

## ANÁLISE CONJUNTA BASEADA EM NOTAS VIA MODELO DE REGRESSÃO BETA

Vanessa Silva RESENDE<sup>1</sup>  
Eduardo Yoshio NAKANO<sup>1</sup>

- RESUMO: O modelo tradicional de análise conjunta supõe que as notas atribuídas seguem uma distribuição Normal. No entanto, a suposição de normalidade pode não ser razoável visto que as notas são limitadas. Como todas as notas limitadas podem ser padronizadas a valores entre 0 e 1, esse trabalho apresenta uma nova metodologia que permite modelar essas notas por meio de uma distribuição Beta. Ademais, esse trabalho apresenta um procedimento geral de cálculo da importância relativa que pode ser aplicado para qualquer outra distribuição adotada para modelar as notas. A metodologia apresentada foi ilustrada em uma aplicação numérica referente à produção de um *smartphone* e em dados simulados. O modelo de regressão Beta se mostrou uma técnica mais precisa para o ajuste dos dados na análise conjunta baseada em notas.
- PALAVRAS-CHAVE: Análise de preferência; importância relativa; função de utilidade.

### 1 Introdução

A Análise Conjunta, também chamada de Análise de Preferência (ARTES, 1991), é uma técnica estatística multivariada utilizada para descrever como os respondentes desenvolvem preferências por quaisquer tipos de objetos, tais como produtos, ideias ou serviços (HAIR, 2005). Este método tem como base a premissa de que os consumidores avaliam o valor de um objeto fazendo uma combinação das quantias separadas de valor fornecidas por cada atributo. Ou seja, diversas vezes as estimativas das preferências dos consumidores são melhores quando julgadas por uma combinação de atributos.

A análise conjunta surge como mais uma técnica para enriquecer os estudos de *marketing* e para auxiliar os profissionais da área a entender a estrutura de preferência do consumidor. Ao identificar os atributos de maior influência no processo de decisão de compra do consumidor, esta técnica auxilia no momento de se elaborar um produto, analisar os produtos já existentes no mercado a fim de aperfeiçoá-los, saber as vantagens e desvantagens em relação a empresas concorrentes, a fim de torná-los mais competitivos no mercado.

A análise conjunta possui três tipos de modelos: o modelo baseado em postos, modelo baseado em escolha e o modelo baseado em notas, sendo este último o enfoque do trabalho. As diferenças entre esses modelos são as formas como os respondentes avaliam

---

<sup>1</sup> Universidade de Brasília - UnB, Departamento de Estatística, CEP: 70910-900, Brasília, DF, Brasil. E-mail: [vanessinha-df@hotmail.com](mailto:vanessinha-df@hotmail.com); [nakano@unb.br](mailto:nakano@unb.br)

os produtos. No modelo baseado em postos, o respondente ordena os produtos apresentados de acordo com sua preferência, no modelo baseado em escolha o respondente escolhe o produto de sua maior preferência, e por fim, no modelo baseado em notas os respondentes atribuem uma nota para cada produto apresentado.

Os modelos usuais baseados em notas consideram que as notas atribuídas pelos respondentes seguem uma distribuição Normal, utilizando-se das técnicas de regressão linear múltipla (NETER et al., 1996) para a obtenção das estimativas dos parâmetros do modelo. No entanto, nem sempre a suposição de normalidade é razoável. Como o valor da nota é sempre limitado em um intervalo, o mesmo pode ser padronizado de forma que assumam valores entre 0 e 1. Neste contexto, este trabalho propõe a modelagem das notas considerando que as mesmas seguem uma distribuição Beta, tendo seus coeficientes estimados por meio de um modelo de regressão Beta (FERRARI et al., 2004; PAOLINO, 2001). Outras aplicações do modelo de regressão beta podem ser vistas em Silva e Souza (2014) e Cribari-Neto e Pereira (2013). Foi também proposto um procedimento geral de cálculo da importância relativa dos atributos dos produtos. Esse procedimento pode ser aplicado em qualquer situação, independente da distribuição adotada para modelar as notas.

A metodologia apresentada foi ilustrada em conjunto de dados fictícios referentes à produção de um *smartphone* que podem ser encontrados no trabalho de Matos (2011) e também em dados simulados. Os resultados obtidos pelo modelo de regressão Beta foram comparados com aqueles obtidos supondo-se a normalidade das notas. Todas as simulações e estimativas foram realizadas pelo *software* livre R (R CORE TEAM, 2013).

## 2 Métodos

### 2.1 Modelo baseado em notas usual

O modelo de Análise Conjunta baseado em notas supõe que a variável resposta segue uma distribuição Normal (ARTES, 1991; DA-SILVA, 2010; KOTRI, 2006). Desta maneira, utiliza-se um modelo de regressão múltipla para a estimação de seus parâmetros e das utilidades parciais. Segundo Hair Jr (2005), a análise conjunta é uma técnica que pode ser aplicada tanto a um único indivíduo como a um grupo e pode fornecer uma perspectiva não só para as preferências por um atributo, mas também para a quantia do mesmo (níveis). O caso em que as estimativas são feitas considerando-se que apenas um indivíduo faz a avaliação dos produtos corresponde ao caso desagregado. Nesta situação os resultados do modelo são analisados separadamente para este indivíduo. Já o nível agregado refere-se a situação em que as estimativas são feitas considerando-se um grupo de indivíduos (MATOS, 2011). A análise desagregada permite que a precisão preditiva seja avaliada para cada pessoa, em vez de ser calculada para a amostra total. Já a análise agregada apresenta maior eficiência estatística, por utilizar mais observações no processo de estimação.

Os produtos (também chamados de tratamentos) são os resultados de uma combinação dos níveis de cada um dos atributos. Portanto, no caso de  $r$  atributos com  $m_s$  níveis cada tem-se um total de  $J = \prod_{s=1}^r m_s$  produtos. A utilidade total atribuída ao  $j$ -ésimo produto pelo  $n$ -ésimo respondente é dada por:

$$U_{nj} = \beta_0 + \sum_{s=1}^r \sum_{i=1}^{m_s} X_{si}^j \beta_{si} + \varepsilon_{nj} \quad (1)$$

em que,  $U_{nj}$  é a utilidade total, ou seja, é a nota dada pelo  $n$ -ésimo respondente para o  $j$ -ésimo produto,  $X_{si}^j$  é a variável binária que indica se um determinado nível de seu respectivo atributo encontra-se no modelo ( $X_{si}^j = 1$ , se o  $i$ -ésimo nível do  $s$ -ésimo atributo estiver presente no  $j$ -ésimo produto;  $X_{si}^j = 0$ , caso contrário),  $\beta_0$  é uma constante correspondente à nota média atribuída pelos respondentes para todos os produtos,  $\beta_{si}$  é a utilidade parcial associada ao  $i$ -ésimo nível do  $s$ -ésimo atributo e;  $\varepsilon_{nj}$  é o erro aleatório não observável no modelo. Em geral, assume-se que  $\varepsilon_{nj} \sim N(0, \sigma^2)$ .

Note que (1) representa um modelo de regressão múltipla para cada respondente. Assim, as estimativas parciais do modelo podem ser facilmente obtidas por meio do Método de Mínimos Quadrados Ordinários (GRAYBILL e IYER, 1994). Neste trabalho essas estimativas foram obtidas pelo procedimento *lm* do *software* R (R CORE TEAM, 2013).

Após obter as estimativas das utilidades parciais pode-se estimar a importância relativa de cada atributo na estrutura de preferência do consumidor. Ou seja, a importância relativa pode ser interpretada como quanto um  $j$ -ésimo atributo impacta na preferência do consumidor por aquele determinado produto. Note que, se  $\beta_{si} < 0$ , há indícios de efeito desfavorável do  $i$ -ésimo nível do  $s$ -ésimo atributo na preferência do respondente, ou seja, que este item diminui a nota de preferência do produto e; se  $\beta_{si} > 0$ , há indícios de efeito favorável na preferência do respondente, ou seja, que este item aumenta a nota de preferência do produto. Assim, a importância relativa (em %) do  $s$ -ésimo atributo para o modelo (1) é dada por:

$$IR_s = \frac{I_s}{\sum_{u=1}^r I_u} \times 100\% , \quad (2)$$

em que  $I_s = \max_i \{\beta_{si}\} - \min_i \{\beta_{si}\}$  é a importância do  $s$ -ésimo atributo,  $s=1, 2, \dots, r$ .

## 2.2 Modelo baseado em notas com distribuição Beta

Como visto anteriormente, o modelo (1) supõe que a nota dos produtos segue uma distribuição Normal. No entanto, nem sempre a suposição de normalidade pode ser razoável para representar as notas dos respondentes. Visto que o respondente sempre representará sua preferência de um produto por meio de uma nota limitada em um intervalo (entre zero e dez, por exemplo), a mesma poderá ser padronizada de forma que assumam valores entre 0 e 1. Neste contexto, propõe-se um modelo de Análise Conjunta para o ajuste destas notas padronizadas admitindo-se que elas têm distribuição Beta. A distribuição Beta pode ser utilizada para modelar variáveis aleatórias contínuas delimitadas no intervalo (0,1), como por exemplo, taxas e proporções. Uma das vantagens

da utilização da distribuição Beta é o fato da mesma ser bastante flexível, podendo a função densidade de probabilidades apresentar diferentes formas.

Ferrari e Cribari-Neto (2004) propuseram um modelo de regressão no qual se propõe que a variável resposta segue uma distribuição Beta. Este modelo de regressão utiliza uma reparametrização do modelo Beta que considera como parâmetros a média,  $0 < \mu < 1$ , e um parâmetro de precisão,  $\phi > 0$ . A função densidade da distribuição Beta com essa parametrização alternativa é dada por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 \leq y \leq 1 \quad (3)$$

que resulta em  $E(Y) = \mu$  e  $\text{Var}(Y) = \mu(1-\mu)/(1+\phi)$ .

Assim, o modelo de regressão Beta possui como componente sistemático:

$$g(\mu_j) = \beta_0 + \sum_{s=1}^r \sum_{i=1}^{m_s} X_{si}^j \beta_{si}, \quad (4)$$

em que  $g(\cdot)$  é a função que faz a ligação entre o componente sistemático e a média do modelo Beta. Há algumas funções de ligação que podem ser utilizadas, como por exemplo, a função de ligação logito, probito, e a função de ligação complemento log-log. Mais detalhes sobre essas funções de ligações podem ser encontrados em (McCULLAGH e NELDER, 1991). Para este trabalho será utilizada a função de ligação logito, cuja função é estritamente crescente em (0,1) e contínua, bem como as outras funções de ligação citadas. A função logito é dada por:

$$g(\mu_j) = \log\left(\frac{\mu_j}{1-\mu_j}\right). \quad (5)$$

Assim, de (4) e (5) tem-se que a utilidade total atribuída ao  $j$ -ésimo produto pelo  $n$ -ésimo respondente é dada por:

$$U_{nj} = \frac{\exp\left\{\beta_0 + \sum_{s=1}^r \sum_{i=1}^{m_s} X_{si}^j \beta_{si}\right\}}{1 + \exp\left\{\beta_0 + \sum_{s=1}^r \sum_{i=1}^{m_s} X_{si}^j \beta_{si}\right\}} + \varepsilon_{nj}, \quad (6)$$

em que,  $U_{nj}$  é a utilidade total, ou seja, é a nota dada pelo  $n$ -ésimo respondente para o  $j$ -ésimo produto,  $X_{si}^j$  é a variável binária que indica se um determinado nível de seu respectivo atributo encontra-se no modelo ( $X_{si}^j = 1$ , se o  $i$ -ésimo nível do  $s$ -ésimo atributo estiver presente no  $j$ -ésimo produto;  $X_{si}^j = 0$ , caso contrário),  $\beta_0$  e  $\beta_{si}$  são os coeficientes obtidos pela regressão Beta associados ao  $i$ -ésimo nível do  $s$ -ésimo atributo e  $\varepsilon_{nj}$  é o erro aleatório não observável no modelo.

Cabe lembrar que as estimativas são feitas considerando-se o nível desagregado e, portanto cada respondente,  $n$ , é modelado separadamente (MATOS, 2011). Assim, a função de verossimilhança para o  $n$ -ésimo respondente é dada por:

$$L_n(\mu_{nj}; \phi_n) = \prod_{j=1}^J \frac{\Gamma(\phi_n)}{\Gamma(\mu_{nj}\phi_n)\Gamma((1-\mu_{nj})\phi_n)} u_{nj}^{\mu_{nj}\phi_n-1} (1-u_{nj})^{(1-\mu_{nj})\phi_n-1}, \quad (7)$$

em que

$$\mu_{nj} = \frac{\exp\left\{\beta_{n0} + \sum_{s=1}^r \sum_{i=1}^{m_s} x_{nsi}^j \beta_{nsi}\right\}}{1 + \exp\left\{\beta_{n0} + \sum_{s=1}^r \sum_{i=1}^{m_s} x_{nsi}^j \beta_{nsi}\right\}}.$$

Note que os estimadores de máxima verossimilhança (EMV) dos parâmetros não podem ser obtidos analiticamente pela maximização direta da função de verossimilhança (7). No entanto, os mesmos podem ser obtidos numericamente através de pacotes estatísticos existentes. Neste trabalho, os EMVs dos parâmetros foram obtidos através do pacote *betareg* (CRIBARI-NETO e ZEILEIS, 2010) do *software* R (R CORE TEAM, 2013).

Ao contrário do modelo Normal, as utilidades parciais do modelo de regressão Beta não apresentam forma fechada, impossibilitando estimar a importância relativa de cada atributo por meio dos métodos usuais.

Neste contexto, este trabalho propõe um procedimento alternativo para se estimar as importâncias relativas dos atributos. Primeiramente calcula-se a variação média da utilidade (6) quando  $X_{si}$  passa de 0 para 1. Ou seja, essa a variação média é dada por:

$$VM_s = \left| \frac{\sum_{j: X_{si}^j=0} U_{nj}}{\sum_{j: X_{si}^j=0} 1} - \frac{\sum_{j: X_{si}^j=1} U_{nj}}{\sum_{j: X_{si}^j=1} 1} \right|, \quad s = 1, 2, 3, \dots, r \quad (8)$$

em que,

$\sum_{j: X_{si}^j=0} U_{nj}$  é a soma das utilidades com  $X_{si}^j = 0$ ;

$\sum_{j: X_{si}^j=1} U_{nj}$  é a soma das utilidades com  $X_{si}^j = 1$ ;

$\sum_{j: X_{si}^j=0} 1$  é a frequência de observações com  $X_{si}^j = 0$  e;

$\sum_{j: X_{si}^j=1} 1$  é a frequência de observações com  $X_{si}^j = 1$ .

Assim, a importância relativa (em %) do  $s$ -ésimo atributo para o modelo é dada por:

$$IR_s^* = \frac{VM_s}{\sum_{u=1}^r VM_u}, \quad (9)$$

em que  $VM_s$  é a variação média do  $s$ -ésimo atributo dado por (7),  $s = 1, 2, 3, \dots, r$ .

Note que, ao contrário de (2), a importância relativa definida em (9) pode ser utilizada para qualquer que seja a distribuição ou a função de ligação utilizada para modelagem das notas. A importância relativa (2) pode ser considerada um caso particular de (9). De fato, na suposição de normalidade das notas, (9) é equivalente à (2). Note também que a variação média (8) representa também a importância do  $s$ -ésimo atributo.

### 3 Ilustração numérica

A metodologia apresentada neste trabalho é ilustrada em dois exemplos numéricos e em uma pequena simulação. O primeiro exemplo considera uma análise a nível desagregado e no segundo exemplo, a análise é realizada agregando as notas de dois respondentes. De forma análoga, a análise agregada pode ser realizada para  $n$  respondentes.

Matos (2011) apresenta um exemplo da aplicação da Análise Conjunta baseada em notas sob a pressuposição de normalidade. O conjunto de dados é referente às notas atribuídas por um respondente a fim de se saber qual atributo (tipo de teclado, sistema operacional ou cor do aparelho) um consumidor julga ser mais importante ao comprar um *smartphone*. Imagina-se que uma empresa especializada em celulares pretende fabricar e lançar um novo produto (*smartphone*) no mercado. Mas para isso pretende-se decidir quais atributos apostar para se obter uma vantagem competitiva no mercado. Foram considerados três atributos para uma boa aceitação dos consumidores: teclado (*touch scream* = 1 ou *qwert* = 0), tipo de sistema operacional (*Android* = 1 ou *Windows* = 0), e cores (branca = 1 ou preta = 0). Um planejamento fatorial completo foi feito, e assim obtiveram-se oito diferentes produtos. Para efeito de ilustração, notas de um segundo respondente foram geradas para realizar a análise no nível agregado. A Tabela 1 apresenta as notas atribuídas pelos dois respondentes para os oito produtos analisados.

Tabela 1 - Notas dos dois respondentes para todos os tipos de *smartphone*

Teclado	Sistema Operacional	Cor	Nota*	
			Respondente 1	Respondente 2
<i>Touch screen</i>	<i>Android</i>	Branco	9,0	8,5
<i>Touch screen</i>	<i>Android</i>	Preto	8,0	7,5
<i>Touch screen</i>	<i>Windows</i>	Branco	7,5	8,2
<i>Touch screen</i>	<i>Windows</i>	Preto	6,5	6,0
<i>Qwert</i>	<i>Android</i>	Branco	7,0	5,9
<i>Qwert</i>	<i>Android</i>	Preto	6,0	6,5
<i>Qwert</i>	<i>Windows</i>	Branco	5,0	4,5
<i>Qwert</i>	<i>Windows</i>	Preto	4,0	3,5

\*Fonte: Matos (2011)

\*\* Notas geradas aleatoriamente pelo autor

### 3.1 Análise de dados a nível desagregado

Inicialmente a metodologia é aplicada no nível desagregado para o Respondente 1. A Tabela 2 apresenta os coeficientes estimados para o modelo baseado em notas com distribuição Normal (1) e Beta (2). Visto que as notas foram limitadas a valores entre 0 e 10, as mesmas devem ser padronizadas para o ajuste do modelo Beta. A padronização adotada aqui foi simplesmente dividir por dez os valores originais das notas.

Tabela 2 - Estimativas dos coeficientes dos modelos Normal e Beta para os dados do Respondente 1

Atributo*	Coeficientes	Estimativas dos coeficientes	
		Modelo Normal	Modelo Beta
	$\beta_0$	4,125	-0,467
Teclado	$\beta_1$	2,250	1,093
Sistema Operacional	$\beta_2$	1,750	0,861
Cor	$\beta_3$	1,000	0,498

\* Teclado *qwert*, sistema operacional *Windows* e cor preto são os níveis de referência dos atributos.

Estimados os coeficientes dos modelos, as utilidades estimadas podem ser obtidas a partir (1) e (6). A título de ilustração, as utilidades estimadas do modelo Beta foram multiplicadas por 10 de forma obter as notas em sua escala original. As notas estimadas para os modelos Normal e Beta encontram-se na Tabela 3 a seguir, bem como o erro cometido na estimação.

Tabela 3 - Estimativas das notas, erro (resíduo) das estimativas e soma de quadrados dos resíduos calculados pelos modelos Normal e Beta para os dados do Respondente 1

Produto			Nota estimada (erro)	
Teclado	Sistema Operacional	Cor	Normal	Beta
<i>Touch screen</i>	<i>Android</i>	Branco	9,125 (0,125)	8,792 (0,208)
<i>Touch screen</i>	<i>Android</i>	Preto	8,125 (0,125)	8,156 (0,156)
<i>Touch screen</i>	<i>Windows</i>	Branco	7,375 (0,125)	7,548 (0,048)
<i>Touch screen</i>	<i>Windows</i>	Preto	6,375 (0,125)	6,516 (0,016)
<i>Qwert</i>	<i>Android</i>	Branco	6,875 (0,125)	7,093 (0,093)
<i>Qwert</i>	<i>Android</i>	Preto	5,875 (0,125)	5,973 (0,027)
<i>Qwert</i>	<i>Windows</i>	Branco	5,125 (0,125)	5,078 (0,078)
<i>Qwert</i>	<i>Windows</i>	Preto	4,125 (0,125)	3,854 (0,146)
Soma de Quadrados Resíduos			0,125	0,107

Pode-se notar que para este respondente o produto de sua maior preferência é o aparelho celular com teclado *touch*, com sistema operacional *Android*, e cor branca. Neste exemplo ambos os modelos, Normal e Beta, apresentaram boas estimativas para as notas,

tendo o modelo Beta apresentado menor Soma de Quadrados Resíduos – SQRes (soma do quadrado das diferenças entre o a nota observada e estimada pelo modelo) e portanto apresentando um resultado sensivelmente melhor do que o modelo Normal.

A Tabela 4 apresenta a importância bruta e relativa, obtidas por (8) e (9), de cada atributo considerando o modelo Normal e Beta.

Tabela 4 - Importância bruta e relativa dos atributos para os dados do Respondente 1

Atributo	Modelo Normal		Modelo Beta	
	Importância	Importância Relativa	Importância	Importância Relativa
Teclado	2,25	45%	2,253	44,98%
Sistema Operacional	1,75	35%	1,754	35,01%
Cor	1,00	20%	1,003	20,01%

O Teclado foi o atributo mais importante, cuja importância relativa foi estimada em 45% pelo modelo Normal e 44,98% pelo modelo Beta. Por outro lado, a cor do aparelho foi o atributo de menor importância, tendo a importância relativa estimada em 20% pelo modelo Normal e 20,01% pelo modelo beta. Resultados similares podem ser observados para a análise desagregada para o Respondente 2.

### 3.2 Análise de dados a nível agregado

Nesta seção, a metodologia é aplicada agregando as notas dos dois respondentes (Tabela 1). Após padronização das notas para o ajuste do modelo Beta, os modelos Normal (1) e Beta (2) foram ajustados e as estimativas de seus coeficientes são apresentadas na Tabela 5.

Tabela 5 - Estimativas dos coeficientes dos modelos Normal e Beta para os dados agregados dos dois respondentes

Atributo*	Coeficientes	Estimativas dos coeficientes	
		Modelo Normal	Modelo Beta
	$\beta_0$	4,000	-0,498
Teclado	$\beta_1$	2,350	1,106
Sistema Operacional	$\beta_2$	1,650	0,789
Cor	$\beta_3$	0,950	0,464

\* Teclado *qwert*, sistema operacional *Windows* e cor preto são os níveis de referência dos atributos.

As notas estimadas para os modelos Normal e Beta na análise agregada são apresentadas na Tabela 6. De forma similar à análise desagregada, o produto de maior preferência é o aparelho celular com teclado *touch*, com sistema operacional *Android*, e cor branca. Ambos os modelos apresentaram boas estimativas para as notas, tendo o modelo Beta apresentado novamente um resultado sensivelmente melhor (menor SQRes) do que o modelo Normal.

Tabela 6 - Estimativas das notas e soma de quadrados dos resíduos calculados pelos modelos Normal e Beta para os dados agregados dos dois respondentes

Teclado	Produto		Nota estimada	
	Sistema Operacional	Cor	Normal	Beta
<i>Touch screen</i>	<i>Android</i>	Branco	8,950	8,655
<i>Touch screen</i>	<i>Android</i>	Preto	8,000	8,018
<i>Touch screen</i>	<i>Windows</i>	Branco	7,300	7,451
<i>Touch screen</i>	<i>Windows</i>	Preto	6,350	6,476
<i>Qwert</i>	<i>Android</i>	Branco	6,600	6,804
<i>Qwert</i>	<i>Android</i>	Preto	5,650	5,724
<i>Qwert</i>	<i>Windows</i>	Branco	4,950	4,916
<i>Qwert</i>	<i>Windows</i>	Preto	4,000	3,781
Soma de Quadrados Resíduos			3,400	3,044

A Tabela 7 apresenta a importância bruta e relativa, obtidas por (8) e (9), de cada atributo considerando o modelo Normal e Beta para os dados agregados. Os resultados obtidos pelos dois modelos foram similares. O Teclado foi o atributo mais importante, cuja importância relativa foi estimada em 47,48% pelo modelo Normal e 47,41% pelo modelo Beta. A cor do aparelho foi o atributo de menor importância, tendo a importância relativa estimada em 19,19% pelo modelo Normal e 19,34% pelo modelo Beta.

Tabela 7 - Importância bruta e relativa dos atributos para os dados agregados dos dois respondentes

Atributo	Modelo Normal		Modelo Beta	
	Importância	Importância Relativa	Importância	Importância Relativa
Teclado	2,35	47,48%	2,344	47,41%
Sistema Operacional	1,65	33,33%	1,644	33,25%
Cor	0,95	19,19%	0,956	19,34%

A semelhança entre os resultados obtidos para modelos Normal e Beta apresentados na ilustração foi possivelmente devido à simetria das notas observadas (note que os resíduos das estimativas do modelo Normal para a análise desagregada são constantes). Neste contexto, é apresentado na próxima seção o desempenho dos modelos Normal e Beta em dados assimétricos gerados por meio da distribuição Beta.

### 3.3 Estudo de simulação

Esta seção apresenta o ajuste dos modelos Normal e Beta considerando dados simulados. A amostra considerada consistiu de um total de 100 respondentes. Cada nota

foi gerada pela distribuição Beta descrita em (3) com média dada pela nota do Respondente 1 em cada um dos 8 produtos apresentados pela Tabela 1 e parâmetro de precisão  $\phi = 1025$ . Desta forma, a amostra foi gerada através de uma distribuição Beta( $a_j, b_j$ ), com  $a_j = \phi\mu_j = 1025\mu_j$  e  $b_j = \phi(1 - \mu_j) = 1025(1 - \mu_j)$ , com  $\mu_j$  igual a nota padronizada do  $j$ -ésimo produto apresentado pelo Respondente 1,  $j=1,2,\dots,8$ . O valor do parâmetro de precisão foi escolhido de forma a aumentar a variância da distribuição Beta e também para diferenciar seus parâmetros e torná-la assimétrica. O ajuste dos modelos Beta e Normal foi realizado no nível agregado considerando, e como critério de comparação utilizou-se o critério de Akaike – AIC (AKAIKE, 1974) e a soma de quadrados resíduos. Os resultados são apresentados na Tabela 8.

Tabela 8 - Comparação dos modelos com base nas simulações das notas de 100 respondentes

	Modelo Normal	Modelo Beta
AIC	-3.704,940	-4.625,465
SQRes	45,071	14,723

Observa-se pela Tabela 8 um melhor ajuste do modelo Beta para os dados simulados (menor AIC). A soma de quadrados resíduos foi menor ao se utilizar o novo modelo proposto, ou seja, o ajuste do modelo baseado em notas em que se utiliza a regressão Beta apresentou um resultado mais preciso do que aquele que utiliza a regressão Normal.

## Conclusões

A análise conjunta é uma técnica estatística que pode ser utilizada para entender a estrutura de preferência de um consumidor por um determinado produto. As estimativas das preferências dos consumidores são melhores, em geral, quando se avalia um conjunto de atributos. Com essa técnica multivariada pode-se identificar os atributos de maior influência sobre a preferência do consumidor. O modelo baseado em notas é um dos principais modelos da análise conjunta. Este consiste em um modelo no qual os respondentes atribuem uma nota para cada produto apresentado, ou seja, para cada conjunto de atributos e níveis estabelecidos pelo pesquisador. A partir destas notas atribuídas é possível estimar as utilidades parciais, utilidades totais e a importância relativa. Tradicionalmente, estas estimativas são feitas através de modelos de regressão múltipla, supondo que as notas dadas seguem distribuição Normal. Este trabalho propôs a modelagem das notas por meio de uma distribuição Beta, obtendo as estimativas a partir de uma regressão Beta. Estas estimativas podem ser obtidas facilmente por meio do pacote *betareg* do *software* R.

Os dois modelos abordados neste trabalho foram aplicados a um conjunto de dados fictícios relativos à produção de um celular (*smartphone*). Foram considerados três atributos: teclado, sistema operacional e, cor; com dois níveis cada. Com base nos resultados obtidos, concluiu-se que o atributo de maior importância foi o teclado. Observa-se ainda que o modelo baseado em notas com distribuição Beta obteve resultados semelhantes aos do modelo baseado em notas com distribuição Normal, entretanto o modelo com notas com distribuição Beta foi sensivelmente melhor. A semelhança entre os

resultados dos dois modelos apresentados na ilustração foi, possivelmente, devido à simetria das notas, algo não percebido quando se gerou dados assimétricos através da distribuição Beta. Nesta última situação o modelo Normal apresentou um ajuste inferior.

Neste contexto, observa-se que o modelo de regressão Beta se mostrou uma técnica mais precisa para o ajuste dos dados na análise conjunta baseada em notas. Isto porque a suposição de normalidade das observações pode não ser razoável em algumas situações, visto que as notas são limitadas e nem sempre apresentam um comportamento simétrico. A estimação dos parâmetros do modelo Beta envolve maximização numérica e não é tão simples como no modelo Normal. Entretanto, no *software* R há a biblioteca *betareg* que fornece as estimativas dos parâmetros de forma simples, viabilizando o uso do modelo Beta na Análise Conjunta. A metodologia proposta neste trabalho para o cálculo da importância relativa é um procedimento intuitivo e computacionalmente simples, e pode ser utilizada qualquer que seja a distribuição utilizada para a modelagem das notas.

Portanto, para o modelo da análise conjunta baseado em notas há uma nova metodologia desenvolvida que pode ser aplicada quando se deseja, por exemplo, criar um novo produto ou aperfeiçoar um já existente de tal forma que este tenha vantagem competitiva no mercado. Pode-se considerar para futuros trabalhos o uso de outras distribuições para modelar as notas e/ou a generalização do cálculo da importância relativa para qualquer número de níveis de cada atributo.

## Agradecimentos

A primeira autora agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio concedido para a realização desse trabalho.

RESENDE, V. S.; NAKANO, E. Y. Score-based conjoint analysis via beta regression model. *Rev. Bras. Biom.*, São Paulo, v.33, n.1, p.51-62, 2015.

- *ABSTRACT: The traditional model of conjoint analysis assumes that attributed scores follow a Normal distribution. However, normality assumption may not be reasonable since the scores are limited. Since any limited scores can be standardized to values from 0 to 1, this paper presents a new methodology to allow modeling these scores by a Beta distribution. In addition, the paper presents a general procedure to calculate the relative importance, which can be applied to any other distribution. The presented methodology is illustrated on a numerical application of a smartphone production and on simulated data. The Beta regression model showed a more accurate technique for adjusting the traditional conjoint analysis models.*
- *KEYWORDS: Stated preference analysis; relative importance; utility functions.*

## Referências

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.2, n.6, p. 716-723, 1974

ARTES, R. *Análise de preferência (“conjoint analysis”)*. 1991. 189f. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), São Paulo, 1991.

- CRIBARI-NETO, F.; PEREIRA, T.L. Avaliação da eficiência de administrações municipais no Estado de São Paulo: Uma nova abordagem via modelos de regressão Beta. *Rev. Bras. Biom.*, São Paulo, v.31, n.2, p.270-294, 2013.
- CRIBARI-NETO, F.; ZEILEIS, A. Beta Regression in R. *Journal of Statistical Software*, v. 34, n.2, p. 1-24, 2010.
- DA SILVA, C. H.; BASTOS, F. S. Introdução à Conjoint Analysis. In: IX ENCONTRO MINEIRO DE ESTATÍSTICA, Viçosa–MG. Minicurso ministrado, 2010.
- FERRARI, S.L.P.; CIRBARI-NETO, F. Beta Regression For Rates and Proportions, *Journal of Applied Statistics*, v.31, n. 7, p. 799-815, 2004.
- GRAYBILL; F.A.; IYER, H.K. *Regression Analysis: Concepts and Applications*. 1.ed. Duxbury: Duxbury Press, 1994. 650p.
- HAIR JR, J. F. *Análise multivariada de dados*. 5.ed. Porto Alegre: Bookman, 2005. 688p.
- KOTRI, A. Analyzing customer value using Conjoint Analysis: the example of a packaging company. *Working Paper Series*, Faculty of Economics and Business Administration, University of Tartu (Estonia). n.46, p. 1-38, 2006.
- MATOS, T.O. *Conjoint Analysis: Uma Aplicação ao Marketing*. Monografia (Bacharelado em Estatística) – Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora (UFJF). Minas Gerais, 2011.
- McCULLAGH, P.; NELDER, J.A. *Generalized Linear Models*. 2.ed. Londres: Chapman & Hall, 1991. 532p.
- NETER, J.; KUTNER, M.; NACHTSHEIM, C.; WASSERMAN, W. *Applied Linear Statistical Models*, 4. ed. Estados Unidos: Irwin Series in Statistics, 1996. 1408p.
- PAOLINO, P. Maximum Likelihood Estimation of Models with Beta Distributed Dependent Variables. *Political Analysis*, v.9, n.4, p. 325-346, 2001.
- R 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>.
- SILVA, C.R.; SOUZA, T.C. Modelagem da taxa de analfabetismo no Estado da Paraíba via modelo de regressão Beta. *Rev. Bras. Biom.*, São Paulo, v.32, n.3, p.345-359, 2014.

Recebido em 07.11.2014

Aprovado após revisão em 10.01.2015