A cadeia de Markov na determinação de indicadores educacionais

Applying Markov chain to determine educational indicators

***Jair Mendes Marques*[1]**
***Sani de Carvalho Rutz da Silva*[2]**

Resumo

A cadeia de Markov é um caso particular de processos estocásticos com estados discretos. O objetivo deste artigo é aplicar a cadeia de Markov para determinar e interpretar indicadores educacionais. Para sua aplicação, foram utilizados dados do Ensino Médio da Região Sul do Brasil entre os anos de 2009 a 2011, obtidos no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). As análises foram realizadas considerando-se os ensinos público e privado por estado (Paraná, Santa Catarina e Rio Grande do Sul). Os indicadores desenvolvidos foram: o tempo médio esperado de permanência em cada série, o tempo médio esperado de permanência no sistema e probabilidades dos estados estacionários (abandono e término). Os resultados mostraram que, para o ensino público e o privado, os tempos médios esperados de permanência em cada série diminuem com o progresso nas séries para os três estados. O Rio Grande do Sul apresentou o maior tempo médio de permanência em cada série, e Santa Catarina apresentou o menor tempo médio. Para os tempos médios esperados para sair do sistema, o Rio Grande do Sul apresentou maior tempo, enquanto Santa Catarina apresentou menor tempo para o ensino público e o Paraná, o menor tempo para o ensino privado. As análises das probabilidades de terminar o curso mostraram que estas aumentaram com os progressos nas séries. Com relação à probabilidade de abandono, constatou-se que ela decresce com o progresso na série e que o Rio Grande do Sul tem as maiores probabilidades, independentemente da série.

Palavras-chave: Cadeia de Markov. Processos Estocásticos. Indicadores Educacionais.

Abstract

Markov chain is a particular case of stochastic processes with discrete states. The aim of this paper was to apply Markov chain to determine and interpret educational indicators. In order to apply it, data from high school in southern Brazil from the years 2009 to 2011 was used. The data was obtained through the INEP website. The analyses were performed considering public and private schools in each state (Parana, Santa Catarina and Rio Grande do Sul). The indicators developed were: the expected average time spent in each grade, the expected average time spent in the system and probabilities of permanence (dropout and termination). The results for public and private schools showed that the average time expected of permanence in each grade decreases with the progress in the grades for the three states. Rio Grande do Sul had the highest average time spent in each grade, while Santa Catarina had the lowest average time. For the average time expected to leave the system, Rio Grande do Sul had a higher average, while Santa Catarina had a lower average for public education and Paraná a lower average for private education. The analysis of the probability of finishing the course showed that it increased with the progress in the grades. Regarding the probability of dropping out, we have found that it decreases with the progress in the grade and the state of Rio Grande do Sul has the highest probability, regardless of the grades.

Keywords: Markov Chain. Stochastic Processes. Educational Indicators.

1 Doutor em Ciências Geodésicas pela Universidade Federal do Paraná. Professor da FAE Centro Universitário e do Bom Jesus. Professor Sênior da Universidade Federal do Paraná no curso de Pós-Graduação de Métodos Numéricos em Engenharia. *E-mail*: jair.m.marques@gmail.com.

2 Doutora em Ciências dos Materiais pela Universidade Federal do Rio Grande do Sul. Professora da Universidade Tecnológica Federal do Paraná. Coordenadora do Programa de Pós-Graduação em Ensino de Ciência e Tecnologia. *E-mail*: sani@utfpr.edu.br.

1 Cadeias de Markov

1.1 Processos Estocásticos

Definição: um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória.

De acordo com Morettin e Toloi (1987), processos estocásticos servem para descrever o procedimento de um sistema operando sobre algum período de tempo, com isso, em termos formais, a variável aleatória $Z(t)$ representa o estado do sistema no parâmetro *t* (geralmente tempo). Portanto, pode-se afirmar que $Z(t)$ é definido como um espaço de estados.

Os processos estocásticos podem ser classificados da seguinte forma (ROSS, 1996):

(1) Em relação ao estado

— Estado discreto (cadeia): $Z(t)$ é definido sobre um conjunto enumerável ou finito.

— Estado contínuo (sequência): $Z(t)$ é definido sobre o conjunto dos reais.

(2) Em relação ao tempo (parâmetro)

— Tempo discreto: t é finito ou enumerável.

— Tempo contínuo: *t* é real não negativo.

1.2 Processos Markovianos

Definição: um processo estocástico é markoviano se a ocorrência de um estado futuro depender somente do estado imediatamente precedente, ou seja,

$$P\{Z(t_{k+1}) \leq z_{k+1} \mid Z(t_k) = z_k, Z(t_{k-1}) = z_{k-1}, \dots , Z(t_1) = z_1, Z(t_0) = z_0\}$$
$$= P\{Z(t_{k+1}) \leq z_{k+1} \mid Z(t_k) = z_k\} \qquad (1)$$

Um processo estocástico é markoviano se a ocorrência de um estado futuro depender somente do estado imediatamente precedente.

Essa expressão pode ser interpretada como a probabilidade condicional de qualquer evento futuro, dado qualquer evento passado, e o estado presente $Z(t_k) = z_k$ é independente do evento passado e depende somente do estado presente (NORRIS, 1998; DERMAN, 1970).

As probabilidades condicionais $P\{Z(t_{k+1}) = z_{k+1}|Z(t_k) = z_k\}$ são denominadas probabilidades de transição e representam, portanto, a probabilidade de o estado $Z(t_{k+1})$ ser z_{k+1} no instante t_{k+1}, dado que o estado $Z(t_k)$ é z_k no instante t_k (HOWARD, 1960).

1.3 Cadeias de Markov

Definição: a cadeia de Markov é um processo markoviano quando as variáveis aleatórias Z(t) estão definidas em um espaço de estados discretos.

Quando o tempo é discreto, a cadeia de Markov é considerada uma cadeia de Markov em tempo discreto. Assim, tem-se que:

$$P\{Z(k+1) = z_{k+1} \mid Z(k) = z_k, Z(k-1) = z_{k-1}, \dots, Z(1) = z_1, Z(0)=z_0\}$$
$$= P\{Z(k+1) = z_{k+1} \mid Z(k) = z_k\} \text{ para } \forall \text{ sequência } 0, 1, \dots, k-1, k, k+1. \qquad (2)$$

As probabilidades de transição $P\{Z(k+1) = z_{k+1}|Z(k) = z_k\}$ representam, então, a probabilidade de o estado Z(k+1) ser z_{k+1} no tempo k+1, dado que o estado Z(k) é z_k no tempo k.

Em um processo markoviano com *n* estados (resultados) exaustivos e mutuamente exclusivos, as probabilidades em um ponto específico do tempo t = 0, 1, 2, ... são habitualmente expressas por:

$$P_{ij} = P[Z_t = j|Z_{t-1} = i], (i,j) = 1, 2, \dots, n; t= 0, 1, 2, \dots, T \qquad (3)$$
$$\text{em que } \sum_j p_{ij} = 1, i = 1, 2, \dots, n \text{ e } p_{ij} \geq 0, (i,j) = 1, 2, \dots, n$$

Um modo conveniente de resumir as probabilidades de transição em uma etapa é usar a seguinte notação matricial (MOREIRA, 2010; WEBER, 1986):

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

Essa matriz é denominada matriz de transição. Cada linha da matriz de transição é chamada de vetor de probabilidade, sendo o i-ésimo vetor de probabilidade (i-ésima linha) representado por

$V_i = [p_{i1} \ \ p_{i2} \ \dots \ p_{in}]$, em que $0 \leq p_{ij} \leq 1$ e $\sum_{j=1}^{n} p_{ij} = 1$, para i = 1, 2, ... , n.

A cadeia de Markov é um processo markoviano quando as variáveis aleatórias Z(t) estão definidas em um espaço de estados discretos.

Considere agora os vetores de probabilidades iniciais V_i^0 de iniciar no estado *i* e a matriz de transição P de uma cadeia de Markov, os vetores de probabilidades V_i^n de estar no estado *i* após n transições (n > 0) são calculados da seguinte maneira:

$$
\begin{aligned}
V_i^1 &= V_i^0 P \\
V_i^2 &= V_i^1 P = V_i^0 P^2 \\
V_i^3 &= V_i^2 P = V_i^0 P^3 \\
&\vdots \\
V_i^n &= V_i^{n-1} P = V_i^0 P^n
\end{aligned} \tag{4}
$$

A matriz P^n é conhecida como matriz de transição em n passos.

1.4 Classificação dos Estados em uma Cadeia de Markov

A classificação pode ser feita com base na probabilidade de transição pij da matriz de transição P (TAHA, 2008).

- Estado absorvente: um estado *i* é dito absorvente se retornar para ele mesmo, com certeza, em uma transição, ou seja, $p_{ii} = 1$.
- Estado transiente: um estado *i* é dito transiente se puder alcançar outro estado, mas não puder voltar ao mesmo estado em que estava com base em outro estado.
- Estado recorrente: um estado *i* é dito recorrente se a probabilidade de voltar ao estado em que estava com base em outros estados for 1. Isso ocorrerá se, e somente se, o estado não for transiente.
- Estado periódico: um estado i é dito periódico com período t > 0 se um retorno só for possível em t, 2t, 3t, ... , passos.

1.5 Cadeias Ergódicas de Markov

Se o sistema ou processo submetido ao modelo da cadeia de Markov tiver certas propriedades, é possível determinar as probabilidades de sucesso, depois de chegar às condições de regime estacionário. No regime estacionário, as probabilidades de transição tornam-se constantes ao longo do tempo (MARQUES, 2010).

Para garantir que se chegou às condições de regime estacionário, a cadeia de Markov deve ser ergódica (às vezes chamada de irredutível).

Uma cadeia de Markov é ergódica quando descreve matematicamente um processo no qual é possível ir de um estado a qualquer outro estado, não sendo necessário que isso seja feito em apenas um passo.

1.6 Cadeias de Markov Regulares

A cadeia regular é um caso mais restrito de uma cadeia ergódica. A cadeia regular é definida como uma cadeia que tem uma matriz de transição P, que para determinada potência de P tem apenas elementos positivos (MARQUES, 2010).

1.7 Regime Estacionário

Nas condições de regime estacionário, as probabilidades se tornam constantes ao longo do tempo. Para garantir que se chegou às condições de regime estacionário, a cadeia deve ser ergódica.

A existência de condições de regime estacionário em uma cadeia ergódica regular pode ser demonstrada facilmente calculando-se P^n para vários valores de *n*. Note-se que, quando *n* aumenta, os valores de p_{ij} tendem a um limite fixo e cada vetor de probabilidade V_i^n tende a tornar-se igual para todos os valores de *i* (WEBER, 1986).

Assim, prova-se que:

- Para um valor suficientemente grande de n, o vetor probabilidade Vin torna-se igual para todo i e não se altera para valores maiores que n.
- Como Vin+1 = Vin.P e Vin+1 = Vin, então existe um vetor V* tal que V* = V*.P, em que o vetor V* contém as probabilidades existentes nas condições de regime estacionário.

Como V^* é um vetor de probabilidade, a seguinte condição deve ser satisfeita:

$$\sum_{j=1}^{m} v_j = 1.$$

1.8 Análise das Cadeias Absorventes de Markov

Por meio das cadeias absorventes, é possível determinar:

— o número esperado de passos antes de o processo ser absorvido;

— o número esperado de vezes que o processo se encontra em qualquer estado não absorvente;

— a probabilidade de absorção por qualquer estado absorvente.

Para que essas determinações sejam possíveis, deve-se inicialmente reorganizar a matriz de transição em quatro submatrizes, como segue:

$$P = \begin{bmatrix} I & \vdots & O \\ \cdots & \vdots & \cdots \\ A & \vdots & N \end{bmatrix}$$

Supondo que o número de estados absorventes seja a, o número de estados não absorventes seja *n* e *a + n = m* o número de estados totais, então:

— I = matriz identidade a x a, representando as probabilidades de permanecer dentro de um estado absorvente.

— O = matriz nula a x n, que reflete as probabilidades de ir de um estado absorvente para um estado não absorvente.

— A = matriz n x a, contendo as probabilidades de ir de qualquer estado não absorvente para um estado absorvente.

— N = matriz n x n, contendo as probabilidades de ir de um estado não absorvente para qualquer outro estado não absorvente.

Como consequência das submatrizes vistas, podemos inferir que:

— N^n fornece as probabilidades de ir de um estado não absorvente para outro estado não absorvente em n passos.

— A soma de cada linha da matriz $(I-N)^{-1}$ fornece o número de vezes que um processo está em cada estado não absorvente antes da absorção.

— $(I-N)^{-1}$ fornece as probabilidades de absorção de um estado absorvente.

2 Metodologia

O trabalho foi desenvolvido com aplicação da cadeia de Markov finita nas três primeiras séries do Ensino Médio, envolvendo os três estados da Região Sul (Paraná, Santa Catarina e Rio Grande do Sul) e considerando-se como sujeitos da pesquisa o ensino público, o privado e o total (público mais privado).

A matriz de transição para o curso médio, três anos de duração, pode ser escrita da seguinte forma:

$$P_j = \begin{bmatrix} p_{11} & p_{12} & 0 & 0 & p_{1a} \\ 0 & p_{22} & p_{23} & 0 & p_{2a} \\ 0 & 0 & p_{33} & p_{3t} & p_{3a} \\ 0 & 0 & 0 & p_{tt} & 0 \\ 0 & 0 & 0 & 0 & p_{aa} \end{bmatrix} \qquad (5)$$

em que j = 1, 2, 3 (ensino público, ensino privado e total).

As probabilidades da matriz de transição são dadas por:

p_{ii} = probabilidade de permanecer na mesma série (i = 1, 2, 3).

p_{ij} = probabilidade de passar de uma série *i* para a série seguinte j (i = 1, 2 e j = i+1).

p_{3t} = probabilidade de término.

p_{ia} = probabilidade de abandono da série *i* (i = 1, 2, 3).

$p_{tt} = p_{aa} = 1$ (estados absorventes).

Com base nas matrizes de transição, determinaram-se:

(1) Os tempos médios esperados de permanência em cada série, por meio dos elementos diagonais da matriz $M_j = (I-N)^{-1}$ = 1, 2, 3.

(2) Os tempos médios requeridos para o estado (série) ser absorvido, por meio da soma de cada linha da matriz M_j, j = 1, 2, 3.

(3) As probabilidades dos estados estacionários (término e abandono), por meio da matriz

$E_j = (I-N)^{-1}.A$, j = 1, 2, 3.

As TAB. 1, 2 e 3 mostram os indicadores (aprovação, reprovação e abandonos) para as três primeiras séries do Ensino Médio para cada estado da Região Sul, abrangendo os anos de 2009 (1ª série), 2010 (2ª série) e 2011 (3ª série).

TABELA 1 – Indicadores educacionais (%) do estado do Paraná – 2009 a 2011

ENSINO	APROVADOS			REPROVADOS			ABANDONOS		
	Séries			Séries			Séries		
	1ª	2ª	3ª	1ª	2ª	3ª	1ª	2ª	3ª
Público	72,2	81,0	86,7	16,8	11,7	8,0	11,0	7,3	5,3
Privado	92,4	96,4	98,5	7,0	3,4	1,1	0,6	0,2	0,4
Total	74,2	82,8	88,3	15,9	10,7	7,1	9,9	6,5	4,6

FONTE: INEP (2011)

TABELA 2 – Indicadores educacionais (%) do estado de Santa Catarina – 2009 a 2011

ENSINO	APROVADOS			REPROVADOS			ABANDONOS		
	Séries			Séries			Séries		
	1ª	2ª	3ª	1ª	2ª	3ª	1ª	2ª	3ª
Público	75,1	83,1	90,3	15,2	10,3	3,2	9,7	6,6	6,5
Privado	93,0	96,6	97,9	6,7	3,2	1,9	0,3	0,2	0,2
Total	77,1	84,9	91,4	14,2	9,4	3,0	8,7	5,7	5,6

FONTE: INEP (2011)

TABELA 3 – Indicadores educacionais (%) do estado do Rio Grande do Sul – 2009 a 2011

ENSINO	APROVADOS			REPROVADOS			ABANDONOS		
	Séries			Séries			Séries		
	1ª	2ª	3ª	1ª	2ª	3ª	1ª	2ª	3ª
Público	53,8	70,9	84,0	29,5	18,2	9,4	16,7	10,9	6,6
Privado	86,5	91,9	96,6	12,7	7,6	3,0	0,8	0,5	0,4
Total	56,6	73,4	85,8	28,0	16,9	8,5	15,4	9,7	5,7

FONTE: INEP (2011)

3 Resultados e Discussão

3.1 Matrizes de Transição

Dos indicadores fornecidos pelas TAB. 1, 2 e 3 resultaram as seguintes matrizes de transição:

— Para o estado do Paraná

(1) Ensino público

$$P_1 = \begin{bmatrix} 0{,}168 & 0{,}722 & 0 & 0 & 0{,}110 \\ 0 & 0{,}117 & 0{,}810 & 0 & 0{,}073 \\ 0 & 0 & 0{,}080 & 0{,}867 & 0{,}053 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Na matriz P_1, a probabilidade de um aluno do ensino público (Paraná) ser reprovado na 1ª série é de 0,168 (16,8%), de ser promovido para a 2ª série é de 0,722 (72,2%) e de abandonar a 1ª série é de 0,110 (11,0%). Para um aluno que se encontra na 2ª série, a probabilidade de reprovação é 0,117 (11,7%), de promoção para a 3ª série é 0,810 (81,0%) e de abandono é 0,073 (7,3%). No caso do aluno da 3ª série, a probabilidade de reprovação é 0,080 (8,0%), de término é 0,867 (86,7%) e de abandono é 0,053 (5,3%). Essa mesma análise pode ser feita para as demais matrizes de transição.

(2) Ensino privado

$$P_2 = \begin{bmatrix} 0{,}070 & 0{,}924 & 0 & 0 & 0{,}006 \\ 0 & 0{,}034 & 0{,}964 & 0 & 0{,}002 \\ 0 & 0 & 0{,}011 & 0{,}985 & 0{,}004 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(3) Total (público + privado)

$$P_3 = \begin{bmatrix} 0{,}159 & 0{,}742 & 0 & 0 & 0{,}099 \\ 0 & 0{,}107 & 0{,}828 & 0 & 0{,}065 \\ 0 & 0 & 0{,}071 & 0{,}883 & 0{,}046 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

— Para o estado de Santa Catarina

(1) Ensino público

$$P_1 = \begin{bmatrix} 0{,}152 & 0{,}751 & 0 & 0 & 0{,}097 \\ 0 & 0{,}103 & 0{,}831 & 0 & 0{,}066 \\ 0 & 0 & 0{,}032 & 0{,}903 & 0{,}065 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(2) Ensino privado

$$P_2 = \begin{bmatrix} 0{,}067 & 0{,}930 & 0 & 0 & 0{,}003 \\ 0 & 0{,}032 & 0{,}966 & 0 & 0{,}002 \\ 0 & 0 & 0{,}019 & 0{,}979 & 0{,}002 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(3) Total (público + privado)

$$P_3 = \begin{bmatrix} 0{,}142 & 0{,}771 & 0 & 0 & 0{,}087 \\ 0 & 0{,}094 & 0{,}849 & 0 & 0{,}057 \\ 0 & 0 & 0{,}030 & 0{,}914 & 0{,}056 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

— Para o estado do Rio Grande do Sul

(1) Ensino público

$$P_1 = \begin{bmatrix} 0{,}295 & 0{,}538 & 0 & 0 & 0{,}167 \\ 0 & 0{,}182 & 0{,}709 & 0 & 0{,}109 \\ 0 & 0 & 0{,}094 & 0{,}840 & 0{,}066 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(2) Ensino privado

$$P_2 = \begin{bmatrix} 0{,}127 & 0{,}865 & 0 & 0 & 0{,}008 \\ 0 & 0{,}076 & 0{,}919 & 0 & 0{,}005 \\ 0 & 0 & 0{,}030 & 0{,}966 & 0{,}004 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(3) Total (público + privado)

$$P_3 = \begin{bmatrix} 0{,}280 & 0{,}566 & 0 & 0 & 0{,}154 \\ 0 & 0{,}169 & 0{,}734 & 0 & 0{,}097 \\ 0 & 0 & 0{,}085 & 0{,}858 & 0{,}057 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

De maneira geral, observando todas as matrizes de transição, verifica-se que:

— O índice de reprovação sempre é maior na 1ª série, reduzindo com o avanço da série. Isso ocorre para os três estados da Região Sul tanto para o ensino público quanto para o privado.

— O índice de abandono também segue a mesma tendência da reprovação, ou seja, é maior na 1ª série, reduzindo com o avanço dela. Essa tendência ocorre em todos os estados tanto para o ensino público quanto para o ensino privado.

— O Rio Grande do Sul apresenta os maiores índices de reprovação tanto no ensino público quanto no ensino privado. Apresenta também os maiores índices de abandono em todas as séries.

3.2 Matrizes de Tempo Médio de Permanência no Sistema

— Para o estado do Paraná

(1) Ensino público

$$M_1 = \begin{bmatrix} 1{,}2019 & 0{,}9828 & 0{,}8653 \\ 0 & 1{,}1325 & 0{,}9971 \\ 0 & 0 & 1{,}0870 \end{bmatrix}$$

Portanto, o tempo médio de permanência dos alunos na 1ª série é 1,2019 ano; na 2ª série esse tempo é de 1,1325 ano; e na 3ª série é de 1,0870 ano.

(2) Ensino privado

$$M_2 = \begin{bmatrix} 1{,}0753 & 1{,}0285 & 1{,}0025 \\ 0 & 1{,}0352 & 1{,}0090 \\ 0 & 0 & 1{,}0111 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,0753 ano; na 2ª série esse tempo é de 1,0352 ano; e na 3ª série é de 1,0111 ano.

(3) Ensinos público e privado

$$M_3 = \begin{bmatrix} 1{,}1891 & 0{,}9880 & 0{,}8806 \\ 0 & 1{,}1198 & 0{,}9981 \\ 0 & 0 & 1{,}0764 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,1655 ano; na 2ª série esse tempo é de 1,1198 ano; e na 3ª série é de 1,0764 ano.

— Para o estado de Santa Catarina

(1) Ensino público

$$M_1 = \begin{bmatrix} 1{,}1792 & 0{,}9873 & 0{,}8476 \\ 0 & 1{,}1148 & 0{,}9570 \\ 0 & 0 & 1{,}0331 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,1792 ano; na 2ª série esse tempo é de 1,1148 ano; e na 3ª série é de 1,0331 ano.

(2) Ensino privado

$$M_2 = \begin{bmatrix} 1{,}0718 & 1{,}0297 & 1{,}0140 \\ 0 & 1{,}0331 & 1{,}0173 \\ 0 & 0 & 1{,}0194 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,0718 ano; na 2ª série esse tempo é de 1,0331 ano; e na 3ª série é de 1,0194 ano.

(3) Ensinos público e privado

$$M_3 = \begin{bmatrix} 1{,}1655 & 0{,}9918 & 0{,}8681 \\ 0 & 1{,}1038 & 0{,}9661 \\ 0 & 0 & 1{,}0309 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,1655 ano; na 2ª série esse tempo é de 1,1038 ano; e na 3ª série é de 1,0309 ano.

— Para o estado do Rio Grande do Sul

(1) Ensino público

$$M_1 = \begin{bmatrix} 1{,}4184 & 0{,}9329 & 0{,}7301 \\ 0 & 1{,}2225 & 0{,}9567 \\ 0 & 0 & 1{,}1038 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,4184 ano; na 2ª série esse tempo é de 1,2225 ano; e na 3ª série é de 1,1038 ano.

(2) Ensino privado

$$M_2 = \begin{bmatrix} 1,1455 & 1,0723 & 1,0160 \\ 0 & 1,0823 & 1,0253 \\ 0 & 0 & 1,0309 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,1455 ano; na 2ª série esse tempo é de 1,0823 ano; e na 3ª série é de 1,0309 ano.

(3) Ensinos público e privado

$$M_3 = \begin{bmatrix} 1,3889 & 0,9460 & 0,7589 \\ 0 & 1,2034 & 0,9653 \\ 0 & 0 & 1,0929 \end{bmatrix}$$

O tempo médio de permanência dos alunos na 1ª série é 1,3889 ano; na 2ª série esse tempo é de 1,2034 ano; e na 3ª série é de 1,0929 ano.

Os resultados dos tempos médios de permanência em cada série foram resumidos na TAB. 4.

TABELA 4 – Tempo médio de permanência dos alunos em cada série de acordo com os estados da Região Sul do Brasil

SÉRIE	Ensino público			Ensino privado			Ensinos público e privado		
	PR	SC	RS	PR	SC	RS	PR	SC	RS
1ª	1,20	1,18	1,42	1,08	1,07	1,15	1,19	1,17	1,39
2ª	1,13	1,11	1,22	1,04	1,03	1,08	1,12	1,10	1,20
3ª	1,09	1,03	1,10	1,01	1,02	1,03	1,08	1,03	1,09

FONTE: Os autores (2012)

A análise da TAB. 4 mostra que, nas três situações (ensino público, ensino privado e público mais privado), o tempo médio esperado de permanência reduz da 1ª para a 3ª série, e isso ocorre para os três estados da Região Sul. Essa ocorrência está relacionada aos maiores índices de reprovação nas séries iniciais. De uma maneira geral, o Rio Grande do Sul apresenta os maiores tempos médios esperados de permanência em cada uma das três séries, enquanto Santa Catarina apresenta os menores tempos.

3.3 Tempo Médio Requerido para o Estado ser Absorvido

— Para o estado do Paraná

Ensino público (T_1), ensino privado (T_2) e ensinos público e privado (T_3).

$$T_1 = \begin{bmatrix} 3,0500 \\ 2,1296 \\ 1,0870 \end{bmatrix} \quad T_2 = \begin{bmatrix} 3,1063 \\ 2,0442 \\ 1,0111 \end{bmatrix} \quad T_3 = \begin{bmatrix} 3,0576 \\ 2,1179 \\ 1,0764 \end{bmatrix}$$

— Para o estado de Santa Catarina

Ensino público (T_1), ensino privado (T_2) e ensinos público e privado (T_3).

$$T_1 = \begin{bmatrix} 3,0141 \\ 2,0718 \\ 1,0331 \end{bmatrix} \quad T_2 = \begin{bmatrix} 3,1155 \\ 2,0504 \\ 1,0194 \end{bmatrix} \quad T_3 = \begin{bmatrix} 3,0254 \\ 2,0699 \\ 1,0309 \end{bmatrix}$$

— Para o estado do Rio Grande do Sul

Ensino público (T_1), ensino privado (T_2) e ensinos público e privado (T_3).

$$T_1 = \begin{bmatrix} 3,0814 \\ 2,1792 \\ 1,1038 \end{bmatrix} \quad T_2 = \begin{bmatrix} 3,2338 \\ 2,1076 \\ 1,0309 \end{bmatrix} \quad T_3 = \begin{bmatrix} 3,0938 \\ 2,1687 \\ 1,0929 \end{bmatrix}$$

As matrizes T_j (j = 1, 2, 3), soma dos elementos de cada linha das matrizes M_j (j = 1, 2, 3), respectivamente, representam os tempos médios esperados (em anos) para um aluno sair do sistema, dado que iniciou no estado (série) i.

TABELA 5 – Tempos médios esperados (em anos) para um aluno sair do sistema, dado que iniciou na série i, para os estados da Região Sul do Brasil

SÉRIE	Ensino público			Ensino privado			Ensinos público e privado		
	PR	SC	RS	PR	SC	RS	PR	SC	RS
1ª	3,05	3,01	3,08	3,11	3,12	3,23	3,06	3,03	3,09
2ª	2,13	2,07	2,18	2,04	2,05	2,11	2,12	2,07	2,17
3ª	1,09	1,03	1,10	1,01	1,02	1,03	1,08	1,03	1,09

FONTE: Os autores (2012)

Verifica-se, na TAB. 5, que os tempos médios esperados para sair do sistema, dado que se encontra na 1ª série, sempre é maior para o estado do Rio Grande do Sul, enquanto Santa Catarina apresenta menor tempo para o ensino público e no total e o Paraná, para o ensino privado. Partindo da 2ª série, novamente o tempo médio esperado para sair do sistema é sempre maior para o estado do Rio Grande do Sul, sendo menor para Santa Catarina no ensino público e total, enquanto para o Paraná esse tempo é menor para o ensino privado. Na 3ª série, o tempo maior para sair do sistema novamente ocorre no Rio Grande do Sul, nos três casos; esse tempo é menor para Santa Catarina no ensino público e no total; e para o Paraná esse tempo é menor para o ensino privado.

3.4 Probabilidades dos Estados Estacionários

Em seguida, são apresentadas as probabilidades dos estados estacionários para os três estados, levando-se em conta o ensino público, o ensino privado e os ensinos público e privado.

— Para o estado do Paraná

Ensino público (E_1), ensino privado (E_2) e ensinos público e privado (E_3).

$$E_1 = \begin{bmatrix} 0,7502 & 0,2498 \\ 0,8645 & 0,1355 \\ 0,9224 & 0,0576 \end{bmatrix} \quad E_2 = \begin{bmatrix} 0,9875 & 0,0125 \\ 0,9939 & 0,0061 \\ 0,9960 & 0,0040 \end{bmatrix} \quad E_3 = \begin{bmatrix} 0,7766 & 0,2224 \\ 0,8813 & 0,1187 \\ 0,9505 & 0,0495 \end{bmatrix}$$

— Para o estado de Santa Catarina

Ensino público (E_1), ensino privado (E_2) e ensinos público e privado (E_3).

$$E_1 = \begin{bmatrix} 0,7654 & 0,2346 \\ 0,8642 & 0,1358 \\ 0,9329 & 0,0671 \end{bmatrix} \quad E_2 = \begin{bmatrix} 0,9927 & 0,0073 \\ 0,9959 & 0,0041 \\ 0,9980 & 0,0020 \end{bmatrix} \quad E_3 = \begin{bmatrix} 0,7935 & 0,2065 \\ 0,8830 & 0,1170 \\ 0,9423 & 0,0577 \end{bmatrix}$$

— Para o estado do Rio Grande do Sul

Ensino público (E_1), ensino privado (E_2) e ensinos público e privado (E_3).

$$E_1 = \begin{bmatrix} 0,6132 & 0,3868 \\ 0,8036 & 0,1964 \\ 0,9272 & 0,0728 \end{bmatrix} \quad E_2 = \begin{bmatrix} 0,9814 & 0,0186 \\ 0,9905 & 0,0095 \\ 0,9959 & 0,0041 \end{bmatrix} \quad E_3 = \begin{bmatrix} 0,6511 & 0,3489 \\ 0,8282 & 0,1718 \\ 0,9377 & 0,0623 \end{bmatrix}$$

TABELA 6 – Probabilidade (em %) de término a partir da i-ésima série, para os estados da Região Sul do Brasil

SÉRIE	Ensino público			Ensino privado			Ensinos público e privado		
	PR	SC	RS	PR	SC	RS	PR	SC	RS
1ª	75,02	76,54	61,32	98,75	99,27	98,14	77,66	79,35	65,11
2ª	86,45	86,42	80,36	99,39	99,59	99,05	88,13	88,30	82,82
3ª	92,24	93,29	92,72	99,60	99,80	99,59	95,05	94,23	93,77

FONTE: Os autores (2012)

Na TAB. 6, percebe-se que, em todos os casos, as probabilidades de término do curso aumentam com o avanço da série em que o aluno se encontra. Com relação ao ensino público, as maiores probabilidades de término são para o estado de Santa Catarina, com os alunos que se encontram na 1ª e na 3ª séries, e para o Paraná, com os alunos da 2ª série. No ensino privado, as maiores probabilidades de término, para as três séries, são todas de Santa Catarina. No geral, as maiores probabilidades de término são para Santa Catarina (1ª e 2ª séries) e Paraná (3ª série). Note-se que, praticamente em todos os casos, o estado do Rio Grande do Sul apresenta as menores probabilidades de término.

TABELA 7 – Probabilidade (em %) de abandono a partir da i-ésima série, para os estados da Região Sul do Brasil

SÉRIE	Ensino público			Ensino privado			Ensinos público e privado		
	PR	SC	RS	PR	SC	RS	PR	SC	RS
1ª	24,98	23,46	38,68	1,25	0,73	1,86	21,80	22,24	34,89
2ª	13,55	13,58	19,64	0,61	0,41	0,95	11,87	11,70	17,18
3ª	5,76	6,71	7,28	0,40	0,20	0,41	4,95	5,77	6,23

FONTE: Os autores (2012)

De acordo com a TAB. 7, verifica-se que as probabilidades de abandono, em todos os casos, decrescem com o avanço das séries. Em todos os casos, as maiores probabilidades de abandono são do estado do Rio Grande do Sul. No ensino público, Santa Catarina tem a menor probabilidade de abandono na 1ª série e o Paraná, na 2ª e na 3ª séries. No ensino privado, as menores probabilidades de abandono ocorrem em Santa Catarina, para as três séries, e no geral (ensinos público e privado) as menores probabilidades são de Santa Catarina, nas duas séries iniciais, e Paraná, na 3ª série.

Conclusão

Por meio deste estudo, apresentaram-se as possibilidades de análises e previsões de indicadores educacionais utilizando o método conhecido como cadeia de Markov.

O desenvolvimento teórico do método e as aplicações envolvendo os ensinos públicos e privados na Região Sul do Brasil possibilitaram conhecer a realidade dos indicadores educacionais propostos, levando em conta os dados obtidos no portal do Inep para os anos de 2009 (1ª série), 2010 (2ª série) e 2011 (3ª série).

O método proposto proporcionou as obtenções dos tempos médios esperados de permanências em cada série, os tempos médios esperados de permanências no sistema, os tempos médios requeridos para sair do sistema e as probabilidades de absorção dos estados (término e abandono).

As aplicações dos métodos para as três séries do Ensino Médio da Região Sul (Paraná, Santa Catarina e Rio Grande do Sul) mostraram que, para os ensinos público e privado, os tempos médios esperados de permanências em cada série são reduzidos de acordo com o avanço desta, e isso ocorre para os três estados. O Rio Grande do Sul apresenta o maior tempo médio de permanência em cada série, enquanto Santa Catarina apresenta o menor tempo médio.

Com relação aos tempos médios esperados para sair do sistema, constatou-se que o Rio Grande do Sul apresenta maior tempo em todos os casos; Santa Catarina apresenta menor tempo para o ensino público em relação às três séries; e o Paraná tem o menor tempo para o ensino privado também em relação às três séries.

As análises das probabilidades de término do curso mostraram que essas probabilidades aumentam com os avanços das séries e que, de uma maneira quase geral, o Rio Grande do Sul apresenta as menores probabilidades de término, enquanto Paraná e Santa Catarina apresentam probabilidades muito próximas.

Com relação às probabilidades de abandono, constatou-se que ela decresce com o avanço da série e que o estado do Rio Grande do Sul tem as maiores probabilidades independentemente da série. Santa Catarina apresentou as menores probabilidades de abandono em todas as séries do ensino privado e na 1ª e na 2ª série do ensino público; o Paraná apresentou as menores probabilidades de abandono na 3ª série do ensino público.

- **Recebido em: 03/09/2012**
- **Aprovado em: 19/02/2013**

Referências

DERMAN, C. **Finite state Markovian decisions processes**. New York: Academic Press, 1970.

HOWARD. R. **Dynamic programming and Markov processes**. Cambridge: Technology Press of Massachusetts Institute, 1960.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Disponível em: <http://portal.inep.gov.br/>. Acesso em: 26 jun. 2013.

MARQUES, J. M. **Pesquisa operacional II**. Curitiba, 2010. Notas de Aula.

MOREIRA, DANIEL A. **Pesquisa operacional**: curso introdutório. 2. ed. São Paulo: Cengage Learning, 2010.

MORETTIN, Pedro A.; TOLOI, Clélia Maria C. **Previsão de séries temporais**. 2. ed. São Paulo: Atual, 1987.

NORRIS, James N. **Markov chains**. New York: Cambridge University Press, 1998.

ROSS, Sheldon M. **Stochastic processes**. New York: Wiley, 1996.

TAHA, Hamdy A. **Pesquisa operacional**. 8. ed. São Paulo: Pearson Prentice Hall, 2008.

WEBER, Jean E. **Matemática para economia e administração**. 2. ed. São Paulo: Harbra, 1986.