



Universidade de Brasília
IE - Departamento de Estatística
Trabalho de Conclusão de Curso 2

**Identificação de tipologias por meio do método
Grade of Membership: Uma aplicação aos
estudantes da Universidade de Brasília**

Marina Barros de Oliveira

Orientadora: Prof^ª. Maria Teresa Leão Costa

Brasília
11 de julho de 2018

Marina Barros de Oliveira

Identificação de tipologias por meio do método *Grade of Membership*:
Uma aplicação aos estudantes da Universidade de Brasília

Orientadora: Prof^a. Maria Teresa Leão Costa

Trabalho de conclusão de curso apresentado
para obtenção do título de Bacharel em
Estatística ao Departamento de Estatística
da Universidade de Brasília.

Brasília

11 de julho de 2018

AGRADECIMENTOS

A realização desse trabalho foi um grande desafio pessoal. Desse modo, eu devo um agradecimento a todos aqueles que me ajudaram e me apoiaram durante essa fase complicada. Primeiramente eu quero agradecer minha professora e orientadora Maria Teresa, que sempre acreditou em mim, me motivou, teve paciência para me orientar e esteve de portas abertas para todos os meus momentos de aflição. Eu também quero oferecer meus sinceros agradecimentos para a professora Ana Maria, por todas as oportunidades que me permitiram crescer durante a graduação e por me mostrar a magia da estatística.

Também sou muito grata pela minha família incrível. Meus pais que sempre me apoiaram e entenderam meus dias mais difíceis. E não posso esquecer dos meus amigos maravilhosos que tornaram esse período menos estressante e me proporcionaram grandes momentos durante a graduação. Esses eu sei que são para a vida toda.

Finalmente, eu sou tão sortuda de ter a pessoa mais maravilhosa do meu lado. Obrigada, Rodrigo, não só por todo apoio, mas por sempre ter acreditado no meu potencial, e, mais importante, por me fazer acreditar em mim.

“N3o se pode falar de educa33o sem amor.”
(Paulo Freire)

RESUMO

O presente estudo tem como objetivo descrever perfis de estudantes ingressantes na Universidade de Brasília, por meio do método *Grade of Membership* – GoM, com vistas a subsidiar políticas de acompanhamento de desempenho e de assistência estudantil. Os dados referem-se aos ingressantes em 2017 extraídos da base de dados da pesquisa “Perfil dos Estudantes da Universidade de Brasília – Etapa Registro” conduzida pelo Observatório da Vida Estudantil desde 2012. Chegou-se à identificação de três perfis extremos: 1) ingressantes por cotas raciais, pretos ou pardos, baixa renda, ensino médio em escola pública, pais com até ensino fundamental completo, cursos de baixo prestígio; 2) ingressantes por cotas sociais, renda intermediária, ensino médio em escola pública, pais com ensino médio completo; e 3) ingressantes pelo sistema universal, brancos ou amarelos, renda elevada, ensino médio em escola particular, pais com nível superior completo, cursos de alto prestígio. Conclui-se que ainda que a Universidade de Brasília tenha ampliado o acesso a estudantes de baixa renda por meio da implementação de cotas sociais e raciais, desigualdades ao acesso a cursos de maior prestígio persistem.

Palavras Chave: *Grade of Membership*; Perfis socioeconômicos, Universidade de Brasília;

LISTA DE ABREVIATURAS E SIGLAS

ABEP Associação Brasileira de Estudos de Pesquisa

DF Distrito Federal

GoM Grade of Membership

OVE Observatório da Vida Estudantil

PE Perfil Extremo/Tipo puro

RLFM Razão Lambda/Frequência Marginal

SM Salários Mínimos

TM Tipo Misto

UnB Universidade de Brasília

LISTA DE ILUSTRAÇÕES

1	Percentual de ingressantes na UnB segundo sexo - 2017.	32
2	Percentual de ingressantes na UnB segundo raça/cor da pele - 2017. .	33
3	Percentual de ingressantes na Universidade de Brasília segundo sistema de ingresso - 2017.	34
4	Percentual de ingressantes na UnB segundo realização de curso preparatório - 2017.	35
5	Percentual de ingressantes na UnB segundo renda mensal da família - 2017.	36
6	Percentual de ingressantes na UnB segundo a classificação do Critério Brasil - 2017.	37
7	Percentual de ingressantes na Universidade de Brasília segundo escolaridade do pai - 2017.	38
8	Percentual de ingressantes na Universidade de Brasília segundo escolaridade da mãe - 2017.	38
9	Proporção de ingressantes na UnB segundo sexo e prestígio do curso - 2017.	39
10	Proporção de ingressantes na UnB segundo raça/cor e prestígio do curso - 2017.	40
11	Proporção de ingressantes na UnB segundo o tipo de ensino médio e prestígio do curso - 2017.	41
12	Percentual de ingressantes na Universidade de Brasília segundo a classificação dos perfis - 2017.	46
13	Percentual de ingressantes na Universidade de Brasília segundo o perfil e categoria de curso de prestígio - 2017.	47
14	Percentual de ingressantes na UnB segundo o perfil e categoria modalidade de ingresso - 2017.	48
15	Percentual de ingressantes na UnB segundo o perfil o campus - 2017.	49
16	Proporção de ingressantes na UnB segundo o número de tentativas de ingresso e prestígio do curso - 2017.	50
17	Proporção de ingressantes na UnB segundo a classificação dos perfis extremos e mistos - 2017.	51
18	Arquivo GOMK3 para a execução da programação de classificação dos perfis.	60
19	Formato do banco para a utilização da validação para variáveis contínuas.	65

LISTA DE QUADROS E TABELAS

Quadro 1 - Variáveis e Categorias.	29
Quadro 2 - Renda média domiciliar em cada classe socioeconômica segundo o Critério Brasil	30
Tabela 1 - Percentual de ingressantes na UnB segundo raça/cor da pele - 2017.	32
Tabela 2 - Valores de AIC para os modelos GoM com k perfis - UnB, 2017.	42
Tabela 3 -Frequências marginais, probabilidades estimadas e fatores deli- neadores de perfis, segundo variáveis socioeconômicas - Universidade de Brasília, 2017.	44
Tabela 4 - Comparação entre os perfis gerados pelo critério da moda e da preponderância relativa.	45
Tabela 5 - Comparação entre os perfis gerados pelo critério de Melo e de Guedes.	51
Tabela 6 - Resultados dos p-valores da comparação de médias entre cada perfil.	52
Tabela 7 - Distribuição das categorias segundo os tipos mistos.	54

Sumário

1	INTRODUÇÃO	12
2	OBJETIVOS	13
2.1	Objetivos Gerais	13
2.2	Objetivos Específicos	13
3	ACESSO AO ENSINO SUPERIOR	14
4	<i>GRADE OF MEMBERSHIP</i>	16
4.1	Estimação do grau de pertencimento (g_{ik})	18
4.2	Estimação dos parâmetros estruturais ou locais (λ_{kjl})	19
4.3	Caracterização e descrição dos perfis extremos	20
4.3.1	Critérios para a definição do número ótimo de perfis extremos	20
4.3.2	Critério de preponderância	22
4.4	Criação de tipos mistos a partir dos perfis extremos	22
4.5	Classificação dos perfis	23
4.5.1	Modelo de três perfis extremos	23
4.5.2	Critério de agrupamento pela moda	24
4.5.3	Critério de agrupamento pela preponderância relativa	25
4.5.4	Validação dos tipos mistos criados	25
5	BASE DE DADOS	27
5.1	Pesquisa sobre o perfil dos estudantes da Universidade de Brasília	27
5.2	Variáveis do estudo	27
5.2.1	Critério Basil	30
5.2.2	Sistema de Ingresso	30
5.2.3	Cursos de prestígio	31
5.2.4	Escolaridade dos pais	31
5.2.5	Raça/cor da pele	31
5.3	Análise dos dados	31
6	RESULTADOS	32
6.1	Retrato socioeconômico dos estudantes que ingressaram na UnB em 2017	32
6.2	Identificando perfis: <i>Grade of Merbership</i>	42
6.2.1	Tipos puros	42

6.2.2	Tipos mistos	50
7	CONCLUSÃO	56
8	REFERÊNCIAS BIBLIOGRÁFICAS	57
9	APÊNDICE	60
9.1	Código em R	60
9.1.1	Pacote GoMRcpp	60
9.1.2	Classificação dos perfis	60
9.1.3	Validação dos tipos mistos	65

1 INTRODUÇÃO

O sistema educacional de um país tem importância crucial para seu desenvolvimento socioeconômico. Investimentos em capital humano são componentes fundamentais do processo de crescimento econômico e são frequentemente associados a melhorias no estilo de vida, maior coesão social e melhores igualdades de oportunidades. Isso se dá, principalmente, para as parcelas mais carentes da população, atuando ainda como elementos de grande importância em questões de redução das desigualdades regionais e sociais (BARROS et al, 2001; WILLMS, 1997).

Diante da importância desempenhada pelo sistema educacional, viu-se a necessidade de se fazer um estudo sobre o corpo discente da Universidade de Brasília. Ao se observar, ao longo dos anos, o cenário delineado para o sistema de ensino da UnB, pode-se obter resultados importantes sobre o impacto proporcionado aos alunos da Universidade pelas transformações desse sistema em seus diversos aspectos. Dentre eles se destacam a implementação do sistema de cotas raciais e sociais e a adesão do Sistema de Seleção Unificada (Sisu) a partir do desempenho dos estudantes no Exame Nacional do Ensino Médio (ENEM).

Por outro lado, mesmo com essas transformações, ainda há sérias desigualdades em termos de acesso e condições de permanência na universidade (SILVA et al, 2006). Assim, verifica-se a importância de conhecer o perfil dos estudantes da UnB para o auxílio na elaboração de políticas de assistência estudantil e para a compreensão dos motivos de tais desigualdades e vulnerabilidades no meio acadêmico.

Desta maneira, para a identificação desses perfis e tipologias dos estudantes, optou-se pelo *Grade of Membership* (GoM), que é um método estatístico utilizado para delinear perfis multidimensionais com base em dados complexos, permitindo identificar grupos, ou *clusters*, e descrever as diferenças entre eles (GUEDES et al, 2016). Este método também é chamado de modelo de conjuntos nebulosos (*fuzzy sets*) ou de partição difusa (WOODBURY; CLIVE; GARSON JR., 1978), pois um mesmo estudante pode ter distintos graus de pertencimento a múltiplas categorias, diferentemente de outros métodos estatísticos, muitos dos quais assumem que esse estudante faz parte de um único grupo.

A proposta principal desse estudo é construir uma tipologia dos estudantes da Universidade de Brasília. Para isso, será aplicado o método *Grade of Membership* a partir de um conjunto de informações obtidas pelo Observatório da Vida estudantil (OVE), através da aplicação de um questionário sobre o perfil do estudante no momento do registro estudantil na Universidade.

2 OBJETIVOS

2.1 Objetivos Gerais

Traçar tipologias dos estudantes da Universidade de Brasília por meio do método *Grade of Membership*.

2.2 Objetivos Específicos

- Fazer análise descritiva dos dados, com foco nas variáveis sexo, renda, curso, forma e sistema de ingresso, raça, etc;
- Aplicar o método GoM para obter as tipologias, com base nas variáveis mais interessantes, utilizando o software R;
- Identificar como o perfil socioeconômico dos estudantes ingressantes na Universidade de Brasília pode refletir em algumas características estudantis por meio do método *Grade of Membership*.
- Verificar a adequação do método aos dados.

3 ACESSO AO ENSINO SUPERIOR

O número de matrículas nos cursos de graduação no Brasil cresceu consideravelmente nas últimas três décadas de acordo com os dados dos Censos da Educação Superior, passando de 1,5 milhão em 1991 para 8,04 milhões em 2016 (INEP, 2016).

A expansão do ensino superior foi estimulada por diversas políticas públicas, muitas das quais incluíam iniciativas que buscavam reduzir as desigualdades no acesso e na permanência de estratos sociais menos privilegiados na universidade.

Dentre essas políticas, pode-se destacar o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), que visa melhorar as condições necessárias para ampliação do acesso e permanência dos estudantes na educação superior das universidades federais. Assim, há consonância com as metas estabelecidas pelo Plano Nacional de Educação, de estabelecer o fornecimento da oferta de educação superior para, pelo menos, 30% dos jovens de 18 a 24 anos até 2020.

Além do REUNI, houve a implementação do Sistema de Cotas Sociais e Raciais (Lei nº 12.711/2012), que estabelece a reserva de 50% das matrículas em todos os cursos e turnos das instituições federais para estudantes oriundos de escolas públicas, segundo critérios socioeconômicos e étnicos-raciais.

Simultaneamente às transformações ocorridas no país, o ensino de graduação da Universidade de Brasília também sofreu algumas transformações. Uma delas foi a adoção de formas alternativas de ingresso: o Programa de Avaliação Seriada (PAS), implantado em 1996 e, mais recentemente, a adesão do Sistema de Seleção Unificada (Sisu), a partir do desempenho dos estudantes no Exame Nacional do Ensino Médio (ENEM). Além disso, houve a introdução do sistema de cotas raciais em 2004, a criação de novos campi (Planaltina em 2006, Ceilândia e Gama em 2008) e diversos cursos novos, e, mais recentemente, a implementação do Sistema de Cotas Sociais e Raciais.

Com auxílio dessas políticas de ampliação de acesso, houve um crescimento de 82,51% no número de ingressantes entre os anos de 2006 a 2016 na Universidade de Brasília. No 2º semestre de 2016 haviam 37.071 alunos regulares matriculados nos cursos de graduação, sendo que 27% desse valor corresponde ao número total de ingressantes nesse ano (INEP, 2016).

Diante desse contexto de transformações, surge o questionamento de como essas mudanças podem refletir nas universidades que aderiram a elas. As desigualdades ainda estão muito evidentes ou há indícios de um avanço nesse sentido? Quem são as pessoas que ingressam na Universidade? As condições socioeconômicas in-

3. ACESSO AO ENSINO SUPERIOR

fluenciam na escolha do curso? Existe diferença de perfil entre os cursos? Qual o perfil de maior vulnerabilidade social? E, por fim, como identificar esses perfis e as diferenças que podem estar presentes em cada um deles?

Deste modo, este estudo tem o propósito de responder algumas dessas perguntas, na dimensão da UnB, por meio do método *Grade of Membership*.

4 GRADE OF MEMBERSHIP

Dentre as alternativas metodológicas para a construção de tipologias, analisaram-se duas frentes importantes: a representada pelos conjuntos discretos, ou *crisp sets*, e a representada pelos conjuntos difusos, ou *fuzzy sets*. A maioria dos métodos estatísticos baseia-se no primeiro conjunto, que assume que um indivíduo pode pertencer ou não a um determinado grupo, enquanto a segunda alternativa admite que um mesmo indivíduo pode pertencer parcialmente a múltiplas categorias. Desse modo, optou-se pelo método Grade of Membership, que se baseia no segundo conjunto e ainda apresenta a vantagem de ser indicado para os casos em que há poucas observações e muitas variáveis (MANTON, 1994).

O modelo dos conjuntos difusos pode ser entendido como um sistema em que há conjuntos fechados, bem delineados e bem definidos (GUEDES et al, 2016). A situação que esse autor utiliza para exemplificar esse modelo é a classificação de vários indivíduos de acordo com sua altura em centímetros. O interesse é dividir os grupos da seguinte forma: baixa estatura (até 150 cm), estatura mediana (de 151 a 165 cm), alta estatura (de 166 a 190 cm) e muito altos (acima de 190 cm). No entanto, pode-se ter interesse nas pessoas que estão nos limites entre uma categoria e outra. Ao observar 2 casos, uma pessoa que tem 150 cm e outra que tem 151 cm, é realmente plausível dizer que elas são diferentes a ponto de serem classificadas em grupos diferentes? Essa distinção pode ser detectada a partir do modelo de conjuntos difusos. Nele, essas duas pessoas poderiam pertencer, em algum nível, aos dois grupos.

Ao trazer essa situação para a classificação dos estudantes da UnB, teria-se como objetivo identificar o grau de pertencimento ao grupo k de cada estudante i com base nas respostas de um questionário que possui J questões com L categorias.

O grau de pertencimento do i -ésimo estudante ao k -ésimo grupo - ou categoria - é medido através do escore de pertinência, ou escore GoM, denotado por g_{ik} . Os escores variam no intervalo $[0,1]$, sendo que 0 (zero) indica que um aluno i não pertence à categoria K , e 1 (um) indica total pertencimento ao k -ésimo grupo, ou seja, possui todas as características desse determinado grupo. Como o escore pode assumir valores entre 0 e 1, existe a possibilidade de um estudante pertencer parcialmente a mais de um perfil, ou seja, pode ocorrer a criação de perfis mistos. Esse escore de pertinência possui as seguintes restrições:

1. $g_{ik} \geq 0$, para cada i e j ;
2. $\sum_{k=1}^K g_{ik} = 1$, para cada i .

Além disso, considera-se que as respostas do questionário são representadas pelas variáveis aleatórias binárias Y_{ijl} , onde i se refere ao estudante i , j à questão e l à categoria de resposta. E são necessárias as seguintes suposições para a formulação do modelo e estimação dos parâmetros (escores):

1. As respostas dos diferentes estudantes são independentes;
2. Os valores de g_{ik} , $k = 1, 2, \dots, K$, são realizações das componentes do vetor aleatório $\delta_i = (\delta_{i1}, \dots, \delta_{iK})$ com função de distribuição $H(x) = P(\delta_i \geq x)$. Isso significa que os escores GoM são realizações de variáveis aleatórias quando um indivíduo é selecionado na população. Os escores na amostra fornecem estimativas da função de distribuição $H(x)$;
3. Se o grau de pertinência g_{ik} é conhecido, as respostas do estudante i para as várias questões Y_{ijl} são independentes para as categorias de cada variável;
4. A probabilidade de resposta l , para a j -ésima questão, pelo indivíduo com o k -ésimo perfil extremo, é dada pelo parâmetro λ_{kjl} . Por pressuposto do modelo, existe pelo menos um indivíduo que é um membro bem definido do k -ésimo perfil. Esse pressuposto dá a probabilidade de resposta, para este indivíduo, para os vários níveis de cada questão. Os valores de λ_{kjl} obedecem as seguintes restrições:
 - $\lambda_{kjl} \geq 0$, para cada j , k e l ;
 - $\sum_{l=1}^L \lambda_{kjl} = 1$, para cada k e j .
5. A probabilidade de uma resposta de nível l , da j -ésima questão, pelo indivíduo i , condicionada ao escore g_{ik} será dada por:

$$0 \leq P(Y_{ijl} = 1) = \sum_{k=1}^K (g_{ik} \lambda_{kjl}). \quad (1)$$

A construção do modelo de probabilidade para a estimação de máxima verossimilhança pode ser feita com base nos pressupostos mencionados. Ao considerar uma amostra aleatória, este modelo é o produto do modelo multinomial com a probabilidade de cada cédula:

$$E(Y_{ijl}) = \sum_{k=1}^K (g_{ik} \lambda_{kjl}) \quad (2)$$

Onde g_{ik} é, por pressuposto, conhecido e maior ou igual a 0 e menor ou igual a 1. A partir disso, o modelo de máxima verossimilhança pode ser escrito como:

$$L(y) = \prod_{i=1}^I \prod_{j=1}^J \prod_{l=1}^L \left(\prod_{k=1}^K g_{ik} \lambda_{kjl} \right)^{Y_{ijl}} \quad (3)$$

Com base na expressão (3) e considerando a probabilidade dada em (1), juntamente com as restrições mencionadas, é possível escrever a log-verossimilhança do modelo GoM como:

$$l(y) = \sum_i \sum_j \sum_l y_{ijl} \log \sum_k g_{ik} \lambda_{kjl} + \sum_i \gamma_i (1 - \sum_{gik}) + \sum_k \sum_j v_{kj} (1 - \sum_l \lambda_{kjl}) \quad (4)$$

Onde γ_i e v_{kj} são os multiplicadores de Lagrange.

4.1 Estimação do grau de pertencimento (g_{ik})

Os parâmetros gamma g_{ik} informam a distância de cada indivíduo em relação a cada um dos padrões de probabilidade de resposta. Para a obtenção dos estimadores de máxima verossimilhança desses parâmetros, é necessário buscar a função *score* correspondente a eles (CERQUEIRA, 2004). Para isso, deriva-se parcialmente a log-verossimilhança do modelo em (4), em relação aos parâmetros referentes ao grau de pertencimento, obtendo-se:

$$U(G_{ik}) = \frac{\partial l(y)}{\partial g_{ik}} \quad (5)$$

$$= \sum_j \sum_l \frac{y_{ijl}}{\sum_k g_{ik} \lambda_{kjl}} \lambda_{kjl} - \gamma_i \quad (6)$$

Depois de calcular a derivada parcial, deve-se igualar $U(g_{ik})$ a zero, multiplicar ambos os lados de (6) por g_{ik} e aplicar o somatório em k :

$$\sum_k \left[\sum_j \sum_l \frac{y_{ijl}}{\sum_k g_{ik} \lambda_{kjl}} \lambda_{kjl} g_{ik} - \gamma_i g_{ik} \right] = 0 \quad (7)$$

$$\begin{aligned} &= \sum_j \sum_l \frac{y_{ijl}}{\sum_k g_{ik} \lambda_{kjl}} \sum_k g_{ik} \lambda_{kjl} - \gamma_i \sum_k g_{ik} \\ &= \sum_j \sum_l y_{ijl} - \gamma_i \end{aligned} \quad (8)$$

Então o estimador de máxima verossimilhança será:

$$\hat{\gamma}_i = \sum_j \sum_l y_{ijl} \quad (9)$$

Substituindo (9) em (6) e colocando $\sum_j \sum_l y_{ijl}$ em evidência, teremos:

$$\sum_j \sum_l y_{ijl} \left(\frac{\lambda_{kjl}}{\sum_k g_{ik} \lambda_{kjl}} - 1 \right) \quad (10)$$

A resolução de (10) é encontrada de forma iterativa. O conjunto de parâmetros g_{ik} , por exemplo, é estimado inicialmente, mantendo-se o conjunto λ_{kjl} constante (GUEDES et al., 2016). Então, o primeiro conjunto de escores estimados, relativos ao grau de pertinência, é obtido pela expressão:

$$\hat{g}_{ik} = \frac{1}{y_{i++}} \sum_{i=1}^I \sum_{l=1}^L y_{ijl} \frac{g_{ik}^* \lambda_{kjl}^*}{p_{ijl}} \quad (11)$$

em que $y_{i++} = \sum_j \sum_l y_{ijl}$ e $p_{ijl} = \sum_k g_{ik}^* \lambda_{kjl}^*$

4.2 Estimação dos parâmetros estruturais ou locais (locacionais)

Os parâmetros locais λ_{kjl} descrevem padrões de probabilidade de resposta l , para a j -ésima questão, pelo indivíduo do k -ésimo grupo. Os estimadores de máxima verossimilhança desses parâmetros podem ser obtidos de forma semelhante aos EMV's de g_{ik} . Primeiro deriva-se parcialmente a log-verossimilhança do modelo em (3) em relação aos parâmetros estruturais:

$$U(\lambda_{kjl}) = \frac{\partial l(y)}{\partial \lambda_{kjl}} \quad (12)$$

$$\sum_i \frac{y_{ijl}}{\sum_k g_{ik} \lambda_{kjl}} g_{ik} - v_{kji} \quad (13)$$

Depois de encontrar a derivada, deve-se multiplicar (13) por λ_{kjl} , aplicar o somatório em i e igualar a zero:

$$\sum_i \left(\sum_l \frac{y_{ijl}}{\sum_k g_{ik} \lambda_{kjl}} g_{ik} \lambda_{kjl} - v_{kji} \lambda_{kjl} \right) = 0 \quad (14)$$

Então é possível obter:

$$v_{kj} = \sum_i \sum_l y_{ijl} \frac{g_{ik} \lambda_{kjl}}{\sum_k g_{ik} \lambda_{kjl}} \quad (15)$$

Ao substituir (15) em (13), encontra-se a expressão:

$$U(\lambda_{kjl}) = \sum_i \left[\frac{y_{ijl}}{\sum_k g_{ik} \lambda_{kjl}} - \frac{y_{ij+}}{\sum_l g_{ik} \lambda_{kjl}} \right] \quad (16)$$

Como feito anteriormente, aplica-se o algoritmo numérico para a obtenção do conjunto complementar de escores, utilizado para determinar os parâmetros λ_{kjl} :

$$\hat{\lambda}_{kjl} = \frac{\sum_i y_{ijl} \frac{g_{ik}^* \lambda_{kjl}^*}{P_{ijl}}}{\sum_i y_{ij} + \sum_{l=1} \frac{g_{ik}^* \lambda_{kjl}^*}{P_{ijl}}} \quad (17)$$

Em ambos os conjuntos, g_{ik}^* e λ_{kjl}^* são os estimadores dos parâmetros de uma iteração anterior. Essas iterações devem parar quando $L(Y)$ atingir o seu valor máximo.

4.3 Caracterização e descrição dos perfis extremos

“Perfil extremo” pode ser entendido como cada padrão de probabilidades de respostas (PINTO, 2013). Para a caracterização desses perfis, segue-se 2 etapas:

1. Definição do número ótimo de perfis extremos; e
2. Definição do critério de preponderância.

4.3.1 Critérios para a definição do número ótimo de perfis extremos

Existem diferentes critérios para a definição do número ideal de perfis extremos, mas este estudo utilizará três dos métodos quantitativos mais utilizados nos estudos empíricos que se baseiam no GoM (GUEDES et al., 2016): o teste da razão das verossimilhanças, o critério de informação de Akaike e o Critério de Informação Bayesiano (BIC). Além disso, considerou-se a relevância substantiva conjuntamente com os resultados desses métodos para a seleção do melhor modelo.

4.3.1.1 Teste da Razão das Verossimilhanças

A escolha do teste da razão das verossimilhanças para a identificação do número ótimo de perfis foi sugerida por Manton et al (2006). Este teste requer que

o modelo com k perfis extremos seja um subconjunto do modelo $k + 1$, significando que todas as variáveis internas em k devem estar contidas em $k + 1$. A estatística do teste em questão pode ser obtida pela fórmula

$$D = -2\ln \frac{L(\text{Modelo}_{k+1})}{L(\text{Modelo}_k)} \sim \chi^2_{(p_{k+1}-p_k)}, \quad (18)$$

onde L é o valor da verossimilhança e p é o número de parâmetros estimados do modelo, sendo que D segue uma distribuição com os graus de liberdade igual à diferença entre o número de parâmetros a serem estimados entre os modelos (ANDREOTTI et al., 2009).

O GoM gera uma grande quantidade de parâmetros, e, para corrigir isso, este critério penaliza o número excessivo de parâmetros de um modelo de uma dada dimensão.

4.3.1.2 Critério de Informação de Akaike (AIC)

O outro critério para identificação do número ideal de perfis extremos é o Critério de Informação de Akaike. A regra de decisão corresponde ao modelo com menor valor da estatística AIC, o que representa o modelo que melhor descreve a estrutura latente dos dados. A estatística AIC pode ser obtida pela fórmula

$$AIC = 2p - 2\ln L \quad (19)$$

Onde L é o valor da verossimilhança e p é o número de parâmetros estimados do modelo. Este modelo deve atender à restrição $AIC_{(k-1)} > AIC_k$.

4.3.1.3 Critério de Informação Bayesiano (BIC)

O Critério de Informação Bayesiano (BIC) também é bastante utilizado na literatura. Ele possui a fórmula

$$BIC = -2\ln(L) + p\ln(n) \quad (20)$$

sendo L o valor máximo da função de verossimilhança, p o número de parâmetros estimados e n o tamanho da amostra.

4.3.2 Critério de preponderância

A preponderância é utilizada para o estabelecimento de um critério objetivo para a descrição dos perfis gerados de acordo com suas características predominantes. O critério de preponderância utilizado neste estudo foi sugerido por Sawyer, Leite e Alexandrino (2002). Ele é calculado pela razão entre a frequência marginal estimada e a frequência marginal observada:

$$RLFM = \frac{\lambda_{kjl}}{(n_{kjl})^{-1}N_{kj}}, \quad (21)$$

em que RLFM significa Razão Lambda/Frequência Marginal¹, N_{kjl} é o número de observações na categoria l e N_{kj} é o número de observações na variável j .

O valor de corte para a descrição desses perfis é arbitrário e depende do grau de heterogeneidade que se queira captar da amostra. Para isso, deve-se levar em consideração que quanto maior o seu valor, mais restritivo será o processo de inclusão de categorias das variáveis a um determinado perfil. O valor de corte sugerido pelos autores citados anteriormente é de 20%, indicando que a probabilidade de ocorrência de uma l -ésima resposta a uma j -ésima variável em um k -ésimo perfil deve ser pelo menos 20% superior à probabilidade de ocorrência dessa mesma resposta l na amostra. Assim, somente as características marcantes em relação aos dados amostrais serão destacadas como caracterizadoras dos tipos puros dos perfis extremos.

4.4 Criação de tipos mistos a partir dos perfis extremos

Como a estimação pela técnica GoM permite que o indivíduo pertença parcialmente a múltiplas categorias, os indivíduos que não se enquadrarem integralmente nos perfis extremos serão reagrupados em grupos que possuam características mistas desses perfis extremos.

Os tipos mistos são conjuntos de elementos que possuem características de um determinado perfil combinadas com um conjunto secundário de características de outro perfil extremo.

O reagrupamento pode ser realizado a partir de expressões *booleanas* que definem o grau de manifestação de cada perfil específico na construção final do grupo (GUEDES et al, 2016).

¹Alguns autores, como Assis, Machado e Rodrigues (2008), também chamam de Razão E/O.

4.5 Classificação dos perfis

4.5.1 Modelo de três perfis extremos

Existem vários critérios de classificação dos grupos baseados nos escores de pertencimento. Considerando o modelo de 3 perfis ($K = 3$), temos, como por exemplo, a classificação de Melo (2007); Guedes, Costa e Brondízio (2009); e Guedes (2010), que serão apresentados a seguir.

1. Critério de classificação de Melo (2007):

- Tipo puro do perfil extremo k :

a) preponderância muito alta de k :

$$g_{ik} \geq 0,75 \text{ com } k = 1, 2, 3$$

b) preponderância alta de k :

$$0,5 < g_{ik} < 0,75 \cap (g_{im} \leq 0,25 \cap g_{in} \leq 0,25) \text{ com } k \neq m \neq n.$$

- Tipo misto de dois perfis extremos, em que predominam as características de k :

$$0,5 < g_{ik} < 0,75 \cap (0,25 \leq g_{iz} < 0,5) \text{ com } k \neq z = m, n.$$

- Tipo amorfo:

a) sem predominância total:

$$(g_{ik} < 0,5) \cap (g_{im} < 0,5) \cap (g_{in} < 0,5) \text{ com } k \neq m \neq n.$$

b) sem predominância pareada:

$$(g_{ik} = 0,5) \cap (g_{iz} = 0,5) \text{ com } k \neq z = m, n.$$

2. Critério de classificação de Guedes, Queiroz e VanWey (2009):

- Tipo puro do perfil extremo k :

$$g_{ik} = 1 \text{ com } k = 1, 2, 3.$$

- Tipo misto com predominância de k :

a) nível 1:

$$0,8 \leq g_{ik} < 1 \text{ com } k = 1, 2, 3$$

b) nível 2:

$$(0,6 \leq g_{ik} < 0,8) \cap (g_{im} + g_{in} \leq 0,4) \text{ com } k \neq m \neq n$$

- Tipo misto com predominância:

$$(0, 6 \leq g_{ik} < 0, 8) \cap (0, 3 \leq g_{iz} \leq 0, 4) \text{ com } k \neq z = m, n$$

- Tipo misto sem predominância:

$$(0, 4 \leq g_{ik} < 0, 5) \cap (0, 4 \leq g_{iz} \leq 0, 5) \text{ com } k \neq z = m, n$$

∪

$$(0, 4 \leq g_{iz} < 0, 5) \cap (0, 4 \leq g_{ik} \leq 0, 5) \text{ com } k \neq z = m, n$$

3. Critério de classificação de Guedes, Costa e Brondízio (2009):

- Tipo puro do perfil extremo k :

$$g_{ik} = 1 \text{ com } k = 1, 2, 3.$$

- Tipo misto com alta preponderância absoluta:

$$0, 7 < g_{ik} < 1 \text{ com } k = 1, 2, 3.$$

- Tipo misto com alta preponderância relativa:

$$(0, 5 \leq g_{ik} \leq 0, 7) \cap (0, 1 \leq g_{iz} \leq 0, 4) \text{ com } k \neq z = m, n.$$

- Tipo misto com preponderância pareada relativa:

$$g_{ik} + g_{iz} \geq 0, 66 \cap [(g_{ik} < 1, 5g_{iz}) \cup (g_{iz} < 1, 5g_{ik})] \text{ com } k \neq z = m, n.$$

4. Critério de classificação de Guedes (2010):

- Tipo puro do perfil extremo k :

$$g_{ik} \geq 0, 75 \text{ com } k = 1, 2, 3.$$

- Tipo misto:

$$(0, 5 \leq g_{ik} < 0, 75) \cap (0, 25 < g_{im} \leq 0, 5) \cap (g_{in} < 0, 25) \text{ com } k \neq m \neq n.$$

- Tipo amorfo:

$$g_{ik} < 0, 5 \text{ com } k = 1 \cup 2 \cup 3.$$

Independentemente do número de perfis delineados no estudo, tem-se que os indivíduos não classificados pelos critérios sugeridos são denominados de tipo amorfos. Então, indivíduos que recebem essa classificação não possuem características significativas de qualquer perfil extremo que predomine sobre sua classificação.

4.5.2 Critério de agrupamento pela moda

Neste critério, o indivíduo será designado para o perfil extremo que apresentar o maior valor modal dos graus de pertencimento, então:

$$- i \in k_1 \text{ se } g_{i1} = \max(g_{ik})$$

- $i \in k_2$ se $g_{i2} = \max(g_{ik})$

⋮

- $i \in k_n$ se $g_{in} = \max(g_{ik})$

Com $k = 1, 2, \dots, n$.

Se ocorrer uma relação bimodal, uma das soluções é classificar esses casos como um perfil amorfo.

4.5.3 Critério de agrupamento pela preponderância relativa

Este critério considera que um indivíduo pertence a um determinado perfil extremo k se o grau de pertencimento daquele perfil é maior que a somatória dos demais:

- $i \in k_1$ se $g_{i1} > \sum g_{ik}$, com $k \neq 1$

- $i \in k_2$ se $g_{i2} > \sum g_{ik}$, com $k \neq 2$

- $i \in k_n$ se $g_{in} > \sum g_{ik}$, com $k \neq n$

4.5.4 Validação dos tipos mistos criados

Nem sempre os critérios citados anteriormente vão fornecer o melhor resultado de classificação dos elementos heterogêneos. Dessa forma, é importante testar a validade interna dos tipos mistos criados. Este teste pode ser feito a partir da aplicação do teste de média das variáveis internas entre cada perfil extremo e tipos mistos e entre os próprios tipos mistos.

Para as variáveis contínuas que foram categorizadas para a aplicação da técnica GoM, utilizou-se o teste t para comparação de médias de dois grupos independentes. Este teste compara duas populações com variâncias desconhecidas e possui as seguintes hipóteses

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad (22)$$

Para o caso das variáveis que originalmente eram qualitativas nominais, utilizou-se o teste para comparação entre duas proporções com as seguintes hipóteses

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases} \quad (23)$$

Desta maneira, este estudo utilizará a avaliação sugerida por Guedes (2010). Ela determina que pelo menos 70% das médias das variáveis entre dois perfis devem ser significativamente diferentes, o que deve justificar a existência dos

4. *GRADE OF MEMBERSHIP*

perfis isoladamente. Caso isso não aconteça, o autor aconselha que esses perfis sejam agregados e que sejam feitos novos testes de média.

5 BASE DE DADOS

5.1 Pesquisa sobre o perfil dos estudantes da Universidade de Brasília

Este estudo utilizou os dados provenientes da Pesquisa sobre o perfil dos estudantes da UnB, realizada pelo Observatório da Vida Estudantil (OVE) e apoiada pelo Laboratório de População e Desenvolvimento (LPD), ambos vinculados ao Centro de Estudos Avançados Multidisciplinares (CEAM).

O Observatório tem como objetivo conhecer o corpo discente da Universidade de Brasília em seus diferentes aspectos e obter informações que permitam subsidiar políticas institucionais. Nesse sentido, está construindo um banco de dados sobre os estudantes desde 2012, em que abrange os seguintes aspectos: origem social, formação escolar anterior, expectativas de inserção profissional e características sociodemográficas.

O levantamento de dados sobre o perfil estudantil é realizado a partir de um questionário aplicado na etapa de registro estudantil. Este instrumento de coleta é composto por cerca de 60 perguntas e é dividido em 10 blocos temáticos que abrangem os fatores citados acima.

A aplicação do questionário teve seu início no primeiro semestre de 2012. Todo semestre, na etapa de registro estudantil, esse questionário é disponibilizado via internet para os estudantes preencherem e é obrigatório para completar o processo de registro na Universidade.

A base de dados é composta pelos estudantes que ingressaram na UnB no ano de 2017, em ambos os semestres, correspondendo a 7.571 respondentes no total.

5.2 Variáveis do estudo

Para as análises deste estudo, utilizou-se 14 variáveis relacionadas a características socioeconômicas, modalidade de ingresso, curso, categorizadas da seguinte forma (Quadro 1):

- Sexo: Masculino - Feminino;
- Raça/Cor da pele: Brancos e Amarelos - Pretos e Pardos;
- Sistema de ingresso na Universidade: Cotas raciais - Cotas Sociais - Sistema Universal;

5. BASE DE DADOS

- Classe Socioeconômica (Critério Brasil): A (45 a 100 pontos) - B (29 a 44 pontos) - C (17 a 28 pontos) - D (0 a 16 pontos);
- Escolaridade do pai: Até o ensino fundamental - Ensino médio completo - Ensino superior completo;
- Escolaridade da mãe: Até o ensino fundamental - Ensino médio completo - Ensino superior completo;
- Categoria administrativa da instituição de ensino médio: Escola pública - Escola Particular;
- Classificação dos cursos de acordo com o prestígio: Alto - Médio - Baixo;
- Renda mensal da família: até 3 Salários Mínimos - 3 a 10 Salários Mínimos - 10 a 20 Salários Mínimos - 20 ou mais Salários Mínimos;
- Realização de curso preparatório: Sim - Não;
- Número de tentativas de ingresso: Primeira tentativa - Uma ou mais - Já fez outros cursos;
- Modalidade de ingresso: ENEM - PAS - Vestibular;
- Trocaria de curso: Sim - Não; e
- Campus: Ceilândia - Darcy Ribeiro - Gama - Planaltina.

Quadro 1: Variáveis e Categorias.

Variáveis	Categorias	Variáveis	Categorias
Escolaridade do pai	Até o ensino Fundamental Completo	Sistema de ingresso na Universidade	Cotas raciais
	Ensino médio completo		Sistema Universal
	Ensino superior completo		Cotas sociais
Escolaridade da mãe	Até o ensino fundamental Completo	Raça/ Cor da pele	Amarelos e Brancos
	Ensino médio completo		Pretos e Pardos
	Ensino superior completo	Classe Socioeconômica (Critério Brasil)	A (45 a 100 pontos)
Tipo de ensino médio	Escola Pública		B (29 a 44 pontos)
	Escola Particular		C (17 a 28 pontos)
Prestígio do curso	Alto		D (0 a 16 pontos)
	Médio	Sexo	Masculino
	Baixo		Feminino
Renda mensal da família	Até 3 SM	Modalidade de Ingresso	ENEM
	3 a 10 SM		PAS
	10 a 20 SM		Vestibular
	Mais de 20 SM	Número de tentativas	Primeira vez
Campus	Darcy Ribeiro		1 ou mais
	Gama		Já fez outros cursos
	Ceilândia	Cursinho/Trocaria de curso	Sim
	Planaltina		Não

Para a criação das tipologias do presente estudo, considerou-se apenas as variáveis relacionadas a características socioeconômicas: sexo, raça, escolaridade dos pais, tipo de ensino médio, realização de curso preparatório, renda mensal da família, classe socioeconômica e sistema de ingresso.

As observações que possuíam algum valor faltante em pelo menos uma das categorias, foram descartadas para que houvesse uma maior extração de informação das análises por meio da aplicação do método GoM. Desta forma, a base final foi composta por 6.351 estudantes.

5.2.1 Critério Brasil

A Associação Brasileira de Empresas de Pesquisa (ABEP) estabeleceu um critério utilizado por muitas empresas de pesquisa para classificar os estratos sociais, chamado de Critério Brasil. Ele foi construído com base na posse de bens duráveis (carro, TV, geladeira, etc), nas condições de moradia (quantidade de banheiros na casa), na oferta de serviços públicos (água encanada, rua pavimentada, etc) e no grau de instrução do chefe da família. A partir dessas informações, atribui-se uma pontuação para estimar a renda média da família.

Diante desse contexto e tendo como base o estudo de Costa (2015), a variável Classe Socioeconômica foi criada de acordo com as pontuações atribuídas às respostas sobre essas temáticas contidas no questionário. O agrupamento dessa pontuação varia entre zero e cem e define as classes socioeconômicas dos estudantes, como mostra o Quadro 2.

Quadro 2: Renda média domiciliar em cada classe socioeconômica segundo o Critério Brasil - ABEP, 2015

Pontuação com base nos bens	Estrato Sócio-econômico	Renda média domiciliar R\$
45 a 100	A	Mais de 20 mil
29 a 44	B	de 4 até 20 mil
17 a 28	C	de 1 até 4 mil
0 a 16	D	até mil

5.2.2 Sistema de Ingresso

Para avaliar se existe diferenças entre os perfis de acordo com as formas de ingresso, fez-se a seguinte categorização:

- Sistema de cotas raciais: compreende os estudantes que entraram na UnB por meio do sistema de cotas raciais e os que entraram por meio do sistema de cotas para escolas públicas e raciais (com renda inferior ou superior a 1,5 salários mínimos);
- Sistema de cotas sociais: alunos que ingressaram por meio do sistema de cotas para escolas públicas (com renda superior ou inferior a 1,5 salários mínimos); e
- Sistema universal: estudantes que ingressaram pelo sistema universal.

5.2.3 Cursos de prestígio

Para estudar as diferenças no perfil socioeconômico dos estudantes em função do curso, fez-se uma análise da demanda e da nota de corte dos processos seletivos anteriores.

Diante dessa análise, os cursos foram classificados da seguinte forma:

- Maior prestígio: Medicina, Direito, Relações Internacionais, as Engenharias, Odontologia, Psicologia, Comunicação Social e Ciências Econômicas.
- Menor prestígio: Biblioteconomia, Pedagogia, Serviço Social, Letras e Matemática.
- Médio prestígio: demais cursos.

5.2.4 Escolaridade dos pais

Quanto à escolaridade dos pais, considerou-se 3 categorias:

- Até o ensino fundamental completo: compreende os pais que não sabem ler nem escrever até pais que chegaram no nível médio mas não completaram;
- Ensino médio completo: pais que completaram o ensino médio até os pais que começaram a fazer o ensino superior mas não completaram; e
- Ensino superior completo: pais que completaram o ensino superior até os pais que possuem pós-graduação.

5.2.5 Raça/cor da pele

Para a variável raça/cor da pele, agregou-se brancos com amarelos e pretos com pardos. Como a porcentagem de indígenas representa um valor muito baixo, essa categoria foi desconsiderada.

5.3 Análise dos dados

Para a implementação do método GoM, utilizou-se *Software* livre R para a execução do arquivo R-script *GoMRcpp*², desenvolvido por Julimar Pinto e André Caetano da Universidade Católica de Minas Gerais (2013).

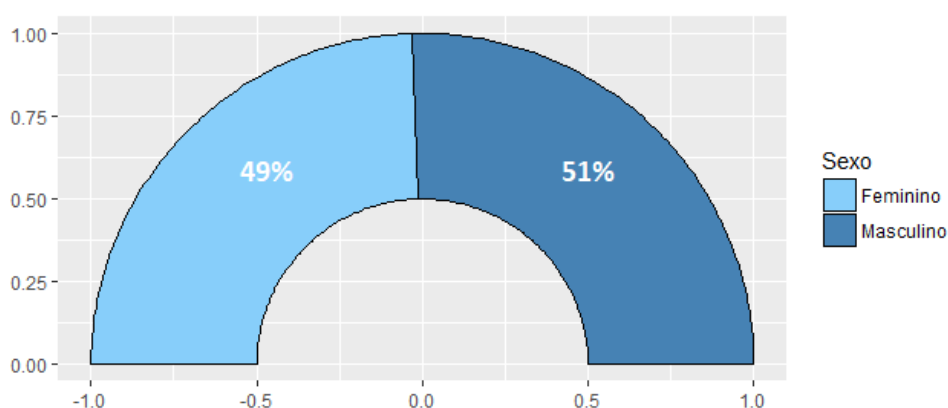
²Para mais informações sobre essa programação, acesse <http://gilvanguedes.com/publications/gombook/>.

6 RESULTADOS

6.1 Retrato socioeconômico dos estudantes que ingressaram na UnB em 2017

Dos alunos que ingressaram em 2017 nos cursos de graduação, há pouca diferença em relação ao sexo, sendo 51% do sexo masculino e 49% do feminino (Figura 1).

Figura 1: Percentual de ingressantes na UnB segundo sexo - 2017.



Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Ao observar a raça/cor dos ingressantes (Tabela 1), tem-se que apenas 2,31% se declarou como amarelo e 10,77% o fizera como preto. Como brancos (45,15%) e pardos (41,82%) representam a grande maioria e o sistema de cotas abrange pardos e pretos, optou-se por trabalhar com a agregação das categorias: amarelos/brancos e pretos/pardos.

Tabela 1: Percentual de ingressantes na UnB segundo raça/cor da pele - 2017.

Raça/ Cor da pele	Frequência	Percentual
Amarelos	147	2,31%
Branco	2.868	45,15%
Pardos	2.656	41,82%
Pretos	680	10,72%

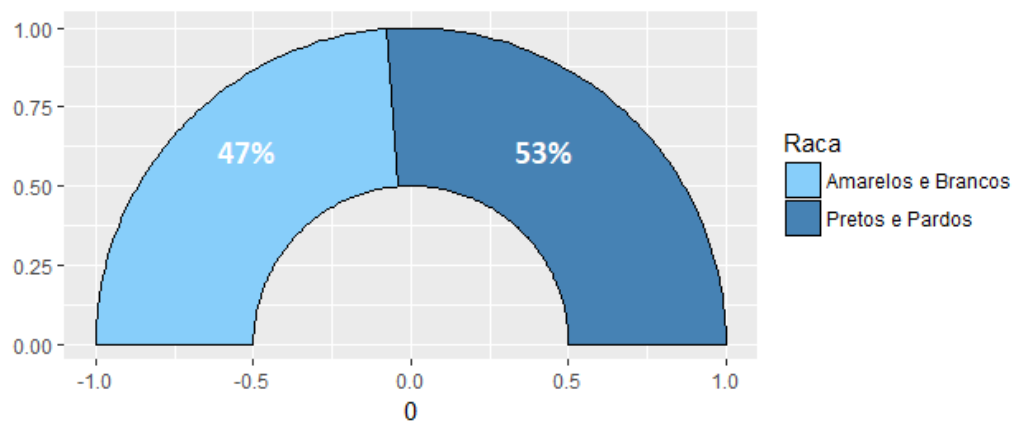
Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017.

Portanto, 53% se declararam pretos ou pardos e 47% amarelos ou brancos

6. RESULTADOS

(Figura 2).

Figura 2: Percentual de ingressantes na UnB segundo raça/cor da pele - 2017.

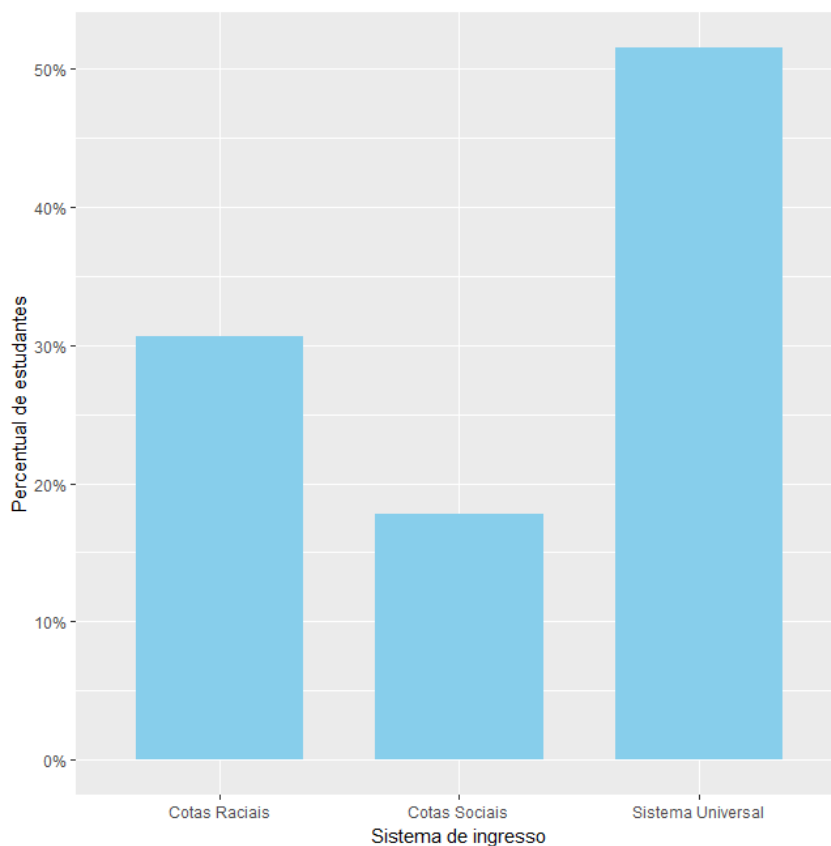


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Quanto ao sistema de ingresso, a maioria dos estudantes ingressaram na universidade por meio do sistema universal (52%), enquanto 31% pelo sistema de cotas raciais e apenas 18% pelo sistema de cotas sociais (Figura 3).

6. RESULTADOS

Figura 3: Percentual de ingressantes na Universidade de Brasília segundo sistema de ingresso - 2017.



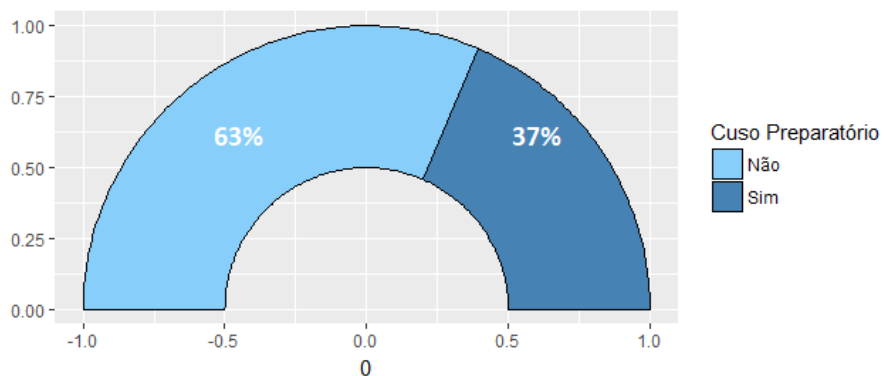
Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Por causa da implementação do sistema de cotas sociais, que estabelece que 50% das vagas devem ser destinadas a alunos de escolas públicas, não há diferença entre o número de alunos oriundos de escolas públicas e particulares.

Ao analisar se estes alunos fizeram ou não curso preparatório, pode-se observar que a grande maioria não fez (63%), enquanto apenas 37% fizeram cursinho (Figura 4).

6. RESULTADOS

Figura 4: Percentual de ingressantes na UnB segundo realização de curso preparatório - 2017.

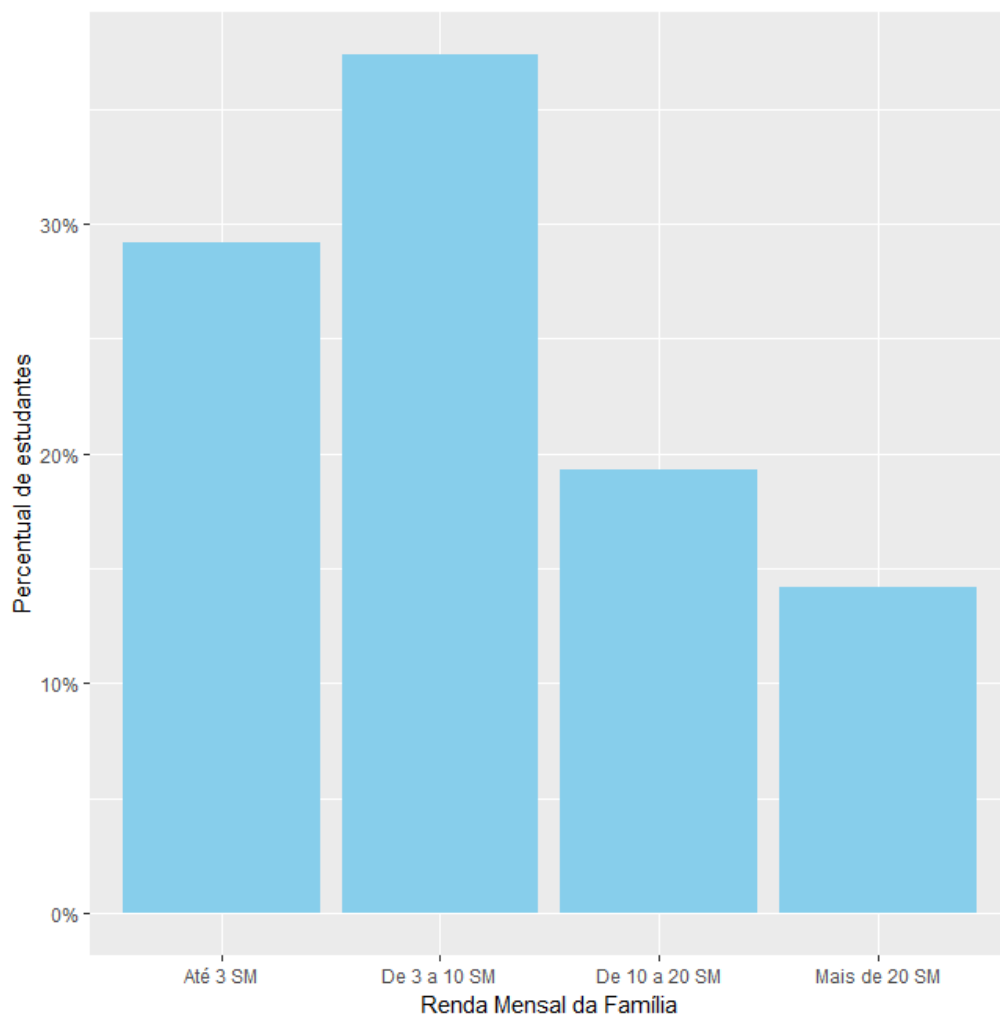


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Quanto à renda, fez-se uma análise da renda mensal da família e da variável criada com base no Critério Brasil de 2015. A Figura 5 mostra que a maioria dos estudantes possui renda entre 3 e 10 salários mínimos (37%), enquanto a minoria (14%) recebe mais de 20 salários mínimos. 29% faz parte da classe mais baixa, que recebe até 3 salários mínimos e, 19% recebe entre 10 e 20.

6. RESULTADOS

Figura 5: Percentual de ingressantes na UnB segundo renda mensal da família - 2017.

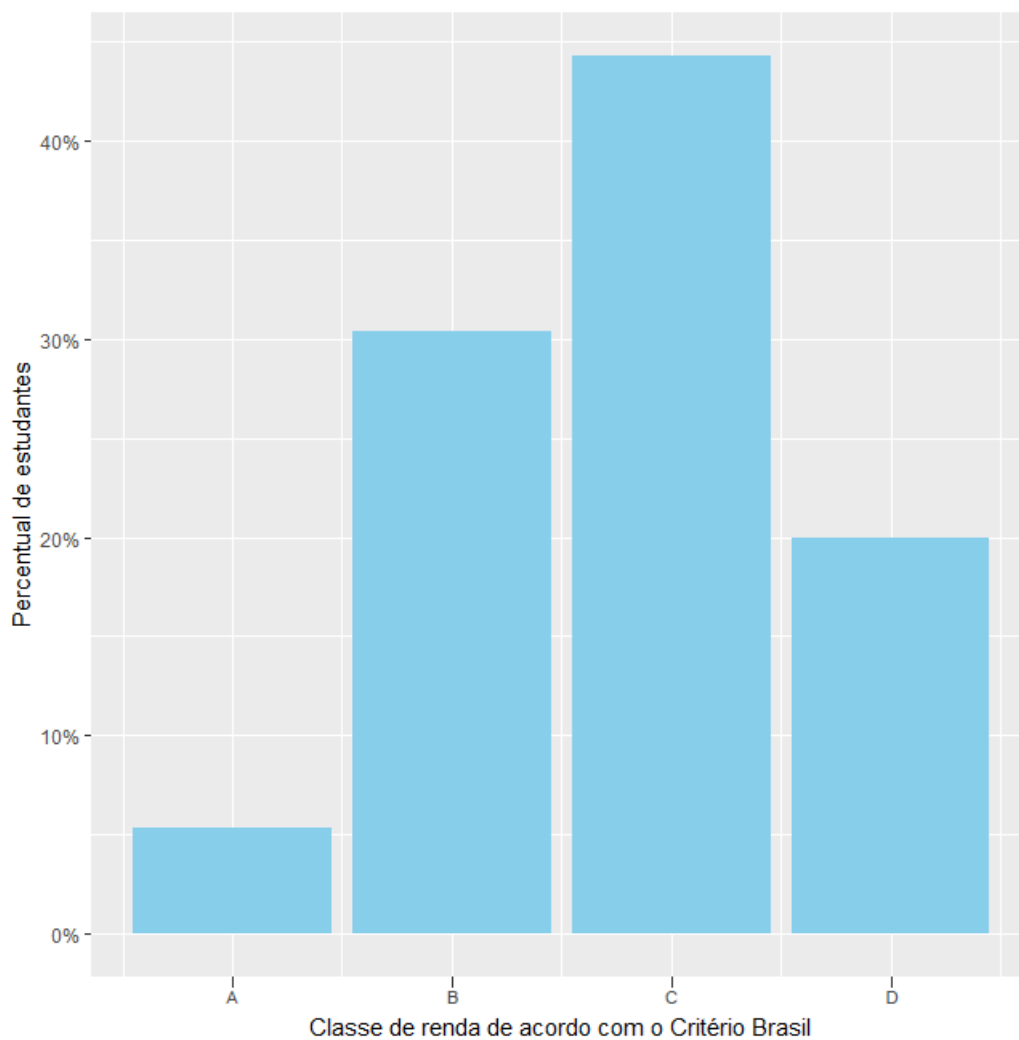


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

A Figura 6 mostra a classificação com base no Critério Brasil. Observa-se que a maioria (44%) faz parte da classe C, e a minoria da classe A (5%), que representa a classe mais alta. A classe B é representada por 30%, e a classe D por 20%.

6. RESULTADOS

Figura 6: Percentual de ingressantes na UnB segundo a classificação do Critério Brasil - 2017.

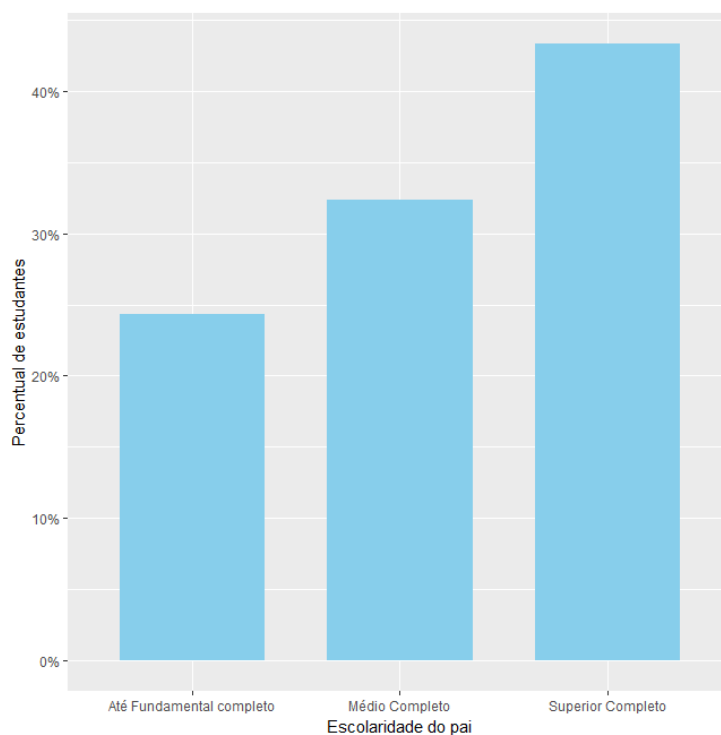


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

A escolaridade dos pais é outra variável importante para o estudo. É um dos maiores determinantes do sucesso educacional dos filhos, pois está diretamente relacionada com a renda, a qual auxilia no investimento educacional desses estudantes (BARROS et al, 2001). As figuras 7 e 8 mostram a relação de distribuição do último nível de escolaridade atingido pelos pais dos alunos. Observa-se que, em ambos os casos, a maioria dos alunos possui pais com nível de ensino superior completo, depois ensino médio completo e, por último, ensino fundamental completo.

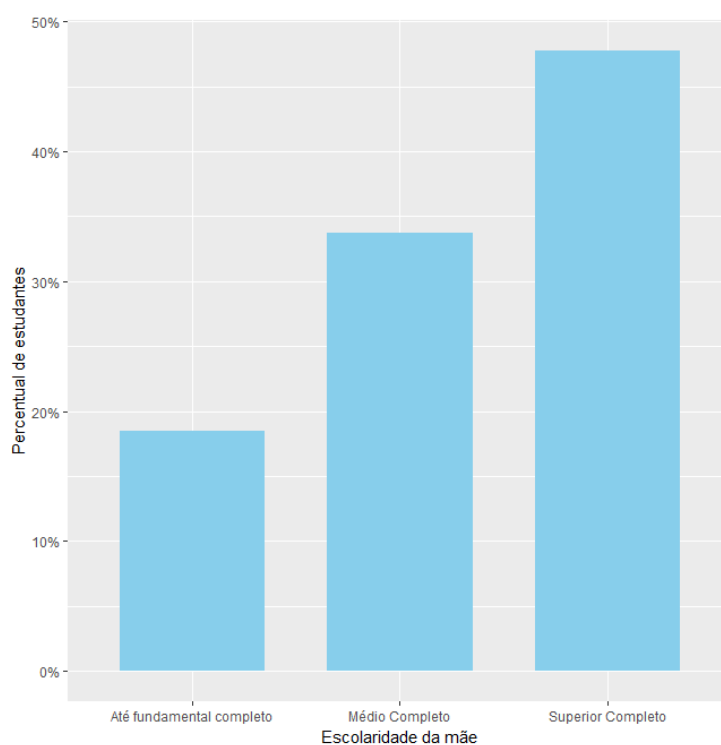
6. RESULTADOS

Figura 7: Percentual de ingressantes na Universidade de Brasília segundo escolaridade do pai - 2017.



Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Figura 8: Percentual de ingressantes na Universidade de Brasília segundo escolaridade da mãe - 2017.



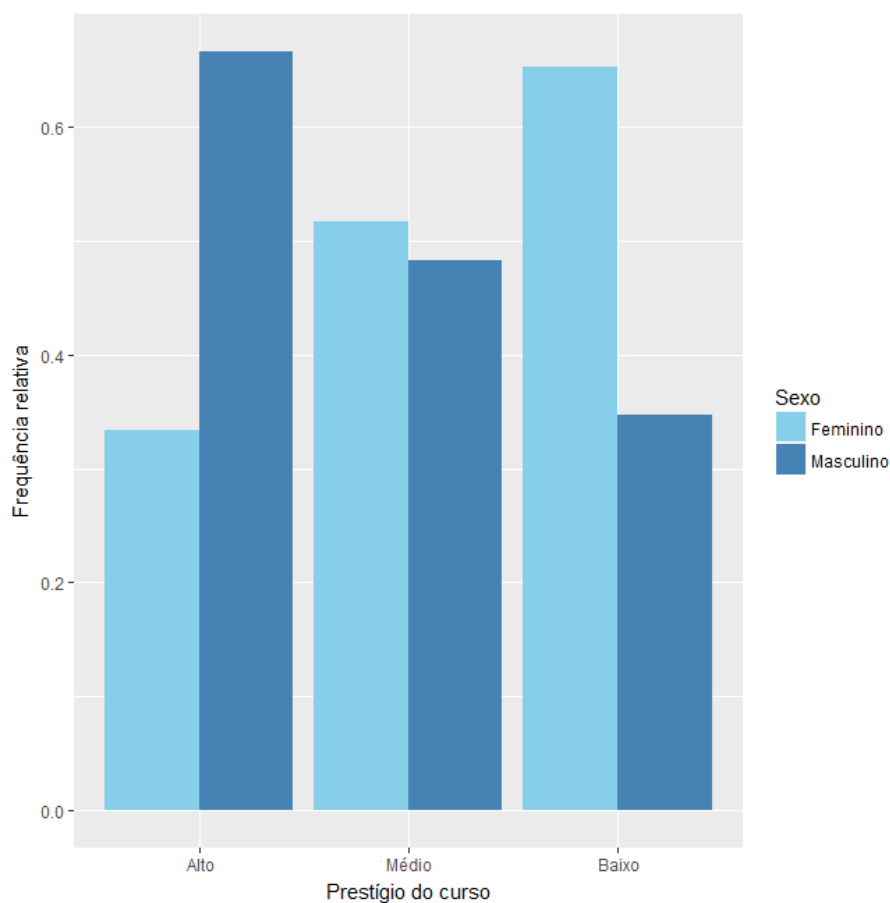
Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

6. RESULTADOS

Ao observar a distribuição dessas variáveis selecionadas como principais para a análise do perfil socioeconômico segundo as categorias de prestígio do curso, pode-se obter alguns resultados interessantes.

O sexo é uma das variáveis mais controversas quando é comparada com o curso. Mesmo apresentando distribuição uniforme com relação ao total de ingressantes, nota-se que mais de 60% dos estudantes do sexo masculino fazem parte dos cursos de alto prestígio, enquanto essa relação é invertida para mulheres que ingressaram em cursos de baixo prestígio (Figura 9).

Figura 9: Proporção de ingressantes na UnB segundo sexo e prestígio do curso - 2017.



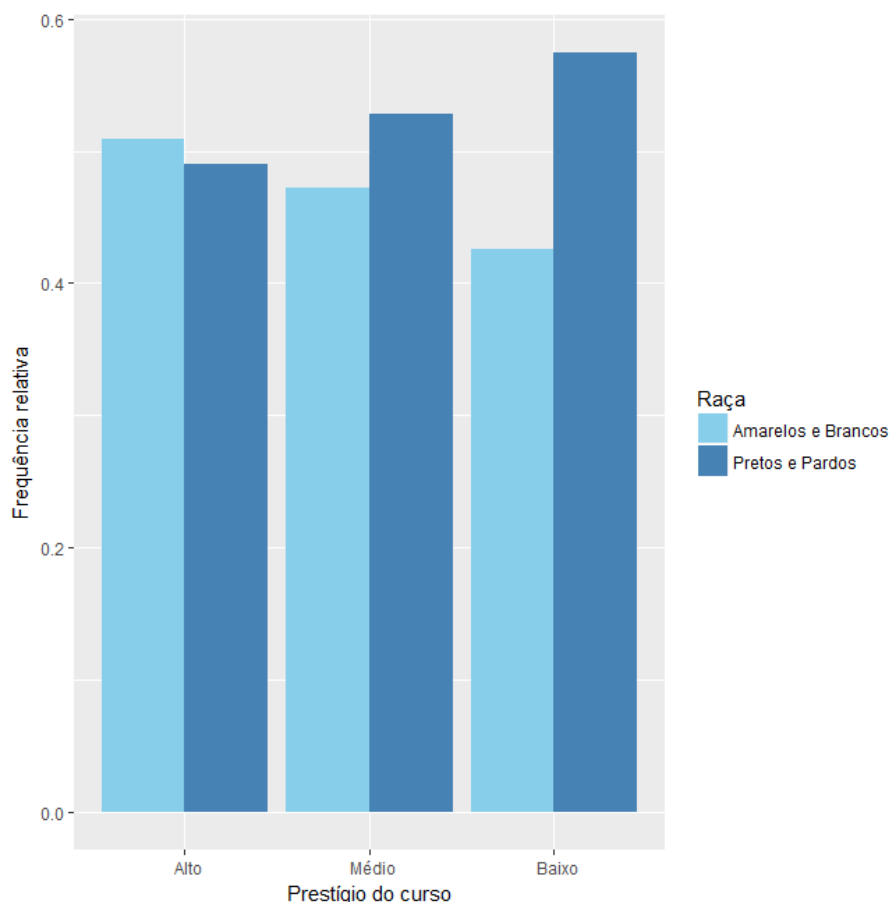
Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

No que se refere à raça, observa-se na Figura 10 que não há muita diferença entre brancos e negros na distribuição dos cursos, mesmo que tenham mais brancos (51%) nos cursos de alto prestígio. O que chama mais atenção é a distribuição nos cursos de baixo prestígio, já que quase 60% dos estudantes que entraram para os cursos dessa categoria são representados por pretos e pardos. Uma possível justificativa para a distribuição equilibrada na categoria de alto prestígio é a implementação

6. RESULTADOS

do sistema de cotas raciais.

Figura 10: Proporção de ingressantes na UnB segundo raça/cor e prestígio do curso - 2017.

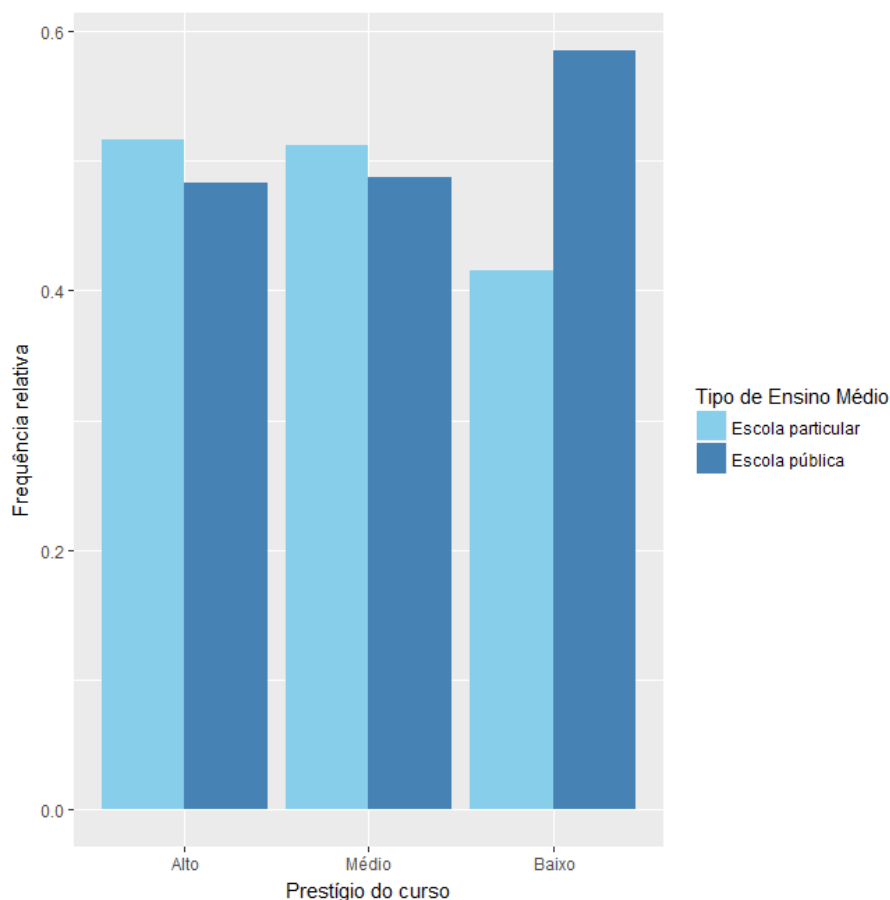


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

A categoria administrativa da escola de formação do aluno no ensino médio é uma variável importante para a análise, pois funciona como um indicador de qualidade dos insumos educacionais, os quais podem fazer a diferença na aprovação dos candidatos (CESAR, 2013). Portanto, com a implementação das cotas sociais, espera-se que haja um equilíbrio na distribuição do prestígio segundo o tipo de escola, que é o que acontece com os cursos mais prestigiados. No entanto, com relação aos cursos menos prestigiados, observa-se que (59%) desses estudantes são oriundos de escola pública (Figura 11).

6. RESULTADOS

Figura 11: Proporção de ingressantes na UnB segundo o tipo de ensino médio e prestígio do curso - 2017.



Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Diante desses resultados, pode-se ter uma ideia das principais características socioeconômicas dos ingressantes de 2017 e de como elas podem contribuir para a escolha do curso, por exemplo. Porém, pertencer a um fator específico de uma das variáveis analisadas (estudar em escola particular, por exemplo) pode não determinar totalmente o pertencimento a outra variável (possuir renda alta, por exemplo). Em outras palavras, um fator isoladamente pode não influenciar na caracterização de um perfil.

Desta forma, a combinação dos efeitos das características traduzem o efeito conjunto dessas variáveis que podem contribuir para elucidar as relações entre alguma característica socioeconômica e a seleção de curso, por exemplo. Portanto, o objetivo principal desse estudo é analisar todas essas variáveis em conjunto, para entender como elas contribuem para o perfil socioeconômico dos estudantes, bem como se esses perfis podem influenciar em outros fatores relacionados ao curso, modalidade de ingresso, campus, etc.

6.2 Identificando perfis: *Grade of Membership*

6.2.1 Tipos puros

Compreender os perfis sociodemográficos dos estudantes da UnB é essencial para identificar fatores que podem influenciar na trajetória acadêmica. Desta forma, a identificação desses perfis é importante no auxílio de políticas que gerem oportunidades e incentivos, principalmente para os estudantes mais vulneráveis.

Para a análise, considerou-se 2 variáveis relacionadas com o estrato socioeconômico dos estudantes. A renda mensal da família é uma medida que depende do conhecimento do respondente sobre o quanto os pais - ou representantes da família - ganham e nem sempre essa informação é de fácil acesso, refletindo em respostas imprecisas. Já o critério Brasil atribui pontos com base na posse de alguns bens estipulados e estima a renda média do domicílio. Portanto, é mais fácil para os estudantes responderem a quantidade de banheiros presentes na casa, por exemplo, do que a própria renda da família. Como a primeira variável tende, em geral, a subestimar a renda, o critério Brasil pode ser utilizado para proporcionar um equilíbrio, refletindo em resultados mais precisos.

Desse modo, construiu-se um modelo com 3 perfis. Esse número ótimo foi definido com auxílio dos métodos quantitativos (Tabela 2) conjuntamente com a relevância substantiva (teórica, qualitativa). O valor do AIC quando $k = 3$ foi bem próximo do $k = 2$, e a escolha de 2 perfis pode resultar em uma agregação excessiva de estudantes de modo a não mostrar as diferenças entre eles, portanto a utilização de $k = 3$ se mostrou satisfatória. Além disso, muitos autores, como Cerqueira (2004), optam pelo delineamento de 3 perfis, pois facilita na interpretação dos resultados relacionados aos tipos mistos. Desta forma, este estudo optou pela utilização do número ótimo de 3 perfis extremos, que resulta na criação de 6 tipos mistos, fazendo com que 10 perfis de classificação sejam gerados, considerando o tipo amorfo.

Tabela 2: Valores de AIC para os modelos GoM com k perfis - UnB, 2017.

k	ln(L)	AIC
2	-38.835,50	103.175,01
3	-34.631,06	107.518,11
4	-32.132,59	115.273,19
5	-29.870,46	123.500,93

A aplicação do modelo GoM resulta na estimação dos parâmetros λ_{kjl} , que consistem na probabilidade de resposta l para a j-ésima categoria do indivíduo do k-ésimo grupo. Essa estimação permite a descrição dos perfis gerados em função

de suas características predominantes por meio da razão lambda/frequência marginal (RLFM). Esses resultados se encontram descritos abaixo e podem ser melhor visualizados na Tabela 3, em que os valores pintados se referem às características prevalentes, ou seja, a $RLFM \geq 1.2$.

- **Perfil 1:** É composto em sua maioria por alunos pretos ou pardos que ingressaram na Universidade por meio do sistema de cotas raciais; não há distinção quanto ao sexo; realizaram ensino médio em escola pública e não fizeram curso preparatório; tanto o pai quanto a mãe desses alunos possuem apenas ensino fundamental completo; e, por fim, estão nas classes de renda mais baixas, tanto pela classificação do Critério Brasil (D) quanto pela renda mensal da família (até 3 SM).
- **Perfil 2:** Constituído por estudantes que ingressaram na UnB por meio do sistema de cotas sociais, porém não houve distinção de raça/cor; cursaram o ensino médio em escola pública e, possuem pais com o ensino médio completo; fazem parte da classe C e possuem família com renda mensal entre 3 e 10 salários mínimos.
- **Perfil 3:** Alunos brancos ou amarelos que ingressaram por meio do sistema universal; provenientes de escolas particulares e realizaram curso preparatório; os pais possuem nível superior completo; em relação à renda, esses alunos fazem parte das classes sociais mais altas, em ambas as variáveis, sendo que a classificação do Critério Brasil (B) e a renda mensal entre 10 e 20 SM são as mais predominantes.

6. RESULTADOS

Tabela 3: Frequências marginais, probabilidades estimadas e fatores delineadores de perfis, segundo variáveis socioeconômicas - Universidade de Brasília, 2017.

Variáveis	Categorias	Frequência		Probabilidades de resposta			Perfis		
		n	%	λ_1	λ_2	λ_3	1	2	3
Sistema de Ingresso	Cotas raciais	1947	30,66	1,00	0,36	0,00	3,26	1,17	0,00
	Sistema universal	3273	51,54	0,00	0,00	1,00	0,00	0,00	1,94
	Cotas sociais	1131	17,81	0,00	0,64	0,00	0,00	3,60	0,00
Sexo	Masculino	3238	50,98	0,47	0,46	0,57	0,93	0,91	1,11
	Feminino	3113	49,02	0,53	0,54	0,44	1,08	1,09	0,89
Raça/cor	Amarelos e Brancos	3015	47,47	0,00	0,45	0,75	0,00	0,96	1,58
	Pretos e Pardos	3336	52,53	1,00	0,55	0,25	1,90	1,04	0,48
Escolaridade do pai	Fundamental completo	1543	24,30	1,00	0,00	0,00	4,12	0,00	0,00
	Médio completo	2057	32,39	0,00	1,00	0,00	0,00	3,09	0,00
	Superior completo	2751	43,32	0,00	0,00	1,00	0,00	0,00	2,31
Escolaridade da mãe	Fundamental	1174	18,49	0,91	0,00	0,00	4,91	0,00	0,00
	Completo	2144	33,76	0,09	1,00	0,00	0,27	2,96	0,00
	Médio completo	3033	47,76	0,00	0,00	1,00	0,00	0,00	2,09
Cursinho	Sim	2354	37,07	0,19	0,34	0,49	0,51	0,91	1,33
	Não	3997	62,94	0,81	0,66	0,51	1,29	1,05	0,80
Ensino Médio	Escola pública	3191	50,24	1,00	1,00	0,00	1,99	1,99	0,00
	Escola particular	3160	49,76	0,00	0,00	1,00	0,00	0,00	2,01
Critério Brasil	A	341	5,37	0,00	0,00	0,13	0,00	0,00	2,38
	B	1929	30,37	0,00	0,00	0,87	0,00	0,00	2,87
	C	2813	44,29	0,00	1,00	0,00	0,00	2,26	0,00
	D	1268	19,97	1,00	0,00	0,00	5,01	0,00	0,00
Renda mensal da família	até 3 SM	1855	29,21	1,00	0,00	0,00	3,42	0,00	0,00
	3 a 10 SM	2372	37,35	0,00	1,00	0,00	0,00	2,68	0,00
	10 a 20 SM	1225	19,29	0,00	0,00	0,60	0,00	0,00	3,13
	20 ou mais SM	899	14,16	0,00	0,00	0,40	0,00	0,00	2,80

6. RESULTADOS

Observa-se que os perfis são representados por alunos que entraram pelo sistema universal, cotas sociais e cotas raciais. Há indícios de que o perfil 3, representado pelo sistema universal, é o mais privilegiado socioeconomicamente, enquanto o perfil 1, de cotas raciais, é o que possui condições socioeconômicas mais baixas. Além disso, percebe-se que a técnica não captou distinção em relação ao sexo.

Após verificar os perfis socioeconômicos dos estudantes que entraram na Universidade de Brasília em 2017, pode-se investigar como é a distribuição dessas tipologias com o intuito de mostrar onde se encontram os perfis mais vulneráveis.

Desse modo, pode-se utilizar os critérios de preponderância relativa e de agrupamento pela moda para quantificar o pertencimento dos estudantes em cada perfil extremo, apresentados na Tabela 4.

Tabela 4: Comparação entre os perfis gerados pelo critério da moda e da preponderância relativa.

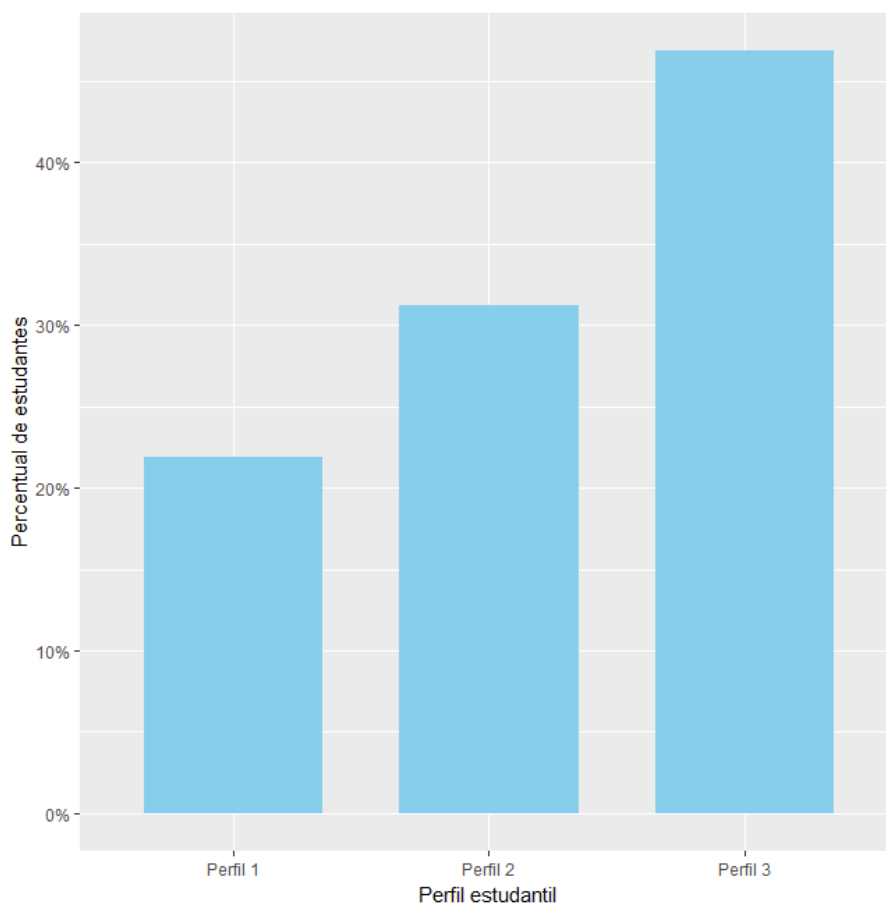
Categoria	Critério da Moda	Critério da Preponderância
	% de estudantes classificados	
Tipo Amorfo	0%	11%
Perfil Extremo 1	22%	20%
Perfil Extremo 2	31%	26%
Perfil Extremo 3	47%	43%

O critério da preponderância relativa possui a desvantagem de considerar os tipos amorfos (categoria que não possui predominância de nenhum perfil). Analisando os resultados mostrados na Tabela 4 pode-se observar que não há muita diferença na distribuição dentro de cada perfil entre os critérios. Portanto, este estudo utilizará o critério de agrupamento pela moda.

Este critério considera apenas os três perfis extremos para classificar todos os indivíduos, baseado no maior escore de pertencimento. Porém pessoas com escores próximos poderiam ser classificadas em perfis diferentes, e essa situação será abordada em breve a partir da criação dos tipos mistos.

Utilizando o critério de agrupamento pela moda, pode-se fazer uma análise geral sobre a distribuição dos perfis construídos. Pela Figura 12, é possível observar que o perfil 3 representa a maior parte dos estudantes que ingressaram na UnB (47%). Já o perfil 2 corresponde a 31% e a minoria faz parte do perfil mais vulnerável (22%).

Figura 12: Percentual de ingressantes na Universidade de Brasília segundo a classificação dos perfis - 2017.



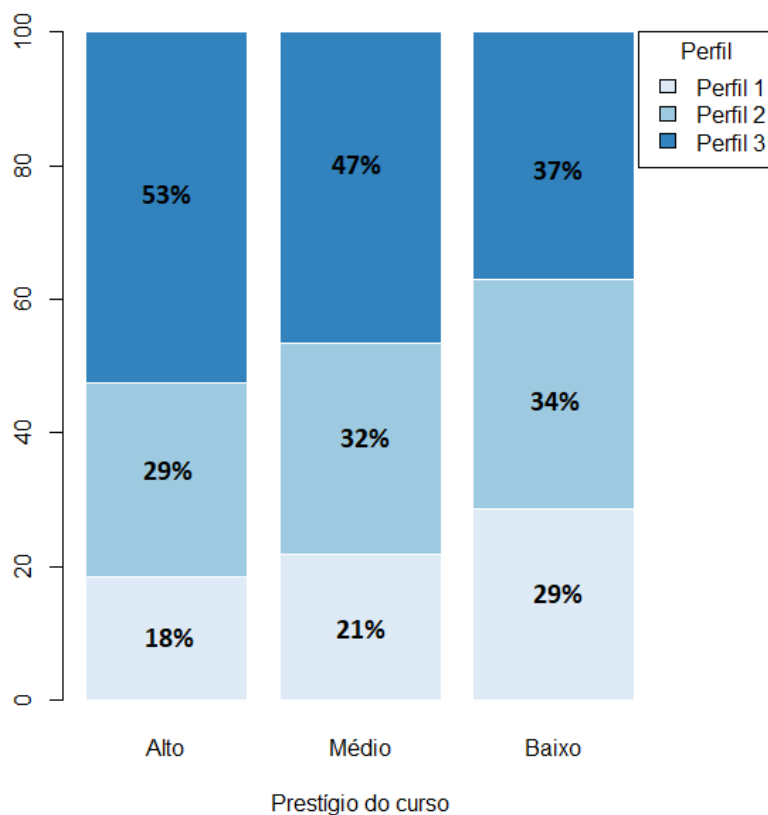
Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Diante desse resultado, surge o questionamento de como os perfis socioeconômicos podem estar associados com algumas variáveis relacionadas ao curso e à modalidade de ingresso, por exemplo.

Ao relacionar os perfis puros com o grau de prestígio do curso, observa-se pela Figura 13 uma possível associação. Apesar de o perfil 3 ser o mais presente em todas as categorias de prestígio do curso, é evidente o quanto essa porcentagem vai diminuindo conforme o prestígio diminui. Já os alunos que fazem parte do perfil 1 são minoria, nota-se que, quando o grau de prestígio diminui, a frequência relativa aumenta.

6. RESULTADOS

Figura 13: Percentual de ingressantes na Universidade de Brasília segundo o perfil e categoria de curso de prestígio - 2017.

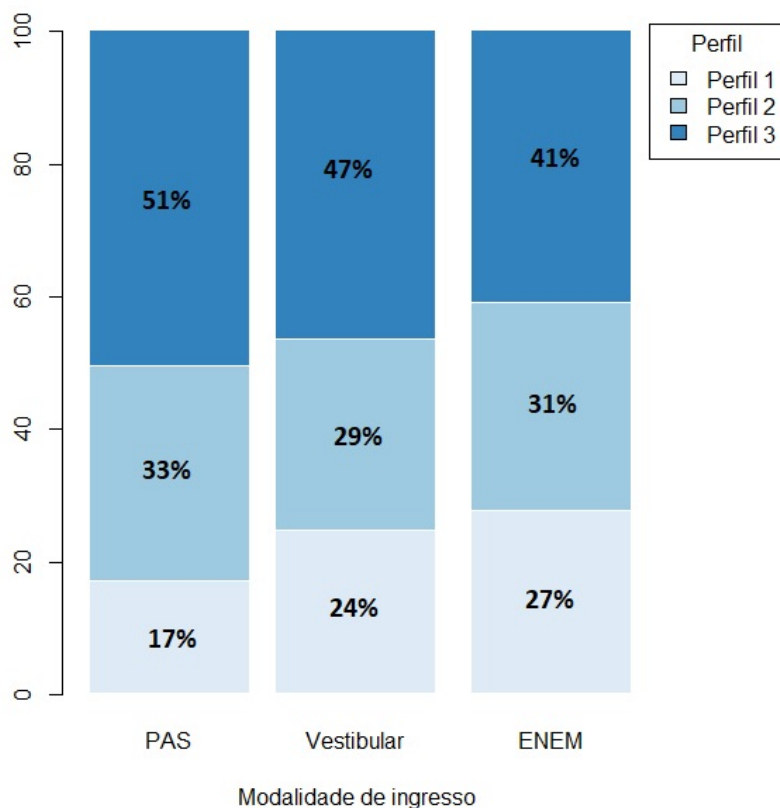


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Há 3 formas principais de ingressar na Universidade de Brasília: PAS, ENEM e vestibular. Observa-se na Figura 14 que o PAS é a modalidade representada por mais características do perfil de alto poder aquisitivo e menos características do perfil mais vulnerável socioeconomicamente. O ENEM parece ser caracterizado de forma equilibrada dentre os perfis e também é o que possui mais alunos do perfil 1.

6. RESULTADOS

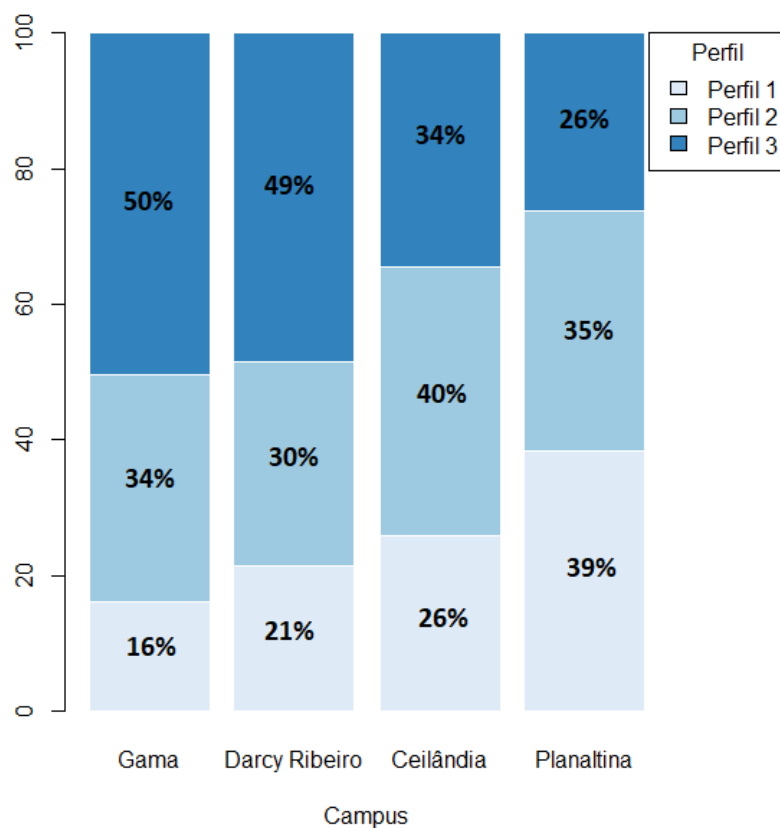
Figura 14: Percentual de ingressantes na UnB segundo o perfil e categoria modalidade de ingresso - 2017.



Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Ao analisar o perfil socioeconômico dos estudantes em cada campus da universidade tem-se resultados interessantes e esperados (Figura 15). O campus Gama é constituído por apenas 6,7% dos ingressantes na UnB e é representado por estudantes de alta renda, já que se trata apenas dos cursos de engenharias. Além disso, pode-se observar que há uma representatividade muito baixa do perfil 1. A Ceilândia também apresentou 6,7% alunos ingressantes e possui mais estudantes do perfil 2. O campus Darcy Ribeiro, que possui 83% dos alunos ingressantes, possui um perfil parecido com o do Gama. O campus de Planaltina é o que mais chama atenção: a maior parte dos estudantes é do perfil de baixa renda, enquanto o perfil 3 é minoria, mostrando possuir estudantes com perfis mais vulneráveis socioeconomicamente.

Figura 15: Percentual de ingressantes na UnB segundo o perfil o campus - 2017.

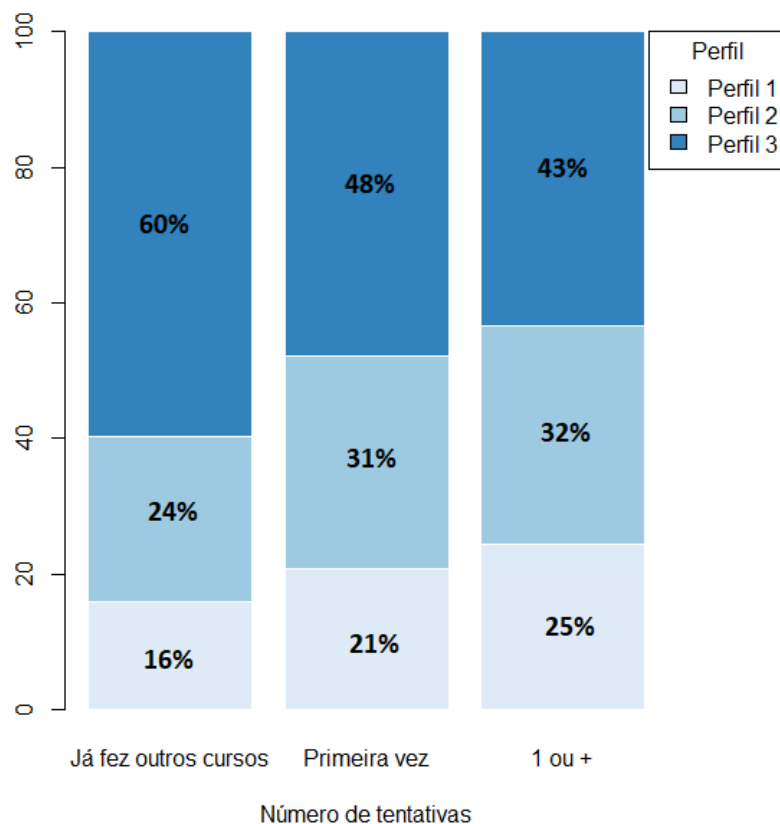


Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

Quanto ao número de tentativas de ingresso, pela Figura 16, observa-se que não há muita diferença de perfil entre as pessoas que entraram de primeira e as que fizeram uma ou mais tentativas, apesar de nesse último caso haver mais presença do perfil 1.

Com relação às pessoas que já tinham feito outros cursos anteriormente, nota-se que a grande maioria faz parte do perfil 3, enquanto apenas 16% do perfil 1. Uma possível explicação é que, dado as características socioeconômicas do perfil 3, pode existir pessoas que queiram dar continuidade aos estudos para melhorar sua formação acadêmica devido ao mercado de trabalho. Uma maneira de conseguir isso seria buscar outras formações.

Figura 16: Proporção de ingressantes na UnB segundo o número de tentativas de ingresso e prestígio do curso - 2017.



Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

6.2.2 Tipos mistos

Uma vez que a técnica GoM permite o pertencimento parcial do indivíduo i à múltiplas categorias a partir da definição dos tipos mistos com base nos perfis extremos, a análise pode ser mais aprofundada. Desta forma, será possível identificar os estudantes que podem apresentar boas condições em determinados aspectos, porém carentes em outros.

Os tipos mistos podem ser definidos com base em alguns critérios de classificação. Dentre os critérios citados nesse estudo, fez-se uma avaliação dos resultados obtidos com o critério de Melo (2007) e o Guedes (2010). A partir dessa avaliação, tem-se a comparação apresentada na Tabela 5.

Observa-se que, pelo critério de Guedes, mais estudantes são classificados como Tipo Amorfos (TA), também conhecido como a categoria que não possui predominância de nenhum perfil. Além disso, proporcionalmente falando, as outras categorias são distribuídas igualmente pelos dois critérios. Portanto, como o propósito desse estudo é classificar ao máximo de modo a mostrar a heterogeneidade

6. RESULTADOS

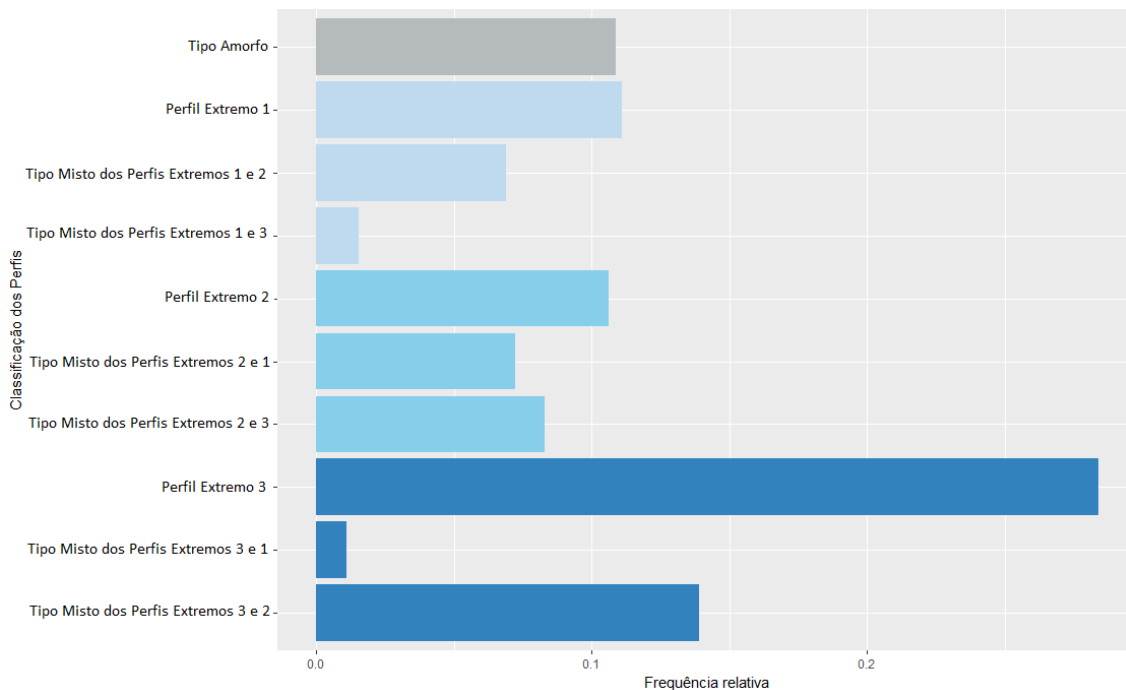
dos indivíduos, optou-se pelo critério de Melo, que pode ser melhor visualizado na Figura 17.

Tabela 5: Comparação entre os perfis gerados pelo critério de Melo e de Guedes.

Categoria (1)	Critério de Melo	Critério de Guedes
	% de estudantes classificados	
TA	11%	17%
PE1	11%	9%
TM12	7%	7%
TM13	2%	2%
PE2	11%	9%
TM21	7%	7%
TM23	8%	8%
PE3	28%	26%
TM31	1%	1%
TM32	14%	14%

Nota: (1) Em que TA significa Tipo Amorfo, PE Perfil Extremo e TM Tipo Misto.

Figura 17: Proporção de ingressantes na UnB segundo a classificação dos perfis extremos e mistos - 2017.



Fonte: Pesquisa sobre os estudantes da UnB - OVE, 2017

6. RESULTADOS

Após a identificação dos perfis mistos, faz-se necessária a verificação da validade dos resultados. Desse modo, realizou-se um teste de médias, com 5% de significância, para avaliar se pelo menos 70% das médias obtidas em cada variável comparadas entre cada perfil são significativamente diferentes (Tabela 6).

Tabela 6: Resultados dos p-valores da comparação de médias entre cada perfil (1).

Perfis	S. de ingresso	Sexo	Raça	Esc. Pai	Esc. Mae	Curso Prep.	Ens. médio	Crit. Brasil	Renda Fam.
PE1 x TM12	*	0,42	*	*	*	0,13	0,01	*	*
PE1 x TM13	*	0,74	*	*	*	0,04	*	0,00	*
PE1 x PE2	*	0,54	*	*	*	*	0,02	*	*
PE1 x TM21	*	0,09	*	*	*	0,02	*	*	*
PE1 x TM23	*	0,04	*	*	*	*	*	*	*
PE1 x PE3	*	0,02	*	*	*	*	*	*	*
PE1 x TM31	*	0,94	*	*	*	0,07	*	*	*
PE1 x TM32	*	0,94	*	*	*	*	*	*	*
TM12 x TM13	*	0,90	0,81	*	*	0,22	*	0,55	*
TM12 x PE2	*	0,82	0,00	*	*	*	0,84	*	*
TM12 x TM21	0,00	0,41	0,81	*	*	0,00	0,12	*	*
TM12 x TM23	*	0,27	0,50	*	*	*	*	*	*
TM12 x PE3	0,00	*	*	*	*	*	*	*	*
TM12 x TM31	*	0,64	0,00	*	*	0,01	*	*	*
TM12 x TM32	*	0,42	*	*	*	*	*	*	*
TM13 x PE2	*	1,00	0,05	*	*	*	*	*	*
TM13 x TM21	*	0,53	0,71	*	*	0,00	*	0,00	*
TM13 x TM23	*	0,43	0,88	*	*	0,63	*	*	*
TM13 x PE3	*	0,15	*	*	*	*	*	*	*
TM13 x TM31	0,26	0,77	0,00	*	*	0,00	0,41	*	*
TM13 x TM32	*	0,76	*	*	*	*	*	*	*
PE2 x TM21	0,00	0,27	0,00	*	*	0,02	0,06	*	*
PE2 x TM23	*	0,15	*	*	*	0,76	*	*	*
PE2 x PE3	*	0,00	*	*	*	0,00	*	*	*
PE2 x TM31	*	0,72	0,07	*	*	0,59	*	0,01	*
PE2 x TM32	*	0,54	*	*	*	0,26	*	*	*
TM21 x TM23	*	0,78	0,35	*	*	0,01	*	*	*
TM21 x PE3	*	*	*	*	*	*	*	*	*
TM21 x TM31	*	0,37	0,00	*	*	0,56	*	*	*
TM21 x TM32	*	0,08	*	*	*	0,00	*	*	*
TM23 x PE3	*	*	*	*	*	0,00	*	*	*
TM23 x TM31	*	0,30	0,00	*	*	0,50	0,55	0,41	*
TM23 x TM32	*	0,03	*	*	*	0,45	*	*	*
PE3 x TM31	*	0,39	0,00	*	*	0,04	*	*	*
PE3 x TM32	*	0,00	0,00	*	*	0,00	*	*	*
TM31 x TM32	*	0,92	*	*	*	0,30	*	0,66	*

Nota:(1) Os campos preenchidos com * indicam p-valor < 0,0001 e os valores em negrito indicam igualdade entre as médias.

Diante dos resultados dos p-valores, nota-se que a variável sexo apresentou a maior ocorrência de igualdade entre as médias. Além disso, observa-se que, como os perfis extremos tendem a diferenciar mais os níveis das variáveis entre si, os p-valores relacionados a comparações entre esses perfis são os que rejeitam ou chegam mais próximo de rejeitar, indicando diferença significativa entre as médias. Por outro lado, os tipos mistos não conseguiram fazer uma boa segregação dessa variável sexo. Isso também vale para as outras variáveis só que com maior quantidade de rejeições, já que essas se mostraram mais heterogêneas.

Ademais, a escolaridade dos pais e a renda mensal da família rejeitaram todas as hipóteses de igualdade entre as médias de cada perfil. Isso indica que cada nível dessas variáveis foi bem discriminado em cada grupo.

Portanto, foi inferido do teste de médias que a técnica GoM conseguiu alocar bem os estudantes em grupos diferentes. A um nível de significância de 5%, pode-se observar que 81% das médias são significativamente diferentes entre os grupos.

Após a validação, pode-se descrever os tipos mistos por meio de uma análise exploratória com base nas características mais predominantes em cada grupo, listadas abaixo conforme a Tabela 7:

- Perfil misto de baixa e média condição socioeconômica, com predominância da baixa: estudantes do sexo feminino, negras e pardas, que ingressaram por meio do sistema de cotas raciais; a maior parte fez ensino médio em escola pública, com pais que possuem até ensino fundamental completo; não fizeram cursinho e fazem parte das classes de renda mais baixas.
- Perfil misto de baixa e alta condição socioeconômica, com predominância da baixa: Estudantes do sexo feminino, negras ou pardas, que ingressaram por meio do sistema universal; o tipo de ensino médio é representado por escola pública e possuem pais com ensino fundamental completo; não fizeram cursinho e assim como o perfil anterior, pertencem às classes de renda mais baixas.
- Perfil misto de média e baixa condição socioeconômica, com predominância da média: Estudantes do sexo feminino, negras ou pardas, que entraram na Universidade por meio do sistema de cotas raciais e cursaram o ensino médio em escolas públicas; Os pais possuem nível médio completo e não fizeram curso preparatório; fazem parte da classe C ou renda menor que 3 salários mínimos.
- Perfil misto de média e alta condição socioeconômica, com predominância da média: Alunas negras e pardas que ingressaram por meio do sistema universal;

6. RESULTADOS

pais com nível médio de escolaridade; não realizaram curso preparatório e fazem parte da Classe C ou renda mensal familiar entre 3 e 10 salários mínimos.

- Perfil misto de alta e baixa condição socioeconômica, com predominância da alta: Alunos amarelos e brancos do sexo masculino que entraram por meio do sistema universal; fizeram o ensino médio em escolas particulares e não realizaram curso preparatório; possuem pais com nível superior completo e pertencem à Classe B.
- Perfil misto de alta e média condição socioeconômica, com predominância da alta: Estudantes brancos do sistema universal que estudaram em escola particular e não realizaram cursinho; pais com alto grau de instrução e renda média familiar entre 3 e 10 salários mínimos ou classe C.

Tabela 7: Distribuição das categorias segundo os tipos mistos.

Perfil	Sexo		S. Ingresso			Raça		Escola	
	Fem	Masc	C. Raciais	C. Sociais	Universal	Brancos/ Amarelos	Pretos/ Pardos	Particular	Pública
TM12	52%	48%	60%	40%	0%	34%	66%	2%	98%
TM13	52%	48%	34%	1%	65%	33%	67%	44%	56%
TM21	55%	45%	53%	44%	3%	35%	65%	4%	96%
TM23	56%	44%	27%	22%	50%	32%	68%	47%	53%
TM31	49%	51%	45%	0%	55%	55%	45%	51%	49%
TM32	50%	50%	2%	13%	85%	77%	23%	80%	20%

Perfil	Esc Mãe			Esc. Pai			Cursinho	
	Fund	Méd	Sup	Fund	Médio	Sup	Sim	Não
TM12	55%	45%	0%	66%	34%	0%	21%	79%
TM13	61%	13%	26%	72%	7%	21%	15%	85%
TM21	25%	75%	1%	40%	59%	1%	32%	68%
TM23	3%	53%	44%	1%	55%	45%	39%	61%
TM31	21%	4%	75%	32%	4%	63%	35%	65%
TM32	2%	32%	67%	3%	37%	60%	41%	59%

Perfil	Renda família				Critério Brasil			
	<3	3-10	10-20	20+	A	B	C	D
TM12	82%	18%	0%	0%	0%	0%	45%	55%
TM13	77%	7%	11%	4%	0%	20%	4%	76%
TM21	59%	41%	0%	0%	0%	1%	76%	24%
TM23	2%	84%	13%	2%	1%	23%	76%	1%
TM31	34%	6%	34%	27%	10%	46%	7%	37%
TM32	4%	62%	26%	8%	2%	38%	59%	2%

A descrição dos tipos mistos permitiu a observação de uma interessante questão. Com os resultados apenas dos tipos puros (perfis extremos), não foi possível verificar como o sexo se encaixava na vulnerabilidade socioeconômica. Apesar de o

teste das médias não ter sido significativo com relação a variável sexo, a caracterização dos tipos mistos dá uma identificação que o sexo feminino está mais presente nos perfis mais vulneráveis, enquanto o sexo masculino, no perfil mais beneficiado.

Outra questão interessante é a visualização da distribuição da raça nos perfis mistos em que o segundo perfil é prevalente, o qual não foi possível descrever utilizando apenas os tipos puros. Observa-se que negros e pardos são a maioria nesses grupos.

Nos grupos relacionados com o perfil 3, o que seria o mais privilegiado, foi possível ver uma predominância de pessoas de classes mais baixas, evidenciando que esse grupo não é exclusivamente de pessoas com alto poder aquisitivo.

7 CONCLUSÃO

Os resultados obtidos nesse trabalho vão de encontro de estudos anteriores realizados no país. Para os estudantes pretos e pardos, o perfil socioeconômico seria relacionado a escolas públicas, baixa renda e a predominância de acesso a cursos de menor prestígio. Já os alunos brancos e amarelos possuem o perfil mais beneficiado, em que fazem parte da classe social mais alta, com ensino médio em escolas particulares, pais com nível superior completo e cursos de alto prestígio.

Além disso, é importante ressaltar que, mesmo que a Universidade de Brasília tenha a maior parte dos ingressantes de 2017 representados pelo perfil de maior privilégio, esse perfil não é exclusivo de estudantes de alta renda. Isso foi melhor observado a partir do delineamento dos perfis mistos por meio do GoM, em que foi possível destacar estudantes de baixa renda dentro dessa categoria.

Também foi possível observar que a implementação do sistema de cotas sociais e raciais contribuiu para uma maior inserção de estudantes negros, de classes sociais mais baixas e de escolas públicas, que eventualmente podem possuir uma base escolar deficiente. Desta forma, acentua-se a relevância da reforma no sistema educacional, pois a base de ensino pré universitário é de extrema importância para uma boa formação dentro da graduação.

Os resultados aqui apresentados destacam a importância de conhecer o perfil discente da UnB, a fim de implementar ações e suportes para que estes alunos, principalmente os mais vulneráveis, tenham melhores condições de permanecer e concluir os cursos pretendidos com sucesso, pois a evasão estudantil acarreta em prejuízo para os próprios estudantes, a Universidade e o país.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ANDREOTTI, A. et al. Multidimensional profiles of health status: an application of the grade of membership model to the world health survey. **PloS one**, v. 4, n. 2, p. e4426, 2009.

BARROS, R. P. et al. Determinantes do desempenho educacional no Brasil. 2001.

COSTA, B.S.P. Medidas de desigualdade: uma análise de segregação socioespacial na área metropolitana de Brasília. 2015.

CERQUEIRA, C.A. Tipologia e características dos estabelecimentos escolares brasileiros. **Belo Horizonte: Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais**, 2004.

CESAR, L. J. T. Mecanismos de seleção para o ensino superior e desigualdade educacional: um estudo sobre o PAS e o vestibular na Universidade de Brasília. 2013.

GILES, R. The concept of grade of membership. **Fuzzy Sets and Systems**. v.25, n.3, p.297-323,1988.

GUEDES, G. R. Ciclo de vida domiciliar, ciclo do lote e mudança no uso da terra na Amazônia Rural Brasileira—um estudo de caso para Altamira, Pará. **Federal University of Minas Gerais, Dissertation**, 2010.

GUEDES, G.; COSTA, S.; BRONDIZIO, E. Revisiting the hierarchy of urban areas in the Brazilian Amazon: a multilevel approach. **Population and environment**, v. 30, n. 4-5, p. 159-192, 2009.

GUEDES, G. R. et al. Identificabilidade e estabilidade dos parâmetros no método Grade of Membership (GoM): considerações metodológicas e práticas. **Revista Brasileira de Estudos de População**, v. 27, n. 1, 2010.

GUEDES, G. R. et al. **Grade of Membership: conceitos básicos e aplicação empírica usando o programa GoM para Windows, Linux e Stata**. mimeo, 2013.

8. REFERÊNCIAS BIBLIOGRÁFICAS

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP) /Ministério da Educação. **Censo da Educação Superior (2016)**. Brasília: INEP/Ministério da Educação. Outubro de 2017. Disponível em:
<http://portal.inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior>. Acesso em 15/06/2018.

JÚNIOR, N.; FERREIRA, J. E. **Perfil sócio demográfico e profissional dos graduados em Estatística no Brasil**. 2016. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Norte.

MANTON, K. G. **Statistical applications using fuzzy sets**. Wiley-Interscience, 1994.

MELO, F. L. B. Casais na Grande São Paulo: investigando a diversidade. *Nova economia*, v. 17, n. 2, p. 207-240, 2007.

PINTO, J. S.; CAETANO, A. J. A Heterogeneidade da Vulnerabilidade Social das Juventudes: Uma Perspectiva Empírica Através do Método Grade Of Membership. *Mediações-Revista de Ciências Sociais*, v. 18, n. 1, p. 164-182, 2013.

SAWYER, D. O.; LEITE, I. C.; ALEXANDRINO, R. Perfis de utilização de serviços de saúde no Brasil. **Ciência Saúde Coletiva**, v. 7, p. 757-776, 2002.

SILVA, J. S.; BARBOSA, J. L.; SOUSA, A. I. Desigualdade e diferença na universidade: gênero, etnia e grupos sociais populares. **Rio de Janeiro: UFRJ, Pró-Reitoria de Extensão**, 2006.

WILLMS, J. D. Literacy Skills and Social Class: How some schools overcome socio-economic handicaps. **POLICY OPTIONS-MONTREAL**, v. 18, p. 22-26, 1997.

WOODBURY, M. A.; CLIVE, J. Clinical pure types as a fuzzy partition. **Journal of Cybernetics and Systems**, v. 4, n. 3, p. 111-121, 1974.

WOODBURY, M. A. et al. Mathematical typology: a grade of membership

8. REFERÊNCIAS BIBLIOGRÁFICAS

technique for obtaining disease definition. **Computers and Biomedical Research**, v.11, p.277-298, 1978.

9 APÊNDICE

9.1 Código em R

9.1.1 Pacote GoMRcpp

O arquivo R-script desenvolvido por Julimar Pinto e André Caetano se encontra no site <http://gilvanguedes.com/publications/gombook/>. A execução desse programa gera 2 documentos:

- **LogGoMK_n**: Permite a descrição dos perfis extremos por meio da RLFM.
- **GoMK_n**: Possui informação do grau de pertencimento do indivíduo *i* ao perfil *k* (g_{ik}).

9.1.2 Classificação dos perfis

Após a identificação dos perfis extremos, tem-se os critérios de classificação que permitem alocar os indivíduos nos grupos segundo seus escores de pertencimento (g_{ik}). Utilizando o arquivo “GoMK_n”, como mostrado na Figura 18, utilizei a seguinte programação:

Figura 18: Arquivo GOMK3 para a execução da programação de classificação dos perfis.

ID	Variáveis										Frequência	Grau (escore) de pertencimento						
	V003C	V008C	V014C	V028C	V029C	V041C	V036C	abep	V024C1	Patterns		initial_gik	initial_gik	initial_gik	final_gik1	final_gik2	final_gik3	
1	1	1	2	3	2	2	2	2	2	4	112322224	1	0,3333	0,3333	0,3333	0,1675	0,2174	0,6151
3	2	2	2	3	3	2	2	2	3	2	222332232	24	0,3333	0,3333	0,3333	0	0,3846	0,6154
4	2	2	2	3	2	1	2	2	2	2	222312222	2	0,3333	0,3333	0,3333	0	0,3594	0,6406
5	1	1	2	1	1	2	1	4	1	1	1112112141	93	0,3333	0,3333	0,3333	1	0	0
6	2	2	1	2	3	2	2	3	2	2	221232232	14	0,3333	0,3333	0,3333	0	0,4995	0,5005
7	2	2	2	3	3	1	2	3	3	3	222331233	12	0,3333	0,3333	0,3333	0	0,1881	0,8119
8	2	1	2	3	3	2	2	1	4	1	212332214	7	0,3333	0,3333	0,3333	0	0	1
9	2	1	2	3	2	2	1	3	2	2	212322132	3	0,3333	0,3333	0,3333	0	0,6916	0,3084
10	2	1	1	2	3	2	2	1	4	2	211232214	2	0,3333	0,3333	0,3333	0	0,1598	0,8402
11	2	1	1	3	3	2	2	2	2	2	211332222	27	0,3333	0,3333	0,3333	0	0,1598	0,8402
12	2	2	1	3	3	2	2	1	4	4	221332214	31	0,3333	0,3333	0,3333	0	0	1
13	3	1	1	3	3	2	1	2	2	3	311332122	7	0,3333	0,3333	0,3333	0	0,483	0,517
14	1	1	2	3	2	2	1	3	1	1	112322131	6	0,3333	0,3333	0,3333	0,4276	0,4241	0,1482
15	3	1	1	3	3	1	1	2	3	3	311331123	11	0,3333	0,3333	0,3333	0	0,2989	0,7011
16	1	1	1	2	2	1	1	2	3	3	111221123	1	0,3333	0,3333	0,3333	0	0,6229	0,3771
17	3	2	1	3	1	2	1	3	1	3	321312131	1	0,3333	0,3333	0,3333	0,36	0,4498	0,1902
18	1	2	2	2	2	2	1	3	2	2	122222132	21	0,3333	0,3333	0,3333	0	1	0
19	2	1	1	2	2	2	2	3	3	3	211222233	3	0,3333	0,3333	0,3333	0	0,483	0,517
20	2	2	2	2	2	2	2	3	2	2	222222232	16	0,3333	0,3333	0,3333	0	0,7047	0,2953
21	2	2	2	3	3	1	2	2	3	3	222331223	21	0,3333	0,3333	0,3333	0	0	1
22	1	1	2	3	2	1	1	4	2	2	112321142	2	0,3333	0,3333	0,3333	0,376	0,4532	0,1709
23	2	2	2	2	3	2	2	3	2	2	222232232	11	0,3333	0,3333	0,3333	0	0,5489	0,4511

```
library(xlsx)
setwd("diretório")
giks = dados_gam[,c(16,17,18)] #posição dos giks

##Critério de Guedes

perfil_guedes = numeric(nrow(giks))

um_quatro = function(x){
  result = numeric(1)
  if(x>=0.75){
    result = 4
  } else if (x>=0.5){
    result = 3
  } else if(x>=0.25){
    result = 2
  } else result = 1
  return(result)
}

for(i in 1:nrow(dados_gam)){
  g1k = giks$final_gik1[i]
  g1k = um_quatro(g1k)
  g2k = giks$final_gik2[i]
  g2k = um_quatro(g2k)
  g3k = giks$final_gik3[i]
  g3k = um_quatro(g3k)
  perfil_guedes[i] = 0
  if(g1k == 4){
    perfil_guedes[i] = 1
  }
  if(g2k == 4){
    perfil_guedes[i] = 4
  }
  if(g3k == 4){
    perfil_guedes[i] = 7
  }
}
```

```
if(g1k == 3 & g2k == 2 & g3k == 1){
  perfil_guedes[i] = 2
}
if(g1k == 3 & g3k == 2 & g2k == 1){
  perfil_guedes[i] = 3
}
if(g2k == 3 & g1k == 2 & g3k == 1){
  perfil_guedes[i] = 5
}
if(g2k == 3 & g3k == 2 & g1k == 1){
  perfil_guedes[i] = 6
}
if(g3k == 3 & g2k == 2 & g1k == 1){
  perfil_guedes[i] = 9
}
if(g3k == 3 & g1k == 2 & g2k == 1){
  perfil_guedes[i] = 8
}
}

dados_gam$perfil_guedes = perfil_guedes

table(dados_gam$perfil_guedes)

# Critério de agrupamento pela Moda

perfil_moda = numeric(nrow(dados_gam))
for(i in 1:nrow(dados_gam)){
  g1k = giks$final_gik1[i]
  g2k = giks$final_gik2[i]
  g3k = giks$final_gik3[i]
  if(g1k == max(g1k,g2k,g3k)){
    perfil_moda[i] = 1
  }
  if(g2k == max(g1k,g2k,g3k)){
    perfil_moda[i] = 2
  }
}
```

```
    if(g3k == max(g1k,g2k,g3k)){
      perfil_moda[i] = 3
    }
  }
}
dados_gam$perfil_moda = perfil_moda
table(dados_gam$perfil_moda)

# Critério da Preponderância Relativa

perfil_prep_rel = numeric(nrow(dados_gam))
for(i in 1:nrow(dados_gam)){
  g1k = giks$final_gik1[i]
  g2k = giks$final_gik2[i]
  g3k = giks$final_gik3[i]
  perfil_prep_rel[i] = 0
  if(g1k > 0.5){
    perfil_prep_rel[i] = 1
  }
  if(g2k > 0.5){
    perfil_prep_rel[i] = 2
  }
  if(g3k > 0.5){
    perfil_prep_rel[i] = 3
  }
}
}
dados_gam$perfil_prep_rel = perfil_prep_rel
table(dados_gam$perfil_prep_rel)

# Critério de Melo

perfil_melo = numeric(nrow(dados_gam))
for(i in 1:nrow(dados_gam)){
  g1k = giks$final_gik1[i]
  g1k = um_quatro(g1k)
  g2k = giks$final_gik2[i]
  g2k = um_quatro(g2k)
  g3k = giks$final_gik3[i]
```

```
g3k = um_quatro(g3k)
if(g1k >= 3 & g2k == 1 & g3k == 1){
  perfil_melo[i] = 1
}
if(g1k == 1 & g3k >= 3 & g2k == 1){
  perfil_melo[i] = 7
}
if(g2k >= 3 & g1k == 1 & g3k == 1){
  perfil_melo[i] = 4
}
if(g1k == 3 & g2k == 2 & g3k == 1){
  perfil_melo[i] = 2
}
if(g1k == 3 & g3k == 2 & g2k == 1){
  perfil_melo[i] = 3
}
if(g2k == 3 & g1k == 2 & g3k == 1){
  perfil_melo[i] = 5
}
if(g2k == 3 & g3k == 2 & g1k == 1){
  perfil_melo[i] = 6
}
if(g3k == 3 & g2k == 2 & g1k == 1){
  perfil_melo[i] = 9
}
if(g3k == 3 & g1k == 2 & g2k == 1){
  perfil_melo[i] = 8
}
}
dados_gam$perfil_melo = perfil_melo
table(dados_gam$perfil_melo)

##Exportando para o Excel
write.csv2(dados_gam, file="dados_gam_tm.csv", row.names = F)
```


9.1.3 Validação dos tipos mistos

Depois da definição do critério a ser utilizado no trabalho (Melo) deve-se fazer a validação.

- **Variáveis contínuas**

Para variáveis que eram contínuas e foram categorizadas para a aplicação da metodologia GoM, deve-se transformar elas em contínuas novamente e aplicar o teste t. Portanto deve seguir o formato mostrado na Figura 19 e utilizar a seguinte programação¹

Figura 19: Formato do banco para a utilização da validação para variáveis contínuas.

	Variável contínua	Classificação do Perfil
	v1	perfil
1	34	7
2	27	9
3	32	9
4	8	1
5	20	9
6	18	7
7	49	7
8	22	6
9	46	7
10	41	7

```
# Rotina para Teste de Médias entre Perfis #
construir.t.test = function(dados,mu = 0){
  t = sort(unique(dados$perfil))
  n = length(t)
  name = list()
  for (i in 1:n){
    name2 = list()
    c = 1
    for (j in 1:n){
      name2[[c]] = paste(paste(paste("dados_",
```

¹Programação disponível no site*+ <http://gilvanguedes.com/publications/gombook/> adaptada.

```

                                t[i],sep=""), "_",sep=""),t[j],sep="")
assign(name2[[c]],dados[dados$perfil == t[i] | dados$perfil
                                == t[j],])

    c = c + 1
}
name[[i]] = name2
}

# Mudança no nome das variáveis #
v29 = list()
vec = 1:ncol(dados)
tam = vec[names(dados) == "perfil"]
for (i in 1:n){
  v = list()
  for (j in 1:n){
    k = get(name[[i]][[j]])
    v[[j]] = paste(paste(paste(paste(paste(paste(
      paste("Perfil_",
            t[i],sep=""), "_",sep=""),"x",sep=""), "_",
            sep = ""),"Perfil",sep = ""),"_",sep = ""),t[j],sep = "")
    names(k)[tam] = v[[j]]
    assign(name[[i]][[j]],k)
  }
  v29[[i]] = v
}

# Criação das fórmulas #
fmla = list()
for (i in 1:n){
  fmla2 = list()
  for (j in 1:n){
    fmla3 = list()
    for (k in 1:(dim(dados)[2]-1)){
      fmla3[[k]] = as.formula(paste(paste(paste("v",k,sep="")
    }
    fmla2[[j]] = fmla3
  }
  fmla[[i]] = fmla2
}
```

```
}
#### Fazendo os Testes # Confere
lista = list()
for (i in 1:n){
  perfis = list()
  for (j in 1:n){
    perfis2 = list()
    for(k in 1:(dim(dados)[2]-1)){
      if( i != j){
        perfis2[[k]] = t.test(fmla[[i]][[j]][[k]],
                             data = get(name[[i]][[j]]),mu = mu)
      }
    }
    perfis[[j]] = perfis2
  }
  lista[[i]] = perfis
}
return(lista)
}
```

```
lista = construir.t.test(dados)
```

```
lista[[2]][[3]][1] #PE1 com TM12
lista[[2]][[4]][1] #PE1 com TM13
lista[[2]][[5]][1] #PE1 com PE2...
```

- **Variáveis qualitativas**

Para o uso da programação a seguir, deve-se usar o banco gerado na classificação dos perfis (no meu exemplo: “dados_gam_tm.csv”), e utilizar a variável do critério escolhido (no meu exemplo: “perfil_melo”).

```
library(readr)
dados <- read_delim("C:../dados_gam_tm.csv",
                   ";", escape_double = FALSE, trim_ws = TRUE)
```

```
##i indica a posição da variável
##[,c(2,5)]: significa fazer um teste para os perfis 1 e 2,
##tem que fazer a modificação para cada comparação...
```

```
lista_dados=list()
for (i in 2:10) {
  lista_dados[[i]]=prop.test(table(dados[,i],dados$perfil_melo)[,c(2,5)],
                             alternative='two.sided',
                             conf.level=0.95,
                             correct=FALSE)
}
```