

UTILIZAÇÃO DA DISTRIBUIÇÃO INVERSA DE CHEN PARA A MODELAGEM DOS TEMPOS DE SOBREVIVÊNCIA DE COBAIAS INOCULADAS PELO VÍRUS DA TUBERCULOSE

USE OF THE CHEN INVERSE DISTRIBUTION TO THE MODELING OF THE SURVIVAL TIMES OF GUINEA PIGS INOCULATED BY VIRUS TUBERCULOSIS.

Paulo Roberto de Lima Gianfelice, Sérgio Minoru Oikawa, Fernando Antônio Moala

Universidade Estadual Paulista - UNESP - Faculdade de Ciências e Tecnologias, Departamento de Estatística, Presidente Prudente, São Paulo.
E-mail: geanfelice@hotmail.com

RESUMO - Com o intuito de modelar o tempo em que o curso de uma patologia se desenvolve, o presente trabalho empregou os conceitos de análise de sobrevivência no desenvolvimento de um modelo estatístico para estudar os tempos de vida dos hospedeiros do vírus da tuberculose. Com base nos resultados, constatou-se que o modelo probabilístico Inversa de Chen forneceu um excelente ajuste aos dados de sobrevivência, quando eles apresentam função de risco unimodal. Além disso, concluiu-se que, numa comparação com outros modelos propostos pela literatura, o modelo em questão se mostrou o mais adequado para analisar e descrever os dados sobre os tempos de vida no desenvolvimento da infecção por tuberculose.

Palavras-chave: Ajustamento de Curva; Análise de Sobrevivência; Análise Estatística; Modelagem Estatística; Modelo de Risco.

ABSTRACT - With the intention to model the time on what the course of an pathology if develops, in this study we used the survival analysis concepts to obtain a statistical model to be used in life span study of host of the vírus the tuberculosis. Based on the results, it was found that the probabilistic model Chen Inverso provided an excellent fit to survival data, since these showed unimodal risk curve. Besides that, it was concluded in a comparison with other models proposed in the literature that the model in question if showed to be most suitable to analyze and describe the life-time data in development of infection tuberculosis.

Keywords: Adjustment Curve; Statistical Modeling; Statistical Analysis; Survival Analysis; Risk Model.

1. INTRODUÇÃO

A infecção por bacilo da tuberculose (TB) representa um grave problema de saúde pública, que de acordo com Sant'Anna e Hijjar (2007), só no Brasil, estima-se a ocorrência de 111.000 novos casos hospitalares por ano, com cerca de 6.000 óbitos, indicando a necessidade e importância de estudos em caráter preventivo e de combate desta patologia.

No entanto, uma análise dos resultados de um experimento de inoculação deste vírus realizados em 1414 cobaias animais (porcos da índia) foi interpretada por *Bjerkedal* (1960) como um mostrador de que as diferenças de virulência dos bacilos infectantes só mudam o aspecto quantitativo da relação entre os números iniciais de bacilos infectantes e o tempo de sobrevivência das cobaias. O estudo, portanto, deixa claro o uso de cobaia animal e a sua dependência direta da análise de sobrevivência em seus respectivos tempo de vida.

Segundo Morales (2008, p. 33-36), ao contrário do que muitos pensam, o interesse por métodos alternativos em experimentação tem crescido na comunidade científica no intuito de reduzir o número de animais utilizados em experimentos. Dessa forma, como método alternativo, Morales destaca em suas propostas que,

[...] os modelos matemáticos podem contribuir para o trabalho experimental através da definição de variáveis e testes de teorias, reduzindo o custo desses experimentos e os tornando mais eficazes. [...] computadores processam e armazenam conhecimentos já existentes e muitos deles foram adquiridos com a utilização de animais na pesquisa.

Nesse contexto, a partir de um conjunto de dados correspondentes aos tempos de sobrevivência (em dias) de cobaias animais, o presente estudo teve como objetivo propor um modelo estatístico com a finalidade de analisar os tempos de vida dos porcos da índia infectados com o bacilo da tuberculose. E em particular, associar-se às técnicas de análise de sobrevivência como um método alternativo no refinamento dos modelos matemáticos tomados para simular a utilização de cobaia animal em pesquisas científicas.

2. METODOLOGIA

Análise de Sobrevivência

A análise de sobrevivência é a metodologia adotada neste trabalho, e é definida como um método estatístico usado para análise de dados de tempo de vida, comumente derivados de estudos de laboratórios ou de clínicas relacionadas à Biomedicina em tratamentos de doenças agudas, severas ou fatais, e em áreas de conhecimento em geral.

Geralmente, estuda-se o tempo em que uma unidade experimental sobrevive a um

determinado tratamento, isto é, o tempo de resposta a um dado evento ou o tempo em que uma unidade experimental desenvolveu uma dada doença específica.

Comumente, no tratamento dos dados efetuam-se as estatísticas descritivas e gráficas e em paralelo realizam-se inferências e testes das estimativas levantadas sobre os dados de sobrevivência pelos métodos do EMV (estimador de máxima verossimilhança) e o EKM (estimador de Kaplan - Meyer) a fim de tomar os resultados obtidos à cerca da experimentação de interesse para se alcançar as conclusões mais precisas e confiáveis

$$f(x|\alpha; \beta) = \alpha\beta x^{\beta-1} \exp\{x^\beta + \alpha[1 - \exp(x^\beta)]\} \quad (1)$$

Então, X segue uma distribuição de Chen com parâmetros α e β . Denota-se $X \sim \text{Chen}(\alpha; \beta)$.

Agora, considere T como sendo uma variável aleatória tal que $T = X^{-1}$. Então, utilizando-se o método de transformação de variáveis aleatórias, neste caso, a função de

sobre o caso praticado. Para maiores detalhes, sugere-se consultar Colosimo e Giolo (2006).

O Modelo Proposto: A Distribuição Inversa de Chen com Dois Parâmetros

Inicialmente, citado por Srivastava e Srivastava (2014), Chen (2000) propôs uma distribuição de probabilidade com dois parâmetros a fim de estudar os tempos de vida, denominada distribuição de Chen. Se X é uma variável aleatória com função densidade de probabilidade dada por:

distribuição acumulada (fda) e a função densidade de probabilidade (fdp) de T são dadas, respectivamente, por:

$$F(t|\alpha; \beta) = \exp\{\alpha[1 - \exp(y^{-\beta})]\} \quad (2)$$

$$f(t|\alpha; \beta) = F'(y|\alpha; \beta) = \alpha\beta y^{-(\beta+1)} \exp\{y^{-\beta} + \alpha[1 - \exp(y^{-\beta})]\} \quad (3)$$

em que $\alpha > 0$ e $\beta > 0$. Dessa forma, t segue uma distribuição inversa de Chen e é denotada por $T \sim \text{Chen}^{-1}(\alpha; \beta)$.

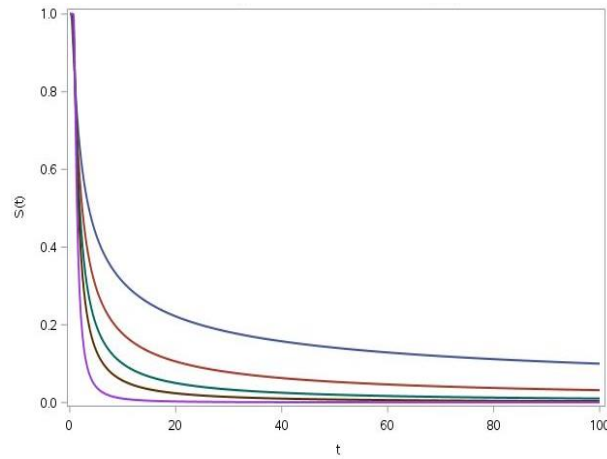


Figura 1. Gráficos ilustrativos de $S(t)$ para Chen^{-1}

A sua função de sobrevivência, denotada por $S(t)$ é definida como $S(t) = P(T \geq t)$, isto é, a probabilidade de ocorrência do evento T no instante t , em que $S(t) = 1 -$

$F(t)$. Assim, a função de sobrevivência e a função de risco (taxa de falha) são dadas, respectivamente, pelas expressões (4) e (5).

$$S(t|\alpha; \beta) = 1 - \exp\{\alpha[1 - \exp(t^{-\beta})]\} \tag{4}$$

$$\lambda(t|\alpha; \beta) = \frac{\alpha\beta t^{-(\beta+1)} \exp\{t^{-\beta} + \alpha[1 - \exp(t^{-\beta})]\}}{1 - \exp\{\alpha[1 - \exp(t^{-\beta})]\}} \tag{5}$$

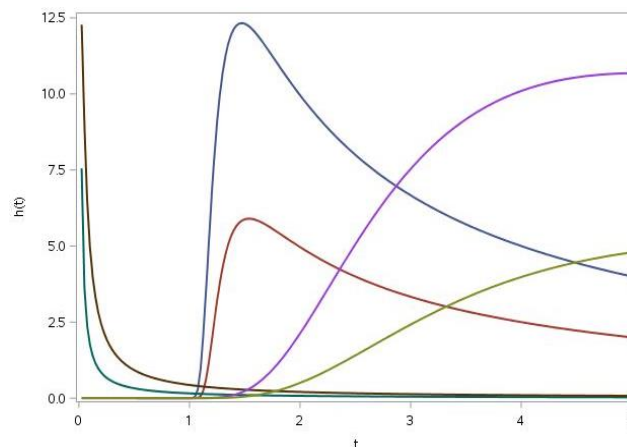


Figura 2. Gráficos ilustrativos de $\lambda(t)$ para Chen^{-1}

A Figura 1 ilustra a função de sobrevivência para diferentes valores dos parâmetros. De acordo com a Figura 2, para

qualquer valor de α e β , observa-se que a distribuição de inversa de Chen apresenta funções de risco, $\lambda(t)$, unimodal, monótona

decrecente e monótona crescente para alguns intervalos de tempo $t \in [t_1; t_2]$. Neste caso, este modelo é predominantemente côncavo, significando que a distribuição inversa de Chen é adequada para analisar dados com taxa de falha unimodal. Mais detalhes sobre o modelo probabilístico de $\text{Chen}^{-1}(\alpha; \beta)$ pode ser obtido em Srivastava e Srivastava (2014).

O Gráfico do Tempo Total em Teste (TTT)

Existem diversos procedimentos preliminares para identificar as formas da função de risco em análise de sobrevivência. Dentre eles, destaca-se o método gráfico conhecido como tempo total em teste (TTT) que é comumente empregado em contextos onde existem informações qualitativas sobre a curva de risco.

$$L(\alpha, \beta | x) = \prod_{i=1}^n \alpha \beta x_i^{-(\beta+1)} \exp\{x_i^{-\beta} + \alpha[1 - \exp(x_i^{-\beta})]\} \quad (6)$$

O logaritmo da verossimilhança é dado por:

$$l(\alpha, \beta | x) = n(\log(\alpha\beta) - (\beta + 1) \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \{x_i^{-\beta} + \alpha[1 - \exp(x_i^{-\beta})]\}) \quad (7)$$

Os estimadores de máxima verossimilhança são obtidos como solução das equações dadas por:

$$\frac{\partial l(\alpha, \beta | x)}{\partial \alpha} = \frac{n}{\alpha} + n - \sum_{i=1}^n \exp(x_i^{-\beta}) = 0 \quad (8)$$

$$\frac{\partial l(\alpha, \beta | x)}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n (x_i^{-\beta}) + \alpha \sum_{i=1}^n [\exp(x_i^{-\beta}) x_i^{-\beta} \log(x_i)] = 0 \quad (9)$$

Como não existe uma solução fechada para (8) e (9), foram usadas métodos numéricos para encontrar as estimativas de

A informação qualitativa diz respeito ao comportamento da função de risco correspondente aos dados que, em geral, são curvas que podem assumir formas unimodais, banheira, constante, monótona crescente ou decrescente.

Estimação de Máxima Verossimilhança

Utilizou-se o método de máxima verossimilhança para estimar os parâmetros desconhecidos α e β . Supondo que X_1, X_2, \dots, X_n é uma amostra aleatória de uma distribuição $\text{Chen}^{-1}(\alpha; \beta)$, então a função de verossimilhança dos dados observados é dada por:

máxima verossimilhança de α e β . Neste trabalho, utilizaram-se diversos pacotes do

software R, Livre (www.r-project.org), dentre eles, `maxLik()`.

O Conjunto de Dados

No presente trabalho, a metodologia proposta foi utilizada para analisar o

conjunto de dados reais correspondentes aos tempos de sobrevivência, em dias, de 72 cobaias infectadas com doses do bacilo da tuberculose. Tais dados foram extraídos de Bjerkedal (1960) e podem ser visualizados no quadro a seguir.

Quadro 1. Tempo de sobrevivência das cobaias infectadas

12	15	22	24	24	32	32	33	34	38	38	43	44	48	52	53	54	54
54	56	57	58	58	59	60	60	60	60	61	62	63	65	65	67	68	70
70	72	73	75	76	76	81	83	84	85	87	91	95	96	98	99	109	110
121	127	129	131	143	146	146	175	175	211	233	258	258	263	297	341	341	376

Fonte: Bjerkedal (1960).

3. RESULTADOS

Inicialmente, verificou-se o comportamento da função de risco resultante dos dados de sobrevivência do Quadro 1, através do gráfico TTT (Figura 3). Constata-se que o mesmo apresenta a forma unimodal para a função taxa de risco.

Neste caso, assumiu-se a distribuição de inversa de Chen pelo fato dos dados de sobrevivência interceptar o eixo anti-diagonal do gráfico TTT na forma de S invertido, sendo uma parte da curva côncava seguida de outra parte convexa, necessariamente, nesta

ordem ao longo do eixo das abscissas. Na sequência, a fim de realizar comparações dos ajustes da distribuição de Chen^{-1} com outras distribuições comumente usadas na análise de sobrevivência, consideraram-se as distribuições Log-Normal e Log-Logística que também apresentam risco unimodal. Para tanto, utilizou-se o método de máxima verossimilhança na obtenção dos valores estimados dos parâmetros de forma e escala dos modelos considerados. Os resultados são apresentados no Quadro 1.

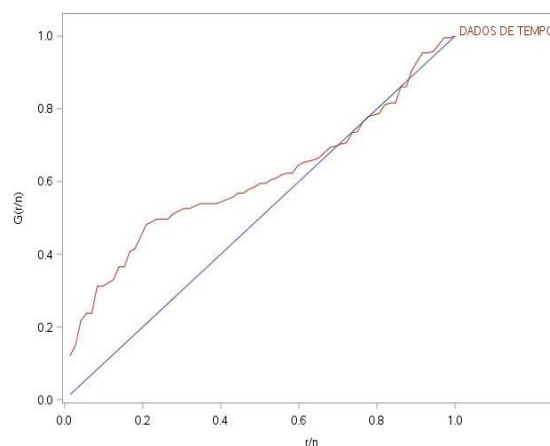


Figura 3. Gráfico TTT para os dados reais

Com base nestas estimativas, foi possível modelar os comportamentos das funções de sobrevivência e de taxa de risco dos modelos

Log-Normal, log-Logística e Chen^{-1} a fim de compará-los analiticamente e graficamente.

Tabela 1. Estimativas dos parâmetros de forma e escala e os respectivos IC(95%)

Modelo	$(\hat{\alpha}; \hat{\beta})$	Erros padrões de $(\hat{\alpha}; \hat{\beta})$	Intervalo 95% de Confiança
CHEN^{-1}	(282,971; 1,415)	(4,224; 0,031)	(274,67; 291,25) e (1,35; 1,48)
LOG-NORMAL	(4,344; 1,005)	(0,119; 0,118)	(4,11; 4,58) e (0,77; 1,24)
LOG-LOGISTICA	(75,221; 2,482)	(2,966; 0,248)	(69,41; 81,03) e (1,99; 2,96)

Analiticamente, a comparação entre os três modelos considerados foram realizados através das respectivas funções penalizadoras, denominadas de critérios de informação (veja a Tabela 2) que identificam aqueles que geram os menores valores e, portanto, indicam os melhores ajustes. Neste

trabalho foram considerados os critérios de informação de Akaike (AIC) e o Bayesiano de Schwarz (BIC), conforme são apresentados na tabela a seguir:

Tabela 2. Critérios de Informação

Modelo	AIC	BIC
CHEN^{-1}	795,47	808,58
LOG-NORMAL	784,67	897,78
LOG-LOGISTICA	783,98	797,09

Embora os modelos Log-Normal e Log-Logística apresentem os valores dos critérios de informações menores, as diferenças não são significantes entre os três valores e por isso nenhuma conclusão deve ser tomada em relação ao melhor ajuste.

Graficamente, os modelos ajustados cujos comportamentos das curvas de sobrevivência se encontram dentro da região de confiança fornecida pelo estimador de Kaplan-Meyer (Figura 4) foram considerados aceitáveis. O modelo Chen^{-1} se mostrou um pouco

melhor no ajustamento para a região de confiança gerada devido o fato da sua curva de sobrevivência romper os limites da região de confiança durante um intervalo de tempo menor em relação aos outros dois modelos. Ou seja, aproximadamente nos intervalos de tempo $t \in [40; 50]$ no modelo Chen^{-1} , em $t \in [0, 115]$ no Log-Normal e $t \in [150, 350]$ no Log-Logística.

Embora a decisão em assumir que o melhor ajuste da curva de sobrevivência seja atribuído ao modelo Chen^{-1} , tal decisão

pode ser considerada subjetiva ao leitor através da análise gráfica, uma vez que, mesmo que a curva se encontre dentro da região de confiança durante mais tempo de duração, os três modelos rompem os limites

estipulados. Além disso, os critérios de informação dos três modelos não são significativos para reforçar tal decisão.

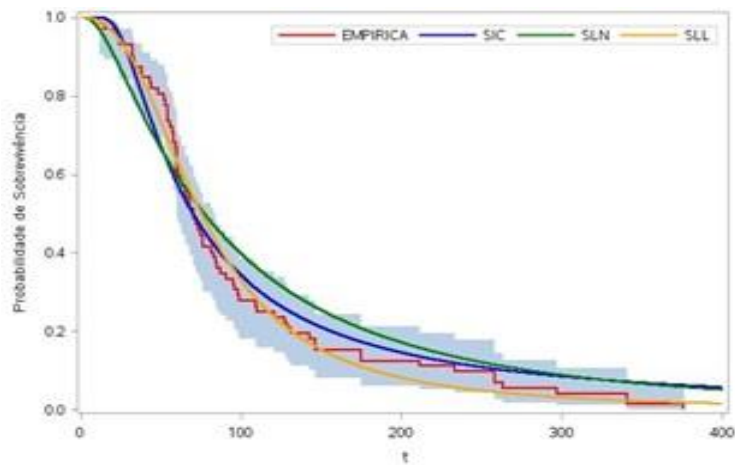


Figura 4. Gráfico de sobrevivência para as Cobaias

Contudo, uma terceira comparação se fez necessária para a conclusão do melhor ajustamento e, por este motivo, uma segunda análise gráfica foi realizada para verificar a forma da função de risco estimada pelos três modelos e comparadas com a função de risco empírica obtida dos dados do Quadro 1.

A Figura 5 apresenta as formas das funções de risco dos três modelos comparados com a estimativa da função de risco empírica, usando o método baseado em kernel proposto por Hess et al. (1999).

Na sequência é apresentado o *EQM* (erro quadrático médio) dos três modelos de risco em relação a taxa de falha empírica.

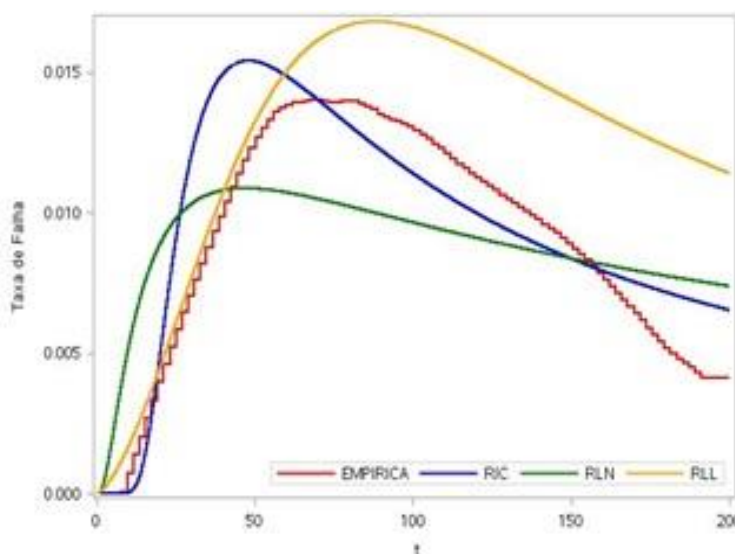


Figura 5. Curvas de risco estimadas

Tabela 3. EQM entre os modelos de risco estimados e o empírico

Modelo	CHEN ⁻¹	LOG-NORMAL	LOG-LOGISTICA
EQM	4,6X10 ⁻⁶	8,0X10 ⁻⁶	1,65X10 ⁻⁵

4. CONCLUSÃO

Portanto, os resultados das comparações gráficas e analíticas das funções de sobrevivência e de risco evidenciam o melhor desempenho do modelo Chen⁻¹ em relação aos modelos, Log-Normal e Log-Logística, considerado os dados sobre os tempos de vida das cobaias infectadas com doses do bacilo da tuberculose.

Tal conclusão é evidenciada pelo EQM da tabela 3, onde constata-se que $4,6 \times 10^{-6} < 8,0 \times 10^{-6} < 1,65 \times 10^{-5}$. Essa desigualdade garante que o modelo Chen⁻¹ apresenta o melhor ajustamento aos dados do experimento considerado.

AGRADECIMENTOS

Os autores agradecem o revisor da comissão científica pelas valiosas sugestões.

REFERÊNCIAS

BJERKEDAL, T. Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. **American Journal Epidemiology**, Oxford, v.72, p.130–148, 1960.
<https://doi.org/10.1093/oxfordjournals.aje.a120129>

COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. 1. ed. São Paulo: Edgard Blucher, 2006. 392p.

HESS, K.R.; SERACHITOPOL, D.M.; BROWN, B.W. Hazard function estimators: a simulation study. **Statistics in Medicine**, v.18, p.3075-3088, 1999.
[https://doi.org/10.1002/\(SICI\)1097-0258\(19991130\)18:22<3075::AID-SIM244>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-0258(19991130)18:22<3075::AID-SIM244>3.0.CO;2-6)

MORALES, M.M. Métodos alternativos à utilização de animais em pesquisa científica: mito ou realidade?. **Ciência e Cultura**, São Paulo, v.60, p.33-36, 2008.

SANT'ANNA, C. C.; HIJJAR, M. A. Recente contribuição da Organização Mundial de Saúde para o controle da tuberculose na infância. **Revista de Saúde Pública**, São Paulo, v.41, p.117-120, 2007.
<https://doi.org/10.1590/S0034-89102007000800016>

SRIVASTAVA, P. K.; SRIVASTAVA R. S. Two parameter inverse chen distribution as survival model . **International Journal of Statistika and Matematika**, Gorakhpur, v.11, p.12-16, 2014.