

Um estudo de caso de análise de comportamento animal usando GAMLSS

Prof. Me. Lineu Alberto Cavazani de Freitas

Departamento de Estatística
Laboratório de Estatística e Geoinformação



Introdução

Estudos de comportamento animal

- ▶ O **comportamento de um animal** reflete seu **estado emocionais**.
- ▶ Estudos de comportamento animal forneceram informações úteis sobre o bem-estar animal em uma ampla variedade de situações.
- ▶ Tais estudos costumam produzir **dados complexos e não estruturados**, registrados a partir de áudios, imagens e vídeos, entre outros.
- ▶ Estes dados não estruturados, quando tratados, são fontes de **diversas variáveis aleatórias**.

Estudos de comportamento animal

Por exemplo:

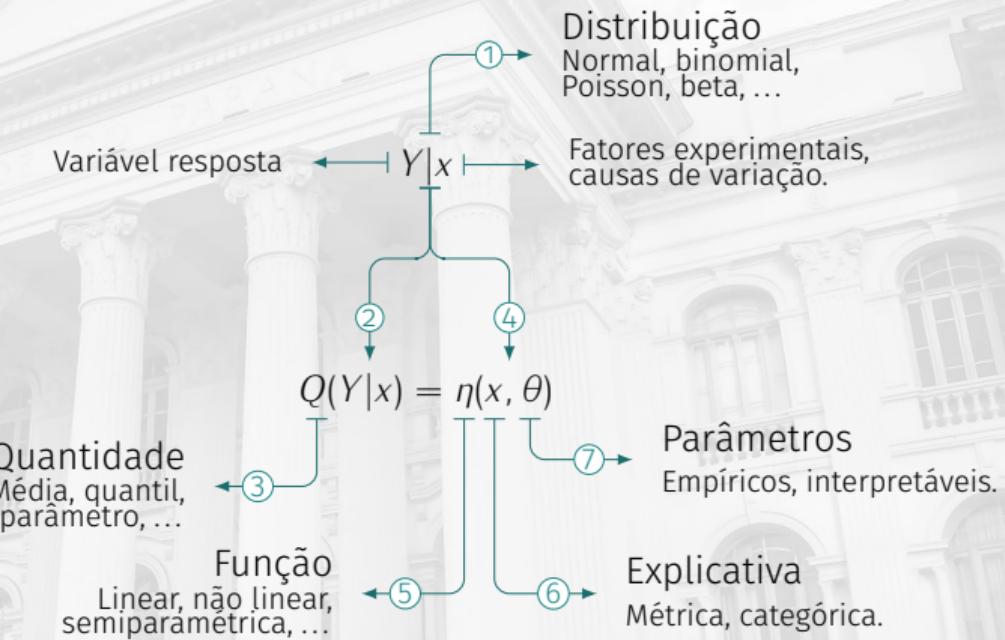
- ▶ Considere que a movimentação de cabeça de um animal em contato com um ser humano reflete algum estado emocional.
- ▶ Para avaliar a movimentação de cabeça podemos:
 1. Verificar se cada animal apresentou ou não o comportamento específico, configurando um **resultado binário**.
 2. Avaliar o número de vezes que o comportamento foi repetido, gerando uma **variável de contagem**.
 3. Avaliar a proporção do tempo em que o animal realizou o movimento, gerando uma **variável contínua**.

Estudos de comportamento animal

- ▶ Estudos de comportamento animal costumam apresentar diversos outros fatores que devem ser levados em consideração na análise estatística.
- ▶ Unidades agrupadas (ninhadas, rebanhos).
- ▶ Estudos longitudinais (cada animal é avaliado em diferentes momentos).
- ▶ Estudos espaciais (coordenadas espaciais são relevantes para a análise).
 - ▶ Relações não lineares.
 - ▶ Heterocedasticidade.
 - ▶ Assimetria e/ou curtose.
 - ▶ Super ou subdispersão.
 - ▶ Inflação zeros.

Modelos de regressão

- ▶ A análise estatística de dados de comportamento animal requer **modelos flexíveis**.
- ▶ Em geral, são usados **modelos de regressão**.



Modelos de regressão

1. Definição do problema.

- ▶ Qual o fenômeno aleatório de interesse?
- ▶ Que fatores externos podem afetar este fenômeno?

2. Análise dos dados via regressão.

- ▶ Distribuição de probabilidade.
- ▶ Especificação do modelo.
- ▶ Obtenção dos parâmetros (ajuste).
- ▶ Diagnóstico.

3. Interpretação dos resultados.

- ▶ Quais os fatores externos apresentam ou não impacto sobre o fenômeno.
- ▶ Qual a dimensão desse impacto.

- ▶ Existem diversas classes de modelos de regressão, para os mais diversos fins.
- ▶ Uma destas classes é o GAMLSS.
- ▶ GAMLSS: **generalized additive models for location, scale and shape**.
- ▶ É um **framework** para ajuste de modelos de regressão.
- ▶ Dispõe de um grande conjunto de **distribuições**.
- ▶ Permite modelar cada parâmetro de distribuição incluindo **covariáveis, efeitos aleatórios e suavizadores**.

- ▶ Desta forma, várias das restrições relacionadas aos dados de comportamento animal podem ser adequadamente tratadas.
- ▶ Os **efeitos aleatórios** são uma ferramenta adequada para lidar com dados multiníveis e permitem acomodar adequadamente a **estrutura de correlação** resultante de projetos de medidas repetidas.
- ▶ Os **suavizadores** são úteis na modelagem de **relacionamentos não lineares** entre os parâmetros de distribuição e as covariáveis contínuas.

Pontos vistos até aqui

- ▶ Dados de comportamento animal.
 - ▶ Objetivo.
 - ▶ Tipos de resposta.
 - ▶ Especificidades dos dados.
- ▶ Modelos de regressão.
 - ▶ Ideia geral.
- ▶ GAMLSS.
- ▶ Potencial do GAMLSS para problemas de comportamento animal.

Objetivo

- ▶ O estudo teve como objetivo explorar o potencial do **GAMLSS** na **análise de comportamento animal**.
- ▶ Os dados que motivaram o estudo é de **comportamento de ovinos** em que buscou-se avaliar se **intervenção humana** e **isolamento social** influenciam o comportamento dos animais.
- ▶ A amostra era composta por **20 ovinos** pertencentes a **duas linhagens genéticas** diferentes.

Objetivo

- ▶ Dentre as principais informações resultantes deste experimento podemos destacar dois tipos de variáveis:
 1. O **número de mudanças** posturais.
 2. A **proporção** (ou percentual) de tempo que cada postura foi expressa.

- ▶ Uma série de respostas foi observada, como
 1. Posturas de **orelha** (levantadas, horizontais e assimétricas).
 2. Postura dos **olhos** (fechados, semicerrados e abertos).
 3. Comportamento **alimentar** (comendo, ruminando, não comendo ou ruminando).
- entre outros.

Objetivo

- ▶ A título de ilustração, consideramos duas variáveis de tipos distintos:
 - ▶ O **número de mudanças** de postura de orelha.
 - ▶ A **proporção** de tempo realizando a postura de orelha horizontal.
- ▶ Ovelhas submetidas a eventos positivos:
 - ▶ Apresentam altas proporções de posturas de orelha horizontais.
 - ▶ Menos alterações de postura de orelha do que ovelhas em situações negativas.

Estudo de caso

Estudo de caso

- ▶ Foram observadas **20 ovelhas** pertencentes a **2 linhagens genéticas**, que foram **submetidas à escovação** 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.



Figura 1. Fotografia dos animais sob análise.

Estudo de caso



Figura 2. Fotografia dos animais sob análise.

Estudo de caso

O experimento foi conduzido em **três sessões experimentais**:

1. Na primeira havia uma grade de metal separando o animal testado dos demais animais, **sem distância** entre eles.
2. Na segunda havia duas grades de metal separando o animal testado dos demais animais a uma **distância de 1,7 m**.
3. Na terceira sessão os animais voltaram a ser separados por apenas uma grade, **sem distanciamento** dos demais animais.

Estudo de caso

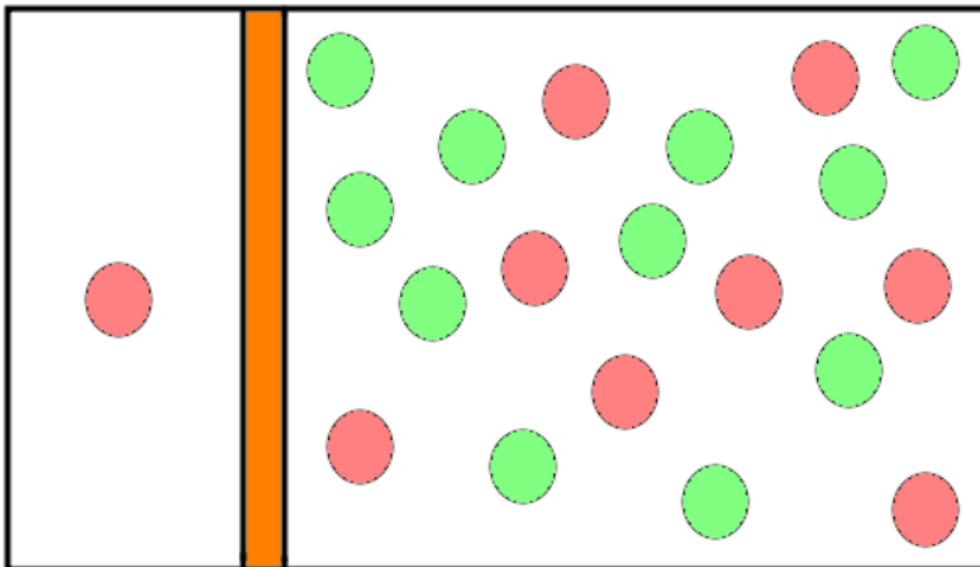


Figura 3. Esquema sessões 1 e 3 (sem isolamento).

Estudo de caso

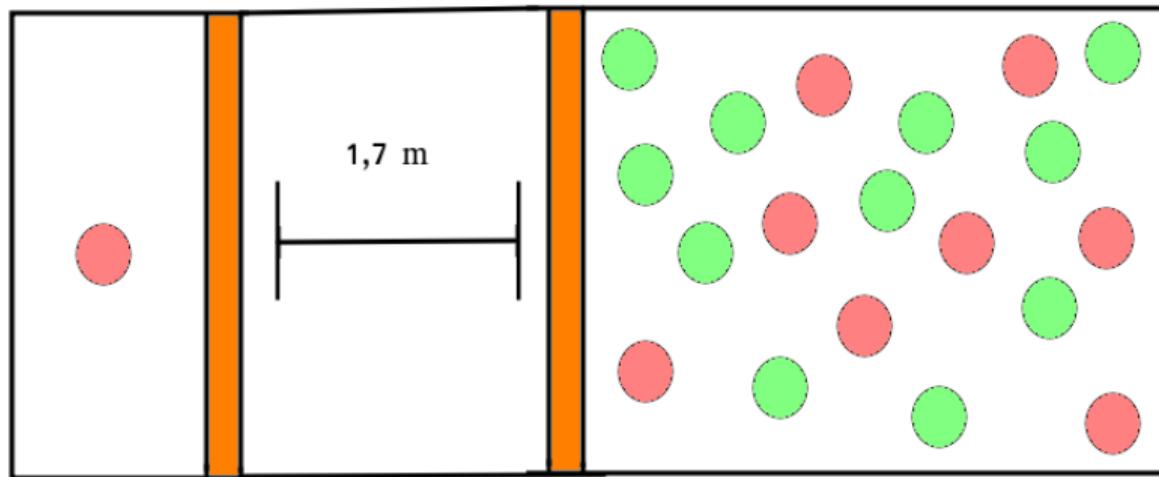


Figura 4. Esquema sessão 2 (com isolamento).

Estudo de caso



Figura 5. Foto sessão.

Estudo de caso

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

1. Fase de **pré escovação**, com duração de 2 minutos e 30 segundos.
2. Fase de **escovação**, com duração de 3 minutos.
3. Fase de **pós escovação**, com duração de 2 minutos e 30 segundos.

Estudo de caso



Figura 6. Fotografia animal sob intervenção humana (escovação).

Estudo de caso

Temos **3 variáveis categóricas**:

1. **Linhagem**: fator de 2 níveis que classifica os animais como reativos (R+) ou não reativos (R-).
2. **Sessão**: fator de 3 níveis que indica a sessão experimental de acordo com o isolamento social (sessões 1 e 3 sem isolamento, sessão 2 com isolamento).
3. **Momento**: fator de 3 níveis que indica se o animal está ou não sob intervenção humana (antes, durante ou depois da intervenção).

Temos **2 respostas** distintas:

1. Número de mudanças de postura de orelha.
2. Proporção do tempo com as orelhas em posição neutra.

Estudo de caso

- ▶ Considerando as combinações entre as variáveis, **cada animal** contribui com **9 medidas** ao conjunto de dados.
- ▶ Portanto existe uma estrutura de medidas repetidas.

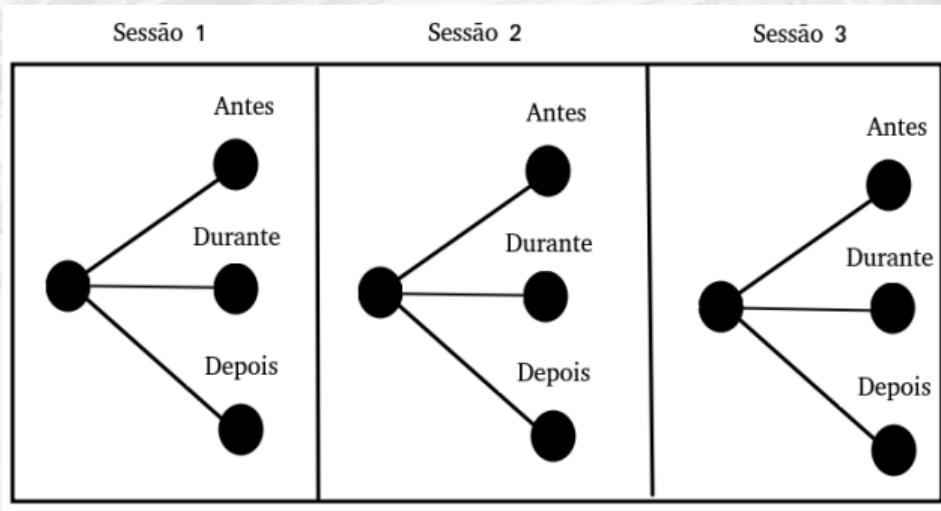


Figura 7. Combinação entre sessão e momento.

GAMLSS

Para definição de um GAMLSS considere:

- ▶ Um conjunto de observações Y_i , $i = 1, 2, \dots, n$.
- ▶ As observações seguem função (densidade) de probabilidade $f_Y(y_i|\theta)$ parametrizada por $\theta = (\theta_1, \dots, \theta_p)^T$.
- ▶ θ é um vetor de p até 4 parâmetros distribucionais denotados por $(\mu, \sigma, \nu, \tau)^T$ se $p = 4$.
- ▶ Geralmente (mas não necessariamente) μ e σ representam os parâmetros de locação e dispersão, enquanto que ν and τ representam parâmetros de forma.

$$g_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} Z_{jk} \gamma_{jk}, \quad k = 1, 2, 3, 4, \quad (1)$$

- ▶ $g_k(\cdot)$ são funções de ligação que relacionam o k -ésimo parâmetro distribucional ao preditor η_k .
- ▶ $\beta_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J'_k k})^T$ é um vetor de parâmetros de regressão de dimensão J'_k ,
- ▶ X_k e Z_{jk} são matrizes de delineamento de ordem $n \times J'_k$ e $n \times q_{jk}$, respectivamente.
- ▶ γ_{jk} é um vetor aleatório de dimensão q_{jk} para os quais assume-se que $\gamma_{jk} \sim N_{q_{jk}}(0, G_{jk}^{-1})$.
- ▶ G_{jk}^{-1} é a inversa generalizada da matriz simétrica $G_{jk} = G_{jk}(\lambda_{jk})$ de ordem $q_{jk} \times q_{jk}$
- ▶ λ_{jk} é um vetor de hiperparâmetros.

- ▶ Os parâmetros de regressão (β_k) e os efeitos aleatórios (γ_{jk}) são usualmente estimados na maximização de uma log-verossimilhança penalizada (l_p), dada por

$$l_p = l - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \gamma'_{jk} G_{jk} \gamma_{jk}, \quad (2)$$

- ▶ $l = \sum_{i=1}^n \log(f(y_i|\theta))$ representa a log-verossimilhança.
- ▶ A log-verossimilhança penalizada se reduz à usual log-verossimilhança quando não há efeitos aleatórios ou termos suavizadores no modelo.

Aplicação

Procedimento de análise

Respostas

1. Número de mudanças de postura de orelha.
2. Proporção do tempo com as orelhas em posição neutra.

Procedimento para o número de mudanças de postura de orelha:

1. Seleção de distribuições compatíveis com o problema.
2. Ajuste de modelos com o mesmo preditor para todas as distribuições.
3. Seleção do modelos via medidas de qualidade de ajuste OU análise gráfica.
4. Busca do menor modelo possível que não difrisse do modelo inicial.
5. Interpretação dos resultados.

Número de mudanças de postura de orelha

Número de mudanças de postura de orelha

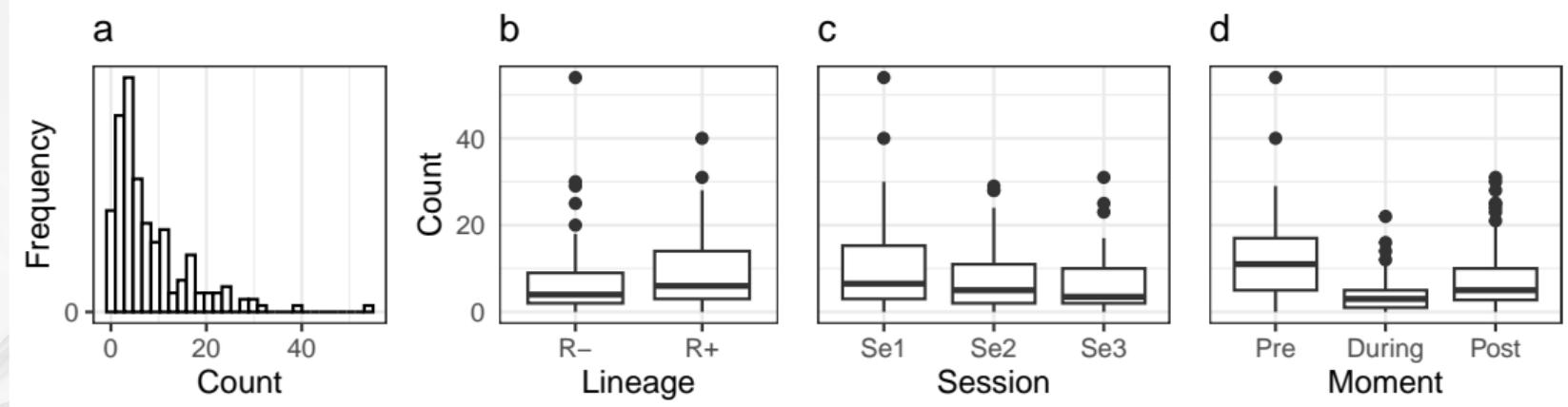


Figura 8. Análise exploratória do número de mudanças de postura de orelha.

Distribuições

- ▶ Poisson (PO).
- ▶ Negative binomial distribution (NBI).
- ▶ Zero inflated Poisson distribution (ZIP).
- ▶ Zero inflated negative binomial distribution (ZINBI).
- ▶ Zero adjusted Poisson distribution (ZAP).
- ▶ Zero adjusted negative binomial distribution (ZANBI).

Preditores iniciais

- ▶ Para cada distribuição os modelos foram especificados com os efeitos fixos de sessão, momento, linhagem e as interações duas a duas.
- ▶ Adicionalmente foram incluídos 2 efeitos aleatórios: um a nível de animal e outro a nível de animal dentro de sessão.
- ▶ Os efeitos aleatórios foram usados devido à necessidade de incorporar as correlações causadas pela estrutura de medidas repetidas (9 medidas por animal).
- ▶ Nas distribuições com parâmetros associados à inflação de zeros foram incluídos os efeitos fixos de sessão, momento e linhagem.

Seleção de modelos

- ▶ A distribuição foi selecionada por meio de medidas de qualidade de ajuste.
- ▶ A distribuição com melhor desempenho foi a ZANBI.
- ▶ Buscou-se o menor modelo que não diferisse do original.
 - ▶ Testou-se a exclusão de variáveis explicativas do modelo.
 - ▶ A comparação entre o modelo original e os reduzidos foi feito por meio do teste da razão de verossimilhanças.

Modelo final

Considere Y_{ijkl} a variável resposta.

- ▶ i ($i = 1, 2, \dots, 20$) representa o animal.
- ▶ j ($j = 1, 2, 3$) representa a sessão.
- ▶ k ($k = 1, 2, 3$) representa o momento.
- ▶ l ($l = 1, 2$) representa a linhagem.

Modelo final

$$Y_{ijkl}|u_j, v_{jk} \sim ZANBI(\mu_{ijkl}, \sigma, v_{ikl}), \quad (3)$$

- ▶ μ_{ijkl} representa o parâmetro de locação da distribuição condicional de y
- ▶ σ representa o parâmetro de dispersão.
- ▶ v_{ikl} está associado ao excesso de zeros.

Modelo final

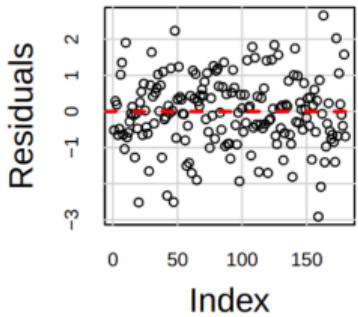
$$\log(\mu_{ijkl}) = \alpha^{(1)} + \beta_j^{(1)} + \gamma_k^{(1)} + \theta_l^{(1)} + u_i + v_{ik} \quad (4)$$

$$\text{logit}(v_{jkl}) = \alpha^{(2)} + \beta_j^{(2)} + \gamma_k^{(2)} + \theta_l^{(2)} \quad (5)$$

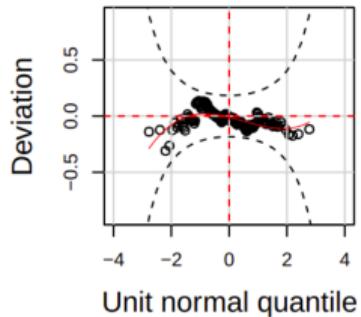
- ▶ $\alpha^{(1)}$ representa o intercepto.
- ▶ $\beta_j^{(1)}$ representa o efeito de sessão.
- ▶ $\gamma_k^{(1)}$ representa o efeito de momento.
- ▶ $\theta_l^{(1)}$ representa o efeito de linhagem.
- ▶ u_i representa o efeito aleatório a nível de animal.
 - ▶ $u_i \sim N(0, \sigma_U^2)$
- ▶ v_{ik} representa o efeito aleatório a nível de animal dentro de sessão.
 - ▶ $v_{ik} \sim N(0, \sigma_V^2)$

Diagnóstico

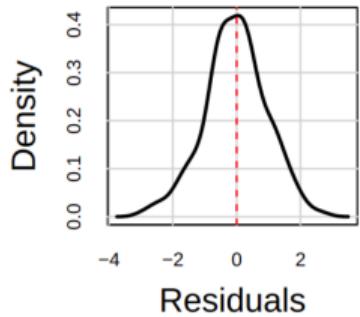
Residuals plot



Worm-Plot



Density



Q-Q Plot

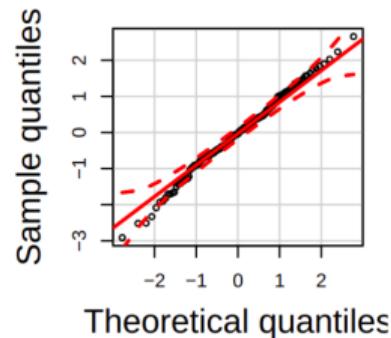


Figura 9. Análise exploratória da proporção do tempo com as orelhas em posição neutra.

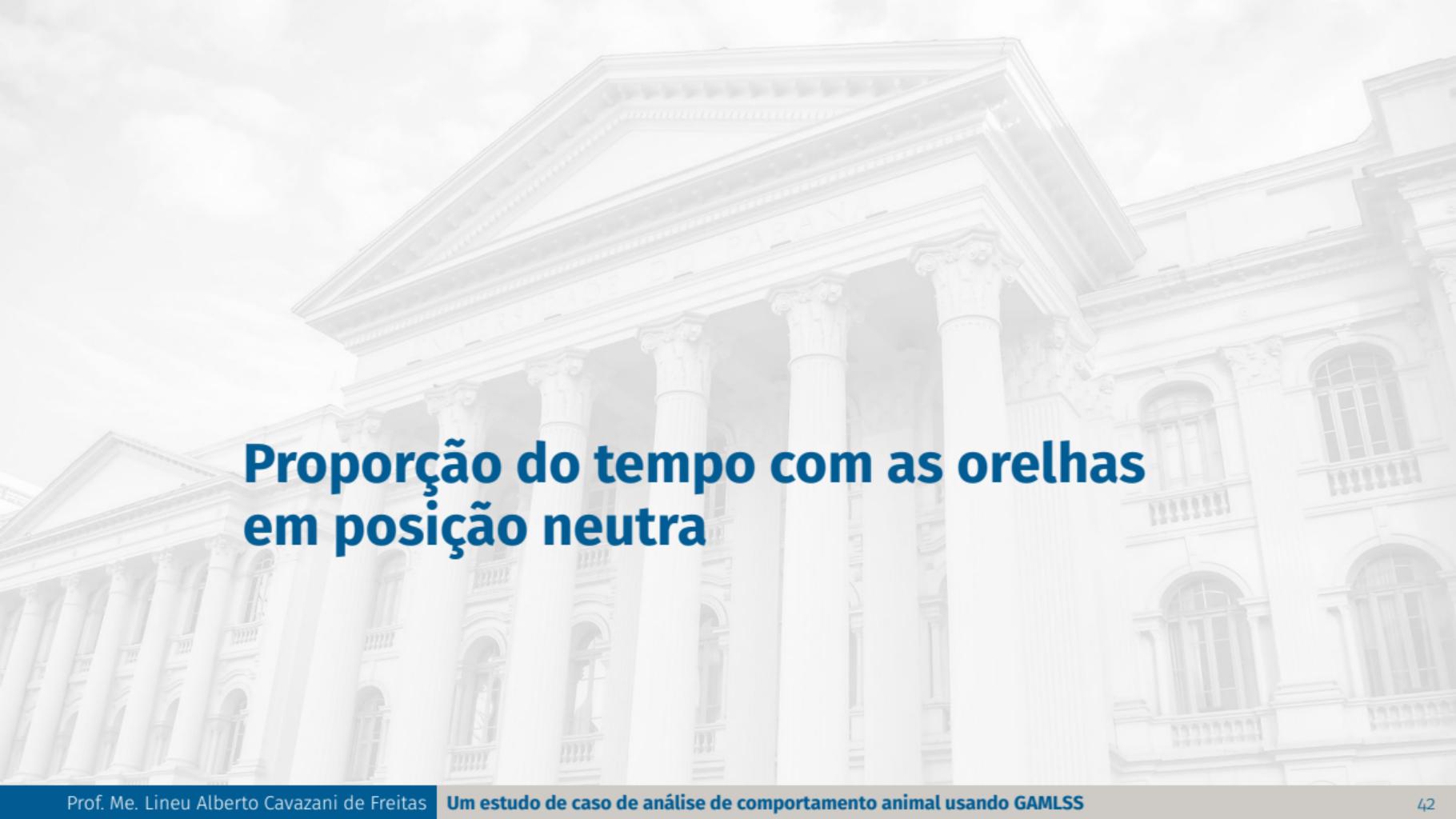
Estimativas

Tabela 1. Relative rates, odds ratios, confidence intervals and p-values for the final ZANBI model. Session 1, moment before brushing and lineage $R-$ are taken as reference categories

Parm.	μ			ν		
	RR	CI(95%)	P-value	OR	CI(95%)	P-value
α	15.77	(13.12; 18.95)	<0.001	0.004	(0.00; 0.06)	<0.001
β_{se2}	0.69	(0.56; 0.84)	<0.001	5.83	(0.64; 53.54)	0.121
β_{se3}	0.66	(0.53; 0.81)	<0.001	14.31	(1.69; 121.49)	0.016
γ_{dur}	0.35	(0.28; 0.43)	<0.001	15.80	(1.89; 131.99)	0.012
γ_{post}	0.65	(0.54; 0.79)	<0.001	4.43	(0.47; 42.12)	0.197
θ_{R+}	1.33	(1.12; 1.57)	0.001	0.33	(0.09; 1.14)	0.081

Interpretação





Proporção do tempo com as orelhas em posição neutra

Proporção do tempo com as orelhas em posição neutra

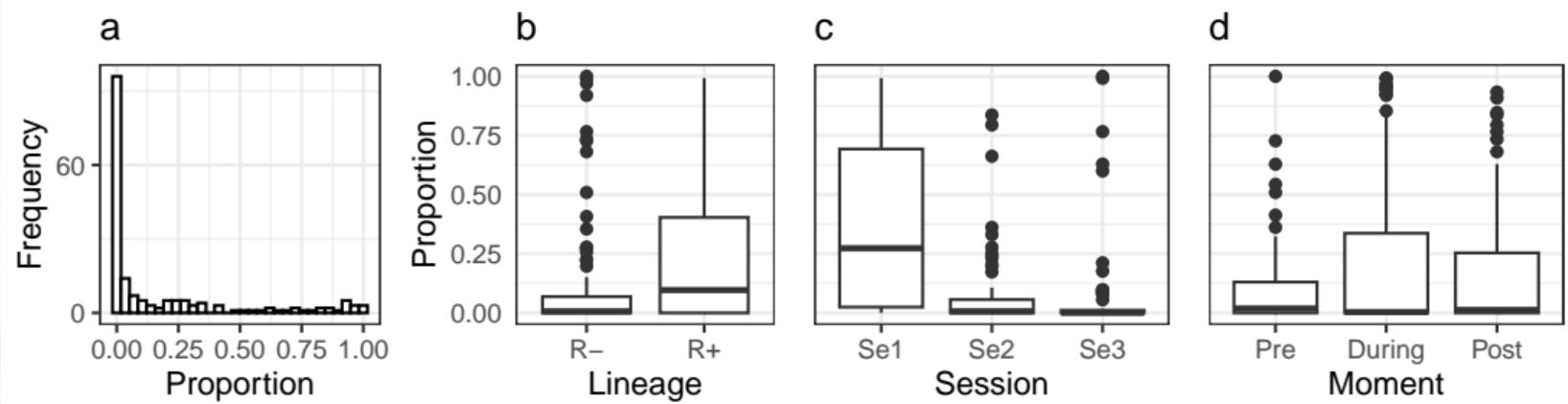


Figura 10. Análise exploratória da proporção do tempo com as orelhas em posição neutra.



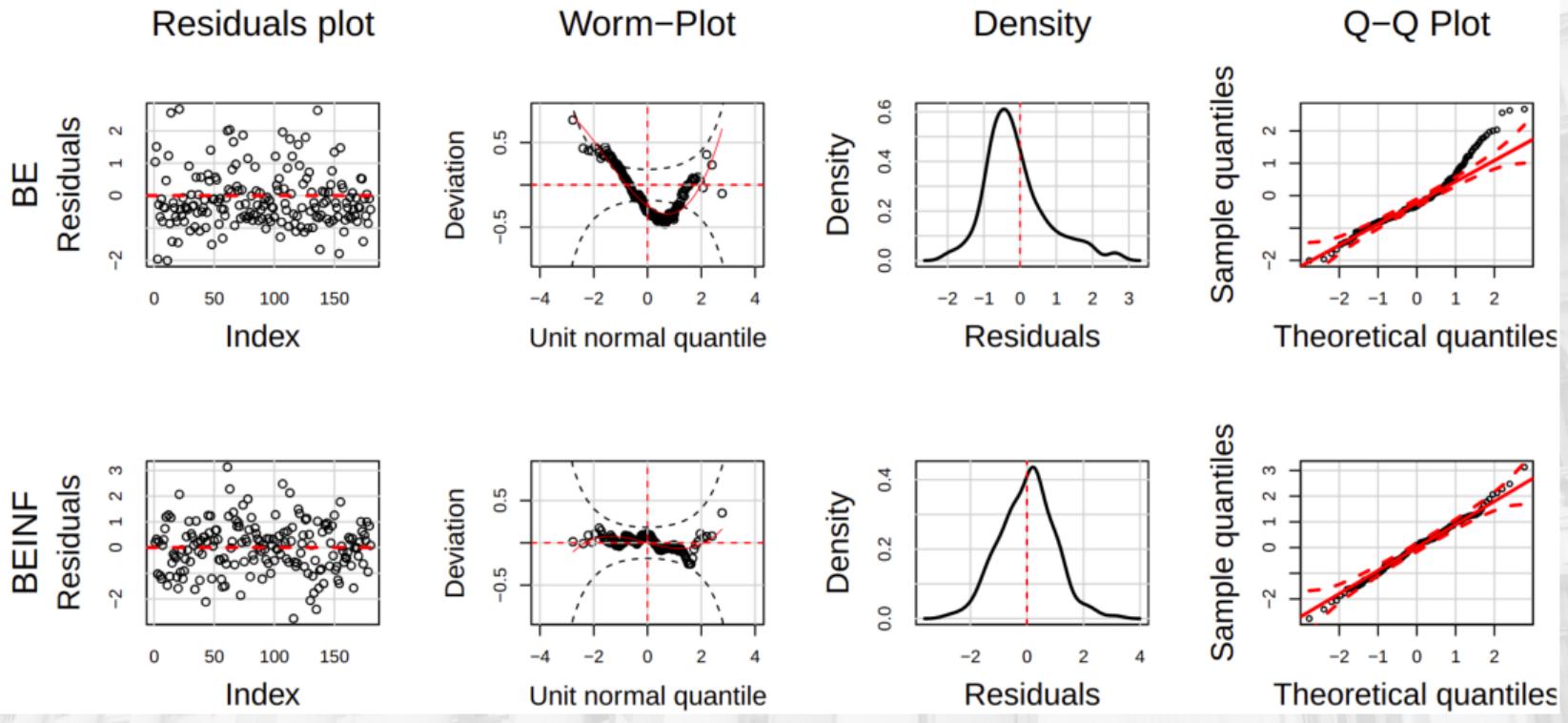


Figura 11. Análise exploratória da proporção do tempo com as orelhas em posição neutra.

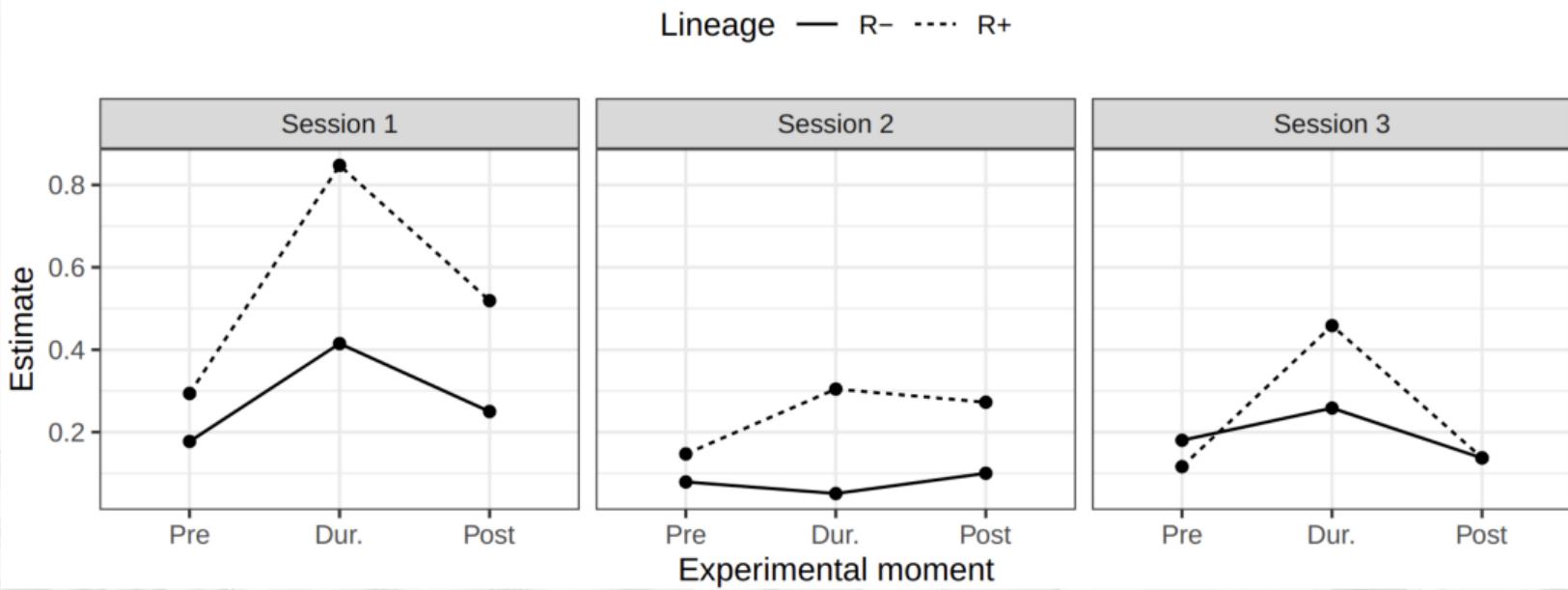


Figura 12. Análise exploratória da proporção do tempo com as orelhas em posição neutra.

Considerações finais

Considerações finais

- ▶ Neste trabalho exploramos a flexibilidade do GAMLSS na análise de dados de comportamento animal.
- ▶ Dois tipos de variáveis comuns a este tipo de estudo foram analisadas: um contagem e uma proporção.
- ▶ No estudo de caso o objetivo era investigar o efeito de isolamento social, intervenção humana e linhagem dos animais.
- ▶ Os dados apresentavam uma estrutura multinível, super-dispersão, inflação de zeros e efeito de covariáveis em mais de um parâmetro.
- ▶ Com GAMLSS foi possível analisar adequadamente ambas as respostas considerando todas as especificidades do problema e provendo resultados confiáveis e consistentes.