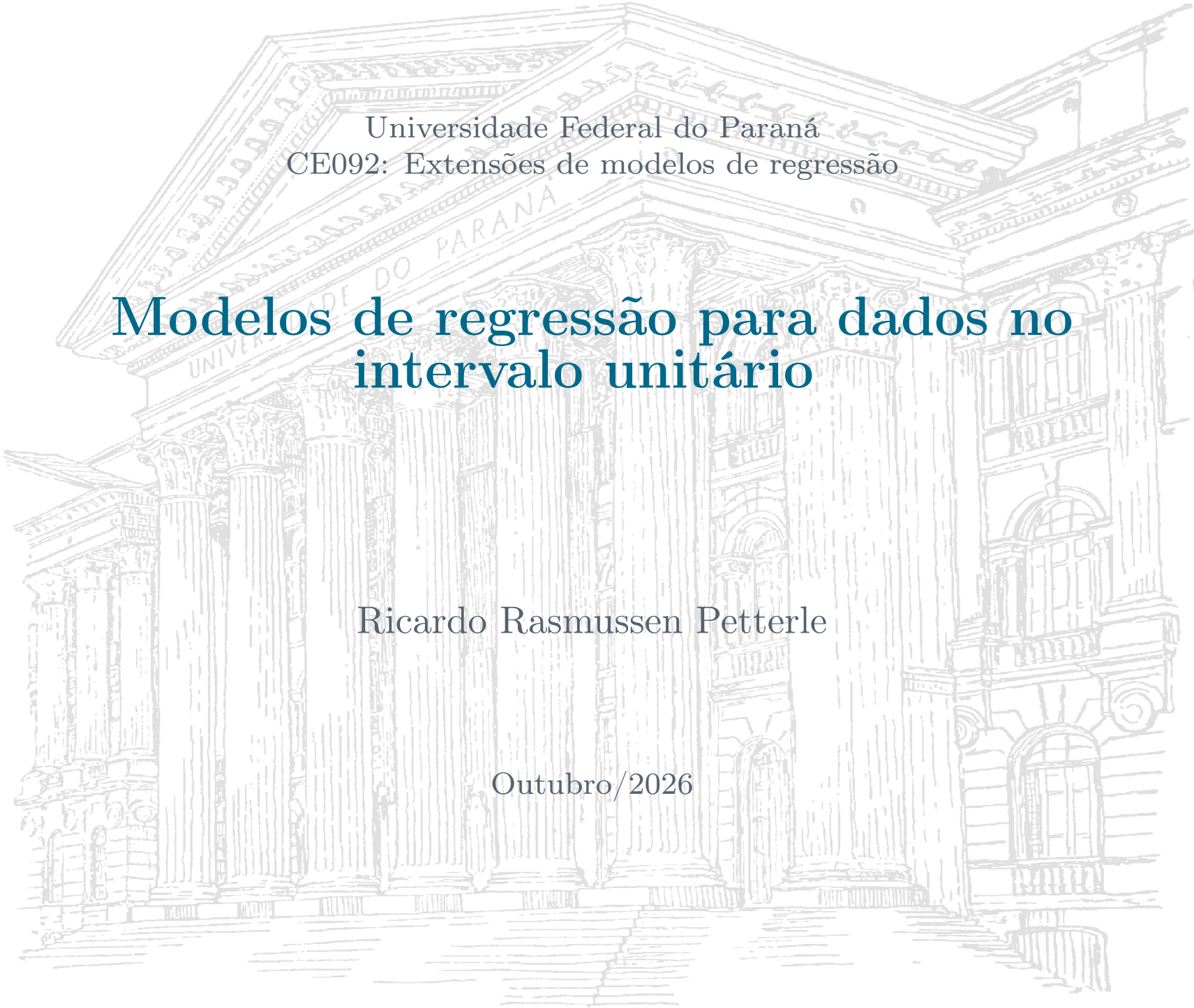


Universidade Federal do Paraná
CE092: Extensões de modelos de regressão

Modelos de regressão para dados no intervalo unitário

Ricardo Rasmussen Petterle

Outubro/2026

Que tipo de dado estamos modelando?

Muitos problemas aplicados geram respostas limitadas:

- **Exemplos**:
  - Índice de qualidade da água;
  - Percentual de gordura corporal;
  - Proporção de locais doentes para um determinado tipo de dente;
  - Percentual de consumo de álcool nos últimos 30 dias.
- Variáveis restritas aos intervalos: $(0, 1)$, $[0, 1)$, $(0, 1]$, ou $[0, 1]$
- **Principais áreas**:
  1. Medicina, psicologia, biologia;
  2. Ciências ambientais;
  3. Economia, ciências sociais e políticas;
  4. Engenharias, ciências da computação e etc.

Proporção binomial versus resposta contínua unitária

Proporção binomial

$$Y_i = \frac{X_i}{n_i}, \qquad X_i \sim \text{Bin}(n_i, p_i).$$

Existe um **número conhecido** de tentativas n_i e **contamos** quantos **sucessos** ocorreram.

Exemplos:

- 18 sementes germinaram entre 25 plantadas;
- 7 pacientes responderam ao tratamento entre 20 tratados;
- 42 peças defeituosas entre 500 produzidas.

Resposta contínua unitária

$$Y_i \in (0, 1).$$

Não existe necessariamente uma **contagem de sucessos** nem um denominador natural conhecido.

Exemplos:

- Índice de qualidade da vida;
- Percentual da renda que uma família gasta com alimentação;
- Índice de desenvolvimento humano municipal (IDHM).

Dados entre 0 e 1: qual é a natureza da resposta?

Contagem de sucessos
X_i sucessos em n_i tentativas

↓

Binomial
ou **beta-binomial**

Resposta unitária contínua
$Y_i \in (0, 1)$

↓

Beta
e **outras distribuições** unitárias

O suporte em [0, 1] não determina sozinho o modelo: é preciso considerar **como a resposta foi gerada**.

O primeiro diagnóstico é conceitual

Antes de escolher o modelo, pergunte:

1. A resposta é **contínua** ou **discreta**?
2. O intervalo é **aberto** $(0, 1)$ ou **fechado** $[0, 1]$?
3. Existem **zeros** ou **uns** exatos?
4. Existem **medidas repetidas**, **dados agrupados** ou **municípios**?
5. O interesse é na **média**, na **mediana**, em um **quantil** ou no valor **modal**?

Essas respostas já sugerem o tipo de modelo.

Por que a regressão linear parece tentadora?

- O modelo linear normal é familiar:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

  Ele permite:

  - interpretação direta dos coeficientes;
  - estimação simples por mínimos quadrados;
  - diagnóstico conhecido;
  - implementação em qualquer software.

- Mas essas vantagens podem vir com um custo: **o modelo não respeita o suporte da resposta**.

Problema 1: valores ajustados fora do intervalo

- Se usamos
$$\widehat{Y}_i = \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}},$$
- O valor ajustado pode ser menor que 0 ou maior que 1.
- Um modelo pode prever índice de qualidade igual a $\mathbf{1.08}$ ou proporção igual a $-\mathbf{0.04}$.
- Isso não é apenas um problema estético: é uma **incoerência** entre o **modelo estatístico** e a **variável observada**.

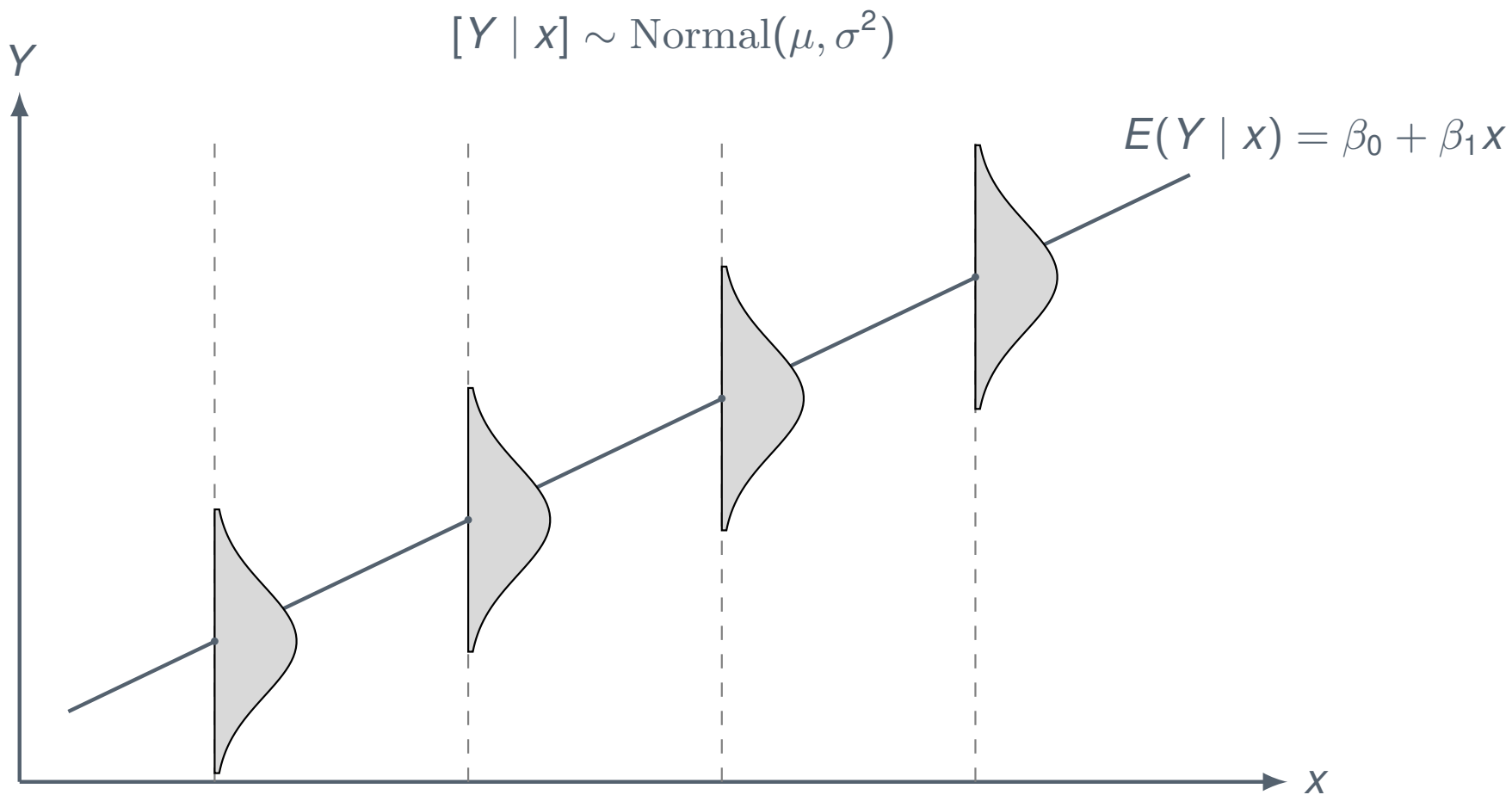
Problema 2: variância constante raramente é adequada

- O modelo linear assume:

$$\mathrm{Var}(Y_i \mid \mathbf{x}_i) = \sigma^2.$$

- Se μ_i está perto de 0 ou 1, a resposta tem **menos espaço para variar**;
- Uma resposta com média 0.95 não pode **variar simetricamente** como uma resposta com média 0.50;
- No modelo linear gaussiano, a média depende de x;
- Mas, a **variância condicional** permanece **constante**.

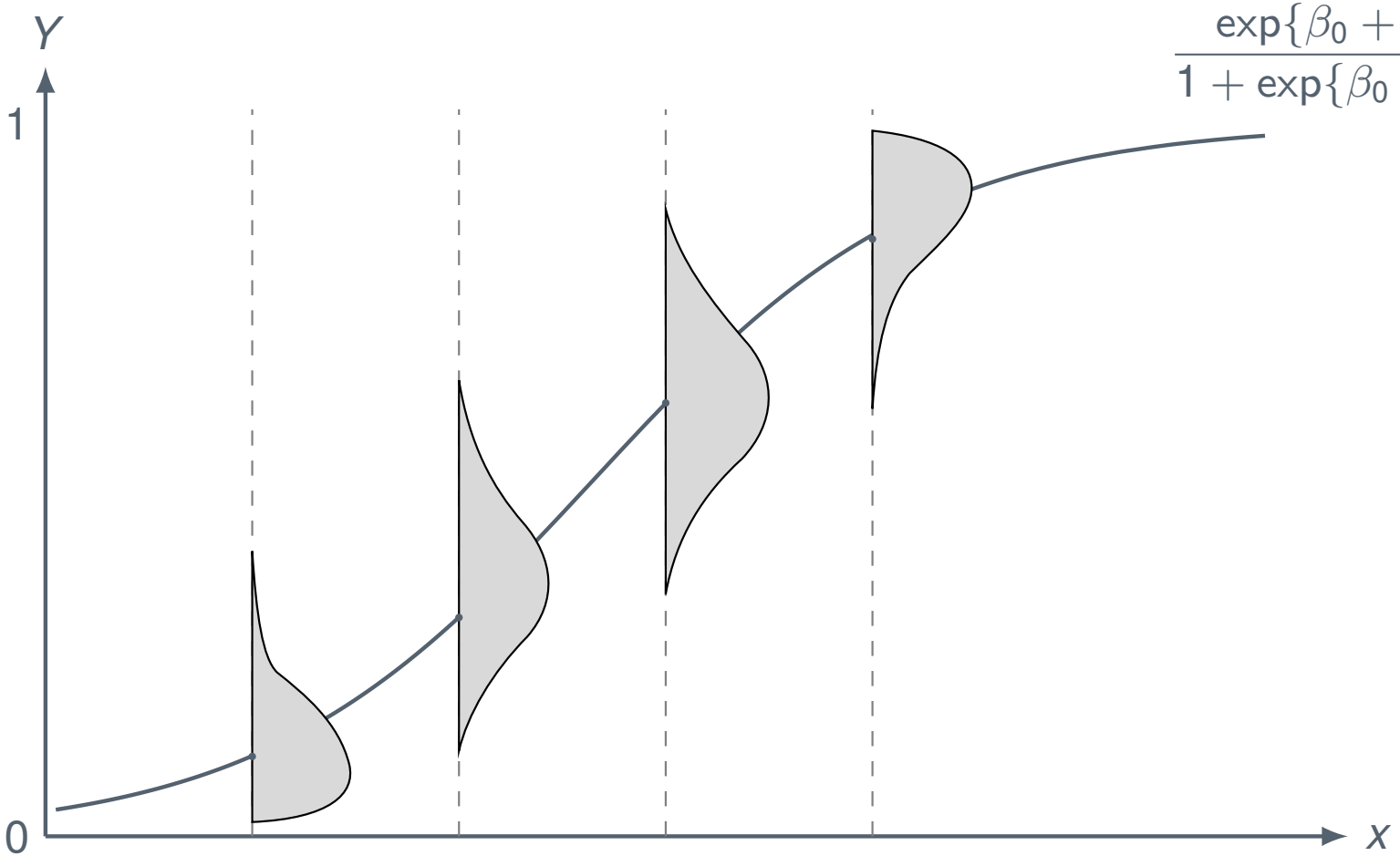
Modelo linear gaussiano



Regressão beta

$$[Y \mid x] \sim \text{Beta}(\mu, \phi)$$

$$\frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}$$



Problema 3: assimetria

Distribuições para **dados no intervalo unitário** podem ser:

- assimétricas à direita;
- assimétricas à esquerda;
- aproximadamente simétricas;
- em forma de J, J invertido ou U;
- concentradas perto de **zero** ou perto de **um**.

A forma da densidade importa, especialmente perto dos limites.

Exemplo em R: simulando dados em (0, 1)

- Vamos gerar uma **resposta beta** com uma **covariável**.

```
set.seed(123)
n <- 300
x <- rnorm(n, mean = 0, sd = 1)
eta <- -0.2 + 1.1 * x
mu <- plogis(eta)
phi <- 5
y <- rbeta(n, shape1 = mu * phi,
           shape2 = (1 - mu) * phi)
dat <- data.frame(y = y, x = x, mu = mu)
```

- A função `plogis()` garante que $0 < \mu_i < 1$.

Ajustando um modelo linear inadequado

Mesmo que os dados estejam em $(0, 1)$, podemos ajustar:

```
fit_lm <- lm(y ~ x, data = dat)
summary(fit_lm)
range(predict(fit_lm))
par(mfrow = c(2,2), oma = c(0, 0, 1.1, 0))
plot(fit_lm)
```

- O que observar:
  - Os valores ajustados **podem sair do intervalo**.
  - A variabilidade dos resíduos muda ao longo de x.

Ideia geral: modelar uma quantidade em (0, 1)

- Em vez de modelar diretamente Y_i por uma **reta**, modelamos um **parâmetro de locação**:

$$\mu_i = \mathrm{E}(Y_i \mid \mathbf{x}_i), \qquad 0 < \mu_i < 1.$$

- Usamos uma **função de ligação**:

$$g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

- Assim, η_i pode assumir qualquer **valor real**, enquanto μ_i permanece no **intervalo unitário**.

Ligação logit

- A ligação **logit** é definida por:

$$\text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \eta_i.$$

- A **inversa** é:

$$\mu_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}.$$

- E o **preditor linear** é:

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

Funções de ligação

- Para uma resposta com média $\mu_i \in (0, 1)$, especificamos

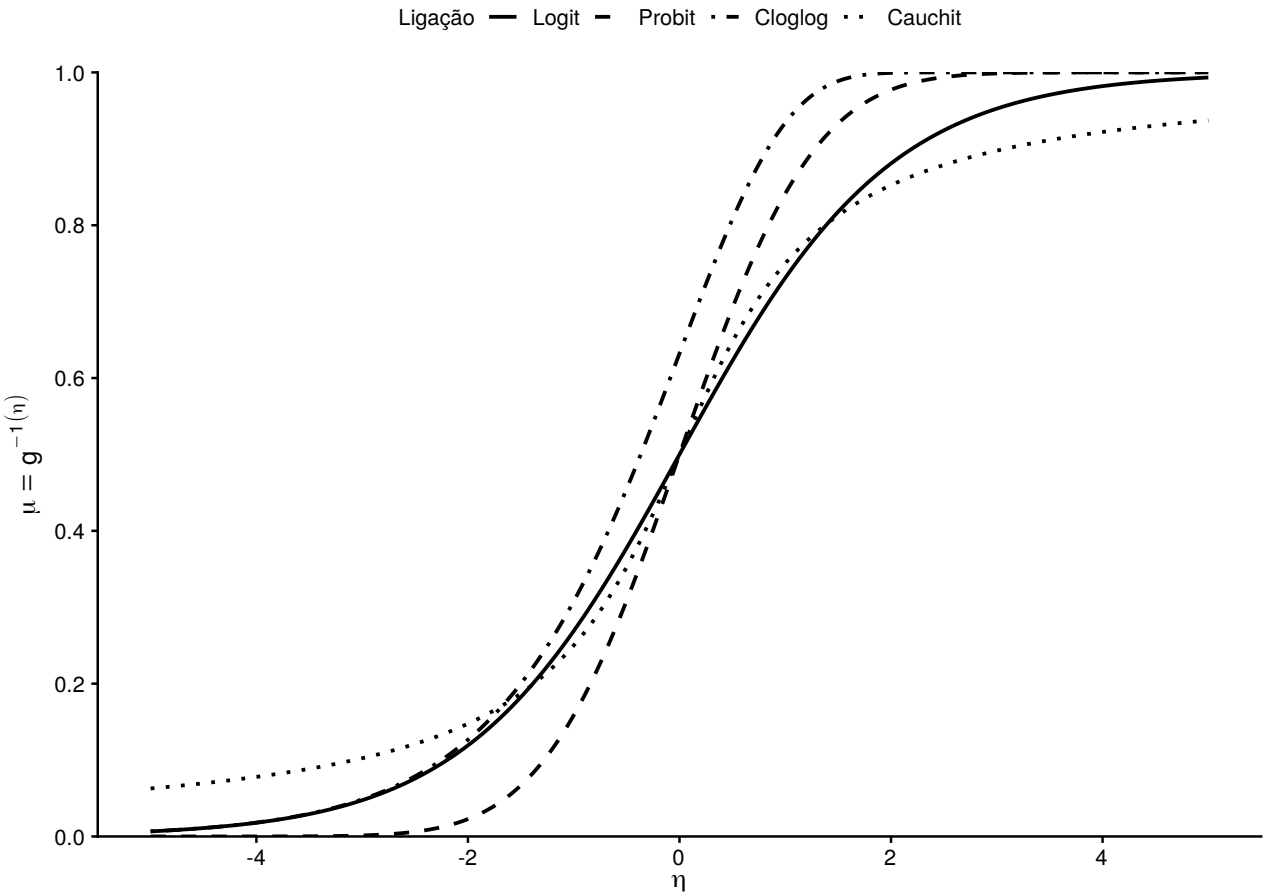
$$g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

- As ligações mais usuais incluem,

$$\begin{aligned}
&\textbf{logit}: && g(\mu) = \log\left(\frac{\mu}{1-\mu}\right),\\
&\textbf{probit}: && g(\mu) = \Phi^{-1}(\mu),\\
&\textbf{cloglog}: && g(\mu) = \log\{-\log(1-\mu)\},\\
&\textbf{cauchit}: && g(\mu) = \tan\{\pi(\mu - 0.5)\}.
\end{aligned}$$

- $g(\cdot) : (0, 1) \to \mathbb{R}$;
- $\Phi(\cdot)$ é a função distribuição da normal padrão.

Comparação das funções de ligação



- Todas as **funções inversas** transformam $\eta_i \in \mathbb{R}$ em $\mu_i \in (0, 1)$, mas diferem quanto à **simetria** e ao comportamento nas **caudas**.

Distribuição beta

- Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias **independentes** com função densidade de probabilidade dada por

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma[(1-\mu)\phi]} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},$$

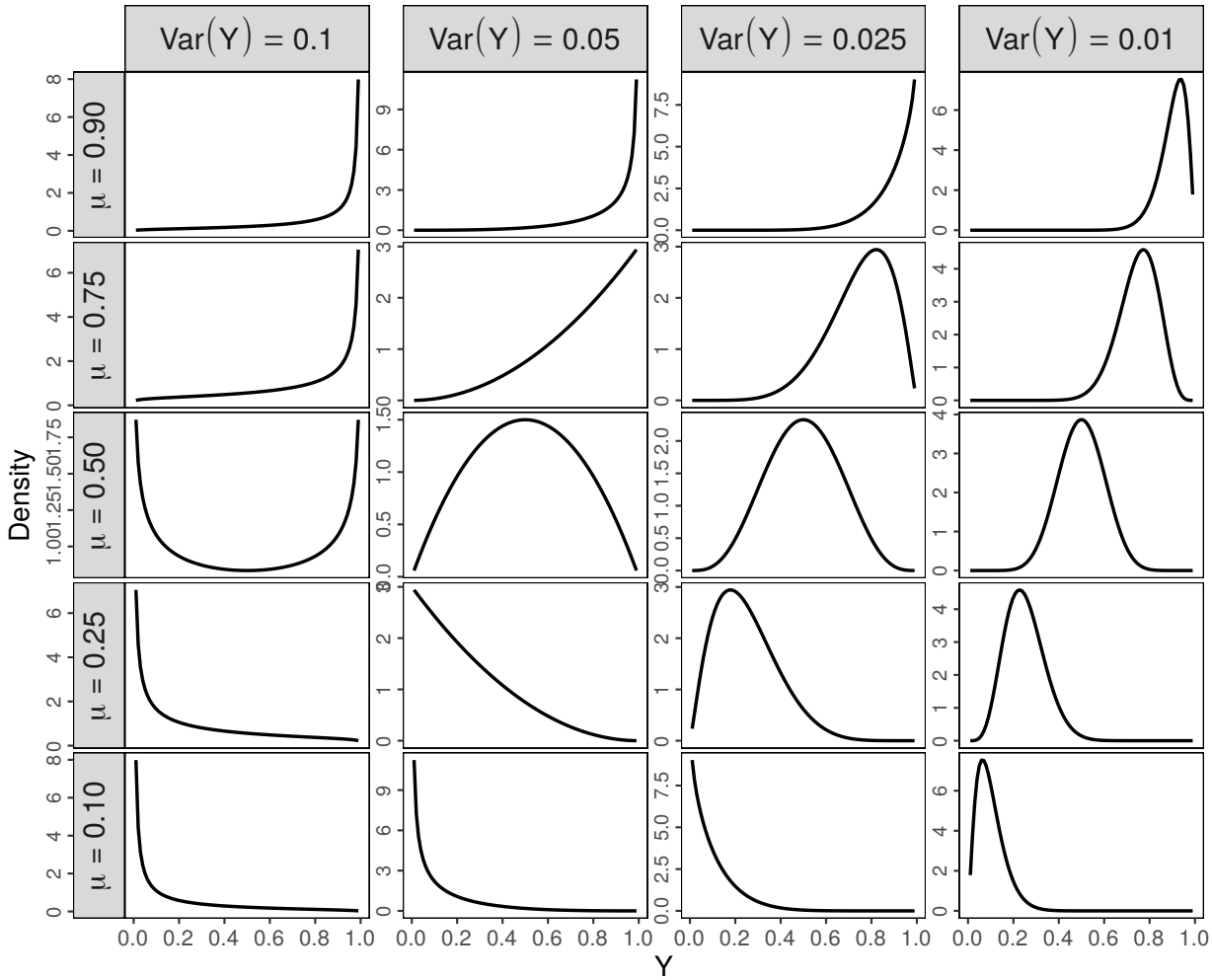
em que $y, \mu \in (0,1)$ e $\phi > 0$ denota o parâmetro de **precisão**.

- De acordo com Ferrari e Cribari-Neto (2004) a **esperança** e a **variância** são dadas, respectivamente, por

$$\mathrm{E}(Y) = \mu, \qquad \mathrm{Var}(Y) = \frac{\mu(1-\mu)}{1+\phi}.$$

- Quando $\phi \to 0$, a $\mathrm{Var}(Y) \to \mu(1-\mu)$.
- Relação média e variância da distribuição Bernoulli;

Densidade beta



- Para $\mathrm{Var}(Y) \in \{0.1, 0.05, 0.025, 0.01\}$ e $\mu = 0.5$;
- Tem-se $\phi \in \{1.5, 4, 9, 24\}$.

Modelo de regressão beta

- Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes, tais que

$$Y_i \sim \text{Beta}(\mu_i, \phi), \qquad i = 1, \ldots, n.$$

- Assim, o modelo de regressão beta é especificado por

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta},$$

- $g(\cdot)$ é uma função de ligação;
- $\mathbf{x}_i$ é um vetor $p \times 1$ de covariáveis conhecidas;
- $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros de regressão desconhecidos;
- Esse modelo assume **precisão constante**.

Verossimilhança do modelo beta

- Para observações independentes,

$$L(\beta, \phi) = \prod_{i=1}^{n} f(y_i; \mu_i, \phi),$$

em que

$$\mu_i = g^{-1}(\mathbf{x}_i^\top \beta).$$

- A log-verossimilhança é

$$\ell(\beta, \phi) = \sum_{i=1}^{n} \ell_i(\mu_i, \phi),$$

em que

$$\begin{aligned}\ell_i(\mu_i, \phi) = &\log \Gamma(\phi) - \log \Gamma(\mu_i\phi) - \log \Gamma((1-\mu_i)\phi) + (\mu_i\phi - 1)\log y_i \\ &+ \{(1-\mu_i)\phi - 1\}\log(1-y_i).\end{aligned}$$

O pacote `betareg`

- O pacote `betareg` é a implementação usual da regressão beta no R.
  - Ajusta modelos beta por **máxima verossimilhança**;
  - Permite diferentes funções de ligação para a média;
  - Permite precisão **constante** ou precisão **variável**;
  - Possui métodos usuais: `summary()`, `predict()`, `residuals()`, `logLik()`, `AIC()`.

Exemplo em R: regressão beta com `betareg`

- Ajustando o **modelo beta** aos dados simulados;

```
# install.packages("betareg")
library(betareg)

fit_beta <- betareg(y ~ x, data = dat)
summary(fit_beta)
range(predict(fit_beta))
```

- O modelo ajustado é:

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 x_i.$$

- E os coeficientes estimados são:

$$\hat{\beta}_0 = -0.229, \quad \hat{\beta}_1 = 1.103, \quad \hat{\phi} = 4.869.$$

Exemplo 1: `FoodExpenditure`

- O conjunto `FoodExpenditure` contém informações sobre a **proporção da renda familiar gasta com alimentação**.
- **Variáveis**:
  - `food:` gasto familiar com alimentação;
  - `income:` renda familiar;
  - `persons`: número de pessoas no domicílio.
- Definimos a **resposta** como:

$$Y_i = \frac{\texttt{food}_i}{\texttt{income}_i}, \qquad i = 1, \ldots, 38.$$

- Isto é, a proporção da renda destinada à alimentação;
- Como $Y_i \in (0, 1)$, a **regressão beta** é uma escolha natural.

Exemplo 1: construindo a resposta


```
library(betareg)

data("FoodExpenditure", package = "betareg")
FoodExpenditure$prop <- with(FoodExpenditure, food / income)

summary(FoodExpenditure$prop)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1075  0.2269  0.2611  0.2897  0.3469  0.5612

 head(FoodExpenditure)
    food income persons      prop
1 15.998 62.476       1 0.2560663
2 16.652 82.304       5 0.2023231
3 21.741 74.679       3 0.2911260
4  7.431 39.151       3 0.1898036
5 10.481 64.724       5 0.1619337
6 13.548 36.786       3 0.3682923
```

- Cada observação representa a fração da **renda familiar** destinada à alimentação.

Exemplo 1: modelo linear como ponto de partid

Antes da regressão beta, podemos ajustar um **modelo linear** apenas como referência.

```
fit_lm <- lm(prop ~ income + persons,
             data = FoodExpenditure)

summary(fit_lm)
```

Modelo ajustado:

$$Y_i = \beta_0 + \beta_1 \texttt{income}_i + \beta_2 \texttt{persons}_i + \varepsilon_i.$$

Limitação

A **média** condicional do **modelo linear** não é restrita ao intervalo $(0, 1)$.

Exemplo 1: ajuste com modelo de regressão beta

- Agora, especificamos,

$$Y_i \sim \text{Beta}(\mu_i, \phi),$$

com

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \texttt{income}_i + \beta_2 \texttt{persons}_i.$$

```
fit_beta <- betareg(prop ~ income + persons,
                    data = FoodExpenditure)

summary(fit_beta)
```

- Coeficientes estimados:

$$\hat{\beta}_0 = -0.623, \qquad \hat{\beta}_1 = -0.012, \qquad \hat{\beta}_2 = 0.118.$$

Exemplo 1: como interpretar os coeficientes?

- Com ligação logit,

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \texttt{income}_i + \beta_2 \texttt{persons}_i,$$

- **Renda**:

$$\exp(-0.012) = 0.988.$$

  Mantido o número de pessoas constante, o aumento de uma unidade na renda reduz em aproximadamente **1,2%** a razão $\mu_i/(1-\mu_i)$.

- **Número de pessoas**:

$$\exp(0.118) = 1.125.$$

  Mantida a renda constante, uma pessoa adicional aumenta em aproximadamente **12,5%** a razão $\mu_i/(1-\mu_i)$.

Até aqui: modelo de regressão beta com precisão constante

- O modelo usado no primeiro exemplo foi

$$Y_i \sim \text{Beta}(\mu_i, \phi), \qquad \text{logit}(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta},$$

  com **precisão constante**.

- Nesse caso,

$$\text{Var}(Y_i) = \frac{\mu_i(1-\mu_i)}{1+\phi}.$$

- E se a variabilidade também mudar de acordo com covariáveis?

Precisão variável

- O modelo de regressão beta pode ter duas **estruturas de regressão**:

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \qquad h(\phi_i) = \mathbf{z}_i^\top \boldsymbol{\gamma}.$$

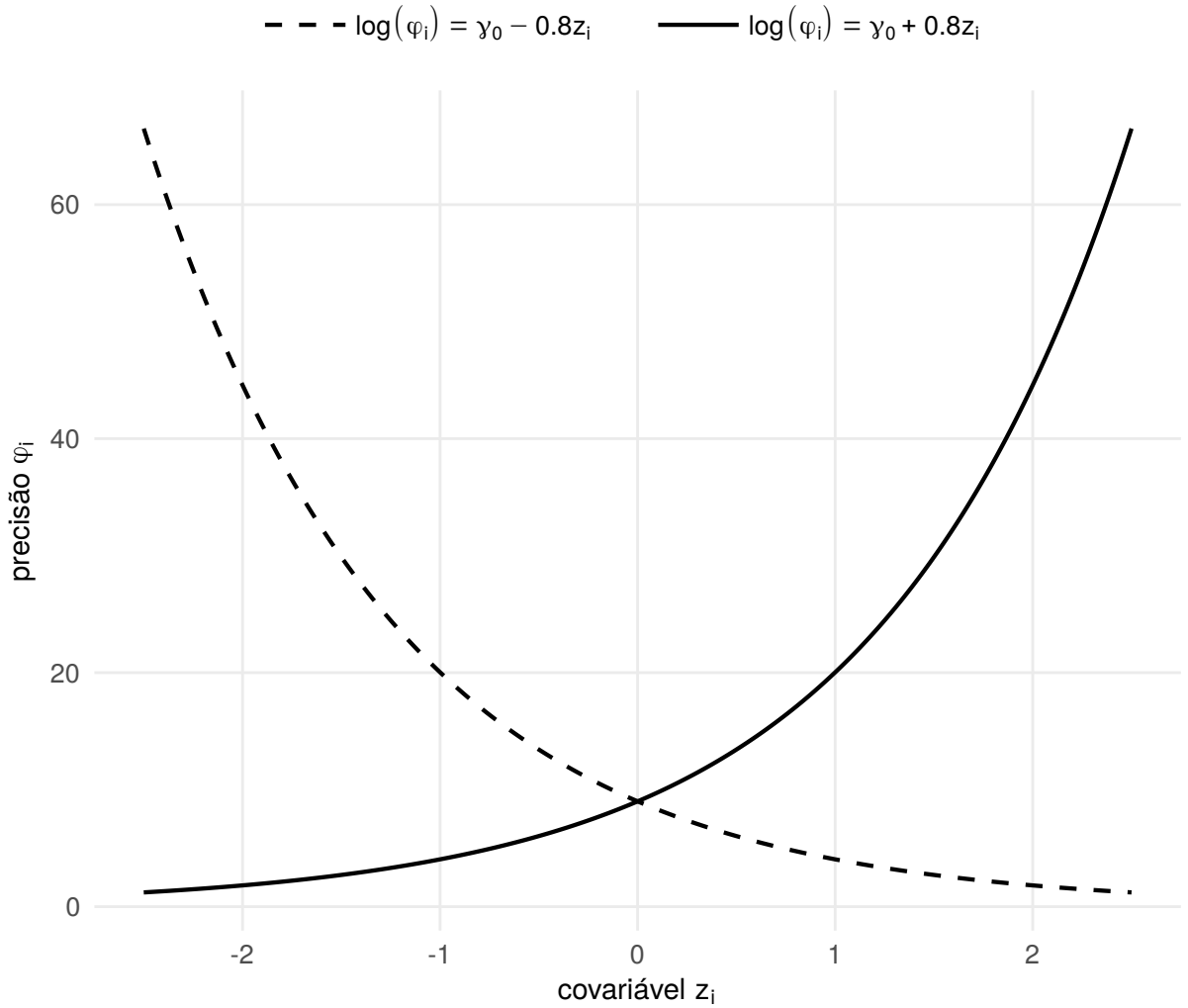
- Usualmente,

$$h(\phi_i) = \log(\phi_i),$$

  pois $\phi_i > 0$.

- **Interpretação**:
  - A estrutura da **média** descreve a locação da resposta;
  - A estrutura da **precisão** descreve a concentração em torno da média;
  - **Maior** ϕ_i implica **menor** variabilidade condicional.

Precisão variável: efeito de uma covariável



Quando $\log(\phi_i) = \mathbf{z}_i^\top \boldsymbol{\gamma}$, os efeitos são multiplicativos na precisão.

Sintaxe do `betareg` com precisão variável

- No `betareg`, a barra vertical separa a estrutura da **média** e a estrutura da **precisão**.

```
fit_beta2 <- betareg(y ~ x1 + x2 | z1 + z2,
                     data = dados)
```

- Isso representa

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2},$$
$$\log(\phi_i) = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2}.$$

- Caso particular `y ~ x1 + x2` equivale a usar **precisão constante**.

Exemplo 2: `GasolineYield`

- O conjunto `GasolineYield` é frequentemente usado para ilustrar regressão beta com precisão variável.

Variável	Papel no exemplo
`yield`	resposta contínua no intervalo unitário
`batch`	fator associado ao lote do petróleo bruto
`temp`	temperatura final de destilação

A ideia é permitir que a temperatura afete tanto a média quanto a precisão.

Exemplo 2: ajuste com precisão constante

- Primeiro, ajustamos um modelo beta com precisão constante:

```
data("GasolineYield", package = "betareg")

fit_gy1 <- betareg(yield ~ batch + temp,
                   data = GasolineYield)

summary(fit_gy1)
```

- Esse modelo assume

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \, \texttt{batch} + \beta_2 \, \texttt{temp}_i,$$

com ϕ constante.

Exemplo 2: ajuste com precisão variável

- Agora deixamos a precisão variar com a temperatura:

```
fit_gy2 <- betareg(yield ~ batch + temp | temp,
                   data = GasolineYield)

summary(fit_gy2)
```

- O modelo é

$$\text{logit}(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \qquad \log(\phi_i) = \gamma_0 + \gamma_1 \texttt{temp}_i.$$

- Se $\gamma_1 > 0$, maiores temperaturas estão associadas a maior precisão e menor variabilidade condicional.

Comparando modelos no exemplo 2

- Uma comparação natural é verificar se o submodelo de precisão melhora o ajuste:

$$\text{precisão constante} \quad \times \quad \text{precisão variável.}$$

  No R:

  - `summary(fit_gy1)` e `summary(fit_gy2)`;
  - `AIC(fit_gy1, fit_gy2)`;
  - análise de resíduos;
  - predições para **média** e **precisão**.

- A regressão beta permite separar efeitos sobre a média de efeitos sobre a dispersão.

Síntese da regressão beta

A regressão beta combina:

Modelo para a média

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

- Locação da resposta.
- Interpretação via ligação.
- Predições em $(0, 1)$.

Modelo para a precisão

$$h(\phi_i) = \mathbf{z}_i^\top \boldsymbol{\gamma}.$$

- Concentração em torno da média.
- Heterogeneidade da variabilidade.
- Estrutura opcional.

Mas continuamos assumindo que a forma distribucional é beta.

Por que considerar outras distribuições?

- A distribuição beta é flexível, mas não é a única distribuição possível em $(0, 1)$.
- Distribuições alternativas podem ser úteis quando queremos:
  - representar assimetrias diferentes;
  - obter outras relações média-variância;
  - modelar caudas de forma distinta;
  - parametrizar diretamente média, quantil ou moda;
  - melhorar o ajuste sem alterar o suporte da resposta.
- A escolha distribucional afeta a **forma da densidade** e a relação entre **média e variância**.

Algumas distribuições para dados em $(0, 1)$

Distribuição	Dispersão/precisão	Comentário
Beta	precisão	referência clássica
Unit-gamma	precisão	assimetria alternativa
Simplex	dispersão	relação média–variância distinta
Bessel	precisão	alternativa flexível
Vasicek	dispersão	construção via normal bivariada

- O ponto principal é entender que **diferentes distribuições** induzem diferentes **formas** e diferentes **variâncias**.

Diferentes relações entre média e variância

Em todas as distribuições,

$$\mathrm{E}(Y) = \mu, \qquad 0 < \mu < 1,$$

mas a variância depende de μ e ϕ de formas distintas.

Beta: $\mathrm{Var}(Y) = \dfrac{\mu(1-\mu)}{1+\phi},$

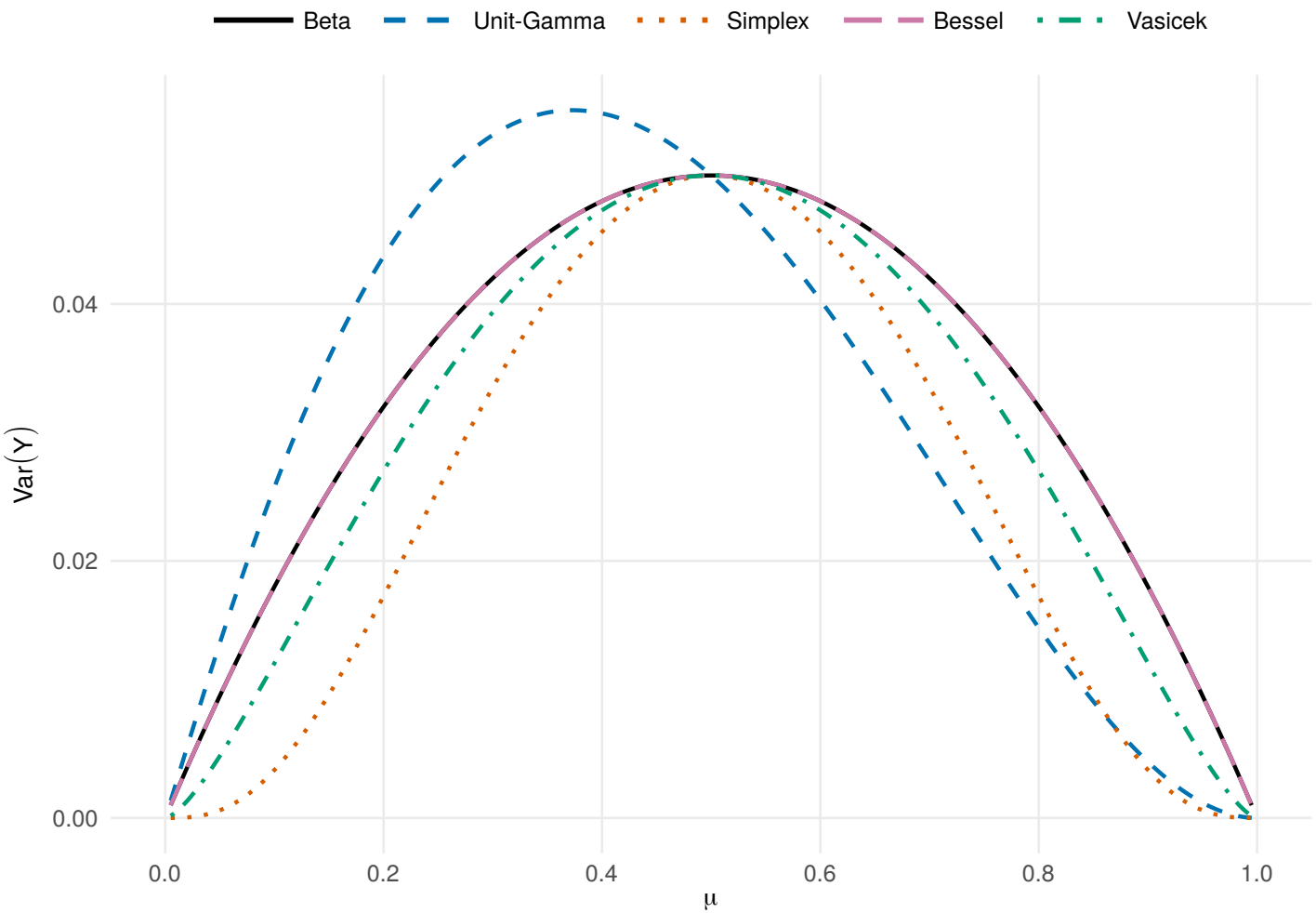
Unit-gamma: $\mathrm{Var}(Y) = \mu\left[\left(2-\mu^{1/\phi}\right)^{-\phi} - \mu\right],$

Simplex: $\mathrm{Var}(Y) = \mu(1-\mu) - \sqrt{\dfrac{1}{2\phi^2}}\exp\left\{\dfrac{1}{2\phi^2\mu^2(1-\mu)^2}\right\} \times \Gamma\left(\dfrac{1}{2}, \dfrac{1}{2\phi^2\mu^2(1-\mu)^2}\right),$

Bessel: $\mathrm{Var}(Y) = \dfrac{\mu(1-\mu)}{2}\left[1-\phi+\phi^2 e^{\phi} E_1(\phi)\right],$

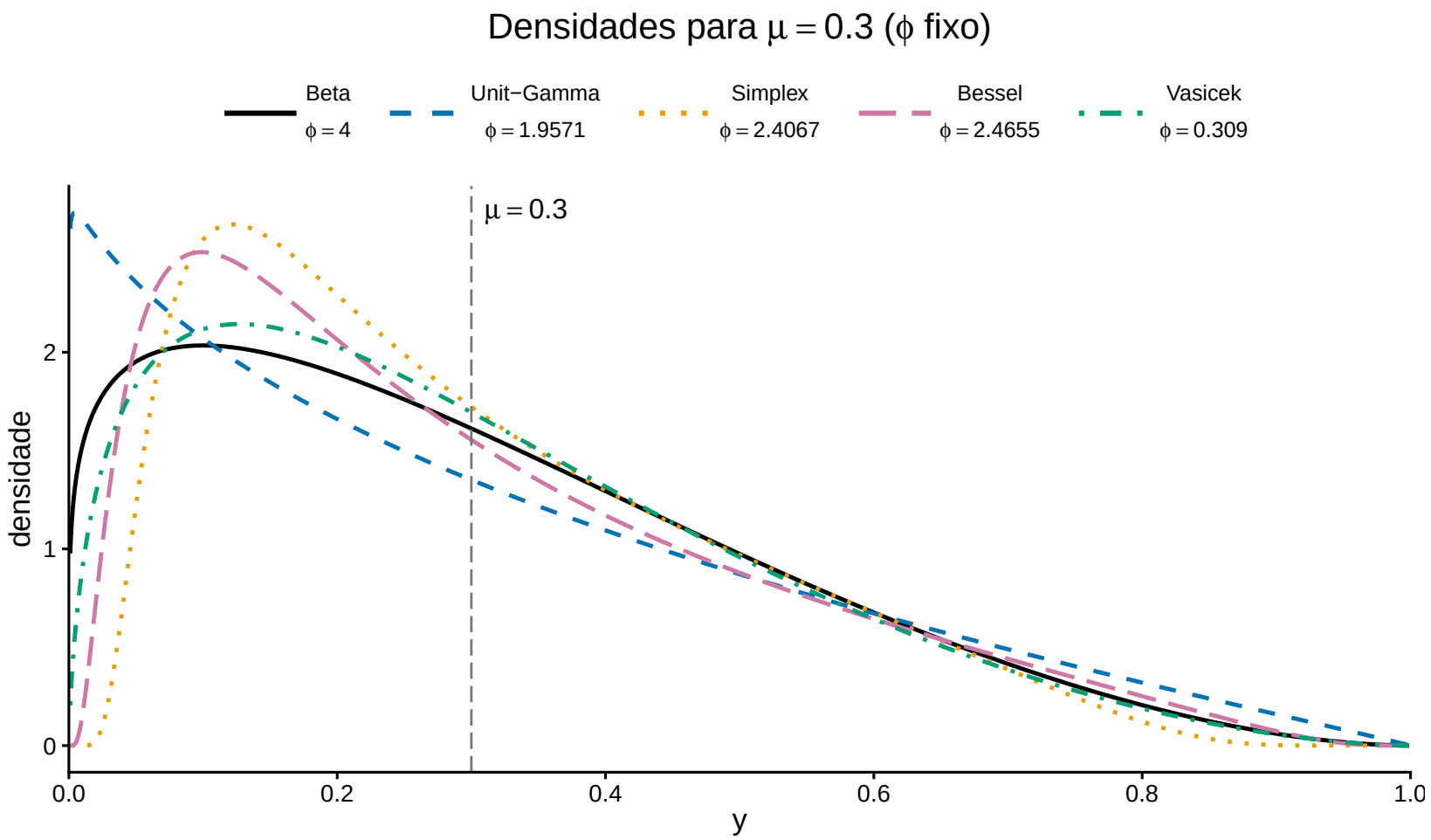
Vasicek: $\mathrm{Var}(Y) = \Phi_2\left\{\Phi^{-1}(\mu), \Phi^{-1}(\mu); \phi\right\} - \mu^2.$

Relação média e variância



- Fixar a mesma **média** e a mesma **variância** em um ponto não torna as distribuições equivalentes.

Densidade das cinco distribuições



O pacote `unitregTMB`

O pacote `unitregTMB` fornece uma interface comum para **modelos de regressão para respostas no intervalo unitário**.

- Para regressão na média,

$$g(\mu_i) = \mathbf{x}_i^\top \beta, \qquad \mu_i = E(Y_i \mid \mathbf{x}_i),$$

- Podemos alterar a distribuição condicional mantendo a mesma estrutura sistemática.

Beta Unit-gamma Simplex Bessel Vasicek

- Mesma resposta, mesmas covariáveis e mesma função de ligação;
- Muda apenas a distribuição assumida para $Y_i \mid \mathbf{x}_i$.

Instalação do pacote `unitregTMB`

- O pacote `unitregTMB` está disponível em:
  `https://github.com/rpetterle/unitregTMB`

- Instale o pacote `remotes`, caso ainda não esteja instalado:

```
install.packages("remotes")
```

- Instale a dependência `regcore`:

```
remotes::install_github("rpetterle/regcore")
```

- Instale o pacote `unitregTMB`:

```
remotes::install_github("rpetterle/unitregTMB")
```

Exemplo 3: Índice de qualidade da água

- O IQA foi desenvolvido nos EUA na década de 70;
- Avaliar a qualidade da água destinada ao abastecimento após tratamento;
- Em 1975 a CETESB começou a usar este índice;
- Principal indicador de qualidade da água;
- Obtido por meio de 9 parâmetros físico-químicos e biológicos;

$$\text{IQA} = \prod_{i=1}^{9} q_i^{w_i},$$

onde,

1. q_i: valor entre 0 e 100;
2. w_i: valor entre 0 e 1 (pesos);
3. Fácil interpretação;
4. Comparável com outras localidades;
5. Agência Nacional de Águas (ANA) e IAP.

Exemplo 3: Índice de qualidade da água

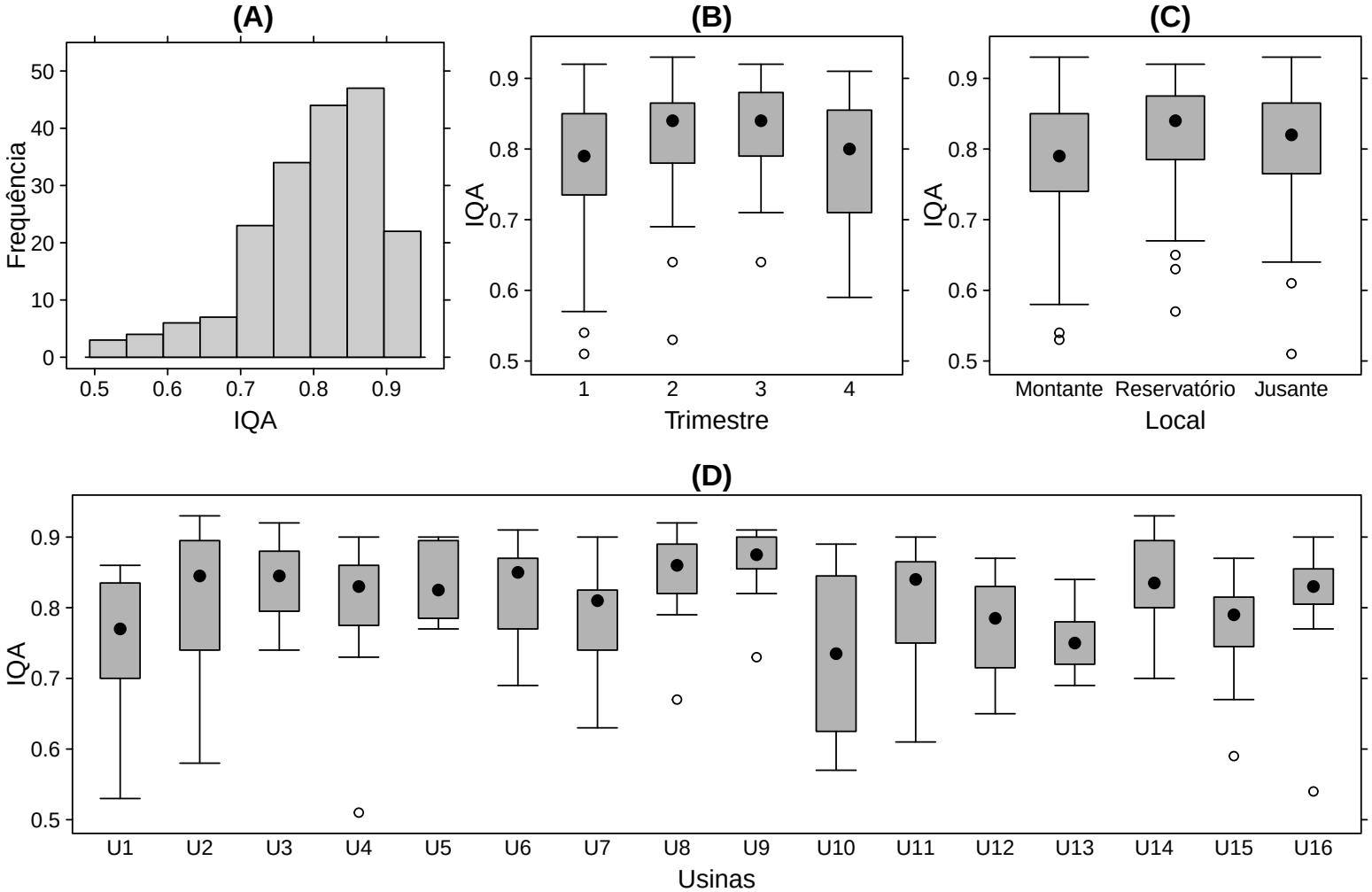
- A COPEL opera 16 usinas hidrelétricas no Estado do Paraná;
- A principal finalidade é geração de energia elétrica;
- Outras finalidades: pesca, navegação e abastecimento das cidades;
- Monitoramento **trimestral** da qualidade da água a montante, a jusante e nos reservatórios dos rios represados;
- **Principal objetivo**: detectar mudanças na qualidade da água atribuíveis à presença das barragens.

Exemplo 3: Índice de qualidade da água

- São 12 medidas (3 locais $\times$ 4 trimestres) em cada uma das 16 usinas;
- 190 observações com apenas dois dados faltantes.



Exemplo 3: Índice de qualidade da água



O que a análise exploratória sugere?

- A resposta está no intervalo unitário e se concentra em valores relativamente altos.
- O histograma sugere assimetria à esquerda.
- Os valores médios tendem a ser maiores nos trimestres 2 e 3.
- O reservatório apresenta valores de IQA maiores do que os demais locais.
- Há heterogeneidade entre usinas e, como existem medidas repetidas em cada usina, é plausível haver dependência intraunidade.

Consequência para modelagem

O conjunto é adequado para discutir modelos para respostas em $(0, 1)$ e, posteriormente, efeitos aleatórios.

Dados no `unitregTMB`


```
library(unitregTMB)
data("wqi")

head(wqi)
##   id    y quarter   location
## 1 U1 0.68       1   Upstream
## 2 U1 0.83       1  Reservoir
## 3 U1 0.76       1 Downstream
## 4 U1 0.53       2   Upstream
## 5 U1 0.75       2  Reservoir
## 6 U1 0.64       2 Downstream

dim(wqi)
## 190   4
```

Variável	Descrição
`id`	unidade amostral/usina
`y`	IQA reescalado em $(0, 1)$
`quarter`	trimestre de medição
`location`	local de coleta

Nesta etapa, vamos inicialmente ignorar a dependência dentro de `id` para focar na escolha distribucional.

Mesma média, diferentes distribuições

Vamos comparar modelos com a mesma estrutura sistemática para a média:

$$\text{logit}(\mu_i) = \beta_0 + \boldsymbol{\beta}_1^\top \texttt{location}_i + \boldsymbol{\beta}_2^\top \texttt{quarter}_i.$$

Em R:

```
y ~ location + quarter.
```

A única mudança será a distribuição condicional:

Beta	Unit-Gamma	Simplex	Bessel	Vasicek

Comparação justa

Mesma resposta, mesmas covariáveis e mesma função de ligação; muda apenas a distribuição assumida para $Y_i \mid \mathbf{x}_i$.

Ajustando os modelos


```
## Modelo beta
fit_beta <- unitregTMB(y ~ location + quarter,
                       family = beta(model_for = "mean"),
                       data = wqi)

## Modelo unit-gamma
fit_ug <- unitregTMB(y ~ location + quarter,
                     family = unitgamma(model_for = "mean"),
                     data = wqi)

## Modelo simplex
fit_simplex <- unitregTMB(y ~ location + quarter,
                          family = simplex(model_for = "mean"),
                          data = wqi)
```

- A estrutura de regressão permanece a mesma; apenas a **distribuição condicional** é **modificada**.

Outras alternativas

A mesma especificação pode ser usada com outras distribuições parametrizadas pela média:

```
fit_bessel <- unitregTMB(y ~ location + quarter,
                         family = bessel(model_for = "mean"),
                         data = wqi)

fit_vasicek <- unitregTMB(y ~ location + quarter,
                          family = vasicek(model_for = "mean"),
                          data = wqi)
```

Beta, Unit-gamma, Simplex, Bessel, Vasicek.

Comparando os modelos

Em vez de comparar cinco saídas completas, podemos organizar os critérios de ajuste em uma tabela:

```
gof_tab(fit_beta, fit_ug, fit_simplex, fit_bessel, fit_vasicek)
```

Modelo	logLik	AIC	BIC
Beta	224.62	-435.25	-412.52
Unit-Gamma	227.79	-441.57	-418.84
Simplex	**232.44**	**-450.88**	**-428.15**
Bessel	228.39	-442.77	-420.04
Vasicek	227.65	-441.29	-418.56

O que aprendemos com o exemplo?

- A média condicional pode ter a mesma interpretação nos cinco modelos:

$$\mu_i = E(Y_i \mid \mathbf{x}_i).$$

- A escolha distribucional altera a forma da **distribuição condicional** e a **relação média-variância**.
- Critérios como **log-verossimilhança, AIC e BIC** ajudam a comparar os modelos.
- O conjunto `wqi` também será útil para estudar **efeitos aleatórios**.
- Mais adiante, retomaremos o `wqi` com

```
y ~ location + quarter + (1 | id).
```

Contato

- Nome: Ricardo Rasmussen Petterle.
- Estatístico.
- Setor de Ciências da Saúde.
- Universidade Federal do Paraná.
- E-mail: ricardopetterle@ufpr.br

- Obrigado!