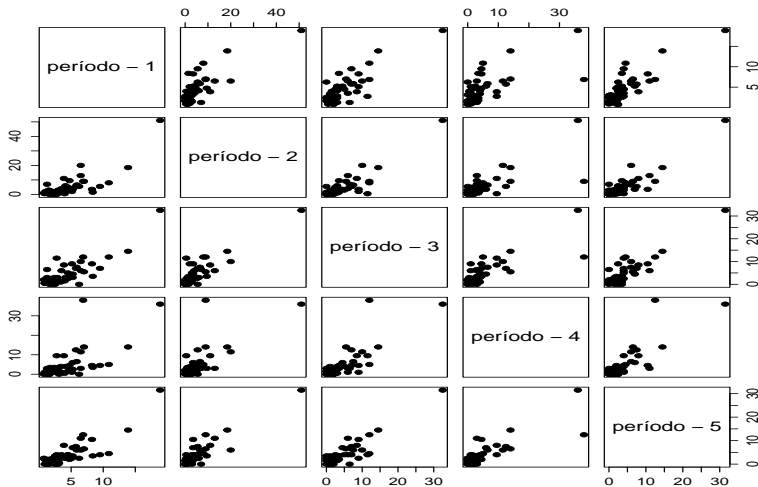
Perfis individuais e médios



Box plot



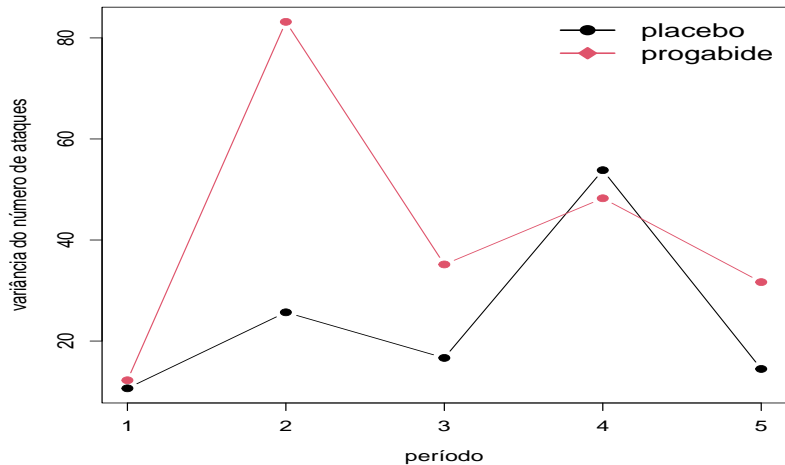
Matriz de diagramas de dispersão



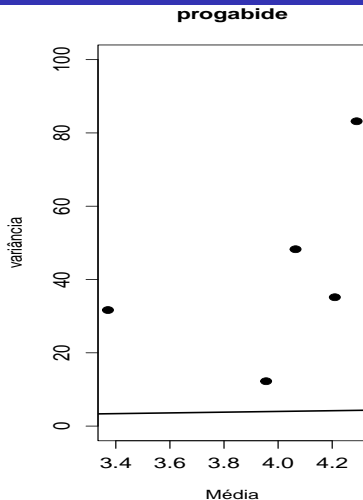
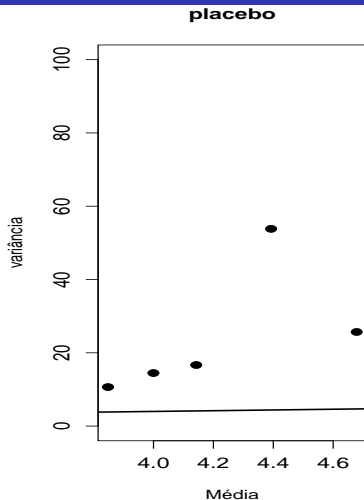
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo)

Período				
1	2	3	4	5
11,29	0,80	0,83	0,67	0,84
19,81	55,02	0,87	0,74	0,89
14,23	32,90	25,95	0,80	0,90
15,98	38,71	28,92	50,05	0,82
13,61	31,92	21,96	28,10	23,22

Variâncias em cada condição



Dispersão entre as médias e as variâncias amostrais



Modelo marginal de Poisson

$$Y_{ijk} \sim \text{Poisson}(\mu_{ijk}/t_k),$$

$$i = 1, 2, (\text{grupo1} - \text{placebo}, 2 - \text{progabide}),$$

$$j = 1, \dots, n_i (\text{indivíduo}, n_1 = 28; n_2 = 31),$$

$$k = 1, 2, 3, 4, 5 (\text{período}),$$

$$\ln(\mu_{ijk}) = \mu + \alpha_i + (\beta + \delta_i)(x_{ijk} - 1) + \ln(t_k),$$

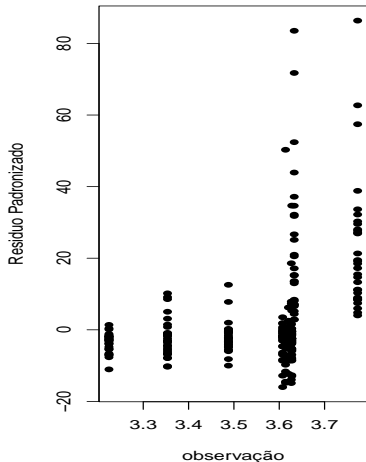
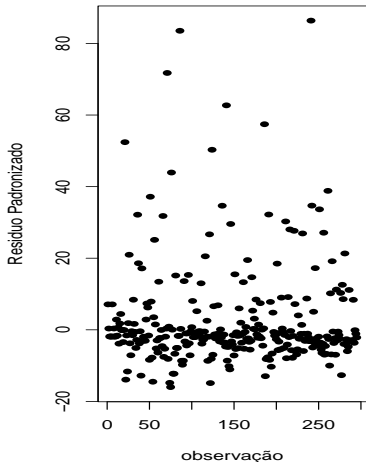
$$\alpha_1 = 0, \delta_1 = 0,$$

Matriz de correlação de trabalho: simétrica ou permutável.

Modelo marginal de Poisson

- Y_{ijk} : número de ataques do j -ésimo paciente, do i -ésimo grupo no k -ésimo período.
- $x_{ijk} = k$, ou seja, o valor do período no qual o número de ataques do indivíduo j do grupo i , no período k , fora medido.
- t_k : número de semanas relativas ao k -ésimo período.
- $\ln(t_k)$: “offset”.
- Exercício: apresentar pelo menos uma estimativa para cada parâmetro e/ou funções deles.

Diagnóstico do Resíduo de Pearson



Modelo marginal de Poisson

Parâmetro	Est.	EP	IC(95%)	Esta. Z	p-valor
μ	1,29	0,17	[0,97;1,62]	7,78	<0,0001
α_2	0,04	0,21	[-0,37;0,45]	0,18	0,8563
β	-0,00	0,05	[-0,10;0,09]	-0,04	0,9708
δ_2	-0,04	0,06	[-0,15;0,08]	-0,65	0,5143
ϕ^{-1}	22,55				
ρ	0,78				

Comentários

- O modelo, de acordo com o resíduo de Person, não se ajustou bem aos dados, apresenta uma variabilidade crescente, em função dos valores preditos e uma variância bem distante de 1. Além disso, há um número muito grande de resíduos candidatos a outliers.
- Sugestão: ajustar um modelo EEG usando como distribuições marginais, binomiais negativas.

Comentários

- (supondo que o modelo de ajustou bem) A interação não é significativa ($p=0,5143$). Dessa forma, podemos avaliar os p-valores dos efeitos de tempo e tratamento em separados os quais, também, não são significativos ($p=0,9708$ e $p=0,8563$), respectivamente. A estimativa de ϕ^{-1} reforça a existência de superdispersão nos dados.