

Residual analysis for joint modeling of longitudinal binary data and survival data

Sandro L. L. Lucena¹, Jalmar M. F. Carrasco²

¹ Federal University of Bahia, Salvador, Bahia, Brazil

² Federal University of Bahia, Salvador, Bahia, Brazil



Abbreviated abstract: Survival analysis is broadly used in many areas of science. Models were developed to estimate the time until the occurrence of an event. In this paper, we considered the mixed generalized linear model to fit the longitudinal binary data with a link function logit and survival data were fitted by Cox model. We use the two-stage method to estimate the vector of the parameters. A simulation study was conducted to evaluate the asymptotic behavior of the maximum likelihood estimators obtained by two-stage method and to verify the empirical distribution of the Martingale, quantile, deviance, NRSP and NMSP residuals.

Related publications:

- Hwang et al., Journal of Applied Statistics (46), 2357-2371 (2019)
- Wulfsohn and Tsiatis, Biometrics (53), 330-339 (1997)

Model

Stage 1

Longitudinal data – Mixed linear generalized model (MLGM), extension of GLM Nelder and Wedderburn (1972),
 $\mathbf{Z} = (Z_{i1}, Z_{i2}, \dots, Z_{im_i})$;

$$Z_{ij} | \boldsymbol{\psi}_i \sim \text{Bernoulli}(\pi(\boldsymbol{\psi}_i; t_i)) \text{ (i.i.d)}$$

$$g(\pi(\boldsymbol{\psi}_i; t_i)) = \sum_{l=1}^{q-1} \psi_{il} t^l$$

$$\boldsymbol{\psi}_i \sim \text{Normal}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \text{ (i.i.d)}$$

$$\log(L(\mathbf{Z}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)) = \sum_{i=1}^n \log \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{m_i} f(Z_{ij} | \boldsymbol{\psi}_i) \right\} f(\boldsymbol{\psi}_i) d\boldsymbol{\psi}_i \quad (1)$$

Stage 2

Suppose we can observe $(t_i, \delta_i, \mathbf{x}_i)$ for a variable \mathbf{z}_i in time. We define $\bar{Z}_i = \{Z_i(s), s \leq t\}$ as the average trajectory of $\mathbf{z}_i(t)$. The risco for the i-th individual at time t , given \mathbf{X}_i and \bar{Z}_i , can be expressed using the proportional hazards model (Cox, 1972).

$$\lambda(t | \mathbf{X}_i, \bar{Z}_i) = \lambda_0(t) \exp\{\mathbf{X}_i^T \boldsymbol{\beta} + \kappa \mathbf{Z}_i(t)\} \quad (2)$$

Randomized quantile residual

$$r_i^q = \begin{cases} \Phi^{-1} \left(1 - \exp \left\{ - \int_0^{t_i} \lambda_0(u) \exp[\mathbf{X}^T \boldsymbol{\beta} + \kappa \hat{\pi}(\tilde{\psi}_i, t)] du \right\} \right), & \text{if } \delta_i = 1 \\ \Phi^{-1} \left(U_i \left(1 - \exp \left\{ - \int_0^{t_i} \lambda_0(u) \exp[\mathbf{X}^T \boldsymbol{\beta} + \kappa \hat{\pi}(\tilde{\psi}_i, t)] du \right\} \right) \right), & \text{if } \delta_i = 0 \end{cases}$$

lins.lucena@gmail.com - 2

Application

n=739

Repeated measures: Satisfaction (No=0, yes=1)

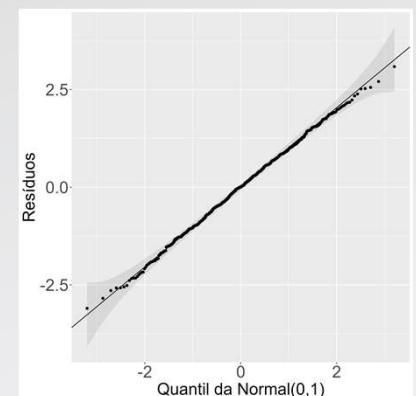
Variables: Age (<75 = 0; $\geq 75=1$); Gender; Hypertension (HBP)

Table 1. Model

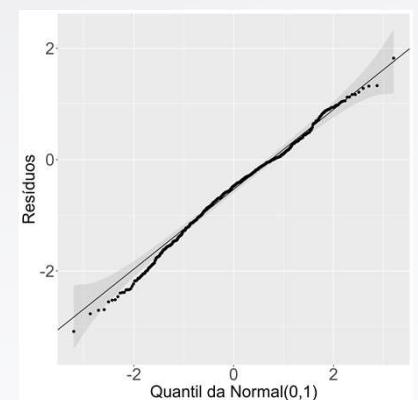
Parameter	Estimate	exp(estimate)	E.P	$IC_{2,5\%}$	$IC_{97,5\%}$	$Pr(>z)$
Age	0,962	2,614	0,0986	0,768	1,155	<0,0001***
Gender	0,408	1,504	0,0948	0,222	0,542	0,0014**
HBP	0,336	1,399	0,1052	0,129	0,594	<0,0008***
κ	-0,954	0,385	0,3676	-1,674	-0,233	0,009**

Table 2. Jackknife

Parameter	Mean	Bies	E.P.	$IC_{2,5\%}$	$IC_{97,5\%}$
Age	0,9605	0,012	0,1281	0,9512	0,9697
Gender	0,4106	0,022	0,1108	0,41226	0,4186
HBP	0,3267	0,009	0,1704	0,3144	0,3389
κ	-1,1063	-0,048	2,1988	-1,2652	-0,9476

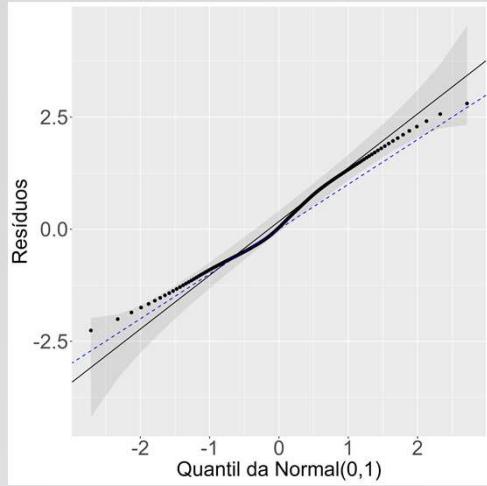


Randomized quantile

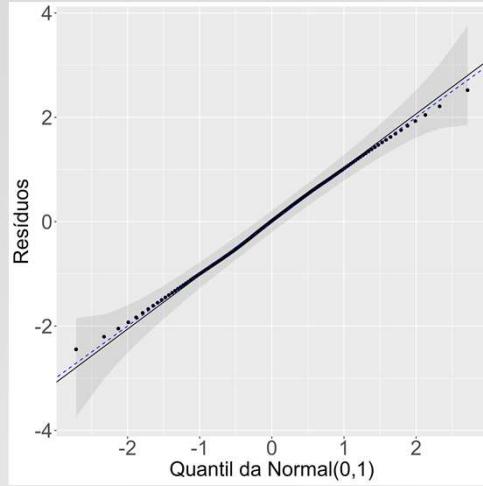


NRSP

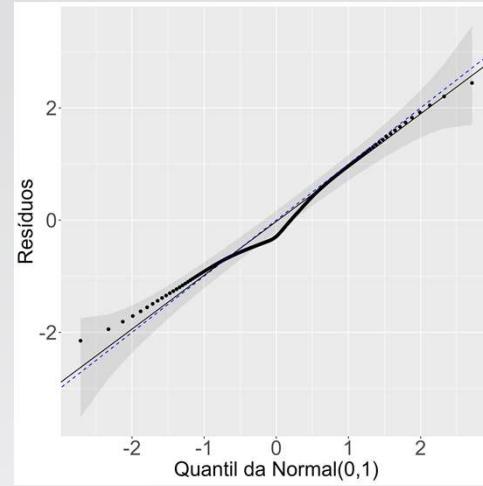
Results and Conclusions



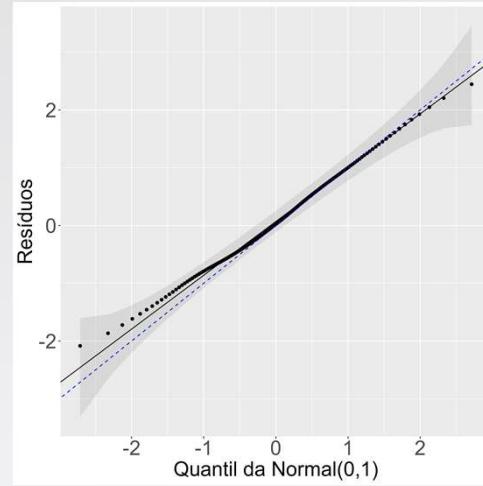
Deviance



Randomized quantile



NMSP



NRSP

- Exemple: $n=150$, $m_i=11$, Monte Carlo replicas = 2000, censoring ratio = 30% , $X_{i1} \sim Bernoulli(0,5)$ and $X_{i2} \sim Uniform(0,1)$

$$\lambda_i(t) = 0,05 \exp \left(0,3x_{i1} + 2x_{i2} - 0,1\hat{\pi}(\tilde{\psi}_i, t) \right)$$

- Residuals are affected when the censoring ratio is 30% or higher;
- The maximum likelihood estimator for κ is biased, independent of the censoring ratio;