

CE092- Extensões do modelo de regressão

Modelos generalizados aditivos para locação, escala e forma (GAMLSS)

Cesar Augusto Taconeli

17 de setembro, 2026

O que vamos estudar nesta aula

O que estudaremos nesta aula

- Compreender as limitações de modelos que descrevem apenas a média da resposta;

O que estudaremos nesta aula

- Compreender as limitações de modelos que descrevem apenas a média da resposta;
- Revisar extensões para relações não lineares e dispersão não constante;

O que estudaremos nesta aula

- Compreender as limitações de modelos que descrevem apenas a média da resposta;
- Revisar extensões para relações não lineares e dispersão não constante;
- Introduzir os modelos GAMLSS para locação, escala e forma;

O que estudaremos nesta aula

- Compreender as limitações de modelos que descrevem apenas a média da resposta;
- Revisar extensões para relações não lineares e dispersão não constante;
- Introduzir os modelos GAMLSS para locação, escala e forma;
- Aplicar e comparar diferentes estratégias de regressão em um mesmo conjunto de dados.

Introdução

- Modelos de regressão são usados na análise da relação entre uma variável aleatória de interesse (resposta) e uma ou mais covariáveis por meio de uma função.

Introdução

- Modelos de regressão são usados na análise da relação entre uma variável aleatória de interesse (resposta) e uma ou mais covariáveis por meio de uma função.
- A análise de regressão tem duas finalidades principais:

- Modelos de regressão são usados na análise da relação entre uma variável aleatória de interesse (resposta) e uma ou mais covariáveis por meio de uma função.
- A análise de regressão tem duas finalidades principais:
 - **Exploratória:** analisar o efeito das covariáveis na resposta.

- Modelos de regressão são usados na análise da relação entre uma variável aleatória de interesse (resposta) e uma ou mais covariáveis por meio de uma função.
- A análise de regressão tem duas finalidades principais:
 - **Exploratória:** analisar o efeito das covariáveis na resposta.
 - **Preditiva:** prever o valor desconhecido da resposta para indivíduos com base nos valores das covariáveis.

- Modelos de regressão são usados na análise da relação entre uma variável aleatória de interesse (resposta) e uma ou mais covariáveis por meio de uma função.
- A análise de regressão tem duas finalidades principais:
 - **Exploratória:** analisar o efeito das covariáveis na resposta.
 - **Preditiva:** prever o valor desconhecido da resposta para indivíduos com base nos valores das covariáveis.
- Modelos de regressão, em geral, envolvem um conjunto de parâmetros, que devem ser estimados usando dados amostrais.

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).

Introdução

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).
- Na regressão linear, assumimos que a relação entre a resposta e as covariáveis seja linear, além de um conjunto de suposições adicionais.

Introdução

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).
- Na regressão linear, assumimos que a relação entre a resposta e as covariáveis seja linear, além de um conjunto de suposições adicionais.
- Basicamente, supõe-se que dados os valores das covariáveis, a distribuição da resposta tenha as seguintes propriedades:

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).
- Na regressão linear, assumimos que a relação entre a resposta e as covariáveis seja linear, além de um conjunto de suposições adicionais.
- Basicamente, supõe-se que dados os valores das covariáveis, a distribuição da resposta tenha as seguintes propriedades:
 - Média linear nas covariáveis;

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).
- Na regressão linear, assumimos que a relação entre a resposta e as covariáveis seja linear, além de um conjunto de suposições adicionais.
- Basicamente, supõe-se que dados os valores das covariáveis, a distribuição da resposta tenha as seguintes propriedades:
 - Média linear nas covariáveis;
 - Variância constante;

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).
- Na regressão linear, assumimos que a relação entre a resposta e as covariáveis seja linear, além de um conjunto de suposições adicionais.
- Basicamente, supõe-se que dados os valores das covariáveis, a distribuição da resposta tenha as seguintes propriedades:
 - Média linear nas covariáveis;
 - Variância constante;
 - Independência entre as observações;

- Dentre as diferentes classes de modelos de regressão, originalmente surgiram os modelos de regressão linear (século XIX).
- Na regressão linear, assumimos que a relação entre a resposta e as covariáveis seja linear, além de um conjunto de suposições adicionais.
- Basicamente, supõe-se que dados os valores das covariáveis, a distribuição da resposta tenha as seguintes propriedades:
 - Média linear nas covariáveis;
 - Variância constante;
 - Independência entre as observações;
 - Distribuição normal da resposta.

- Embora amplamente utilizada, a regressão linear baseia-se em suposições que nem sempre são compatíveis com as características dos dados, limitando sua aplicação.

Introdução

- Embora amplamente utilizada, a regressão linear baseia-se em suposições que nem sempre são compatíveis com as características dos dados, limitando sua aplicação.
- Ao longo das últimas décadas, uma variedade de modelos de regressão foi proposta, visando estender as técnicas de análise de regressão para problemas mais gerais.

- Embora amplamente utilizada, a regressão linear baseia-se em suposições que nem sempre são compatíveis com as características dos dados, limitando sua aplicação.
- Ao longo das últimas décadas, uma variedade de modelos de regressão foi proposta, visando estender as técnicas de análise de regressão para problemas mais gerais.
- Dentre as principais classes de modelos de regressão, destacamos:

- Embora amplamente utilizada, a regressão linear baseia-se em suposições que nem sempre são compatíveis com as características dos dados, limitando sua aplicação.
- Ao longo das últimas décadas, uma variedade de modelos de regressão foi proposta, visando estender as técnicas de análise de regressão para problemas mais gerais.
- Dentre as principais classes de modelos de regressão, destacamos:
 - Modelos lineares generalizados;

- Embora amplamente utilizada, a regressão linear baseia-se em suposições que nem sempre são compatíveis com as características dos dados, limitando sua aplicação.
- Ao longo das últimas décadas, uma variedade de modelos de regressão foi proposta, visando estender as técnicas de análise de regressão para problemas mais gerais.
- Dentre as principais classes de modelos de regressão, destacamos:
 - Modelos lineares generalizados;
 - Modelos generalizados aditivos;

- Embora amplamente utilizada, a regressão linear baseia-se em suposições que nem sempre são compatíveis com as características dos dados, limitando sua aplicação.
- Ao longo das últimas décadas, uma variedade de modelos de regressão foi proposta, visando estender as técnicas de análise de regressão para problemas mais gerais.
- Dentre as principais classes de modelos de regressão, destacamos:
 - Modelos lineares generalizados;
 - Modelos generalizados aditivos;
 - Modelos generalizados aditivos para localização, escala e forma...

- Na prática, qual modelo de regressão devo usar?

- Na prática, qual modelo de regressão devo usar?
- Fundamental levar em conta as características dos dados, em especial a distribuição da variável resposta e sua relação com as demais variáveis.

- Algumas questões pertinentes na identificação do modelo a ser usado.

- Algumas questões pertinentes na identificação do modelo a ser usado.
 - Qual a escala da variável resposta? Trata-se de uma variável discreta ou contínua? Se discreta, é uma variável dicotômica? Politômica? De contagem? Se contínua, tem distribuição simétrica? Normal? Assimétrica...

- Algumas questões pertinentes na identificação do modelo a ser usado.
 - Qual a escala da variável resposta? Trata-se de uma variável discreta ou contínua? Se discreta, é uma variável dicotômica? Politômica? De contagem? Se contínua, tem distribuição simétrica? Normal? Assimétrica...
 - A relação entre a resposta e as covariáveis é linear? Caso seja não linear, há algum modelo disponível na literatura capaz de descrever essa relação?

- Algumas questões pertinentes na identificação do modelo a ser usado.
 - Qual a escala da variável resposta? Trata-se de uma variável discreta ou contínua? Se discreta, é uma variável dicotômica? Politômica? De contagem? Se contínua, tem distribuição simétrica? Normal? Assimétrica...
 - A relação entre a resposta e as covariáveis é linear? Caso seja não linear, há algum modelo disponível na literatura capaz de descrever essa relação?
 - Os dados exibem algum tipo de correlação? Foram coletados num mesmo indivíduo em diferentes momentos (correlação temporal)? Ou coletados em domicílios vizinhos de certa localidade (correlação espacial)?

Estudo de caso - Modelagem dos preços de aluguel de imóveis

Dados sobre preços de aluguel em Munique, 1980

- A base de dados dispõe de informações de 1969 imóveis disponíveis para aluguel em nove variáveis, das quais cinco serão usadas na análise:

Dados sobre preços de aluguel em Munique, 1980

- A base de dados dispõe de informações de 1969 imóveis disponíveis para aluguel em nove variáveis, das quais cinco serão usadas na análise:
 - R: variável resposta, é o aluguel mensal (em marcos alemães) menos o custo utilitário do imóvel;

Dados sobre preços de aluguel em Munique, 1980

- A base de dados dispõe de informações de 1969 imóveis disponíveis para aluguel em nove variáveis, das quais cinco serão usadas na análise:
 - R: variável resposta, é o aluguel mensal (em marcos alemães) menos o custo utilitário do imóvel;
 - Fl: Área construída, em metros quadrados;

Dados sobre preços de aluguel em Munique, 1980

- A base de dados dispõe de informações de 1969 imóveis disponíveis para aluguel em nove variáveis, das quais cinco serão usadas na análise:
 - R: variável resposta, é o aluguel mensal (em marcos alemães) menos o custo utilitário do imóvel;
 - Fl: Área construída, em metros quadrados;
 - A: Ano de construção;

Dados sobre preços de aluguel em Munique, 1980

- A base de dados dispõe de informações de 1969 imóveis disponíveis para aluguel em nove variáveis, das quais cinco serão usadas na análise:
 - R: variável resposta, é o aluguel mensal (em marcos alemães) menos o custo utilitário do imóvel;
 - Fl: Área construída, em metros quadrados;
 - A: Ano de construção;
 - H: Fator com dois níveis, se o imóvel tem (0) ou não (1) aquecimento central;

Dados sobre preços de aluguel em Munique, 1980

- A base de dados dispõe de informações de 1969 imóveis disponíveis para aluguel em nove variáveis, das quais cinco serão usadas na análise:
 - R: variável resposta, é o aluguel mensal (em marcos alemães) menos o custo utilitário do imóvel;
 - Fl: Área construída, em metros quadrados;
 - A: Ano de construção;
 - H: Fator com dois níveis, se o imóvel tem (0) ou não (1) aquecimento central;
 - loc: um fator que classifica a locação do imóvel como abaixo da média (1), na média (2) ou acima da média (3).

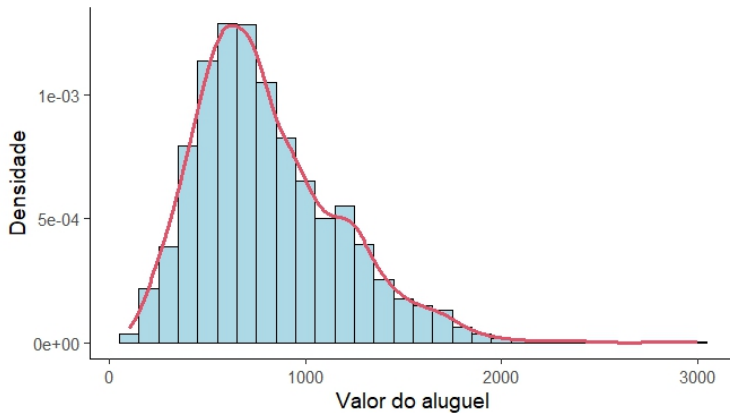


Figura 1: Distribuição dos valores de aluguel

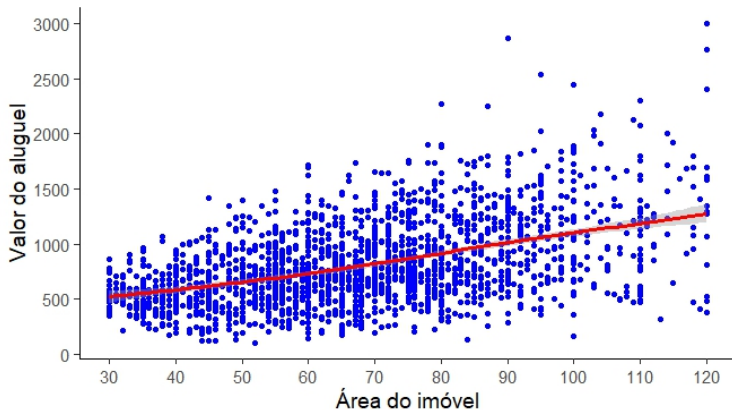


Figura 2: Valor do aluguel vs área do imóvel

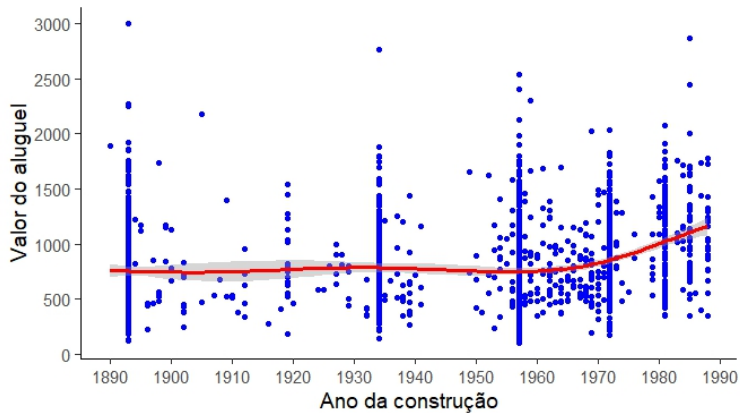


Figura 3: Valor do aluguel vs ano de construção

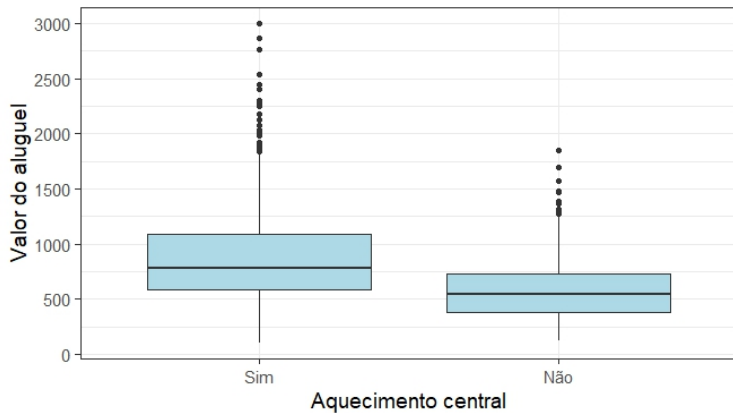


Figura 4: Valor do aluguel vs aquecimento central

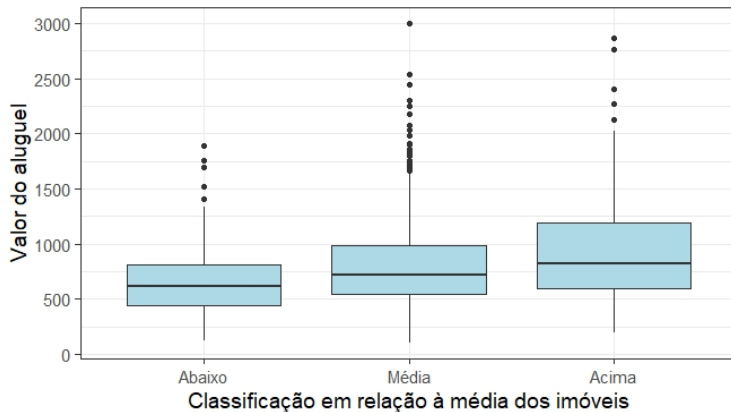


Figura 5: Valor do aluguel vs classificação do imóvel

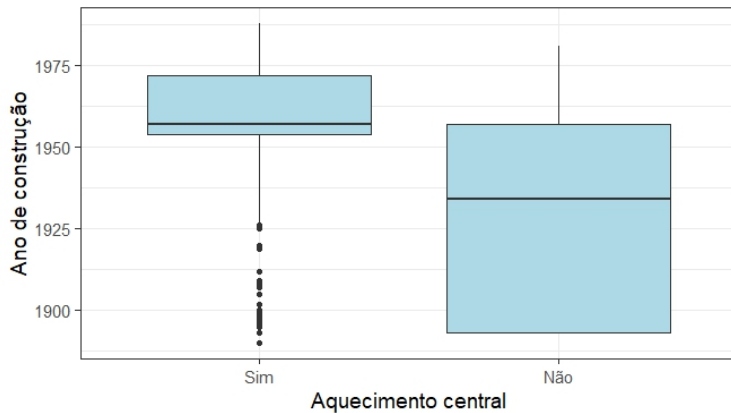


Figura 6: Ano de construção vs aquecimento central

O que os dados parecem exigir do modelo?

- **Resposta positiva e assimétrica:** distribuição mais flexível;

O que os dados parecem exigir do modelo?

- **Resposta positiva e assimétrica:** distribuição mais flexível;
- **Relações não lineares:** suavizadores;

O que os dados parecem exigir do modelo?

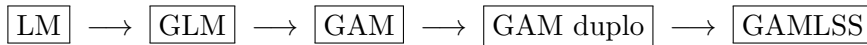
- **Resposta positiva e assimétrica:** distribuição mais flexível;
- **Relações não lineares:** suavizadores;
- **Dispersão variável:** modelagem do parâmetro de escala;

O que os dados parecem exigir do modelo?

- **Resposta positiva e assimétrica:** distribuição mais flexível;
- **Relações não lineares:** suavizadores;
- **Dispersão variável:** modelagem do parâmetro de escala;
- **Assimetria aparentemente variável:** eventualmente modelagem da forma.

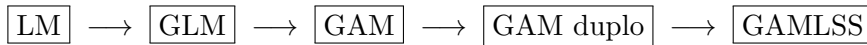
Sequência da análise

- Na análise desses dados vamos visitar uma sequência de modelos, com nível crescente de complexidade:



Sequência da análise

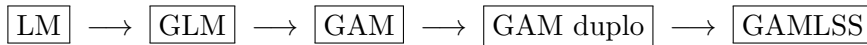
- Na análise desses dados vamos visitar uma sequência de modelos, com nível crescente de complexidade:



- LM:** média linear + normalidade + variância constante;

Sequência da análise

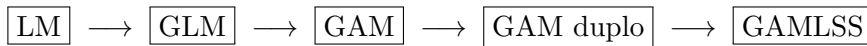
- Na análise desses dados vamos visitar uma sequência de modelos, com nível crescente de complexidade:



- LM:** média linear + normalidade + variância constante;
- GLM:** flexibiliza a distribuição da resposta;

Sequência da análise

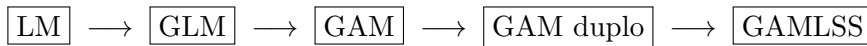
- Na análise desses dados vamos visitar uma sequência de modelos, com nível crescente de complexidade:



- **LM:** média linear + normalidade + variância constante;
- **GLM:** flexibiliza a distribuição da resposta;
- **GAM:** flexibiliza também a forma da relação com as covariáveis;

Sequência da análise

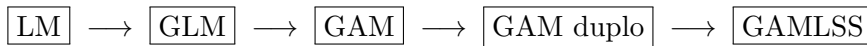
- Na análise desses dados vamos visitar uma sequência de modelos, com nível crescente de complexidade:



- **LM:** média linear + normalidade + variância constante;
- **GLM:** flexibiliza a distribuição da resposta;
- **GAM:** flexibiliza também a forma da relação com as covariáveis;
- **GAM duplo:** média + dispersão dependem das covariáveis;

Sequência da análise

- Na análise desses dados vamos visitar uma sequência de modelos, com nível crescente de complexidade:



- **LM:** média linear + normalidade + variância constante;
- **GLM:** flexibiliza a distribuição da resposta;
- **GAM:** flexibiliza também a forma da relação com as covariáveis;
- **GAM duplo:** média + dispersão dependem das covariáveis;
- **GAMLSS:** permite modelar vários parâmetros da distribuição.

Regressão linear

Modelo de regressão linear

- Para o modelo de regressão linear assumimos que a variável resposta, condicional aos valores das covariáveis, tem distribuição normal.

Modelo de regressão linear

- Para o modelo de regressão linear assumimos que a variável resposta, condicional aos valores das covariáveis, tem distribuição normal.
- A média da distribuição pode ser expressa como uma combinação linear de covariáveis e parâmetros associados.

Modelo de regressão linear

- Para o modelo de regressão linear assumimos que a variável resposta, condicional aos valores das covariáveis, tem distribuição normal.
- A média da distribuição pode ser expressa como uma combinação linear de covariáveis e parâmetros associados.
- A variância, no entanto, é assumida constante, para quaisquer valores das covariáveis sob estudo.

Modelo de regressão linear

- Considere y a variável resposta e $x_1, x_2, \dots, x_k$ um conjunto de k covariáveis.

Modelo de regressão linear

- Considere y a variável resposta e $x_1, x_2, \dots, x_k$ um conjunto de k covariáveis.
- O modelo de regressão linear fica definido, por:

$$y|\mathbf{x} \stackrel{ind}{\sim} \text{Normal}(\mu_{\mathbf{x}}, \sigma^2),$$

em que

$$\mu_{\mathbf{x}} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Modelo de regressão linear

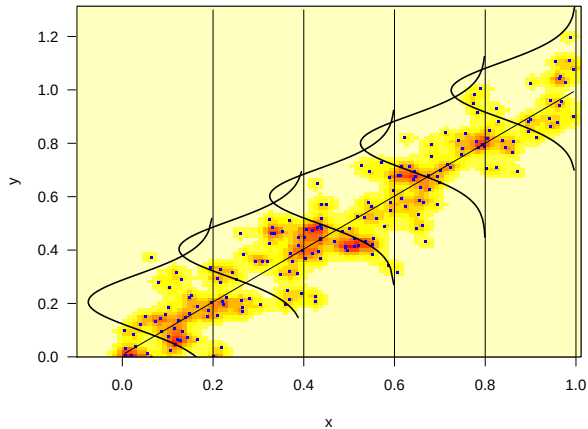


Figura 7: Ilustração de modelo de regressão linear com uma covariável x .

Modelo de regressão linear

- Voltando à análise dos dados de aluguéis de imóveis, o seguinte modelo de regressão linear é proposto:

$$R|x \sim \text{Normal}(\mu_x, \sigma^2),$$

Modelo de regressão linear

- Voltando à análise dos dados de aluguéis de imóveis, o seguinte modelo de regressão linear é proposto:

$$R|x \sim \text{Normal}(\mu_x, \sigma^2),$$

em que

$$\begin{aligned}\mu_x = & \beta_0 + \beta_1 \times \text{Área} + \beta_2 \times \text{Ano const.} + \beta_3 \times I(\text{Aquec: não}) \\ & + \beta_4 \times I(\text{Loc: média}) + \beta_5 \times I(\text{Loc: acima}),\end{aligned}$$

sendo $I(\cdot)$ a função indicadora, tal que, por exemplo, $I(\text{Aquec:não}) = 1$ para os imóveis sem aquecimento central e $I(\text{Aquec:não}) = 0$ para os imóveis com aquecimento central.

- O modelo de regressão linear ajustado tem a seguinte expressão:

$$\hat{\mu}_x = -2775.04 + 8.84 \times \text{Área} + 1.48 \times \text{Ano const.} - 204.76 \times I(\text{Aquec: não}) \\ + 134.05 \times I(\text{Loc: média}) + 209.58 \times I(\text{Loc: acima}).$$

Modelo de regressão linear

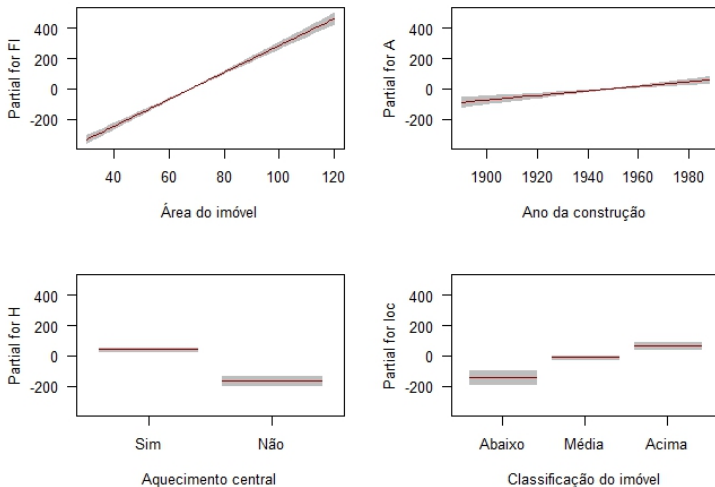


Figura 8: Gráfico de efeitos- regressão linear

Regressão linear

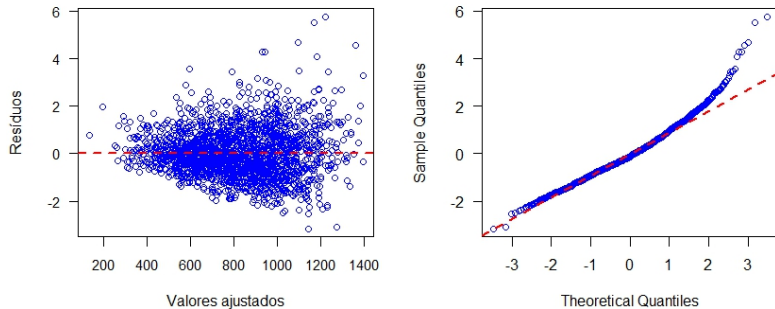


Figura 9: Análise de resíduos- regressão linear

Modelo linear generalizado

Modelo linear generalizado

- Modelos lineares generalizados configuram extensões dos modelos de regressão linear.

Modelo linear generalizado

- Modelos lineares generalizados configuram extensões dos modelos de regressão linear.
- Dentre seus principais diferenciais, destaca-se a possibilidade de modelar respostas com distribuições diferentes da normal, como exemplo:

Modelo linear generalizado

- Modelos lineares generalizados configuram extensões dos modelos de regressão linear.
- Dentre seus principais diferenciais, destaca-se a possibilidade de modelar respostas com distribuições diferentes da normal, como exemplo:
 - Binomial;

Modelo linear generalizado

- Modelos lineares generalizados configuram extensões dos modelos de regressão linear.
- Dentre seus principais diferenciais, destaca-se a possibilidade de modelar respostas com distribuições diferentes da normal, como exemplo:
 - Binomial;
 - Poisson;

Modelo linear generalizado

- Modelos lineares generalizados configuram extensões dos modelos de regressão linear.
- Dentre seus principais diferenciais, destaca-se a possibilidade de modelar respostas com distribuições diferentes da normal, como exemplo:
 - Binomial;
 - Poisson;
 - Gama;

Modelo linear generalizado

- Modelos lineares generalizados configuram extensões dos modelos de regressão linear.
- Dentre seus principais diferenciais, destaca-se a possibilidade de modelar respostas com distribuições diferentes da normal, como exemplo:
 - Binomial;
 - Poisson;
 - Gama;
 - Normal inversa, dentre outras.

Modelo linear generalizado

- Uma variável aleatória y tem distribuição pertencente à família exponencial se sua função (densidade) de probabilidade puder ser expressa na seguinte forma:

$$f(y; \mu, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\},$$

em que θ e ϕ são parâmetros canônico e de dispersão, respectivamente.

Modelo linear generalizado

- Uma variável aleatória y tem distribuição pertencente à família exponencial se sua função (densidade) de probabilidade puder ser expressa na seguinte forma:

$$f(y; \mu, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\},$$

em que θ e ϕ são parâmetros canônico e de dispersão, respectivamente.

- A média e a variância de y são dadas, respectivamente, por $E(y) = b'(\theta)$ e $\text{Var}(y) = \phi V(\mu)$, onde $V(\mu) = b''[\theta(\mu)]$ é a chamada *função de variância*.

Modelo linear generalizado

- Dentre as principais distribuições contempladas pela teoria de MLG estão a normal ($V(\mu) = 1$), binomial ($V(\mu) = \mu(1 - \mu)$), Poisson ($V(\mu) = \mu$), gama ($V(\mu) = \mu^2$) e normal inversa ($V(\mu) = \mu^3$).

Modelo linear generalizado

- Dentre as principais distribuições contempladas pela teoria de MLG estão a normal ($V(\mu) = 1$), binomial ($V(\mu) = \mu(1 - \mu)$), Poisson ($V(\mu) = \mu$), gama ($V(\mu) = \mu^2$) e normal inversa ($V(\mu) = \mu^3$).
- Um modelo linear generalizado pode ser representado, genericamente, da seguinte forma:

$$y|\mathbf{x} \stackrel{ind}{\sim} f(\mu_{\mathbf{x}}, \phi),$$

em que $f(\cdot)$ denota uma particular distribuição da família exponencial, ϕ é o parâmetro de dispersão e

$$g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

Modelo linear generalizado

- Na aplicação dos dados sobre os valores de aluguel de imóveis, vamos considerar a distribuição gama, que é uma alternativa para a modelagem de dados contínuos assimétricos.

Modelo linear generalizado

- Na aplicação dos dados sobre os valores de aluguel de imóveis, vamos considerar a distribuição gama, que é uma alternativa para a modelagem de dados contínuos assimétricos.
- Um modelo linear generalizado com distribuição gama para a resposta fica definido por:

$$\mathbf{R}|\mathbf{x} \stackrel{ind}{\sim} \text{gama}(\mu_{\mathbf{x}}, \phi), \quad \mu > 0, \phi > 0,$$

Modelo linear generalizado

- Na aplicação dos dados sobre os valores de aluguel de imóveis, vamos considerar a distribuição gama, que é uma alternativa para a modelagem de dados contínuos assimétricos.
- Um modelo linear generalizado com distribuição gama para a resposta fica definido por:

$$\mathbf{R}|\mathbf{x} \stackrel{ind}{\sim} \text{gama}(\mu_{\mathbf{x}}, \phi), \quad \mu > 0, \phi > 0,$$

com

$$\begin{aligned} \log(\mu_{\mathbf{x}}) = & \beta_0 + \beta_1 \times \text{Área} + \beta_2 \times \text{Ano const.} + \beta_3 \times I(\text{Aquec: não}) \\ & + \beta_4 \times I(\text{Loc: média}) + \beta_5 \times I(\text{Loc: acima}). \end{aligned}$$

Modelo linear generalizado

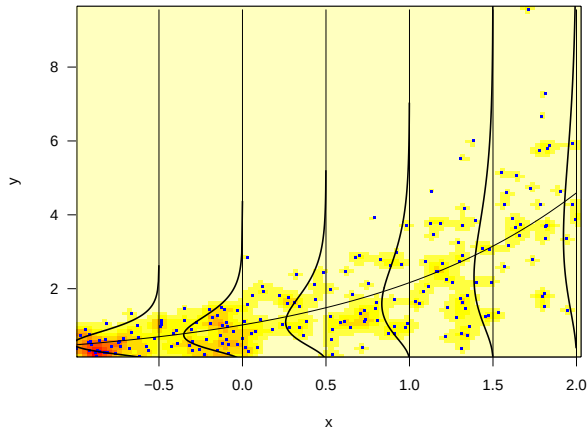


Figura 10: Ilustração de modelo de regressão com resposta gama e função de ligação logarítmica.

- O ajuste do modelo linear generalizado com resposta gama resulta em:

$$\log(\hat{\mu}_x) = 2.8649 + 0.0106 \times \text{Área} + 0.0015 \times \text{Ano const.} - 0.3001 \times I(\text{Aquec: não}) \\ + 0.1907 \times I(\text{Loc: média}) + 0.2641 \times I(\text{Loc: acima}),$$

Modelo linear generalizado

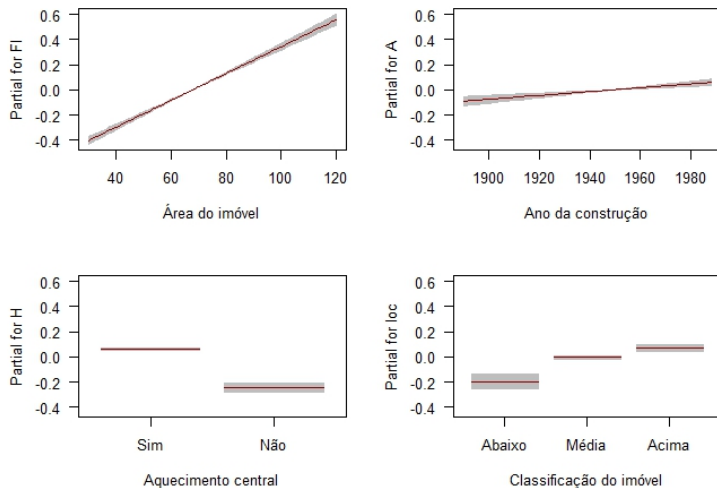


Figura 11: Gráfico de efeitos- modelo linear generalizado gama

Modelo linear generalizado

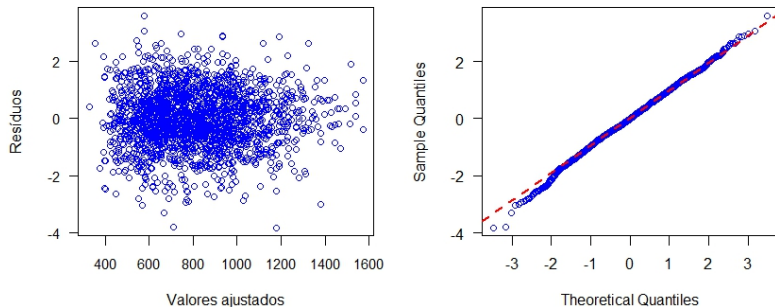


Figura 12: Análise de resíduos- modelo linear generalizado gama

Modelo generalizado aditivo

- Modelos aditivos permitem explicar relações não lineares entre covariáveis e a resposta.

Modelo generalizado aditivo

- Modelos aditivos permitem explicar relações não lineares entre covariáveis e a resposta.
- Os efeitos das variáveis inseridas ao preditor por meio de funções suaves podem ser interpretados usando gráficos para os efeitos.

Modelo generalizado aditivo

- Modelos aditivos permitem explicar relações não lineares entre covariáveis e a resposta.
- Os efeitos das variáveis inseridas ao preditor por meio de funções suaves podem ser interpretados usando gráficos para os efeitos.
- Na aplicação dos valores de aluguel vamos considerar a inclusão de efeitos aditivos (não paramétricos) para a área e o ano de construção do imóvel.

Modelo generalizado aditivo

- Modelos generalizados aditivos configuram extensões dos MLGs em que o efeito de ao menos uma das covariáveis é incorporado ao preditor linear através de uma função suavizadora (não paramétrica).

Modelo generalizado aditivo

- Modelos generalizados aditivos configuram extensões dos MLGs em que o efeito de ao menos uma das covariáveis é incorporado ao preditor linear através de uma função suavizadora (não paramétrica).
- Um modelo generalizado aditivo pode ser representado, de forma geral, por:

$$y|\mathbf{x} \stackrel{ind}{\sim} f(\mu_{\mathbf{x}}, \phi),$$

em que

$$g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + s_{j+1}(x_{j+1}) + \dots + s_k(x_k),$$

em que s_q é uma função suave não paramétrica aplicada à variável x_q , $q = j + 1, \dots, k$.

Modelo generalizado aditivo

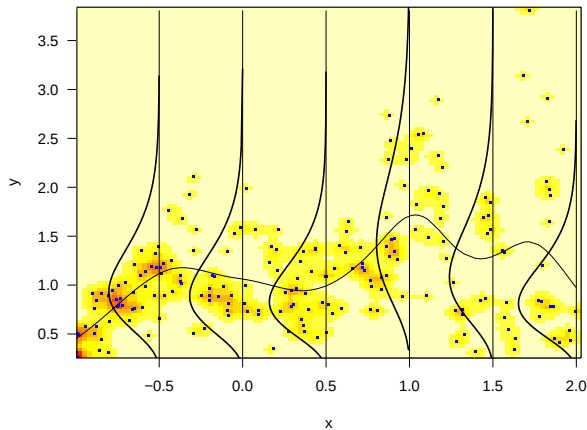


Figura 13: Modelo de regressão com resposta gama e suavizador para a covariável

Modelo generalizado aditivo

- O modelo generalizado aditivo a ser ajustado aos dados de preços de aluguel, considerando resposta gama, é especificado da seguinte forma:

$$R|\mathbf{x} \stackrel{ind}{\sim} \text{gama}(\mu_{\mathbf{x}}, \phi), \quad \mu > 0, \phi > 0,$$

Modelo generalizado aditivo

- O modelo generalizado aditivo a ser ajustado aos dados de preços de aluguel, considerando resposta gama, é especificado da seguinte forma:

$$\mathbf{R}|\mathbf{x} \stackrel{ind}{\sim} \text{gama}(\mu_{\mathbf{x}}, \phi), \quad \mu > 0, \phi > 0,$$

com

$$\begin{aligned} \log(\mu_{\mathbf{x}}) = & \beta_0 + s_1(Fl) + s_2(A) + \beta_1 \times I(\text{Aquec: não}) \\ & + \beta_2 \times I(\text{Loc: média}) + \beta_3 \times I(\text{Loc: acima}), \end{aligned}$$

em que s_1 e s_2 são suavizadores não paramétricos aplicados às variáveis **F1** e **A** (ano de construção), respectivamente.

- O modelo ajustado fica dado por:

$$\begin{aligned}\log(\hat{\mu}_x) = & 3.0851 + s_1(Fl) + s_2(A) - 0.3008 \times I(\text{Aquec: não}) \\ & + 0.1887 \times I(\text{Loc: média}) + 0.2720 \times I(\text{Loc: acima}),\end{aligned}$$

Modelo generalizado aditivo

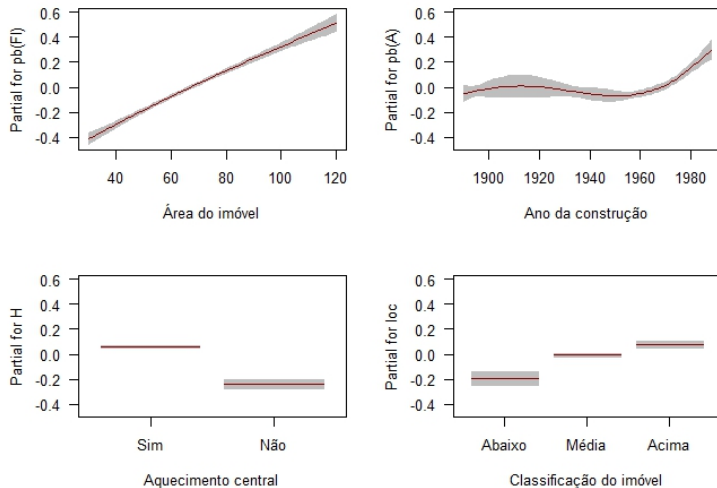


Figura 14: Gráfico de efeito- modelo generalizado aditivo gama

Modelo generalizado aditivo

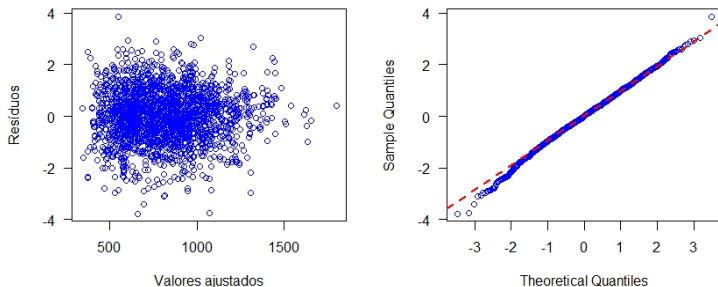


Figura 15: Análise de resíduos- modelo generalizado aditivo gama

Modelo generalizado aditivo duplo

Modelo generalizado aditivo duplo

- Até o momento consideramos apenas modelos em que a média da distribuição varia conforme os valores das covariáveis.

Modelo generalizado aditivo duplo

- Até o momento consideramos apenas modelos em que a média da distribuição varia conforme os valores das covariáveis.
- Modelos mais gerais permitem incluir covariáveis também na modelagem de outros parâmetros da distribuição (por exemplo, para o parâmetro de dispersão).

Modelo generalizado aditivo duplo

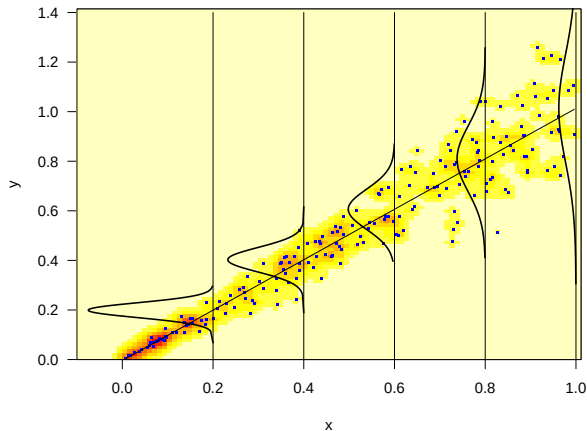


Figura 16: Ilustração - distribuição normal com média e variância dependentes de uma covariável x .

Modelo generalizado aditivo duplo

- Uma formulação geral para o modelo, neste caso, seria da seguinte forma:

$$y|\mathbf{x} \stackrel{ind}{\sim} f(\mu_{\mathbf{x}}, \phi_{\mathbf{x}}),$$

onde

$$\begin{aligned} g_1(\mu_{\mathbf{x}}) &= \beta_{10} + \beta_{11}x_1 + \dots + \beta_{1j_1}x_{j_1} + \dots + s_{j_1+1}(x_{j_1+1}) + \dots + s_{r_1}(x_{k_1}) \\ g_2(\phi_{\mathbf{x}}) &= \beta_{20} + \beta_{21}x_1 + \dots + \beta_{2j_2}x_{j_2} + \dots + s_{j_2+1}(x_{j_2+1}) + \dots + s_{r_2}(x_{k_2}), \end{aligned}$$

em que f representa alguma distribuição de probabilidades em que seus dois parâmetros (μ e ϕ) são funções de covariáveis, inseridas no preditor por meio de termos paramétricos ou não paramétricos.

Modelo generalizado aditivo duplo

- Vamos retomar a análise dos dados de preços de aluguel com o ajuste de modelos com resposta gama e normal inversa, inserindo covariáveis também na modelagem do parâmetro de dispersão (ϕ).

Modelo generalizado aditivo duplo

- Vamos retomar a análise dos dados de preços de aluguel com o ajuste de modelos com resposta gama e normal inversa, inserindo covariáveis também na modelagem do parâmetro de dispersão (ϕ).
- Em ambos os casos vamos considerar função de ligação logarítmica tanto para μ quanto para ϕ , produzindo:

Modelo generalizado aditivo duplo

- Vamos retomar a análise dos dados de preços de aluguel com o ajuste de modelos com resposta gama e normal inversa, inserindo covariáveis também na modelagem do parâmetro de dispersão (ϕ).
- Em ambos os casos vamos considerar função de ligação logarítmica tanto para μ quanto para ϕ , produzindo:

$$\begin{aligned}\log(\mu_{\mathbf{x}}) = & \beta_{10} + s_{11}(Fl) + s_{12}(A) + \beta_{11} \times I(\text{Aquec: não}) \\ & + \beta_{12} \times I(\text{Loc: média}) + \beta_{13} \times I(\text{Loc: acima}),\end{aligned}$$

e

$$\begin{aligned}\log(\phi_{\mathbf{x}}) = & \beta_{20} + s_{21}(Fl) + s_{22}(A) + \beta_{21} \times I(\text{Aquec: não}) \\ & + \beta_{22} \times I(\text{Loc: média}) + \beta_{23} \times I(\text{Loc: acima}).\end{aligned}$$

Modelo generalizado aditivo duplo

- O modelo duplo com resposta de distribuição gama, ajustado aos dados de valores de aluguel, é o seguinte:

$$\log(\hat{\mu}_{\mathbf{x}}) = 2.8844 + s_{11}(Fl) + s_{12}(A) - 0.2918 \times I(\text{Aquec: não}) \\ + 0.1938 \times I(\text{Loc: média}) + 0.2734 \times I(\text{Loc: acima}),$$

com

$$\log(\hat{\phi}_{\mathbf{x}}) = 5.9225 + s_{21}(Fl) + s_{22}(A) + 0.0659 \times I(\text{Aquec: não}) \\ - 0.1166 \times I(\text{Loc: média}) - 0.1702 \times I(\text{Loc: acima}).$$

Modelo generalizado aditivo duplo

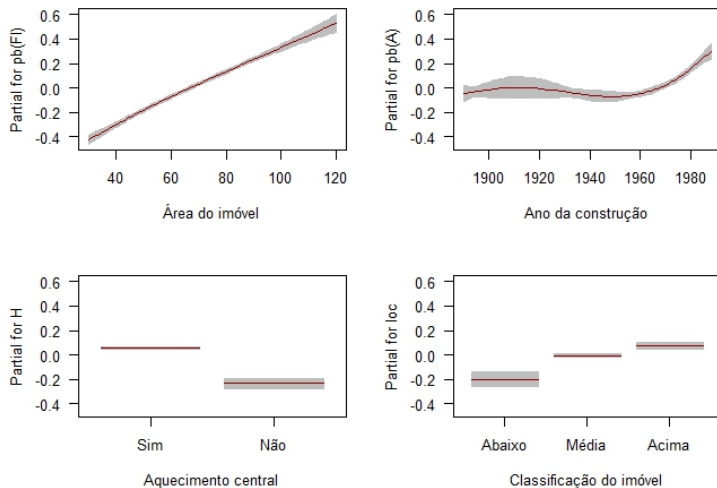


Figura 17: Gráfico de efeito- modelo generalizado aditivo duplo gama (μ)

Modelo generalizado aditivo duplo

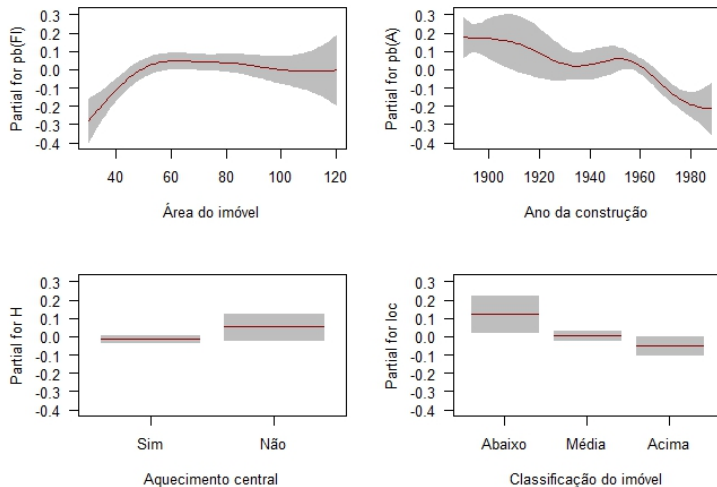


Figura 18: Gráfico de efeito- modelo generalizado aditivo duplo gama (ϕ)

Modelo generalizado aditivo duplo

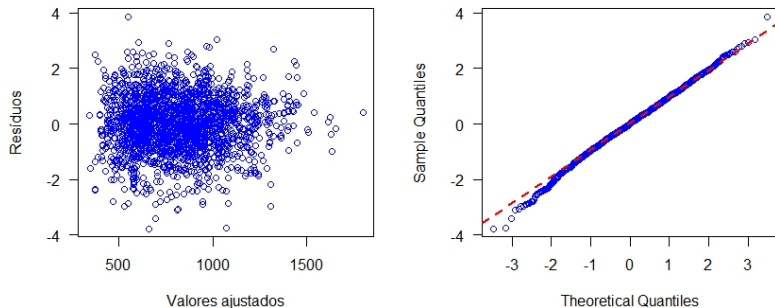


Figura 19: Análise de resíduos- modelo generalizado aditivo duplo gama

Modelo generalizado aditivo para locação, escala e forma (GAMLSS)

Modelo generalizado aditivo para locação, escala e forma

- Os modelos considerados anteriormente baseiam-se em distribuições com dois parâmetros, associados à locação e à escala da distribuição da variável resposta.

Modelo generalizado aditivo para locação, escala e forma

- Os modelos considerados anteriormente baseiam-se em distribuições com dois parâmetros, associados à locação e à escala da distribuição da variável resposta.
- A metodologia GAMLSS contempla distribuições de probabilidade com até quatro parâmetros, permitindo modelar parâmetros associados à locação, escala e forma em função de covariáveis.

Modelo generalizado aditivo para locação, escala e forma

- Os modelos considerados anteriormente baseiam-se em distribuições com dois parâmetros, associados à locação e à escala da distribuição da variável resposta.
- A metodologia GAMLSS contempla distribuições de probabilidade com até quatro parâmetros, permitindo modelar parâmetros associados à locação, escala e forma em função de covariáveis.
- Desta forma, dispõe-se de uma ampla variedade de modelos para a análise de dados com distribuições de diferentes formas.

Modelo generalizado aditivo para locação, escala e forma

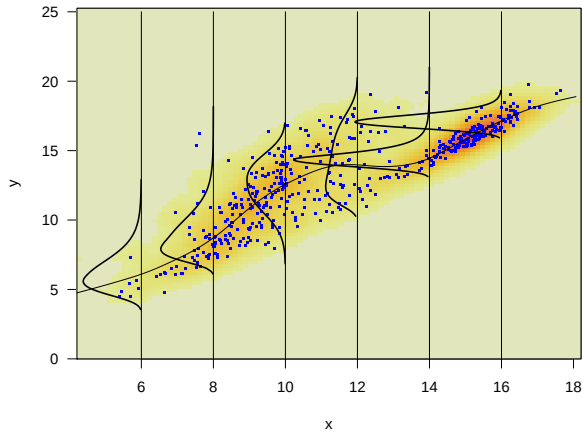


Figura 20: Ilustração - distribuição com locação, dispersão e forma dependentes de x .

Modelo generalizado aditivo para localização, escala e forma

- Um modelo generalizado aditivo para localização, escala e forma é definido, de forma geral, por:

$$y|\mathbf{x} \stackrel{ind}{\sim} D(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}}, \nu_{\mathbf{x}}, \tau_{\mathbf{x}}),$$

onde

$$\begin{aligned} g_1(\mu_{\mathbf{x}}) &= \beta_{10} + \beta_{11}x_1 + \dots + \beta_{1j_1}x_{j_1} + \dots + s_{j_1+1}(x_{j_1+1}) + \dots + s_{k_1}(x_{k_1}) \\ g_2(\sigma_{\mathbf{x}}) &= \beta_{20} + \beta_{21}x_1 + \dots + \beta_{2j_2}x_{j_2} + \dots + s_{j_2+1}(x_{j_2+1}) + \dots + s_{k_2}(x_{k_2}) \\ g_3(\nu_{\mathbf{x}}) &= \beta_{30} + \beta_{31}x_1 + \dots + \beta_{3j_3}x_{j_3} + \dots + s_{j_3+1}(x_{j_3+1}) + \dots + s_{k_3}(x_{k_3}), \\ g_4(\tau_{\mathbf{x}}) &= \beta_{40} + \beta_{41}x_1 + \dots + \beta_{4j_4}x_{j_4} + \dots + s_{j_4+1}(x_{j_4+1}) + \dots + s_{k_4}(x_{k_4}) \end{aligned}$$

em que $D(\mu, \sigma, \nu, \tau)$ é uma distribuição de (até) quatro parâmetros e ν e τ são parâmetros de forma.

Modelo generalizado aditivo para locação, escala e forma

- Dando sequência à análise dos dados de preços de aluguel, vamos considerar, como alternativa à distribuição gama, a distribuição Box-Cox Cole e Green (BCCGo).

Modelo generalizado aditivo para locação, escala e forma

- Dando sequência à análise dos dados de preços de aluguel, vamos considerar, como alternativa à distribuição gama, a distribuição Box-Cox Cole e Green (BCCGo).
- A distribuição BCCGo tem três parâmetros (μ , σ e ν), sendo ν o parâmetro de forma da distribuição.

Modelo generalizado aditivo para locação, escala e forma

- Dando sequência à análise dos dados de preços de aluguel, vamos considerar, como alternativa à distribuição gama, a distribuição Box-Cox Cole e Green (BCCGo).
- A distribuição BCCGo tem três parâmetros (μ , σ e ν), sendo ν o parâmetro de forma da distribuição.
- Cada um dos parâmetros pode ou não ser modelado em função de covariáveis. Além disso, diferentes conjuntos de variáveis podem ser incluídos para cada parâmetro.

Modelo generalizado aditivo para locação, escala e forma

- Dando sequência à análise dos dados de preços de aluguel, vamos considerar, como alternativa à distribuição gama, a distribuição Box-Cox Cole e Green (BCCGo).
- A distribuição BCCGo tem três parâmetros (μ , σ e ν), sendo ν o parâmetro de forma da distribuição.
- Cada um dos parâmetros pode ou não ser modelado em função de covariáveis. Além disso, diferentes conjuntos de variáveis podem ser incluídos para cada parâmetro.
- **Nota:** Note que usamos uma especificação deliberadamente flexível para ilustrar as possibilidades da classe GAMLSS; em uma análise aplicada, a estrutura de cada parâmetro deve ser selecionada e validada.

Modelo generalizado aditivo para locação, escala e forma

- O seguinte modelo foi especificado para os valores de imóveis:

$$R|x \stackrel{ind}{\sim} BCCGo(\mu_x, \sigma_x, \nu_x),$$

Modelo generalizado aditivo para locação, escala e forma

- O seguinte modelo foi especificado para os valores de imóveis:

$$\mathbf{R}|\mathbf{x} \stackrel{ind}{\sim} BCCGo(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}}, \nu_{\mathbf{x}}),$$

onde

$$\begin{aligned} \log(\mu_{\mathbf{x}}) = & \beta_{10} + s_{11}(Fl) + s_{12}(A) + \beta_{11} \times I(\text{Aquec: não}) \\ & + \beta_{12} \times I(\text{Loc: média}) + \beta_{13} \times I(\text{Loc: acima}), \end{aligned}$$

$$\begin{aligned} \log(\sigma_{\mathbf{x}}) = & \beta_{20} + s_{21}(Fl) + s_{22}(A) + \beta_{21} \times I(\text{Aquec: não}) \\ & + \beta_{22} \times I(\text{Loc: média}) + \beta_{23} \times I(\text{Loc: acima}), \end{aligned}$$

$$\begin{aligned} \nu_{\mathbf{x}} = & \beta_{30} + s_{31}(Fl) + s_{32}(A) + \beta_{31} \times I(\text{Aquec: não}) \\ & + \beta_{32} \times I(\text{Loc: média}) + \beta_{33} \times I(\text{Loc: acima}). \end{aligned}$$

Modelo generalizado aditivo para locação, escala e forma

- Como resultado temos o seguinte modelo ajustado:

$$\begin{aligned}\log(\hat{\mu}_{\mathbf{x}}) = & 2.0285 + s_{11}(Fl) + s_{12}(A) - 0.3213 \times I(\text{Aquec: não}) \\ & + 0.1853 \times I(\text{Loc: média}) + 0.2742 \times I(\text{Loc: acima}),\end{aligned}$$

$$\begin{aligned}\log(\hat{\sigma}_{\mathbf{x}}) = & 6.6534 + s_{21}(Fl) + s_{22}(A) + 0.0819 \times I(\text{Aquec: não}) \\ & - 0.0851 \times I(\text{Loc: média}) - 0.1410 \times I(\text{Loc: acima}),\end{aligned}$$

$$\begin{aligned}\hat{\nu}_{\mathbf{x}} = & -3.1601 + s_{31}(Fl) + s_{32}(A) - 0.2866 \times I(\text{Aquec: não}) \\ & - 0.1711 \times I(\text{Loc: média}) - 0.0845 \times I(\text{Loc: acima}).\end{aligned}$$

Modelo generalizado aditivo

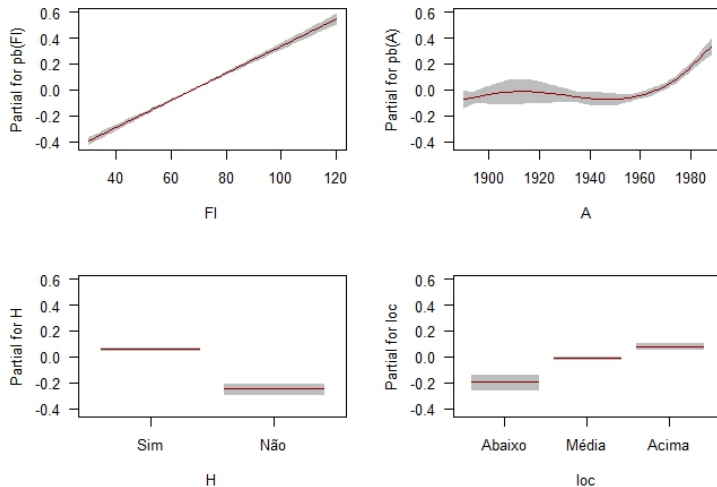


Figura 21: Gráfico de efeitos- GAMLSS (μ)

Modelo generalizado aditivo

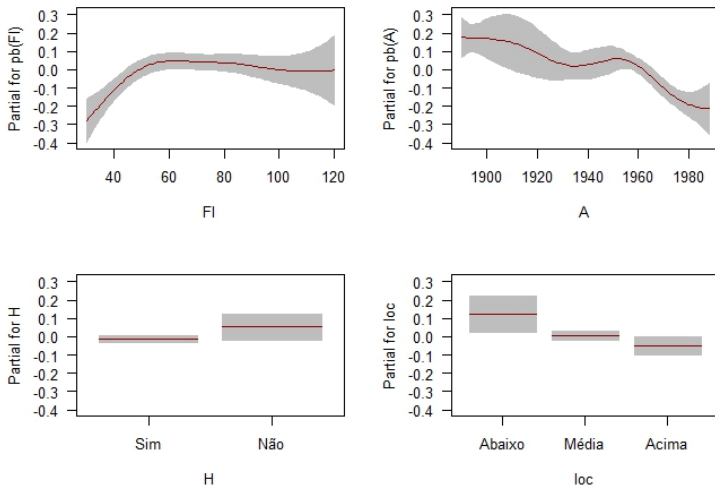


Figura 22: Gráfico de efeitos- GAMLSS (σ)

Modelo generalizado aditivo

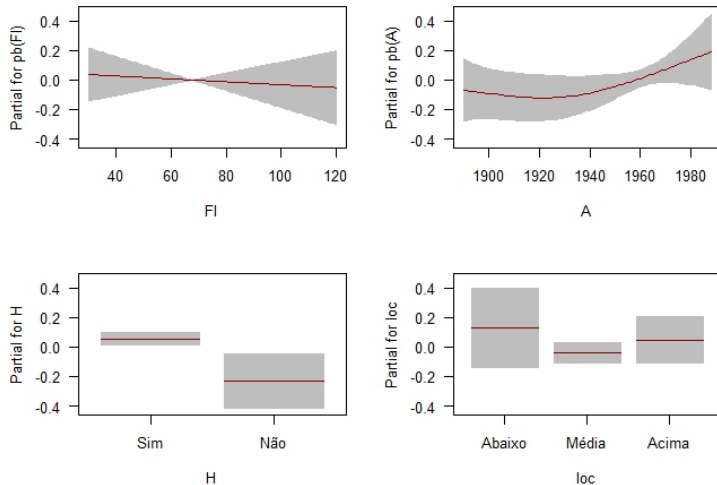


Figura 23: Gráfico de efeitos- GAMLSS (ν)

Modelo generalizado aditivo para locação, escala e forma

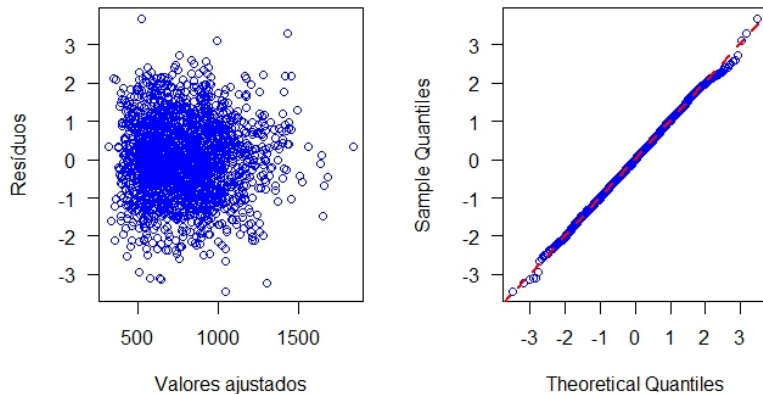


Figura 24: Análise de resíduos- GAMLSS

Comparação dos modelos

- A avaliação e comparação dos modelos devem considerar:

- A avaliação e comparação dos modelos devem considerar:
 - Adequação da distribuição;

- A avaliação e comparação dos modelos devem considerar:
 - Adequação da distribuição;
 - Estrutura dos resíduos;

- A avaliação e comparação dos modelos devem considerar:
 - Adequação da distribuição;
 - Estrutura dos resíduos;
 - Capacidade de representar relações não lineares;

- A avaliação e comparação dos modelos devem considerar:
 - Adequação da distribuição;
 - Estrutura dos resíduos;
 - Capacidade de representar relações não lineares;
 - Necessidade de modelar escala e forma;

- A avaliação e comparação dos modelos devem considerar:
 - Adequação da distribuição;
 - Estrutura dos resíduos;
 - Capacidade de representar relações não lineares;
 - Necessidade de modelar escala e forma;
 - Complexidade do modelo;

- A avaliação e comparação dos modelos devem considerar:
 - Adequação da distribuição;
 - Estrutura dos resíduos;
 - Capacidade de representar relações não lineares;
 - Necessidade de modelar escala e forma;
 - Complexidade do modelo;
 - Critérios de informação.

Tabela 1: Critério de informação de Akaike para os diferentes modelos ajustados

Modelo	AIC
Regressão linear	28173
Modelo linear generalizado	27778
Modelo generalizado aditivo	27705
Modelo generalizado aditivo duplo	27614
Modelo generalizado aditivo para locação, escala e forma	27609

O que estudamos nesta aula

O que estudamos nesta aula

- Partimos das limitações do modelo linear e ampliamos a distribuição da resposta por meio dos MLGs;

O que estudamos nesta aula

- Partimos das limitações do modelo linear e ampliamos a distribuição da resposta por meio dos MLGs;
- Utilizamos modelos aditivos para representar relações não lineares;

O que estudamos nesta aula

- Partimos das limitações do modelo linear e ampliamos a distribuição da resposta por meio dos MLGs;
- Utilizamos modelos aditivos para representar relações não lineares;
- Permitimos que localização, escala e forma variem de acordo com as covariáveis;

O que estudamos nesta aula

- Partimos das limitações do modelo linear e ampliamos a distribuição da resposta por meio dos MLGs;
- Utilizamos modelos aditivos para representar relações não lineares;
- Permitimos que localização, escala e forma variem de acordo com as covariáveis;
- Comparamos modelos considerando ajuste, diagnóstico e parcimônia.

Conclusão

Conclusão

- Lembre-se: os dados demandam o modelo de regressão, e não o modelo se impõe aos dados.

Conclusão

- Lembre-se: os dados demandam o modelo de regressão, e não o modelo se impõe aos dados.
- A modelagem deve se pautar no princípio da parcimônia, buscando-se, na medida do possível, um modelo simples que explique *honestamente* os dados.

- Lembre-se: os dados demandam o modelo de regressão, e não o modelo se impõe aos dados.
- A modelagem deve se pautar no princípio da parcimônia, buscando-se, na medida do possível, um modelo simples que explique *honestamente* os dados.
- Não existe modelo de regressão único para qualquer problema.

Conclusão

- Lembre-se: os dados demandam o modelo de regressão, e não o modelo se impõe aos dados.
- A modelagem deve se pautar no princípio da parcimônia, buscando-se, na medida do possível, um modelo simples que explique *honestamente* os dados.
- Não existe modelo de regressão único para qualquer problema.

All models are wrong but some are useful

George Box