METODOLOGIA COM BASE PROBABILÍSTICA PARA A MEDIÇÃO DA EFICÁCIA DO PROCESSO DE ENSINO NOS SISTEMAS PRESENCIAL E A DISTÂNCIA

Fortaleza, abril/2014

Graziella Batista de Moura – Universidade de Fortaleza – graziella@unifor.br

Mateus Mosca Viana – Universidade de Fortaleza – mosca@unifor.br

Experiência inovadora

Educação superior

Teorias e modelos – Inovação e mudança – Interação e comunicação em comunidades de aprendizagem

Relatório de estudo concluído

RESUMO

A atividade de ensino se desenvolve por meio de processos que precisam ser acompanhados e avaliados, tal qual acontece em qualquer setor de natureza produtiva. Para tanto, devem-se identificar variáveis associadas a essa atividade, cuja representação se realize por meio de elementos de alguma escala de natureza quantitativa.

Neste trabalho utilizam-se as notas parciais de alunos de uma turma, obtidas ao cursar uma disciplina, como sendo os elementos usados na quantificação da atividade. Tais notas são separadas em grupos, de modo que, a partir do conjunto das frequências relativas dos mesmos, se pode construir uma função de densidade de probabilidade.

De posse dessa função, é possível construir um índice de relacionamento para a análise do desempenho, seja em uma turma considerada individualmente, seja na comparação entre duas turmas distintas. Trata-se de uma abordagem cuja principal característica é a de proporcionar uma visão abrangente dos dados, diferente daquela obtida por meio de análises de natureza paramétrica. Esta é uma proposta apresentada como sugestão ao suporte da gestão acadêmica em cursos de nível superior, quer sejam ministrados na forma presencial, que na forma de educação a distância (EaD).

Palavras-chave: **educação a distância; probabilidade; modelo de métrica**

1. INTRODUÇÃO

A atividade de ensino se desenvolve por meio de processos que precisam ser acompanhados e avaliados, tal qual acontece em qualquer setor de natureza produtiva. Ao longo da evolução dos processos pedagógicos e andragógicos, a identificação de dados, juntamente com a sua observação e medição, torna-se uma tarefa indispensável para a obtenção de informações destinadas à gestão.

Neste trabalho são utilizados os conceitos de "probabilidade" (Kolmogorov, 1956, pp. 3-4; DeGroot, Schervish, 2012, pp. 16) e de "probabilidade condicional" (Feller, 1968, pp. 114-115), além daquele de "produto escalar entre dois vetores" (Anton, Rorres, 2012, pp. 258-260), com a finalidade de quantificar processos da atividade de ensino. Esta é uma proposta apresentada como sugestão ao suporte da gestão acadêmica em cursos de nível superior, quer sejam ministrados na forma presencial, quer na forma de educação a distância (EaD).

A estrutura do trabalho se apresenta com a seguinte composição. Na seção "2. CENÁRIO" se descreve o ambiente no qual serão recolhidos os dados, os critérios para a sua classificação, bem como o modo utilizado para organizá-los e medi-los. Por seu turno, na seção "3. MODELO" se estabelecem as medidas de natureza probabilística que se podem obter a partir dos dados coletados, além de metodologias sugeridas para o tratamento comparativo entre os mesmos. O modelo matemático é aplicado sobre dados obtidos por meio de gerador de números aleatórios e tratados da seção "4. ESTUDO DE CASO SIMULADO". Finalmente, são apresentadas as seções "5. CONCLUSÕES" e "6. REFERÊNCIAS".

2. CENÁRIO

2.1 APRESENTAÇÃO

As instituições de nível superior (IES) atualmente trabalham tanto no sistema presencial, quanto no sistema de educação a distância (EaD), cada um desses sistemas com as suas características próprias. Em qualquer dessas situações, os alunos cursam disciplinas, em geral com avaliações em duas etapas, de onde se originam duas notas parciais. A média ponderada entre essas duas notas parciais é a nota obtida pelo aluno na disciplina.

Nas IES existem critérios que estabelecem valores críticos com respeito à nota obtida por um aluno, de modo a situá-lo em um de três estados distintos, aqui denominados: (A) – **aprovação direta**; (B) - **necessidade de alguma recuperação para ser aprovado**; (C) - **reprovação direta**.

2.2 COLETA DE DADOS

Sendo o foco do presente estudo a evolução do processo de aprendizagem de alunos de uma turma, o dado fundamental a ser utilizado é o conjunto dos resultados obtidos pelo aluno, nas avaliações às quais o mesmo é submetido. Desse modo, para uma determinada disciplinas serão definidos os seguintes conjuntos de dados: (N_1) - **arquivo das notas dos alunos na avaliação 1**; (N_2) - **arquivo das notas dos alunos na avaliação $\mathbf{2}$**.

Agregando os arquivos N_1 e N_2 é possível construir um novo arquivo, denominado N_{12}, cujo conteúdo é um **conjunto de registros incluindo os campos com as notas das avaliações $\mathbf{1}$ e $\mathbf{2}$**, para cada turma. Em seguida, o arquivo N_{12} deverá ser ordenado de modo decrescente, de acordo com o campo referente à nota na avaliação 1. Depois dessa operação, o arquivo N_{12} passa a receber a denominação de D_{12}.

Estado	Critério
A	8,0 ≤ nota ≤ 10,0
B	5,0 ≤ nota < 8,0
C	0,0 ≤ nota < 5,0

Tabela 1. – Estados e critérios

A notação D_{12} significa **conjunto de registros de notas do período, dado que a ordem é decrescente com respeito à nota** 1. A tabela 1 contém os intervalos de valores críticos, a partir dos quais são definidos os estados A, B e C, conforme citados no subtópico 2.1. Considerando os limites da tabela 1, pode-se construir uma partição do conjunto D_{12}, de modo que se terá a expressão:

$$D_{12} = D_{1A2} \cup D_{1B2} \cup D_{1C2}. \quad (1)$$

Os elementos no membro à direita de (1) são respectivamente:

D_{1X2} – subconjunto de registros de notas 1 e 2 do período, dado que a ordem é decrescente com respeito à nota 1, cujo valor na nota 1 se encontra em X (que pode ser A, B, ou C).

Dentro de cada um desses subconjuntos deverão ser contados os registros cujos valores da nota 2 se encontram nos estados A, B, ou C. No conjunto D_{1A2}, por exemplo, a quantidade de cada um desses registros será denotada, respectivamente como sendo d_{1A2A}, d_{1A2B}, d_{1A2C}, ou de modo mais explícito:

d_{1A2X} – quantidade de notas 2 no estado X (que pode ser A, B, ou C), dado que a nota 1 se encontra no estado A.

Repetindo-se esse processo para o subconjunto de registros cuja nota 1 se encontra no estado B e para aquele cuja nota 1 se encontra no estado C, resume-se o resultado na tabela 2 a seguir.

Estado da nota 1	**Qtd. nota 1**	**Qtd. nota 2 dada a nota 1, por estado da nota 1**		
		A	B	C
A	d_{1A2}	d_{1A2A}	d_{1A2B}	d_{1A2C}
B	d_{1B2}	d_{1B2A}	d_{1B2B}	d_{1B2C}
C	d_{1C2}	d_{1C2A}	d_{1C2B}	d_{1C2C}

Tabela 2 – Distribuição absoluta de notas 2 por estado da nota 1

Pode-se ver na tabela 2 que, em cada linha, a soma dos valores dos elementos das três últimas colunas é igual ao valor do elemento na segunda coluna. De modo geral, fazendo X tomar respectivamente os valores A, B e C, cada linha satisfaz à expressão:

$$d_{1X2} = d_{1X2A} + d_{1X2B} + d_{1X2C} \tag{2}$$

Enfim, a tabela 2 se constitui na forma como se devem resumir os dados obtidos a partir dos resultados de uma disciplina.

3. MODELO

3.1 REPRESENTAÇÕES RELATIVAS

As expressões formadas a partir de (2), no subtópico 2.2, representam as **frequências absolutas** dos valores em estudo. Tal representação não é apropriada, quando se pretendem realizar comparações entre valores de períodos letivos diferentes. A fim de viabilizar tal comparação deve-se alterar a representação dos valores para a forma de **frequência relativa**. Nesta forma, os valores absolutos em uma linha são substituídos pelas razões entre cada um deles e o valor de sua soma.

Desse modo, cada linha referente a um estado na tabela 2, se transforma em uma nova linha, na tabela 3, conforme se pode ver a seguir.

Estado da nota 1	**Frequência relativa da nota 2, dada a nota 1, por estado da nota 1**		
	A	B	C
A	p_{1A2A}	p_{1A2B}	p_{1A2C}
B	p_{1B2A}	p_{1B2B}	p_{1B2C}
C	p_{1C2A}	p_{1C2B}	p_{1C2C}

Tabela 3 – Distribuição relativa de notas 2 por estado da nota 1

Considerando a linha referente ao estado A, por exemplo, fazendo X tomar respectivamente os valores A, B e C, os seus elementos são definidos de acordo com expressões do tipo:

$$p_{1A2X} = d_{1A2X}/d_{1A2} \tag{3}$$

Os elementos das linhas B e C podem ser calculados de modo análogo.

Verifica-se de imediato que, em cada linha da tabela 3, os seus elementos componentes possuem as seguintes propriedades (Kolmogorov, 1956, pp. 3-4; DeGroot, Schervish, 2012, pp. 16):

a) São números cujos valores se encontram no intervalo fechado com extremos zero e um (isto é, [0,1]).

b) A soma dos elementos de uma linha é igual a “$1{,}0$”.

Isso significa que, em cada linha, os elementos das três últimas colunas da tabela 3 são medidas de probabilidades. Em particular, são “**probabilidades condicionais**” (Feller, 1968, pp. 114-115) dos valores da nota 2 se situarem em cada um dos três estados, dado que a nota 1 se encontra em um estado determinado. Tomando por exemplo a linha do estado A da tabela 3, fazendo X tomar respectivamente os valores A, B e C, pode-se dizer que:

$$p_{1A2X} = Prob(nota\ 2\ no\ estado\ X | nota\ 1\ no\ estado\ A) \tag{4}$$

Essa representação permite que se possam realizar estudos comparativos entre períodos de tempo diferentes, visto que cada período se caracteriza por meio de uma “**distribuição de probabilidade**” (Kolmogorov, 1956, pp. 21-24).

3.2 COMPARANDO DENSIDADES DE PROBABILIDADES

A construção da tabela 3 foi o resultado de uma série de operações simples, envolvendo contagem de dados brutos, relativização de medidas e padronização de valores. Neste ponto, convém representar a tabela 3 na forma de uma matriz quadrada, denominada “**matriz de transição**” (Hoel, Port Stone, 1972, pp.16; Yates, Goodman, 2005, pp. 445-451), como a seguir.

$$
\begin{array}{ccc}
 & & \text{Nota 2} \\
 & & \begin{array}{ccc} \text{A} & \text{B} & \text{C} \end{array} \\
\text{Nota 1} & \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array} & \begin{bmatrix} p_{1A2A} & p_{1A2B} & p_{1A2C} \\ p_{1B2A} & p_{1B2B} & p_{1B2C} \\ p_{1C2A} & p_{1C2B} & p_{1C2C} \end{bmatrix}
\end{array}
$$

Matriz de Transição

Essa denominação se deve ao fato de que os elementos dessa matriz são valores de probabilidades condicionais de um estado da Nota 2 com respeito a um estado da Nota 1. Assim, por exemplo, p_{1B2A} significa a “**probabilidade de um aluno se encontrar no estado A da nota** 2**, sabendo-se que antes ele se encontrava no estado B da nota** 1.

De acordo com a subseção 3.1, sabe-se que cada linha da matriz de transição é composta pelos elementos de uma “**função densidade de probabilidade**” (Kolmogorov, 1956, p. 21; Yates, Goodman, 2005, pp. 105-110). Em outras palavras, cada uma das linhas é um vetor de probabilidades que apresenta a descrição de como os alunos se distribuem pelos estados A, B e C, de acordo com o valor da nota 2. Fazendo X e Y tomarem respectivamente os valores A, B e C, os vetores serão representados do seguinte modo:

$$p_{1Y} = (p_{1Y2X} \quad p_{1Y2X} \quad p_{1Y2X}) \tag{5}$$

Essa forma de representação permite que se possam realizar comparações sobre as distribuições de frequências dos alunos, tomando por base os estados de origem, e a partir daí tirar conclusões sobre a evolução dos mesmos. Essa comparação deve ser expressa na forma de um valor numérico, entre zero e um, cujo significado vai desde a total disparidade entre as distribuições, até a completa semelhança.

Tomando por base o conceito de “**produto escalar entre dois vetores**” (Anton, Rorres, 2012, pp. 258-260), considerando os vetores p_{1A} e p_{1B} como exemplo, sabe-se que

$$\langle p_{1A}, p_{1B}\rangle = \|p_{1A}\|.\|p_{1B}\|.\cos(p_{1A}, p_{1B}) \qquad (6)$$

seguindo-se que

$$\cos(p_{1A}, p_{1B}) = \frac{\langle p_{1A}, p_{1B}\rangle}{\|p_{1A}\|.\|p_{1B}\|}. \qquad (7)$$

Apesar de o conjunto de valores da função “$\cos(p_{1A}, p_{1B})$” ser o intervalo “$[-1;+1]$”, o fato de os vetores “p_{1A}” e “p_{1B}” não conterem valores negativos, pois seus componentes são notas de provas que variam de zero a dez, o conjunto de valores de (7) estará limitado ao intervalo “$[0;+1]$”. A expressão (7) permite definir um “**indicador de relacionamento**” entre os dois vetores, que será doravante denominado “$\boldsymbol{IR}$”.

Apenas para ilustrar, sob o ponto de vista geométrico (Anton, Rorres, 2012, pp. 280-285), quando “$IR = 0$”, os vetores são perpendiculares e não apresentam correlação entre si. Por outro lado, sendo “$IR = 1$”, então os vetores são coincidentes, havendo nesse caso total dependência entre os mesmos.

3.3 COMPARANDO VETORES

A simples comparação de duas densidades de probabilidades não é suficiente para estabelecer qual das duas se encontra em vantagem com respeito à outra. A densidade de probabilidades oriunda das notas de alunos será um importante indicador para a percepção da eficácia do processo de ensino-aprendizagem, se for possível utilizá-la para medir a distância entre a situação atual e a ideal. Essa última é aquela na qual o estado final de cada aluno seja “A”, independente do estado original, sendo representado pelo vetor:

$$p_A = (1 \quad 0 \quad 0)\ . \qquad (8)$$

Tomando como exemplo o vetor

$$p_{1A} = (p_{1A2A} \quad p_{1A2B} \quad p_{1A2C}),$$

o indicador de relacionamento (IR) será representado como

$$IR(p_{1A}, p_A) = \cos(p_{1A}, p_A) = \frac{\langle p_{1A}, p_A\rangle}{\|p_{1A}\|.\|p_A\|} = \frac{p_{1A2A}}{\sqrt{(p_{1A2A}{}^2 + p_{1A2B}{}^2 + p_{1A2C}{}^2)}}\ . \qquad (9)$$

Depois de calcular para cada vetor o IR entre ele e o vetor de referência p_A, o vetor cujo IR for maior será aquele que contém os melhores resultados.

4. ESTUDO DE CASO SIMULADO

Utilizando um método simples de simulação, por meio de uma planilha eletrônica, foi gerado um conjunto de registros contendo dois campos cujos valores correspondem aos das notas das duas provas. Ao todo foram gerados cento e vinte nove registros, de modo a simular uma situação semelhante à real, tratada pelos autores deste trabalho.

Construiu-se a Tabela 4 tomando por base os critérios estabelecidos na Tabela 1.

Estado da nota 1	**Qtd. nota 1**	**Qtd. nota 2 dada a nota 1, por estado da nota 1**		
		A	B	C
A	75	45	23	7
B	27	9	11	7
C	27	4	2	21

Tabela 4 - Distribuição absoluta de notas 2 por estado da nota 1

A Tabela 5 contém os valores da distribuição relativa de notas 2, por estado da nota 1.

Estado da nota 1	**Frequência relativa da nota 2, dada a nota 1, por estado da nota 1**		
	A	B	C
A	0,6000	0,3067	0,0933
B	0,3333	0,4074	0,2593
C	0,1481	0,0741	0,7778

Tabela 5 – Distribuição relativa de notas 2 por estado da nota 1

Na realidade, a Tabela 5 é a própria matriz de transição, apresentada logo a seguir.

$$\mathbf{T} = \begin{bmatrix} 0{,}6000 & 0{,}3067 & 0{,}0933 \\ 0{,}3333 & 0{,}4074 & 0{,}2593 \\ 0{,}1481 & 0{,}0741 & 0{,}7778 \end{bmatrix}$$

Uma vez de posse da matriz de transição é possível agora realizar os cálculos necessários para obter o IR, entre os vetores linha e os vetores de referência. Na seção 3, apenas foi considerado o vetor de referência para comparações com o estado A, isto é:

$$p_A = (1 \quad 0 \quad 0).$$

No entanto, nada impede que se possam utilizar vetores de referência para comparações com os estados B e C, cujas configurações são:

$$p_B = (0 \quad 1 \quad 0) \text{ e } p_C = (0 \quad 0 \quad 1).$$

Considerando a matriz de transição **T**, o valor do IR calculado para cada uma das três linhas dessa matriz, denominadas respectivamente p_{1A}, p_{1B}, p_{1C}, com respeito ao vetor de referência p_A, serão os seguintes:

$$\mathbf{IR}(\boldsymbol{p_{1A}}, \boldsymbol{p_A}) = \frac{0{,}6000}{\sqrt{(0{,}6000^2 + 0{,}3067^2 + 0{,}0933^2)}} = 0{,}8820$$

$$\mathbf{IR}(\boldsymbol{p_{1B}}, \boldsymbol{p_A}) = \frac{0{,}3333}{\sqrt{(0{,}3333^2 + 0{,}4074^2 + 0{,}2593^2)}} = 0{,}5677$$

$$\mathbf{IR}(\boldsymbol{p_{1C}}, \boldsymbol{p_A}) = \frac{0{,}1481}{\sqrt{(0{,}1481^2 + 0{,}0741^2 + 0{,}7778^2)}} = 0{,}1862$$

Os três valores calculados para o indicador permitem que se conclua que, cerca de oitenta e oito por centos alunos que se encontravam no estado “A”, com respeito ao resultado da primeira prova, permanecem ainda nesse estado. Quanto aos alunos cujos resultados obtidos na primeira prova os colocavam no estado “B”, cerca de cinquenta e sete por cento passaram para o estado “A”. Por fim, os alunos originalmente no estado “C”, foram aqueles com a menor capacidade de recuperação, pois menos de dezenove por cento ascenderam ao estado “A”.

5. CONCLUSÃO

As formas tradicionais para quantificar a avaliação se apoiam fortemente em métricas fundamentadas em medidas de tendência central, como média e mediana e moda. Sem pretender simplesmente descartar essa prática consagrada, o presente trabalho propõe uma metodologia complementar sobre como realizar o processo de avaliação, por meio de uma abordagem baseada na Teoria das Probabilidades. O foco considerado tomou por base os sistemas com memória, que é o caso daqueles de notas consecutivas em uma turma de alunos, nos quais se aplicam o conceito de probabilidade condicional. O aspecto operacional é simples de ser realizado, em especial nos ambientes que contam com os recursos da Tecnologia da Informação.

Enfim, também se pretende com o trabalho motivar novas pesquisas que possam ser realizadas levando em conta a possibilidade de acompanhar a evolução de tais sistemas por meio de modelos de processos estocásticos.

6. REFERÊNCIAS

Anton, Howard; Rorres, Chris; **Elementary Linear Algebra – Applications.** 10th ed. New Jersey. John Wiley, 2010.

DeGroot, Morris H.; Schervish, Mark J.; **Probability and Statistics**. 4th ed. Boston. Addison-Wesley. 2012.

Feller, William; **An Introduction to Probability Theory and its Applications – Vol. 1**. New York. Wiley. 1968.

Hoel, Paul G.; Port, Sidney C.; Stone, Charles J.; **Introduction to Stochastic Processes**. Los Angeles. Waveland Press, 1972.

King, Christopher; **Notes on Probability Theory**; Department of Mathematics Northeastern University, www.hamilton.ie/ollie/Downloads/ProbMain.pdf, 2009, acesso em março/2014.

Knill, Oliver; **Probability and Stochastic Processes with Applications**, www.math.havard.edu/~knill/teaching/math144_1994/probability.pdf, 2008, acesso em março/2014.

Kolmogorov, A. N.; **Foundations of The Theory of Probability**;. 2d. ed. New York. Chelsea Publishing Co. 1956.

Yates, Roy D.; Goodman, David J.; **Probability and Stochastic Processes**. 2d. ed. New Jersey. Wiley. 2005.