Universidade de Brasília

Modelos de regressão Birnbaum-Saunders parametrizados pela mediana, média, moda e quantis: Uma aplicação comparativa em dados de germinação de sementes

Tailine J. S. Nonato

Programa de Pós-Graduação em Estatística
Universidade de Brasília

Novembro, 2025

Sumário

1. Introdução

Contexto

- Modelos Birnbaum–Saunders têm se tornado uma alternativa robusta para analisar tempos de vida e variáveis positivas assimétricas.
- No entanto, diferentes parametrizações do modelo coexistem na literatura, com implicações distintas para a interpretação e ajuste.
- Entender como essas variações se comportam na prática é essencial para o uso adequado do modelo.

Literatura

- Embora comparações entre parametrizações do modelo BS já tenham sido conduzidas, ainda há poucas aplicações em dados de agricultura, o que pode revelar comportamentos diferentes devido à natureza da variabilidade experimental.

Objetivos

- Comparar o desempenho de diferentes parametrizações do modelo de regressão Birnbaum–Saunders;
- Avaliar diferenças em termos de ajuste, precisão e interpretabilidade dos parâmetros;
- Atualização dos códigos em R para facilitar a interpretação.

2. Referencial teórico

Distribuição Birnbaum-Saunders (BS)

Considere T uma variável aleatória definida por

$$T = \frac{\beta}{4}\left(\alpha Z + \sqrt{\alpha^2 Z^2 + 4}\right)^2,$$

em que $Z \sim N(0,1)$, $\alpha > 0$ e $\beta > 0$.
Assim, T segue uma distribuição BS, denotada por $T \sim BS(\alpha, \beta)$.
A função densidade de probabilidade da distribuição BS pode ser expressa como

$$f_T(t) = \frac{1}{2\alpha\beta\sqrt{2\pi}}\left[\sqrt{\frac{\beta}{t}} + \sqrt{\left(\frac{\beta}{t}\right)^3}\right]\exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right], \qquad t > 0,$$

em que $\alpha > 0$ e $\beta > 0$ representam, respectivamente, os parâmetros de forma e escala. Nesta parametrização, o parâmetro β corresponde à **mediana** da distribuição, ou seja, $P(T \leq \beta) = 0{,}5$.

Modelo de regressão (mediana)

Considerando T_i como a variável resposta, o modelo de regressão BS pode ser escrito como

$$T_i = \beta_i \varphi_i, \qquad i = 1, 2, \ldots, n,$$

em que β_i denota a mediana associada à observação i, e φ_i é uma variável aleatória com distribuição BS padrão, isto é, $\varphi_i \sim BS(1, \alpha)$, tal que $E(\varphi_i) \neq 1$, mas $med(\varphi_i) = 1$.

A estrutura de regressão é usualmente definida em termos da mediana condicional,

$$\beta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\theta}),$$

onde $\mathbf{x}_i$ é o vetor de covariáveis associado à observação i, e $\boldsymbol{\theta}$ é o vetor de parâmetros de regressão.

O modelo permite que a mediana da distribuição de T_i varie de acordo com as covariáveis, mantendo o parâmetro de forma α constante para todas as observações [1] [2].

Parametrização pela média

A média desta distribuição é

$$\mathbb{E}(T) = \beta\left(1 + \frac{\alpha^2}{2}\right).$$

Definindo $c = 1 + \frac{\alpha^2}{2}$, a reparametrização em termos da média $\mu = \mathbb{E}(T)$ é dada por

$$\beta = \frac{\mu}{c}.$$

Substituindo $\beta = \frac{\mu}{c}$ na expressão da densidade, obtemos a função densidade de T em termos de (α, μ):

$$f_T(t; \alpha, \mu) = \frac{c}{2\alpha\mu\sqrt{2\pi}}\left[\sqrt{\frac{\mu}{c\,t}} + \left(\frac{\mu}{c\,t}\right)^{\frac{3}{2}}\right]\exp\left\{-\frac{1}{2\alpha^2}\left(\frac{c\,t}{\mu} + \frac{\mu}{c\,t} - 2\right)\right\}, t > 0.$$

Modelo de regressão (média)

Para o modelo de regressão, parametrizamos a média condicional μ_i em função das covariáveis $\mathbf{x}_i$ por meio do vínculo logarítmico (garantindo positividade):

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\theta}, \qquad i = 1, \ldots, n,$$

ou equivalentemente

$$\mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\theta})$$

.

Assim, a densidade condicional de T_i é $f_{T_i}(t_i; \alpha, \mu_i) = f_T(t_i; \alpha, \mu_i)$, com α tipicamente assumido constante entre observações.

Parametrização pela moda

Seguindo Bourguignon e Gallardo, consideramos a reparametrização da distribuição BS em termos da *moda* $\tilde{\mu}$ e do parâmetro de dispersão ϕ, onde

$$\phi = \alpha^2, \qquad \alpha = \sqrt{\phi}, \qquad \tilde{\mu} \approx \beta(1 - \alpha^2).$$

A moda não possui expressão fechada em termos elementares. Para obter a densidade em termos de $(\phi, \tilde{\mu})$ usa-se a relação inversa
$\beta = \frac{\tilde{\mu}}{1-\phi}$, com $0 < \phi < 1$. Logo,

$$f_T(t; \phi, \tilde{\mu}) = \frac{1}{2\sqrt{\phi}\,\frac{\tilde{\mu}}{1-\phi}\sqrt{2\pi}} \left[\sqrt{\frac{\tilde{\mu}/(1-\phi)}{t}} + \left(\frac{\tilde{\mu}/(1-\phi)}{t} \right)^{\frac{3}{2}} \right]$$
$$\cdot \exp\left\{ -\frac{1}{2\phi} \left(\frac{t(1-\phi)}{\mu} + \frac{\tilde{\mu}}{t(1-\phi)} - 2 \right) \right\}, \quad t > 0.$$

Modelo de regressão (moda)

Parametrizando a moda condicional $\tilde{\mu}_i$ por covariáveis $\mathbf{x}_i$ através do vínculo logarítmico (garantindo positividade):

$$\log(\tilde{\mu}_i) = \mathbf{x}_i^\top \boldsymbol{\theta}, \qquad i = 1, \ldots, n,$$

ou seja,

$$\tilde{\mu}_i = \exp(\mathbf{x}_i^\top \boldsymbol{\theta})$$

A densidade condicional de T_i é então $f_{T_i}(t_i; \phi, \tilde{\mu}_i)$ com f_T definida acima.

Tipicamente ϕ (ou uma transformação de ϕ) pode ser mantida constante ou também modelada por covariáveis, dependendo da estrutura inferencial desejada.

Parametrização pelos quantis

Seja $q \in (0,1)$ um nível de quantil fixado. De acordo com Sánchez et al., o $q \times 100^{\text{th}}$ quantil da distribuição BS pode ser definido como

$$Q = t_q = \frac{\beta}{4}\left(\alpha Z_q + \sqrt{\alpha^2 Z_q^2 + 4}\right)^2, \qquad q \in (0,1),$$

em que Z_q representa o $q \times 100^{\text{th}}$ quantil da distribuição normal padrão. Considerando a transformação um a um $(\alpha, \beta) \mapsto (\alpha, Q)$, a função densidade de probabilidade pode ser expressa como

$$f_T(t) = \frac{1}{\alpha\gamma_\alpha\sqrt{8\pi Q_t}}\left(\frac{c}{2} + \frac{2Q}{t}\right)\exp\left[-\frac{2Q}{\alpha^2 ct}\left(\frac{tc}{4Q} - 1\right)^2\right], \qquad t > 0.$$

onde $c = 1 + \frac{\alpha^2}{2}$.

Modelo de regressão (quantis)

Na formulação de regressão por quantil, o quantil Q varia com as observações por meio de covariáveis $\mathbf{x}_i$. A parametrização mais natural e usual para garantir positividade é a ligação logarítmica:

$$\log(Q_i) = \mathbf{x}_i^\top \boldsymbol{\theta}, \qquad i = 1, \ldots, n,$$

ou equivalentemente

$$Q_i = \exp(\mathbf{x}_i^\top \boldsymbol{\theta}).$$

onde $\boldsymbol{\theta}$ é o vetor de parâmetros de regressão que relaciona as covariáveis ao q-ésimo quantil condicional de T_i.

Cada componente θ_j do vetor $\boldsymbol{\theta}$ tem interpretação multiplicativa sobre o q-ésimo quantil condicional: um incremento unitário em x_{ij} altera Q_i por um fator $\exp(\theta_j)$, mantendo as demais covariáveis constantes.

3. Aplicação

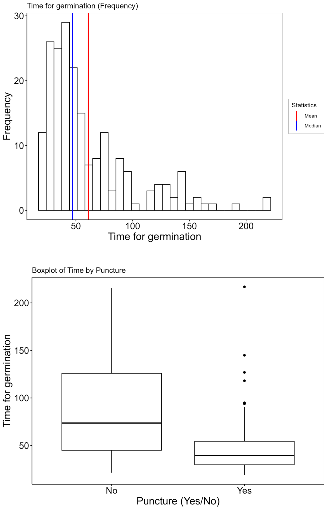
Visão geral dos dados

A semente de mamona é a fonte do óleo de rícino, amplamente conhecido por seu uso cosmético. Esse óleo é também amplamente utilizado na indústria devido a uma característica química única, que não é encontrada em nenhum outro óleo vegetal produzido comercialmente.

Os dados utilizados neste estudo foram obtidos em um experimento cujo objetivo foi avaliar o progresso da hidratação das sementes de mamona sob diferentes temperaturas, variando entre 16 °C e 34 °C, e verificar se a perfuração na casca da semente acelera o processo de hidratação [5]. O tempo até a germinação das sementes também foi registrado.

Visão geral dos dados

Figura: Exploração da variável resposta e covariáveis



Estimações

Tabela: Estimativas dos coeficientes dos modelos

	Média	Moda	q=0.1	Mediana	q=0.9
(Intercept)	3.397*	3.365*	3.153*	3.401*	3.650*
Temperature	-0.014*	-0.013*	-0.013*	-0.013*	-0.013*
Puncture [Yes]	-0.028	-0.029	-0.029	-0.029	-0.029
Time under hydration	0.013*	0.012*	0.012*	0.012*	0.012*

Tabela: Erro padrão das estimativas dos coeficientes dos modelos

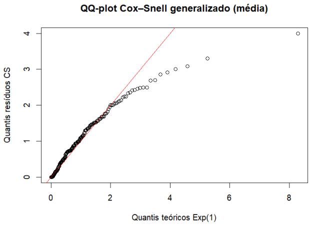
	Média	Moda	q=0.1	Mediana	q=0.9
(Intercept)	0.119	0.118	0.119	0.119	0.112
Temperature	0.003	0.003	0.003	0.003	0.003
Puncture [Yes]	0.034	0.034	0.034	0.034	0.034
Time under hydration	0.010	0.0003	0.0004	0.0004	0.0004

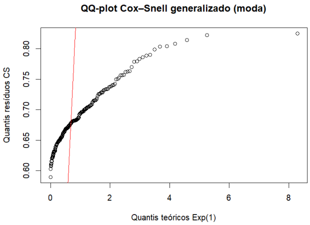
Comparação de modelos

Tabela: AIC/BIC dos modelos

	Média	Moda	q=0.1	Mediana	q=0.9
AIC	-80.48	1503.47	1503.47	1503.47	1503.47
BIC	-63.96	1519.98	1519.98	1519.98	1519.98

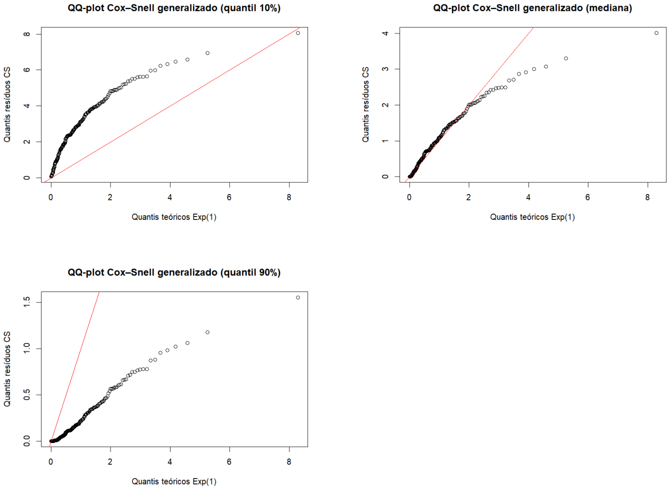
Figura: Diagnóstico dos resíduos: Cox–Snell e quantílicos normalizados





Análise dos resíduos

Figura: Diagnóstico dos resíduos: Cox–Snell e quantílicos normalizados



Atualizações práticas nos códigos

- Quantis: q como argumento da função e garantia de convergência;
- Cálculo de AIC e BIC (revisão necessária).

4. Considerações finais

Considerações finais

- Coeficientes estáveis entre média, moda e quantis $\Rightarrow$ efeito consistente das covariáveis.
- Temperatura reduz o tempo de hidratação; tempo sob hidratação o aumenta.
- Resíduos indicam falta de ajuste adequada do modelo BS (mesmo formato dos erros em todos os casos).
- AIC/BIC favorecem o modelo da média, mas sem ajuste satisfatório.
- Próximos passos: testar distribuições alternativas e modelos BS estendidos.

Referências I

[1] Artur J. Lemonte e Gauss M. Cordeiro. "Birnbaum–Saunders regression models: A review and new developments". Em: *Journal of Statistical Computation and Simulation* 83.5 (2013), pp. 977–1001. DOI: 10.1080/00949655.2011.610721.

[2] James R. Rieck e Jack R. Nedelman. "A Log-Linear Model for the Birnbaum–Saunders Distribution". Em: *Technometrics* 33.1 (1991), pp. 51–60. DOI: 10.1080/00401706.1991.10484652.

[3] Marcelo Bourguignon e Diego I Gallardo. "A new look at the Birnbaum–Saunders regression model". Em: *Applied Stochastic Models in Business and Industry* 38.6 (2022), pp. 935–951.

[4] Luis Sánchez et al. "Birnbaum-Saunders quantile regression and its diagnostics with application to economic data". Em: *Applied Stochastic Models in Business and Industry* 37.1 (2021), pp. 53–73.

[5] Liv Soares Severino. "Castor seed hydration and germination influenced by temperature and puncture". Versão V1. Em: *Redape* (2024). DOI: 10.48432/6N3WQA.

Referências II

[6] Zygmunt W Birnbaum e Sam C Saunders. "Estimation for a family of life distributions with applications to fatigue". Em: *Journal of applied probability* 6.2 (1969), pp. 328–347.

[7] Manoel Santos-Neto et al. "On new parameterizations of the Birnbaum-Saunders distribution". Em: *Pakistan Journal of Statistics* 28.1 (2012), pp. 1–26.

[8] Liv S Severino. "Studies on the Germination and Emergence of Castor Seedlings". Em: *Seeds* 3.2 (2024), pp. 251–268.