

# Regressão Quantílica:

Definição, estimação, aplicações e extensões

André F. B. Menezes | Outubro, 2020

# Organização

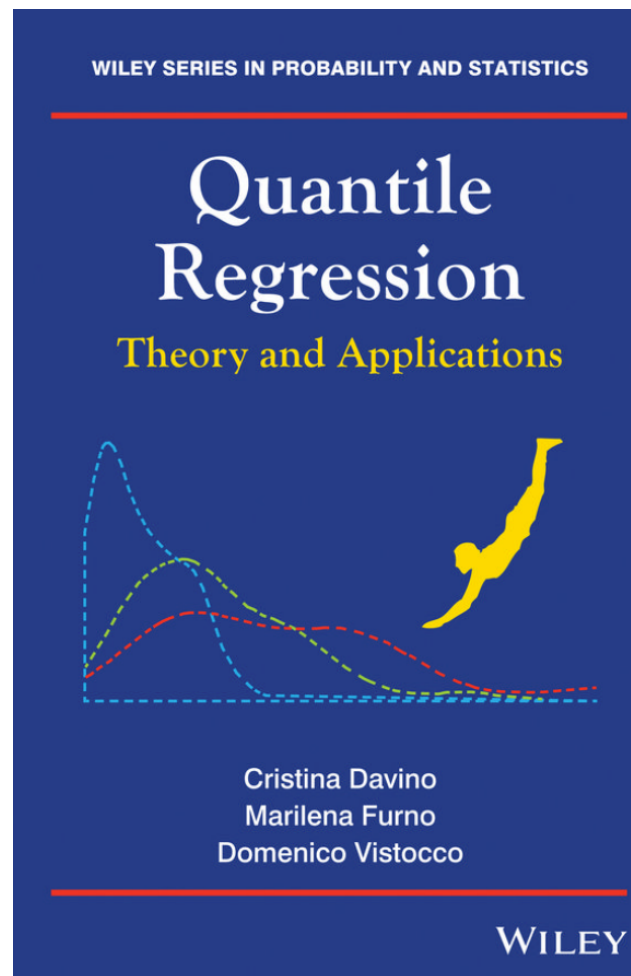
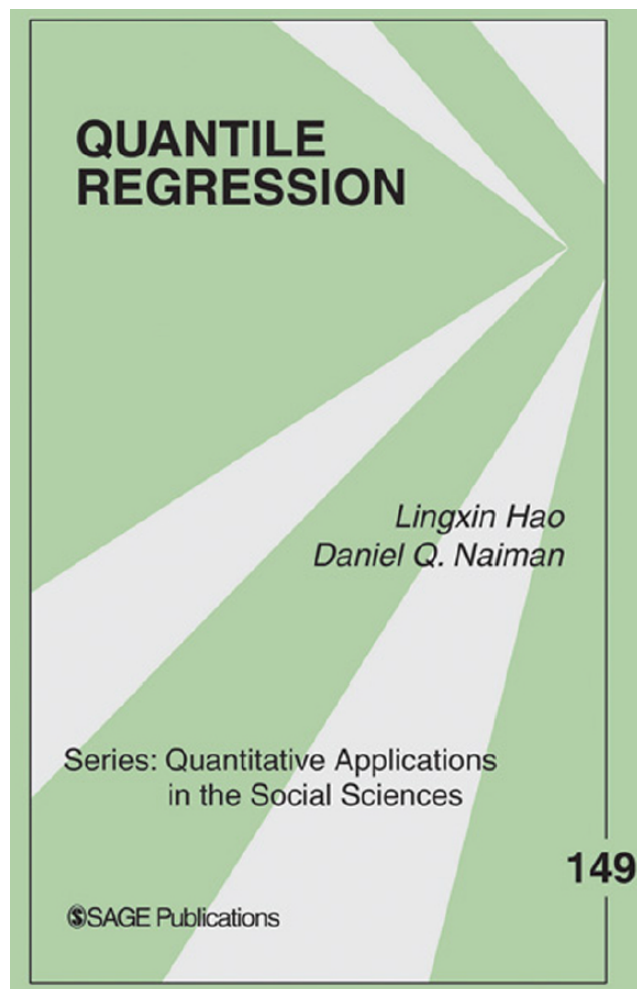
- Referências, motivações e definições
- Estimação e inferência
- Recursos computacionais
- Exemplos empíricos
- Extensões

# Referências, motivações e definições

# Livros de referência



# Livros de referência



# Outras referências

- Manual do SAS: **QUANTREG Procedure**.
- Dissertação e Tese de Bruno Ramos dos Santos:
  - **Modelos de regressão quantílica**
  - **Extensões dos modelos de regressão quantílica**

# Motivação inicial

- Muitas aplicações reais sob o ponto de vista da estatística são formulações mais elaboradas do clássico modelo de regressão linear.

What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of x's. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions.

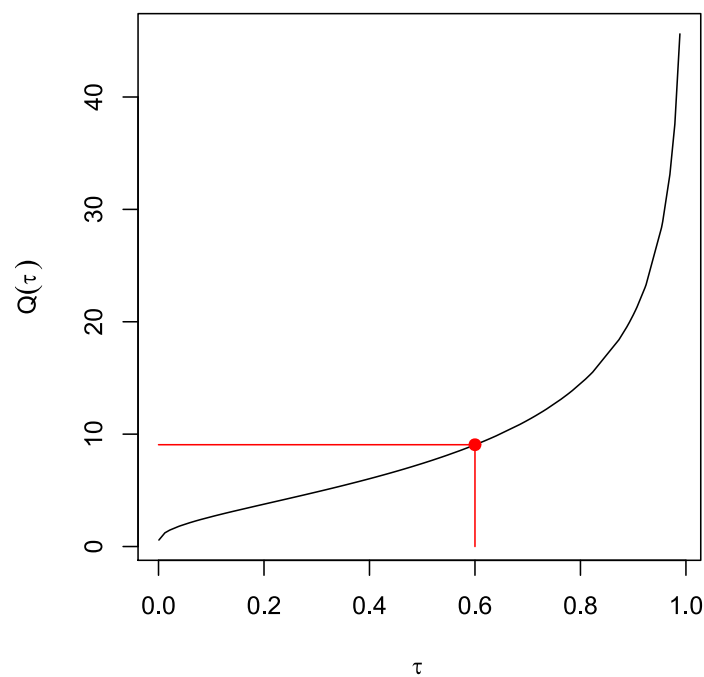
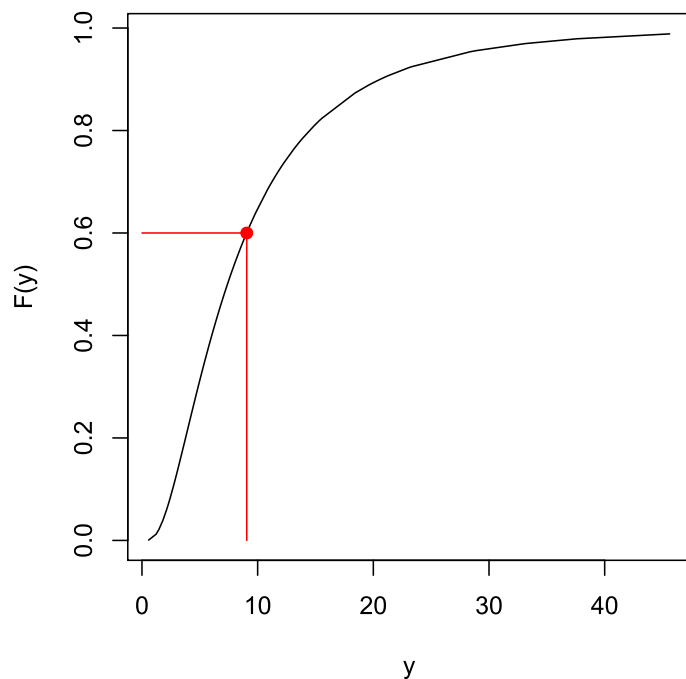
-- *Mosteller and Tukey (1977)*

- Como ir além ("*go further*") do clássico modelo de regressão linear?  
**Regressão Quantílica!**

# Quantil: Definição probabilística

Seja  $Y$  uma variável aleatória com função de distribuição  $F_Y(y) = P(Y \leq y)$  o quantil de ordem  $\tau \in (0, 1)$  é definido por:

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$$





# Quantil: Definição alternativa

Um quantil também pode ser considerado uma solução para um problema de minimização.

- Seja  $\mu = EY$ , sabemos que

$$\mu = \arg \min_c E(Y - c)^2.$$

- Seja  $m$  uma mediana de  $Y$ , sabemos que

$$m = \arg \min_c E|Y - c|.$$

- A prova dessas proposições podem ser vistas, por exemplo, em [James \(2008\)](#).
- Este resultado pode ser generalizado para qualquer quantil  $\tau \in (0, 1)$ .

# Quantil: Definição alternativa

Considere o problema de **teoria da decisão** de prever o valor da variável aleatória  $Y$  adotando a seguinte **função perda**:

$$\rho_\tau(y) = [\tau - I(y < 0)] y, \quad \tau \in (0, 1)$$

sendo  $I(\cdot)$  a função indicadora.

- Em particular, um quantil  $q_\tau$  minimiza a perda esperada, isto é,

$$q_\tau = \arg \min_c E[\rho_\tau(Y - c)].$$

- **Prova:**

- Observe que

$$E[\rho_\tau(Y - c)] = (\tau - 1) \int_{-\infty}^c (y - c) dF(y) - \tau \int_c^{\infty} (y - c) dF(y).$$

# Quantil: Definição alternativa

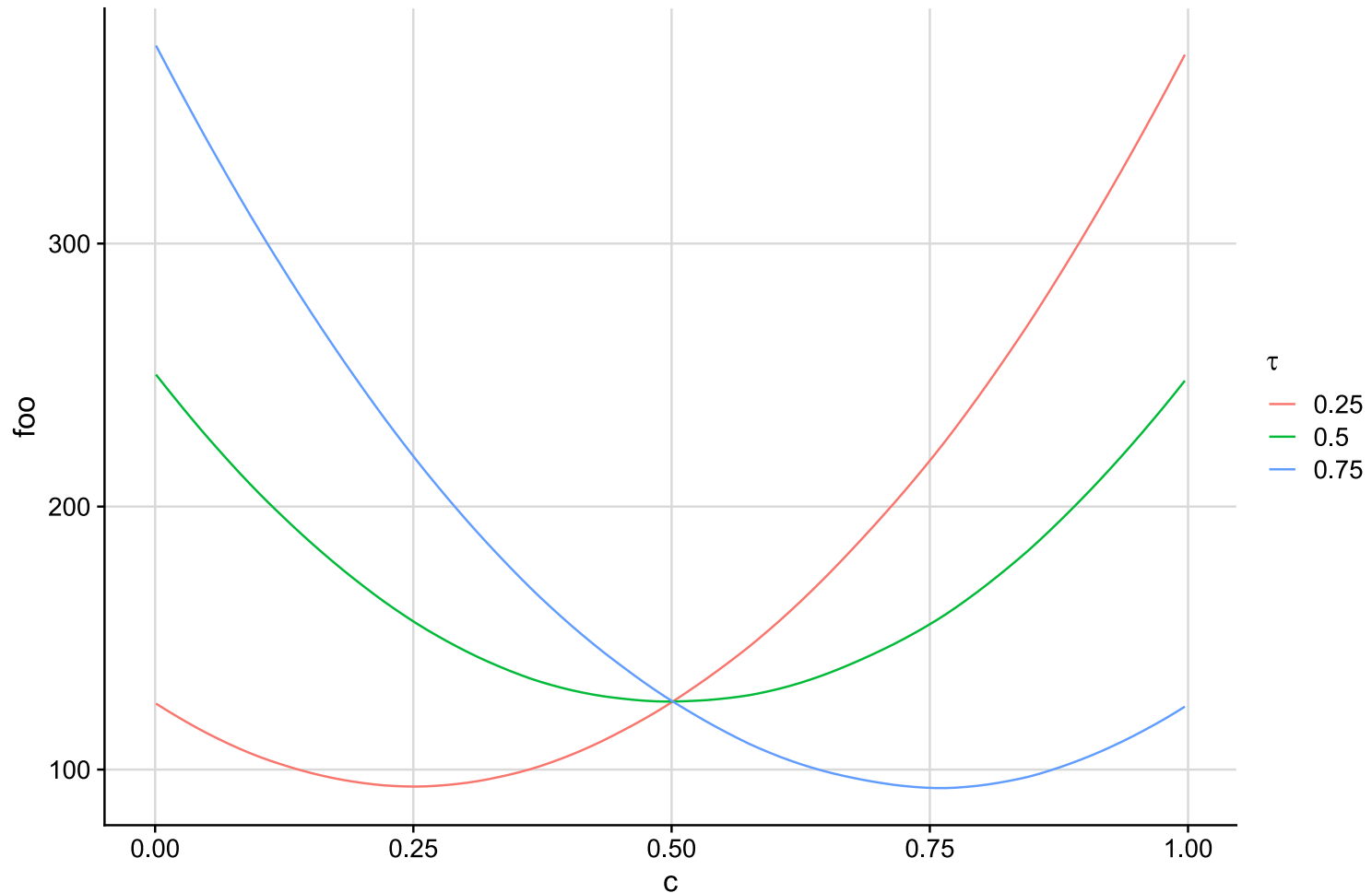
- **Prova:**

- Diferenciando em relação a  $c$  temos

$$0 = (1 - \tau) \int_{-\infty}^c dF(y) - \tau \int_c^{\infty} dF(y) = F(c) - \tau.$$

- Como  $F$  é uma função monôtona, qualquer  $q_\tau$  que satisfaça  $q_\tau : F(q_\tau) = \tau$  minimiza a perda esperada.
  - Quando a solução é única temos que  $q_\tau = F^{-1}(\tau)$ , caso contrário teremos um intervalo de quantis.
- Essa definição é fundamental para especificação do modelo de regressão quantílica.

# Comportamento da função perda



# Esperança e Quantil

Seja  $h$  uma função monótona sabemos que

- $E[h(Y)] \neq h(EY)$ .
- $Q_\tau[h(Y)] = h[Q_\tau(Y)]$ .

A esperança de  $Y$  pode ser calculada da seguinte forma:

$$\mu = E(Y) = \int_0^1 Q(y) dy.$$

- Detalhes podem ser encontrados em [Gilchrist \(2000\)](#).

# Esperança condicional

- No contexto de regressão estamos interessados em estimar uma função média condicional aos preditores  $\mathbf{x}_i$ .
- Assim, uma extensão do resultado apresentado pode ser enunciado da seguinte forma:

$$\hat{\mu}(\mathbf{x}_i, \boldsymbol{\beta}) = \arg \min_{\mu} E[Y - \mu(\mathbf{x}_i, \boldsymbol{\beta})]^2.$$

em que  $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = E(Y | X = \mathbf{x}_i)$ .

- Naturalmente, no caso do modelo linear temos que  $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i^\top \boldsymbol{\beta}$ , resultando nos estimadores de mínimos quadrados:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} E[Y - \mathbf{x}_i^\top \boldsymbol{\beta}]^2.$$

# Quantil condicional

- Analogamente, podemos estar interessado em estimar o quantil condicional de  $Y$  dado  $\mathbf{x}$  como uma função linear nos parâmetros, isto é,

$$Q_\tau(Y | \mathbf{x}) = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau)$$

- Por que estimar quantis condicionais?
  - Desigualdade econômica interesse intrínseco nos pobres e/ou ricos (cauda inferior e/ou superior);
  - Estudos educacionais buscam compreender grupos em níveis de desempenho pré-estabelecidos.

# Regressão quantílica

- O modelo de regressão quantílica foi introduzido por **Koenker and Bassett (1978)** e é definido por:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau) + u_i$$

em que  $u_i$  são variáveis aleatórias i.i.d. com  $Q_\tau(u_i | \mathbf{x}_i) = 0$ .



# Comparação

- Preserva transformação da variável resposta:
  - $E(Y | X) \neq h^{-1}(E(h(Y) | X))$ .
  - $Q_\tau(Y | X) = h^{-1}(Q_\tau(h(Y) | X))$ .
- Uma aproximação para os coeficiente de regressão na média:

$$\beta_\mu = \frac{1}{M} \sum_{\tau} \beta(\tau)$$

sendo  $M$  total de quantis utilizados.

Linear Regression	Quantile Regression
Predicts the conditional mean $E(Y X)$	Predicts conditional quantiles $Q_\tau(Y X)$
Applies when $n$ is small	Needs sufficient data
Often assumes normality	Is distribution agnostic
Does not preserve $E(Y X)$ under transformation	Preserves $Q_\tau(Y X)$ under transformation
Is sensitive to outliers	Is robust to response outliers
Is computationally inexpensive	Is computationally intensive

# Estimação e inferência

# Estimação frequentista

- Sob o ponto de vista da estatística frequentista a estimação dos parâmetros da regressão quantílica é realizada utilizando técnicas de programação linear para minimização da seguinte função:

$$\sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i^{\top} \boldsymbol{\beta}).$$

- Para exemplificação considere regressão L1 (mediana), ou seja,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n |y_i - \mathbf{x}_i^{\top} \boldsymbol{\beta}|$$

- Considere a seguinte notação:

1.  $\boldsymbol{\mu} = [\mathbf{y} - A^{\top} \boldsymbol{\beta}]_+$  e  $\boldsymbol{\nu} = [A^{\top} \boldsymbol{\beta} - \mathbf{y}]_+$

2.  $\boldsymbol{\phi} = [\boldsymbol{\beta}]_+$  e  $\boldsymbol{\varphi} = [-\boldsymbol{\beta}]_+$

em que  $\mathbf{y} = (y_1, \dots, y_n)^{\top}$ ,  $A^{\top} = (\mathbf{x}_1, \dots, \mathbf{x}_p)^{\top}$  e  $[z]_+$  é a parte não negativa de  $z$ .

# Estimação frequentista

- Reformulando o problema de minimização temos que

$$\min_{\beta} \{ \mathbf{1}_n^\top \boldsymbol{\mu} + \mathbf{1}_n^\top \boldsymbol{\nu} \mid \mathbf{y} = \mathbf{A}^\top \boldsymbol{\beta} + \boldsymbol{\mu} - \boldsymbol{\nu} \}$$

- Seja  $\mathbf{B} = [\mathbf{A}^\top - \mathbf{A}^\top \mathbf{I} - \mathbf{I}]$ ,  $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\nu})^\top$  e  $\mathbf{d} = (\mathbf{0}, \mathbf{0}, \mathbf{1}_n, \mathbf{1}_n)^\top$ .
- Transformando a um problema de programação linear (LP) padrão:

$$(P) \quad \min_{\boldsymbol{\theta}} \mathbf{d}^\top \boldsymbol{\theta}; \text{ sujeito à } \mathbf{B}\boldsymbol{\theta} = \mathbf{y}, \boldsymbol{\theta} \geq 0.$$

- Com a seguinte formulação *dual*:

$$(D) \quad \max_{\mathbf{z}} \mathbf{y}^\top \mathbf{z}; \text{ sujeito à } \mathbf{B}^\top \mathbf{z} \leq \mathbf{d}.$$

- Que pode ser simplificada e generalizada para qualquer quantil:

$$\max_{\mathbf{z}} \mathbf{y}^\top \mathbf{z}; \text{ sujeito à } \mathbf{A}\mathbf{z} = (1 - \tau)\mathbf{A}\mathbf{1}_n, \mathbf{z} \in [0, 1]^n.$$

# Algoritmos

- Simplex proposto por [Barrodale and Roberts \(1973\)](#) e [Koenker and d'Orey \(1994\)](#):
  - Estável, pois encontra uma única solução.
  - Computacionalmente intensivo para grandes amostras ( $n > 5000$ ).
- Ponto interior proposto por [Portnoy and Koenker \(1997\)](#):
  - Computacionalmente mais eficiente.
  - Apresenta instabilidade dependendo do comportamento das preditoras.
- *Smoothing* proposto por [Chen \(2007\)](#):
  - Aproxima a função objetivo por uma função de suavização (função Huber).

# Inferência

- Existem três abordagens para inferência frequentista em regressão quantílica:
  1. Teoria assintótica (esparsidade).
  2. Bootstrap.
  3. Inversão de testes de escores ordinais.
- Em geral, resultados assintóticos são práticos e menos custoso computacionalmente, porém são sensíveis a suposição i.i.d.
- Veremos um pouco sobre 1. e 2.
- Detalhes sobre o método 3. podem ser encontrados em [Koenker \(2005\)](#).

# Inferência: Teoria Assintótica

- Seja  $F$  a função de distribuição de  $u_i$  e  $f = F'$ .
- Sob certas condições de regularidades **Koenker and Bassett (1982a)** enunciaram e provaram um teorema com o seguinte resultado:

$$\sqrt{n} \left[ \hat{\beta}(\tau) - \beta(\tau) \right] \rightarrow N(0, \omega^2(\tau, F) \Omega^{-1})$$

em que  $\omega^2(\tau, F) = \tau(1 - \tau)/f^2(F^{-1}(\tau))$  e  $\Omega = \lim_{n \rightarrow \infty} n^{-1} \sum \mathbf{x}_i \mathbf{x}_i^\top$ .

- Naturalmente, a partir da distribuição assintótica obtemos intervalos de confiança e testes de hipóteses.
- Como obter uma estimativa para o termo  $s(\tau) = [f(F^{-1}(\tau))]^{-1}$ , denominado função de esparsidade?

# Inferência: Teoria Assintótica

- Note que

$$s(t) = \frac{d}{dt} F^{-1}(t).$$

- Assim  $s(t)$  pode ser estimado pelo quociente de diferença da função de quantil empírica, isto é,

$$\hat{s}_n(t) = \frac{1}{2h_n} \left[ \hat{F}_n^{-1}(t + h_n) - \hat{F}_n^{-1}(t - h_n) \right]$$

em que

- $\hat{F}_n^{-1}$  é uma estimativa empírica de  $F^{-1}$ .
- $h_n$  é uma largura (*bandwith*) que tende a 0 quando  $n \rightarrow \infty$ .



# Inferência: Teoria Assintótica

- Como estimar  $F^{-1}$ ?
  - Quantil empírico dos resíduos:

$$\hat{F}^{-1}(t) = r_{(i)}, \quad t \in [(i-1)/n, i/n),$$

- Quantil empírico da regressão:

$$\hat{F}^{-1}(t) = \bar{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}(t).$$

# Inferência: Teoria Assintótica

- Qual *bandwidth* utilizar?
  - Bofinger:

$$h_n = n^{-1/5} \left[ \frac{4.5 s^2(t)}{(s^{(2)}(t))^2} \right]^{1/5} .$$

- Hall-Sheather:

$$h_n = n^{-1/3} z_\alpha^{2/3} \left[ \frac{1.5 s(t)}{s^{(2)}(t)} \right]^{1/3} .$$

em que  $s^{(2)}(t)$  é a segunda derivada de  $s(t)$  e  $z_\alpha$  é o quantil de ordem  $\alpha$  da  $N(0, 1)$ .

# Inferência: Bootstrap

- Obter a distribuição do estimador  $\hat{\beta}(\tau)$ :
  1. Gera uma amostra Bootstrap.
  2. Estima-se  $\beta(\tau)$ .
  3. Repete-se várias vezes 1. e 2. e obtêm-se uma distribuição para  $\hat{\beta}(\tau)$ .
- Como gerar uma amostra Bootstrap?
  - Reamostra aleatoriamente o par  $(y_i, \mathbf{x}_i)$ .
  - Reamostra aleatoriamente os resíduos e obtém  $\hat{u}_i^*$ , utiliza como resposta  $y_i^* = \hat{y}_i + \hat{u}_i^*$ .
  - Método proposto por **He and Hu (2002)** *Markov Chain Marginal Bootstrap* (MCMB).

# Inferência: Testando efeito das covariáveis

Queremos testar  $H_0 : \beta_j(\tau) = 0$ , para um dado  $\tau$ .

- Estatística de Wald:

$$S_W(\tau) = \beta_j^\top(\tau) \hat{\Sigma}(\tau)^{-1} \beta_j(\tau)$$

em que  $\hat{\Sigma}(\tau)^{-1}$  pode ser estimado via teoria assintótica ou Bootstrap.

- Estatística da razão de verossimilhanças:

$$S_{LR}(\tau) = 2 [\tau(1 - \tau)\hat{s}(\tau)]^{-1} [D_1(\tau) - D_0(\tau)]$$

em que  $D_0(\tau)$  e  $D_1(\tau)$  são as funções objetivos do modelo irrestrito e restrito, respectivamente e  $\hat{s}(\tau)$  é a estimativa da função de esparsidade.

- Estatística baseada em escores ( $S_R$ ): ver [Gutenbrunner, Jureckova, Koenker, and Portnoy \(1993\)](#).
- [Koenker and Machado \(1999\)](#) mostram que sob  $\mathcal{H}_0$  as três estatística são assintoticamente  $\chi_q$ , sendo  $q$  a dimensão de  $\beta_j(\tau)$ .

# Inferência: Teste para Heteroscedasticidade

- Existe diferença estatística entre as estimativas  $\beta_i(\tau)$  de para vários valores de  $\tau$ ?

$$\mathcal{H}_0 : \beta_i(\tau_1) = \dots = \beta_i(\tau_k)$$

- **Koenker and Bassett (1982b)** propuseram uma estatística cuja a distribuição assintótica é qui-quadrado.

# Paradigma Bayesiano: Modelo

- **Yu and Moyeed (2001)** utilizaram a distribuição Laplace assimétrica para introduzir o paradigma bayesiano na regressão quantílica.
- Uma variável aleatória  $Y$  segue distribuição Laplace assimétrica ( $LA$ ) se sua densidade é dada por

$$f(y \mid \mu, \sigma, \tau) = \frac{\tau(1 - \tau)}{\sigma} \exp \left\{ -\rho_{\tau} \left( \frac{y - \mu}{\sigma} \right) \right\}$$

em que  $\mu \in \mathbb{R}$  é o quantil de ordem  $\tau \in (0, 1)$  e  $\sigma > 0$  é um parâmetro de escala.

- No contexto de regressão temos que  $Y_i \mid \mathbf{x}_i \sim LA(\mu_i, \sigma; \tau)$ .
- Supondo que  $\mu_i = \mathbf{x}_i^{\top} \boldsymbol{\beta}(\tau)$  para um dado  $\tau \in (0, 1)$ .
- Maximizar a verossimilhança do modelo é equivalente a minimização no contexto frequentista.

# Paradigma Bayesiano: Estimação e Inferência

- Para estimação e inferência **Yu and Moyeed (2001)** fixaram  $\sigma = 1$
- Assumiram uma distribuição Uniforme imprópria para cada componente  $\beta$  e métodos MCMC foram utilizados para obter a distribuição a posteriori.
- **Kozumi and Kobayashi (2011)** propuseram uma representação por mistura da distribuição Laplace assimétrica da seguinte forma

$$Y \mid \nu \sim N(\mu + \theta \nu, \varphi^2 \sigma \nu)$$
$$\nu \sim \text{Exp}(\sigma)$$

em que  $\theta = \frac{1 - 2\tau}{\tau(1 - \tau)}$  e  $\varphi^2 = \frac{2}{\tau(1 - \tau)}$ .

- A distribuição marginal de  $Y$  é  $LA(\mu, \sigma; \tau)$

# Paradigma Bayesiano: Estimação e Inferência

- Incorporando preditores, isto é,  $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau)$  o modelo de regressão quantílica sob o paradigma Bayesiano introduzido por [Kozumi and Kobayashi \(2011\)](#) é definido por

$$Y_i \mid \nu \sim N(\mu_i + \theta \nu, \varphi^2 \sigma \nu)$$

$$\nu \sim \text{Exp}(\sigma)$$

$$\boldsymbol{\beta}(\tau) \sim N(b_0, B_0)$$

$$\sigma \sim \text{GI}(n_0/2, s_0/2)$$

- Sob esta representação é possível obter expressões mais eficientes para amostrar das distribuições a posteriori.



# Recursos computacionais

# Recursos computacionais

- **R**
  - **quantreg**: regressão quantílica.
  - **quantgen**: regressão quantílica com penalização (implementação mais eficiente).
  - **uwquantreg**: regressão quantílica para variáveis respostas restritas ao intervalo  $(0, 1)$ .
- **python**
  - **statsmodels**: regressão quantílica.
- **SAS**
  - **QUANTREG Procedure**: todo ferramental de regressão quantílica.
  - **QUANTSELECT Procedure**: métodos para seleção de preditoras em regressão quantílica.
  - **QUANTLIFE Procedure**: regressão quantílica para dados com censura.

# Exemplos empíricos

# Exemplo 1

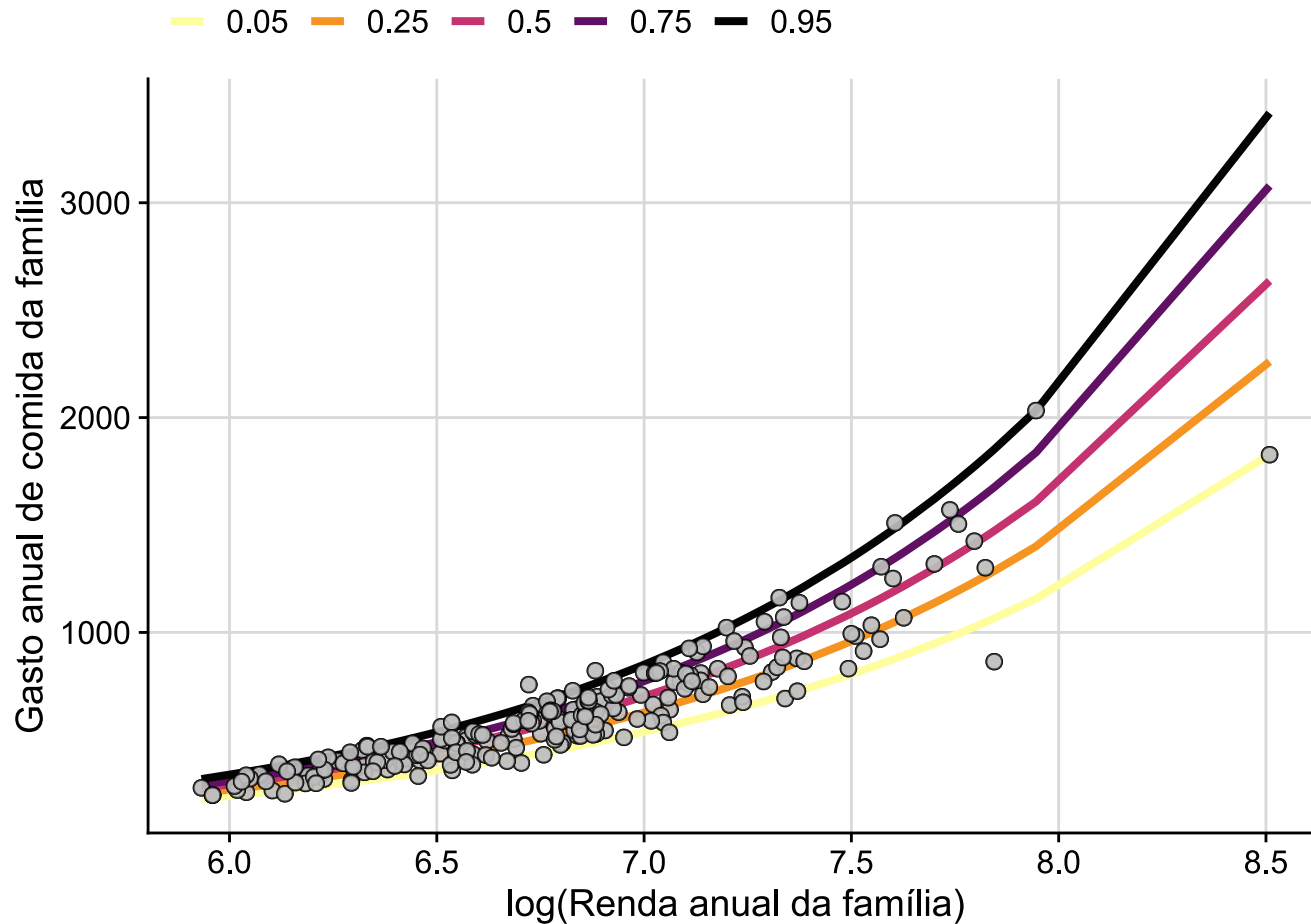
- Amostra de 235 famílias da Bélgica.
- Utilizado por **Koenker and Bassett (1982b)** para estudar:
  - $y_i$ : gasto anual de comida da família.
  - $x_i$ : renda anual da família.
- O modelo é dado por

$$Q_\tau(\log(y_i) \mid x_i) = \beta_0(\tau) + \beta_1(\tau) \log(x_i)$$

- Resultados:

	<b>tau= 0.05</b>	<b>tau= 0.25</b>	<b>tau= 0.50</b>	<b>tau= 0.75</b>	<b>tau= 0.95</b>
(Intercept)	0.607	0.495	0.418	0.241	0.290
log(income)	0.811	0.849	0.877	0.916	0.922

# Exemplo 1



# PNAD Covid19

# Estudo

- **Pesquisa Nacional por Amostra de Domicílios:** Maio/2020 a Ago/2020.
- **Plano amostral:** amostra estratificada abrangendo todas as UF.
- Total de 180.584 domicílios, totalizando 386.520 entrevistados.
- 145 variáveis observadas:
  - Identificação e controle: UF, localização do domicílio, etc.
  - Características gerais dos moradores: idade, sexo, raça, escolaridade, **salário**, etc.
  - Covid19: sintomas, internação, etc.
  - Características de trabalho das pessoas com mais de 14 anos: tipo de trabalho, recebeu auxílio emergencial, etc.

# Objetivo

- É senso comum que os rendimentos dos brasileiros são baixos e insuficientes.
- Nexo Jornal: **O seu salário diante da realidade brasileira**
  - Dados da PNAD apontam que um bolsista de mestrado ganha mais do que 67% de todos os brasileiros!
- O que mais impacta o baixo/alto salário dos brasileiros?

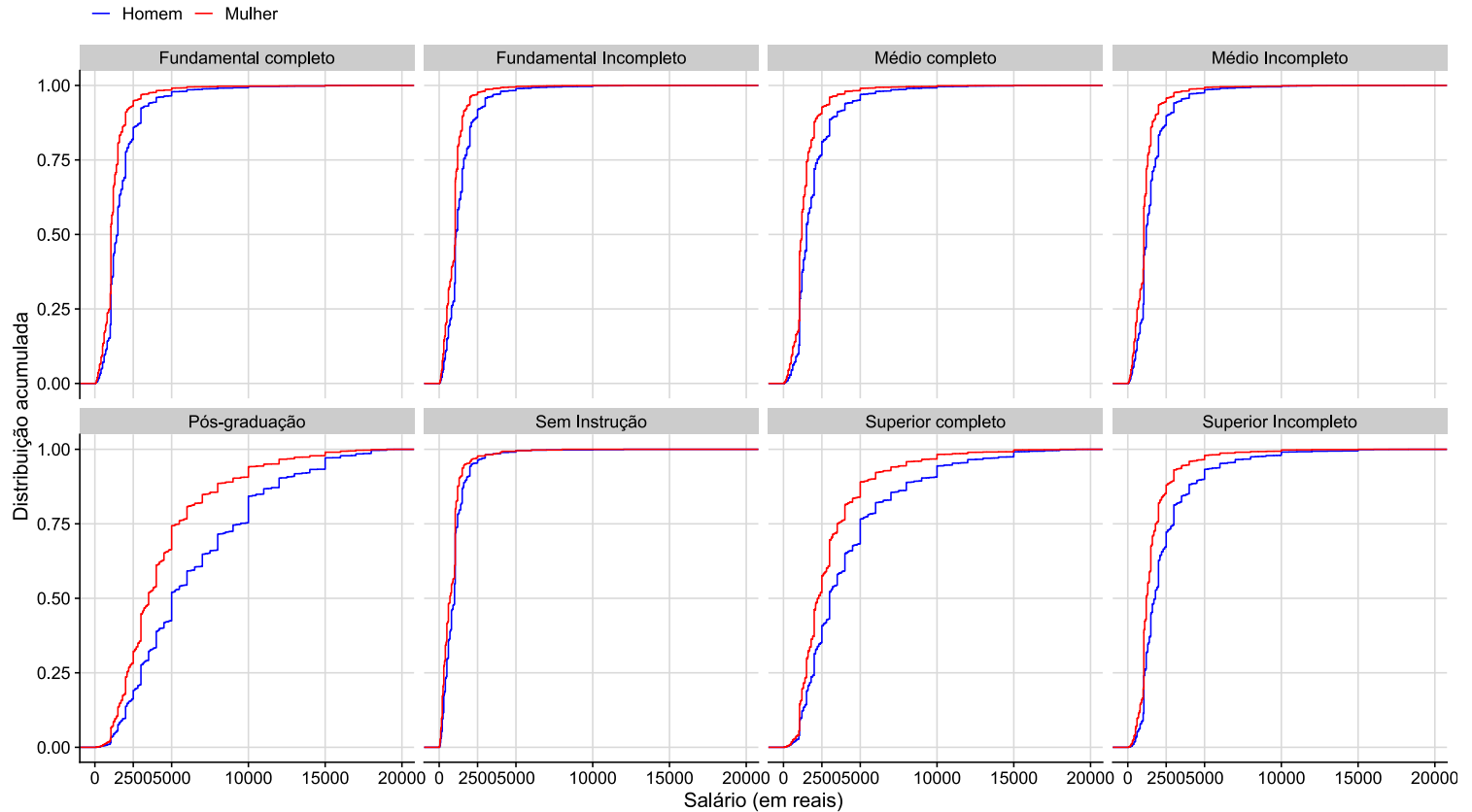


# Análise descritiva

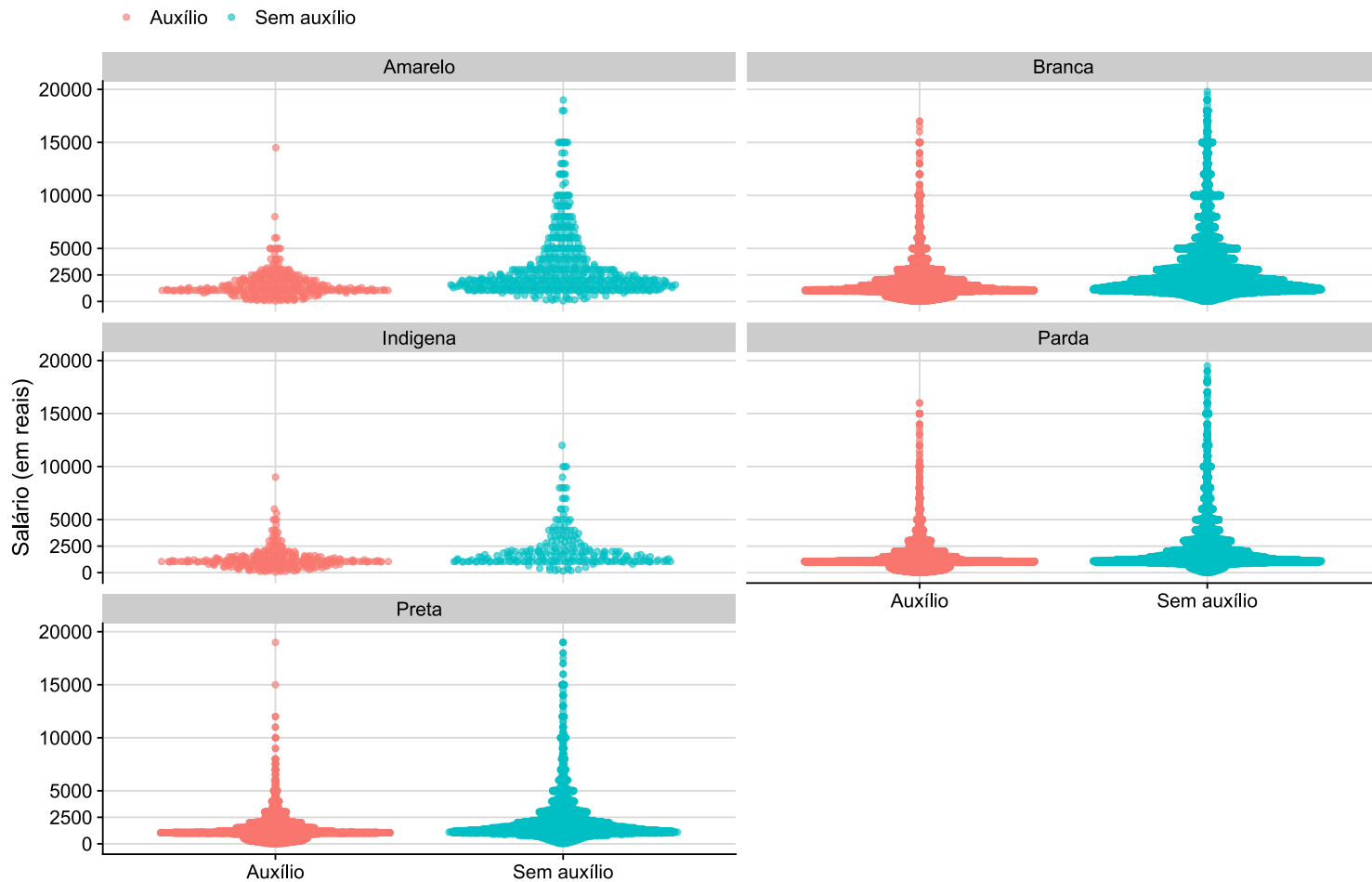
Quantis empíricos conforme sexo e idade.

Categoria	1%	5%	20%	35%	50%	65%	80%	95%	99%
Homem entre [0-15)	50.00	68.0	200	300	500.0	594.8	800	1180.00	2780
Mulher entre [0-15)	42.10	130.5	188	200	300.0	352.5	468	1060.00	1332
Homem entre [15-25)	100.00	300.0	700	1045	1045.0	1200.0	1500	2500.00	4500
Mulher entre [15-25)	100.00	292.8	600	1000	1045.0	1200.0	1500	2300.00	4000
Homem entre [25-34)	200.00	500.0	1045	1200	1500.0	1901.2	2500	5000.00	10000
Mulher entre [25-34)	150.00	400.0	1045	1050	1300.5	1700.0	2500	5000.00	10000
Homem entre [35-50)	200.00	500.0	1045	1300	1600.0	2100.0	3000	7306.75	13000
Mulher entre [35-50)	150.00	400.0	1045	1050	1400.0	2000.0	3000	6000.00	11000
Homem entre [50-65)	160.00	400.0	1045	1200	1600.0	2080.0	3000	8000.00	15000
Mulher entre [50-65)	120.00	300.0	1000	1045	1300.0	2000.0	3000	6155.00	12000
Homem entre [65-112)	100.00	300.0	900	1045	1500.0	2000.0	3700	10000.00	15000
Mulher entre [65-112)	59.88	200.0	600	1045	1200.0	2000.0	3000	8000.00	13060

# Análise descritiva



# Análise descritiva



# Modelagem

- Modelo proposto:

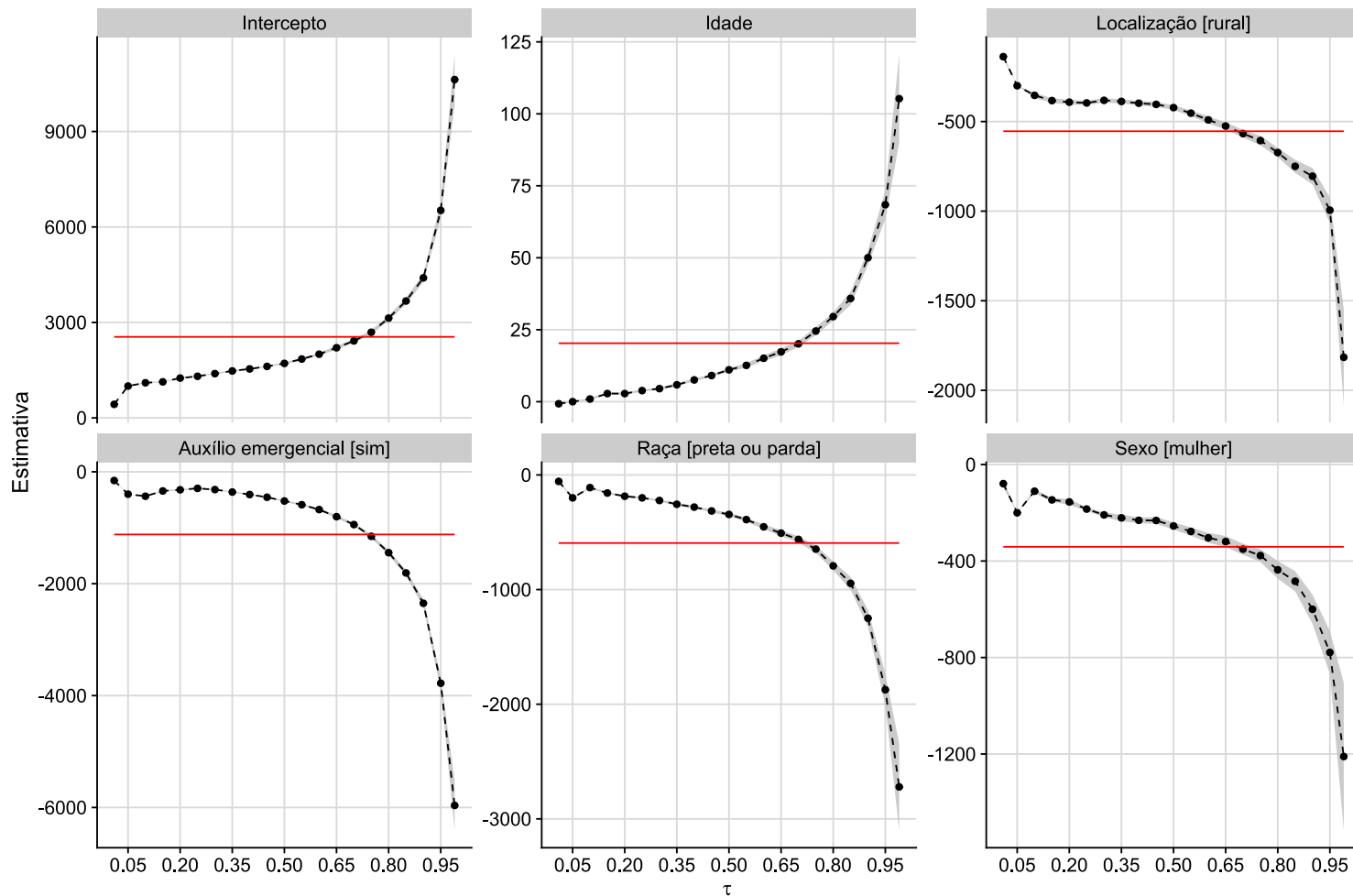
$$Q_\tau(\text{salario}_i | \mathbf{x}_i) = \beta_0(\tau) + \beta_1(\tau)\text{idade}_i + \beta_2(\tau)\text{mulher}_i + \beta_3(\tau)\text{preta\_parda}_i \\ + \beta_4(\tau)\text{auxilio}_i + \beta_5(\tau)\text{rural}_i$$

- **Problema/Limitação:** temos uma amostra estratificada.
  - Cada indivíduo recebeu um peso de estratificação associado a seu domicílio. Como incluir no modelo?
  - Seja  $w_i$  o peso da  $i$  observação, então a estimativa da regressão quantílica foi obtida minimizando a função:

$$\sum_{i=1}^n w_i \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}).$$

- Pacote `quantreg` foi utilizado considerando:
  - `method="pfn"`: Algoritmo para estimação proposto por **Portnoy and Koenker (1997)**.

# Modelagem: Resultados



# Extensões

# Regressão quantílica paramétrica

- Similar aos modelos lineares generalizados:

$$Y_i | \mathbf{x}_i \sim \mathcal{F}(\mu_i, \phi; \tau), \quad \text{com} \quad \mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta}),$$

- $\mu_i$  é o parâmetro de localização referente ao quantil de ordem  $\tau$  e  $\phi$  é um parâmetro de escala/forma.
- $g(\cdot)$  é uma função de ligação que depende do espaço paramétrico de  $\mu_i$ .
- Alguns trabalhos:
  - **Noufaily and Jones (2013)**: regressão quantílica Gama generalizada.
  - **Smithson and Shou (2017)** e **Shou and Smithson (2019)**: regressão quantílica para dados no  $(0, 1)$  considerando um classe de distribuições baseadas na função distribuição acumulada.
  - **Sanchez, Leiva, Galea, and Saulo (2020)**: regressão quantílica Birnbaum-Saunders.
  - **Mazucheli, Menezes, Fernandes, de Oliveira, and Ghitany (2020)**: regressão quantílica unit-Weibull.

# Regressão quantílica com censura

- Seja  $T_i$  a variável resposta, e.g., tempo de sobrevivência.
- Censura: **nem todas as observações  $T_i$  são completamente observadas.**
- Seja  $C_i$  a variável de censura. Então observamos  $Y_i = \min(T_i, C_i)$ .
- Em geral, a censura é incorporada no modelo por meio da função  $F_Y(t)$  ou  $S_Y(t) = 1 - F_Y(t)$ .
- **Portnoy (2003)** propôs uma adaptação do estimador de **Kaplan-Meier** para estimar quantis condicionais na presença de censura.
- **Peng and Huang (2008)** propuseram uma adaptação do estimador de **Nelson-Aalen** para estimar quantis condicionais na presença de censura.
- Detalhes teóricos podem ser vistos no manual da *procedure* **QUANTLIFE**.



# Modelo dinâmico de regressão quantílica

- Proposta de **Goncalves, Migon, and Bastos (2020)** que incorpora a teoria de modelos dinâmicos (espaço de estados) em regressão quantílica.
- Construção do modelo sob o paradigma Bayesiano.
- Seja  $y_t$  uma variável aleatória no tempo  $t$ , o modelo proposto tem como objetivo estimar

$$Q_\tau(y_t) = \mathbf{F}_t^\top \boldsymbol{\theta}_t^\tau$$

em que  $\mathbf{F}_t$  é um vetor preditoras no tempo  $t$  e  $\boldsymbol{\theta}_t^\tau$  são os coeficientes que dependem do tempo  $t$  e do quantil  $\tau$ .

- Utilizando a representação de mistura os autores definem o modelo como

$$\begin{aligned} y_t \mid \boldsymbol{\theta}_t, \phi, \tau &\sim LA(\mathbf{F}_t^\top \boldsymbol{\theta}_t^\tau, \phi^{-1/2}, \tau) \\ \boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1} &\sim N_p(\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t), \quad t = 1, \dots, T. \end{aligned}$$

- Métodos MCMC são empregados para obter as distribuições a posteriori.

# Referências

**Barrodale, I. and F. D. K. Roberts** (1973). "An Improved Algorithm for Discrete L1 Linear Approximation". In: *SIAM Journal on Numerical Analysis* 10, pp. 839-848.

**Chen, C.** (2007). "A Finite Smoothing Algorithm for Quantile Regression." In: *Journal of Computational and Graphical Statistics* 16, pp. 136-164.

**Gilchrist, W. G.** (2000). *Statistical Modelling with Quantile Functions*. Ed. by N. York. Chapman & Hall/CRC.

**Goncalves, K. C. M, H. S. Migon, and L. S. Bastos** (2020). "Dynamic Quantile Linear Models: A Bayesian Approach". In: *Bayesian Analysis* 15.2, pp. 335-362.

**Gutenbrunner, C, J. Jureckova, R. Koenker, et al.** (1993). "Tests of Linear Hypotheses Based on Regression Rank Scores". In: *Journal of Nonparametric Statistics* 2, pp. 307-331.

**He, X. and F. Hu** (2002). "Markov Chain Marginal Bootstrap". In: *Journal of the American Statistical Association* 97, pp. 783-795.

# Referências

James, B. (2008). *Probabilidade: Um curso em Nível Intermediário*. IMPA.

Koenker, R. (2005). *Quantile Regression*. Ed. by C. U. Press.

Koenker, R. and G. Bassett (1978). "Regression Quantiles". In: *Econometrica* 46.1, pp. 33-50.

Koenker, R. and G. W. Bassett (1982b). "Robust Tests for Heteroscedasticity Based on Regression Quantiles". In: *Econometrica* 50, pp. 43-61.

Koenker, R. and G. W. Bassett (1982a). "Tests of Linear Hypotheses and L1 Estimation". In: *Econometrica* 50, pp. 1577-1583.

Koenker, R. and V. d'Orey (1994). "A Remark on Algorithm AS 229: Computing Dual Regression Quantiles and Regression Rank Scores". In: *Journal of the Royal Statistical Society, Series C* 43, pp. 410-414.

Koenker, R. and A. F. Machado (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression." In: *Journal of the American Statistical Association* 94, pp. 1296-1310.

# Referências

Kozumi, H. and G. Kobayashi (2011). "Gibbs sampling methods for Bayesian quantile regression". In: *Journal of Statistical Computation and Simulation* 81.11, pp. 1565-1578.

Mazucheli, J, A. F. B. Menezes, L. B. Fernandes, et al. (2020). "The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates". In: *Journal of Applied Statistics* 47.6, pp. 954-974.

Mosteller, F. and J. Tukey (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley.

Noufaily, A. and M. C. Jones (2013). "Parametric quantile regression based on the generalized gamma distribution". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62.5, pp. 723-740.

Peng, L. and Y. Huang (2008). "Survival Analysis with Quantile Regression Models". In: *Journal of the American Statistical Association* 103, pp. 637-649.

Portnoy, S. (2003). "Censored Regression Quantiles". In: *Journal of the American Statistical Association* 98, pp. 1001-1012.

# Referências

Portnoy, S. and R. Koenker (1997). "The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators". In: *Statistical Science* 12.4, pp. 279-300.

Sanchez, L, V. Leiva, M. Galea, et al. (2020). "Birnbbaum-Saunders quantile regression and its diagnostics with application to economic data". In: *Applied Stochastic Models in Business and Industry*.

Shou, Y. and M. Smithson (2019). "cdfquantreg: An R Package for CDF-Quantile Regression". In: *Journal of Statistical Software* 88.1, pp. 1-30.

Smithson, M. and Y. Shou (2017). "CDF-Quantile Distributions for Modeling Random Variables on the Unit Interval." In: *British Journal of Mathematical and Statistical Psychology* 70.3, pp. 412-438.

Obrigado!