

Análise de Homogeneidade do Desempenho de Instituições de Ensino Superior no Rio de Janeiro no ENADE 2007

Giovani Glaucio de Oliveira Costa

Universidade Federal Rural do Rio de Janeiro

Instituto Multidisciplinar

Rua Capitão Chaves, 60. CEP 26221-010. Centro. Nova Iguaçu. Rio de Janeiro. Brasil

giovaniglaucio@ufrj.br

Resumo:

O objetivo deste artigo foi realizar um estudo sobre o desempenho das instituições de ensino superior no Rio de Janeiro no ENADE 2007 através da busca de perfis diferenciados de desempenho e suas inter relações no espaço de análise. Como se procurou expor, a Análise de Homogeneidade é particularmente adequada para reproduzir num espaço de menores dimensões a multiplicidade que sustenta o espaço original. Não sendo este o único objetivo que leva à seleção de uma HOMALS é, sem dúvida, um dos mais habituais e cujo enfoque foi neste artigo privilegiado. A análise de dados citada permitiu visualizar num espaço de menores dimensões três configurações ou perfis de desempenho das instituições no Rio de Janeiro: baixo, mediano e alto desempenho. A grande maioria das instituições deste ENADE pode ser considerada como de desempenho mediano (67,1%), com destaque também para as que tiveram baixo desempenho (28,4%). Numa perspectiva de completabilidade, foi proposto classificar as instituições estudadas em função de seu grau de desempenho na avaliação em foco, o que pode ser viabilizado através da Análise de Clusters. A articulação da Análise de Homogeneidade com a Análise de Clusters permitiu validar a consistência da classificação. A tipologia de segmentação revelou-se coerente com a configuração topológica revelada pela Análise de Homogeneidade. Os dados deste estudo mostram que muitas instituições de ensino superior precisam melhor ainda em muito para oferecer cursos de graduação de excelência no Rio de Janeiro que promova a capacidade intelectual e profissional necessária para que os jovens brasileiros possam se auto realizar em suas carreiras, além impulsionarem o devido desenvolvimento do país.

Palavras Chaves: desempenho do ensino superior, ENADE 2007, Análise de homogeneidade, análise de clusters.

1-Introdução

O principal objetivo deste artigo é indicar um critério para classificar as instituições do município do Rio de Janeiro, participantes do ENADE 2007 nos grupos de baixo, médio e ótimo desempenho no ENADE 2007.

Para concretizar o objetivo proposto, será utilizada a articulação da análise da homogeneidade com a análise de clusters.

Em primeiro lugar, será definido o ENADE e as características do realizado em 2007.

Em seguida, serão apresentadas as metodologias da análise da homogeneidade e da análise de clusters, e como estas duas técnicas da análise multivariada podem ser articuladas para gerar os resultados esperados.

Todas as etapas necessárias para a realização da análise de homogeneidade serão apresentadas, bem como da análise de clusters e da articulação dessas duas técnicas.

Com as análises de dados geradas da articulação da análise de homogeneidade/análise de clusters, pode-se ter um critério de classificar o desempenho das instituições em função do perfil que ela apresenta.

O artigo termina expondo as principais conclusões da pesquisa e as referências bibliográficas utilizadas.

Este artigo tem a vantagem de realizar um estudo estatístico além do descritivo para os resultados do ENADE 2007, com análises de topologia e tipologia, utilizando análise multivariada, gerando uma classificação útil para as autoridades competentes nas tomadas de decisão necessárias com respeito a cada instituição na busca pela qualidade do ensino superior no Brasil, mas também para as instituições participantes, que devem se basear para melhorar a qualidade do ensino que oferecem.

2- O ENADE 2007

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep/MEC, por intermédio de sua Diretoria de Estatísticas e Avaliação da Educação Superior – Deaes, realiza cada ano o Exame Nacional de Desempenho dos Estudantes – ENADE como parte integrante do *Sistema Nacional de Avaliação da Educação Superior - Sinaes*, criado pela Lei no 10.861/2004.

Por ser componente curricular obrigatório dos cursos de graduação, o ENADE deve ser plicado a todos os estudantes do final do primeiro e do último ano do curso, de acordo com as portarias nº 107, de 22/7/2004 e nº 603, de 7/3/2006. Cabe, portanto, aos dirigentes das instituições de educação superior – IES, e aos coordenadores dos cursos que estarão participando do ENADE observar atentamente as suas atribuições, os prazos e os procedimentos que devem ser adotados pelas IES para garantir aos estudantes o acesso às informações pertinentes à sua participação no Exame.

O ENADE 2007 avaliou 16 áreas de conhecimento, incluindo os cursos de agronomia, biomedicina, educação físico, enfermagem, farmácia, fisioterapia, fonoaudiologia, medicina, medicina veterinária, nutrição, odontologia, serviço

social, tecnologia em radiologia, tecnologia em agroindústria, terapia ocupacional e zootecnia. A prova foi realizada em novembro 2008. Ao todo, 258.342 universitários foram convocados para fazer o exame.

Este artigo se concentrou nas instituições de ensino superior participantes da cidade do Rio de Janeiro, que reuniu um total de 88 instituições.

Foram avaliados 3.239 cursos de 753 instituições do país. De acordo com Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), responsável pela aplicação do exame, foram convocados 145.380 ingressantes e 112.962 concluintes, além de 1.365 voluntários, 31.104 candidatos irregulares de 2004, 2005 e 2006 e 10.454 inscritos fora do prazo.

A prova é a mesma para ingressantes e concluintes da mesma área e conta com 40 questões de múltipla escolha e dissertativas, 30 específicas do curso e dez de formação geral. Com base no desempenho dos estudantes, os cursos recebem conceitos de 1 a 5.

O ENADE é uma avaliação obrigatória para os alunos selecionados por amostragem pelo Inep. Caso o aluno escolhido pelo Inep não faça a prova, ele perderá o direito de receber o diploma de conclusão de curso.

A avaliação faz parte do Sistema Nacional de Avaliação da Educação Superior (Sinaes) e visa avaliar a qualidade dos cursos de graduação mediante o desempenho dos estudantes. Também compõem o Sinaes, a Avaliação de Curso e a Avaliação Institucional.

O exame busca medir a contribuição do curso na formação do estudante. Dessa forma, aplica a mesma prova para alunos ingressantes e para alunos concluintes de cada um dos cursos avaliados - medindo, assim, a situação em que o estudante chegou ao curso e a situação em que está saindo, o que permite avaliar o que foi agregado em termos de conhecimento, tanto geral quanto específico.

O ENADE 2007 calcula três conceitos que motivarão a classificação pretendida das instituições neste artigo: ENADE Conceito, o IDD e o CPC.

O *ENADE Conceito* é calculado pela média ponderada da nota padronizada dos concluintes no componente específico e da nota padronizada em formação geral (concluintes e ingressantes), possuindo estas, respectivamente, os seguintes pesos: 75% e 25%. Assim, a parte referente ao componente específico contribui com 75% da nota final, enquanto a referente à formação geral contribui com 25%. O conceito é apresentado em cinco categorias (1 a 5) sendo que 1 é o resultado mais baixo e 5 é o melhor resultado possível. SC: quando não tem ingressante ou não tem concluinte que participou efetivamente do ENADE através da realização da prova. Para o caso das engenharias tinha menos de 10 cursos participantes. Em alguns cursos de Engenharia não é atribuído conceito aos estudantes ingressantes, aparecendo o conceito apenas para os concluintes. Estes são os cursos de Engenharia que adotam o sistema de curso básico, no qual os estudantes escolhem a habilitação ou ênfase apenas quatro ou mais semestres após o ingresso. Assim, em alguns cursos, os ingressantes fizeram à prova no Grupo VII, denominado "Engenharia" e os concluintes em um grupo diferente onde estava situado o curso e a ênfase. Como os estudantes ingressantes fizeram provas diferentes dos concluintes os resultados não são comparáveis.

O *IDD Conceito*, Indicador de Diferença entre os Desempenhos Observados e Esperados, é uma transformação do *IDD Índice*, de forma que ele seja apresentado em cinco categorias (1 a cinco) sendo que 1 é o resultado mais baixo e 5 é o melhor resultado possível no *IDD Conceito*. SC: menos de 10 ingressantes e/ou concluintes e nota zero.

O *CPC*, Conceito Preliminar do Curso, é indicador prévio que assume valores de 1 a 5, em que 5 é o resultado mais alto. O cálculo do Conceito Preliminar pondera os resultados dos cursos no Conceito ENADE (40%), Conceito *IDD* (30%) e variáveis de insumos (30%) que informam a situação dos cursos em relação à infra-estrutura e equipamentos, aos recursos didático-pedagógicos e ao corpo docente. Essas variáveis de insumo assumem os seguintes pesos respectivamente: 10,2%; 27,1% e 62,7%. As respectivas ponderações foram definidas conforme resultados de um modelo que estuda de que forma e quais variáveis influenciam a qualidade oferecida pelos cursos medida pelo *IDD*.

3-A Análise de Homogeneidade

O método de Análise de Homogeneidade (*HOMALS*-Homogeneity Analysis by Menos of Least Square), também conhecida como Análise de Correspondência Múltipla (*ACM*), permite resumir um grande número de variáveis qualitativas de uma base de dados de estudo em um pequeno número de variáveis quantitativas, facilitando o estudo das relações entre as diversas características existentes num determinado espaço de análise.

A Análise de Homogeneidade integra os métodos de “*Optimal Scaling*”. Estes métodos consistem, na prática, numa intervenção sobre as categorias com o objetivo de proceder a sua quantificação.

A *HOMALS* foi formalizada por Gifi que também foi responsável pela operacionalização dos seus procedimentos no software *SPSS*.

A *HOMALS* viabiliza uma Análise de Homogeneidade por via de um método algébrico do tipo “*Alternating Least Squares*”. O algoritmo da *HOMALS* é, então, do tipo *ALS*, o que significa que em cada uma das iterações vão sendo determinadas, alternativamente, estimativas quantitativas para os objetos e categorias da base de dados. Este processo é um procedimento de transformação ótima porque vai sendo minimizada uma função perda, cuja convergência corresponderá às quantificações das categorias e dos objetos. Estas quantificações têm a propriedade de guardarem e resumirem as características essenciais das categorias e objetos de análise da matriz de dados de partida, mantendo a variação do espaço original intacta no espaço gerado de menores dimensões.

Em sentido restrito, a Análise de Homogeneidade é um método de análise quantitativa para variáveis nominais e tem por objetivo atingir a solução ótima na quantificação de dados qualitativos. Em sentido amplo, a Análise de Homogeneidade refere-se a um conjunto de critérios para analisar dados multivariados, com o objetivo de otimizar a homogeneidade das variáveis.

A *HOMALS* constitui, portanto, num conjunto de critérios para analisar dados multivariados, com o objetivo de se atingir a homogeneidade das variáveis. A técnica trata múltiplas variáveis categoriais e desenvolve análises estruturais. È

um método de análise quantitativa para variáveis qualitativas e tem por objetivo atingir a solução ótima na quantificação de dados qualitativos ou categoriais.

A análise de homogeneidade submete os dados qualitativos de *input* a um processo de quantificação. A transformação desencadeada tem por fim estimar quantificações ótimas para os parâmetros em análises: categorias e objetos.

Na elaboração de um plano de pesquisa que preveja uma coleta extensa de informações, a seleção dos métodos de análise de dados é uma tarefa de importância decisiva, quaisquer que sejam os fenômenos em estudo. É essencial garantir que os instrumentos que venham a ser selecionados permitam uma análise adequada do modelo e sejam consistentes com a natureza das informações colhidas.

A HOMALS é o instrumento analítico que privilegia as situações nas quais é necessário lidar conceitualmente e metodologicamente com objetos de configuração complexa. A Análise de Homogeneidade consiste numa abordagem multifacetada e relacional sobre o objeto em estudo indicando traços de sua configuração complexa.

A Análise de Homogeneidade é particularmente interessante para reproduzir num espaço de menores dimensões multiplicidade que sustenta o espaço original. Uma das premissas para a Análise de Homogeneidade é a de que a complexidade é multidimensional.

A configuração complexa do espaço de análise leva a que o seu processo de conhecimento perspetive uma abordagem multifacetada e relacional.

Na análise de dados pela HOMALS, pretende-se interagir com situações de pesquisa empírica nas quais é necessário operar a multidimensionalidade, o que implica trabalhar com amplo sistema de variáveis.

Na Análise de Homogeneidade, procura-se visualizar através da disposição relativa dos indivíduos ou das categorias em planos, as relações estatísticas delineadas entre as múltiplas variáveis em análise.

A partir das configurações definidas pelas categorias das variáveis espera-se conhecer de forma aproximada a estrutura que sustenta o espaço em análise.

As configurações nos planos ao exibirem diferentes combinações das propriedades envolvidas refletem a presença de grupos de indivíduos relativamente homogêneos. Seja, então, qual for a natureza das unidades de análise, o que determina a opção pela Análise de Homogeneidade é propósito de realizar uma abordagem relacional sobre as múltiplas variáveis que caracterizam os objetos, podendo vir a serem definidos diferentes grupos.

A descrição desses grupos pode contemplar dois vetores analíticos:

- ❖ *Identificação do perfil de cada grupo definido;*
- ❖ *Observação do posicionamento relativo dos vários grupos que permita detectar a existência de relação de associação ou de oposição.*

A Análise de Homogeneidade, portanto, trata múltiplas variáveis categoriais e desenvolve análises estruturais.

Fala-se em homogeneidade no sentido em que a proximidade de um certo número de categorias de diferentes variáveis induz a presença de indivíduos na base de dados que compartilham tendencialmente as mesmas características, isto é, tem o mesmo perfil. A diferentes núcleos de homogeneidade correspondem grupos de indivíduos com perfis distintos, mas que coexistem, com maior ou menor proximidade, no mesmo espaço.

Os resultados ou output da HOMALS incluem:

- 1) *Estatísticas-valores próprios (variâncias explicadas de eixos estruturantes), medidas de discriminação ou importância das variáveis, quantificações das categorias e escores de objetos;*
- 2) *Imagens gráficas (planos bidimensionais) que permitem a visualização da informação contida na matriz de entrada num espaço de menores dimensões.*

A par do suporte quantitativo, é também essencial a atribuição de sentido a esses elementos numéricos. Contemplando todos estes elementos, a interpretação incidirá sobre dois elementos distintos: dimensões e planos.

A interpretação das dimensões deve ser apoiada em elementos estatísticos, os quais incluem os valores próprios, as medidas de discriminação, as quantificações das categorias e eventualmente os escores de objetos.

Os planos gerados devem ser sustentados por todos os elementos considerados relevantes por dimensões ortogonais.

Numa perspectiva mais pragmática, dir-se-ia que a análise das dimensões permite desde logo saber quais as variáveis e categorias devem ser consideradas quando da interpretação dos planos.

As diferentes configurações correspondem a associações privilegiadas das categorias, sobretudo de diferentes variáveis, permitindo assim conhecer não só o perfil dos diferentes grupos, mas inclusive, a sua localização relativa ao espaço em estudo.

Nos planos nos quais estão projetados os objetos, a análise toma por referência a interpretação das distâncias mas, neste caso, analisam-se as associações e as oposições observando diretamente quem são os objetos e não as suas características, os seus perfis. Esta última informação é peculiar aos gráficos das categorias.

A interpretação das dimensões e dos planos constituem dois vetores analíticos que cumprem diferentes objetivos.

De forma esquemática, dir-se-ia que da interpretação das dimensões selecionadas, a HOMALS permite conhecer os principais eixos analíticos responsáveis pela configuração da estrutura relacional que subjaz esses espaços de complexa definição.

No caso dos planos, especificamente a leitura do gráfico das categorias, procura-se identificar e caracterizar configurações. Cada configuração tem por referência um conjunto de objetos tendencialmente homogêneos.

O método da HOMALS tem as seguintes fases:

1-Identificação dos Eixos Estruturais do Espaço de Análise

1.1-Seleção das Dimensões mais Representativas

1.1.1-Identificação das dimensões com valores próprios mais elevados, com a ajuda do gráfico em barras, feito em Excel, destas estatísticas;

1.1.2-Identificação das variáveis com medidas de discriminação mais elevadas, privilegiando as dimensões da análise feita em 1.1.1, podendo ser feito também com o auxílio do gráfico de variáveis, que também é chamado de gráficos das medidas de discriminação.

1.2-Interpretação/Nomeação das Dimensões

1.2.1-Leitura do significado das dimensões por via das variáveis selecionadas em 1.1.2. O objetivo desta fase é nomear as dimensões;

1.2.2-Leitura do significado das dimensões por via também das categorias das variáveis selecionadas usando as suas quantificações. O objetivo desta fase é analisar as associações e as oposições entre as categorias selecionadas. As associações são analisadas a partir da presença de categorias que detenham quantificações com mesmo sinal e as oposições pela presença de categorias que possuem sinais contrários.

2-Identificação de Configurações ou de Grupos Homogêneos no Espaço de Análise

2.1-Interpretação dos Planos das Categorias

2.1.1-Leitura de diferentes configurações de categorias por via das projeções de todas as categorias identificadas em 1.2.2. O Objetivo desta fase é identificar grupos homogêneos.

2.2-Interpretação dos Planos de Objetos

2.2.1-Observação da intensidade da presença dos grupos já identificados;

2.2.2-Identificação dos objetos quando tal for possível;

2.2.3-Observações de outliers.

3- A Articulação da HOMALS com a Análise de Clusters

O que determina a utilização prévia da HOMALS é a necessidade de interagir com contextos de análise que se caracterizam por contemplar estruturas multifacetadas, relacionais e definidas segundo variáveis qualitativas. O resultado desta análise implica na identificação de configurações e perfis.

A análise de Clusters também permite gerir múltiplas variáveis qualitativas, mas somente a nível classificatório. A análise de Clusters não define perfis, como a HOMALS.

Contudo, pode-se fazer a aproximação da configuração topológica delimitada pela Análise de Homogeneidade com a construção de tipologias classificatórias da Análise de Clusters, articulando as potencialidades das duas técnicas numa única análise de dados.

A classificação dos objetos aos clusters é determinada em função da distância que existem entre os objetos, relativamente a um conjunto de variáveis definidas.

Na articulação HOMALS/Análise de Homogeneidade, a Análise de Clusters toma como variáveis de *input*, para a classificação, as dimensões que sustentam o plano em estudo na Análise de Homogeneidade. Quando se faz referência às dimensões enquanto novas variáveis, isso já pressupõe que depois de realizada a HOMALS e tomada a decisão sobre quantas dimensões interpretar, foram salvos na base de dados os escores de objetos por dimensão. Com este procedimento, serão acrescentadas à base de dados novas variáveis, em número igual ao indicado para as dimensões.

As dimensões são transformadas em novas variáveis com características particularmente interessantes: variáveis compostas e contínuas, que têm enorme potencial analítico. As dimensões são constituídas pelos escores de objetos e estes, algebricamente são determinadas a partir das múltiplas variáveis qualitativas que figuram na matriz de *input* e que definem o perfil de cada indivíduo. Daí afirmar-se que as dimensões são variáveis compostas, garantindo que a multidimensionalidade foi preservada. Quanto à sua natureza, e apesar do *input* ser qualitativas, as dimensões são variáveis contínuas.

Para realizar a Análise de Clusters começa-se por recorrer-se ao Método *K-Means* Clusters. Assim, será utilizada como variáveis de *input* as dimensões consideradas para a análise e solicitar o número de grupos que se quer formar, correspondente ao número de configurações reveladas no gráfico das categorias da Análise de Homogeneidade.

Para garantir que a partição dos indivíduos nos clusters gerados pela análise fique definida como nova variável é necessário salvar esta variável na base de dados. Em seguida, vem à caracterização dos grupos obtidos, a fim de efetuar a correspondência destes com os perfis previamente definidos via HOMALS.

Na prática, a variável de classificação, salva na base de dados, classifica cada objeto em um dos clusters gerados. Contudo, o método de clusters só diz quem é cada indivíduo, a que grupo pertence, mas não informa como eles são. A associação com a informação da HOMALS vai justamente traçar o perfil de todos os clusters gerado pela Análise de Clusters.

Usa-se o Método *K-Means*, um método não hierárquico, pois geralmente já se sabe, através da análise do plano da análise da HOMALS, o número de grupos que a Análise de Clusters deve segmentar o espaço de análise original.

Caso não se tenha a informação prévia de quantos grupos deve a Análise de Clusters segmentar o espaço de análise ou para validar se aquela que se considera a melhor partição é consistente com as conclusões retiradas do plano em estudo da HOMALS, sugere-se, antes, usar um método hierárquico de aglomeração, que indique o número de clusters ideal para a classificação do espaço de análise.

Para se concluir quanto ao número de clusters, pode-se usar a leitura do dendograma, estatística gerada pelo método hierárquico. Em casos em que o número de objetos da base de dados é alto, sugere-se usar um gráfico com as distâncias dos coeficientes de fusão (*Screen Plots*), estatística fornecida pelo *output* da análise. Este gráfico relaciona números de clusters em função das distâncias ou coeficientes de fusão. Observa-se neste gráfico a partir de quantos clusters as distâncias deixam de ser tão significativas. Os resultados da realização de um método de classificação hierárquico permite, assim, validar a solução sugerida pela HOMALS.

Classificando os casos, é, então, feita a correspondência entre a solução dos clusters determinados e os perfis configurados pelo plano da HOMALS. Este procedimento pode ser conseguido pelo cruzamento de cada uma das variáveis usadas na definição dos perfis, via HOMALS, com a variável na base de dados que define os clusters.

Para tanto, são retomadas as variáveis de *input* da HOMALS que mais discriminam nas dimensões retidas e efetuar-se o cruzamento com a variável de clusters salva na base de dados. A partir das percentagens mais elevadas para cada um dos clusters pode-se fazer uma síntese interpretativa de suas principais características.

Para validar graficamente a consistência desta classificação, pode ainda proceder-se à projeção dos clusters no plano das categorias.

Do ponto de vista técnico, consiste numa representação gráfica na qual estão simultaneamente representadas as variáveis que configuram os perfis, via HOMALS, como variáveis ativas, e a variável que identifica a tipologia de segmentação, tendo esta o status de variável passiva. O objetivo é observar a disposição dos clusters formados pela Análise de Clusters com a realização da Análise de Homogeneidade.

A representação gráfica deve evidenciar a tipologia de segmentação coerente com os resultados desde logo obtidos com a realização da Análise de Homogeneidade.

A exemplificação sistematizada evidencia como a gestão complementar entre a HOMALS e a Análise de Clusters pode de fato, desempenhar um papel importante na operacionalização de grupos, na perspectiva de definir tipologias.

Assim, num primeiro momento, afere-se sobre a configuração topológica do espaço de análise, o que implica a operar com múltiplas variáveis qualitativas que interagem de forma sistêmica. Este processo analítico resulta na identificação de diferentes perfis. Num segundo momento, parte-se para o ensaio de tipologias dos objetos do espaço de análise. De cada tipo gerado pelo método não hierárquico de aglomeração é traçado um perfil coerente com o evidenciado pela Análise de Homogeneidade. É este encadeamento operatório que se considera ter uma concretização possível da articulação da Análise de Homogeneidade com a Análise de Clusters.

4-Análise de Homogeneidade do Espaço de Desempenho das IES do ENADE 2007

O objetivo desta seção é reduzir o espaço de desempenho das instituições de ensino superior, IES, participantes do ENADE 2007, constituídas das variáveis tratadas como qualitativas e descritas abaixo em um espaço de menores dimensões, representado por um plano bidimensional, que sustenta a multidimensionalidade do referido espaço de partida.

O quadro 1 descreve variáveis consideradas na Análise de Homogeneidade.

Quadro 1

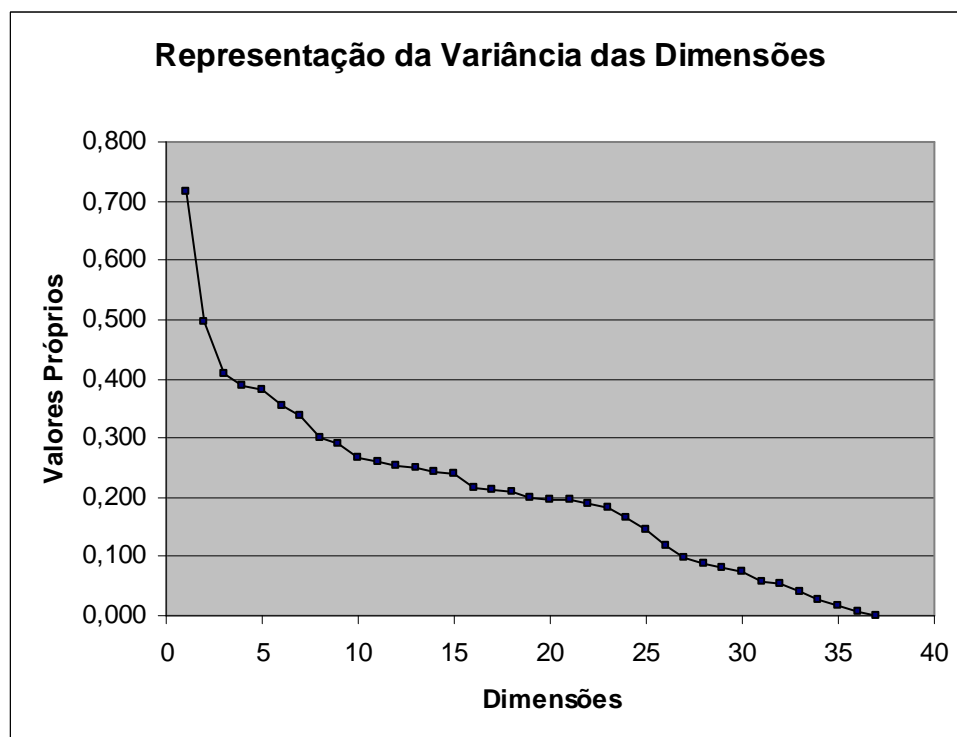
Variáveis	Categorias
ENADE	<i>Sem Conceito</i> <i>Muito Baixo ENADE</i> <i>Baixo ENADE</i> <i>Médio ENADE</i> <i>Bom ENADE</i> <i>Ótimo ENADE</i>
IDD	<i>Sem Conceito</i> <i>Muito baixo IDD</i> <i>Baixo IDD</i> <i>Médio IDD</i> <i>Alto IDD</i> <i>Altíssimo IDD</i>
CPC	<i>Sem Conceito</i> <i>Muito Baixo CPC</i> <i>Baixo CPC</i> <i>Médio CPC</i> <i>Alto CPC</i> <i>Altíssimo CPC</i>

4.1-Identificação dos Eixos Estruturais do Espaço de Análise

4.1.1-Seleção das Dimensões mais Representativas

4.1.1.1-Identificação das dimensões com valores próprios mais elevados, com a ajuda do gráfico em barras, feito em Excel, destas estatísticas:

Gráfico 1- Representação da Variâncias das Dimensões



A partir da representação dos valores próprios(gráfico 1) é possível aferir sobre o destaque faz duas primeiras dimensões.

4.1.1.2-Identificação das variáveis com medidas de discriminação mais elevadas, privilegiando as dimensões da análise feita em 4.1.1.1, podendo ser feito também com o auxílio do gráfico de variáveis, que também é chamado de gráficos das medidas de discriminação.

Como se distribuem as três variáveis nas duas primeiras dimensões é o que se passa analisar por via da leitura das medidas de discriminação (quadro 3). Deve proceder-se à seleção das variáveis mais importantes para cada dimensão. Para tal, sugere-se usar como referência o valor próprio de cada uma das duas primeiras dimensões(quadro 2). As variáveis que mais discriminam em uma dada dimensão é aquela cuja sua medida de discriminação é maior ou igual ao valor próprio da dimensão.

Quadro 2-Valores Próprios

Dimensão	Valor Próprio
1	0,814
2	0,678

As variáveis assinaladas no quadro 3 são as que mais discriminam em cada uma das duas dimensões, uma vez que seus valores são maiores que os valores próprios.

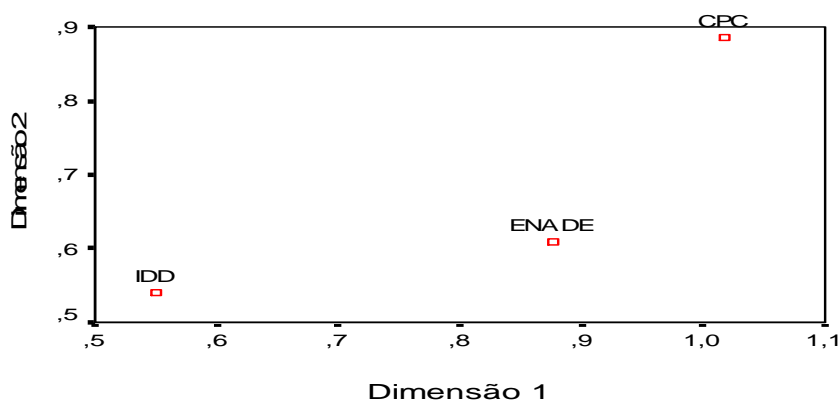
Quadro 3-Medidas de Discriminação

Variáveis	Dimensão	
	1	2
ENADE	0,876	0,609
IDD	0,550	0,540
CPC	1,017	0,885

Tendo por base o cálculo da percentagem da variância explicada para cada uma das dimensões 1 e 2, é possível perceber que, na generalidade, as variáveis ENADE e CPC contribuem da mesma forma praticamente semelhante para a definição dessas dimensões (ver gráfico 2). Para a dimensão 1, são mais determinantes as variáveis ENADE e CPC. Para a dimensão 2, se sobressai a variável CPC. Pela análise das medidas de discriminação, a variável IDD parece não discriminar em nenhuma das duas dimensões retidas. O que pode ser confirmado pela sua proximidade da origem, observado no gráfico 2, apesar dela ter praticamente o mesmo poder de discriminação nas duas dimensões retidas. Deve-se investir, então, na interpretação ao nível das categorias e também dos objetos, se assim se considerar interessante.

Gráfico 2-Gráfico das Variáveis

Gráficos das Variáveis



4.2-Interpretação/Nomeação das Dimensões:

4.1.2.1-Leitura do significado das dimensões por via das variáveis selecionadas em 4.1.1.2. O objetivo desta fase é nomear as dimensões:

A dimensão 1 é discriminada pelas variáveis ENADE e CPC e a dimensão 2 somente é discriminada pela variável CPC. O que diferencia as duas dimensões é a variável ENADE, que prioriza diretamente o conhecimento delegado à formação do estudante e só discrimina diretamente a dimensão 1. A dimensão 2 é discriminada unicamente pela variável CPC que se refere além do conhecimento, componentes de insumo na avaliação. Como a dimensão 1 é bastante reforçada pelo conhecimento específico e geral dado ao estudante (presentes na composição tanto da variável ENADE como na CPC), podemos nomeá-la como **qualidade de formação** delegada ao estudante e a dimensão 2 como é unicamente discriminada pela variável CPC, que envolve em sua composição além do conhecimento específico e geral fornecido, características de insumo, como infra-estrutura, recursos pedagógicos e corpo docente, devem sugerir referir à **qualidade de insumo** delegada ao estudante.

4.1.2.2-Leitura do significado das dimensões por via também das categorias das variáveis selecionadas usando as suas quantificações.

O objetivo desta fase é analisar as associações e as oposições entre as categorias selecionadas. As associações são analisadas a partir da presença de categorias que detenham quantificações com mesmo sinal e as oposições pela presença de categorias que possuem sinais contrários.

Quadro 4-ENADE

Variáveis	Frequências	Quantificação das Categorias	
		1	2
Muito Baixo ENADE	2	-0,833	-0,907
Baixo ENADE	27	-0,753	-0,825
Médio ENADE	28	-0,093	0,601
Bom ENADE	16	0,739	0,774
Ótimo ENADE	4	3,584	-1,861
Missing	11	-	-

Quadro 5- CPD

Variáveis	Frequências	Quantificação das Categorias	
		1	2
Muito Baixo CPC	1	-1,347	-2,924
Baixo CPC	24	-0,839	-0,992
Médio CPC	34	-0,146	0,621
Alto CPC	14	1,118	0,789
Altíssimo CPC	3	4,188	-2,825
Missing	12	-	-

Pretende-se identificar as categorias que efetivamente geram maior nível de diferenciação entre os objetos em análise. Para tanto, não são consideradas as quantificações próximas de zero. Ainda que se lhes possa ser feita referência, o que se pretende ressaltar é que não são elas as responsáveis pela tradução das diferenças.

Tomando-se por referência o sinal das quantificações, assinaladas em cada um dos quadros 4 e 5, aferiu-se sobre o tipo de associações como de oposição que existem entre as diferentes categorias dessas variáveis. O objetivo é poder avaliar como é que cada uma das duas dimensões diferencia as IES do ENADE 2007.

Para facilitar a apresentação dessa análise, optou-se por apresentar os resultados de forma esquemática, conforme elucida o quadro 6.

Quadro 6- Dimensão 1-Tipos de Discriminação

Variáveis	Quantificação das Categorias(QC)	
	QC<0	QC>0
<ul style="list-style-type: none"> ❖ ENADE ❖ CPC 	<ul style="list-style-type: none"> ❖ Muito baixo ENADE ❖ Baixo ENADE ❖ Muito Baixo CPC ❖ Baixo CPC ❖ Médio CPC 	<ul style="list-style-type: none"> ❖ Bom ENADE ❖ Ótimo ENADE ❖ Alto ENADE ❖ Altíssimo ENADE

A dimensão 1 reflete uma associação entre as categorias de pior desempenho das IES e uma associação entre as categorias de melhor desempenho das IES. O grupo de instituições de pior desempenho se opõe ao grupo de instituições de melhor desempenho. Assim a dimensão 1 afasta as IES em função de seus desempenhos no ENADE e CPC.

Quadro 7- Dimensão 2-Tipos de Discriminação

Variáveis	Quantificação das Categorias(QC)	
	QC<0	QC>0
❖ CPC	<ul style="list-style-type: none"> ❖ Muito baixo CPC ❖ Baixo CPC ❖ Altíssimo CPC 	<ul style="list-style-type: none"> ❖ Médio CPC ❖ Alto CPC

A dimensão 2 aponta a associação de instituições com muito baixo CPC, baixo CPC e altíssimo CPC. Uma outra associação existe entre as variáveis médio CPC e alto CPC. Em oposição, aparecem predominantemente as categorias de baixo desempenho com as de melhor desempenho.

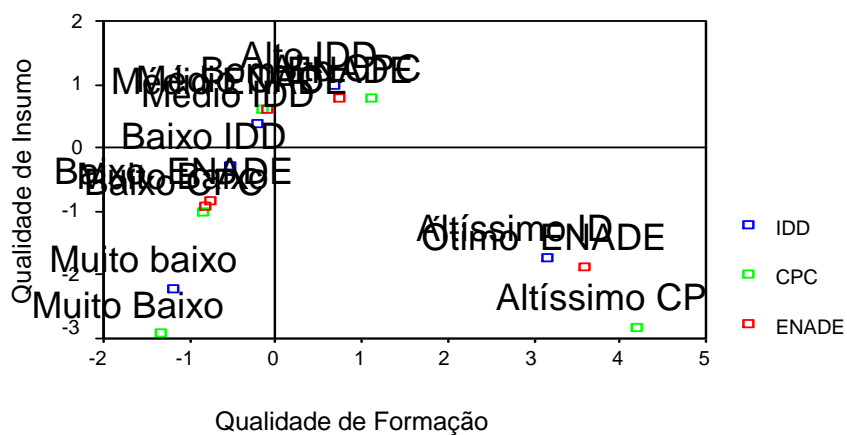
4.2-Identificação de Configurações ou de Grupos Homogêneos no Espaço de Análise

4.2.1-Interpretação dos Planos das Categorias

4.2.1.1-Leitura de diferentes configurações de categorias por via das projeções de todas as categorias identificadas em 4.1.2.2. O Objetivo desta fase é identificar grupos homogêneos.

Gráfico 3- Gráfico das Categorias

Gráfico das Categorias



O gráfico 3 representa o plano que cruza as duas primeiras dimensões. Tem-se assim o espaço do desempenho das instituições de ensino superior no Rio de Janeiro segundo os dois principais eixos estruturantes: **qualidade de formação** e **qualidade de insumo**.

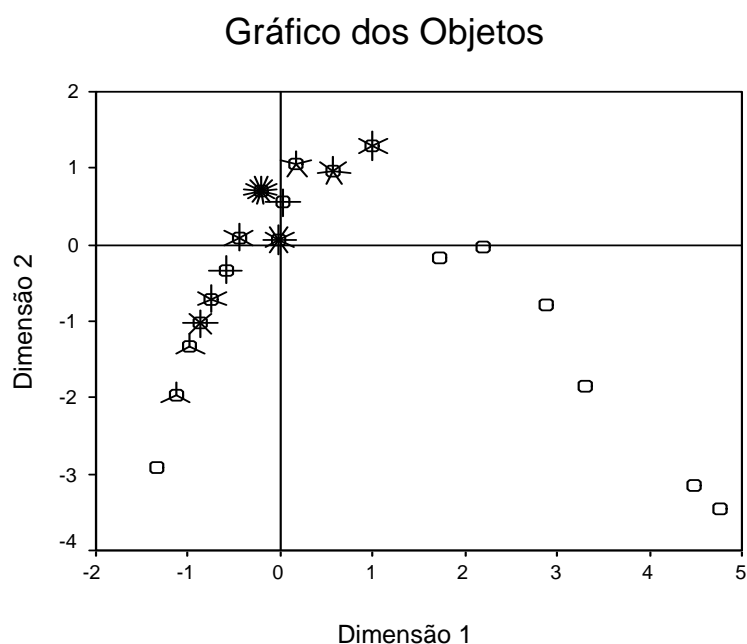
Como se destaca, este plano reflete a presença de três perfis de instituição, cuja identificação foi privilegiada pela proximidade entre as categorias, procurando-se núcleos de homogeneidade.

No 4º quadrante, pode-se identificar um perfil de instituições com altíssimo desempenho no ENADE 2007. No 2º quadrante, observa-se um grupo de instituições com alto e médio desempenho e finalmente no 3º quadrante observa-se um grupo de instituições com baixa e baixíssima avaliação.

4.2.2-Interpretação dos Planos de Objetos

4.2.2.1-Observação da intensidade da presença dos grupos já identificados:

Gráfico 4- Gráfico dos Objetos



Verifica-se pela observação do gráfico 4 que existe uma concentração grande de instituições localizadas nos 2º e 3º quadrantes, onde pelo gráfico 3 foram identificadas as instituições com mau e médio desempenho no ENADE 2007. No 4º quadrante, existe uma quantidade bem menor de instituições, que é justamente as com altíssimo desempenho.

A configuração das IES no plano de desempenho na qualidade de ensino superior é preocupante, pois revela que uma quantidade expressiva delas forma grupos com perfil de mediana e baixa avaliação de qualidade de ensino no Rio de Janeiro. A grande maioria das instituições no Rio de Janeiro precisa realizar efetivas tomadas de decisão, esforços, no sentido de proporcionar a qualidade de formação, corpo docente e infra-estrutura que a sociedade precisa para impulsionar o desenvolvimento social-econômico da cidade e do país.

5- A Articulação da HOMALS com a Análise de Clusters

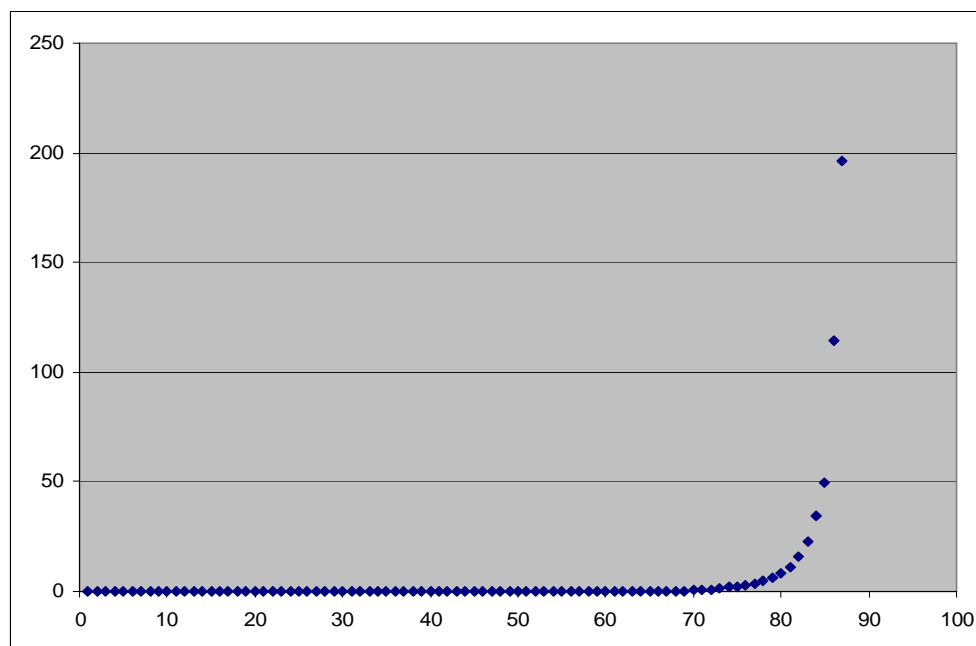
O objetivo desta seção é identificar cada instituição da base de dados em função dos perfis identificados na HOMALS. Para tanto foi realizada uma análise classificatória das IES aplicando a Análise de Clusters, em dois estágios:

- ❖ *Método Hierárquico, para validar o número de clusters a ser segmentada a base de dados;*
- ❖ *Método não hierárquico K-Means, para a classificação propriamente dita.*

Aplicando o método hierárquico, foram gerados os coeficientes de fusão, para qual foi construído o gráfico 5. Verifica-se que para mais de três grupos as

distâncias deixam de ser significativas. Pode observar-se que o declive é bastante menos acentuado a partir do terceiro agrupamento.

Gráfico 5-Coefficientes de Fusão



Concluída esta demonstração retoma-se a solução otimizada com a realização do método de *K-Means*. Classificados os casos, é então necessário fazer a correspondência entre a solução de três clusters e os três perfis configurados pelo plano da HOMALS. Esta operação pode ser realizada através do cruzamento de cada uma das variáveis usadas na definição dos perfis, via HOMALS com os clusters.

O resultado encontra-se no quadro 8. A partir das percentagens em negrito neste quadro, as mais elevadas em termos relativos para cada um dos três grupos, pode fazer-se uma síntese das suas principais características.

Como se pretende caracterizar cada um deles, no sentido de identificar os perfis configurados pela Análise de Homogeneidade, procedeu-se a uma análise comparativa. Tem-se então:

Clusters 1 com apenas 4,5%- identifica o segmento de instituições de ensino superior no Rio de Janeiro com excelente desempenho no ENADE 2007. Se observar a base de dados estas instituições são todas formadas por instituições públicas de ensino;

Clusters 2 com 28,4%- representa o segmento de instituições com baixo e baixíssimo desempenhos nos cursos contemplados neste ENADE. Inclui totalmente universidade privadas de ensino superior no Rio de Janeiro participantes do ENADE 2007;

Clusters 3 com 67,1%- identifica as instituições com desempenho mediano. É neste grupo que se concentra a grande maioria das instituições de ensino superior no Rio de Janeiro, inclusive a PUC - Rio e algumas universidades públicas.

Quadro 8- Caracterização dos Três Clusters

		Clusters		
		1	2	3
Variáveis	Categorias	Col %	Col %	Col %
ENADE	<i>Sem Conceito</i>	0	0	19
	<i>Muito Baixo ENADE</i>	0	8	0
	<i>Baixo ENADE</i>	0	76	14
	<i>Médio ENADE</i>	0	16	41
	<i>Bom ENADE</i>	25	0	25
	<i>Ótimo ENADE</i>	75	0	2
Sutotal		100	100	100
IDD	<i>Sem Conceito</i>	50	12	20
	<i>Muito baixo IDD</i>	0	16	0
	<i>Baixo IDD</i>	0	48	14
	<i>Médio IDD</i>	0	24	42
	<i>Alto IDD</i>	0	0	22
	<i>Altíssimo IDD</i>	50	0	2
Sutotal		100	100	100
CPC	<i>Sem Conceito</i>	0	0	20
	<i>Muito Baixo CPC</i>	0	4	0
	<i>Baixo CPC</i>	0	96	0
	<i>Médio CPC</i>	0	0	58
	<i>Alto CPC</i>	25	0	22
	<i>Altíssimo CPC</i>	75	0	0
Sutotal		100	100	100

Para validar graficamente a consistência desta classificação, pode ainda proceder-se à projeção dos três clusters no plano cuja configuração os havia sugerido. É o que mostra o gráfico 6. O gráfico permite verificar que a posição ocupada por cada um dos clusters está muito próxima das diversas categorias que caracterizam os respectivos perfis. Conforme evidencia a representação gráfica, a tipologia de segmentação afigura-se coerente com os resultados logo obtidos com a realização da Análise de Homogeneidade.

uma HOMALS é, sem dúvida, um dos mais habituais e cujo enfoque foi neste artigo privilegiado.

A análise de dados citada acima permitiu visualizar num espaço de menores dimensões três configurações ou perfis de desempenho das instituições no Rio de Janeiro: baixo, mediano e alto desempenho. A grande maioria das instituições deste ENADE podem ser consideradas como de desempenho mediano (67,1%), com destaque também para as que tiveram baixo desempenho (28,4%).

Numa perspectiva de completabilidade, foi proposto classificar as instituições estudadas em função de seu grau de desempenho na avaliação em foco, o que pode ser viabilizado através da Análise de Clusters. A articulação da Análise de Homogeneidade com a Análise de Clusters permitiu validar a consistência da classificação. A tipologia de segmentação revelou-se coerente com a configuração topológica revelada pela Análise de Homogeneidade.

Os dados deste estudo mostram que muitas instituições de ensino superior precisam melhor ainda em muito para oferecer cursos de graduação de excelência no Rio de Janeiro que promova a capacidade intelectual e profissional necessária para que os jovens brasileiros possam se auto realizar em suas carreiras, além impulsionarem o devido desenvolvimento do país.

7-Bibliografia

- [1] **Carvalho**, Helena. Análise Multivariada de Dados Qualitativos- Utilização da HOMALS com SPSS. Edições Silabo. Lisboa, 2004.
- [2] **Geer**, John Van de. Multivariate Analysis of Categorical Data: Theory. Sage Publications. USA,1993.
- [3] **Greenacre**, Michel. Theory and Applications of Correspondence Analysis. Academic Press. London, 1996.
- [4] **Reis**, Elizabeth. Estatística Multivariada Aplicada. Edições Silabo. Lisboa, 1997.
- [5] **Young**, Forrest. Quantitative Analysis of Qualitative Data. Psychometrika, Vol.46, no 4, pp.357-388.
- [6] SPSS Statistical Algorithms (sd). Chicago, SPSS Inc.2003.
- [7] **Maroco**, João. Análise Estatística com Utilização do SPSS. Edições Silabo. Lisboa, 2003.
- [8] **Reis**, Elizabeth & **Moreira**, Raul. Pesquisa de Mercado. Edições Silabo.Lisboa,1993
- [9] INDICADOR PRELIMINAR DOS CURSOS DE GRADUAÇÃO. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ano Referência 2007.
- [10] **Meulman**, Jacqueline. The Integration of Multidimensional Scaling and Multivariate Analysis with Optimal Transformations. Psychometrika, Vol.57, no 4, pp.539-565.