

Análise de Variância Multivariada para Dados Não Gaussianos via Teste Wald

Jhenifer Caetano Veloso
Lineu Alberto Cavazani de Freitas
Prof. PhD Wagner Hugo Bonat

Trabalho de Conclusão de Curso
Departamento de Estatística
Universidade Federal do Paraná

17 de junho de 2019

Sumário

- 1. Introdução**
- 2. Metodologia**
- 3. Estudo de Simulação**
- 4. Resultados e Discussão**
- 5. Considerações Finais**

1

Introdução

Modelos de regressão univariados

- ▶ Relacionam a média de uma variável resposta às covariáveis.
- ▶ O caso mais conhecido é o modelo linear Normal.
- ▶ Pressupostos do modelo:
 - ▶ Normalidade.
 - ▶ Homocedasticidade.
 - ▶ Independência.
- ▶ Quando a suposição de normalidade não é atendida.
 - ▶ Transformação de variáveis.
 - ▶ Modelos Lineares Generalizados (NELDER; WEDDERBURN, 1972).
- ▶ Teste dos efeitos das covariáveis sobre a variável resposta.
 - ▶ Análise de variância (ANOVA).

Modelos de regressão multivariados

- ▶ Relacionam a média de duas ou mais variáveis respostas às covariáveis.
- ▶ O mais conhecido é o modelo linear Normal multivariado.
- ▶ Pressupostos do modelo:
 - ▶ Normalidade multivariada.
 - ▶ Homogeneidade das matrizes de variâncias e covariâncias.
 - ▶ Independência entre as observações.
- ▶ Quando os pressupostos não são atendidos.
 - ▶ Modelos Multivariados de Covariância Linear Generalizada (MCGLMs) proposto por Bonat e Jørgensen (2016).
- ▶ Teste dos efeitos das covariáveis sobre a variável resposta.
 - ▶ Análise de variância multivariada (MANOVA).

- ▶ Esta classe de modelagem comporta:
 - ▶ Múltiplas respostas.
 - ▶ Respostas de diferentes naturezas.
 - ▶ Respostas correlacionadas.
 - ▶ Observações não independentes.
 - ▶ Extensões multivariadas para modelos de:
 - ▶ Séries temporais.
 - ▶ Dados longitudinais.
 - ▶ Dados espaciais.
- ▶ Faz uso de apenas duas funções:
 - ▶ Função de variância.
 - ▶ Função de ligação.

Para esta classe, não existe discussão a respeito de teste de hipóteses multivariados.

1.1

Introdução **Objetivos**

Proposta

Colaborar com a literatura estatística

- ▶ Propondo e implementando o teste Wald para análise de variância multivariada para dados não gaussianos no contexto dos MCGLMs.
- ▶ Discutindo as propriedades e comportamento do teste proposto com base em
 - ▶ Estudo de simulação.
 - ▶ Aplicação a conjuntos de dados reais.

Proposta

Verificar o poder do teste proposto sob diferentes distribuições de variável resposta e tamanhos de amostra.

Conjuntos de Dados:

1. Processos Movidos contra Grandes Litigantes.
 - ▶ Dados coletados via Consulta de Processos de Primeiro Grau (CJPG) do Tribunal de Justiça de São Paulo.
 - ▶ 3 variáveis respostas.
2. Comportamento de Ovelhas Submetidas à Intervenção Humana.
 - ▶ Estudo delineado com o objetivo de verificar o efeito de intervenção humana e isolamento social no comportamento de ovelhas.
 - ▶ 8 variáveis respostas.

2

Metodologia

2.1

Metodologia

Modelos Multivariados de Covariância Linear Generalizada

Modelos Lineares Generalizados

Sejam

- ▶ \mathbf{Y} um vetor $N \times 1$ de variáveis resposta.
- ▶ \mathbf{X} uma matriz $N \times k$ de delineamento.
- ▶ $\boldsymbol{\beta}$ um vetor $k \times 1$ de parâmetros de regressão desconhecidos.

Um Modelo Linear Generalizado (MLG) pode ser definido da forma

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta}), \\ \text{Var}(\mathbf{Y}) &= \Sigma = V(\boldsymbol{\mu}; p)^{1/2} (\tau_0 \mathbf{I}) V(\boldsymbol{\mu}; p)^{1/2}, \end{aligned} \tag{1}$$

em que

- ▶ g é a função de ligação.
- ▶ $V(\boldsymbol{\mu}; p) = \text{diag}(\vartheta(\boldsymbol{\mu}; p))$ em que $\vartheta(\boldsymbol{\mu}; p)$ é a função de variância.
- ▶ p e τ_0 são os parâmetros de potência e dispersão.
- ▶ \mathbf{I} é a matriz Identidade.

Modelos Lineares Generalizados

- ▶ Os MLGs são uma forma de modelagem univariada para dados de diferentes naturezas.
- ▶ Duas principais restrições:
 - ▶ Incapacidade de lidar com observações dependentes.
 - ▶ Incapacidade de lidar com múltiplas respostas.

Funções de variância.

- ▶ Função de variância potência (JØRGENSEN, 1987; JØRGENSEN, 1997).
 - ▶ $\vartheta(\mu; p) = \mu^p$.
 - ▶ Caracteriza a família Tweedie de distribuições.
 - ▶ Acomoda dados assimétricos e com massa de probabilidade em zero.
 - ▶ Casos particulares: Gaussiana ($p = 0$), Gama ($p = 2$) e Inversa - Gaussiana ($p = 3$).
- ▶ Função de dispersão Poisson-Tweedie (JØRGENSEN; KOKONENDJI, 2015).
 - ▶ $\vartheta(\mu; p) = \mu + \mu^p$.
 - ▶ Acomoda contagens equi, super e subdispersas, cauda pesada e excesso de zeros.
 - ▶ Casos particulares: Hermite ($p = 0$), Neyman Tipo A ($p = 1$) e Binomial Negativa ($p = 2$).
- ▶ Função de variância binomial.
 - ▶ $\vartheta(\mu) = \mu(1 - \mu)$.
 - ▶ Acomoda dados binários e de proporção.

Correlação entre as observações.

- ▶ Substituir a matriz identidade I por uma matriz não diagonal $\Omega(\tau)$ descrita por

$$h\{\Omega(\tau)\} = \tau_0 Z_0 + \dots + \tau_D Z_D, \quad (2)$$

em que

- ▶ $h(\cdot)$ é a função de ligação de covariância.
- ▶ Z_d com $d = 0, \dots, D$ são matrizes que representam a estrutura de covariância.
- ▶ $\tau = (\tau_0, \dots, \tau_D)$ é um vetor $(D + 1) \times 1$ de parâmetros de dispersão.
- ▶ Chamado de preditor linear matricial.

Múltiplas respostas.

Seja

- ▶ $\mathbf{Y}_{N \times R} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_R\}$ uma matriz de variáveis respostas.
- ▶ $\mathbf{M}_{N \times R} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_R\}$ uma matriz de valores esperados.
- ▶ Σ_b uma matriz de ordem $R \times R$ que descreve a correlação entre as variáveis respostas.

A variância e covariância dentro de cada resposta é expressa por

$$\Sigma_r = V_r(\boldsymbol{\mu}_r; p)^{1/2} \boldsymbol{\Omega}_r(\boldsymbol{\tau}) V_r(\boldsymbol{\mu}_r; p)^{1/2}. \quad (3)$$

Um MCGLM é descrito como

$$\begin{aligned} E(\mathbf{Y}) &= \mathbf{M} = \{g_1^{-1}(\mathbf{X}_1\boldsymbol{\beta}_1), \dots, g_R^{-1}(\mathbf{X}_R\boldsymbol{\beta}_R)\} \\ \text{Var}(\mathbf{Y}) &= \mathbf{C} = \boldsymbol{\Sigma}_R \stackrel{G}{\otimes} \boldsymbol{\Sigma}_b, \end{aligned} \tag{4}$$

em que

- ▶ $\boldsymbol{\Sigma}_R \stackrel{G}{\otimes} \boldsymbol{\Sigma}_b = \text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1, \dots, \tilde{\boldsymbol{\Sigma}}_R)(\boldsymbol{\Sigma}_b \otimes \mathbf{I})\text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1^\top, \dots, \tilde{\boldsymbol{\Sigma}}_R^\top)$ é o produto generalizado de Kronecker.
- ▶ A matriz $\tilde{\boldsymbol{\Sigma}}_r$ denota a matriz triangular inferior da decomposição de Cholesky da matriz $\boldsymbol{\Sigma}_r$.
- ▶ O operador Bdiag denota a matriz bloco-diagonal.
- ▶ \mathbf{I} uma matriz identidade de ordem $N \times N$.

Funções de estimação

A estimação dos parâmetros de regressão (função quasi-score) e de dispersão (função de estimação de Pearson) são dadas por

$$\begin{aligned}\psi_{\beta}(\beta, \lambda) &= \mathbf{D}^{\top} \mathbf{C}^{-1} (\mathcal{Y} - \mathcal{M}) \\ \psi_{\lambda_i}(\beta, \lambda) &= \text{tr}(W_{\lambda_i}(\mathbf{r}^{\top} \mathbf{r} - \mathbf{C})), i = 1, \dots, Q,\end{aligned}\tag{5}$$

em que

- ▶ β é um vetor $K \times 1$ de parâmetros de regressão.
- ▶ λ é um vetor $Q \times 1$ de parâmetros de dispersão.
- ▶ \mathcal{Y} é um vetor $NR \times 1$ com os valores da matriz de variáveis respostas $\mathbf{Y}_{N \times R}$ empilhados.
- ▶ \mathcal{M} é o vetor $NR \times 1$ com os valores da matriz de valores esperados $\mathbf{M}_{N \times R}$ empilhados.
- ▶ $\mathbf{D} = \nabla_{\beta} \mathcal{M}$ é uma matriz $NR \times K$ e ∇_{β} denota o operador gradiente.
- ▶ $W_{\lambda_i} = -\frac{\partial \mathbf{C}^{-1}}{\partial \lambda_i}$.
- ▶ $\mathbf{r} = (\mathcal{Y} - \mathcal{M})$.

Distribuição assintótica e algoritmo de estimação

Algoritmo de estimação:

- ▶ Para resolver o sistema de equações $\psi_{\beta} = 0$ e $\psi_{\lambda} = 0$ faz-se uso do algoritmo chaser modificado:

$$\begin{aligned}\boldsymbol{\beta}^{(i+1)} &= \boldsymbol{\beta}^{(i)} - S_{\boldsymbol{\beta}}^{-1} \psi_{\beta}(\boldsymbol{\beta}^{(i)}, \boldsymbol{\lambda}^{(i)}), \\ \boldsymbol{\lambda}^{(i+1)} &= \boldsymbol{\lambda}^{(i)} \alpha S_{\boldsymbol{\lambda}}^{-1} \psi_{\lambda}(\boldsymbol{\beta}^{(i+1)}, \boldsymbol{\lambda}^{(i)}).\end{aligned}\tag{6}$$

Distribuição assintótica:

- ▶ Seja $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\lambda}}^T)^T$ o estimador baseado em funções de estimação de $\boldsymbol{\theta}$.
- ▶ A distribuição assintótica de $\hat{\boldsymbol{\theta}}$ é

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, J_{\boldsymbol{\theta}}^{-1}),\tag{7}$$

em que $J_{\boldsymbol{\theta}}^{-1}$ é a inversa da matriz de informação de Godambe, dada por

$$J_{\boldsymbol{\theta}}^{-1} = S_{\boldsymbol{\theta}}^{-1} V_{\boldsymbol{\theta}} S_{\boldsymbol{\theta}}^{-T},\tag{8}$$

em que $S_{\boldsymbol{\theta}}^{-T} = (S_{\boldsymbol{\theta}}^{-1})^T$.

2.2

Metodologia

Teste Wald para Análise de Variância Multivariada para Dados Não Gaussianos

MANOVA

- ▶ Vasta discussão na literatura.
- ▶ Tem o objetivo de verificar a nulidade conjunta dos parâmetros de um modelo de regressão multivariado.
- ▶ Existem diversas propostas, tais como:
 - ▶ Lambda de Wilk's (WILKS, 1932).
 - ▶ Traço de Hotelling-Lawley (LAWLEY, 1938; HOTELLING, 1951).
 - ▶ Traço de Pillai (PILLAI et al., 1955).
 - ▶ Maior raiz de Roy (ROY, 1953).

Hipóteses

Sob a equação que descreve um McGLM, as hipóteses de interesse são definidas como

$$\begin{aligned} H_0 &= \mathbf{L}\boldsymbol{\beta} = \mathbf{0} \\ H_1 &= \mathbf{L}\boldsymbol{\beta} = \mathbf{n}, \end{aligned} \tag{9}$$

em que

- ▶ \mathbf{n} é um vetor qualquer não nulo.
- ▶ Sob hipótese nula, $\mathbf{L} = \mathbf{G} \otimes \mathbf{F}$.
 - ▶ \mathbf{G} tem dimensão $R \times R$ e especifica as hipóteses referentes às respostas.
 - ▶ \mathbf{F} tem dimensão $s \times p$ e indica as hipóteses entre tratamentos.
 - ▶ s é o número de restrições lineares.
 - ▶ p é o número de coeficientes de regressão para cada variável resposta.

Sendo assim, a matriz \mathbf{L} tem dimensão $(sR \times P)$.

Extensão do Teste Wald

Para avaliar a hipótese linear tem-se a estatística de Wald representada por

$$W_s = (\mathbf{L}\boldsymbol{\beta})^T \left(\mathbf{L}J_{\boldsymbol{\beta}}^{-1} \mathbf{L}^T \right)^{-1} (\mathbf{L}\boldsymbol{\beta}), \quad (10)$$

em que

- ▶ $J_{\boldsymbol{\beta}}^{-1}$ é a parte da inversa da matriz de informação de Godambe que considera apenas os parâmetros de regressão.
- ▶ Sob hipótese nula, a estatística segue assintoticamente distribuição Qui-quadrado com sR graus de liberdade.

Exemplificação do teste proposto

Considerando

- ▶ 2 variáveis respostas.
- ▶ 1 variável explicativa X categórica com três níveis (A, B e C).

O preditor linear para cada variável de resposta é dado por

$$\mu_r = \beta_{r0} + \beta_{r1} [X = B] + \beta_{r2} [X = C],$$

em que

- ▶ β_{r0} é o intercepto.
- ▶ β_{r1} é a diferença do tratamento A para B.
- ▶ β_{r2} é a diferença do tratamento A para C.

A matriz de delineamento tem a forma

$$D = \begin{pmatrix} 1 & . & . & . & . & . \\ 1 & 1 & . & . & . & . \\ 1 & . & 1 & . & . & . \\ . & . & . & 1 & . & . \\ . & . & . & 1 & 1 & . \\ . & . & . & 1 & . & 1 \end{pmatrix}.$$

Exemplificação do teste proposto

A hipótese nula do teste pode ser escrita como

$$H_0 = \beta_{r1} = \beta_{r2} = 0, \text{ para } r = 1, 2.$$

O que equivale, em notação matricial a

$$H_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

de forma que é possível verificar que

$$\mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ e } \mathbf{F} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

3

Estudo de Simulação

Descrição do estudo

Objetivo: Verificar a eficiência do teste de hipóteses proposto para testar a nulidade simultânea dos parâmetros.

- ▶ Foram consideradas 3 variáveis respostas.
- ▶ Distribuição das variáveis respostas: Poisson, Binomial ($n=10$) e Beta.
- ▶ Uma covariável de 5 níveis: 1, 2, 3, 4 e 5.
- ▶ Correlação entre as variáveis respostas:

$$\Sigma_b = \begin{bmatrix} 1 & 0,75 & 0,5 \\ 0,75 & 1 & 0,25 \\ 0,5 & 0,25 & 1 \end{bmatrix}.$$

- ▶ Coeficientes de regressão: entre $-0,5$ e $0,5$, variando de $0,1$ em $0,1$.
- ▶ 1000 amostras de tamanhos: 50, 100, 250, 500 e 1000.
- ▶ Considerando todas as possibilidades, foram simulados, para cada uma das variáveis respostas, 55 cenários.

Especificação dos modelos (MCGLM)

- ▶ Função de ligação e de variância:

		F. de ligação	F. de variância
Poisson	→	Logarítmica	Tweedie
Binomial	→	Logito	Binomial
Beta	→	Logito	Binomial

- ▶ O preditor matricial foi especificado como

$$h\{\boldsymbol{\Omega}(\boldsymbol{\tau})\} = \tau_0 Z_0, \quad (11)$$

em que Z_0 é uma matriz Identidade de ordem dada pelo tamanho amostral.

- ▶ Os modelos são dados por

$$g_r(\mu_{ijr}) = \beta_{0r} + \beta_{1r} X_{ij}, \quad (12)$$

em que

- ▶ r varia de 1 até 3 e refere-se às variáveis resposta.
- ▶ i diz respeito ao número de observações.
- ▶ j varia de 1 até 5 e diz respeito aos níveis da variável categórica.

Tabela: Valores esperados para os parâmetros de regressão fixados.

Parâmetro	Valor esperado	
	Poisson	Binomial/Beta
Intercepto	10	0,5
-0,5	6,07	0,38
-0,4	6,70	0,40
-0,3	7,41	0,43
-0,2	8,19	0,45
-0,1	9,05	0,48
0	10	0,50
0,1	11,05	0,52
0,2	12,21	0,55
0,3	13,50	0,57
0,4	14,92	0,60
0,5	16,49	0,62

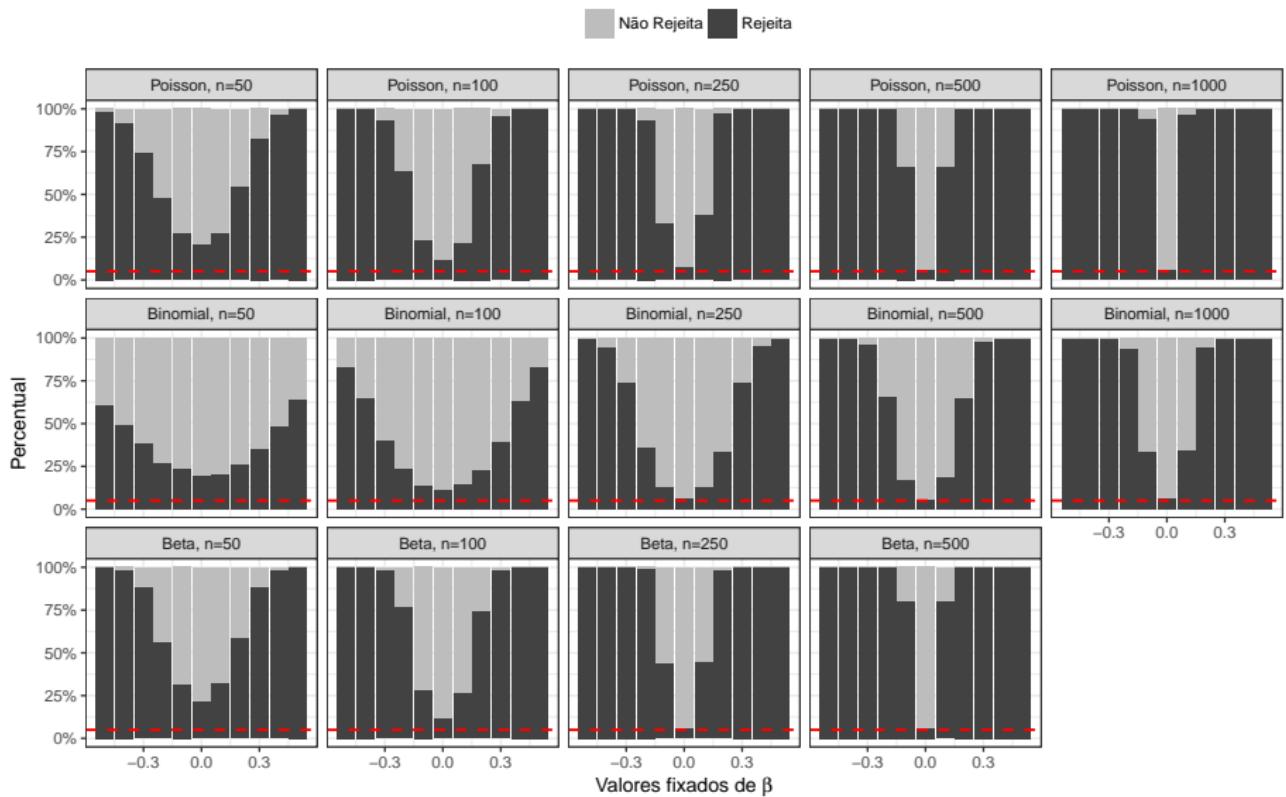


Figura: Percentual de rejeição e não rejeição da hipótese de nulidade conjunta dos parâmetros dos modelos para as amostras simuladas.

4

Resultados e Discussão

4.1

Resultados e Discussão Análise Comportamental de Ovelhas Submetidas à Intervenção Humana

Descrição do experimento

- ▶ Experimento realizado na fazenda experimental INRA La Fage, Roqueford, França, em setembro de 2015.
- ▶ Foram observadas 20 ovelhas pertencentes a 2 linhagens genéticas:
 - ▶ Reativa ao isolamento social, com 9 animais.
 - ▶ Não reativa ao isolamento social, com 11 animais.
- ▶ Os animais foram escovados por um humano que lhes era familiar.
- ▶ Verificou-se o número de alterações de postura e as proporções de tempo em que os animais permaneciam em determinadas posturas.

Objetivo:

- ▶ Verificar o efeito da linhagem genética, intervenção humana e isolamento social nas respostas que configuram contagens e proporções (TAMIOSO et al., 2017).

Descrição do experimento



Figura: Fotografia dos animais sob análise.

Descrição do experimento



Figura: Fotografia dos animais sob análise.

Condução do experimento

O experimento foi conduzido em três sessões experimentais:

1. Uma grade de metal separando o animal testado dos demais animais, sem distância entre eles.
2. Duas grades de metal separando o animal testado dos demais animais a uma distância de 1,7 metros.
3. Animais separados por apenas uma grade, sem distanciamento dos demais animais.

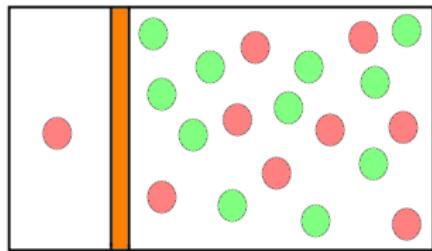


Figura: Ilustração das sessões 1 e 3.

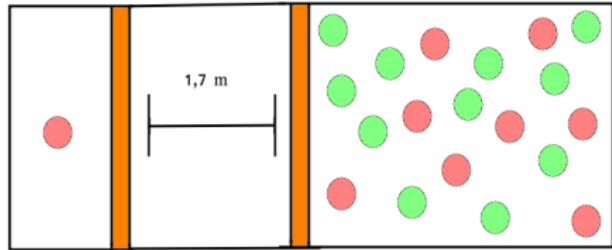


Figura: Ilustração da sessão 2.

Condução do experimento



Figura: Fotografia do procedimento nas sessões 1 e 3.

Condução do experimento

Em cada sessão, as ovelhas foram observadas em 3 momentos distintos:

- ▶ Pré escovação: duração de 2 minutos e 30 segundos.
- ▶ Escovação: duração de 3 minutos.
- ▶ Pós escovação: duração de 2 minutos e 30 segundos.

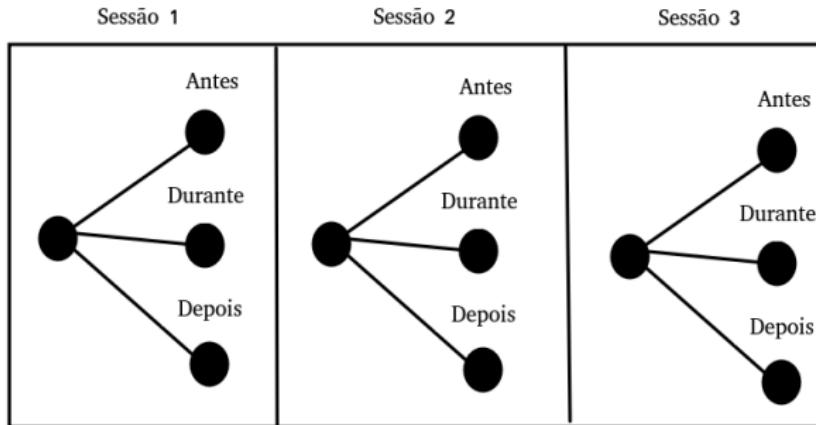


Figura: Delineamento experimental para um animal considerando as diferentes seções e momentos de observação.

Condução do experimento



Figura: Fotografia de um dos animais sob análise durante a escovação.

Variáveis do estudo

► Variáveis respostas:

Contagens	Proporções
<i>Número de mudanças de postura</i>	<i>Proporção do tempo nas posturas</i>
Corpo (ncorpo)	Orelhas levantadas (resporlev)
Cabeça (ncabeca)	Orelhas levantadas ou assimétricas (resporelha2)
Orelha (norelha)	Olhos fechados ou semi-cerrados (respolho)
Olho (nolho)	Com a cauda em movimento (respcauda)

► Variáveis explicativas:

- Sessão: fator de 3 níveis referente a condição de isolamento social.
- Momento: fator de 3 níveis referente a intervenção humana (escovação).
- Linhagem: fator de 2 níveis classificando animais (reativos ou não reativos).

Análise das variáveis respostas de contagem

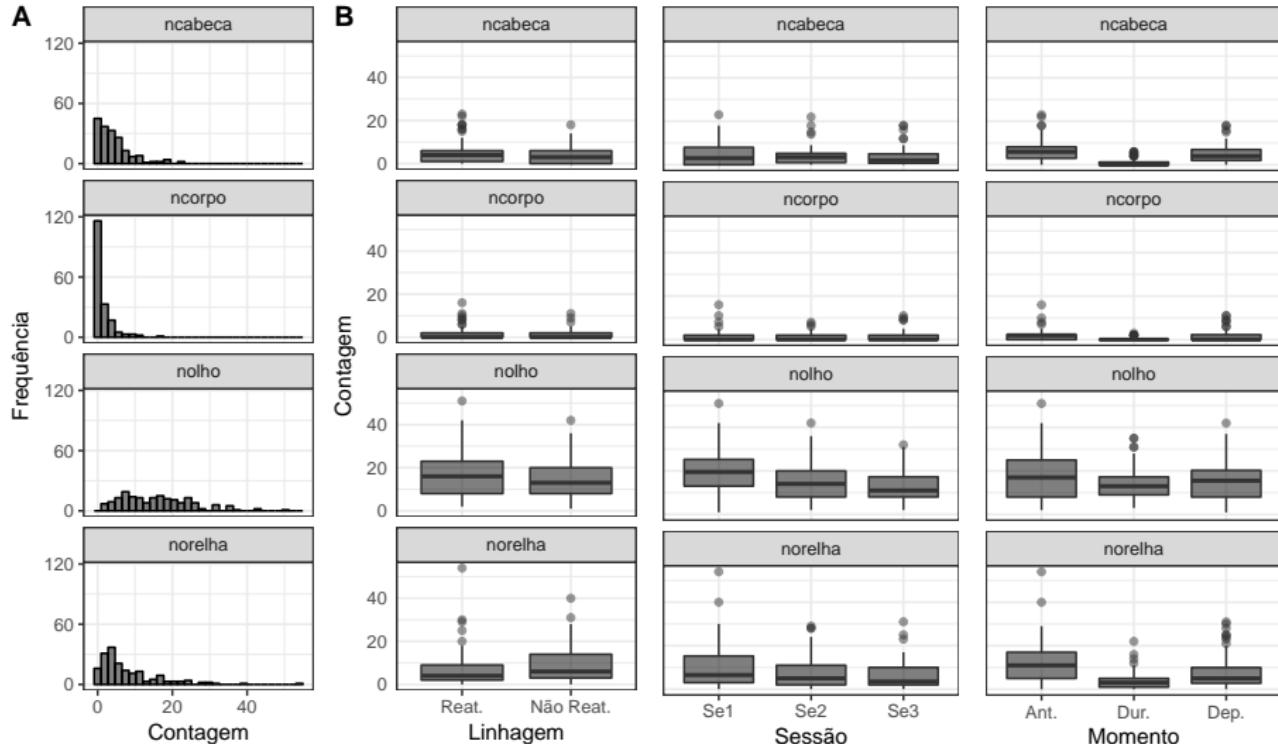


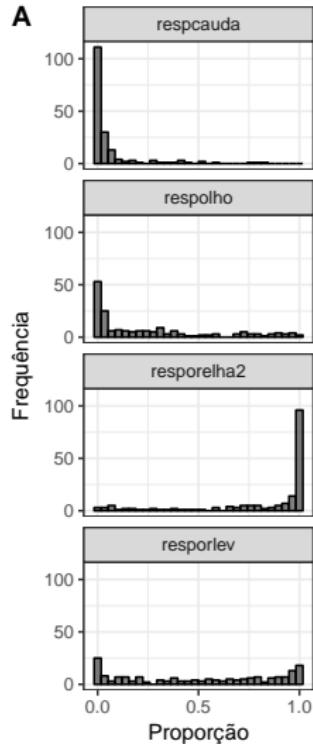
Figura: Comportamento das variáveis respostas (A) e comportamento das respostas em função das covariáveis (B).

Análise das variáveis respostas de contagem

- ▶ O número de mudanças de postura de corpo e cabeça são inflacionadas em 0.
- ▶ O número de mudanças de postura de orelha é assimétrica à direita.
- ▶ Nenhuma resposta igual a 0 foi observada para as variações de postura de olho.
- ▶ Marginalmente, há efeito de momento para o número de mudanças de postura de cabeça, corpo e orelha.
- ▶ Não é possível verificar efeito das variáveis sob as mudanças de postura de olho.

Análise das variáveis respostas de proporção

A



B

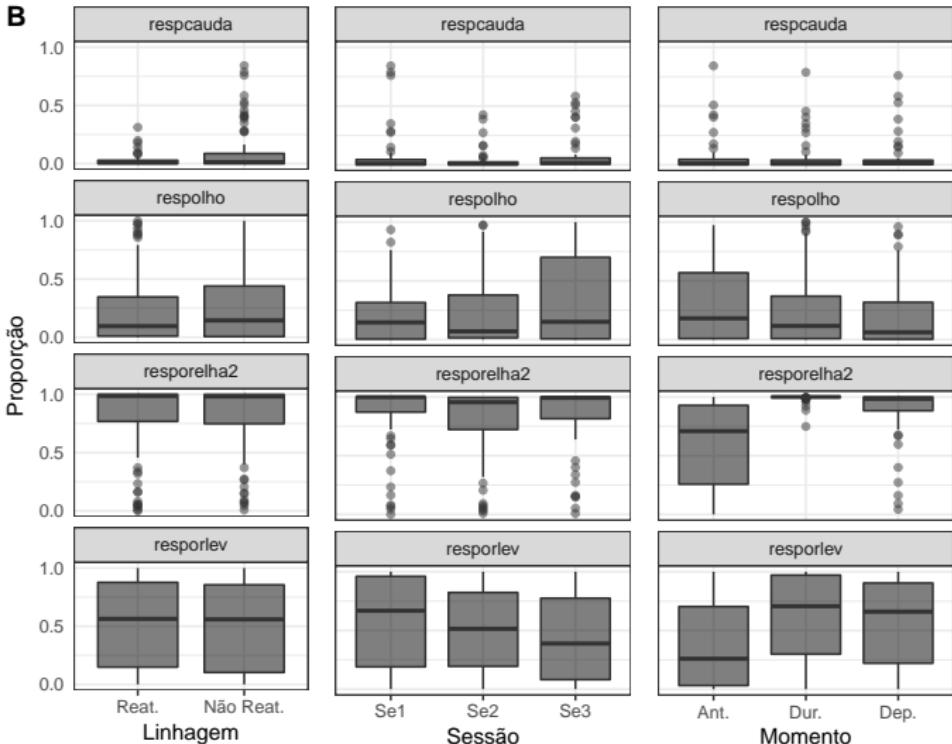


Figura: Comportamento das variáveis respostas (A) e comportamento das respostas em função das covariáveis (B).

Análise das variáveis respostas de proporção

- ▶ O tempo com as orelhas levantadas ou assimétricas é uma variável inflacionada em 1.
- ▶ O tempo com a cauda em movimento e a proporção do tempo com os olhos fechados ou semi-cerrados são inflacionadas em 0.
- ▶ O tempo com as orelhas levantadas apresenta frequências elevadas em 0 e em 1 e frequências menores entre estes dois valores.
- ▶ Observa-se efeito marginal de momento para as duas variáveis respostas referentes à posturas de orelha.
- ▶ Para o tempo com a cauda em movimento nota-se que há um efeito marginal de linhagem.
- ▶ Para o tempo com os olhos fechados ou semi-cerrados há efeito de todas as covariáveis.

Modelagem

- ▶ Foram consideradas todas as 8 respostas apresentadas.
- ▶ Foi ajustado um modelo multivariado com os efeitos fixos de sessão experimental, momento e linhagem dispostos de forma aditiva.
- ▶ Trata-se de um experimento em que as observações não são independentes.

- ▶ Função de ligação e de variância:

		F. de ligação	F. de variância
Contagem	→	Logarítmica	Poisson-Tweedie
Proporção	→	Logito	Binomial

Modelagem - preditor linear

Os preditores lineares são descritos como

$$g_r(\mu_{ijklr}) = \beta_{0r} + \beta_{1j} \text{Sessao}_{ij} + \beta_{2k} \text{Momento}_{ik} + \beta_{3r} \text{Linhagem}_i, \quad (13)$$

em que

- ▶ r refere-se às variáveis respostas.
- ▶ $g_1(\cdot)$ a $g_4(\cdot)$ refere-se à função de ligação logarítmica para as respostas de contagem.
- ▶ $g_5(\cdot)$ a $g_8(\cdot)$ refere-se à função de ligação logito para as respostas de proporção.
- ▶ $\beta_0, \beta_{1j}, \beta_{2k}$ e β_{3r} representam o intercepto, efeitos de sessão, momento e linhagem, respectivamente.
- ▶ i varia de 1 até 180 e diz respeito ao número de observações.
- ▶ j assume os níveis de sessão experimental.
- ▶ k assume os níveis de momento experimental.
- ▶ l assume os níveis de linhagem.

Modelagem - preditor matricial

- ▶ São comuns a todas as respostas.
- ▶ Evidencia que as medidas de um mesmo animal são correlacionadas, bem como as medidas dentro da mesma sessão.

Os preditores são dados por

$$h\{\Omega(\tau)\} = \tau_0 Z_0 + \tau_1 Z_1 + \tau_2 Z_2, \quad (14)$$

em que

- ▶ $h(\cdot)$ é a função de ligação de variância identidade.
- ▶ τ_0, τ_1 e τ_2 representam os parâmetros de dispersão.
- ▶ Z_0 é uma matriz identidade de ordem 180×180 .
- ▶ Z_1 é uma matriz 180×180 composta por blocos 9×9 de uns na diagonal principal especificando que as medidas provenientes do mesmo animal são correlacionadas.
- ▶ Z_2 é uma matriz 180×180 composta por blocos 3×3 de uns na diagonal principal indicando que medidas na mesma sessão são correlacionadas.

Estimativas dos parâmetros

Tabela: Estimativas dos parâmetros de dispersão e potência, e erro-padrão (ep).

Resposta	τ_0	τ_1	τ_2	p
	Est. (ep)	Est. (ep)	Est. (ep)	Est. (ep)
ncorpo	2.34 (0.69)	0.42 (0.20)	-0.22 (0.22)	1.28 (0.26)
ncabeca	1.62 (0.38)	0.79 (0.32)	-0.11 (0.17)	1.11 (0.17)
norelha	1.75 (0.88)	0.23 (0.20)	0.14 (0.21)	1.42 (0.28)
nolho	1.67 (3.55)	0.16 (0.38)	0.18 (0.45)	1.21 (0.77)
resporelha2	0.75 (0.14)	-0.02 (0.03)	0.07 (0.07)	1.20 (0.04)
respcauda	0.59 (0.32)	-0.08 (0.06)	0.38 (0.18)	1.29 (0.11)
respolho	1.02 (0.48)	-0.05 (0.08)	0.40 (0.26)	1.54 (0.31)
resporlev	0.06 (0.09)	0.001 (0.01)	0.01 (0.02)	-0.27 (1.01)

Qualidade de ajuste

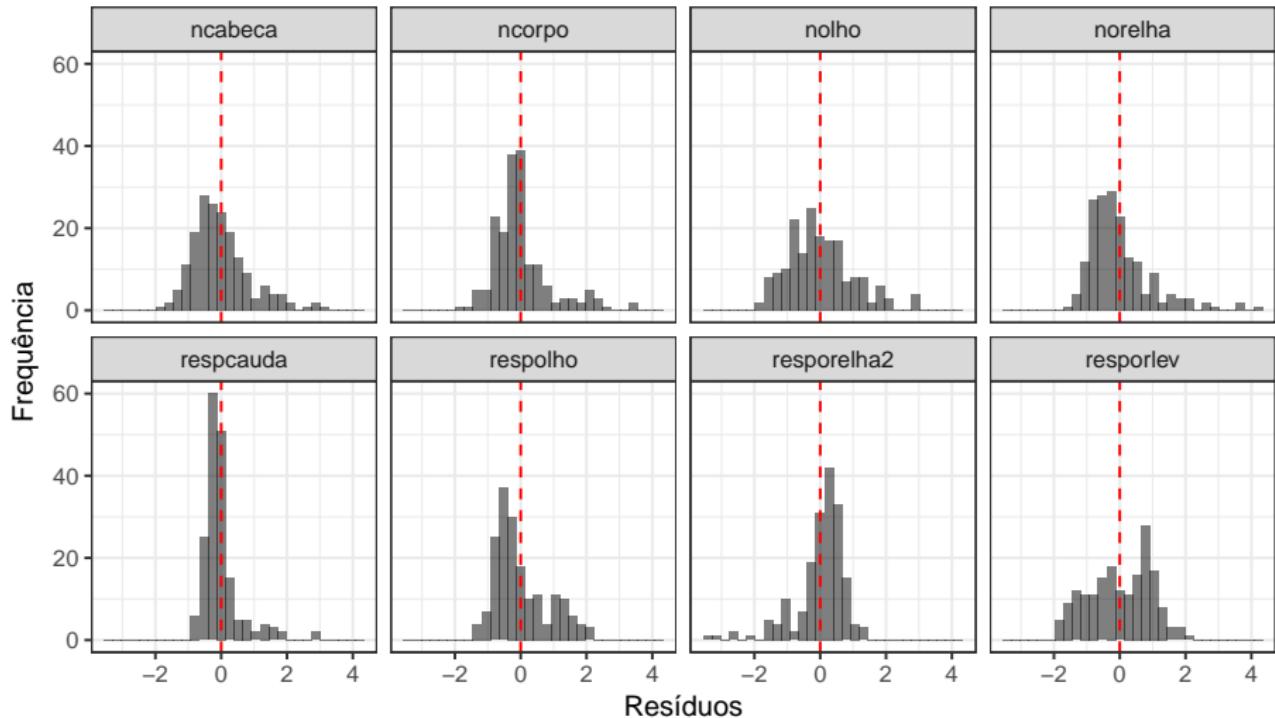


Figura: Histograma dos resíduos de Pearson por resposta.

Qualidade de ajuste

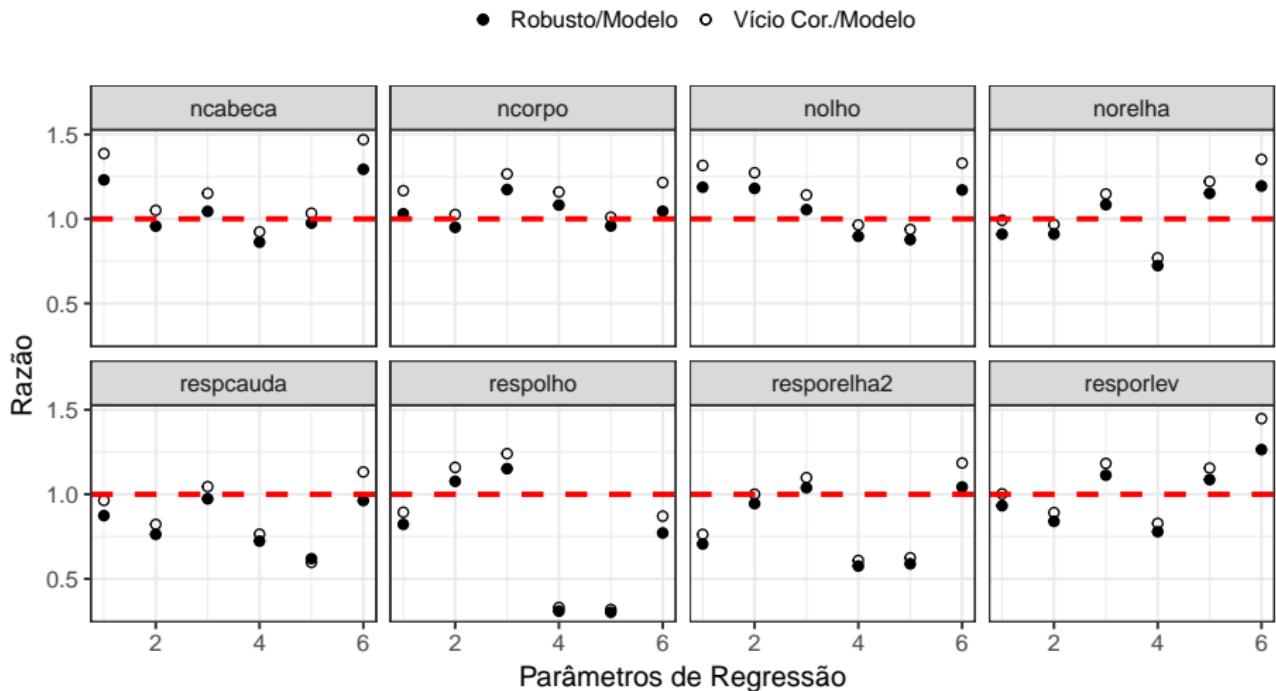


Figura: Razão entre os erros robusto e vício corrigido com os erros do modelo.

MANOVA para dados não gaussianos

$$H_0 = L\beta = \mathbf{0}$$
$$H_1 = L\beta \neq \mathbf{0}$$

Tabela: Resumo do teste de hipóteses para o modelo ajustado.

Efeitos	GL	Hotelling Lawley	Qui-quadrado	p-valor
Intercepto	8	6.98	1257.72	< 0.01
Sessão	16	0.18	33.27	< 0.01
Momento	16	0.78	140.57	< 0.01
Linhagem	8	0.10	19.41	< 0.01

ANOVA

Tabela: Medidas de ajuste de cada resposta para avaliação individual dos efeitos das covariáveis.

Resposta	χ^2	gl	p-valor	χ^2	gl	p-valor	χ^2	gl	p-valor
Contagens									
Sessão				Momento				Linhagem	
ncorpo	0,83	2	0,65	20,26	2	<0,01	0,22	1	0,63
ncabeça	3,47	2	0,17	54,69	2	<0,01	0	1	0,99
norelha	8,97	2	0,01	48,85	2	<0,01	4,05	1	0,04
nolho	16,37	2	<0,01	6,25	2	0,04	1,39	1	0,23
Proporções									
Sessão				Momento				Linhagem	
resporelha2	1,58	2	0,45	47,50	2	<0,01	0,13	1	0,70
respcauda	1,47	2	0,47	0,02	2	0,98	12,31	1	<0,01
respolho	2,43	2	0,29	2,97	2	0,22	0,34	1	0,55
resporlev	4,59	2	0,10	19,13	2	<0,01	0,01	1	0,91

ANOVA

- ▶ Para a variação de posturas de corpo, cabeça, proporção com as orelhas levantadas ou assimétricas e proporção com as orelhas levantadas constatou-se efeito de momento.
- ▶ Para o número de mudanças de postura de olho, sessão experimental e momento se mostraram significativas.
- ▶ Na proporção do tempo com a cauda em movimento apenas o efeito de linhagem foi detectado.
- ▶ No número de mudanças de postura de orelha todas as variáveis explicativas tiveram efeito constatado.
- ▶ Nenhuma das variáveis se mostrou significativa na proporção do tempo com os olhos fechados e semi-cerrados.

5

Considerações Finais

Teste de hipóteses

O teste pode ser aplicado em conjuntos de dados:

- ▶ Com observações independentes ou não e com medidas repetidas.
- ▶ Em que todas as respostas são ou não de mesma natureza.
- ▶ Cujo estudo é observacional ou experimental.
- ▶ Com as respostas sujeitas às mesmas variáveis explicativas.

Com base no estudo de simulação:

- ▶ Atingiu-se os 95% de confiança para os casos com β fixados em 0 e tamanho amostral maior que 250.
- ▶ O poder do teste para todos os cenários com distribuição binomial foi inferior aos cenários com as demais distribuições.
- ▶ Para resposta seguindo distribuição beta com tamanho amostral 1000 não obteve-se resultados.

Trabalhos futuros

Sugestões para continuidade da pesquisa:

- ▶ Construção de outros testes multivariados para dados não gaussianos no contexto dos MCGLMs.
- ▶ Ampliar o estudo de simulação:
 - ▶ Ampliar o número de respostas.
 - ▶ Considerar casos com respostas seguindo diferentes distribuições.
 - ▶ Simular amostras com observações não independentes.
 - ▶ Aumentar o tamanho amostral.
 - ▶ Testar diferentes funções de ligação e variância.

Referências

-  BONAT, W. H.; JØRGENSEN, B. Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 65, n. 5, p. 649–675, 2016.
-  HOTELLING, H. *A generalized T test and measure of multivariate dispersion*. [S.l.], 1951.
-  JØRGENSEN, B. Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 49, n. 2, p. 127–145, 1987.
-  JØRGENSEN, B. *The theory of dispersion models*. [S.l.]: CRC Press, 1997.
-  JØRGENSEN, B.; KOKONENDJI, C. C. Discrete dispersion models and their tweedie asymptotics. *ASyA Advances in Statistical Analysis*, Springer, v. 100, n. 1, p. 43–78, 2015.
-  LAWLEY, D. A generalization of fisher's z test. *Biometrika*, JSTOR, v. 30, n. 1/2, p. 180–187, 1938.
-  NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, v. 135, p. 370–384, 1972.
-  PILLAI, K. et al. Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 26, n. 1, p. 117–121, 1955.
-  ROY, S. N. On a heuristic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics*, JSTOR, p. 220–238, 1953.

Referências

-  TAMIOSO, P. R. et al. Does emotional reactivity influence behavioral and cardiac responses of ewes submitted to brushing? *Behavioural Processes*, p. np, 2017.
-  WILKS, S. S. Certain generalizations in the analysis of variance. *Biometrika*, JSTOR, p. 471–494, 1932.

Obrigado.