



APLICAÇÕES DE DISTRIBUIÇÕES DISCRETIZADAS NA ANÁLISE DE DADOS DE SOBREVIVÊNCIA

Danielle Peralta¹, Josmar Mazucheli¹, Isabelle Picada Emanuelli² e Robson Marcelo Rossi¹

¹Departamento de Estatística, Universidade Estadual de Maringá.

²Departamento de Pós-Graduação, UNICESUMAR, Mestrado em Tecnologias Limpas.

RESUMO

A proposta deste trabalho é apresentar a distribuição Burr XII discreta de três parâmetros a partir do procedimento proposto por Nakagawa e Osaki (1975). A distribuição Burr XII tem se mostrado bastante útil na análise de dados de sobrevivência, tendo como caso limite a distribuição Weibull. Neste trabalho, analisou-se um conjunto de dados em que a variável resposta representa o número de dias transcorridos entre o nascimento e a primeira postura de três linhagens de codornas submetidas a dois níveis de dieta alimentar. O objetivo foi selecionar uma distribuição de probabilidade que descreva o comportamento da idade (em dias) até a primeira postura e verificar a qualidade do ajuste via o teste qui-quadrado. Além da distribuição Burr XII, adotou-se também como candidatas as distribuições Weibull, Pareto tipo II e Gama. Os procedimentos de discriminação mostraram que a distribuição Burr XII discreta é a mais indicada para descrever a variável em questão.

Palavras chave: Análise de sobrevivência, função de probabilidade, função de verossimilhança, produção animal, seleção de modelos.

1 INTRODUÇÃO

Em análise de sobrevivência a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Na área médica, esses tempos podem ser o tempo até a morte do paciente, da cura ou da recorrência da doença. Em confiabilidade de sistemas, pode-se ter o tempo até a falha de um produto ou de um componente. Os tempos de sobrevivência são, em sua grande maioria, contínuos. No entanto, existem casos que os dados são estritamente discretos, seja por não ser possível ou conveniente medir esses tempos em uma escala contínua, seja devido a imprecisões intrínsecas ao processo de mensuração ou simplesmente por que os dados são de origem discreta. Como, por exemplo, o número de peças que uma máquina produz ao longo da sua vida útil ou o número de oscilações de tensão que um item elétrico suporta antes de fracassar.

Diversos trabalhos envolvendo a discretização de distribuições contínuas tem sido publicados nos últimos anos, alguns dos exemplos mais importantes são: Weibull discreta [4][8], distribuições Burr discreta e Pareto discreta [2], Burr XII Generalizada discreta [6] e Pareto Generalizada discreta e Lomax discreta [7].

Neste trabalho, a distribuição Burr XII discreta é utilizada para descrever o número de dias transcorridos entre o nascimento e a primeira postura (evento) de três linhagens de codornas submetidas a dois níveis de dieta alimentar.

2 DISTRIBUIÇÃO BURR XII DISCRETA

Uma variável aleatória T , não negativa, com distribuição Burr XII contínua tem função de sobrevivência na forma:

$$S(t | \Theta) = \left[1 + \left(\frac{t}{\theta} \right)^\alpha \right]^{-\gamma}, \quad (1)$$

A função de probabilidade da distribuição Burr XII discreta, é obtida através da relação

$$P(T = t) = S(t) - S(t + 1) \text{ para}$$

$$t=0,1,2,\dots(2)$$

e portanto,

$$P(T = t) = \left[1 + \left(\frac{t}{\theta} \right)^\alpha \right]^{-\gamma} - \left[1 + \left(\frac{t+1}{\theta} \right)^\alpha \right]^{-\gamma} \quad (3)$$

$$= \phi^{\log[1+(\frac{t}{\theta})^\alpha]} - \phi^{\log[1+(\frac{t+1}{\theta})^\alpha]} \quad (4)$$

em que $\phi = e^{-\gamma}$, $0 < \phi < 1$ influencia a escala (eixo dos tempos) da função de probabilidade, $\theta > 0$ é o parâmetro de escala e $\alpha > 0$ é o parâmetro de forma. A função de probabilidade (4) é unimodal e seu ponto de máximo, T_{\max} , deve ser obtido via método numérico.

2.1 ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA

Sejam t_1, \dots, t_n os tempos de sobrevivência observados de uma variável aleatória T com distribuição Burr XII discreta as funções de verossimilhança e log-verossimilhança [5, 3] são escritas, respectivamente, como:

$$L(\Theta | \mathbf{t}) = \prod_{i=1}^n \left\{ \phi^{\log[1+(\frac{t_i}{\theta})^\alpha]} [1 - \phi^{\Lambda(t_i, \alpha, \theta)}] \right\} \quad (5)$$

e

$$l(\Theta | \mathbf{t}) = \sum_{i=1}^n \left\{ \log \left[1 + \left(\frac{t}{\theta} \right)^\alpha \right] \log(\phi) + \log [1 - \phi^{\Lambda(t_i, \alpha, \theta)}] \right\} \quad (6)$$

em que $\Lambda(t_i, \alpha, \theta) = \log \left\{ \left[1 + \left(\frac{t+1}{\theta} \right)^\alpha \right] \left[1 + \left(\frac{t}{\theta} \right)^\alpha \right]^{-1} \right\}$.

As estimativas de máxima verossimilhança $\hat{\Theta} = (\hat{\theta}, \hat{\alpha}, \hat{\phi})$ de $\Theta = (\theta, \alpha, \phi)$ são obtidas pela maximização do logaritmo da função verossimilhança. O sistema de equações formado não admitem soluções analíticas e devem ser resolvidas numericamente por meio de um método iterativo, como por exemplo, algoritmo de Newton-Rhapson.

3 MATERIAIS E MÉTODOS

Os dados utilizados referem-se ao tempo, em dias, do nascimento até a primeira postura de codornas (*Coturnix coturnix japonica*) provenientes de três linhagens e submetidas a duas dietas alimentares. Esses dados foram obtidos em um experimento conduzido entre agosto de 2006 e janeiro de 2007 no Setor de Coturnicultura da Fazenda Experimental da Universidade Estadual de Maringá. Aos 28 dias de idade as codornas passaram pelo processo de sexagem e as fêmeas foram divididas em dois grupos e alojadas em gaiolas individuais. Cada grupo passou a receber um de dois tipos de dietas: a primeira com nível preconizado de energia metabolizável de 2.900 kcal/kg (dieta I) e a segunda com nível de energia de 2.500 kcal/kg (dieta II).

Além da distribuição Burr XII discreta, foram ajustadas outras três distribuições de probabilidade também discretas para descrever o tempo até a primeira postura de codornas, entre elas estão as distribuições discretas Weibull [4] e Pareto II [7] e a Gama [1]

A escolha da distribuição mais apropriada baseia-se nos valores das estatísticas, tais como: Critério da função log-verossimilhança e os critérios de informação, tais como: Akaike (AIC), Akaike corrigido (AICc) e Bayesiano (BIC). O critério de decisão é favorável ao modelo que apresenta o menor valor da estatística.

O teste de hipóteses qui-quadrado foi adotado para verificar a qualidade do ajuste.

Tabela 1: Valores da estatísticas usadas como critério de discriminação — linhagem amarela.

Modelo	$-2 \log(L)$	AIC	AICc	BIC	R
Dieta I					
1	1008,4 ¹	1014,4 ¹	1014,5 ¹	1023,6 ¹	4 ¹
2	1133,9 ⁴	1137,9 ⁴	1138,0 ⁴	1144,1 ⁴	16 ⁴
3	1088,9 ³	1092,9 ³	1092,9 ³	1099,0 ³	12 ³
4	1050,4 ²	1054,4 ²	1054,5 ²	1060,6 ²	8 ²
Dieta II					
1	911,7 ¹	917,7 ¹	917,9 ¹	926,7 ¹	4 ¹
2	1032,1 ⁴	1036,1 ⁴	1036,2 ⁴	1042,1 ⁴	16 ⁴
3	979,3 ³	983,3 ³	983,4 ³	989,3 ³	12 ³
4	960,3 ²	964,3 ²	964,4 ²	970,2 ²	8 ²

1: Burr XII, 2: Weibull, 3: Pareto e 4: Gama.

Tabela 2: Valores da estatísticas usadas como critério de discriminação — linhagem azul.

Modelo	$-2 \log(L)$	AIC	AICc	BIC	R
Dieta I					
1	1243,3 ¹	1249,3 ¹	1249,5 ¹	1259,0 ¹	4 ¹
2	1308,8 ³	1312,8 ³	1312,8 ³	1319,2 ³	12 ³
3	1311,3 ⁴	1315,3 ⁴	1315,3 ⁴	1321,7 ⁴	16 ⁴
4	1250,6 ²	1254,6 ²	1254,7 ²	1261,1 ²	8 ²
Dieta II					
1	1313,3 ¹	1319,3 ¹	1319,5 ¹	1329,1 ¹	4 ¹
2	1400,1 ³	1404,1 ³	1404,2 ³	1410,6 ³	12 ³
3	1413,6 ⁴	1417,6 ⁴	1417,7 ⁴	1424,1 ⁴	16 ⁴
4	1347,8 ²	1351,8 ²	1351,8 ²	1358,3 ²	8 ²

1: Burr XII, 2: Weibull, 3: Pareto e 4: Gama.

4 RESULTADOS E DISCUSSÕES

As Tabelas 1, 2 e 3 apresentam os valores das estatísticas usadas como critério de seleção de modelos. Na coluna nomeada por R , tem-se a indicação da distribuição que mais vezes foi selecionada. Observa-se que a distribuição Burr XII discreta apresenta o melhor ajuste em relação às outras distribuições.

Dado que a distribuição Burr XII discreta foi a distribuição escolhida para descrever o tempo até a primeira postura de codornas em todos os grupos, as Tabelas 4, 5 e 6 apresentam as estimativas de máxima verossimilhança, pontuais e intervalares, dos parâmetros da distribuição para cada grupo (linhagem *versus* dieta).

O teste qui-quadrado de aderência e a análise subjetiva das curvas de sobrevivência teóricas sobrepostas às empíricas foram usados para verificar a qualidade do ajuste. A Tabela 7 apresentam as estatísticas do teste qui-quadrado. Observa-se que a hipótese H_0 : os dados seguem distribuição Burr XII discreta, foi rejeitada

Tabela 3: Valores da estatísticas usadas como critério de discriminação — linhagem vermelha.

Modelo	$-2 \log (L)$	AIC	AICc	BIC	R
Dieta I					
1	1329,7 ¹	1335,7 ¹	1335,8 ¹	1345,6 ¹	4 ¹
2	1400,8 ⁴	1404,8 ⁴	1404,9 ⁴	1411,4 ⁴	16 ⁴
3	1397,8 ³	1401,8 ³	1401,9 ³	1408,4 ³	12 ³
4	1345,1 ²	1349,1 ²	1349,2 ²	1355,7 ²	8 ²
Dieta II					
1	1268,6 ¹	1274,6 ¹	1274,7 ¹	1284,1 ¹	4 ¹
2	1376,4 ⁴	1380,4 ⁴	1380,5 ⁴	1386,8 ⁴	16 ⁴
3	1324,6 ³	1328,6 ³	1328,7 ³	1335,0 ³	12 ³
4	1305,6 ²	1309,6 ²	1309,6 ²	1315,9 ²	8 ²

1: Burr XII, 2: Weibull, 3: Pareto e 4: Gama.

Tabela 4: Estimativas de máxima verossimilhança distribuição Burr XII discreta — linhagem amarela.

Dieta	Parâmetro	Estimativa	E.P.	I.C. de 95%	
				L.I.	L.S.
I	$\hat{\theta}$	37,1243	0,7027	35,7471	38,5015
I	$\hat{\alpha}$	20,4649	2,9545	14,6741	26,2557
I	$\hat{\gamma}$	0,4070	0,0939	0,2229	0,5911
II	$\hat{\theta}$	36,5446	0,6572	35,2566	37,8326
II	$\hat{\alpha}$	23,3931	3,8410	15,8648	30,9214
II	$\hat{\gamma}$	0,3146	0,0767	0,1643	0,4649

Tabela 5: Estimativas de máxima verossimilhança distribuição Burr XII discreta — linhagem azul.

Dieta	Parâmetro	Estimativa	E.P.	I.C. de 95%	
				L.I.	L.S.
I	$\hat{\theta}$	49,7602	1,3748	47,0657	52,4547
I	$\hat{\alpha}$	17,0846	2,5546	12,0777	22,0915
I	$\hat{\gamma}$	0,5655	0,1600	0,2519	0,8791
II	$\hat{\theta}$	47,9006	0,7031	46,5225	49,2787
II	$\hat{\alpha}$	28,0484	5,1697	17,9159	38,1810
II	$\hat{\gamma}$	0,2246	0,05512	0,1166	0,3327

Tabela 6: Estimativas de máxima verossimilhança distribuição Burr XII discreta – linhagem vermelha.

Dieta	Parâmetro	Estimativa	E.P.	I.C. de 95%	
				L.I.	L.S.
I	$\hat{\theta}$	46,1090	0,8636	44,4164	47,8015
I	$\hat{\alpha}$	29,0137	7,1225	15,0538	42,9736
I	$\hat{\gamma}$	0,2474	0,0828	0,0852	0,4096
II	$\hat{\theta}$	48,1991	0,8640	46,5058	49,8925
II	$\hat{\alpha}$	24,7298	4,7665	15,3877	34,0719
II	$\hat{\gamma}$	0,2393	0,0628	0,1162	0,3624

Tabela 7: Resultados dos testes qui-quadrado.

Linhagem	Dieta	χ^2	G.L.	Valor-p
Amarela	I	1,8028	3	0,6143
	II	6,0331	3	0,1100
Azul	I	5,7568	3	0,1240
	II	4,4264	3	0,2189
Vermelha	I	10,0003	2	0,0067
	II	6,7584	3	0,0800

somente para o grupo da linhagem vermelha e dieta I, em nível de significância de 5%. Nos demais grupos, há evidências amostrais que os dados seguem distribuição Burr XII discreta.

Na Figura 1 tem-se as curvas de sobrevivência empíricas e as baseadas nas estimativas de máxima verossimilhança. Tanto as curvas empíricas como as provenientes dos ajustes das distribuições podem ser nomeadas de curvas de postura e estimam a probabilidade do tempo até a primeira postura ser superior a um número de dias quaisquer. Nota-se que a seleção da distribuição segundo a inspeção informal das curvas de sobrevivência teóricas e empíricas, estão de acordo com a seleção baseada nos valores das estatísticas apresentadas nas análises anteriores.

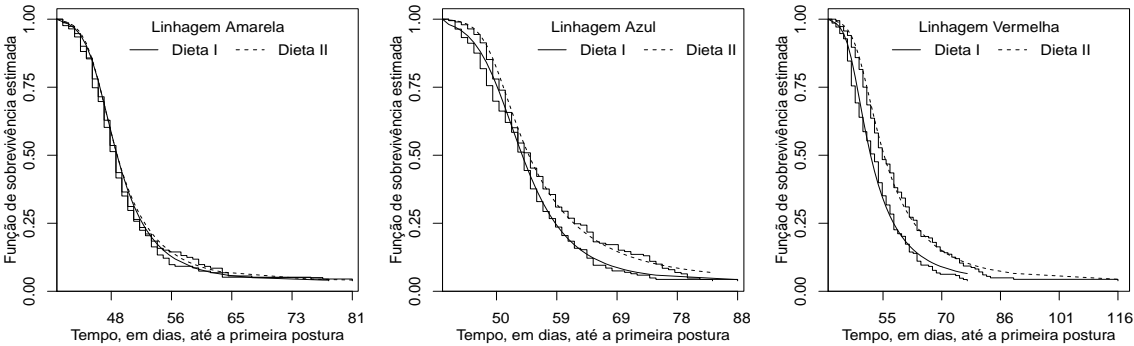


Figura 1: Função de sobrevivência empírica e teórica.

5 CONSIDERAÇÕES FINAIS

Neste trabalho, buscou-se encontrar uma distribuição de probabilidade discreta para explicar o comportamento da idade até a primeira postura de codornas. Segundo vários critérios de discriminação, a distribuição Burr XII discreta foi escolhida, dentre 3 distribuições rivais. Para cada idade provável de primeira postura, calculou-se a probabilidade de ocorrência. Na linhagem amarela, a idade com maior probabilidade de ocorrência da primeira postura é de 38 dias, para ambas dietas. As linhagens azul e vermelha, apresentaram os mesmos resultados: os tempos mais prováveis são $T = 50,5$ dias para a dieta I e $T = 50$ dias, para a dieta II.

Referências

- [1] BAKOUCH, H. S., JAZI, M. A., NADARAJAH, S., 2014. A new discrete distribution. *Statistics* 48 (1), 200–240.
- [2] KRISHNA, H., PUNDIR, P. S., 2009. Discrete burr and discrete pareto distributions. *Statistical Methodology* 6 (2), 177 – 188.
- [3] MILLAR, R. B., 2011. *Statistics in Practice*. John Wiley Sons, Ltd.
- [4] NAKAGAWA, T., OSAKI, S., Dec 1975. The discrete weibull distribution. *Reliability, IEEE Transactions on R-24* (5), 300–301.
- [5] PAWITAN, Y., 2001. In *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, Oxford.
- [6] PARA, B., JAN, T., 2014. Discretization of burr-type iii distribution. *Journal of Reliability and Statistical Studies* 7 (2), 87–94.
- [7] PRIETO, F., GÓMEZ-DÉNIZ, E., SARABIA, J. M., 2014. Modelling road accident blackspots data with the discrete generalized pareto distribution. *Accident Analysis Prevention* 71 (0), 38 – 49.
- [8] STEIN, W. E., DATTERO, R., 1984. A new discrete weibull distribution. *Reliability, IEEE Transactions on* 33 (2), 196–197.