Modelos de Tempo de Falha Acelerado para Modelar Tempos Entre Falhas em Poços de Petróleo

Patrícia Borchardt Santos, Renata Santana Fonseca
Programa de Pós-Graduação em Matemática Aplicada e Estatística, UFRN
email: pborchardt@hotmail.com, natfosa@yahoo.com.br

Dione Maria Valença
Departamento de Estatística, UFRN
email: dione@ccet.ufrn.br

1 Introdução

Propomos neste trabalho modelos de tempo de falha acelerado com efeito aleatório para tratar dados de sobrevivência correlacionados e um procedimento para estimação do modelo, que utiliza a quadratura Gauss-Hermite adaptada para aproximar a verossimilhança marginalizada. Usaremos o procedimento proposto para analisar um conjunto de dados referente ao tempo até a falha de poços petrolíferos terrestres produtores de óleo da Bacia Potiguar (RN/CE).

2 Modelo de Tempo de Falha Acelerado com Efeito Aleatório

Sejam T_{ij} o tempo de vida do indivíduo j no grupo i, com $j = 1, \ldots, n_i$, e $i = 1, \ldots, k$ e $\mathbf{x}_{ij}$ um vetor de covariáveis. Denotemos os tempos de censura por C_{ij}, e considere as respostas dadas por $Y_{ij} = min(logT_{ij}, logC_{ij})$. O indicador de falhas δ_{ij} é definido como $\delta_{ij} = I(T_{ij} \leq C_{ij})$, sendo $I(.)$ a função indicadora. Os dados são então representados pelos pares de variáveis aleatórias (Y_{ij}, δ_{ij}), e as covariáveis $\mathbf{x}_{ij}$. Consideramos aqui que os tempos de falha estão sujeitos a censuras à direita e assumimos que o mecanismo de censura é aleatório e que a censura é não informativa.

Consideramos, então, a seguinte representação para o modelo log-linear com efeito aleatório

$$logT_{ij} = U_i + \beta^T \mathbf{x}_{ij} + \sigma\varepsilon_{ij}, \tag{1}$$

sendo ε_{ij}'s erros aleatórios *iid* , com média e variância conhecidas. O vetor β e σ são parâmetros desconhecidos, e para cada grupo temos um efeito aleatório U_i, representado por *v.a's iid* com densidade g, média α e variância θ.

Assumimos que $Cov(U_i, \varepsilon_{ij}) = 0$, e que, dado o efeito aleatório U_i, as respostas dentro do grupo i são independentes. Assumimos também que os efeitos aleatórios U_i são independentes dos tempos de falha e censura.

3 Estimação

Nos modelos com efeito aleatório, os métodos de máxima verossimilhança baseiam-se, em geral, na verossimilhança marginal.

Considere o modelo descrito em (1) e seja $\lambda = (\alpha, \beta^{\mathbf{T}}, \sigma, \theta)^{\mathbf{T}}$ o vetor de parâmetros desconhecidos. A verossimilhança condicional ao efeito do grupo, para o indivíduo j no grupo i é dada por

$$f(y_{ij}|u_i, \mathbf{x}_{ij})^{\delta_{ij}} S(y_{ij}|u_i, \mathbf{x}_{ij})^{1-\delta_{ij}},$$

com f e S denotando, respectivamente, as funções de densidade e sobrevivência condicionais de $logT_{ij}$ dado U_i.

Pela suposição de independência condicional ao efeito do grupo, temos que a verossimilhança condicional para os indivíduos do grupo i é da forma

$$L_i(\beta,\sigma|u_i)=\prod_{j=1}^{n_i} f(y_{ij}|u_i,\mathbf{x}_{ij})^{\delta_{ij}} S(y_{ij}|u_i,\mathbf{x}_{ij})^{1-\delta_{ij}},$$

para $i=1,\ldots,k$.

Assim, a verossimilhança relativa à distribuição marginal de (Y_i,δ_i), denotada por $L_i(\lambda)$, é representada por

$$L_i(\lambda)=\int L_i(\beta,\sigma|u_i)g(u_i;\alpha,\theta)du_i.$$

Apesar da possível correlação existente dentro dos grupos, assumimos independência entre os vetores $(Y_1,\delta_1),\ldots,(Y_k,\delta_k)$. Desta forma, a log-verossimilhança marginal para toda a amostra, é dada por

$$l(\lambda)=\sum_{i=1}^{k} log \int L_i(\beta,\sigma|u_i)g(u_i;\alpha,\theta)du_i. \tag{2}$$

A solução da integral em (2) e sua maximização com respeito a λ, dependem das distribuições postuladas para o efeito aleatório e para os tempos de vida.

4 Aplicação

Foi analisada uma amostra composta por 450 poços-coluna. No período de 2000 à 2006 em que foi observado o tempo até a falha de cada poço, após detectada a falha, os equipamentos eram reparados e o tempo até uma nova falha era observado, totalizando 2363 observações. Sendo assim, o evento de interesse foi medido mais de uma vez em cada poço, caracterizando eventos recorrentes. Este trabalho visa avaliar a influência de características, no tempo de vida desses poços, tais como produção base do poço, método de elevação, quantidade de água produzida, razão gás óleo, profundidade da instalação da bomba, unidade operacional de origem e idade do poço.

Considerando que observações de um mesmo poço podem estar relacionadas, e assumindo um modelo Weibull com efeito aleatório para os dados. Após uma análise prévia optamos pelo seguinte modelo

$$logT_{ij}=U_i+\beta_{prod}PROD_{ij}+\beta_{bsw}BSW_{ij}+\beta_{BM}BM_i+\beta_{rgo}RGO_{ij}+\beta_{id}IDADE_{ij}+\beta_{unic}CAM_i$$
$$+\beta_{unie}ET_i+\beta_{unir}RFQ_i+\beta_{pb}PROFB_{ij}+\sigma\varepsilon_{ij}, \tag{3}$$

com $U_i \sim N(\alpha,\theta)$, representando o efeito aleatório do poço i, e com ε_{ij} representando o erro aleatório do modelo, com distribuição valor extremo padrão.

Após tratar da correlação dos dados, vimos que a estimativa da variância do efeito aleatório é significativa ($\hat{\theta}=0,6257$), e algumas estimativas obtidas para os demais parâmetros sam próximas das encontradas sob o modelo pressupondo independência.

Notamos que os poços nos quais o método de elavação utilizado foi o Bombeio Mecânico apresentaram maior tempo de funcionamento (p-valor=0,006), como visto no Kaplan-Meier e que o tempo de funcionamento dos poços localizados nas unidades operativas ARG e ET não diferem significativamente (p-valor=0,7158), enquanto que nos poços localizados nas unidades CAM e RFQ apresentaram maior tempo de vida.

Verificamos também que, quanto maiores a profundidade da bomba e o valor BSW, maior é o tempo de funcionamento do poço(p-valor=0,0080 e <0,0001, respectivamente). Entretanto, à medida que a produção aumenta, o funcionamento do poço diminui (p-valor<0,0001).