

Análise de Componentes Principais Aplicada a Avaliação Discente: Um Estudo de Caso em Ambientes Virtuais de Aprendizagem

**T. E. V. Silva, T. I. A. Souza, F. F. Barros Filho, F. J. Santos, P. R. B. Gomes, G. Ribeiro
A. O. Nunes, F. H. L. Vasconcelos**

¹Universidade Federal do Ceará (UFC)
GPEC - Grupo de Pesquisa em Modelagem Computacional Aplicada
Campus do Pici – Fortaleza – CE – Brasil

{thomazveloso, thiagoiachiley}@gmail.com, herbert@virtual.ufc.br

Abstract. *This paper describes the use of principal component analysis (PCA) as a tool for data analysis of student learning assessment in undergraduate course. Linear transformations will be described referring to the mathematical model and the software used for your application. This experiment was applied to data collected in the form of scores obtained by students in activities of forums, chat and portfolios in a Virtual Learning Environment. The results suggest activities and content that have high degrees of correlation, and the degree of significance of the scores obtained in certain activities.*

Resumo. *Este artigo descreve a utilização da análise de componentes principais (PCA) como ferramenta de análise de dados da avaliação de aprendizagem discente em cursos de graduação semi-presencial. Serão descritas as transformações lineares referentes ao modelo matemático, bem como o software utilizado para sua aplicação. Esse experimento foi aplicado em dados coletados na forma de notas obtidas pelos alunos em atividades de fóruns, chat e portfólios em um Ambiente Virtual de Ensino. Os resultados sugerem atividades e conteúdos que apresentam elevados graus de correlação, além dos graus de significância das notas obtidas em determinadas atividades.*

1. Introdução

Avaliação é um ato ou efeito de mensurar, dar valor. De acordo com [Andriola 2002], a avaliação é um processo que implica uma reflexão crítica sobre a prática, no sentido de captar seus avanços, suas resistências, suas dificuldades e possibilitar uma tomada de decisão sobre o que fazer para superar os problemas identificados.

Tratando-se de avaliação que tipos de critérios básicos são utilizados neste processo? O que é avaliado em uma atividade/tarefa? Quais são os resultados que o professor espera de seus alunos? Quais as relações existentes entre conteúdos de mesma área do conhecimento? Portanto, entende-se que a avaliação é um processo de decisão seletiva, um processo que envolve conteúdos e objetivos e que dessa forma permeia com bastante ênfase o campo da subjetividade [Bassani and Behar 2009].

No contexto da avaliação da aprendizagem em Educação a Distância (EaD), diversos trabalhos [Sales et al. 2008][Vasconcelos 2008] apontam para o desenvolvimento

de ferramentas avaliativas dos recursos trabalhados nos Ambientes Virtuais de Aprendizagem (AVA). Para [Bassani 2006] os AVA dispõem de diversas informações de registro, tais como: datas e horários de acesso, participação nos fóruns e chats, testes online, dentre outras. Nessa perspectiva, o objetivo deste trabalho é propor a aplicação da técnica de redução de dimensionalidade, denominada Análise de Componentes Principais (do inglês *Principal Components Analysis - PCA*[Moita-Neto 2009]), em notas obtidas pelos alunos nos fóruns, chat e portfólios bem como a correlação existente entre as atividades/conteúdos em duas disciplinas de mesma componente curricular de uma Instituição Federal de Ensino Superior (IFES).

Este artigo está organizado em sete seções. Na seção 2 será feito um breve estudo da literatura, apontando os trabalhos atuais que utilizam análise de dados. Na seção 3, será apresentada alguns conceitos sobre avaliação da aprendizagem em Ensino a Distância. A seção 4 ilustra a ferramenta matemática de análise de dados, o PCA. Na seção 5, descrevem-se os procedimentos metodológicos de investigação e coleta de dados desta pesquisa, bem como o *software* utilizado para a análise. Na seção 6 é apresentado a análise dos resultados obtidos com a aplicação do método PCA. Por fim, na seção 7 são discutidas as considerações finais deste trabalho e suas perspectivas futuras.

2. Trabalhos Relacionados

Uma motivação para este trabalho foi uma afirmação feita por [Soares 2005] onde ele diz que a "hostilidade em relação aos métodos quantitativos e à estatística (na ciência social brasileira)". Realizando um breve estudo na literatura, encontramos o trabalho de [Santos and Coutinho 2000] o qual corrobora com esse diagnóstico. Isso se deve ao fato que os dados levantados por esse autores apontam para uma mesma inferência: a utilização de técnicas básicas de estatística descritiva e inferencial ainda é bastante limitada nas Ciências Sociais brasileira, entre elas a educação. Alguns trabalhos recentes na área da educação [Almeida and Pimentel 2010] [Baker et al. 2011] investigam base de dados provenientes do processo de ensino aprendizagem. De acordo com [Baker et al. 2011], as informações relacionadas aos dados podem ser úteis para a tomada de decisão mais coerente por parte do professor.

3. Avaliação da Aprendizagem

No entender de [David et al. 2007], para não ser autoritária e conservadora, a avaliação tem a tarefa de ser diagnóstica, ou seja, deverá ser o instrumento dialético do avanço do aluno, e terá de ser o instrumento da identificação de novos rumos. Ainda de acordo com [David et al. 2007], a avaliação deverá, verificar a aprendizagem não só a partir dos requisitos mínimos para aprovação, mas a partir dos requisitos necessários para uma absorção significativa do conteúdo. Enfatiza também a importância dos critérios, pois a avaliação não poderá ser praticada sob dados inventados pelo professor, apesar da definição desses critérios não serem fixos e imutáveis, modificando-se de acordo com a necessidade de alunos e professores.

3.1. Avaliação da Aprendizagem em AVA

De acordo com [Bassani and Behar 2009], a definição de um processo de avaliação coerente tem sido uma das dificuldades que se destacam na modalidade EaD. Um dos contextos atuais da EaD no Brasil, é o programa Universidade Aberta do Brasil (UAB), que

se utiliza de tecnologias computacionais de informação e comunicação através de Ambientes Virtuais de Aprendizagem (AVA) para promover cursos de extensão, graduação e pós-graduação.

Com o crescimento da utilização desses ambientes em diversas instituições de ensino [Sales et al. 2008], [Vasconcelos 2008] abordam o desenvolvimento de ferramentas com o objetivo de melhorar o processo de avaliação da aprendizagem. Essa preocupação é importante, pois nesses ambientes o conhecimento é concretizado a partir de interações entre os participantes, onde cada um constrói seu conhecimento a partir da perspectiva do outro.

Segundo [Bassani 2006] os AVAs possibilitam o acompanhamento da frequência e da produção de cada aluno, dessa forma esses ambientes tornam-se uma grande base de dados que armazenam: frequência e assiduidade (são salvas datas e horários de acesso ao ambiente e ferramentas disponibilizadas), resultados dos testes, trabalhos realizados com datas para entrega pré-estabelecidas pelos professores, dessa forma as atividades entregues após o período pré-estabelecido não são aceitas, e mensagens trocadas entre os participantes de uma determinada atividade ou curso a partir de fóruns de discussão ou chats.

Para [Bassani and Behar 2009] a avaliação em AVAs pode ser entendida a partir das seguintes perspectivas: avaliação por meio de testes on-line apresenta limitações, pois os alunos devem responder um conjunto de questões predefinidas e cabe ao sistema computacional realizar as correções, dessa forma a nota final é vista como um produto do conhecimento; avaliação da produção individual deve ser feita cada vez mais no âmbito coletivo, pois possibilita ao professor o acompanhamento do processo de construção de conceitos por parte dos alunos e torna-o consciente de seu processo de aprendizagem.

4. Análise de Componentes Principais - PCA

A Análise de Componentes Principais (PCA) é um dos métodos da estatística multivariada, que tem por finalidade identificar a relação entre características extraídas dos dados visando sua redução, eliminação de sobreposições e a escolha das formas mais relevantes dos mesmos a partir de combinações lineares das variáveis originais [Moita-Neto 2009]. Sendo também denominada como Transformada de Hotelling, o PCA transforma variáveis discretas em coeficientes descorrelacionados através de uma transformação linear aplicada nos dados, de modo que, os dados resultantes tenham suas componentes mais relevantes nas primeiras dimensões, denominadas como componentes principais [Lay 2007].

Como princípio para o cálculo do PCA, considere um vetor aleatório $X = (X_1, X_2, \dots, X_p)$, contendo p componentes, com um vetor de médias $\mu = E(X) = (\mu_1, \mu_2, \dots, \mu_p)$. A matriz de Covariâncias do vetor aleatório X , quadrada de dimensão p , é denotado por: $Cov(X) = \Sigma_{p \times p}$. A matriz de covariância é uma matriz simétrica, não negativa, ou seja, $a^T \Sigma a > 0$ para todo vetor de constantes $a \in R^p$. Esta condição implica que os autovalores da matriz $\Sigma_{p \times p}$ denotados por $\lambda_1, \lambda_2, \dots, \lambda_p$, são não-negativos, ou seja, $\lambda_i \geq 0$, para qualquer $i = 1, 2, \dots, p$ (Graybill, 1983).

Pelo teorema da Decomposição Espectral [Lay 2007], sendo $\Sigma_{p \times p}$ uma matriz de covariância, existe uma matriz ortogonal $O_{p \times p}$, isto é, $O^T O = O O^T = I$, tal que:

$$O^T \Sigma O = \Theta \quad (1)$$

em que $\lambda_1, \lambda_2, \dots, \lambda_p$, são os autovalores da matriz $\Sigma_{p \times p}$ ordenados em ordem decrescente. Nesse caso, dizemos que a matriz $\Sigma_{p \times p}$ é similar a matriz Θ .

A i -ésima coluna da matriz Θ é o autovetor normalizado e_i correspondente ao autovetor λ_i , com $i = 1, 2, \dots, p$; que é denotado por $e_i = (e_{i1}, e_{i2}, \dots, e_{ip})^T$. Então a matriz Θ é dada por $\Theta = [e_1 e_2 \dots e_p]$ e pelo teorema da decomposição espectral tem-se a seguinte igualdade válida:

$$\Sigma_{p \times p} = \sum_{i=1}^p \lambda_i e_i e_i^T = O \Theta O^T \quad (2)$$

Como $\Theta_1, \Theta_2, \dots, \Theta_p$ formam uma base de R^p , o vetor a pode ser escrito como $\sum_{i=1}^p \alpha_i \Theta_i = \alpha^T O$ para algum $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)^T$.

Sendo Θ ortogonal, $\alpha^T \alpha = 1$ e a variância de $a^T X$ é menor ou igual a λ_1 e tomando $a = O_1$, tem-se que $var(O_1 X) = O_1 \Sigma O_1 = \lambda_1$, e define-se a variável aleatória $U_1 = O_1^T X$ como o primeiro componente principal de X . Para a obtenção de outros componentes principais é feita uma restrição de não correlação do próximo componente U_i com os componentes anteriormente obtidos (U_1, \dots, U_{i-1}). Desta forma as componentes são definidas como vetores aleatórios $U = (U_1, \dots, U_p) = O^T X$, onde as colunas de O são os autovetores de Σ . É importante ressaltar que a matriz de covariância da nova matriz U é diagonal, onde os elementos são os autovalores λ_i .

Verifica-se, portanto, que as variáveis aleatórias que constituem o vetor U são descorrelacionadas entre si consolidando assim o método, que estabelece um novo sistema de coordenadas cujos eixos agora estarão na direção dos autovetores de Σ .

Dentre as principais aplicações do PCA, encontram-se a: compressão de informação (voz e imagem) e redução de dimensionalidade (seleção de atributos e modelos) [Mingoti 2005].

Na aplicação de redução de dimensionalidade, o PCA tem a propriedade de minimizar o erro quadrático médio entre os dados reconstruídos e os dados originais. Supõe-se, por exemplo, que se tem dados de entrada X de dimensionalidade m e dados de saída Y de dimensionalidade m_1 , em que $m_1 < m$.

5. Procedimentos da Pesquisa

5.1. Caracterização dos Dados Coletados

Para o desenvolvimento desta pesquisa, foi realizada uma análise do desempenho de 140 alunos dos cursos de Licenciatura Plena em Química e Matemática na modalidade semi-presencial em uma Instituição Federal de Ensino Superior (IFES). A coleta de dados ocorreu através das notas obtidas em 14 atividades virtuais realizadas pelos alunos (8 fóruns, 1 chat e 5 portfólios). Os 140 alunos estão dispostos em duas disciplinas, Física Introdutória I e Física Introdutória II, onde cada disciplina tem 70 alunos.

5.2. Conteúdos das Atividades Analisadas

Os conteúdos das atividades realizadas abordam assuntos referentes à Física Básica. As Tabelas 1 e 2 apresentam os conteúdos referentes as atividades de fórum em ambas as

disciplinas. Nas Tabelas 3 e 4 são apresentados os conteúdos referentes as atividades de portfólio, os quais tiveram como atividades do curso apenas listas de exercícios.

<i>Fórum</i>	<i>Conteúdo</i>
1	Grandezas Físicas
2	Movimentos
3	Movimentos
4	Leis de Newton
5	Movimento Circular Uniforme
6	Energia Mecânica
7	Ondas Mecânicas
8	Dilatação

Tabela 1. Conteúdos das atividades de Fórum de Física Introdutória I.

<i>Fórum</i>	<i>Conteúdo</i>
1	Eletrização por Contato
2	Quantização da Carga Elétrica
3	Interação entre Cargas
4	Movimento de Cargas em um Potencial Elétrico
5	Carga de um Capacitor
6	Energia no Capacitor
7	Corrente e Diferença de Potencial
8	Associação de Resistores

Tabela 2. Conteúdos das atividades de Fórum de Física Introdutória II.

<i>Portfólio</i>	<i>Conteúdo</i>
1	Lista de Exercícios - Grandezas Físicas
2	Lista de Exercícios - Movimentos
3	Lista de Exercícios - Movimentos
4	Lista de Exercícios - Leis de Newton
5	Lista de Exercícios - Movimento Circular Uniforme

Tabela 3. Conteúdos das atividades de Portfólio de Física Introdutória I.

<i>Portfólio</i>	<i>Conteúdo</i>
1	Lista de Exercícios - Carga Elétrica/Lei de Coulomb
2	Lista de Exercícios - Potencial Elétrico
3	Lista de Exercícios - Capacitores e Dielétricos
4	Lista de Exercícios - Corrente Elétrica
5	Lista de Exercícios - Campo Magnético

Tabela 4. Conteúdos das atividades de Portfólio de Física Introdutória II.

5.3. *Software Statistical Package for the Social Sciences*

O software *Statistical Package for the Social Sciences* (SPSS) é um sistema de análise estatística em um ambiente gráfico de manuseamento de dados, recomendado para a análise de fenômenos ou processos sociais, econômicos, psicológicos e educacionais [SPSS 2007].

Para este trabalho utilizaremos o pacote de análise fatorial do SPSS, que tenta identificar variáveis subjacentes, ou fatores, que explicam o padrão de correlações dentro de um conjunto de variáveis observadas. Análise fatorial é frequentemente utilizada em redução de dados para identificar um pequeno número de fatores que explicam a maior parte da variância que é observado em um número muito maior de variáveis [Lay 2007]. Como método de extração dos fatores, utilizaremos a análise de componentes principais [SPSS 2007].

6. Análise e Discussão dos Resultados

6.1. Validando a Aplicação do PCA

Para determinar se a aplicação da análise em componentes principais tem validade para as variáveis selecionadas, realizamos os testes de Kaiser-Meyer-Olkin e de esfericidade de Bartlett [SPSS 2007]. Esses testes verificam o grau de correlação entre os dados estudados [Pasquali 2005]. Se os dados estiverem muito correlacionados, então é interessante aplicar um método para descorrelacionar os dados como é o caso do PCA.

- Teste Kaiser-Meyer-Olkin (KMO) - Indica a proporção da variância dos dados que pode ser considerada comum a todas as variáveis, ou seja, que pode ser atribuída a um fator comum, então: quanto mais próximo de 1 (unidade) melhor o resultado, ou seja, mais adequada é a amostra à aplicação da análise fatorial.
- Teste de esfericidade de Bartlett - Testa se a matriz de correlação é uma matriz identidade, o que indicaria que não há correlação entre os dados. Dessa forma, procura-se para um nível de significância assumido em 5% rejeitar a hipótese nula de matriz de correlação identidade.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.571
Bartlett's Test of Sphericity	Approx. Chi-Square
	557,840
df	91
Sig.	.000

Figura 1. Teste KMO e Bartlett para Física Introdutória I.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.766
Bartlett's Test of Sphericity	Approx. Chi-Square
	497,252
df	91
Sig.	.000

Figura 2. Teste KMO e Bartlett para Física Introdutória II.

Juntos, esses testes fornecem um padrão mínimo que deve ser passado antes que a análise de componentes principais seja realizada.

Os testes indicam a adequação dos dados para a realização da análise fatorial da Matriz de Correlação em ambas as disciplinas. Em todos os casos reportados nas Figuras 1 e 2, as amostras mostraram-se adequadas para a aplicação da análise fatorial, onde KMO > 0,5 e Bartlett com rejeição de hipótese nula [Moita-Neto 2009].

6.2. Selecionando as Componentes Principais

Para que possamos definir o número de componentes principais que serão utilizadas é necessário estabelecer o percentual da variância total explicada por cada componente associada ao seu respectivo autovalor (Tabelas 5 e 6).

Componentes	Autovalores Iniciais		
	Total	% Variância	% Acumulativa
1	3,572	25,511	25,511
2	2,522	18,014	43,525
3	1,676	11,975	55,500
4	1,338	9,557	65,057
5	1,002	7,158	72,216
6	0,820	5,860	78,076
7	0,708	5,054	83,130
8	0,657	4,695	87,826
9	0,565	4,036	91,862
10	0,436	3,112	94,974
11	0,338	2,411	97,385
12	0,237	1,695	99,080
13	0,120	0,860	99,940
14	0,008	0,060	100

Tabela 5. Variância Explicada dos dados de Física Introdutória I.

Componentes	Autovalores Iniciais		
	Total	% Variância	% Acumulativa
1	5,651	40,361	40,361
2	1,738	12,413	52,774
3	1,288	9,201	61,975
4	1,148	8,197	70,172
5	0,958	6,841	77,013
6	0,653	4,664	81,677
7	0,521	3,722	85,399
8	0,448	3,200	88,598
9	0,425	3,034	91,633
10	0,356	2,540	94,172
11	0,292	2,088	96,260
12	0,271	1,936	98,195
13	0,143	1,021	99,217
14	0,110	0,783	100

Tabela 6. Variância Explicada dos dados de Física Introdutória II.

No intuito de confirmar a fundamentação para a escolha das componentes que serão objetos de análise, aplicamos o Teste Scree sobre o qual obtemos as Figuras 3 e 4. Este gráfico ilustra como a variabilidade dos dados está distribuída entre os eixos de ordenação, permitindo identificar o número de eixos mais significativos [Moita-Neto 2009].

Conforme as Figuras 3 e 4, nota-se que a taxa de variação dos autovalores em relação ao número de componentes decresce abruptamente em um determinado ponto do gráfico. Na Figura 3 este decaimento ocorre até a componente cinco, no qual a partir desse ponto o gráfico fica praticamente na horizontal, ou seja, o acréscimo de componentes na análise representa um incremento relativamente pequeno em termos da representatividade dos autovalores, e na Figura 4 o decaimento abrupto ocorre até a componente quatro.

Pelo critério de Kaiser [Moita-Neto 2009], na nossa análise utilizaremos apenas as componentes principais com autovalores associados maiores do que um. Os dados iniciais demandam um total de 14 componentes principais, relativas a cada atividade do curso, com a aplicação do método PCA, podemos reduzir para 5 componentes para Física

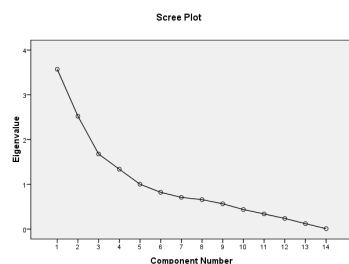


Figura 3. Gráfico Scree para os autovalores referentes a disciplina Física Introdutória I.

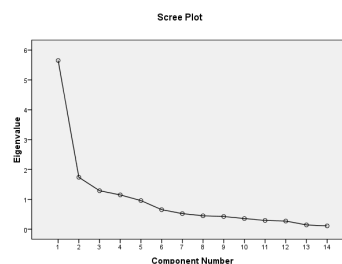


Figura 4. Gráfico Scree para os autovalores referentes a disciplina Física Introdutória II.

Introdutória I, as quais representam uma variância explicada de 72,216% dos dados originais e na disciplina Física Introdutória II reduzimos o número de componentes a 4, as quais representam 70,172% da variância explicada dos dados.

6.3. Análise da Matriz de Correlação

Para estabelecer as relações entres as variáveis avaliadas utilizou-se o coeficiente de correlação linear de Pearson (ρ), o qual é caracterizado por [Dancey and Reidy 2006] pelos seguintes níveis de correlação:

- $\rho > 0.70$ - Forte correlação.
- $0.30 > \rho > 0.70$ - Correlação moderada.
- $0 > \rho > 0.30$ - Fraca correlação.

O coeficiente de correlação Pearson varia de -1 a 1. O sinal indica direção positiva ou negativa do relacionamento e o valor sugere a força da relação entre as variáveis. Uma correlação perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra [Dancey and Reidy 2006].

Nas subseções 6.3.1 e 6.3.2 serão apresentadas as matrizes de correlação obtidas através dos dados coletados. Para a análise foi utilizado os parâmetros do coeficiente de correlação de Pearson apontados anteriormente. Por questão de espaço as atividades tiveram seus nomes abreviados para F, referente aos Fóruns, e P, referente ao portfólio. Serão analisados apenas aqueles casos que tenham forte correlação.

6.3.1. Disciplina Física Introdutória I

A Tabela 7 representa a matriz de correlação entre os dados da disciplina Física Introdutória I, os quais se referem aos conteúdos ministrados na disciplina como foi mostrado anteriormente nas Tabelas 1 e 3.

Podemos destacar dentre os dados obtidos, o alto grau de correlação existente entre os fóruns 6 e 7, com um fator de carregamento de 0,758, e entre os fóruns 2 e 3, com um fator de carregamento de 0,990. Pelos graus de correlação obtidos, podemos inferir que quando as notas dos fóruns 2 e 6 aumentam, as notas dos fóruns 3 e 7 também aumentam, respectivamente [Figueiredo-Filho and Silva-Júnior 2009]. Realizando uma

Matriz de Correlação														
	F1	F2	F3	F4	F5	F6	F7	F8	Chat	P1	P2	P3	P4	P5
F1	1	0,490	0,495	0,660	0,337	0,296	0,323	0,158	0,187	0,197	-0,025	0,080	0,213	-0,029
F2	0,490	1	0,990	0,367	0,149	0,154	0,213	0,436	0,211	-0,010	-0,064	0,034	0,001	-0,064
F3	0,495	0,990	1	0,371	0,150	0,183	0,232	0,399	0,212	-0,012	-0,061	0,033	-0,001	-0,075
F4	0,660	0,367	0,371	1	0,547	0,181	0,362	0,243	0,331	0,143	-0,016	0,082	0,189	-0,081
F5	0,337	0,149	0,150	0,547	1	0,286	0,360	0,036	0,133	-0,095	-0,124	0,039	0,162	-0,210
F6	0,296	0,154	0,183	0,181	0,286	1	0,758	0,197	-0,039	-0,166	0,016	-0,021	-0,033	-0,119
F7	0,323	0,213	0,232	0,362	0,360	0,758	1	0,114	0,119	-0,173	0,243	0,054	-0,031	0,073
F8	0,158	0,436	0,399	0,243	0,036	0,197	0,114	1	0,105	-0,104	-0,050	-0,149	-0,111	-0,186
Chat	0,187	0,211	0,212	0,331	0,133	-0,039	0,119	0,105	1	-0,158	0,145	0,225	0,259	0,393
P1	0,197	-0,010	-0,012	0,143	-0,095	-0,166	-0,173	-0,104	0,158	1	-0,391	0,346	0,158	0,235
P2	-0,025	-0,064	-0,061	-0,016	-0,124	0,016	0,243	-0,050	0,145	0,391	1	0,300	0,169	0,420
P3	0,080	0,034	0,033	0,082	0,039	-0,021	0,054	-0,149	0,225	0,346	0,300	1	0,271	0,410
P4	0,213	0,001	-0,001	0,189	0,162	-0,033	-0,031	-0,111	0,259	0,158	0,169	0,271	1	0,331
P5	-0,029	-0,064	-0,075	-0,081	-0,210	-0,119	0,073	-0,186	0,393	0,235	0,420	0,410	0,331	1

Tabela 7. Matriz de Correlação das atividades de Física Introdutória I.

análise sobre os conteúdos abordados nos fóruns podemos destacar uma relação significativa existente entre os conteúdos de ondas mecânicas e a energia mecânica, no qual são tratados nos fóruns 6 e 7, respectivamente. Os fóruns 2 e 3 tratam de um mesmo assunto, o que corrobora a intenção da análise, a qual atesta que existe essa relação entre as notas obtidas.

6.3.2. Disciplina Física Introdutória II

A Tabela 8 representa a matriz de correlação entre os dados da disciplina Física Introdutória II, os quais tem como conteúdos estudados, aqueles referentes as Tabelas 2 e 4.

Matriz de Correlação														
	F1	F2	F3	F4	F5	F6	F7	F8	Chat	P1	P2	P3	P4	P5
F1	1	0,455	0,641	0,401	0,420	0,343	0,340	0,256	0,148	0,358	0,289	0,342	0,323	0,419
F2	0,455	1	0,554	0,320	0,423	0,135	0,188	0,270	0,035	0,133	0,455	0,288	0,177	0,371
F3	0,641	0,554	1	0,360	0,363	0,160	0,166	0,204	-0,024	0,381	0,390	0,259	0,223	0,369
F4	0,401	0,320	0,360	1	0,611	0,405	0,455	0,589	0,167	0,324	0,441	0,360	0,468	0,575
F5	0,420	0,453	0,363	0,611	1	0,253	0,327	0,611	0,182	0,316	0,687	0,504	0,458	0,501
F6	0,343	0,135	0,160	0,405	0,253	1	0,821	0,389	0,210	0,423	0,191	0,078	0,152	0,152
F7	0,340	0,188	0,166	0,455	0,327	0,821	1	0,577	0,131	0,486	0,341	0,200	0,314	0,205
F8	0,256	0,270	0,204	0,589	0,611	0,389	0,577	1	0,306	0,355	0,507	0,309	0,329	0,312
Chat	0,148	0,035	-0,024	0,167	0,182	0,210	0,131	0,306	1	0,051	-0,045	0,241	-0,035	0,112
P1	0,358	0,133	0,381	0,324	0,316	0,423	0,486	0,355	0,051	1	0,366	0,419	0,460	0,380
P2	0,289	0,455	0,390	0,441	0,687	0,191	0,341	0,507	-0,045	0,366	1	0,450	0,572	0,569
P3	0,342	0,288	0,259	0,360	0,504	0,078	0,200	0,309	0,241	0,419	0,450	1	0,454	0,501
P4	0,323	0,177	0,223	0,468	0,458	0,152	0,314	0,329	-0,035	0,460	0,572	0,454	1	0,544
P5	0,419	0,371	0,369	0,575	0,501	0,152	0,205	0,312	0,112	0,380	0,569	0,501	0,544	1

Tabela 8. Matriz de Correlação das atividades de Física Introdutória II.

Na matriz de correlação da disciplina de Física Introdutória II, podemos ressaltar a relação existente entre os fóruns 6 e 7 ($\rho = 0,821$), que tem como conteúdos chaves a Energia do Capacitor (fórum 6), e o conceito de Corrente Elétrica e Diferença de Potencial (fórum 7). A forte correlação existente indica um padrão de crescimento das notas das atividades de fóruns, ou seja, essas atividades tem correlação positiva. Esse padrão não é observável nas outras atividades, o que indica a falta de uma forte correlação entre as outras notas.

6.4. Análise do PCA

De acordo com o testes feitos anteriormente para a validação da aplicação do PCA na base de dados coletadas, as Tabelas 9 e 10 representam os fatores de cargas relacionados

a importância da atividade diante de todo o conjunto de dados, ou seja, podemos destacar o quanto significativa é a atividade em relação as outras [Moita-Neto 2009].

Componentes					
	1	2	3	4	5
F1	0,345	-0,024	0,096	-0,634	0,287
F2	0,935	0,080	-0,029	-0,076	0,042
F3	0,920	0,071	-0,005	-0,085	0,046
F4	0,211	0,093	-0,080	-0,763	0,132
F5	-0,161	0,051	0,228	-0,786	-0,171
F6	0,031	-0,142	0,859	-0,129	-0,072
F7	0,046	0,052	0,914	-0,137	-0,018
F8	0,662	-0,062	0,074	0,102	-0,126
Chat	0,232	0,783	-0,046	-0,073	-0,112
P1	-0,013	-0,118	-0,231	-0,152	0,905
P2	-0,058	0,150	0,351	0,293	0,636
P3	-0,069	0,371	0,053	-0,017	0,464
P4	-0,174	0,625	-0,114	-0,312	0,014
P5	-0,053	0,742	0,103	0,356	0,185

Tabela 9. Matriz de Componentes de Física Introdutória I.

Componentes				
	1	2	3	4
F1	0,646	-0,109	0,510	0,100
F2	0,552	-0,356	0,417	0,232
F3	0,578	-0,341	0,594	0,001
F4	0,756	0,090	-0,107	0,087
F5	0,786	-0,137	-0,189	0,196
F6	0,511	0,721	0,258	-0,115
F7	0,622	0,661	0,113	-0,203
F8	0,689	0,324	-0,222	0,206
Chat	0,209	0,326	-0,158	0,778
P1	0,621	0,179	0,028	-0,392
P2	0,745	-0,259	-0,213	-0,153
P3	0,617	-0,240	-0,304	0,133
P4	0,649	-0,175	-0,360	-0,374
P5	0,706	-0,316	-0,180	-0,021

Tabela 10. Matriz de Componentes de Física Introdutória II.

Analisando a Tabela 9, podemos observar que não há um padrão na localização dos maiores fatores de carga em relação as componentes obtidas, o que indica que nenhuma atividade se sobressai a outra, no que diz respeito ao seu nível de significância em relação as outras atividades. Ao contrário da Tabela 9, a Tabela 10 tem uma forma bem clara de localização dos maiores fatores de cargas em relação as componentes observadas, o que indica uma presença significativa das notas dos portfólios perante a base de dados. Os fóruns 2, 3 e 7 tem sua significância comprometida por ter fatores de carregamento na terceira componente da matriz, analogamente ao fator referente a atividade de chat, que está na componente menos significativa.

7. Considerações Finais

Este trabalho apresenta uma contribuição da aplicação de uma técnica da estatística multivariada, o método PCA, em dados educacionais relacionados ao processo de avaliação da aprendizagem.

Os resultados dos testes KMO e de esféricidade de Bartlett [SPSS 2007] apontam para a validação da aplicação da técnica na base de dados selecionada. A partir na análise da matriz de correlação podemos inferir sobre o quão correlacionável é uma atividade/conteúdo com outra.

A aplicação do PCA verificou a importância das atividades de portfólio na disciplina de Física Introdutória II, o que nos faz inferir sobre a necessidade de uma análise mais profunda de sua diferença em relação ao curso de Física Introdutória I.

O intuito da aplicação da ferramenta proposta neste artigo é subsidiar o professor responsável pela elaboração de cursos online, tendo em vista que esse tipo de análise trás um *feedback* estatístico que está implícito no conjunto de dados.

Observamos que a da não-trivialidade da análise dos dados não deve servir de desmotivação, tendo em vista que existe um relacionamento entre a matemática aplicada e as relações psicológicas humanas [Kroonenberg 2008]. Como trabalhos futuros, sugere-se uma abrangente análise da matriz de correlação e a aplicação de outras técnicas de análise de dados que trabalham com dimensões de ordem superior a dois [Kroonenberg 2008].

Referências

- Almeida, A. and Pimentel, E. P. (2010). Mineração de dados no ensino a distância para fins de avaliação do uso das ferramentas de aprendizagem do ambiente tidia-ae. In *Anais do III WAValia*.
- Andriola, W. B. (2002). *Detección del funcionamiento diferencial del item en tests de rendimiento. Aportaciones teóricas y metodológicas*. PhD thesis, Universidad Complutense de Madrid.
- Baker, R., Isotani, S., and de Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o brasil. *Revista Brasileira de Informática na Educação*, 19.
- Bassani, P. S. (2006). *Mapemamento das interações em ambiente virtual de aprendizagem: uma possibilidade para avaliação em educação a distância. Tese (Doutorado)*. PhD thesis, Universidade Federal do Rio Grande do Sul - PPGIE, Porto Alegre.
- Bassani, P. S. and Behar, P. A. (2009). *Modelos Pedagógicos em Educação a Distância. Avaliação da Aprendizagem em Ambientais Virtuais*. Artmed.
- Dancey, C. and Reidy, J. (2006). *Estatística Sem Matemática para Psicologia: Usando SPSS para Windows*.
- David, P. B., Pequeno, M. C., da Silva, A., Souza, C. F., Júnior, G. S. V., de Castro Filho, J. A., Ventura, P. P. B., and Maia, S. (2007). Avaliação da aprendizagem em educação a distância numa perspectiva sócio-interacionista. *Anais do XVIII Simpósio Brasileiro de Informática na Educação*.
- Figueiredo-Filho, D. B. and Silva-Júnior, J. A. (2009). Desvendando os mistérios do coeficiente de correlação de pearson. *Revista Política Hoje*, 18.
- Kroonenberg, P. M. (2008). *Applied Multiway Data Analysis*. Wiley.
- Lay, D. C. (2007). *Álgebra Linear e Suas Aplicações*.
- Mingoti, S. A. (2005). *Análises de Dados Através de Métodos de Estatística Multivariada*.
- Moita-Neto, J. M. (2009). *Estatística Multivariada na Pesquisa*, volume 5. Sapiência (FAPEPI).
- Pasquali, L. (2005). *Análise Fatorial para pesquisadores*.
- Sales, G. L., Barroso, G. C., and Soares, J. M. (2008). O indicador de aprendizagem learning vectors como instrumento automatizado de avaliação para suporte a aprendizagem em ead. *Anais do XXVIII Congresso da SBC*.
- Santos, M. H. and Coutinho, M. (2000). Política comparada: estado das artes e perspectivas no brasil. *BIB*, 54:3– 146.
- Soares, G. (2005). O calcanhar metodológico da ciência política no brasil. *Psicologia*, 48:27–52.
- SPSS (2007). *SPSS Statistics Base 17.0 User's Guide*. Inc SPSS.
- Vasconcelos, F. H. L. (2008). Avaliação sócio-interacionista aplicada ao contexto da ead em cursos de graduação semi-presenciais mediado por um ambiente virtual de aprendizagem. *Anais do XXVIII Congresso da SBC*.