Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Distribuições de valores extremos aplicadas na análise sensorial de qualidade de cafés especiais.

Haiany Aparecida Ferreira[1]
Gilberto Rodrigues Liska[2]
Marcelo Ângelo Cirillo[3]
Flávio Meira Borém[4]
Diego Egídio Ribeiro[5]
Ricardo Miguel Cortez[6]

1. Introdução

Para realizar a análise sensorial de cafés é preciso que o café passe por um determinado processo: torração, concentração padrão, preparação para degustação e adição de água. Feito isso, após a degustação é feita a avaliação da amostra em questão ([3]).

Os resultados da avaliação sensorial foram estabelecidos a partir de uma escala que variava de 0 a 10 em que esses valores representam os níveis crescentes de qualidade do café. De acordo com o protocolo de análises sensoriais ([3]), os resultados da avaliação sensorial variam de acordo com uma escala onde as notas 6, 7, 8, 9 correspondem respectivamente a: bom, muito bom, excelente e excepcional. Já quando as notas são menores do que 6 os cafés são declarados com a qualidade abaixo do Grau Specialty, ou seja, o café seja declarado não especial não atendendo aos requisitos conforme as verificações que serão realizadas ([6]).

Sendo a atribuição de notas sensoriais máximas um processo aleatório, no sentido de existir variações no julgamento de diferentes provadores, uma abordagem adequada para esse tipo de estudo se faz necessária.

A Teoria de Valores Extremos (TVE) foi exposta pela primeira vez por Fisher e Tippet (1928), na qual considera três tipos de distribuições de probabilidade, a Gumbel, Fréchet e Weibull. Gumbel foi o primeiro a estudar e formalizar a aplicação estatística, sendo que em 1958 apresentou a distribuição de dupla exponencial, onde a sua função de probabilidade foi definida por meio dos valores de máximos e mínimos([1]).Em 1955,Jenkinsonpropôs que os três tipos de distribuições de valores extremos, ou seja, Gumbel, Fréchet e Weibull poderiam ser representados numa forma paramétrica única, designada por distribuição generalizada de valores extremos, GEV([2]).

---

[1] Graduanda em física na UFLA; e-mail: haianyferreira@yahoo.com.br
[2] Doutorando em Estatística e Experimentação Agropecuária no DEX/UFLA; e-mail: gilbertoliska@hotmail.com
[3] Professor Adjunto IV,o DEX/UFLA; e-mail: macufla @ dex.ufla.br
[4] Professor Associado IV, DEN/UFLA, e-mail: flavioborem@deg.ufla.br
[5] Doutorando em Engenharia Agrícola, UFLA, ; e-mail: diegoagro@hotmail.com
[6] Graduando em Engenharia Agrícola, UFLA, e-mail: ricardocortez@hotmail.com
Agradecimentos ao CNPq/PIBIC, CAPES, FAPEMIG

Nesse sentido e tendo em vista a importância do conhecimento da probabilidade de ocorrência de determinadas notas sensoriais, o presente trabalho consiste no estudo das notas sensoriais de quatro cafés de alta qualidade via teoria de valores extremos, com vista a fornecer a melhor distribuição que se adéqua aos dados e, considerando que notas sensoriais altas constituem eventos raros, determinar as probabilidades de ocorrência de altas notas nos quatro tipos de cafés analisados.

2. Material e métodos

Este experimento obteve seus dados das notas atribuídas para cada tipo de café nos testes realizados na Universidade Federal de Lavras. Foram avaliados quatro atributos em quatro diferentes tipos de café, de forma que a avaliação foi feita por determinado grupo de pessoas em diferentes momentos.

Foi considerado o cenário que avaliou a maior nota atribuída por um provador para cada tipo de café, ou seja, não se considerou o atributo de cada café e o momento em que o teste foi aplicado. Sendo assim, a base de dados é composta de 696 notas sensoriais.

Serão consideradas as distribuições de probabilidade Normal, GEV eGumbel. Um dos motivos para se utilizar essas distribuições de probabilidade, para o caso da normal, seria verificar se a distribuição das notas tem comportamento simétrico, na GEV ocorreria justamente o contrário, ou seja, observar se ela tem um comportamento assimétrico e com relação àGumbel, o fato dela ser muito utilizada para se trabalhar com valores máximos([3],[5]).

A função de distribuição de probabilidade da normal, ou seja, sua f.d.p. é dada por:

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \tag{1}$$

em que $-\infty<x<\infty$, μ é o parâmetro de posição, σ o parâmetro de escala e $\sigma>0$ .Af.d.p. da GEV é dada por:

$$f(x;\mu,\sigma,\xi,)=\frac{1}{\sigma}\left\{\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\left(-\frac{1}{\xi}\right)-1}\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)^{-\left(\frac{1}{\xi}\right)}\right]\right\}\right\}, \tag{2}$$

em que $-\infty<x<\mu=\sigma/\xi$ para $\xi<0$, $\mu-\sigma/\xi<x<\infty$ para $\xi>0$ e (μ,σ,ξ) são os parâmetros posição, escala e forma respectivamente e $\sigma>0$. Fazendo $\lim_{\xi\to 0} f(x;\mu,\sigma,\xi)$ obtém-se a f.d.p. Gumbel, que é dada por:

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, \tag{3}$$

em que $-\infty < x < \infty, -\infty < \mu < \infty$ e $\sigma > 0$.

Os parâmetros das distribuições (1), (2) e (3) foram estimados pelo método da máxima verossimilhança.

Uma vez estimados os parâmetros das distribuições de probabilidades podemos prosseguir ao cálculo de probabilidade. Considerando como *X* a variável aleatória associada às notas sensoriais dos provadores num determinado cenário, a probabilidade de ocorrência de uma nota sensorial ser superior a um valor *x* é dada pelo complementar da função de distribuição acumulada (f.d.a.) das distribuições Normal, Gumbele GEV. A f.d.a da distribuição Normal é dada por

$$F(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{1}{2}\left[\frac{t-\mu}{\sigma}\right]^2\right)dt = \Phi_{\mu,\sigma}\left(x\right) \tag{4}$$

onde $\Phi_{\mu,\sigma}\left(x\right)$ é a notação para a f.d.a. da Normal, uma vez que não existe uma expressão fechada para sua f.d.a.. Af.d.a. da GEV e dada por:

$$F(x;\mu,\sigma,\xi) = \exp\left[-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}\right]. \tag{5}$$

Para a distribuição Gumbel temos que:

$$F(x;\mu,\sigma) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}. \tag{6}$$

Logo, para calcularmos as probabilidades de interesse, basta calcular $P[X > x] = 1 - P[X \leq x] = 1 - F\left(x;\hat{\theta}\right)$,em que $\hat{\theta}$ é um valor ou vetor das estimativas de máxima verossimilhança da f.d.a. da Normal, Gumbel ou GEV.

Para verificar o ajuste das distribuições aos dados foi utilizado o teste de KolmogorovSmirnov (KS) e para verificar a pressuposição de independência das observações para o método da máxima verossimilhança foi utilizado o teste de Ljung-Box. Em ambos os testes foi considerado 5% como nível de significância.

Para realizar o ajuste das distribuições, testes de hipóteses e o cálculo de probabilidades, foi utilizado o Sistema Computacional R ([4]).

3. Resultados e discussão

Os resultados dos testes de KS e Ljung-Box (Q) são apresentados na tabela 1. Considerando-se 5% como nível de significância, observa-se que as distribuições Normal e

Gumbel não se ajustaram aos dados de notas sensoriais máximas, uma vez que o valor p do teste para cada distribuição em cada tipo de café é menor que o nível nominal de significância.

Tabela 1: Estimativas dos parâmetros e resultados dos testes de Kolmogorov-Smirnov (KS) e Ljung-Box (Q) para notas dos provadores do café,

Café	Distribuição	Parâmetros					KS (valor-p)	Q (valor-p)
		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$		
A	Normal	7,0565	1,8784	-	-	-	0,0181	0,3316
	GEV	-	-	6,7754	2,0174	0,6827	0,0540	0,3316
	Gumbel	-	-	6,0261	2,3469	-	0,0209	0,3316
B	Normal	6,7196	2,2707	-	-	-	0,0104	0,0675
	GEV	-	-	6,3719	2,5343	0,6776	0,8539	0,0675
	Gumbel	-	-	5,5149	2,5032	-	0,0005	0,0675
C	Normal	6,4362	2,3959	-	-	-	0,0007	0,0795
	GEV	-	-	6,0101	2,6011	0,6284	0,1975	0,0795
	Gumbel	-	-	5,1373	2,7432	-	0,0074	0,0795
D	Normal	7,4230	2,04799	-	-	-	0,0019	0,0910
	GEV	-	-	7,2541	2,2821	0,8202	0,8905	0,0910
	Gumbel	-	-	6,3089	2,3984	-	0,0114	0,0910

Por outro lado, a distribuição GEV se ajustou aos mesmos dados, indicando que essa distribuição é mais adequada para representar a distribuição das notas sensoriais máximas dos cafés analisados. Com relação ao teste de Ljung-Box, as notas sensoriais máximas podem ser consideradas independentes, uma vez que a hipótese nula do teste foi aceita (*valor-p*> 0,05) para todos os tipos de cafés.

Assim, uma vez estimados os parâmetros de uma distribuição e a mesma se ajustada aos dados, prosseguimos ao cálculo de probabilidade. Foram calculadas as probabilidades das notas de acordo com a distribuição GEV, que foi a única que se ajustou, serem maiores que 5; 6; 7; 8; 9 e 9,5 pontos. Os resultados são apresentados na tabela 2.

Tabela 2: Probabilidades de ocorrência de notas dos cafés obtidas pela distribuição GEV.

Café	Notas					
	5	**6**	**7**	**8**	**9**	**9,5**
A	0,8636	0,7551	0,5896	0,3666	0,1211	0,0236
B	0,7952	0,6834	0,5334	0,3496	0,1537	0,0668
C	0,7572	0,6335	0,4765	0,2970	0,1221	0,0511
D	0,8727	0,7928	0,6712	0,4952	0,2592	0,1257

De acordo com a tabela 2 observa-se que a probabilidade de ocorrência de uma nota alta, ou seja, acima de 9,5 é relativamente baixa quando comparada a probabilidade de uma nota ser maior que 5 para todos os tipos de cafés. Uma interpretação prática das probabilidades na tabela 2 pode ser feita da seguinte forma: a probabilidade de um provador dar uma nota maior do que 9 pontos para o café A ou o café A ser considerado como excepcional por um provador é de 12,11% pela distribuição GEV. O mesmo evento ocorre com maior probabilidade para o café D, ou seja, 25,92%.

4. Conclusões

Dentre as distribuições utilizadas a única que se ajustou foi a GEV já que esta esteve dentro do nível de significância basicamente em todos os momentos. Quanto às probabilidades foi possível verificar de um modo geral que a probabilidade do café D receber notas altas (acima de 9,5) é a maior dos cafés estudados, e que a probabilidade do café A receber notas altas (acima de 9,5) é a menor dentre os cafés estudados.

5. Referências

[1] BEIJO, L. A.; AVELAR, F. G. **Distribuição Generalizada de Valores Extremos no estudo de dados climáticos: uma breve revisão e aplicação**. Revista da Estatística UFOP, v. 1, X Semana da Matemática e II Semana da Estatística, 2010.

[2] LISKA, G.; BORTOLINI, J; SÁFADI, T.; BEIJO, L. **A. Estimativas de velocidade máxima de vento em Piracicaba-SPvia Séries Temporais e Teoria de Valores Extremos**, Revista Brasileira de Biometria, São Paulo, v.31, n.2, p.295-309, 2013.

[3] **PROTOCOLO PARA ANALISE SENSORIAL DE CAFÉ**. Rev.September 2008. Specialty Coffee Association of America. Disponível em: http://www.cafedocerrado.com.br/intranet/docs/SCAA_CuppingProtocols_Port.pdf. Acesso em: 20 Jan. 2014.

[4]R DEVELOPMENT CORE TEAM (2013) R: A language and environment for statistical computing. **R Foundation for Statistical Computing,** Vienna, Austria.ISBN 3-900051-07-0, URL http://www.R-project.org/.

[5] SANSIGOLO, C. **A. Distribuições de Extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006)**. Revista Brasileira de Meteorologia, v.23, n.3, p.341-346, 2008.

[6]**METODOLOGIA SCAA DE AVALIAÇÃO DE CAFÉS ESPECIAIS GUIA RÁPIDO–GREEN COFFEE.** Rev. Março 2009. Disponível em:http://coffeetraveler.net/wp-content/files/903-SCAACuppingMethod_RESUMO_3a.pdf. . Acesso em: 21 Jan. 2014.