

Análise de correspondência como instrumento para descrição do perfil do vestibular 2008 da UESC: aspectos escolares e sociais.

Enio Jelihovschi, Professor adjunto do DCET/UESC, email: eniojelihovs@gmail.com

Marcelo Inácio Ferreira Ferraz, Professor adjunto do DCET/UESC, email: mfferraz@uesc.br

Fernanda Barbosa Silveira, Discente do Curso de Economia, email: fbsninha@hotmail.com

INTRODUÇÃO

Com a introdução no ano de 2008 do sistema de reserva de vagas (cotas), na UESC surgiram indagações de como seria o desempenho dos alunos cotistas em sua caminhada acadêmica. Porém, antes de avaliar o desempenho se faz importante o conhecimento do perfil dos vestibulandos.

Para essa caracterização optou-se pelo uso da Análise de Correspondência que é um método muito poderoso de visualização de dados categóricos organizados em tabela de contingência cujos resultados possibilitam a obtenção de informações com respeito a agrupamentos, relações, em resumo toda a estrutura da interrelação entre as variáveis em questão.

Segundo Greenacre, M, as origens teóricas desta técnica são encontradas há 50 anos, porém, seu ímpeto real para a moderna aplicação se deu pelo lingüista e analista de dados Jean-Paul Benzécri na Universidade de Rennes, França, em princípios da década de 60.

METODOLOGIA

A base de dados do estudo foi composta de dados secundários obtidos do questionário sócio-cultural do vestibular de 2008. Para efeito de análise das mais de 40 questões que compõem o questionário sócio-cultural do vestibular da UESC, foram consideradas apenas dez questões escolares e quatro sociais. Com base nas respostas a essas questões aplicou-se a análise de correspondência e para efeito de construção do gráfico relacionando os aspectos escolares e sociais, foram

considerados apenas os itens de respostas com massa igual ou superior a 10. Por massa entendemos a frequência total do item dividido pelo total de dados.

Como suporte computacional foi utilizado o software R mais especificamente o pacote `ca` e as funções `mjca` e `ca` para a construção dos gráficos.

RESULTADOS

A análise dos aspectos escolares (Tabela 1) revela que a maior parte dos inscritos, concluiu o ensino médio no ano de 2007, seguido do ano de 2006 e no período todo diurno. Realizaram na sua maioria ensino médio em escola particular na modalidade científico tendo frequentado curso pré-vestibular e estão prestando o primeiro vestibular. Os principais motivos por optar pela UESC foram respectivamente a credibilidade da instituição, melhor da milha opção e o ensino gratuito. A escolha do curso para a maior parte dos candidatos teve por base a afinidade pessoal e a vocação, esperando obter com o curso superior formação teórica profissional voltada para o futuro. Quanto à escolaridade do pai e da mãe, a maior parte possui o ensino médio completo.

A análise das informações sócio-culturais (Tabela 2), revelou que a maior parte dos vestibulandos lê de 3 a 5 livros por ano e o meio de informação que mais utilizam é a televisão, seguido da internet. A maior parte dos inscritos acesso à internet em casa.

Após essa análise descritiva dos dados aplicou-se a análise de correspondência. Contudo, para efeito de construção dentre os diversos itens de respostas das questões analisadas foram selecionados as respostas que tiveram uma massa mínima, maior ou igual a 10, assim, para as variáveis escolares foram considerados 25.

A primeira componente principal (PC1) para as informações escolares foi formada basicamente pela contribuição de dez itens, conforme a seguir: $q10C + q12C + q15C + q24B + q25B - q9A - q12D - q15B - q24G - q25G$. Já a segunda componente principal (PC2) pela contribuição de nove itens: $q9G + q15B + q18C + q20C + q22E - q9A - q15A - q18A - q20A$.

Tabela 1 – Descrição das características escolares dos vestibulandos do ano de 2008 da UESC, Ilhéus, BA – 2008.

Variável	Categoria	%	Análise de correspondência									
			Código	Massa			PC 1			PC2		
				mass	Qlt	inr	K=1	cor	ctr	K=1	Cor	Ctr
Ano em que você concluiu ou concluirá o ensino médio	2007	30,2	q9a	30	918	35	-135	808	31	-50	110	98
	2006	21,2	q9b	21	117	3	10	34	0	-16	83	7
	2005	13,0	q9c	13	458	5	60	450	3	8	7	1
	2004	7,9										
	2003	5,2										
	2002	4,1										
1991 a 2001 antes de 1991		14,4	q9g	15	719	14	99	518	8	61	201	71
		3,9										
Em que período você cursou o Ensino Médio?	Todo diurno	69,5	q10a	70	870	16	-62	854	15	9	16	7
	Maior parte diurno	14,0	q10b	14	401	7	57	331	3	-26	69	12
	Todo noturno	10,7	q10c	11	842	41	252	836	38	-21	6	6
	Maior parte noturno	5,9										
Em que tipo de estabelecimento você cursou o Ensino Médio?	Maior parte pública	9,1										
	Maior parte particular	54,5										
	Todo na pública	7,8	q12c	60	992	114	193	992	125	-4	0	1
	Todo na particular	28,6	q12d	31	989	171	-331	987	186	15	2	9
Modalidade de Ensino Médio que concluiu ou está concluindo?	Científico	83,1	q13a	83	711	4	-24	682	3	-5	29	3
	Técnico	5,9										
	Magistério	7,2										
	Supletivo	3,9										
Freqüentou algum curso pré-vestibular?	Não	45,9	q15a	46	812	13	-47	381	6	-50	431	148
	Sim, particular	34,0	q15b	34	953	44	-146	839	41	54	114	130
	Sim, populares	20,0	q15c	20	974	134	359	973	144	11	1	3
Quantos vestibulares você prestou, anteriormente, na UESC?	Nenhum	51,1	q18a	51	806	5	0	0	0	-41	806	109
	Um	23,5	q18b	23	175	1	-14	174	0	1	0	0
	Dois	14,0	q18c	14	707	4	-4	3	0	62	704	69
	Três	6,6										
	Quatro	2,7										
	Cinco	1,1										
Grau de escolaridade do pai ou responsável.	Não alfabetizado	6,37										
	Fundamental incompleto	30,0	q24b	29	991	72	219	989	79	9	2	3
	Fundamental completo	7,43										
	Médio incompleto	8,30										
	Médio completo	28,6	q24e	28	928	21	-117	925	21	-7	3	2
	Superior incompleto	6,36										
Grau de escolaridade da mãe ou responsável.	Superior completo	12,9	q24g	13	975	80	-350	975	86	5	0	0
	Não alfabetizado	5,1										
	Fundamental incompleto	23,3	q25b	23	987	89	273	983	97	18	4	10
	Fundamental completo	5,9										
	Médio incompleto	7,5										
	Médio completo	33,8	q25e	34	794	9	-65	791	8	-4	3	1
Superior incompleto	8,7											
	Superior completo	15,7	q25g	16	988	95	-346	988	104	3	0	0

Nota: mass é proporcional à quantidade de alunos que responderam ao; inr – é a inércia da pergunta; qit – é a qualidade da representação bidimensional (considerando apenas os dois primeiros componentes); k – é a componente principal número k; cor – é a proporção da inércia explicada pela componente principal k, (cor1 + cor2 = qit); ctr – é a contribuição da pergunta à inércia da CP.

Para as variáveis sócio-culturais 17 itens tiveram massa mínima maior ou igual a 10, conforme Tabela 2. A primeira componente para as informações escolares foi formada basicamente pela contribuição de nove itens, conforme a seguir: q27A + q27C - q27f - q28C - q28D + q28e + q29A - q29B + q29C. A segunda componente se formou

pela contribuição de nove itens: q26C + q27C + q27D + q28C + q28E - q26B - q27F - q28D - q29A.

Tabela 2 – Descrição das características sócio-culturais dos vestibulandos do ano de 2008 da UESC, Ilhéus, BA – 2008.

Variável	Categoria	%	Análise de correspondência										
			Código	Massa			PC 1			PC2			
				mass	qlt	inr	K=1	cor	ctr	K=1	Cor	Ctr	
Excetuando os escolares, quantos livros, em média, você lê por ano?	Nenhum	2,1											
	1 a 2	24,4	q26b	61	492	9	24	182	2	-31	311	75	
	3 a 5	41,3	q26c	103	562	3	5	45	0	17	517	37	
	6 a 10	21,9	q26d	55	395	3	-9	88	0	17	308	21	
	11 a 20	7,0	q26e	17	284	5	-36	240	1	15	44	5	
	Mais de 20	3,3											
Qual o meio que você mais utiliza para se manter informado(a)?	Televisão	46,8	q27a	118	933	36	75	922	37	-8	12	11	
	Rádio	0,6											
	Jornal	9,8	q27c	24	804	29	132	743	24	38	61	45	
	Revistas	10,4	q27d	26	931	20	-94	575	13	74	355	185	
	Conversas	1,5											
	Internet	30,9	q27f	77	966	65	-125	933	67	-24	33	56	
Além dos estudos, com quais atividades você mais ocupa seu tempo?	Leitura	20,9	q28a	52	824	22	82	783	19	19	41	23	
	Teatro, dança	2,5											
	Cinema, vídeo	9,1	q28c	23	937	59	-207	834	55	73	103	156	
	Esportes	18,3	q28d	46	951	51	-135	814	46	-55	138	182	
	Religião	12,4	q28e	31	976	82	221	932	85	47	43	91	
	Música	12,2	q28f	31	575	12	-64	510	7	-23	65	21	
	Artesanato, pintura etc	2,0											
	Circulando	1,3											
	Outros	21,2	q28i	53	246	9	25	193	2	-13	53	12	
Você tem acesso à Internet?	Não	5,6	q29a	14	894	71	294	856	67	-62	39	70	
	Sim, em casa	49,7	q29b	125	998	291	-215	996	321	7	1	8	
	Sim, em outros locais	44,8	q29c	113	988	232	201	988	253	-1	0	0	

Nota: mass é proporcional à quantidade de alunos que responderam ao; inr – é a inércia da pergunta; qlt – é a qualidade da representação bidimensional (considerando apenas os dois primeiros componentes); k – é a componente principal número k; cor – é a proporção da inércia explicada pela componente principal k, (cor1 +cor2 = qlt); ctr – é a contribuição da pergunta à inércia da CP.

A tabela de contingência analisada é chamada de tabela empilhada. Primeiramente fazemos a tabulação cruzada de cada questão do quesito INFORMAÇÕES ESCOLARES com cada questão do quesito Informações SOCIAIS. Depois empilhamos todas elas mantendo as INFORMAÇÕES ESCOLARES como linhas e as SOCIAIS como colunas. Neste caso, teoricamente, o total de cada tabela empilhada seria o de todos os alunos inscritos, 13662. Teoricamente, a inércia total deveria ser exatamente a média aritmética das inércias de cada tabela, porém, como há muitos casos faltantes (NA's) os totais na verdade são um pouco diferentes uns dos outros o que aumenta o valor da inércia total em relação à média citada.

A primeira componente principal (PC1) explica 80% da inércia e a segunda PC2 explica 6%. É importante frisar que as informações escolares (linhas) e as sócio-culturais (colunas) estão projetados em dois espaços diferentes no mesmo gráfico, o que significa que não tem sentido analisar agrupamentos de itens provenientes destas duas variáveis. O importante é analisar os caminhos na PC1 que se movem os pontos das linhas e colunas da tabela de contingência e para onde isto aponta.

No sentido da PC1, as linhas (*informações escolares, em preto*) tem mais à esquerda os pontos q12D, q25G e q24G (*Todo na escola particular, Superior completo pai e mãe*).

Depois vêm os pontos q15B, q24E, q9A (*Sim, (cursinho pré-vestibular) particular, Ensino médio completo (pai), concluiu ensino médio 2007*). Do lado direito da origem, q12C, q10C, q25B, q15C (*Todo na escola pública (ensino médio), Todo noturno, Ensino fundamental incompleto (mãe), Sim, programas populares (cursinho pré-vestibular)*).

O caminho das informações escolares denota que as diferenças principais são a de ordem econômica que se polariza na forma de estudo dos pais, e escolas particulares e públicas e os alunos mais pobres tendo de cursar as escolas públicas noturnas pois tem de trabalhar. Apesar de que a maioria dos alunos que cursaram ensino médio em escola pública o fizeram no turno *todo diurno* q10A, esta resposta não ficou polarizada pois todos os estudantes que fizeram o ensino médio particular também o fizeram *todo diurno*. A resposta q10A está perto da origem no lado esquerdo. Os alunos que cursaram no *todo noturno*, todos estudaram em escolas públicas o que põe q10C junto com q12C.

No sentido da PC1, as colunas (*informações sócio-culturais, em vermelho*) tem mais à esquerda os pontos q29B, q28C, depois q28D, q27F, q27D (*Sim, em casa (internet), Cinema-vídeo (atividades), esportes (atividades), internet (informação), revistas (informação)*). Do lado direito da origem estão q27C, q29C, q28E, q29A (*Jornal (informações), Sim apenas em outros locais (internet), Religião (atividades), Não (internet)*).

Da mesma forma a polarização é a econômica de um lado alunos de classe média (*internet em casa etc..*) e os de menor poder aquisitivo (*internet em outros lugares ou não tem acesso etc.*)

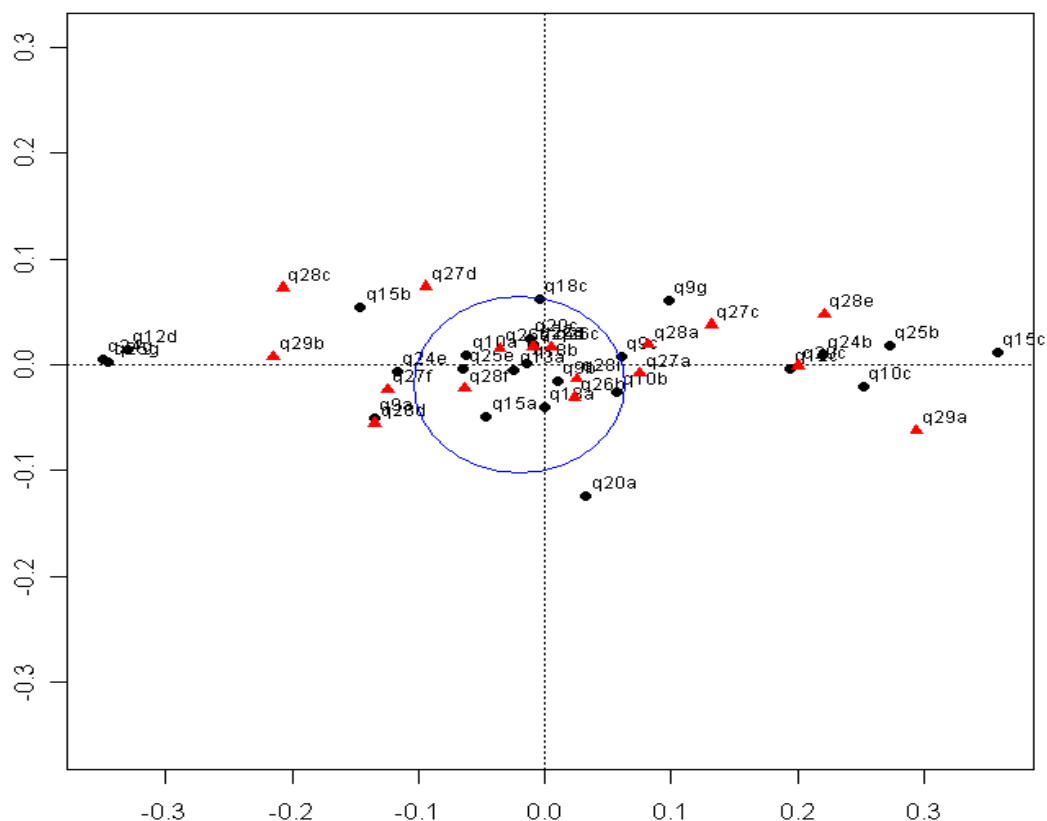


Figura 1 – Primeiro e segundo componente para aspectos escolares e informações escolares para candidatos do vestibular 2008 da UESC, Ilhéus – 2008.

CONCLUSÃO

A análise de correspondência mostrou-se eficiente na análise das informações dos vestibulandos da UESC, e aponta para a possibilidade de uma maior exploração e profundidade de análise das informações dos vestibulandos.

Como já era de se esperar a polarização principal, o que marca a diferença entre os alunos da região da UESC é a diferença econômica que se manifesta tanto na formação escolar fundamental e média do estudante como na sua vida sócio-cultural.

REFERÊNCIAS BIBLIOGRÁFICAS

Greenacre; M.J. (2007). Correspondence Analysis in Practice. Chapman & Hall/CRC, Boca Raton, second edition.