

Regressão Beta para modelagem de taxas e proporções

MAE0006- Modelos de Regressão

Eliardo Guimarães da Costa

Beatriz Ariadna da Silva Ciríaco

Wesley Almeida Cruz

15 de Fevereiro de 2022

Programa de Pós-Graduação em Matemática Aplicada e Estatística (PPgMAE)

1. Motivação
2. Modelo de probabilidade Beta
3. Modelo de Regressão Beta
4. Análise de diagnóstico
5. Aplicação
6. Considerações Finais

Motivação

- Desejamos aplicar uma análise de regressão para uma variável resposta que está contida no intervalo $(0, 1)$;

- Desejamos aplicar uma análise de regressão para uma variável resposta que está contida no intervalo $(0, 1)$;
- Podemos utilizar uma transformação na variável original para que ela esteja contida nos \mathbb{R} ;

- Desejamos aplicar uma análise de regressão para uma variável resposta que está contida no intervalo $(0, 1)$;
- Podemos utilizar uma transformação na variável original para que ela esteja contida nos \mathbb{R} ;
- Porém, esse método tem três defeitos:

- Desejamos aplicar uma análise de regressão para uma variável resposta que está contida no intervalo $(0, 1)$;
- Podemos utilizar uma transformação na variável original para que ela esteja contida nos \mathbb{R} ;
- Porém, esse método tem três defeitos:
 - Perdemos a interpretação da regressão para a variável original;

- Desejamos aplicar uma análise de regressão para uma variável resposta que está contida no intervalo $(0, 1)$;
- Podemos utilizar uma transformação na variável original para que ela esteja contida nos \mathbb{R} ;
- Porém, esse método tem três defeitos:
 - Perdemos a interpretação da regressão para a variável original;
 - Geralmente regressões desse tipo possuem erros heterocedásticos;

- Desejamos aplicar uma análise de regressão para uma variável resposta que está contida no intervalo $(0, 1)$;
- Podemos utilizar uma transformação na variável original para que ela esteja contida nos \mathbb{R} ;
- Porém, esse método tem três defeitos:
 - Perdemos a interpretação da regressão para a variável original;
 - Geralmente regressões desse tipo possuem erros heterocedásticos;
 - As distribuições geralmente são assimétricas e violam condições de normalidade de Gauss-Markov.

- Uma alternativa mais natural para esse problema foi proposta por Ferrari e Cribari-Neto (2004), o modelo de regressão Beta;

- Uma alternativa mais natural para esse problema foi proposta por Ferrari e Cribari-Neto (2004), o modelo de regressão Beta;
- É usado para modelagem de variáveis aleatórias contínuas que assumem valores no intervalo (a, b) de tal forma que $a < b; a, b \in \mathbb{R}$;

- Uma alternativa mais natural para esse problema foi proposta por Ferrari e Cribari-Neto (2004), o modelo de regressão Beta;
- É usado para modelagem de variáveis aleatórias contínuas que assumem valores no intervalo (a, b) de tal forma que $a < b; a, b \in \mathbb{R}$;
- Muito útil para proporções (contínuas no intervalo $(0, 1)$);

- Na regressão Beta assumimos que a variável resposta segue uma distribuição Beta;

- Na regressão Beta assumimos que a variável resposta segue uma distribuição Beta;
- A interpretação dos parâmetros no modelo proposto é em termos da média da variável resposta e o modelo é naturalmente heterocedástico e é robusto a dados assimétricos;

- Na regressão Beta assumimos que a variável resposta segue uma distribuição Beta;
- A interpretação dos parâmetros no modelo proposto é em termos da média da variável resposta e o modelo é naturalmente heterocedástico e é robusto a dados assimétricos;
- Devido à flexibilidade natural da distribuição beta, a sua densidade pode assumir diversos formatos.

Modelo de probabilidade Beta

Seja X uma variável aleatória contínua que pertença a família de distribuição Beta com parâmetros de forma α e β ($X \sim \text{Beta}(\alpha, \beta)$), então a sua função densidade de probabilidade é dada por:

Seja X uma variável aleatória contínua que pertença a família de distribuição Beta com parâmetros de forma α e β ($X \sim \text{Beta}(\alpha, \beta)$), então a sua função densidade de probabilidade é dada por:

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}(x), \quad \alpha, \beta > 0,$$

Seja X uma variável aleatória contínua que pertença a família de distribuição Beta com parâmetros de forma α e β ($X \sim \text{Beta}(\alpha, \beta)$), então a sua função densidade de probabilidade é dada por:

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}(x), \quad \alpha, \beta > 0,$$

em que $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$ e $\Gamma(\cdot)$ é a função gama.

A esperança e variância da distribuição Beta é dada pelas seguintes equações:

A esperança e variância da distribuição Beta é dada pelas seguintes equações:

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \quad \text{e} \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

A esperança e variância da distribuição Beta é dada pelas seguintes equações:

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \quad \text{e} \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

A família de distribuição Beta é bastante versátil e uma gama de situações podem ser modeladas por essa distribuição com sucesso.

A esperança e variância da distribuição Beta é dada pelas seguintes equações:

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \quad \text{e} \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

A família de distribuição Beta é bastante versátil e uma gama de situações podem ser modeladas por essa distribuição com sucesso.

Entretanto, as aplicações dessa distribuição não incluem situações em que o objetivo é conduzir uma análise de regressão para modelar uma variável resposta em relação a covariáveis exógenas.

Modelo de Regressão Beta

- A modelagem e os procedimentos inferenciais propostos por Ferrari e Cribari-Neto (2004) são semelhantes aos modelos lineares generalizados (McCullagh Nelder, 1989);

Modelo de Regressão Beta

- A modelagem e os procedimentos inferenciais propostos por Ferrari e Cribari-Neto (2004) são semelhantes aos modelos lineares generalizados (McCullagh Nelder, 1989);
- Uma alternativa ao modelo de regressão Beta é o modelo Simplex proposto por Jørgensen (1997) que possui quatro parâmetros;

Modelo de Regressão Beta

- A modelagem e os procedimentos inferenciais propostos por Ferrari e Cribari-Neto (2004) são semelhantes aos modelos lineares generalizados (McCullagh Nelder, 1989);
- Uma alternativa ao modelo de regressão Beta é o modelo Simplex proposto por Jørgensen (1997) que possui quatro parâmetros;
- Normalmente, em um contexto de regressão, é normal modelar a média da variável resposta Y . Também é típico que o modelo possua algum parâmetro de dispersão (precisão);

- A modelagem e os procedimentos inferenciais propostos por Ferrari e Cribari-Neto (2004) são semelhantes aos modelos lineares generalizados (McCullagh Nelder, 1989);
- Uma alternativa ao modelo de regressão Beta é o modelo Simplex proposto por Jørgensen (1997) que possui quatro parâmetros;
- Normalmente, em um contexto de regressão, é normal modelar a média da variável resposta Y . Também é típico que o modelo possua algum parâmetro de dispersão (precisão);
- Para obter essa estrutura na modelagem, deve-se reparametrizar a distribuição Beta de forma que os seus parâmetros se tornem parâmetros de média e dispersão.

Modelo de Regressão Beta

Considerando $Y \sim \text{Beta}(\alpha, \beta)$, $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, a reparametrização resulta na nova distribuição $\text{Beta}(\mu\phi, (1 - \mu)\phi)$, em que μ é o parâmetro de média e ϕ o parâmetro de dispersão que serão utilizados para a modelagem de regressão.

Modelo de Regressão Beta

Considerando $Y \sim \text{Beta}(\alpha, \beta)$, $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, a reparametrização resulta na nova distribuição $\text{Beta}(\mu\phi, (1 - \mu)\phi)$, em que μ é o parâmetro de média e ϕ o parâmetro de dispersão que serão utilizados para a modelagem de regressão.

De acordo com a nova reparametrização, a densidade da variável aleatória Y é dada por:

Modelo de Regressão Beta

Considerando $Y \sim \text{Beta}(\alpha, \beta)$, $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, a reparametrização resulta na nova distribuição $\text{Beta}(\mu\phi, (1 - \mu)\phi)$, em que μ é o parâmetro de média e ϕ o parâmetro de dispersão que serão utilizados para a modelagem de regressão.

De acordo com a nova reparametrização, a densidade da variável aleatória Y é dada por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1} I_{(0,1)}(y),$$

Modelo de Regressão Beta

Considerando $Y \sim \text{Beta}(\alpha, \beta)$, $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, a reparametrização resulta na nova distribuição $\text{Beta}(\mu\phi, (1 - \mu)\phi)$, em que μ é o parâmetro de média e ϕ o parâmetro de dispersão que serão utilizados para a modelagem de regressão.

De acordo com a nova reparametrização, a densidade da variável aleatória Y é dada por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1} I_{(0,1)}(y),$$

em que $0 < \mu < 1$ e $\phi > 0$.

Modelo de Regressão Beta

Considerando $Y \sim \text{Beta}(\alpha, \beta)$, $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, a reparametrização resulta na nova distribuição $\text{Beta}(\mu\phi, (1 - \mu)\phi)$, em que μ é o parâmetro de média e ϕ o parâmetro de dispersão que serão utilizados para a modelagem de regressão.

De acordo com a nova reparametrização, a densidade da variável aleatória Y é dada por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1} I_{(0,1)}(y),$$

em que $0 < \mu < 1$ e $\phi > 0$.

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{V(\mu)}{1 + \phi},$$

Modelo de Regressão Beta

Considerando $Y \sim \text{Beta}(\alpha, \beta)$, $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, a reparametrização resulta na nova distribuição $\text{Beta}(\mu\phi, (1 - \mu)\phi)$, em que μ é o parâmetro de média e ϕ o parâmetro de dispersão que serão utilizados para a modelagem de regressão.

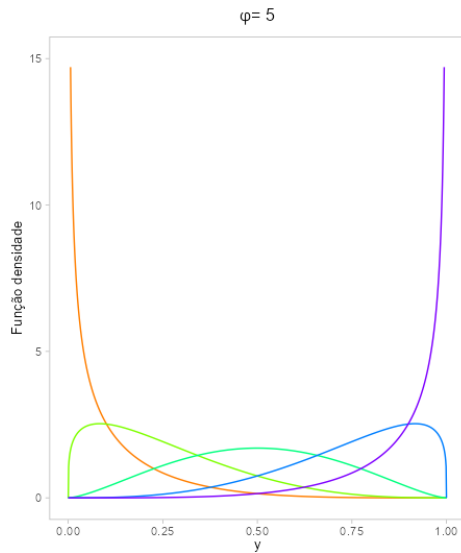
De acordo com a nova reparametrização, a densidade da variável aleatória Y é dada por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1} I_{(0,1)}(y),$$

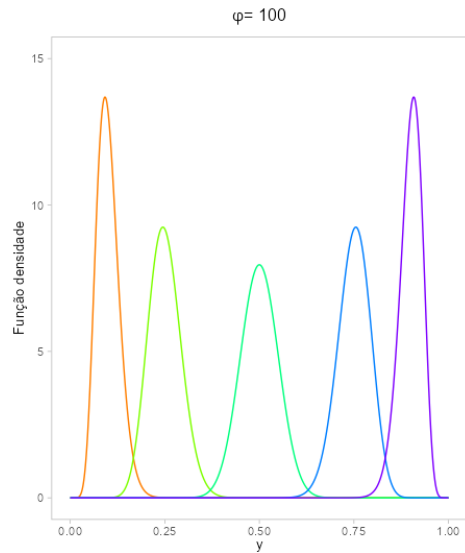
em que $0 < \mu < 1$ e $\phi > 0$.

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{V(\mu)}{1 + \phi},$$

em que $V(\mu) = \mu(1 - \mu)$.



μ — 0.1 — 0.25 — 0.5 — 0.75 — 0.9



Seja Y_1, \dots, Y_n variáveis aleatórias independentes, em que Y_i , $i = 1, \dots, n$, que seguem a distribuição Beta com média μ_i e parâmetro de precisão ϕ desconhecido. A estrutura do modelo de regressão é dado por:

Seja Y_1, \dots, Y_n variáveis aleatórias independentes, em que Y_i , $i = 1, \dots, n$, que seguem a distribuição Beta com média μ_i e parâmetro de precisão ϕ desconhecido. A estrutura do modelo de regressão é dado por:

$$g(\mu_i) = \sum_{t=1}^k x_{it}\beta_t = \eta_i,$$

Seja Y_1, \dots, Y_n variáveis aleatórias independentes, em que Y_i , $i = 1, \dots, n$, que seguem a distribuição Beta com média μ_i e parâmetro de precisão ϕ desconhecido. A estrutura do modelo de regressão é dado por:

$$g(\mu_i) = \sum_{t=1}^k x_{it}\beta_t = \eta_i,$$

em que $\beta = (\beta_1, \dots, \beta_k)^T$ é o vetor dos parâmetros da regressão e x_{i1}, \dots, x_{ik} são as observações das k covariáveis ($k < n$), que assumimos ser fixos e conhecidos.

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Modelo de Regressão Beta

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Existem outras possíveis funções de ligação, como por exemplo:

Modelo de Regressão Beta

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Existem outras possíveis funções de ligação, como por exemplo:

- $g(\mu) = -\log[-\log(\mu)]$ (log-log);

Modelo de Regressão Beta

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Existem outras possíveis funções de ligação, como por exemplo:

- $g(\mu) = -\log[-\log(\mu)]$ (log-log);
- $g(\mu) = \log[-\log(1 - \mu)]$ (complemento log-log);

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Existem outras possíveis funções de ligação, como por exemplo:

- $g(\mu) = -\log[-\log(\mu)]$ (log-log);
- $g(\mu) = \log[-\log(1 - \mu)]$ (complemento log-log);
- $g(\mu) = \Phi^{-1}(\mu)$ (probit).

Modelo de Regressão Beta

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Existem outras possíveis funções de ligação, como por exemplo:

- $g(\mu) = -\log[-\log(\mu)]$ (log-log);
- $g(\mu) = \log[-\log(1 - \mu)]$ (complemento log-log);
- $g(\mu) = \Phi^{-1}(\mu)$ (probit).

Uma função de ligação particularmente útil é a função logito $g(\mu) = \log[\mu(1 - \mu)]$, em que é possível escrever μ_i da seguinte forma:

Modelo de Regressão Beta

A função de ligação $g(\cdot)$ que mapeia $(0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e diferenciável até a segunda ordem.

Existem outras possíveis funções de ligação, como por exemplo:

- $g(\mu) = -\log[-\log(\mu)]$ (log-log);
- $g(\mu) = \log[-\log(1 - \mu)]$ (complemento log-log);
- $g(\mu) = \Phi^{-1}(\mu)$ (probit).

Uma função de ligação particularmente útil é a função logito $g(\mu) = \log[\mu(1 - \mu)]$, em que é possível escrever μ_i da seguinte forma:

$$\mu_k = \frac{e^{x_k^T \beta_k}}{1 + e^{x_k^T \beta_k}}.$$

Dessa forma, a interpretação dos parâmetros da regressão pode ser entendida como, ao aumentar c unidades na k -ésima covariável sem que haja alterações nas demais covariáveis, e seja μ_τ a média de Y restrito as novas covariáveis, denote μ a média de Y com as covariáveis originais. É trivial, notar que a razão de chances nessa situação é dada por:

Dessa forma, a interpretação dos parâmetros da regressão pode ser entendida como, ao aumentar c unidades na k -ésima covariável sem que haja alterações nas demais covariáveis, e seja μ_τ a média de Y restrito as novas covariáveis, denote μ a média de Y com as covariáveis originais. É trivial, notar que a razão de chances nessa situação é dada por:

$$e^{c\beta_k} = \frac{\mu_\tau(1 - \mu_\tau)}{\mu(1 - \mu)}.$$

A função de log-verossimilhança é dada pela seguinte expressão:

A função de log-verossimilhança é dada pela seguinte expressão:

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

A função de log-verossimilhança é dada pela seguinte expressão:

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

em que,

A função de log-verossimilhança é dada pela seguinte expressão:

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

em que,

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

Modelo de Regressão Beta

Seja $y_i^* = \log[y_i/(1 - y_i)]$ e $\mu_i^* = \psi(\mu_i\phi) - \psi[(1 - \mu_i)\phi]$. Então a função escore pode ser obtida pela derivação da função de log-verossimilhança com respeito à parâmetros desconhecidos e são dadas por, respectivamente:

Modelo de Regressão Beta

Seja $y_i^* = \log[y_i/(1 - y_i)]$ e $\mu_i^* = \psi(\mu_i\phi) - \psi[(1 - \mu_i)\phi]$. Então a função escore pode ser obtida pela derivação da função de log-verossimilhança com respeito à parâmetros desconhecidos e são dadas por, respectivamente:

$$U_{\beta}(\beta, \phi) = \phi X^T T(y^* - \mu^*),$$

Modelo de Regressão Beta

Seja $y_i^* = \log[y_i/(1 - y_i)]$ e $\mu_i^* = \psi(\mu_i\phi) - \psi[(1 - \mu_i)\phi]$. Então a função escore pode ser obtida pela derivação da função de log-verossimilhança com respeito à parâmetros desconhecidos e são dadas por, respectivamente:

$$U_\beta(\beta, \phi) = \phi X^T T(y^* - \mu^*),$$

$$U_\phi(\beta, \phi) = \sum_{i=1}^n \{\mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi)\}.$$

Modelo de Regressão Beta

Seja $y_i^* = \log[y_i/(1 - y_i)]$ e $\mu_i^* = \psi(\mu_i\phi) - \psi[(1 - \mu_i)\phi]$. Então a função escore pode ser obtida pela derivação da função de log-verossimilhança com respeito à parâmetros desconhecidos e são dadas por, respectivamente:

$$U_{\beta}(\beta, \phi) = \phi X^T T(y^* - \mu^*),$$

$$U_{\phi}(\beta, \phi) = \sum_{i=1}^n \{\mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi)\}.$$

Sendo X uma matriz $n \times k$, em que a i -ésima linha é x_i^T , $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$ e y^* e μ^* são vetores de y_i^* e μ_i^* , definidos anteriormente

Para obter a expressão da informação de Fisher, vamos considerar $W = \text{diag}(w_1, \dots, w_n)$, com

Para obter a expressão da informação de Fisher, vamos considerar $W = \text{diag}(w_1, \dots, w_n)$, com

$$w_i = \frac{c_i}{(g' \mu_i)^2}.$$

Para obter a expressão da informação de Fisher, vamos considerar $W = \text{diag}(w_1, \dots, w_n)$, com

$$w_i = \frac{c_i}{(g' \mu_i)^2}.$$

em que $c_i = \phi\{\psi'(\mu_i \phi_i) + \psi'[(1 - \mu_i)\phi]\}$ e $c = (c_1, \dots, c_n)^T$, $\psi'(\cdot)$ é a função trigama.

Para obter a expressão da informação de Fisher, vamos considerar $W = \text{diag}(w_1, \dots, w_n)$, com

$$w_i = \frac{c_i}{(g' \mu_i)^2}.$$

em que $c_i = \phi\{\psi'(\mu_i \phi) + \psi'[(1 - \mu_i)\phi]\}$ e $c = (c_1, \dots, c_n)^T$, $\psi'(\cdot)$ é a função trigama. Seja $D = \text{diag}(d_1, \dots, d_n)$, com $d_i = \psi'(\mu_i \phi)\mu_i^2 + \psi'[(1 - \mu_i)\phi](1 - \mu_i^2) - \psi'(\phi)$. Então, a matriz de informação de Fisher é dada por:

Para obter a expressão da informação de Fisher, vamos considerar $W = \text{diag}(w_1, \dots, w_n)$, com

$$w_i = \frac{c_i}{(g' \mu_i)^2}.$$

em que $c_i = \phi\{\psi'(\mu_i \phi_i) + \psi'[(1 - \mu_i)\phi]\}$ e $c = (c_1, \dots, c_n)^T$, $\psi'(\cdot)$ é a função trigama. Seja $D = \text{diag}(d_1, \dots, d_n)$, com $d_i = \psi'(\mu_i \phi) \mu_i^2 + \psi'[(1 - \mu_i)\phi](1 - \mu_i^2) - \psi'(\phi)$. Então, a matriz de informação de Fisher é dada por:

$$\mathbb{K} = \mathbb{K}(\beta, \phi) = \begin{pmatrix} \mathbb{K}_{\beta\beta} & \mathbb{K}_{\beta\phi} \\ \mathbb{K}_{\phi\beta} & \mathbb{K}_{\phi\phi} \end{pmatrix}$$

Para obter a expressão da informação de Fisher, vamos considerar $W = \text{diag}(w_1, \dots, w_n)$, com

$$w_i = \frac{c_i}{(g' \mu_i)^2}.$$

em que $c_i = \phi\{\psi'(\mu_i \phi_i) + \psi'[(1 - \mu_i)\phi]\}$ e $c = (c_1, \dots, c_n)^T$, $\psi'(\cdot)$ é a função trigama. Seja $D = \text{diag}(d_1, \dots, d_n)$, com $d_i = \psi'(\mu_i \phi)\mu_i^2 + \psi'[(1 - \mu_i)\phi](1 - \mu_i^2) - \psi'(\phi)$. Então, a matriz de informação de Fisher é dada por:

$$\mathbb{K} = \mathbb{K}(\beta, \phi) = \begin{pmatrix} \mathbb{K}_{\beta\beta} & \mathbb{K}_{\beta\phi} \\ \mathbb{K}_{\phi\beta} & \mathbb{K}_{\phi\phi} \end{pmatrix}$$

Sendo $\mathbb{K}_{\beta\beta} = \phi X^T W X$, $\mathbb{K}_{\beta\phi} = \mathbb{K}_{\phi\beta}^T = X^T T c$ e $\mathbb{K}_{\phi\phi} = \text{Tr}(D)$.

Os estimadores de máxima verossimilhança $\hat{\beta}$ e $\hat{\phi}$ para β e ϕ são obtidos quando $U_{\beta}(\beta, \phi) = 0$ e $U_{\phi}(\beta, \phi) = 0$, respectivamente. Esses estimadores não possuem forma fechada e devem ser encontrados por meio de algoritmos numéricos iterativos, como por exemplo o Newton–Raphson ou Quasi-Newton.

Os estimadores de máxima verossimilhança $\hat{\beta}$ e $\hat{\phi}$ para β e ϕ são obtidos quando $U_{\beta}(\beta, \phi) = 0$ e $U_{\phi}(\beta, \phi) = 0$, respectivamente. Esses estimadores não possuem forma fechada e devem ser encontrados por meio de algoritmos numéricos iterativos, como por exemplo o Newton–Raphson ou Quasi-Newton.

No processo de otimização destes algoritmos é necessário um valor inicial para a estrutura iterativa, a literatura sugere o uso da estimativa por mínimos quadrados ordinários para β obtido por uma regressão linear de variáveis transformadas $g(Y_1), \dots, g(Y_n)$ como ponto inicial.

O valor inicial de ϕ sugerido no artigo é dado por:

O valor inicial de ϕ sugerido no artigo é dado por:

$$\frac{1}{n} \sum_{i=1}^n \frac{\mu_i^{**}(1 - \mu_i^{**})}{\sigma_i^{**2}} - 1,$$

O valor inicial de ϕ sugerido no artigo é dado por:

$$\frac{1}{n} \sum_{i=1}^n \frac{\mu_i^{**}(1 - \mu_i^{**})}{\sigma_i^{**2}} - 1,$$

em que μ_i^{**} é obtido aplicando $g^{-1}(\cdot)$ na i -ésima predição da regressão linear das variáveis transformadas $g(Y_1), \dots, g(Y_n)$ em X , $\mu_i^{**} = g^{-1}(x_i^T (X^T X)^{-1} X^T z)$ e $\sigma_i^{**} = e^{**T} e^{**} / [(n - k) g'(\mu^{**})^2]$, em que $e^{**} = z - X(X^T X)^{-1} X^T z$ é o vetor dos resíduos da regressão por mínimos quadrados ordinários das variáveis resposta transformadas.

Análise de diagnóstico

Após ajustar o modelo de regressão Beta, é importante realizar uma análise de diagnóstico para checar o ajuste. Uma medida de explicação da variabilidade pode ser obtido calculando o pseudo- $R^2(R_p^2)$ definido como $\text{Cor}(\hat{\eta}, g(y))^2$. A medida (R_p^2) varia entre 0 e 1 e a concordância perfeita ocorre quando $(R_p^2) = 1$. A discrepância do ajuste $(D(y, \mu, \phi))$ pode ser medida utilizando a seguinte expressão:

Após ajustar o modelo de regressão Beta, é importante realizar uma análise de diagnóstico para checar o ajuste. Uma medida de explicação da variabilidade pode ser obtido calculando o pseudo- $R^2(R_p^2)$ definido como $\text{Cor}(\hat{\eta}, g(y))^2$. A medida (R_p^2) varia entre 0 e 1 e a concordância perfeita ocorre quando $(R_p^2) = 1$. A discrepância do ajuste $(D(y, \mu, \phi))$ pode ser medida utilizando a seguinte expressão:

$$D(y, \mu, \phi) = \sum_{i=1}^n (r_i^d)^2,$$

Após ajustar o modelo de regressão Beta, é importante realizar uma análise de diagnóstico para checar o ajuste. Uma medida de explicação da variabilidade pode ser obtido calculando o pseudo- $R^2(R_p^2)$ definido como $\text{Cor}(\hat{\eta}, g(y))^2$. A medida (R_p^2) varia entre 0 e 1 e a concordância perfeita ocorre quando $(R_p^2) = 1$. A discrepância do ajuste ($D(y, \mu, \phi)$) pode ser medida utilizando a seguinte expressão:

$$D(y, \mu, \phi) = \sum_{i=1}^n (r_i^d)^2,$$

em que $r_i^d = \text{sign}(y_i - \hat{\mu}_i)[2\ell_i(\tilde{\mu}_i, \hat{\phi}_i)\ell_i(\hat{\mu}_i, \hat{\phi}_i)]^{1/2}$. Note que a i -ésima observação contribui para $(r_i^d)^2$ no desvio e assim observações com valores nominais altos podem ser vistos como discrepantes.

Os resíduos padrão são definidos da seguinte forma:

Os resíduos padrão são definidos da seguinte forma:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\text{var}}(y_i)}}$$

Os resíduos padrão são definidos da seguinte forma:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\text{Var}}(y_i)}}$$

sendo $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$ e $\hat{\text{Var}}(y_i) = [\hat{\mu}_i(1 - \hat{\mu}_i)]/(1 + \hat{\phi})$. Um gráfico desses resíduos em relação ao indexador das observações não deve mostrar padrões, caso isso ocorra, deve-se verificar se a função de ligação utilizada é a mais adequada.

Os resíduos padrão são definidos da seguinte forma:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\text{Var}}(y_i)}}$$

sendo $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$ e $\hat{\text{Var}}(y_i) = [\hat{\mu}_i(1 - \hat{\mu}_i)]/(1 + \hat{\phi})$. Um gráfico desses resíduos em relação ao indexador das observações não deve mostrar padrões, caso isso ocorra, deve-se verificar se a função de ligação utilizada é a mais adequada.

Como a distribuição dos resíduos é assumida desconhecida, um gráfico de envelope semi-normais simulados é uma ferramenta de diagnóstico bastante útil.

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

- i) Ajuste o modelo e gere uma amostra de observações pseudo-aleatória simulada de tamanho n independente usando os valores preditos como o modelo real;

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

- i) Ajuste o modelo e gere uma amostra de observações pseudo-aleatória simulada de tamanho n independente usando os valores preditos como o modelo real;
- ii) Ajuste o modelo para os dados gerados e compute os valores absolutos dos resíduos ordenados;

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

- i) Ajuste o modelo e gere uma amostra de observações pseudo-aleatória simulada de tamanho n independente usando os valores preditos como o modelo real;
- ii) Ajuste o modelo para os dados gerados e compute os valores absolutos dos resíduos ordenados;
- iii) Repita i) e ii) k vezes;

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

- i) Ajuste o modelo e gere uma amostra de observações pseudo-aleatória simulada de tamanho n independente usando os valores preditos como o modelo real;
- ii) Ajuste o modelo para os dados gerados e compute os valores absolutos dos resíduos ordenados;
- iii) Repita i) e ii) k vezes;
- iv) Compute a média, mínimo e máximo nos n conjuntos de k estatísticas de ordem;

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

- i) Ajuste o modelo e gere uma amostra de observações pseudo-aleatória simulada de tamanho n independente usando os valores preditos como o modelo real;
- ii) Ajuste o modelo para os dados gerados e compute os valores absolutos dos resíduos ordenados;
- iii) Repita i) e ii) k vezes;
- iv) Compute a média, mínimo e máximo nos n conjuntos de k estatísticas de ordem;
- v) Faça o gráfico desses valores e dos resíduos ordenados da amostra original em relação aos escores semi-normais $\Phi^{-1}[(i + n - 1/8)/(2n + 1/2)]$.

O gráfico de envelope pode ser produzido utilizando o seguinte algoritmo:

- i) Ajuste o modelo e gere uma amostra de observações pseudo-aleatória simulada de tamanho n independente usando os valores preditos como o modelo real;
- ii) Ajuste o modelo para os dados gerados e compute os valores absolutos dos resíduos ordenados;
- iii) Repita i) e ii) k vezes;
- iv) Compute a média, mínimo e máximo nos n conjuntos de k estatísticas de ordem;
- v) Faça o gráfico desses valores e dos resíduos ordenados da amostra original em relação aos escores semi-normais $\Phi^{-1}[(i + n - 1/8)/(2n + 1/2)]$.

Os valores mínimos e máximos das k estatísticas de ordem criam os envelopes.

Para verificar observações influentes podemos usar o método de *generalized leverage* proposto por Wei *et. al* (1998), pode ser generalizado pela seguinte expressão avaliada no estimador de máxima verossimilhança de um parâmetro desconhecido θ :

Para verificar observações influentes podemos usar o método de *generalized leverage* proposto por Wei *et. al* (1998), pode ser generalizado pela seguinte expressão avaliada no estimador de máxima verossimilhança de um parâmetro desconhecido θ :

$$GL(\theta) = D_{\theta} \left(-\frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \right)^{-1} \frac{\partial^2 \ell}{\partial \theta \partial \theta^T}$$

Para verificar observações influentes podemos usar o método de *generalized leverage* proposto por Wei *et. al* (1998), pode ser generalizado pela seguinte expressão avaliada no estimador de máxima verossimilhança de um parâmetro desconhecido θ :

$$GL(\theta) = D_{\theta} \left(-\frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \right)^{-1} \frac{\partial^2 \ell}{\partial \theta \partial \theta^T}$$

em que ℓ é a função da log-verossimilhança de um parâmetro desconhecido θ e $D_{\theta} = \partial \mu / \partial \theta^T$.

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

$$GL(\beta) = TX(X^T QX)^{-1}X^T TM$$

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

$$GL(\beta) = TX(X^T QX)^{-1}X^T TM$$

sendo $M = \text{diag}\{m_1, \dots, m_n\}$ com $m_i = 1/[y_i(1 - y_i)]$ e $Q = \text{diag}\{q_1, \dots, q_n\}$ com

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

$$GL(\beta) = TX(X^T QX)^{-1}X^T TM$$

sendo $M = \text{diag}\{m_1, \dots, m_n\}$ com $m_i = 1/[y_i(1 - y_i)]$ e $Q = \text{diag}\{q_1, \dots, q_n\}$ com

$$q_i = \left\{ \phi[\psi'(\mu_i\phi) + \psi'((1 - \mu_i)\phi)] + (y_i^* - \mu_i^*) \frac{g''(\mu_i)}{g'(\mu_i)} \right\} \frac{1}{g'(\mu_i)^2}.$$

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

$$GL(\beta) = TX(X^T QX)^{-1}X^T TM$$

sendo $M = \text{diag}\{m_1, \dots, m_n\}$ com $m_i = 1/[y_i(1 - y_i)]$ e $Q = \text{diag}\{q_1, \dots, q_n\}$ com

$$q_i = \left\{ \phi[\psi'(\mu_i\phi) + \psi'((1 - \mu_i)\phi)] + (y_i^* - \mu_i^*) \frac{g''(\mu_i)}{g'(\mu_i)} \right\} \frac{1}{g'(\mu_i)^2}.$$

Assim, podemos obter a seguinte expressão:

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

$$GL(\beta) = TX(X^T QX)^{-1} X^T T M$$

sendo $M = \text{diag}\{m_1, \dots, m_n\}$ com $m_i = 1/[y_i(1 - y_i)]$ e $Q = \text{diag}\{q_1, \dots, q_n\}$ com

$$q_i = \left\{ \phi[\psi'(\mu_i \phi) + \psi'((1 - \mu_i) \phi)] + (y_i^* - \mu_i^*) \frac{g''(\mu_i)}{g'(\mu_i)} \right\} \frac{1}{g'(\mu_i)^2}.$$

Assim, podemos obter a seguinte expressão:

$$GL(\beta, \phi) = GL(\beta) + \frac{1}{\gamma \phi} TX(X^T QX)^{-1} X^T T f(f^T TX(X^T QX)^{-1} X^T T M - b^T)$$

Para o caso da regressão Beta, $GL(\beta, \phi)$ é dado por:

$$GL(\beta) = TX(X^T QX)^{-1} X^T T M$$

sendo $M = \text{diag}\{m_1, \dots, m_n\}$ com $m_i = 1/[y_i(1 - y_i)]$ e $Q = \text{diag}\{q_1, \dots, q_n\}$ com

$$q_i = \left\{ \phi[\psi'(\mu_i \phi) + \psi'((1 - \mu_i)\phi)] + (y_i^* - \mu_i^*) \frac{g''(\mu_i)}{g'(\mu_i)} \right\} \frac{1}{g'(\mu_i)^2}.$$

Assim, podemos obter a seguinte expressão:

$$GL(\beta, \phi) = GL(\beta) + \frac{1}{\gamma \phi} TX(X^T QX)^{-1} X^T T f(f^T TX(X^T QX)^{-1} X^T T M - b^T)$$

em que $b = (b_1, \dots, b_n)^T$ com $b_i = -(y_i - \mu_i)/[y_i(1 - y_i)]$. Quando ϕ é suficientemente grande, $GL(\beta, \phi) \approx GL(\beta)$.

Uma medida de influência de cada observação nas estimativas dos parâmetros da regressão é a distância de Cook dada por $k^{-1}(\hat{\beta} - \hat{\beta}_{(i)})^T X^T W X (\hat{\beta} - \hat{\beta}_{(i)})$ em que $\hat{\beta}_{(i)}$ é a estimativa do parâmetro sem a i -ésima observação. Para evitar ter que ajustar o modelo $n + 1$ vezes, é recomendado utilizar uma aproximação da distância de Cook, dada por

Uma medida de influência de cada observação nas estimativas dos parâmetros da regressão é a distância de Cook dada por $k^{-1}(\hat{\beta} - \hat{\beta}_{(i)})^T X^T W X (\hat{\beta} - \hat{\beta}_{(i)})$ em que $\hat{\beta}_{(i)}$ é a estimativa do parâmetro sem a i -ésima observação. Para evitar ter que ajustar o modelo $n + 1$ vezes, é recomendado utilizar uma aproximação da distância de Cook, dada por

$$C_i = \frac{h_{ii} r_i^2}{k(1 - h_{ii})^2}$$

Uma medida de influência de cada observação nas estimativas dos parâmetros da regressão é a distância de Cook dada por $k^{-1}(\hat{\beta} - \hat{\beta}_{(i)})^T X^T W X (\hat{\beta} - \hat{\beta}_{(i)})$ em que $\hat{\beta}_{(i)}$ é a estimativa do parâmetro sem a i -ésima observação. Para evitar ter que ajustar o modelo $n + 1$ vezes, é recomendado utilizar uma aproximação da distância de Cook, dada por

$$C_i = \frac{h_{ii} r_i^2}{k(1 - h_{ii})^2}$$

em que h_{ii} é um elemento da matriz $H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$.

Aplicação

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

- **yield:** Proporção de óleo não-refinado convertido para gasolina depois de destilação e fracionamento;

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

- **yield:** Proporção de óleo não-refinado convertido para gasolina depois de destilação e fracionamento;
- **gravity:** Quão pesado é óleo não-refinado em comparação com água (Gravidade API);

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

- **yield:** Proporção de óleo não-refinado convertido para gasolina depois de destilação e fracionamento;
- **gravity:** Quão pesado é óleo não-refinado em comparação com água (Gravidade API);
- **pressure:** Pressão de vapor do óleo não-refinado (lbs/in^2);

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

- **yield**: Proporção de óleo não-refinado convertido para gasolina depois de destilação e fracionamento;
- **gravity**: Quão pesado é óleo não-refinado em comparação com água (Gravidade API);
- **pressure**: Pressão de vapor do óleo não-refinado (lbs/in^2);
- **temp10**: Temperatura em que 10% do óleo foi vaporizado ($^{\circ}\text{F}$);

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

- **yield**: Proporção de óleo não-refinado convertido para gasolina depois de destilação e fracionamento;
- **gravity**: Quão pesado é óleo não-refinado em comparação com água (Gravidade API);
- **pressure**: Pressão de vapor do óleo não-refinado (lbs/in^2);
- **temp10**: Temperatura em que 10% do óleo foi vaporizado ($^{\circ}\text{F}$);
- **temp**: Temperatura em que todo o óleo foi vaporizado ($^{\circ}\text{F}$);

O conjunto de dados possui 32 observações (sem dados faltantes) e foi coletado por Prater (1956), contém as seguintes variáveis:

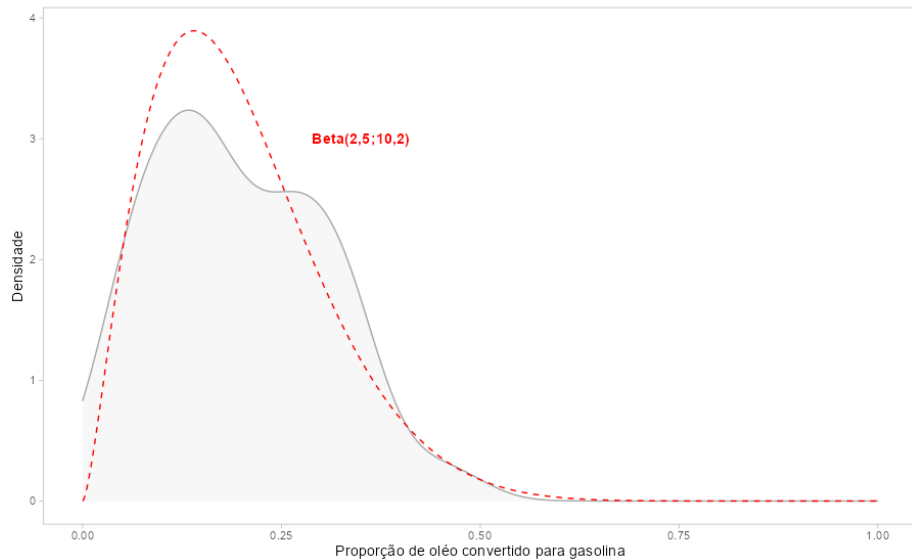
- **yield**: Proporção de óleo não-refinado convertido para gasolina depois de destilação e fracionamento;
- **gravity**: Quão pesado é óleo não-refinado em comparação com água (Gravidade API);
- **pressure**: Pressão de vapor do óleo não-refinado (lbs/in^2);
- **temp10**: Temperatura em que 10% do óleo foi vaporizado ($^{\circ}\text{F}$);
- **temp**: Temperatura em que todo o óleo foi vaporizado ($^{\circ}\text{F}$);
- **batch**: Fator indicando a condição dos lotes.

A variável resposta na análise será **yield** (numérica entre 0 e 1) e as covariáveis serão **temp** (numérica 0 à infinito) e **batch** (categórica com 10 níveis). O conjunto de dados foi analisado por Atkinson (1985), no trabalho foi utilizado um modelo de regressão linear e foi percebido que a distribuição do erro não era simétrica, gerando resíduos discrepantes.

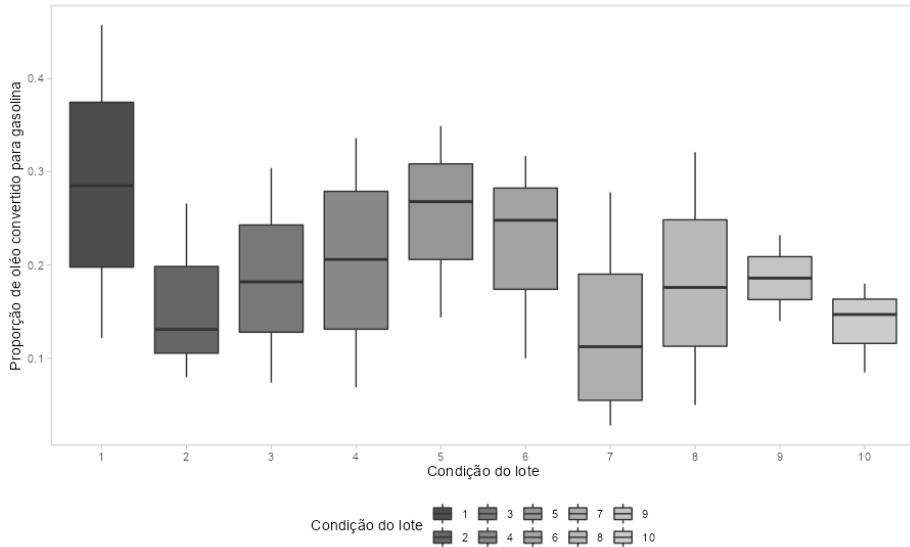
A variável resposta na análise será **yield** (numérica entre 0 e 1) e as covariáveis serão **temp** (numérica 0 à infinito) e **batch** (categórica com 10 níveis). O conjunto de dados foi analisado por Atkinson (1985), no trabalho foi utilizado um modelo de regressão linear e foi percebido que a distribuição do erro não era simétrica, gerando resíduos discrepantes.

Assim, ele aplicou uma transformação na variável resposta de forma que **yield*** pertença a reta Real e, então, aplicou o modelo de regressão linear na variável transformada. Para o conjunto de dados proposto, Ferrari e Cribari-Neto (2004), desenvolveram a regressão Beta e a utilizaram para modelar os dados.

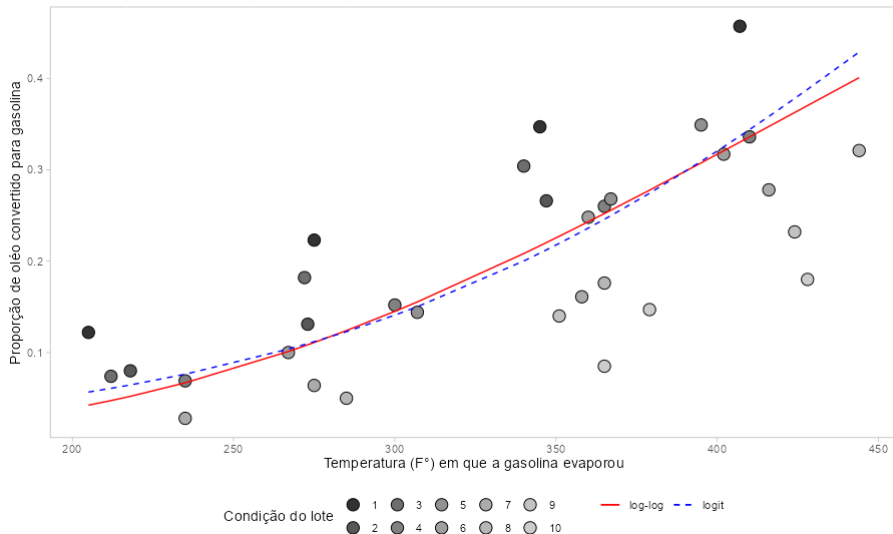
Densidade empírica da variável resposta



Boxplot da proporção de óleo convertido para gasolina por condição de lote



Comparação entre as funções de ligação logito e log-log no ajuste



O valor do pseudo- R^2 foi igual à 0.9617 em 51 iterações (BFGS com algoritmo quasi-Newton).

	Estimativa	Erro Padrão	z value	Pr(> z)	Signif.
(Intercepto)	-6.16	0.18	-33.78	<2e-16	***
temp	0.01	0	26.58	<2e-16	***
batch1	1.73	0.1	17.07	<2e-16	***
batch2	1.32	0.12	11.22	<2e-16	***
batch3	1.57	0.12	13.54	<2e-16	***
batch4	1.06	0.1	10.35	<2e-16	***
batch5	1.13	0.1	10.95	<2e-16	***
batch6	1.04	0.11	9.81	<2e-16	***
batch7	0.54	0.11	4.98	6.29e-07	***
batch8	0.5	0.11	4.55	5.3e-06	***
batch9	0.39	0.12	3.25	0.00114	**

É notável que existe ao menos um ponto que aparenta ser discrepante, a observação 4, nos gráficos da distância de Cook, Resíduos por Índice, Alavanca generalizada, Valores Preditos por Observáveis e Envelope é o que possui maior destaque. A observação 29 também possui destaque no gráfico de alavanca, mas ao investigar esse ponto não apresenta uma diferença muito larga em relação aos demais pontos, com exceção do 4. (**Ver R**)

É notável que existe ao menos um ponto que aparenta ser discrepante, a observação 4, nos gráficos da distância de Cook, Resíduos por Índice, Alavanca generalizada, Valores Preditos por Observáveis e Envelope é o que possui maior destaque. A observação 29 também possui destaque no gráfico de alavanca, mas ao investigar esse ponto não apresenta uma diferença muito larga em relação aos demais pontos, com exceção do 4. (**Ver R**)

Portanto, foi aplicado uma segunda modelagem *post-hoc* sem a quarta observação, dada que ela foi julgada como influente nos dados.

É notável que existe ao menos um ponto que aparenta ser discrepante, a observação 4, nos gráficos da distância de Cook, Resíduos por Índice, Alavanca generalizada, Valores Preditos por Observáveis e Envelope é o que possui maior destaque. A observação 29 também possui destaque no gráfico de alavanca, mas ao investigar esse ponto não apresenta uma diferença muito larga em relação aos demais pontos, com exceção do 4. (**Ver R**)

Portanto, foi aplicado uma segunda modelagem *post-hoc* sem a quarta observação, dada que ela foi julgada como influente nos dados.

As estimativas pontuais entre os dois modelos não foram muito divergentes, mas o parâmetro de precisão ϕ aumentou de 440.3 para 577.8. Porém, a diferença entre os erros padrão assintóticos dos dois modelos foram negligenciáveis.

	Estimativa	Erro Padrão	z value	$\Pr(> z)$	Signif.
(Intercepto)	-6.36	0.17	-37.04	$<2e-16$	***
temp	0.01	0	29.05	$<2e-16$	***
batch1	1.89	0.1	18.83	$<2e-16$	***
batch2	1.37	0.1	13.15	$<2e-16$	***
batch3	1.63	0.1	15.8	$<2e-16$	***
batch4	1.08	0.09	12.04	$<2e-16$	***
batch5	1.15	0.09	12.7	$<2e-16$	***
batch6	1.06	0.09	11.38	$<2e-16$	***
batch7	0.57	0.1	5.91	$3.39e-09$	***
batch8	0.5	0.1	5.25	$5.25e-07$	***
batch9	0.39	0.1	3.71	0.0002	***

Considerações Finais

As análises foram replicadas utilizando o *Software RStudio* (Versão 4.1.1) com auxílio do pacote *tidyverse* para fins de transformação de variáveis e gráficos e o pacote *betareg* para realizar a regressão Beta. Para visualizar os códigos realizados acesse o seguinte repositório do GitHub:

As análises foram replicadas utilizando o *Software RStudio* (Versão 4.1.1) com auxílio do pacote *tidyverse* para fins de transformação de variáveis e gráficos e o pacote *betareg* para realizar a regressão Beta. Para visualizar os códigos realizados acesse o seguinte repositório do GitHub:

https://github.com/wesleyacruzzz/trabalho_regressao_mestrado.

- Bury, K. (1999) Statistical Distributions in Engineering (New York: Cambridge University Press).
- Cribari-Neto F, Zeileis A (2010). “Beta Regression in R.” Journal of Statistical Software, 34(2), 1–24. URL <http://www.jstatsoft.org/v34/i02/>.
- Ferrari SLP, Cribari-Neto F (2004). “Beta Regression for Modelling Rates and Proportions.” Journal of Applied Statistics, 31(7), 799–815.
- Johnson, N. L., Kotz, S. Balakrishnan, N. (1995) Continuous Univariate Distributions, vol. 2, 2nd edn (New York: Wiley).
- Jørgesen, B. (1997) Proper dispersion models (with discussion), Brazilian Journal of Probability and Statistics, 11, pp. 89–140.

- McCullagh, P. Nelder, J. A. (1989) Generalized Linear Models, 2nd edn (London: Chapman and Hall).
- Nocedal, J. Wright, S. J. (1999) Numerical Optimization (New York: Springer-Verlag).
- RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., *Boston*, MA URL. Disponível em: <http://www.rstudio.com/>.
- Wei, B.-C., Hu, Y.-Q. Fung, W.-K. (1998) Generalized leverage and its applications, *Scandinavian Journal of Statistics*, 25, pp. 25–37.