REAMOSTRAGEM- TÉCNICA BOOTSTRAP PARA DADOS DE SOBREVIVÊNCIA ACELERADOS

Sabrina Luzia CAETANO*

RESUMO

O modelo exponencial pode ser usado juntamente com uma relação estresse resposta de potência inversa na determinação do tempo médio de vida. O objetivo deste trabalho consiste do estudo de metodologia para testes acelerados e na proposição de um esquema de utilização desses testes no controle da qualidade industrial.

PALAVRAS-CHAVE: Análise de sobrevivência e confiabilidade. Modelo exponencial. Teoria Assintótica. Técnica Bootstrap.

INTRODUÇÃO

A obtenção de medidas da confiabilidade de produtos manufaturados é imprescindível no controle da qualidade industrial. Em geral, esse controle pode ser feito em meio a linha de produção através dos chamados testes acelerados, que consistem na submissão de um certo número de unidades experimentais a níveis de estresse mais severos que os usualmente utilizados, para assim inferir sob a confiabilidade do produto quando o mesmo é utilizado sob as condições usuais de funcionamento. O principal objetivo consiste do estudo de metodologia para testes acelerados e na proposição de um esquema de utilização desses testes no controle da qualidade industrial, considerando uma distribuição exponencial para os tempos de vida juntamente com uma relação estresse-resposta de potência inversa para a relação entre o parâmetro da exponencial e os níveis de estresse. Técnicas de reamostragem são utilizadas na estimação intervalar dos parâmetros de interesse. Dados simulados ilustram a metodologia proposta.

1 METODOLOGIA

Seja *f(t;θ)* uma função densidade de probabilidade da variável aleatória tempo até a falha de um componente e θ um vetor de parâmetros. A função densidade para os tempos de falha é dada por uma distribuição exponencial da forma: $f(t;\lambda_i) = \lambda_i e^{-\lambda_i t}$,em que $\lambda_i > 0 \;\; e \;\; t \geq 0$ e λ_i é a taxa de risco constante sob o nível de estresse X_i, se $\theta_i = 1/\lambda_i$, então θ_i é o tempo médio de sobrevivência até a falha do componente sob o nível de estresse X_i. Considere que a relação estresse-resposta entre os parâmetros da exponencial e os níveis de estresse é caracterizada pela lei de potência inversa dada por: $\lambda_i = \exp\{-\beta_0 - \beta_1 X_i\}$, em que $-\infty < \beta_0, \beta_1 < \infty$ são parâmetros desconhecidos. Considerando um esquema de censuras tipo II, a função de verossimilhança para k níveis de estresse é: $L(\beta_0, \beta_1) = \exp\{-\beta_0 r - \beta_1 a - e^{-\beta_0} \sum_{i=1}^{k} A_i e^{-\beta_1 X_i}\}$,em que $r = \sum_{i=1}^{k} r_i$ (número

* Professora da Faculdade de Tecnologia de Taquaritinga e doutoranda da Unesp - Jaboticabal em Estatística Aplicada a Genética e Melhoramento Animal.

total de falhas observadas), $a = \sum_{i=1}^{k} r_i X_i$ e $A_i = \sum_{j=1}^{r_i} t_{ij} + (n_i - r_i) t_{ir_i}$. Podemos ter também um estudo com um esquema de censuras do tipo aleatória, a função de verossimilhança permanece a mesma, fazendo $A_i = \sum_{j=1}^{n_i} (\delta_j t_{ij} + (1-\delta_j) t_{ij})$, em que δ_j é igual a 1 se o tempo observado foi uma falha e 0 caso contrário.

1.1 Teoria Assintótica

O intervalo assintótico pode ser construído utilizando os EMV e suas variâncias estimadas. Seja $\beta' = (\beta_0, \beta_1)$ e $L(\beta_0, \beta_1)$ a função de verossimilhança correspondente, e $l(\beta_0, \beta_1) = \log L(\beta_0, \beta_1)$. Seja $\hat{\theta}$ o vetor de EMVs para θ. Para obtermos inferências para θ podemos utilizar a normalidade assintótica dos EMV, $\hat{\beta} \sim N(\beta, I^{-1}(\beta))$, em que $I_{ij}(\beta) = -E\left(\frac{\partial \ln L}{\partial \beta_i \partial \beta_j}\right)$ são os elementos da matriz de informação de Fischer.

1.2 Método Delta

Para a estimação dos tempos médios de sobrevivência θ_i, o método Delta pode ser utilizado porque θ_i é uma função dos parâmetros β_0, β_1. Sabemos que: $\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim N\left[\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right]$.

Então, $\hat{\theta}_i = \exp\{\hat{\beta}_0, \hat{\beta}_1 X_i\} \sim \left[(\theta_i), \left(\sigma_{11}\left(\frac{\partial \theta_i}{\partial \beta_0}\right)^2 + 2\sigma_{12}\left(\frac{\partial \theta_i}{\partial \beta_0}\right)\left(\frac{\partial \theta_i}{\partial \beta_1}\right) + \sigma_{22}\left(\frac{\partial \theta_i}{\partial \beta_1}\right)^2\right)\right]$ e

os intervalos de confiança de 95% são dados por: $\hat{\theta}_1 \pm 1.96\sqrt{\hat{\text{var}}(\hat{\theta}_i)}$.

1.3 Bootstrap

Existem dois tipos básicos de *Bootstrap*. O *Bootstrap* paramétrico em que os EMVs são obtidos através do modelo ajustado, isto é, geramos dados do modelo ajustado com os valores dos parâmetros fixados nos EMV obtidos da amostra original, e o *Bootstrap* não-paramétrico em que os EMVs são baseados em *R* reamostras com reposição obtidas da amostra original.

A simulação via Bootstrap visa à obtenção de estimativas intervalares empíricas para os estimadores dos parâmetros de interesse. Seja β o parâmetro de interesse e para cada amostra calculamos o EMV para β e temos no final de R reamostragens, $\hat{\beta}_1^* < \ldots < \hat{\beta}_R^*$ valores dos EMV ordenados. Utilizamos então: $\hat{\beta}^*_{(R+1)(\frac{\alpha}{2})}$ *e* $\hat{\beta}^*_{(R+1)(1-\frac{\alpha}{2})}$, como sendo os limites inferiores e superiores do intervalo *100(1-α)%* de confiança para β.

2 RESULTADOS E DISCUSSÃO

Aplicando a metodologia acima e considerando uma amostra de 30 elementos com $X_i = -\log(V_i)$, onde Vi é uma variável de voltagem com $V_1=5$ (nível usual), $V_2=10$ e $V_3=30$, com $\beta_0=5.7366$, $\beta_1=0.6$, foram estudados 5 casos para cada tipo de intervalo, observando-se: dados sem censura, com 10% e 30% de censura à direita e com 10% e 30% de censura aleatória.

Dentre os intervalos estudados, existe o interesse de verificarmos qual deles é o intervalo de confiança mais adequado, consideramos para análise a aproximação assintótica, a via *Bootstrap* paramétrico e a via *Bootstrap* não paramétrico. Como a aproximação assintótica é a forma mais utilizada, temos como objetivo principal através da simulação, mostrar que para pequenas amostras ela se torna inviável, ou seja, ela tem uma probabilidade de cobertura muito baixa em relação aos outros métodos testados, sendo os outros métodos, a estimação via *Bootstrap* paramétrico e a estimação via *Bootstrap* não paramétrico.

Devemos deixar claro, que estamos trabalhando com pequenas amostras devido ao fato de estarmos utilizando testes acelerados, pois afinal a confiabilidade de determinados produtos não pode ser testada através de grandes amostras, devido ao custo e ao tempo.
O estudo de simulação realizado baseou-se na geração de 100 conjuntos de dados, desta forma quando geramos cada distribuição exponencial, trabalhamos com amostras de tamanho: 10 (o que acontece geralmente em testes acelerados), 20, 50, 100 e 300.

As tabelas com os resultados obtidos são dadas a seguir:

Tabela 1. Intervalos Assintóticos (prob. de cobertura em porcentagem)

n/censura	não cens.	Censura direita à		Censura aleatória	
		10%	30%	10%	30%
10	12%	21%	31%	16%	18%
20	39%	43%	24%	37%	42%
50	50%	60%	65%	62%	68%
100	72%	74%	55%	70%	76%
300	94%	90%	82%	94%	98%

Através da verificação da tabela 1, podemos visualizar claramente, que a probabilidade de cobertura de amostras de tamanho 10 é baixa, principalmente para amostras não censuradas. Observando agora, as situações nas quais os valores amostrais aumentam, podemos verificar que a probabilidade de cobertura aumenta à medida que a quantidade de amostras também aumenta, comprovando a teoria assintótica.

Tabela 2. Intervalos via Bootstrap não paramétrico (prob. de cobertura em porcentagem)

n/censura	não cens.	Censura direita		à	Censura aleatória	
		10%	30%		10%	30%
10	87%	85%	59%		93%	89%
20	89%	88%	49%		94%	87%
50	91%	85%	50%		92%	89%
100	85%	84%	60%		95%	88%
300	90%	83%	57%		91%	90%

Verificando os intervalos não paramétricos, que é o nosso maior enfoque, devido ao fato de ser um intervalo simples de se manusear e por obtermos resultados satisfatórios, observamos através da tabela 2 que a probabilidade de cobertura para amostras de tamanho 10, que é o nosso principal interesse, é alta em relação aos valores obtidos pela estimação assintótica.

Tabela 3. Intervalos via Bootstrap paramétrico (prob. de cobertura em porcentagem)

n/censura	não cens.	Censura direita		à	Censura aleatória	
		10%	30%		10%	30%
10	39%	71%	64%		68%	56%
20	81%	98%	82%		93%	92%
50	68%	83%	47%		80%	99%
100	95%	99%	99%		93%	91%
300	99%	99%	99%		99%	99%

Agora analisando os intervalos via *Bootstrap* paramétrico podemos verificar através da tabela 3 que a probabilidade de cobertura para amostras de tamanho 10 também é interessante, porém não tão quanto os intervalos via *Bootstrap* não paramétrico. Algo que podemos visualizar é a probabilidade de cobertura quando trabalhamos com amostra de tamanho 100 e 300, a qual é superestimada.

3 CONCLUSÃO

Em suma a utilização das técnicas de reamostragem, apresentadas para construção de intervalos de confiança para os parâmetros de interesse, proporcionam vantagens em relação às técnicas assintóticas uma vez que são baseadas em evidências empíricas. Também se constituem de alternativas eficientes quando a suposição de normalidade assintótica não é válida, assim através dos estudos de simulação podemos verificar claramente que para pequenas amostras a probabilidade de cobertura via intervalos assintóticos é muito baixa, enquanto que para os intervalos *Bootstrap* a probabilidade de cobertura é muito maior e geralmente adequada.

ABSTRACT

The main objective of this paper is to propose a quality control procedure based on an exponential model with a power law stress relationship.

KEYWORDS: *Survival analysis and reliability. Exponential Model. Asymptotic Theory. Bootstrap Technique.*

REFERÊNCIAS

DAVISON, A.C., HINKLEY, D.V. *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press, 1997.
LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York, 1982.
LOUZADA-NETO, F. & ACHCAR, J. A. (1993) - *Uso de Dados Acelerados no Controle de Produtos Industriais Assumindo Uma Dist. Exponencial e Um Modelo Estresse-Resposta Geral*.Est., 45, 144, 145.
NELSON, W. Accelerated *Testing: Statistical Models, Test plans and Data Analysis.* John Wiley and Sons, New York, 1990.