

Testes de hipóteses em Modelos Multivariados de Covariância Linear Generalizada

Semana acadêmica do curso de Estatística 2021

Lineu Alberto Cavazani de Freitas

Orientador: Prof. Dr. Wagner Hugo Bonat

Co-orientador: Prof. Dr. Marco Antonio Zanata Alves

PPG Informática



Conteúdo

- 1 Ciência de dados
- 2 Modelos de regressão
- 3 Modelos multivariados de covariância linear generalizada
- 4 Testes de hipóteses
- 5 Proposta
- 6 Resultados preliminares
 - Adaptação do teste Wald para os McGLM
 - Exemplos de hipóteses
 - ANOVA & MANOVA via teste Wald
 - Funções implementadas
- 7 Próximas etapas
- 8 Considerações finais

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

1

Ciência de dados

Ciência de dados

- ▶ **Ciência de dados** é campo de estudo interdisciplinar que incorpora conhecimento de áreas como:
 1. Estatística.
 2. Ciência da computação.
 3. Matemática.
- ▶ Os **métodos estatísticos** são de fundamental importância em grande parte das etapas da ciência de dados (WEIHS; ICKSTADT, 2018).
- ▶ Neste sentido, os **modelos de regressão** tem papel importante.

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

2

Modelos de regressão

Modelos de regressão

- ▶ Técnica utilizada quando há o interesse em verificar a influência das variáveis explicativas sobre variáveis respostas.
- ▶ Assume-se que a variável resposta condicional às explicativas segue uma distribuição de probabilidades.
- ▶ Um modelo de regressão é uma expressão matemática que relaciona a média da variável resposta às variáveis explicativas.
- ▶ Com modelos de regressão é possível:
 - ▶ **Explicar** a associação entre as variáveis.
 - ▶ **Predizer** valores da resposta com base nos valores observados das explicativas.

Modelos de regressão

1. Definição do problema.

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

2. Planejamento do estudo e coleta de dados.

3. Análise dos dados via regressão.

- ▶ Escolha da distribuição de probabilidade.
- ▶ Especificação do modelo.
- ▶ Obtenção das estimativas dos parâmetros (ajuste).
- ▶ Diagnóstico.

4. Interpretação dos resultados.

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

Ciência de dados

Modelos de regressão

Modelos multivariados de covariância linear generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

3

Modelos multivariados de covariância linear generalizada

Modelos multivariados de covariância linear generalizada

- ▶ Os modelos multivariados de covariância linear generalizada (McGLMs) (BONAT; JØRGENSEN, 2016) configuram uma estrutura geral para análise via modelos de regressão.
- ▶ Comporta múltiplas respostas de diferentes naturezas.
- ▶ Pode-se ajustar modelos com diferentes preditores e distribuições para cada resposta.
- ▶ Pode especificar modelos que levam em conta a correlação entre indivíduos do conjunto de dados.

Modelos multivariados de covariância linear generalizada

- ▶ Os parâmetros são interpretáveis:
 - ▶ **Parâmetros de regressão:** efeito das variáveis explicativas sobre as respostas.
 - ▶ **Parâmetros de dispersão:** impacto da correlação entre unidades.
 - ▶ **Parâmetros de potência:** indicativo de qual distribuição se adequa ao problema.
- ▶ A metodologia do McGLM está implementada no pacote *mcglm* (BONAT, 2018) do software R.

Modelos multivariados de covariância linear generalizada

Para definição de um McGLM considere:

- ▶ $\mathbf{Y}_{N \times R} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_R\}$ uma matriz de variáveis resposta.
- ▶ $\mathbf{M}_{N \times R} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_R\}$ uma matriz de valores esperados.
- ▶ \mathbf{X}_r denota uma matriz de delineamento $N \times k_r$.
- ▶ $\boldsymbol{\beta}_r$ denota um vetor $k_r \times 1$ de parâmetros de regressão.

Modelos multivariados de covariância linear generalizada

Considere ainda:

- ▶ Σ_b uma matriz de correlação entre variáveis resposta, de ordem $R \times R$.
- ▶ Σ_r , $r = 1, \dots, R$, a matriz de variância e covariância para cada resposta r , de dimensão $N \times N$:

$$\Sigma_r = V_r (\mu_r; p_r)^{1/2} (\mathbf{\Omega} (\tau_r)) V_r (\mu_r; p_r)^{1/2}.$$

Em que:

- ▶ $V_r (\mu; p)$ é uma matriz diagonal em que as entradas principais são dadas pela função de variância aplicada ao vetor μ .
- ▶ p_r é o parâmetro de potência.
- ▶ $\mathbf{\Omega} (\tau_r)$ a matriz de dispersão que descreve a parte da covariância dentro de cada variável resposta.

Preditor linear matricial

- ▶ A matriz $\mathbf{\Omega}(\boldsymbol{\tau}_r)$ descreve a estrutura de correlação entre as observações da amostra.
- ▶ É modelada através de um preditor linear matricial combinado com uma função de ligação de covariância:

$$h\{\mathbf{\Omega}(\boldsymbol{\tau}_r)\} = \tau_{r0}Z_0 + \dots + \tau_{rD}Z_D$$

- ▶ $h()$ é a função de ligação de covariância.
- ▶ Z_{rd} com $d = 0, \dots, D$ são matrizes que representam a estrutura de covariância presente em cada variável resposta r .
- ▶ $\boldsymbol{\tau}_r = (\tau_{r0}, \dots, \tau_{rD})$ é um vetor $(D + 1) \times 1$ de parâmetros de dispersão.

Funções de variância

1. Função de variância potência (JØRGENSEN, 1987) e (JØRGENSEN, 1997).

- ▶ Família Tweedie de distribuições.
- ▶ $\vartheta(\mu; p) = \mu^p$.
- ▶ Casos particulares: normal ($p = 0$), Poisson ($p = 1$), gama ($p = 2$) e normal inversa ($p = 3$).

2. Função de dispersão Poisson–Tweedie (JØRGENSEN; KOKONENDJI, 2015).

- ▶ Família Poisson-Tweedie de distribuições.
- ▶ $\vartheta(\mu; p) = \mu + \mu^p$.
- ▶ Casos particulares: Hermite ($p = 0$), Neyman tipo A ($p = 1$), binomial negativa ($p = 2$) e Poisson–inversa gaussiana ($p = 3$).

3. Função de variância binomial.

- ▶ $\vartheta(\mu) = \mu(1 - \mu)$.
- ▶ Acomoda respostas binárias ou restritas a um intervalo.

Modelos multivariados de covariância linear generalizada

Os McGLMs são definidos por:

$$E(\mathbf{Y}) = \mathbf{M} = \{g_1^{-1}(\mathbf{X}_1\boldsymbol{\beta}_1), \dots, g_R^{-1}(\mathbf{X}_R\boldsymbol{\beta}_R)\}$$

$$\text{Var}(\mathbf{Y}) = \mathbf{C} = \boldsymbol{\Sigma}_R \overset{G}{\otimes} \boldsymbol{\Sigma}_b$$

Em que:

- ▶ $\boldsymbol{\Sigma}_R \overset{G}{\otimes} \boldsymbol{\Sigma}_b = \text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1, \dots, \tilde{\boldsymbol{\Sigma}}_R)(\boldsymbol{\Sigma}_b \otimes \mathbf{I})\text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1^T, \dots, \tilde{\boldsymbol{\Sigma}}_R^T)$ é o produto generalizado de Kronecker.
- ▶ $\tilde{\boldsymbol{\Sigma}}_r$ denota a matriz triangular inferior da decomposição de Cholesky da matriz $\boldsymbol{\Sigma}_r$.
- ▶ $\text{Bdiag}()$ denota a matriz bloco-diagonal.
- ▶ \mathbf{I} uma matriz identidade $N \times N$.
- ▶ $g_r()$ são as tradicionais funções de ligação.

Modelos multivariados de covariância linear generalizada

- ▶ Parâmetros estimados nos McGLMs:
 1. Regressão.
 2. Dispersão.
 3. Potência.
 4. Correlação.
- ▶ Todas estas quantidades são interpretáveis e são estimadas com base nos dados.
- ▶ A estimação é feita por meio de **funções de estimação**.
 1. **Função quasi-score** para parâmetros de regressão.
 2. **Função de estimação de Pearson** para os demais parâmetros.
- ▶ Mais detalhes são descritos em (BONAT; JØRGENSEN, 2016).

4

Testes de hipóteses

Testes de hipóteses em modelos de regressão

- ▶ Usados para verificar se a retirada de determinada variável explicativa do modelo geraria uma perda no ajuste.
- ▶ Dentre os testes mais comuns está o teste Wald:
 - ▶ Verifica se existe evidência para afirmar que os parâmetros são iguais a valores postulados.
 - ▶ Avalia quão longe o valor estimado está do valor postulado.
 - ▶ As hipóteses podem ser descritas como:

$$\begin{cases} H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{c} \\ H_1 : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{c} \end{cases}$$

- ▶ A estatística de teste é dada por:

$$WT = (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{c})^T (\mathbf{L} \mathbf{Var}^{-1}(\hat{\boldsymbol{\beta}}) \mathbf{L}^T)^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{c}).$$

- ▶ $WT \sim \chi_q^2$.

ANOVA & MANOVA

- ▶ Formas de **avaliar a significância** de cada uma das variáveis de uma forma procedural.
- ▶ Consiste em efetuar testes sucessivos impondo **restrições ao modelo** original.
- ▶ O objetivo é testar se a ausência de determinada variável **gera perda** ao modelo.
- ▶ Os resultados são sumarizados numa tabela, o chamado **quadro de análise de variância**.

Tópicos abordados até aqui...

1. Importância dos modelos de regressão.
2. McGLMs e importância dos parâmetros de regressão, dispersão e potência.
3. Testes de hipóteses em modelos de regressão, ênfase no teste Wald.
4. Procedimentos baseados em testes de hipóteses: ANOVA e MANOVA.

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

5

Proposta

Proposta

- ▶ Considerando os McGLMs, não há discussão a respeito da construção de testes de hipóteses.
- ▶ Considerando o alto potencial de aplicação dos McGLMs em ciência de dados, nossos objetivos gerais são:
 - ▶ Desenvolvimento de testes de hipóteses para avaliação dos parâmetros de McGLMs.
 - ▶ Adaptação do teste de Wald clássico utilizado em modelos lineares para os McGLMs.

Etapas

1. **Adaptar** o teste Wald para realização de testes de hipóteses gerais sobre parâmetros de McGLMs.
2. **Implementar** funções para efetuar tais testes, bem como funções para efetuar ANOVAs e MANOVAs para os McGLMs.
3. **Avaliar** as propriedades e comportamento dos testes propostos com base em estudos de simulação.
4. **Motivar** o potencial de aplicação das metodologias discutidas com base na aplicação a conjuntos de dados reais.

6

Resultados preliminares

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses

ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

TH MCGLM

Lineu Alberto

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

**Adaptação do teste Wald para
os McGLM**

Exemplos de hipóteses

ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

Adaptação do teste Wald para os McGLMs

Hipóteses

$$H_0 : \mathbf{L}\boldsymbol{\theta}_{\beta,\tau,p} = \mathbf{c} \text{ vs } H_1 : \mathbf{L}\boldsymbol{\theta}_{\beta,\tau,p} \neq \mathbf{c}.$$

Em que:

- ▶ Em que \mathbf{L} é a matriz de especificação das hipóteses a serem testadas, tem dimensão $s \times h$.
- ▶ $\boldsymbol{\theta}_{\beta,\tau,p}$ é o vetor de dimensão $h \times 1$ de parâmetros de regressão, dispersão e potência do modelo.
- ▶ \mathbf{c} é um vetor de dimensão $s \times 1$ com os valores sob hipótese nula.

Estatística de teste

$$W = (\mathbf{L}\hat{\boldsymbol{\theta}}_{\beta,\tau,p} - \mathbf{c})^T (\mathbf{L} \mathbf{J}_{\beta,\tau,p}^{-1} \mathbf{L}^T)^{-1} (\mathbf{L}\hat{\boldsymbol{\theta}}_{\beta,\tau,p} - \mathbf{c}).$$

Em que:

- ▶ \mathbf{L} é a matriz da especificação das hipóteses, tem dimensão $s \times h$.
- ▶ $\hat{\boldsymbol{\theta}}_{\beta,\tau,p}$ é o vetor de dimensão $h \times 1$ com todas as estimativas dos parâmetros de regressão, dispersão e potência.
- ▶ \mathbf{c} é um vetor de dimensão $s \times 1$ com os valores sob hipótese nula.
- ▶ $\mathbf{J}_{\beta,\tau,p}^{-1}$ é a inversa da matriz de informação de Godambe desconsiderando os parâmetros de correlação, de dimensão $h \times h$.

Exemplo de hipóteses nos McGLMs

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses

ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

Problema

- ▶ Considere o exemplo disponível na seção 4.3.6 do livro **Modelos de Regressão com Apoio Computacional** (PAULA, 2004):
- ▶ Os dados referem-se a um estudo sobre demanda de TV's a cabo em 40 regiões dos Estados Unidos.
- ▶ Algumas das variáveis coletadas foram:
 - ▶ Número de assinantes de TV a cabo (em milhares).
 - ▶ Percentual de domicílios com TV a cabo.
 - ▶ Renda per capita por domicílio com TV a cabo (em USD).
 - ▶ Custo médio mensal de manutenção de TV a cabo (em USD).

Ciência de dados

Modelos de regressão

Modelos multivariados de covariância linear generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para os McGLM

Exemplos de hipóteses

ANOVA & MANOVA via teste Wald

Funções implementadas

Próximas etapas

Considerações finais

Variáveis

► Variáveis resposta:

1. Número de assinantes de TV a cabo (em milhares).
2. Percentual de domicílios com TV a cabo.

► Variáveis explicativas:

1. Renda per capita por domicílio com TV a cabo (em USD).
2. Custo médio mensal de manutenção de TV a cabo (em USD).

► No problema existem duas variáveis resposta, de diferentes naturezas: uma contagem e uma proporção.

► A estrutura do problema é ideal para utilizar um modelo multivariado.

Para análise do problema, considere um modelo bivariado:

$$g_r(\mu_r) = \beta_{r0} + \beta_{r1}\text{renda} + \beta_{r2}\text{custo} + \beta_{r3}\text{renda} : \text{custo}.$$

Em que:

- ▶ O índice r denota a variável resposta, $r = 1, 2$.
 - ▶ 1: Número de assinantes de TV a cabo (em milhares).
 - ▶ 2: Percentual de domicílios com TV a cabo.
- ▶ β_{r0} denota o intercepto de cada resposta.
- ▶ Temos três parâmetros de regressão para cada resposta:
 1. β_{r1} é o efeito de renda.
 2. β_{r2} é o efeito de custo.
 3. β_{r3} representa o efeito da interação entre as variáveis renda e custo.
- ▶ Considere ainda que:
 - ▶ Cada resposta possui apenas um parâmetro de dispersão: τ_{r0} .
 - ▶ As unidades em estudo são independentes, logo $Z_0 = \mathbf{I}$.
 - ▶ Os parâmetros de potência foram fixados.

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses
ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

Pergunta

Existe efeito de custo médio mensal de manutenção de TV a cabo (em USD) sobre o número de assinantes de TV a cabo (em milhares) E percentual de domicílios com TV a cabo?

$$H_0 : \beta_{r2} = 0 \text{ vs } H_1 : \beta_{r2} \neq 0.$$

Ou, da mesma forma:

$$H_0 : \begin{pmatrix} \beta_{12} \\ \beta_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ vs } H_1 : \begin{pmatrix} \beta_{12} \\ \beta_{22} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Efeito de custo com manutenção sobre número de assinantes e percentual de domicílios com TV a cabo

A hipótese pode ser reescrita na seguinte notação:

$$H_0 : \mathbf{L}\boldsymbol{\theta}_{\beta,\tau,p} = \mathbf{c} \text{ vs } H_1 : \mathbf{L}\boldsymbol{\theta}_{\beta,\tau,p} \neq \mathbf{c}.$$

Em que:

- ▶ $\boldsymbol{\theta}_{\beta,\tau,p}^T = [\beta_{10} \ \beta_{11} \ \beta_{12} \ \beta_{13} \ \beta_{20} \ \beta_{21} \ \beta_{22} \ \beta_{23} \ \tau_{10} \ \tau_{20}]$.
- ▶ $\mathbf{L} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$.
- ▶ $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, é o valor da hipótese nula.

ANOVA & MANOVA via teste Wald

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses

**ANOVA & MANOVA via
teste Wald**

Funções implementadas

Próximas etapas

Considerações finais

ANOVA & MANOVA via teste Wald

- ▶ Com base na adaptação do teste Wald propostas, buscamos propor ANOVAs e MANOVAs via teste Wald.
- ▶ Propomos 3 tipos diferentes de análises de variância, nomeadas tipo I, II e III.
- ▶ Cada linha do quadro corresponde uma hipótese. Portanto, basta especificar uma matriz L .
- ▶ Os procedimentos para análise de variância retornam um quadro para cada resposta.
- ▶ Os procedimentos para análise variância multivariadas retornam um único quadro.

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses

ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

ANOVA & MANOVA tipo II

- ▶ São feitos testes comparando o modelo completo contra o modelo sem determinada variável.
- ▶ Portanto, considerando o exemplo:
 1. Testa se os parâmetros referentes a **renda** são iguais a 0. Ou seja, é avaliado o impacto da retirada de renda do modelo.
 2. Testa se os parâmetros referentes a **custo** são iguais a 0. Ou seja, é avaliado o impacto da retirada de custo do modelo.
 3. Testa se o efeito de **interação** é 0.

Funções implementadas

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses

ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

Funções implementadas

Baseando-nos nas funcionalidades do pacote *car* (FOX; WEISBERG, 2019) e usando nossa adaptação do teste Wald implementamos uma série de funções:

Função	Descrição
<code>mc_linear_hypothesis()</code>	Hipóteses lineares gerais especificadas pelo usuário
<code>mc_anova_I()</code>	ANOVA tipo I
<code>mc_anova_II()</code>	ANOVA tipo II
<code>mc_anova_III()</code>	ANOVA tipo III
<code>mc_manova_I()</code>	MANOVA tipo I
<code>mc_manova_II()</code>	MANOVA tipo II
<code>mc_manova_III()</code>	MANOVA tipo III
<code>mc_anova_disp()</code>	ANOVA tipo III para dispersão
<code>mc_manova_disp()</code>	MANOVA tipo III para dispersão

Tabela 1. Funções implementadas

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Adaptação do teste Wald para
os McGLM

Exemplos de hipóteses

ANOVA & MANOVA via
teste Wald

Funções implementadas

Próximas etapas

Considerações finais

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

7

Próximas etapas

Próximas etapas

- ▶ Propor e implementar procedimentos para realização de testes de **comparações múltiplas**.
- ▶ Adequar os testes para que sejam válidos para diferentes **contrastes**.
- ▶ Avaliar as propriedades e comportamento dos testes propostos com base em **estudos de simulação**.
- ▶ Motivar o potencial de aplicação das metodologias discutidas com base na **aplicação a conjuntos de dados reais**.

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

8

Considerações finais

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais

Considerações finais

- ▶ O McGLM contorna importantes restrições encontradas nas classes clássicas de modelos.
- ▶ Nossa contribuição vai no sentido de fornecer ferramentas para uma melhor interpretação dos parâmetros estimados na classe.
- ▶ Nossa contribuição visa formas de responder sobre:
 1. Importância das variáveis explicativas no problema.
 2. Impacto das medidas correlacionadas no conjunto de dados.
 3. Qual distribuição se adequa ao problema.

Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses








Proposta

Resultados preliminares

Próximas etapas

Considerações finais

Referências bibliográficas I

-  BONAT, W. H. Multiple response variables regression models in R: The mcglm package. **Journal of Statistical Software**, v. 84, n. 4, p. 1–30, 2018.
-  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.
-  FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**. Third. Thousand Oaks CA: Sage, 2019. Disponível em: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
-  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. **ASTA Advances in Statistical Analysis**, Springer, v. 100, n. 1, p. 43–78, 2015.
-  PAULA, G. A. **Modelos de regressão: com apoio computacional**. [S.l.]: IME-USP São Paulo, 2004.

[Ciência de dados](#)[Modelos de regressão](#)[Modelos multivariados
de covariância linear
generalizada](#)[Testes de hipóteses](#)[Proposta](#)[Resultados preliminares](#)[Próximas etapas](#)[Considerações finais](#)

Referências bibliográficas II



WEIHS, C.; ICKSTADT, K. Data science: the impact of statistics. **International Journal of Data Science and Analytics**, Springer, v. 6, n. 3, p. 189–194, 2018.

Obrigado!

Lineu Alberto Cavazani de Freitas

lineuacf@gmail.com

<https://lineu96.github.io/st/>

PPG Informática



Ciência de dados

Modelos de regressão

Modelos multivariados
de covariância linear
generalizada

Testes de hipóteses

Proposta

Resultados preliminares

Próximas etapas

Considerações finais