

HETEROGENEIDADE DO DESEMPENHO DE ALUNOS DA UNICAMP, DO INGRESSO À CONCLUSÃO

RAFAEL PIMENTEL MAIA

Consultor da Comissão Permanente para o Vestibular
da Universidade Estadual de Campinas
maia@comvest.unicamp.br

HILDETE PRISCO PINHEIRO

Professora do Departamento de Estatística da Universidade Estadual de Campinas
hildete@ime.unicamp.br

ALUISIO DE SOUZA PINHEIRO

Professor do Departamento de Estatística da Universidade Estadual de Campinas
pinheiro@ime.unicamp.br

RESUMO

O objetivo deste trabalho é propor uma metodologia para avaliar o desempenho dos alunos da Universidade Estadual de Campinas, do ingresso à conclusão do curso. A amostra é composta por todos os ingressantes dessa universidade nos anos de 1997 a 2000, e a população que corresponde ao universo dos alunos ingressantes em todo o período de existência da universidade. O conjunto de dados disponível foi obtido a partir dos questionários socioculturais, aplicados pela Comissão Permanente de Vestibulares na inscrição do vestibular, e de informações acadêmicas fornecidas pela Diretoria Acadêmica da Universidade. É proposta uma metodologia com base na variável denominada "ganho relativo" sugerida por Dachs e Maia (2006). Essa nova metodologia fundamenta-se em medidas de diversidades propostas por Rao (1982) e na utilização de U-estatísticas. Propõem-se testes de homogeneidade para avaliar se existe diferença no desempenho entre alunos segundo alguns agrupamentos.

AVALIAÇÃO DE ESTUDANTE – ENSINO SUPERIOR – ESTATÍSTICAS

ABSTRACT

HETEROGENEITY IN THE PERFORMANCE OF UNICAMP STUDENTS FROM ADMISSION TO CONCLUSION OF UNDERGRADUATED STUDIES. The aim of this work is to propose a methodology to evaluate the performance of students at Unicamp [São Paulo State University at Campinas] from admission to graduation. The sample consists of all students enrolled in Unicamp from 1997 to 2000, and the population corresponds to the universe of students throughout the

existence of the university. The available database was gathered from socio-cultural questionnaires applied by the University Commission on College Entrance Examination at the time of enrollment for the examination ("vestibular") and from academic information provided by the Unicamp Academic Studies Board. A methodology is proposed based on the "relative gain" variable suggested by Dachs and Maia (2006). This new methodology is based on the diversity measures proposed by Rao (1982) and on the use of U-statistics. Homogeneity tests have been proposed to assess whether there is a difference in performance among students from different groups.

STUDENTS EVALUATION – HIGHER EDUCATION – STATISTICS

Há alguns anos vêm ocorrendo na sociedade discussões sobre a implantação de políticas de cotas nas universidades. Algumas instituições de ensino superior adotaram tal política, reservando uma determinada percentagem de suas vagas para alunos oriundos do sistema público do ensino médio e/ou autodeclarados pretos ou pardos. Na Universidade Estadual de Campinas – Unicamp –, em 2005, o Conselho Universitário – Consu – aprovou a adoção de políticas de ação afirmativa (Bowen, Bok, 1998) e foi criado o Programa de Ação Afirmativa e Inclusão Social – Paais. Desde o vestibular de 2005 são atribuídos 30 pontos à nota final do vestibular (padronizada para ter média de 500 pontos e desvio padrão de 100 pontos) para os candidatos que cursaram o ensino médio exclusivamente em escolas da rede pública de ensino e mais 10 pontos para os que, além de ter estudado em escolas públicas, se autodeclararam negros ou pardos.

Dachs e Maia (2006) ajustaram modelos de regressão linear e logística para o desempenho dos alunos ingressantes entre 1994 a 1997 na Unicamp, a partir de dados fornecidos pela Comissão Permanente para o Vestibular – Comvest – e pela Diretoria Acadêmica da Unicamp – DAC. O desempenho foi medido através de uma variável chamada de "ganho relativo" que consiste na diferença da colocação do aluno, dentro de sua turma, com relação ao coeficiente de rendimento final – CR – e a sua colocação com relação a nota final padronizada do vestibular – NFP. Os modelos comparam, entre outras características, se há diferença no ganho relativo médio entre estudantes que cursaram o ensino médio em escolas públicas e os estudantes que cursaram em escolas particulares. Neste trabalho consegue-se identificar que os estudantes oriundos de escolas públicas apresentam ganho relativo médio superior aos dos alunos de escolas particulares. Esses resultados serviram como subsídios para a adequação da política de ação afirmativa adotada pela Unicamp.

Este trabalho visa propor uma nova metodologia que possibilite comparar o desempenho de alunos de escolas públicas e privadas. Essa nova metodologia baseia-se na teoria de análise de diversidade proposta por Rao (1982) e em métodos não paramétricos como o *jackknife* (Efron, Tibshirani, 1993). A finalidade é apresentar maior robustez que os modelos de regressão, por incluir duas características importantes: a comparação de todos os indivíduos, 2 a 2, e portanto toda a variabilidade dos dados é mensurada; e a utilização de métodos não paramétricos para estimação dos parâmetros. Aqui não é necessário fazer suposição sobre a verdadeira distribuição de probabilidade dos dados. Serão utilizados dados mais recentes referentes aos alunos ingressantes de 1997 a 2000 fornecidos pela Comvest e pela DAC.

MEDIDA DE DESEMPENHO

Uma das dificuldades do problema em questão é propor uma medida que avalie o desempenho dos alunos da Unicamp. Essa medida, de acordo com Dachs e Maia (2006) precisa ser uma quantidade que mensure tal desempenho de forma comparável entre os diferentes cursos e anos de ingresso. As variáveis de que dispúnhamos eram a NFP e CR dos ingressantes.

A NFP é a média das notas nas provas das diferentes disciplinas, padronizada para ter média de 500 pontos e desvio padrão de 100 pontos. Utilizar a NFP como medida de desempenho não é satisfatório, uma vez que avalia o desempenho do aluno apenas no ingresso à universidade. Uma situação ainda mais complexa ocorre com o CR do aluno. Os processos de avaliação interna são muito distintos para os diferentes cursos das várias áreas e também apresentam variações ao longo do tempo.

Por essas razões, decidiu-se criar uma variável que foi chamada de “ganho relativo”, da seguinte forma: aos alunos de uma mesma turma (ingressantes em mesmo ano e curso) foram atribuídos dois postos (colocação), um baseado na NFP e outro no CR (o aluno com menor nota recebeu o posto 1, segunda menor nota o posto 2, e assim sucessivamente), chamados de posto inicial e final, respectivamente.

Os postos foram divididos pelo número total de alunos em cada turma, para torná-los comparáveis entre turmas, já que elas variam de tamanho de um curso a outro e de um ano a outro, padronizando a escala dos postos para

todas as turmas, de forma que variem entre 0 e 1. Esses novos valores foram chamados de postos relativos. Assim, o aluno com maior NFP/CR recebe posto relativo igual a 1, e quanto mais próximo de 1 for o posto relativo do aluno, melhor foi o seu desempenho com relação à sua turma, e quanto mais próximo de 0, pior. O ganho relativo foi obtido da diferença entre o posto relativo baseado no CR (final) e o posto relativo baseado na NFP (inicial).

O ganho relativo é, por construção, uma variável limitada entre -1 e 1 e simétrica em torno do zero, com média e mediana iguais a zero. Apesar do problema de que os cursos têm métodos de avaliações distintos e as turmas possuem diferentes tamanhos (número de alunos), é razoável assumir que a variável ganho relativo, da maneira como foi construída, é comparável entre turmas. Isso porque se trabalha com postos (ao invés das notas absolutas) relativos à sua turma, ou seja, padronizados para variarem entre 0 e 1.

ANÁLISE DE DIVERSIDADE

A análise de diversidade (Rao, 1982, 1982a) tem sido muito utilizada em várias áreas do conhecimento (Pinheiro, Pinheiro, 2007; Pinheiro, Pinheiro, *sen*, no prelo), tais como: Antropologia, Genética, Economia, Sociologia e outros ramos da Biologia. Ela pode ser considerada como uma generalização da análise clássica de variância – Anova – e vêm sendo muito útil para analisar dados qualitativos que surgem nessas diversas áreas. A partir da análise de diversidade pretende-se avaliar se existem diferenças com relação ao desempenho acadêmico entre alunos que estudaram o segundo grau em escolas públicas e os que estudaram em escolas particulares.

Na análise de variância decompõe-se a variabilidade de um conjunto de dados quantitativos em um número de componentes aditivos. Cada componente é usada para testar certa hipótese nula ou para estimar uma componente da variância. Rao (1982) introduziu uma medida geral de diversidade (variabilidade) aplicável tanto a dados quantitativos quanto a dados qualitativos, estendendo o conceito de Anova para um caso mais geral, chamando de análise de diversidade – Anodiv.

Uma medida de diversidade pode ser usada para decompor a diversidade total dentro de uma determinada população devido a certo número de fatores. No caso em que se tenha uma mistura de populações, pode-se estar

interessado em saber quanto da diversidade da composição das populações é devido à diversidade dentro de cada população e quanto é devido à diversidade entre populações.

Para comparar dois ou mais grupos (alunos de escola pública e particular, por exemplo), foram definidas duas quantidades, uma denominada de Q_g , que representa a variabilidade dentro do grupo g , e outra denominada $C_{gg'}$, que representa a variabilidade entre os grupos g e g' . A variabilidade total dentro da amostra (SQT) pode ser decomposta em duas componentes, uma chamada de variabilidade intragrupos (SQI) e outra chamada de variabilidade entregrupos (SQE). Sob algumas condições o $E(SQE)$, valor esperado de SQE , é maior ou igual a 0, ou seja, positivo, sendo igual a zero apenas quando há homogeneidade entre os grupos testados.

A partir desses conceitos foi proposta como medida de distância entre dois indivíduos, a diferença quadrática do ganho relativo $(x - y)^2$, onde x representa o ganho relativo do indivíduo 1 e y o ganho relativo do indivíduo 2 e calculado o SQT e o SQE .

Para o caso mais geral, onde se comparam G grupos, a SQE é dada pela seguinte fórmula:

$$SQE = \frac{1}{n(n-1)} \left\{ \sum_{g < g'} n_g n_{g'} \left(2\hat{C}_{gg'} - \hat{Q}_g - \hat{Q}_{g'} \right) \right\},$$

$$\text{em que } \hat{Q}_g = \binom{n_g}{2}^{-1} \sum_{i < j} (x_{gi} - x_{gj})^2, \hat{C}_{gg'} = \frac{1}{n_g n_{g'}} \sum_i \sum_j (x_{gi} - x_{g'j})^2 \text{ e } n = \sum_g n_g,$$

para todo $g = 1, 2, \dots, G$.

Foi então proposto um teste de homogeneidade para avaliar duas hipóteses: os grupos de interesse são homogêneos (hipótese nula) e os grupos de interesse não são homogêneos (hipótese alternativa). Formalmente as hipóteses são as seguintes:

$$H_0: 2C_{gg'} - Q_g - Q_{g'} = 0 \forall g = 1, \dots, G;$$

$$H_1: 2C_{gg'} - Q_g - Q_{g'} > 0 \text{ para pelo menos 1 par } (g, g') = 1, \dots, G.$$

Não será assumida nenhuma distribuição para o ganho relativo. As estimativas das variâncias dos estimadores serão obtidas a partir do método de reamostragem de *jackknife* (Davison, Hinkley, 1999). Com isso, garante-se maior robustez às estatísticas propostas, uma vez que não se faz nenhuma suposição a respeito da distribuição dos dados.

APLICAÇÃO

A população de interesse consiste dos alunos já ingressantes na Unicamp e a amostra é composta por todos os ingressantes nos anos de 1997, 1998, 1999 e 2000. O conjunto de dados foi fornecido pela Comvest e pela DAC da Unicamp. Este contém informações de 7.515 alunos ingressantes em todos os cursos de graduação oferecidos (foram excluídos os alunos ingressantes nos cursos tecnológicos ministrados no Campus de Limeira). Essas informações foram atualizadas no final do primeiro semestre de 2007 e referem-se às respostas do questionário sociocultural aplicado pela Comvest no momento da inscrição no vestibular, e informações acadêmicas tais como: nota obtida no vestibular padronizada (NFP), coeficiente de rendimento final do aluno (CR), situação acadêmica, notas obtidas nas disciplinas cursadas.

A amostra é composta, em sua maioria, por estudantes com idades entre 16 e 24 anos (apenas 7,3% dos ingressantes declararam ter mais de 24 anos), e é predominantemente masculina (59,1%).

A situação acadêmica desses alunos, como pode ser vista na tabela I, foi classificada de três formas: graduados (alunos que já haviam concluído o curso), ativos (alunos que não haviam concluído o curso e que ainda estavam matriculados) e outros (em geral alunos jubilados ou desistentes). A maioria dos alunos (76,1%) já havia graduado, 22,4% foram jubilados ou desistiram do curso e apenas 1,5% ainda estavam ativos. Há um aumento do número de ingressantes ao longo dos anos devido à maior oferta de vagas pela universidade. Tomando como base o ano de 1997, em 1998 houve um aumento no ingresso dos alunos em 23,5% (1720/1393 - 1); em 1999, de 57,4% (2192/1393 - 1); e em 2000, de 58,7% (2210/1393 - 1). O maior aumento ocorreu de 1998 para 1999, quando o crescimento foi de 27,4% (2192/1720 - 1).

TABELA I
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO A SITUAÇÃO
ACADÊMICA POR ANO DE INGRESSO

Situação Acadêmica	ano de ingresso								Total	
	1997		1998		1999		2000			
	n	%	n	%	n	%	n	%	n	%
graduado	1.099	78,9	1.315	76,5	1.641	74,9	1.664	75,3	5.719	76,1
ativo	5	0,4	15	0,9	39	1,8	52	2,4	111	1,5
outros	289	20,7	390	22,7	512	23,4	494	22,4	1.685	22,4
Total	1.393	100,0	1.720	100,0	2.192	100,0	2.210	100,0	7.515	100,0

Os alunos são ingressantes em 45 cursos pertencentes a todas as áreas do conhecimento. Os cursos foram classificados segundo a área da seguinte forma:

1. Exatas: Matemática (licenciatura e bacharelado), Matemática Aplicada, Estatística, Física (licenciatura e bacharelado), Química, Ciência da Computação, Química Tecnológica, Ciências da Terra, Geologia e Geografia;
2. Engenharias: Arquitetura e os cursos de Engenharia agrícola, química, mecânica, elétrica, civil, de alimentos, de computação e de controle e automação;
3. Biológicas: Biologia (licenciatura e bacharelado), Odontologia, Medicina, Enfermagem e Educação Física;
4. Humanas: Letras, Ciências Sociais, Ciências Econômicas, Linguística, História, Pedagogia, Filosofia e Licenciatura em Química/Física;
5. Artes: Música, Dança, Artes Visuais e Artes Cênicas.

A tabela 2 apresenta a distribuição da amostra por ano, segundo a área de interesse. Observa-se que as áreas com maior número de ingressantes são as Engenharias e Exatas com cerca de 30% e 22% dos alunos, respectivamente, seguidas das Biológicas e Humanas com cerca de 20% dos ingressantes cada uma. A área de artes é responsável por apenas 6,4% da amostra avaliada.

A principal característica de interesse é a origem dos alunos com relação ao tipo de escola em que cursaram o 2º grau, isto é, se vieram de escolas públicas ou particulares. No questionário sociocultural da Comvest, a pergunta era feita da seguinte forma: "Qual o tipo de estabelecimento que cursou no

ensino médio (2º grau)?", e apresentava as seguintes alternativas: somente particular; somente público; mais público; mais particular; igual tempo em ambas e nenhuma das alternativas. A partir dessa questão foram considerados oriundos de escolas particulares os alunos que declararam ter estudado todo o ensino médio ou a maior parte do tempo em tais escolas. Analogamente, alunos que estudaram todo o ensino médio ou a maior parte dele em estabelecimentos públicos foram considerados como oriundos de escolas públicas. Os alunos que assinalaram uma das duas últimas opções (igual tempo em ambas e nenhuma das alternativas) foram considerados como falta de informação – caso de 117 alunos (1,6%).

TABELA 2
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO A ÁREA
DE INTERESSE POR ANO DE INGRESSO

Área	ano de ingresso								Total	
	1997		1998		1999		2000			
	n	%	n	%	n	%	n	%	n	%
Exatas	235	16,9	416	24,2	516	23,5	515	23,3	1.682	22,4
Engenharias	499	35,8	513	29,8	643	29,3	662	30,0	2.317	30,8
Biológicas	322	23,1	333	19,4	407	18,6	415	18,8	1.477	19,7
Humanas	238	17,1	345	20,1	492	22,5	484	21,9	1.559	20,8
Artes	99	7,1	113	6,6	134	6,1	134	6,1	480	6,4
Total	1.393	100,0	1720	100,0	2.192	100,0	2.210	100,0	7.515	100,0

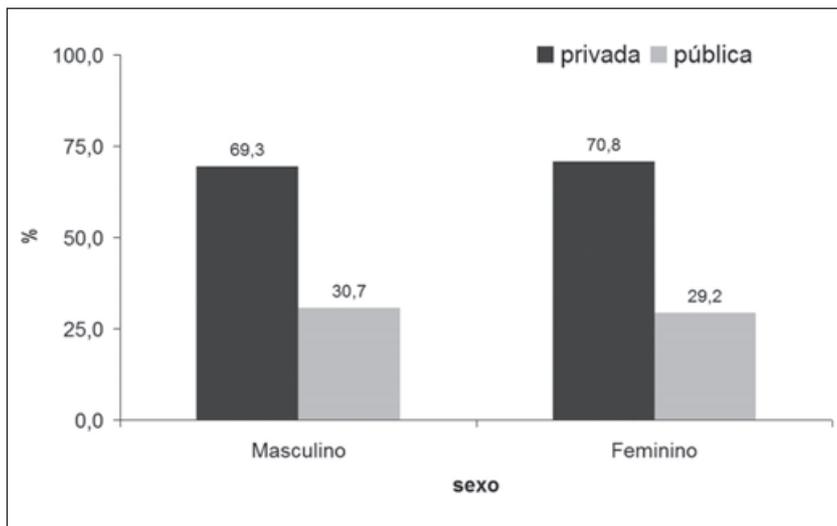
TABELA 3
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO O TIPO DE ESCOLA
CURSADA NO 2º GRAU POR ANO DE INGRESSO

Área	ano de ingresso								Total	
	1997		1998		1999		2000			
	n	%	n	%	n	%	n	%	n	%
privada	965	70,5	1.181	69,7	1.522	70,6	1.500	68,9	5.168	69,9
pública	404	29,5	513	30,3	635	29,4	678	31,1	2.230	30,1
Total	1.369	100,0	1.694	100,0	2.157	100,0	2.178	100,0	7.398	100,0

Como mostra a tabela 3, 30% dos alunos que ingressaram entre 1997 e 2000 são oriundos de escolas públicas, contingente mais de 2 vezes menor

do que o de alunos que cursaram escolas particulares. Essa proporção é válida também quando se observa a distribuição por sexo: cerca de 30% das mulheres e 30% dos homens que ingressam na universidade estudaram em escolas públicas no segundo grau (Fig. 1).

FIGURA 1
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO O TIPO
DE ESCOLA DO ENSINO MÉDIO POR SEXO



Com relação à área, as Exatas apresentam maior percentagem de alunos oriundos de escolas públicas (42,6%), seguida das Humanas (34,8%), e a área com menor percentagem é a das Engenharias (22,1%), como se observa na Figura 2.

Outra característica avaliada foi se o aluno trabalhava ou não ao entrar na universidade. Na amostra total, pouco menos de um terço, 27,1% dos estudantes (1990), declarou que trabalhava ao ingressar na universidade. Quando se faz essa mesma análise separadamente para cada área (Fig. 3), observa-se que a área com maior percentagem de alunos que trabalhavam é Artes (50,6%), seguida de Exatas (37,5%) e de Humanas (34,5%). Nas áreas das Engenharias e Biológicas apenas 16% dos alunos trabalhavam.

FIGURA 2
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO O TIPO
DE ESCOLA DO ENSINO MÉDIO POR ÁREA

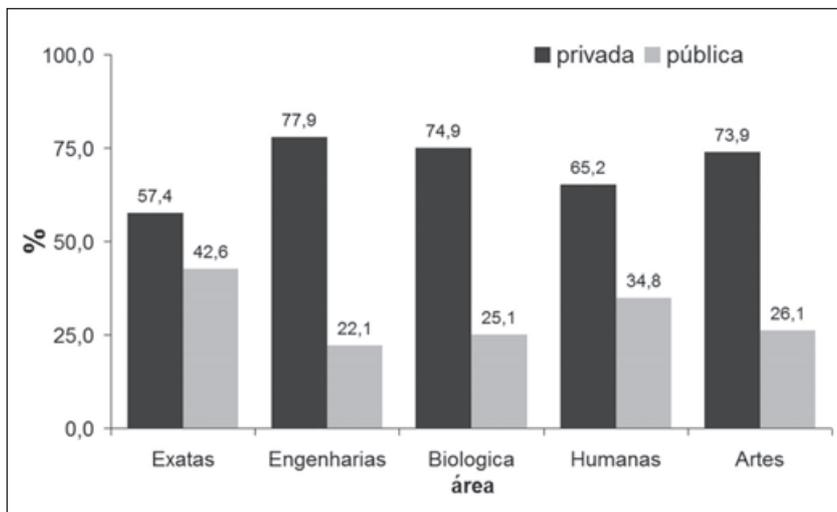
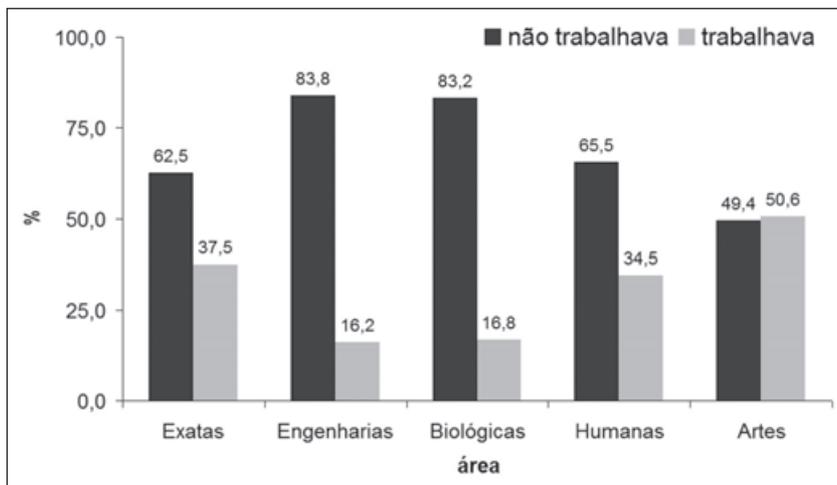


FIGURA 3
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO A VARIÁVEL
TRABALHAVA AO ENTRAR NA UNIVERSIDADE, POR ÁREA



Entre os sexos, a percentagem de rapazes que trabalhavam é maior que a de mulheres, 29,4% contra 23,7%. Já quando se comparam alunos de escolas públicas e privadas, 47,9% dos alunos oriundos de escolas públicas declararam que trabalhavam contra apenas 18,1% que estudaram em escolas particulares (Tab. 4).

TABELA 4
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO A VARIÁVEL TRABALHAVA
OU NÃO POR SEXO E POR TIPO DE ESCOLA CURSADA NO 2º GRAU

Trabalhava ao entrar na universidade	sexo		escola do 2º grau		Total
	masculino	feminino	particular	pública	
não	70,6	76,3	81,9	52,1	72,9
sim	29,4	23,7	18,1	47,9	27,1
Total	100,0	100,0	100,0	100,0	100,0

Com relação ao perfil econômico dos alunos, a única variável de que se dispunha era a renda mensal familiar do aluno em salários mínimos – SM. A amostra total está distribuída da seguinte maneira: declararam que a renda familiar mensal era superior a 20 SM 48,2% dos estudantes; entre 10 e 20 SM, 29,1%; entre 3 e 10 SM, 20,8%; e inferior a 3 SM, apenas 2%.

A Figura 4 apresenta o gráfico com a distribuição dos estudantes segundo a renda por sexo e por tipo de escola do 2º grau, respectivamente. Não há grandes diferenças na distribuição da renda mensal familiar entre os sexos. Entretanto, quando se comparam alunos oriundos de escolas particulares com alunos oriundos de escolas públicas, nota-se que os primeiros apresentam rendas maiores: 58,5% destes alunos declararam renda acima de 20 SM e 26,0% renda familiar mensal entre 10 e 20 SM, ou seja, 84,5% estariam na faixa acima de 10 SM. Entre os alunos que estudaram em escolas públicas, 35,5% declaram renda familiar entre 3 e 10 SM; 36,4%, entre 10 e 20 SM; e apenas 24,4%, superior a 20 SM. Comparado com o grupo anterior (alunos de escolas privadas), 60,8% declararam renda acima 10 SM (23,6% a menos).

A tabela 5 apresenta algumas medidas resumo para o ganho relativo segundo alguns grupos formados de acordo com o sexo e o tipo de escola e se o aluno trabalhava ao entrar na universidade.

FIGURA 4
DISTRIBUIÇÃO DA AMOSTRA SEGUNDO A RENDA MENSAL FAMILIAR EM SALÁRIOS MÍNIMO POR TIPO DE ESCOLA CURSADA NO 2º GRAU E POR SEXO

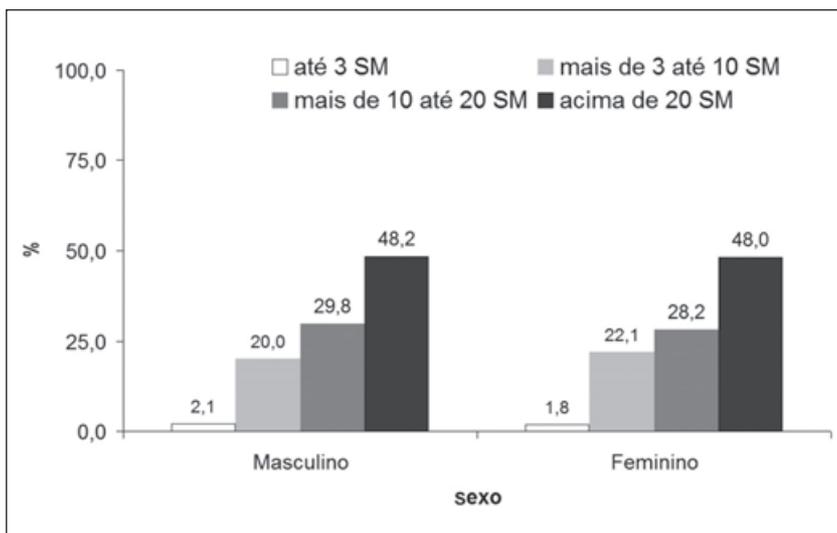
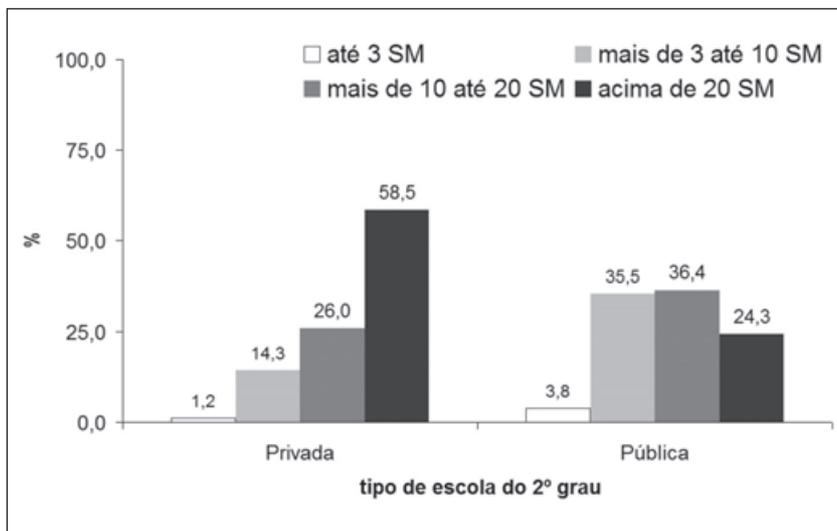


TABELA 5
 MEDIDAS RESUMO PARA O GANHO RELATIVO
 SEGUNDO O GRUPO DE INTERESSE

Grupos	n	média	D.P.	mínimo	máximo
tipo de escola do 2º grau					
particular	5.156	-0,020	0,348	-0,984	0,984
pública	2.223	0,047	0,363	-0,976	0,958
sexo					
masculino	4.356	-0,041	0,354	-0,984	0,984
feminino	3.023	0,059	0,344	-0,947	0,957
trabalhava					
não	5.414	0,002	0,347	-0,984	0,968
sim	2.029	-0,003	0,369	-0,960	0,984
interação sexo e tipo de escola do 2º grau					
masculino - particular	3.017	-0,059	0,348	-0,984	0,984
masculino - pública	1.339	0,001	0,366	-0,976	0,958
feminino - particular	2.139	0,036	0,340	-0,927	0,947
feminino - pública	884	0,116	0,349	-0,947	0,957

Os alunos oriundos de escola pública apresentaram ganho médio (0,047) maior comparado aos alunos que estudaram em escolas particulares (-0,020). Quando se comparam os sexos, as mulheres apresentaram ganho médio igual a 0,059 contra -0,041 dos homens.

Avaliando a interação entre o sexo e o tipo de escola, as disparidades entre tipo de escola são maiores entre as mulheres do que entre os homens. A diferença no ganho relativo médio entre alunos de escolas públicas e alunos de escolas privadas é de 0,080 entre as mulheres e 0,058 entre os homens. Essas divergências também podem ser observadas a partir dos gráficos de caixas apresentados na figura 5.

A tabela 6 apresenta os valores estimados da SQE (variabilidade entregrupos), o desvio padrão (D.P;jack) obtido a partir do método de amostragem de *jackknife* e o respectivo Intervalo de 95% de Confiança (IC95%). Os IC95% foram construídos utilizando a seguinte fórmula: $SQE \pm 1,96 * D.P;jack$.

FIGURA 5
GRÁFICO DE CAIXAS PARA O GANHO RELATIVO
POR SEXO E TIPO DE ESCOLA CURSADA NO SEGUNDO GRAU

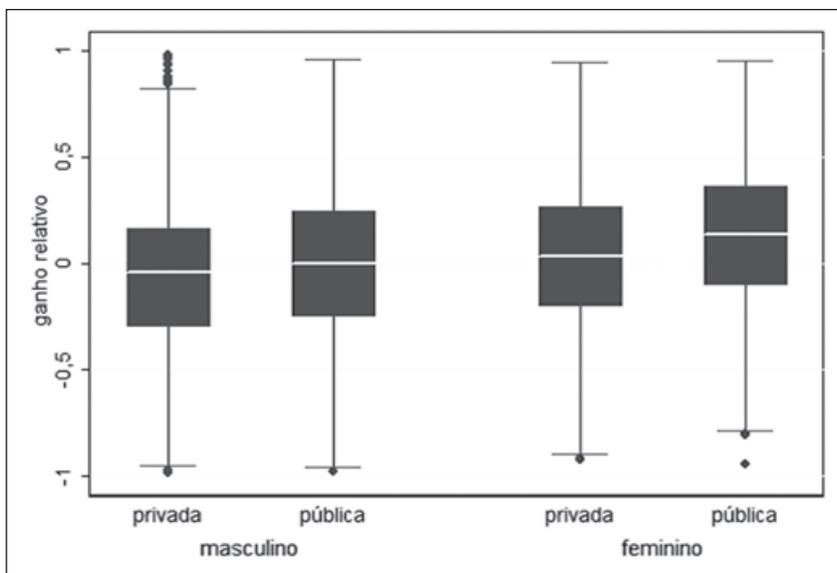


TABELA 6
RESULTADO DA ANÁLISE DE DIVERSIDADE

grupos	SQE	D.Pjack	IC95%	
escola	0,0032	0,0007	0,0019	0,0046*
sexo	0,0028	0,0010	0,0008	0,0048*
trabalha	-0,0003	0,0005	-0,0012	0,0007
Interação tipo escola x sexo	0,0104	0,0013	0,0079	0,0129*
esola sexo	0,0076	0,0016	0,0044	0,0108*
sexo escola	0,0072	0,0015	0,0043	0,0100*

*Diferença significativa ao nível de 5%; escola | sexo = fator tipo de escola condicionado ao sexo; sexo | escola = fator sexo condicionado ao tipo de escola.

A análise de diversidade apresentada na tabela acima mostra que apenas entre os grupos dos que trabalhavam e os dos que não trabalhavam ao entrar na universidade, não houve diferença significativa, no nível de 5%. Portanto, pode-se concluir que os alunos de escolas públicas apresentaram maior ganho

relativo médio, comparados aos alunos oriundos de escolas particulares, e as mulheres apresentaram maior ganho relativo médio comparadas aos homens.

A interação entre sexo e tipo de escola também foi significativa, no nível de 5%. Como pode ser observado na figura 5, a diferença no ganho relativo entre alunos oriundos de escolas públicas e privadas, quando se comparam apenas estudantes do sexo feminino, é maior do que quando comparados apenas os estudantes do sexo masculino.

Quando ajustado o fator tipo de escola pelo sexo (tipo de escola condicionado ao sexo) e o fator sexo pelo tipo de escola (sexo condicionado ao tipo de escola), ambos os fatores foram significativos.

CONSIDERAÇÕES

Uma das vantagens das metodologias baseadas nas medidas de diversidade de propostas por Rao (1982) é que nelas todos os indivíduos são comparados 2 a 2, avaliando-se toda a variabilidade dos dados, o que proporciona estatísticas menos suscetíveis a perturbações.

Com relação à abordagem paramétrica utilizada em Pedrosa et al. (2007) e não paramétrica apresentada nesse trabalho, a segunda é mais robusta, pois não faz suposição sobre a distribuição dos dados, além de que é muito mais simples obter as estimativas das variâncias por técnicas de re-amostragem. Ao se assumir uma distribuição para o conjunto de dados, a escolha de uma distribuição que não seja adequada a eles pode acarretar uma sobre-estimação das variâncias.

REFERÊNCIAS BIBLIOGRÁFICAS

BOWEN, W.; BOK, D. *The Shape of the river: long-term consequences of considering race in Colleg and University admissions*. Princeton: Princeton University Press, 1998.

DACHS, J.N.W.; MAIA, R.P. *Subsídios quantitativos para repensar as políticas de acesso à universidade: aumentando a equidade racial e econômica no ensino do terceiro grau do Brasil e no Estado de São Paulo. (Relatório técnico)*. Campinas: Núcleo de Estudos de Políticas Públicas, Universidade Estadual de Campinas, 2006. Desempenho relativo dos alunos da Unicamp que ingressaram nos anos de 1994, 1995, 1996 e 1997 e descrição dos alunos que prestaram o Provão em 2001, parte 2.

DAVISON, A.C.; HINKLEY, D.V. *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1999.

EFRON, B.; TIBSHIRANI, R. J. *An introduction to bootstrap*. New York: Chapman & Hall, 1993.

PEDROSA, R. H. L. et al. Academic performance, students' background and affirmative action at a Brazilian research university. *Higher Education Management and Policy*, v.19, n.3, p.58-77, 2007.

PINHEIRO, A.; PINHEIRO H. P. Métodos estatísticos não-paramétricos e suas aplicações. 26°. *Colóquio Brasileiro de Matemática*. Rio de Janeiro: Publicações Matemáticas, Instituto Nacional de Matemática Pura e Aplicada, 2007.

PINHEIRO, A.; PINHEIRO, H. P.; SEN, P. K. The Use hamming distance in bioinformatics. no prelo

PINHEIRO, H. P.; PINHEIRO, A.; SEN, P. K. Comparison of genomic sequences using Hamming distance. *Journal of Statistical Planning and Inference*, 130, n.1/2, p.325-339, 2005.

RAO, C. R.. Diversity: its measurement, decomposition, apportionment and analysis. *Sankya A*, n.44, p.1-21, 1982.

_____. Gini-Simpson index of diversity: a characterization, generalization and applications. *Utilitas Mathematica*, n.21, p.273-282, 1982a.

Recebido em: agosto 2008

Aprovado para publicação em: fevereiro 2009