

Exemplo Clientes da Loja

Gilberto A. Paula

Departamento de Estatística
IME-USP, Brasil
giapaula@ime.usp.br

2^o Semestre 2023

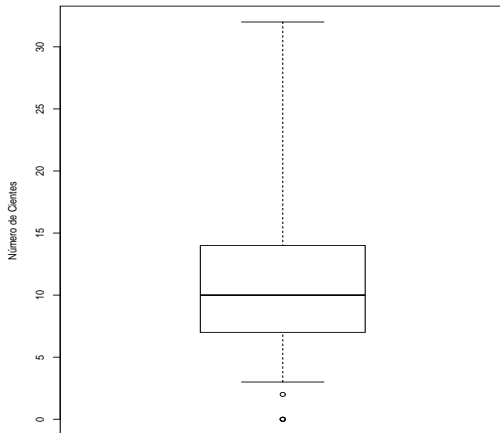
- 1 Clientes da Loja
- 2 Ajuste Modelo de Poisson
- 3 Resultados Modelo Ajustado
- 4 Ajuste Modelo de Poisson Parcial Linear
- 5 Resultados Modelo Ajustado
- 6 Referências

Descrição dos Dados

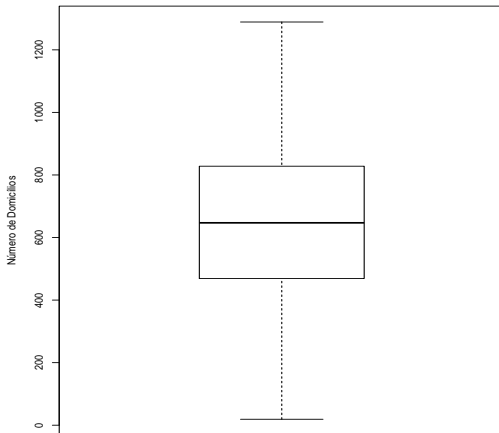
Considere os dados apresentados em Neter et al. (1996) sobre o perfil dos clientes de uma determinada loja oriundos de 110 áreas de uma cidade. O objetivo principal do estudo é relacionar o número esperado de clientes de cada área (**nclientes**) com as seguintes variáveis explicativas em cada área:

- número de domicílios (em mil) (**domic**),
- renda média anual (em mil USD) (**renda**),
- idade média dos domicílios (em anos) (**idade**),
- distância ao concorrente mais próximo (em milhas) (**dist1**),
- distância à loja (em milhas) (**dist2**) (1 milha = 1609,344m).

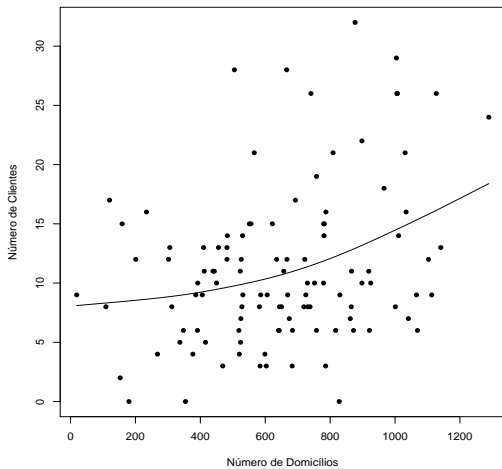
Boxplot Robusto Número de Clientes



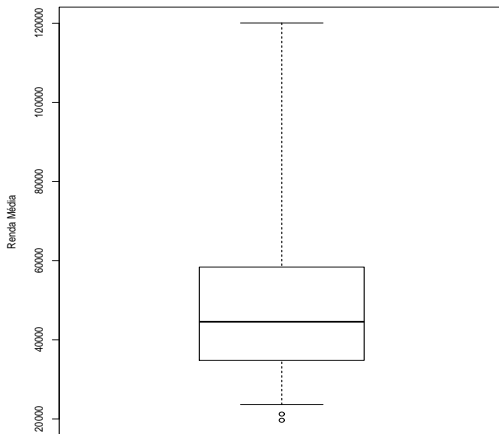
Boxplot Robusto Número de Domicílios



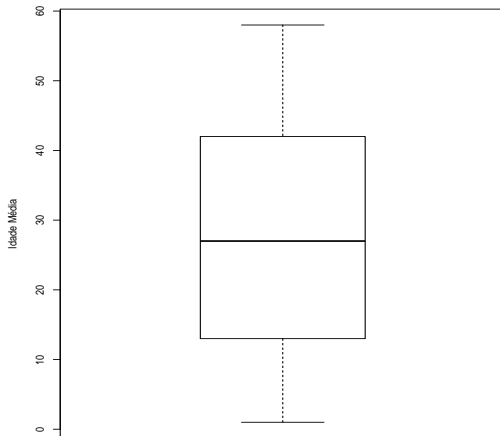
Dispersão Clientes versus Domicílios



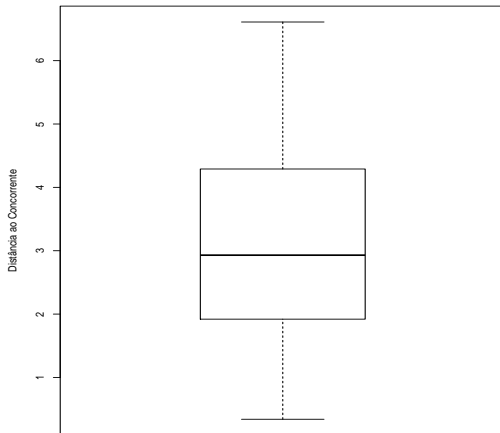
Boxplot Robusto Renda Média



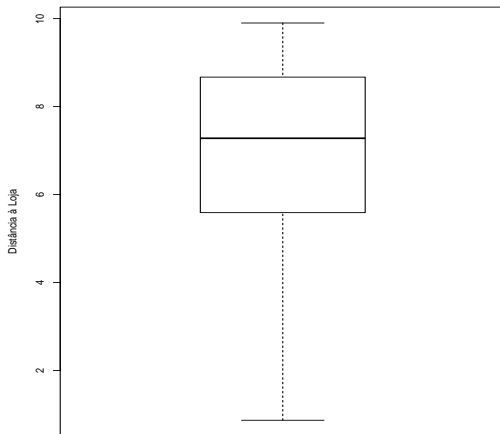
Boxplot Robusto Idade Média



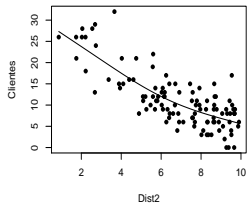
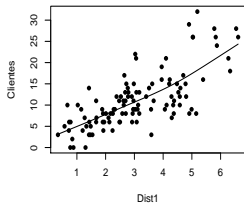
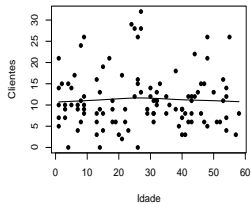
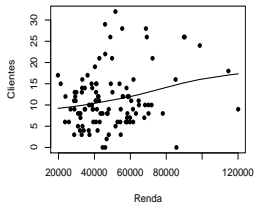
Boxplot Robusto Distância ao Concorrente



Boxplot Robusto Distância à Loja



Diagramas de Dispersão



Comentários

- Nota-se distribuição assimétrica à direita para o número de clientes, renda média domiciliar e distância ao concorrente mais próximo. Tem-se distribuição assimétrica à esquerda para a distância à loja. As demais variáveis têm distribuição aproximadamente simétrica.
- Como esperado, o número de clientes que visitam a loja aumenta à medida que aumenta o tamanho da área, representada pelo número de domicílios.
- Pelos diagramas de dispersão nota-se tendências lineares mais acentuadas entre o número de clientes e as distâncias ao concorrente mais próximo (crescente) e à loja (decrecente). Ligeiro aumento do número de clientes com a renda média.

- 1 Clientes da Loja
- 2 Ajuste Modelo de Poisson**
- 3 Resultados Modelo Ajustado
- 4 Ajuste Modelo de Poisson Parcial Linear
- 5 Resultados Modelo Ajustado
- 6 Referências

Descrição

Seja y_i o número de clientes da i -ésima área que visitaram a loja num determinado período. Vamos supor inicialmente o seguinte modelo:

- $Y_i | \mathbf{x}_i \stackrel{\text{ind}}{\sim} P(\mu_i)$,
- $\log(\mu_i) = \alpha + \beta_1 \times \text{domic}_i + \beta_2 \times \text{renda}_i + \beta_3 \times \text{idade}_i + \beta_4 \times \text{dist1}_i + \beta_5 \times \text{dist2}_i$,

em que $\text{domic} = \text{domic}/1000$ e $\text{renda} = \text{renda}/1000$, para $i = 1, \dots, 110$.

- 1 Clientes da Loja
- 2 Ajuste Modelo de Poisson
- 3 Resultados Modelo Ajustado**
- 4 Ajuste Modelo de Poisson Parcial Linear
- 5 Resultados Modelo Ajustado
- 6 Referências

Descrição

Efeito	Estimativa	E.Padrão	Valor-z	Valor-P
Constante	2,942	0,207	14,20	0,000
Domicílio	0,606	0,142	4,26	0,000
Renda	-0,012	0,002	-5,55	0,000
Idade	-0,004	0,002	-2,09	0,037
Dist1	0,168	0,026	6,53	0,000
Dist2	-0,129	0,016	-7,95	0,000
AIC	571			

Desvio do modelo $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 115$ (para 104 graus de liberdade) com **valor-P=0,217**. Modelo bem ajustado. Desde que $D(\mathbf{y}; \bar{\mathbf{y}}) = 422,22$ tem-se que $R^2 = 1 - \frac{114,99}{422,22} = 0,728$, confirmando excelente ajuste.

Interpretações

Podemos concluir que o número esperado de clientes que frequentam a loja:

- cresce com o aumento do número de domicílios na área,
- cresce com o aumento da distância ao concorrente mais próximo,
- diminui com o aumento da renda média dos domicílios,
- diminui com o aumento da idade média dos domicílios,
- diminui com o aumento da distância da área à loja.

Isso sugere que deve ser uma **loja de conveniência**.

Aumento da Renda Média

Podemos notar pelas estimativas que se aumentarmos, por exemplo, em 1 mil USD a renda média dos domicílios de uma determinada área esperamos aumento relativo no número de clientes que irão à loja de

$$\exp(-0,012) = 0,988.$$

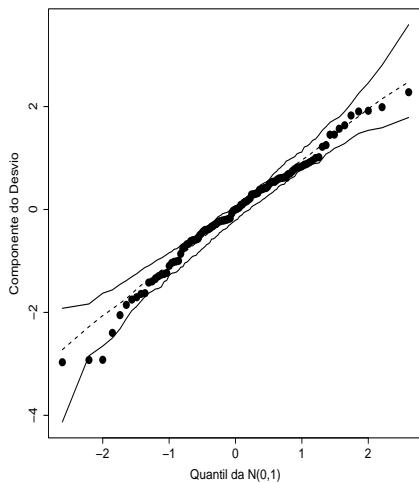
Ou seja, **decréscimo de 1,2%**. Estimativa Intervalar: **[0,8%;1,6%]**.

Aumento da Distância ao Concorrente

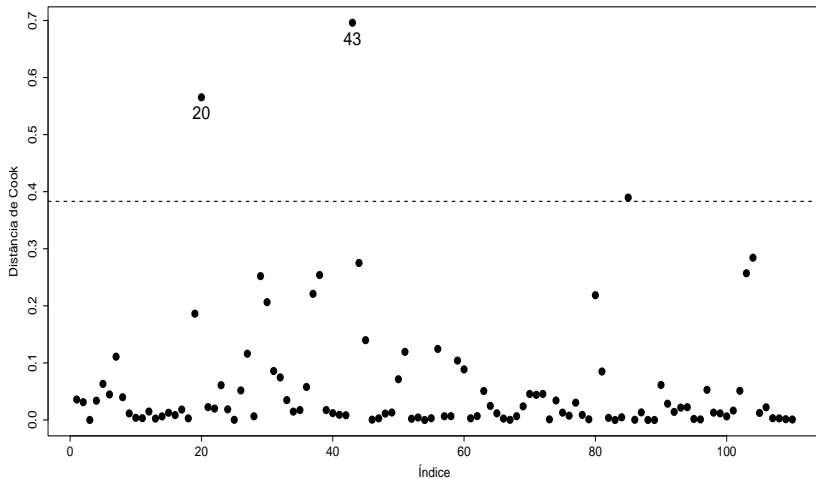
Por outro lado, se a distância ao concorrente mais próximo aumentar em uma milha, esperamos aumento relativo no número de clientes que irão à loja de

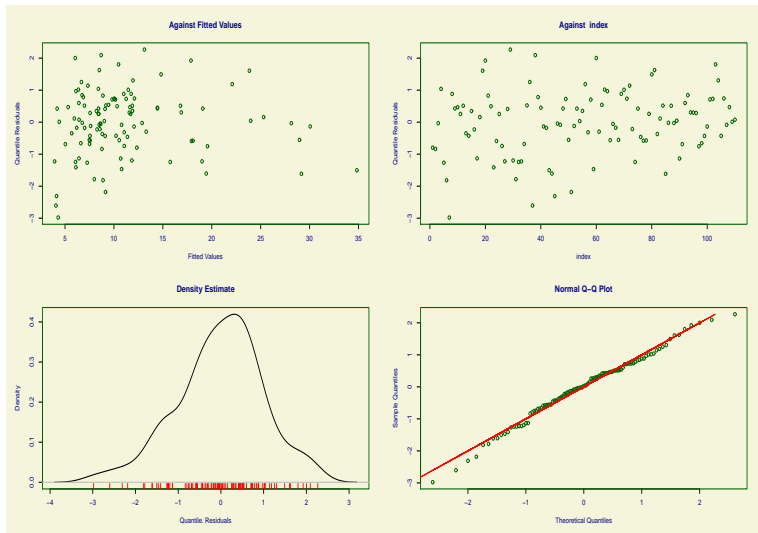
$$\exp(0,168) = 1,183.$$

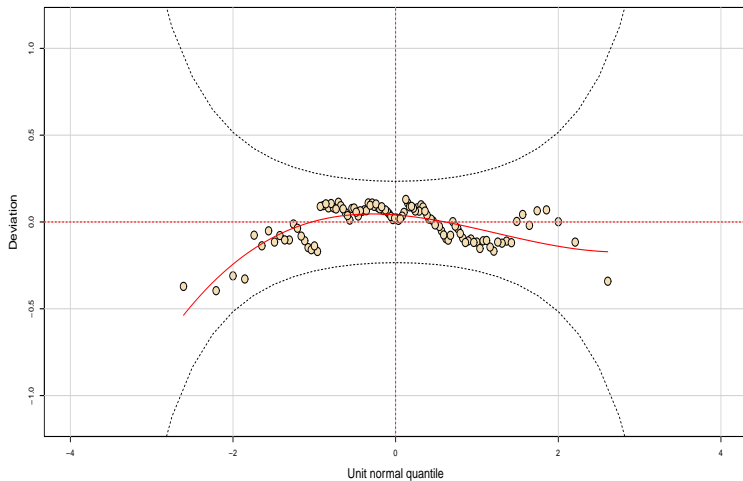
Ou seja, **aumento de 18,3%**. Estimativa Intervalar: **[12,4%;24,5%]**.

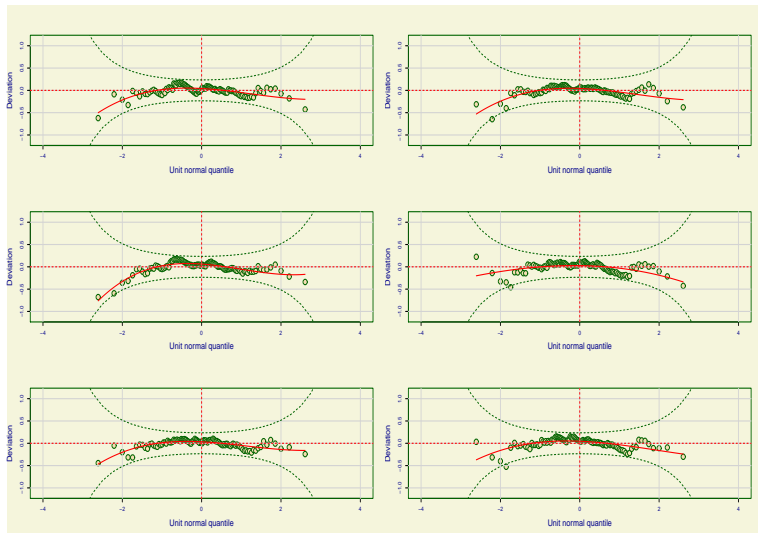


Distância de Cook Modelo Final









Identificação das Observações

- Dentre as observações destacadas pela distância de Cook, apenas as áreas #20 e #43 apresentam algumas variações desproporcionais nas estimativas dos parâmetros, porém sem mudança inferencial.
- A área #20 tem 26/1000 clientes por domicílio (acima da média amostral de 17/1000) porém os domicílios têm idade média (55 anos) e renda média anual (89,9 USD mil) muito acima das respectivas médias amostrais, de 27,43 anos e 48,8 USD mil e está próxima da loja.
- Já a área #43 tem 35/1000 clientes por domicílio o dobro da média amostral, renda média anual de 69,2 USD mil acima da média amostral, idade média de 9 anos bem abaixo da média amostral e está próxima da loja.

Conclusões Parciais

- As observações destacadas como influentes não mudam a inferência quando excluídas individualmente.
- Não há indícios de afastamentos importantes da suposição de distribuição de Poisson para o número de clientes que frequentam a loja.
- Espera-se mais clientes de uma determinado área na loja à medida que o concorrente mais próxima ficar mais distante.
- Por outro lado, espera-se menos clientes na loja de uma determinada área, à medida que a idade dos domicílios aumentar, a renda média domiciliar aumentar ou a distância à loja aumentar.
- A variável explicativa número de domicílios poderia ser controlada de forma não paramétrica através de uma função $f(\text{domic})$.

- 1 Clientes da Loja
- 2 Ajuste Modelo de Poisson
- 3 Resultados Modelo Ajustado
- 4 Ajuste Modelo de Poisson Parcial Linear**
- 5 Resultados Modelo Ajustado
- 6 Referências

Descrição

Seja y_i o número de clientes da i -ésima área que visitaram a loja num determinado período. Vamos supor inicialmente o seguinte modelo:

- $Y_i | \mathbf{x}_i \stackrel{\text{ind}}{\sim} P(\mu_i)$,
- $\log(\mu_i) = \alpha + f(\text{domic}_i) + \beta_2 \times \text{renda}_i + \beta_3 \times \text{idade}_i + \beta_4 \times \text{dist1}_i + \beta_5 \times \text{dist2}_i$,

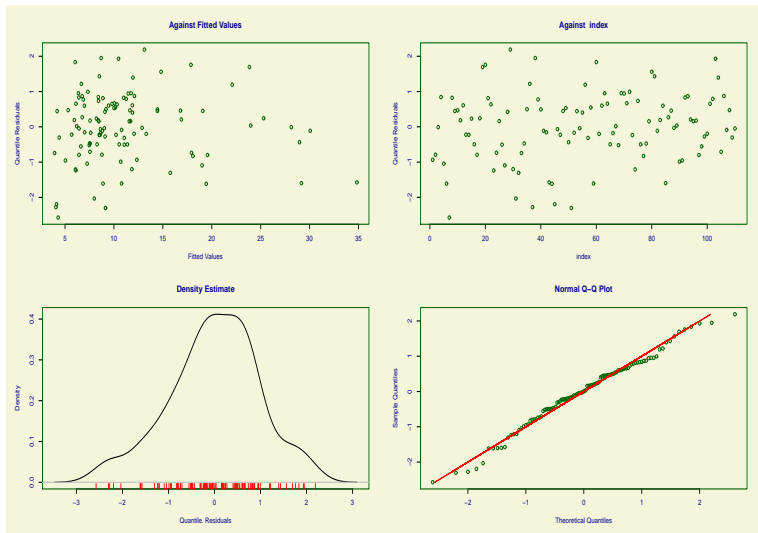
em que em que $\text{domic} = \text{domic}/1000$ e $\text{renda} = \text{renda}/1000$, $f(\cdot)$ é uma função não paramétrica, $i = 1, \dots, 110$.

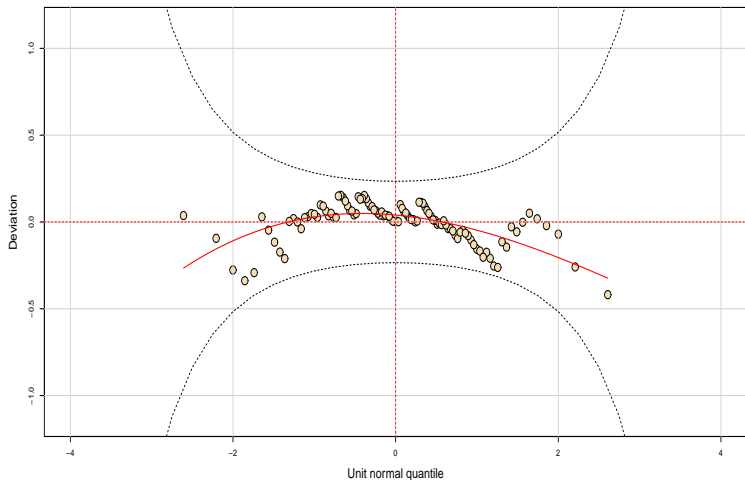
- 1 Clientes da Loja
- 2 Ajuste Modelo de Poisson
- 3 Resultados Modelo Ajustado
- 4 Ajuste Modelo de Poisson Parcial Linear
- 5 Resultados Modelo Ajustado**
- 6 Referências

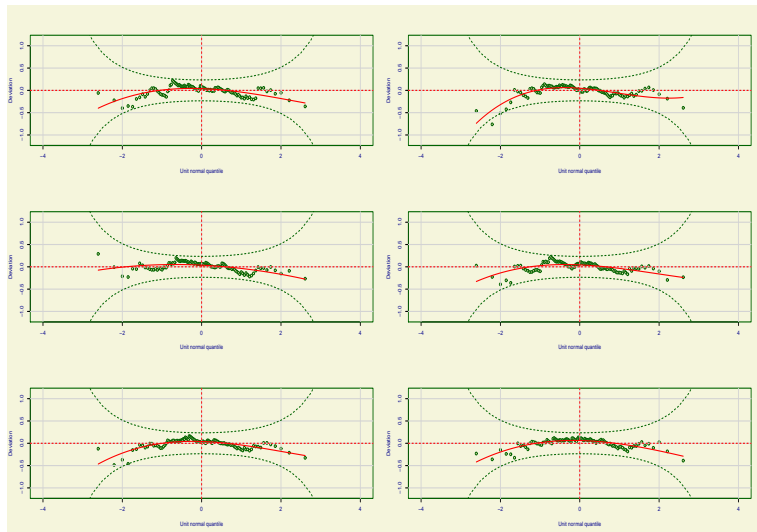
Descrição

Efeito	Estimativa	E.Padrão	Valor-z	Valor-P
Constante	2,943	0,206	14,29	0,000
Renda	-0,012	0,002	-5,49	0,000
Idade	-0,004	0,002	-2,09	0,039
Dist1	0,168	0,026	6,53	0,000
Dist2	-0,129	0,016	-7,95	0,000
pb(Domicílio)	$0,606 \times 10^{-3}$	$0,141 \times 10^{-3}$	5,40	0,000
df ^a	6			
AIC	571			

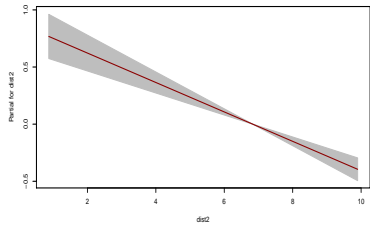
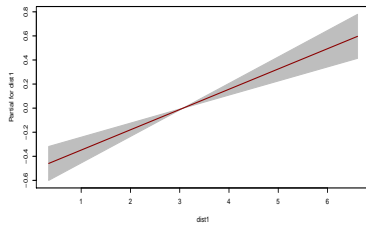
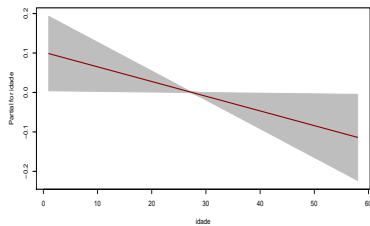
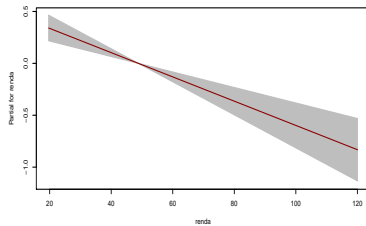
^adf: graus de liberdade efetivos

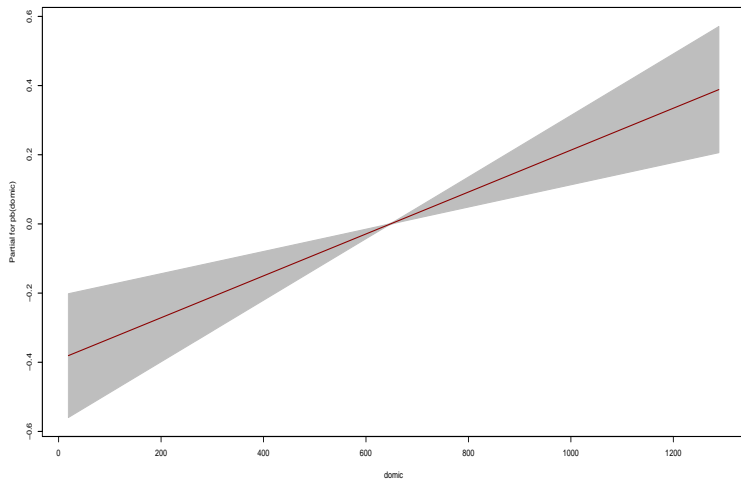






Term Plot Gamlss





Conclusões

- Os ajustes com modelos de Poisson paramétrico e semiparamétrico ficaram praticamente com os mesmos valores para as estimativas dos parâmetros.
- A modelagem paramétrica é mais vantajosa uma vez que permite estimar de forma simples a relação funcional entre o número médio de clientes e o número de domicílios em cada área.

- 1 Clientes da Loja
- 2 Ajuste Modelo de Poisson
- 3 Resultados Modelo Ajustado
- 4 Ajuste Modelo de Poisson Parcial Linear
- 5 Resultados Modelo Ajustado
- 6 Referências**

Referências

- Neter, J.; Kutner, M. H.; Nachtsheim, C. J. e Wasserman, W.(1996). *Applied Linear Regression Models*, 3rd Edition. Irwin, Illinois.
- Stasinopoulos, M. D., Rigby, R. A., Gillian, Z. A., Voudouris, V. e de Bastiani, F. (2017). *Flexible Regression and Smoothing Using GAMLSS in R*. Chapman and Hall/CRC.