

Funções densidade de probabilidade para a estimativa de precipitação mensal

Probability density functions for the estimation of monthly rainfall

Jailson de Araujo Rodrigues¹; Jaime dos Santos Filho²; Lucas Monteiro Chaves³

Resumo

O objetivo deste trabalho foi avaliar o potencial das distribuições gama, log-normal e Weibull na modelagem de precipitação mensal na cidade de Bento Gonçalves, Rio Grande do Sul. Foi utilizada uma série histórica de precipitações pluviométricas mensais no período de janeiro de 2001 a dezembro de 2011. A aderência das distribuições às condições pluviométricas foi verificada via teste de Kolmogorov-Sminorv. As três distribuições apresentaram aderência. Entretanto, os resultados obtidos indicaram que o modelo gama ajustou-se mais adequadamente às condições pluviométricas da cidade de Bento Gonçalves.

Palavras-chave: Comparação de Modelos. Distribuição Gama. Distribuição Log-normal. Distribuição Weibull.

Abstract

The objective of this work was to evaluate the potential of gamma, log-normal and Weibull distributions modeling of monthly precipitation in the city of Bento Gonçalves, Rio Grande do Sul. We used a series of monthly rainfall from January 2001 to December 2011. Adherence to the distributions of rainfall conditions was verified using the Kolmogorov-Sminorv test. The three distributions were adherent. However, the results indicated that the gamma model fit better rainfall conditions in the city of Bento Gonçalves.

Keywords: Comparison of Models. Gamma Distribution. Lognormal Distribution. Weibull Distribution.

Introdução

Medir dados de precipitação pluviométrica é muito importante em diversos contextos, tais como produtividade de culturas agrícolas, manejo dos recursos hídricos, avaliação ambiental, erosão hídrica, etc. A obtenção da correta distribuição temporal para precipitação é relevante no planejamento agrícola, no que diz respeito à instalação de culturas. Além da influência na

agricultura, períodos de estiagens muito longos afetam o nível de água dos mananciais e dos reservatórios das usinas hidrelétricas, trazendo problemas para o abastecimento urbano e na geração de energia elétrica. Segundo Vieira e Carvalho (2001), a quantificação das chuvas é importante no planejamento agrícola e ambiental para o correto dimensionamento das obras, tanto na construção civil quanto na conservação do solo.

¹ Docente do Instituto Federal da Bahia; jailsondearaujo@yahoo.com.br.

² Discente do Doutorado em Estatística e Experimentação Agropecuária, UFLA: jsf_ba@yahoo.com.br.

³ Docente do Departamento de Ciências Exatas, UFLA: lucasi@dex.edu.br.

O papel das precipitações pluviométricas é destacado por Botelho e Morais (1999) como principal fonte de água para a agricultura, que por vezes tem comprometido o desenvolvimento da produção agrícola em função de sua variabilidade não uniforme, ora com grandes períodos de estiagem, ora com consideráveis aguaceiros de curta duração, que superam a capacidade de retenção de água pelo solo, provocando enchentes e inundações. A impossibilidade de saber qual a evolução exata dos valores de precipitação ao longo do tempo e do espaço em função de sua natureza aleatória é alertada por Sampaio *et al.* (2007). A partir dessas dificuldades, diferentes modelos probabilísticos são utilizados para descrever o comportamento tanto da precipitação esperada como do período de precipitação.

Longo *et al.* (2006) avaliaram o uso das distribuições gama e log-normal na estimativa de precipitações pluviais quinzenais no Estado do Paraná, a aderência dos modelos foi verificada via teste qui-quadrado e teste de Kolmogorov-Smirnov. Os resultados mostraram que a distribuição gama obteve melhor ajuste das condições pluviométricas do Estado.

Uma generalização do modelo gama foi introduzida por Nadarajah e Gupta (2007) para estudar a intensidades de secas ocorridas no Estado de Nebraska. Os resultados indicaram que a nova distribuição fornecia um excelente ajuste para os dados de seca, melhor que o ajuste obtido com o modelo gama usual.

Nadarajah e Yang (2006) estudaram a viabilidade da distribuição do produto XY de variáveis aleatórias quando X é exponencialmente distribuída e Y é proveniente da família gama, Weibull, Pareto e log-normal na descrição da magnitude de secas ocorridas em Nebraska-EUA.

Sampaio *et al.* (2007) através dos modelos gama e log-normal estudaram a distribuição da precipitação mensal para o Estado do Paraná. Os resultados obtidos indicaram que a distribuição gama ajustou-

se mais adequadamente à pluviometria do Estado.

Uma análise comparativa entre os modelos gama, Weibull, normal, exponencial e log-normal no ajuste a dados reais de precipitações em Santa Maria no Rio Grande do Sul foi feita por Silva *et al.* (2007), as distribuições gama e Weibull foram as que apresentaram melhores resultados na descrição da variação da probabilidade de ocorrência de precipitação diária.

A distribuição da combinação linear de variáveis gama e Rayleigh foi utilizada por Shakil e Kribia (2009) para modelar dados de precipitações. A boa qualidade de ajuste do modelo foi verificada através do teste qui-quadrado de aderência.

Rodrigues, Silva e Chaves (2011) utilizaram a distribuição gama para modelar o período de dias com ocorrência de precipitações pluviométricas. A base de dados analisada correspondia às medições diárias do índice pluviométrico na cidade de Passo Fundo no Estado do Rio Grande do Sul. Critérios gráficos indicaram o bom ajuste do modelo em relação aos dados de precipitação analisados.

O objetivo deste trabalho foi avaliar as distribuições gama, log-normal e Weibull na descrição da precipitação pluviométrica mensal na cidade de Bento Gonçalves no Rio Grande do Sul, verificando-se as respectivas aderências por meio do teste de Kolmogorov-Smirnov à nível de 0,05 de probabilidade. A escolha do modelo que apresenta melhor ajuste foi feita via Critério de Informação de Akaike, gráfico P-P plot e p-valores do teste de Kolmogorov-Smirnov.

Materiais e Métodos

Os dados explorados foram coletados na estação meteorológica da Embrapa Uva e Vinho com sede local em Bento Gonçalves no nordeste do Estado do Rio Grande do Sul. A estação climatológica está situada à latitude de 29°09'44"S, longitude de 51°31'50"W e altitude de 640m. As observações

correspondem às medições mensais do índice pluviométrico no período de janeiro de 2001 a dezembro de 2011.

Neste estudo realizou-se o ajuste das distribuições gama, log-normal e Weibull (JOHNSON; KOTZ; BALAKRISHNAN, 1995).

Modelo gama: Uma variável aleatória X tem distribuição gama com parâmetros $\alpha > 0$ e $\beta > 0$ quando sua função densidade de probabilidade (fdp) é:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta x) \quad (1)$$

em que $x > 0$ e $\Gamma(\cdot)$ representa função gama,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt \quad (2)$$

Se X_1, \dots, X_n é uma amostra aleatória de (1), as equações de máxima verossimilhança de α e β são:

$$n \log \left(\frac{n \hat{\alpha}}{\sum_{i=0}^n x_i} \right) - n \psi(\hat{\alpha}) + \sum_{i=0}^n \log x_i = 0 \quad (3)$$

$$\hat{\beta} = \frac{n \hat{\alpha}}{\sum_{i=0}^n x_i} \quad (4)$$

em que $\psi(\cdot)$ é a função digama $\psi(\alpha) = d \log(\alpha) / d\alpha$.

Modelo log-normal: Uma variável aleatória X tem distribuição log-normal com parâmetros $-\infty < \mu < +\infty$ e $\sigma > 0$ quando sua fdp é:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right] \quad (5)$$

em que $x > 0$.

Se X_1, \dots, X_n é uma amostra aleatória de (5), as estimativas de máxima verossimilhança de μ e σ^2 são:

$$\hat{\mu} = \frac{1}{n} \sum_{i=0}^n \log x_i \quad (6)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^n \left(\log x_i - \frac{1}{n} \sum_{i=0}^n \log x_i \right)^2 \quad (7)$$

Modelo Weibull: Uma variável aleatória X tem distribuição Weibull com parâmetros $\alpha > 0$ e $\beta > 0$ quando sua fdp é:

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right] \quad (8)$$

em que $x > 0$.

Se X_1, \dots, X é uma amostra aleatória de (8), as estimativas de máxima verossimilhança de α e β são:

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=0}^n x_i^{\hat{\alpha}} \right)^{1/\hat{\alpha}} \quad (9)$$

$$\hat{\alpha} = \left[\left(\sum_{i=0}^n x_i^{\hat{\alpha}} \log x_i \right) \left(\sum_{i=0}^n x_i^{\hat{\alpha}} \right)^{-1} - \frac{1}{n} \sum_{i=0}^n \log x_i \right]^{-1} \quad (10)$$

Para verificar a aderência das distribuições descritas, foi utilizado o teste estatístico de Kolmogorov-Smirnov, que além de não depender do número de classes do agrupamento dos dados, oferece mais vantagens computacionais. Esse teste fundamenta-se na discrepância entre as distribuições, $D_n = \sqrt{n} \sup_x |F_n(x) - F_0(x)|$ em que $F_n(x)$ denota a distribuição teórica e $F_0(x)$ a distribuição ajustada. Essa estatística é usada para testar a hipótese nula $H_0 : F = F_0$ versus hipótese alternativa $H_1 : F \neq F_0$ (KVAM; VIDAKOVIC, 2007).

Ao trabalharmos com modelos é necessário que internalizemos sua condição de aproximação da realidade, ou seja, não existem modelos probabilísticos exatos, sempre haverá perda de informações, fato inerente à presença da componente de aleatoriedade. Dessa forma, é necessário selecionar o "melhor" modelo dentre aqueles utilizados no ajuste de dados. Burnham e Anderson (2004) enfatizam que tal escolha deve ser embasada em princípios científicos. Dentre as metodologias empregadas para esse fim, uma das mais utilizadas é o critério de informação de Akaike (AIC) estabelecido por Akaike (1983), baseado na teoria de decisão, o AIC é definido como a quantidade:

$$AIC = -2L + 2p \quad (11)$$

em que L representa o magnitude máxima da função suporte e p denota o número de parâmetros. De acordo com esse critério, o melhor modelo é aquele que apresenta o menor valor de AIC.

A qualidade de ajuste das distribuições também foi verificada utilizando os gráficos de probabilidade, os P-P plots. Esses gráficos consistem em representar no plano cartesiano a probabilidade observada *versus* a probabilidade estimada (esperada) de determinada distribuição. O procedimento de construção desses gráficos é feito utilizando os seguintes passos:

- 1) Dada uma amostra aleatória X_1, \dots, X_n , rearrumar a amostra em ordem crescente, $X_{(1)}, \dots, X_{(n)}$;
- 2) Determinar as probabilidades observadas $\tilde{P}_i = F(x_{(i)})$ para $i = 1, \dots, n$ em que $F(\cdot)$ é a distribuição proposta;
- 3) Determinar as probabilidades acumuladas $P_i = (i - c)/(n - 2c + 1)$ correspondentes ao valor da função de distribuição teórica para $x_{(i)}$ em que $i = 1, \dots, n$, representa uma estimativa da função de distribuição no ponto $x_{(i)}$. O valor da constante c pode variar para as diferentes distribuições;

- 4) Plotar o gráfico de dispersão para os pares de pontos (\tilde{P}_i, P_i) em que $i = 1, \dots, n$.

Se a pressuposição sobre a distribuição está correta os pares de pontos estarão dispostos sobre uma linha reta, ou próximos dela.

Resultados e Discussão

A precipitação média mensal observada para a cidade de Bento Gonçalves foi de 151,814mm. Observou-se uma variabilidade muito grande nos dados de precipitações mensais, com medição mínima de 17,100mm, máxima de 417,800mm e erro padrão de 83,091mm.

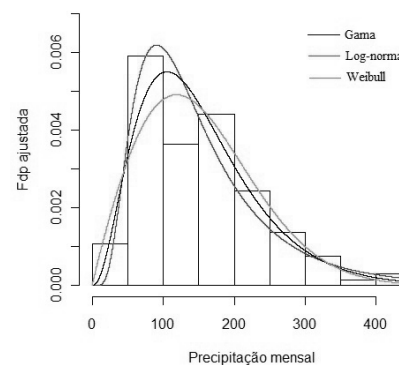
Tabela 1 - Estimativas dos parâmetros das distribuições gama, log-normal e Weibull.

Modelo	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}^2$
Gama	3,264	0,022
Log-normal	4,862	0,596
Weibull	1,945	171,796

Fonte: Elaborada pelo autor.

Na Tabela 1 podem ser observadas as estimativas de máxima verossimilhança encontradas nos ajustes das distribuições gama, Weibull e log-normal, calculadas utilizando o programa R (versão 2011). As fdp ajustadas podem ser visualizadas na Figura 1.

Figura 1 - Histograma de frequência dos dados observados de precipitação mensal sobreposto pelas fdp ajustadas gama, log-normal e Weibull.



Fonte: Elaborada pelo autor.

Ao observar a Figura 1, pode-se notar que as curvas de ajustes das três distribuições acompanharam as condições de moda e assimetria, não sendo possível determinar, visualmente, a que melhor se ajustou aos dados observados.

Tabela 2 - Valores de p do teste de Kolmogorov-Smirnov e AIC.

Modelo	p-valor	AIC
Gama	0,642	1519,356
Log-normal	0,264	1525,613
Weibull	0,575	1522,017

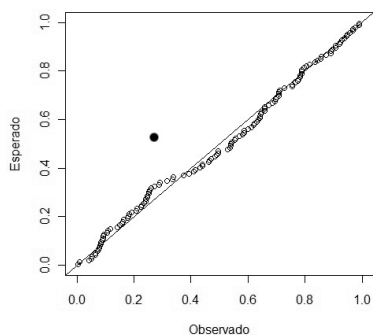
Fonte: Elaborada pelo autor.

A aderência das distribuições foi comprovada segundo o teste de Kolmogorov-Smirnov, todos os p-valores foram superiores a 0,05 indicando que as distribuições descrevem satisfatoriamente os dados observados de precipitação, ver Tabela 2. Em particular, a distribuição gama foi a que obteve maior aderência seguida da distribuição Weibull e a que obteve menor aderência foi à distribuição log-normal.

Os resultados obtidos no teste de Kolmogorov-Smirnov foram confirmados pelos respectivos valores obtidos para os AIC's que podem ser visualizados na Tabela 2. A distribuição gama foi a que teve o melhor desempenho, com o menor AIC em relação aos demais. O maior AIC foi da distribuição log-normal confirmando o status de modelo menos adequado.

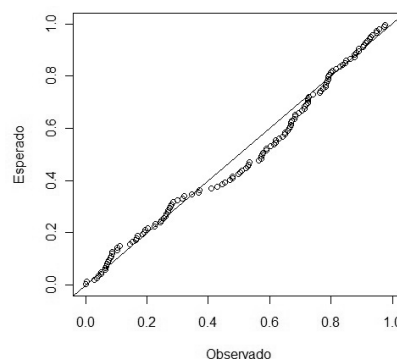
Os P-P plots para as três fdp ajustadas são exibidos nas Figuras 2 a 4.

Figura 2 - P-P plot da fdp gama ajustada.



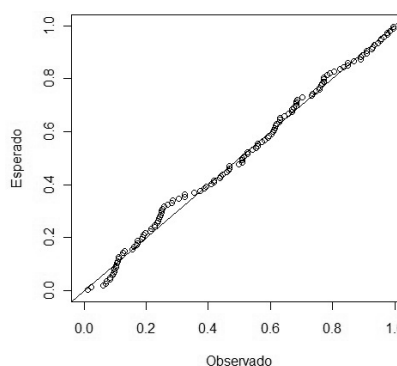
Fonte: Elaborada pelo autor.

Figura 3 - P-P plot da fdp log-normal ajustada.



Fonte: Elaborada pelo autor.

Figura 4 - P-P plot da fdp Weibull ajustada.



Fonte: Elaborada pelo autor.

Ao observar as Figuras de 2 a 4, pode-se notar que a distribuição log-normal foi a que pior se ajustou em relação aos dados observados e os gráficos referentes às distribuições gama e Weibull indicam um bom ajuste. Entretanto, os gráficos são inconclusivos em relação a que apresenta melhor ajuste. Dessa forma, o critério gráfico de P-P plot acompanhou de certa forma os resultados obtidos anteriormente via teste de Kolmogorov-Smirnov e o Critério de Informação de Akaike.

Conclusões

O teste de Kolmogorov-Smirnov indicou a aderência das distribuições gama, log-normal e Weibull no nível de 0,05 de significância. No entanto, a distribuição gama foi a que obteve maior aderência e em seguida a distribuição Weibull. A

distribuição log-normal foi a que apresentou menor aderência.

O Critério de Informação de Akaike confirmou os resultados obtidos pelo teste de Kolmogorov-Smirnov, indicando o modelo gama como o mais adequado na descrição dos dados pluviométricos de Bento Gonçalves. O modelo de pior desempenho segundo esse critério foi o log-normal.

Os P-P plots também indicaram que as três distribuições descrevem adequadamente a série de precipitações. No entanto, foi inconclusivo com respeito ao melhor ajuste, indicando apenas que a distribuição log-normal é a de pior ajuste em relação aos dados.

Referências

- AKAIKE, H. Information measures and model selection. *International Statistical Institute*, Netherlands, v. 44, p. 277-291, 1983.
- BOTELHO, V. A. V. A.; MORAIS, A. R. Estimativas dos parâmetros da distribuição gama de dados pluviométricos de Município de Lavras, Estado de Minas Gerais. *Ciência e Agrotecnologia*, Lavras, v. 23, n. 3, p. 697-706, 1999.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. *Sociological Methods and Research*, Beverly Hills, v. 33, n. 2, p. 261-304, 2004.
- LONGO, J. L.; SAMPAIO, C. S.; QUEIROZ, M. M. F.; SUSZEK, M. Uso das distribuições gama e log-normal na estimativa de precipitação provável quinzenal. *Revista Varia Scientia*, Cascavel, v. 6, n. 3, p. 107-118, 2006.
- NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. *Mathematics and Computers in Simulation*, Amsterdam, v. 74, p. 1-7, 2007.
- NADARAJAH, S.; YANG, D. W. Drought modeling and products of random variables with exponential kernel. *Stochastic Environmental Research and Risk Assessment*, New York, v. 21, n. 2, p. 123-129, 2006.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous univariate distributions*. New York: John Wiley & Sons, 1995.
- KVAM, P. H.; VIDAKOVIC, B. *Nonparametric statistics with applications to science and engineering*. New York: Wiley, 2007.
- R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org. 2011.
- RODRIGUES, J. A.; SILVA, A. P. C. M.; CHAVES, L. M. Distribuição gama bivariada de Smith, Adelfang e Tubbs: soma, produto e proporção das variáveis componentes. *Revista Ciências Exatas e Naturais*, Guarapuava, v. 13, n. 2, p. 185-200, 2011.
- SAMPAIO, S. C.; LONGO, A. J.; QUEIROZ, M. M. F.; GOMES, B. M.; BOAS, M. A.; SUSZEK, M. Estimativa e distribuição da precipitação mensal provável no Estado do Paraná. *Irriga*, Botucatu, v. 12, p. 38-53, 2007.
- SHAKIL, M.; KRIBIA, B. M. G. Exact distributions of the linear combination of gamma and Rayleigh random variables. *Austrian Journal of Statistics*, Sidney, v. 38, p. 33-44, 2009.
- SILVA, J. C.; HELDWEIN, A.; MARTINS, F. B.; TRENTIN, G.; GRIMM, E. L. Análise de distribuição de chuva para Santa Maria, RS. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 11, n. 1, p. 67-72, 2007.
- VIEIRA, S. R.; CARVALHO, J. R. P. Estudo da periodicidade temporal de chuvas em uma bacia hidrográfica dos Rios Turvo/Grande - uma proposta. Campinas: Embrapa Informática Agropecuária, 2001. v. 10.

Recebido em 31 Janeiro 2012- Received on January 9, 2012.
Aceito em 13 Dezembro, 2012 - Accepted on December 13, 2012.