CE043 - GAMLSS

Família GAMLSS - Distribuições tipo binomial

Silva, J.P; Taconeli, C.A.

31 de agosto, 2020

Conteúdo

Introdução

Introdução

- Chamamos *distribuições tipo binomial* as distribuições com suporte no conjunto finito $\{0, 1, 2, ..., n\}$.

- Tais distribuições são usualmente aplicadas na análise de dados de contagens com um limite superior fixo e conhecido (n);

- Nas situações em que temos eventos binários (*sucesso vs fracasso*) sob contagem, independentes e com probabilidade de *sucesso* ($0<\mu<1$), então o número de sucessos (Y) tem **distribuição binomial**, com função de probabilidades:

$$P(Y = y)|n, \mu) = \binom{n}{y}\mu^y(1-\mu)^{n-y}, \quad \text{para } y = 0, 1, ..., n.$$

- Para $n = 1$ temos, como caso particular, a distribuição Bernoulli.

- A média e a variância da distribuição binomial são dadas, respectivamente, por:

$$E(Y) = n\mu; \quad Var(Y) = n\mu(1-\mu).$$

- Assim como ocorre para a distribuição Poisson, dados de contagens do tipo binomial também estão sujeito a super (ou sub) dispersão e excesso ou escassez de zeros.

- A biblioteca `gamlss` dispõe de modelos capazes de acomodar tais problemas na análise de contagens com suporte em $\{0, 1, 2, ..., n\}$.

Tabela 1: Distribuições com suporte em $R_y = \{0, 1, 2, ..., n\}$

Distribuição	Nome gamlss	Parâmetro (função de ligação)			
		μ	σ	ν	τ
binomial	`BI`	logit	-	-	-
beta-binomial	`BB`	logit	log	-	-
double-binomial	`DBI`	logit	log	-	-
zero-adj beta-binomial	`ZABB`	logit	log	logit	-
zero-adj binomial	`ZABI`	logit	logit	-	-
zero-inf beta-binomial	`ZIBB`	logit	log	logit	-
zero-inf binomial	`ZIBI`	logit	logit	-	-

Superdispersão e subdispersão

Superdispersão e subdispersão

- Distribuições tipo binomial podem apresentar super ou subdispersão relativa à distribuição binomial.

- Uma alternativa usual para lidar com superdispersão em dados do tipo binomial é a **distribuição beta-binomial**;

- A beta-binomial é produzida por uma mistura envolvendo as duas distribuições que compõem seu nome.

- Seja $Y|\pi \sim \texttt{BI}(n, \pi)$, e considere que π é uma variável aleatória com distribuição beta: $\pi \sim \texttt{BEo}\left(\frac{\mu}{\sigma}, \frac{1-\mu}{\sigma}\right)$, com $E(\pi) = \mu$.

- Dessa forma, marginalmente Y tem distribuição beta-binomial, denotada por $Y \sim \mathrm{BB}(n, \mu, \sigma)$, com n inteiro conhecido, $0 < \mu < 1$, $\sigma > 0$, com média $\mathrm{E}(Y) = n\mu$ e variância:

$$\mathrm{Var}(Y) = n\mu(1-\mu)\left[1 + \frac{\sigma(n-1)}{(1+\sigma)}\right] > n\mu(1-\mu),$$

de maneira que a beta-binomial se apresenta como alternativa à binomial no caso de superdispersão.

Superdispersão e subdispersão

- A **distribuição double-binomial**, denotada por $\texttt{DBI}(n, \mu, \sigma)$, $0 < \mu < 1$, $\sigma > 0$ é uma alternativa à binomial baseada na família exponencial dupla.

- A média de $Y \sim \texttt{DBI}(n, \mu, \sigma)$ é dada por $\mathrm{E}(Y) = n\mu$ e a variância (aproximada) é $\mathrm{Var}(Y) = \sigma n\mu(1 - \mu)$.

- Assim, a distribuição DBI acomoda superdispersão (relativa à binomial) para $\sigma > 1$, e subdispersão quando $\sigma < 1$.

O problema do excesso ou escassez de zeros

O problema do excesso ou escassez de zeros

- Assim como visto anteriormente, para dados de contagens, distribuições do tipo binomial também podem apresentar excesso ou escassez de zeros;

- Novamente, as versões zero-inflacionadas e as zero-ajustadas dos modelos usuais permitem acomodar esse problema;

- As distribuições relacionadas na sequência estão explicitamente implementadas na biblioteca `gamlss`.

O problema do excesso ou escassez de zeros

- **Binomial zero-inflacionada** `ZIBI`(n, μ, σ): associa probabilidade σ a $Y = 0$ e probabilidade (1-σ) a uma `BI`(n, μ);

- **beta-binomial zero-inflacionada** `ZIBB`(n, μ, σ, ν): associa probabilidade ν a $Y = 0$ e probabilidade (1-ν) a uma `BB`(n, μ, σ);

- Importante reforçar que distribuições zero-inflacionadas permitem modelar apenas dados com excesso de zeros (e não escassez).

O problema do excesso ou escassez de zeros

- **Binomial zero-ajustada** `ZIBI`(n, μ, σ): associa probabilidade σ a $Y = 0$ e probabilidade (1-σ) a uma binomial truncada em zero;

- **beta-binomial zero-inflacionada** `ZIBB`(n, μ, σ, ν): associa probabilidade ν a $Y = 0$ e probabilidade (1-ν) a uma beta-binomial truncada em zero;

- Distribuições zero-ajustadas permitem modelar tanto dados com excesso escassez de zeros.

Sessão R

- Nesta sessão R vamos analisar dados usando as distribuições tipo-binomial implementadas no `gamlss`.