

# Cadeia de Markov aplicada à tendência para a formação acadêmica ou técnica, e comparação com a oferta de mercado

George Mamede

Vânia Santos

## Resumo

De acordo com os estudos realizados por alguns estudiosos e economistas, há uma preocupação em atender às expectativas de mercado, com relação às demandas de novos profissionais que fazem suas escolhas, baseados em diversas motivações. Este trabalho está fundamentado em um Processo Estocástico – a Cadeia de Markov – onde os dados utilizados foram coletados através de um procedimento de pesquisa, sobre algumas escolas das redes pública e privada da cidade de Salvador e região metropolitana, na qual questiona-se aos alunos do terceiro ano do ensino médio quais áreas do conhecimento, em curso técnico ou superior, os mesmos tinham interesse em cursar: Exatas, Biológicas ou Humanas. Para atender à dinâmica da Cadeia de Markov, o espaço amostral foi reorganizado em uma mecânica matricial de porcentagens de escolhas e operados por sistema linear de equações, para serem expressos em vetores de probabilidades. Ainda sobre os resultados de fidelização em permanência na área escolhida, foi também realizada uma comparação com dados de oferta de vagas de trabalho nas áreas citadas, entre os anos de 2017 a 2018. Os resultados foram distribuídos em Momento 1 ( $M_1$ ) e Momento 2 ( $M_2$ ), sobre os quais foram feitas a discussão e análise.

Palavras-chave: Processos Estocásticos. Cadeia de Markov. Sistemas linear de Equações.

## Abstract

According to studies carried out by some scholars and economists, there is a concern to meet market expectations, regarding the demands of new professionals who make their choices, based on different motivations. This work is based on a Stochastic Process – the Markov Chain – where the data used were collected through a research procedure, on some schools of the public and private networks of the city of Salvador and metropolitan region, in which the students are questioned of the third year of high school which areas of knowledge, undergoing technical or superior, they were interested in attending: Exact, Biological or Human. To meet the Markov Chain dynamics, the sample space was reorganized into a matrix mechanics of percentages of choices and operated by linear system of equations, to be expressed in probability vectors. Still on the results of loyalty in permanence in the chosen area, a comparison was also made with data on job vacancies in the mentioned areas, between the years 2017 to 2018. The results were distributed in moment 1 and moment 2, on the which the discussion and analysis were made.

**Keywords:** Stochastic Processes. Markov Chain. Linear Systems of Equations.

## 1. Introdução

No decorrer do período acadêmico, na Universidade Católica do Salvador, durante a formação de Licenciatura em Matemática Plena, foi despertado o quão interessante e abrasador são as disciplinas de álgebra Linear 1 e 2. As diversas aplicações em resolução de sistemas de equações lineares, estudos de probabilidade, operadores ortogonais, teorema espectral para operadores simétricos, entre outras. Também a disciplina Matemática para Educação Básica 3 teve grande importância, onde são vistos, inicialmente, análise combinatória, probabilidades, com suas deduções em um determinado espaço amostral, eventos simples, espaços equiprováveis, probabilidade condicional, e por fim, o estudo das matrizes. Dois conteúdos que serviram de base para a pesquisa realizada nesse trabalho. No entanto, por questão de conveniência, foi utilizado o método básico da substituição para a resolução de sistemas de equações, o que deve facilitar a leitura.

O desafio e objetivo principal desse trabalho foi demonstrar que por intermédio de um processo de Cadeia de Markov seria possível avaliar a tendência para formação ou demanda profissional de mercado, através do produto de uma pesquisa fomentada por questionamentos constantes dos estudantes do ensino fundamental II e médio, que ficavam exaltados depois de algumas elucidações de questões, sejam elas em relação ao nível de conhecimento ou até mesmo de revisão para os exames, dentre as quais cito aqui algumas: “Para que serve a matemática, estatística, partes da biologia, engenharias e afins (Exatas)?”; “Pretendo estudar direito, economia, administração (Humanas)?”. “Pretendo ser médico, enfermagem, professor de educação física (Biológicas)?”, entre outros questionamentos.

Essa situação problema já havia sido despertada em tempos pregressos, pois esse trabalho havia sido iniciado na disciplina Iniciação à Vida Universitária, em 2016, cujo tema era: “As ciências exatas e as escolhas dos alunos para o mercado de trabalho”. Ao ser retomada posteriormente, tal questão trouxe uma metodologia em grau superior e depois com o propósito probabilístico de avaliar como os estudantes do terceiro ano do ensino médio estão avançando em suas escolhas sejam elas nas áreas de Exatas, Biológicas e/ou Humanas.

Pelo recurso de Processo Estocástico utilizado, foi possível obter um resultado relevante, visto que o Processo Markoviano estuda a probabilidade de qualquer comportamento futuro do encadeamento, quando se conhece exatamente o seu estado atual. quando o seu estado atual é conhecido exatamente. Nesses casos, a escolha inicial dos alunos não é alterada pelo conhecimento complementar sobre seu procedimento passado.

Os momentos de escolha dos alunos foram apurados em dois períodos determinados como momento  $M_1$  e momento  $M_2$ , para uma facilitação de informação. São determinados nos cálculos dos sistemas de equações através do resultado do momento  $M_1$ , a fidelização das preferências dos alunos nas áreas questionadas. Com essa primeira solução é feita uma operação de Procedimento Markoviano; com isso, subsequentemente se encontram o momento  $M_2$ , onde encontra-se as possíveis tendências de escolhas dos estudantes do ensino médio. Esse processo poderia estender-se por  $n$ -ésimos momentos, conseqüentemente, entraríamos na Teoria Ergódica, que é um ramo da Matemática que estuda Sistemas Dinâmicos, sendo esse um dos objetivos para uma futura pesquisa.

## 2. Fundamentos teóricos

### 2.1. Processos Aleatórios: Cadeia de Markov

Muitos dos processos que ocorrem na natureza e na sociedade podem ser estudados (pelo menos em primeira aproximação) como se o fenômeno passasse, a partir de um estado inicial, por uma sequência de estados, onde a transição de um determinado estado para o seguinte ocorreria segundo uma certa probabilidade. Pois a probabilidade de transição depende apenas do estado a seguir, sendo assim, o momento anterior é esquecido. Esse Processo Estocástico será chamado processo de Markov, e a sequência de estados seguindo este processo será denominada uma cadeia de Markov. É óbvio que provavelmente em demasia, iremos simplificar ao máximo, uma vez que as probabilidades podem se modificar com o tempo.

Um **Processo Estocástico** é definido como uma coleção de variáveis randômicas (aleatórias)  $(X(t))$  indexadas por um parâmetro  $t$  pertencente a um conjunto  $T$ . Frequentemente  $T$  é tomado para ser o conjunto dos inteiros não negativos (porém, outros conjuntos são perfeitamente possíveis) e  $X(t)$  representa uma característica mensurável de interesse no tempo  $t$ . Considere um sistema que, em cada instante, está em apenas um entre um número finito de estados.

Exemplificando,  $X(t)$  pode representar o nível de estoque de um produto no fim da semana  $t$ ; o tempo em determinada região pode estar chuvoso ou seco, uma pessoa pode ser fumante ou não fumante; podemos frequentar ou não uma universidade; vivemos em uma área urbana ou rural; pertencemos à classe baixa, média ou alta; compramos um carro de marca C, F ou outra marca qualquer. Ao longo do tempo, o sistema pode mudar de um estado para o outro; supondo que o estado do sistema é observado em períodos fixos de tempo (por exemplo, uma vez por dia, em cada hora, e assim por diante). Em muitas aplicações conhecemos o estado atual do sistema e queremos prever o estado no próximo período de observação ou em algum período futuro. Pode-se prever, muitas vezes, a probabilidade de o sistema estar em determinado estado em um período futuro de observação a partir de sua história pregressa.

Os Processos Estocásticos são ferramentas matemáticas que podem ser úteis para descrever o procedimento de um sistema operando sobre algum período de tempo. Com isso, em termos formais, a variável randômica  $X(t)$  representa o estado do sistema no parâmetro (geralmente tempo)  $t$ . Portanto pode-se afirmar que  $X(t)$  é definido em um espaço denominado Espaço de Estados.

Os processos estocásticos podem ser classificados como:

#### 1. Em relação ao estado

- Estado Discreto (cadeia):  $X(t)$  é definido sobre um conjunto enumerável ou finito.
- Estado Contínuo (sequência):  $X(t)$  caso contrário.

#### 2. Em relação ao tempo (Parâmetro)

- Tempo Discreto:  $t$  é finito ou enumerável.
- Tempo Contínuo:  $t$  caso contrário.

Exemplos:

1. Número de usuários em uma fila de banco em um determinado instante: Estado Discreto e Tempo Contínuo.
2. índice pluviométrico diário: Estado Contínuo e Tempo Discreto.
3. Número de dias chuvosos: Estado Discreto e Tempo Discreto.

Nossa intenção, nesta publicação, é trabalharmos nos Estados e Tempos Discretos.

Existem vários “tipos” de Processos Estocásticos, porém, neste trabalho apenas abordaremos um tipo de Processo Estocástico denominado **Processo Markoviano**.

## 2.2. Processos Markovianos

Uma **cadeia de Markov** ou **processo de Markov** é um **Processo Estocástico** no qual a probabilidade de um sistema estar em determinado estado em um dado período de observação depende apenas do estado no período de observação imediatamente anterior. Um Processo Estocástico é dito ser um Processo Markoviano se:

$$P\{X(t_{k+1}) \leq x_{k+1} | X(t_k) = x_k, X(t_{k-1}) = x_{k-1}, \dots, X(t_1) = x_1, X(t_0) = x_0\} = P\{X(t_{k+1}) \leq x_{k+1} | X(t_k) = x_k\} \quad (1)$$

para  $t_0 \leq t_1 \leq t_k \leq t_{k+1} = 0, 1, \dots$  e toda sequência  $k_0, k_1, \dots, k_{t-1}, k_t, k_{t+1}$ .

A expressão (1) pode ser “traduzida”por: a probabilidade condicional de qualquer evento futuro, dado qualquer evento passado e o estado presente  $X(t_k) = x_k$  é independente do evento passado e depende somente do estado presente. Em termos mais simples: um Processo Estocástico é dito ser um Processo Markoviano se o estado futuro depende apenas do estado presente e não dos estados passados. Esse tipo de Processo Estocástico é também denominado de “memoryless process”(processo sem memória), uma vez que o passado é “esquecido”ou desprezado.

As probabilidades condicionais  $P\{X(t_{k+1}) \leq x_{k+1} | X(t_k) = x_k\}$  são denominadas Probabilidades de Transição e representam, portanto, a probabilidade de o estado  $X(t_{k+1})$  ser  $x_{k+1}$  no instante  $t_{k+1}$  dado que o estado  $X(t_k)$  é  $x_k$  no instante  $t_k$ .

O processo aleatório de Markov pode assumir os estados  $a_1, a_2, \dots, a_r$  de tal modo que a probabilidade de transição de um estado  $a_j$  para um estado  $a_i$  seja  $P_{ij}$  (um número que só depende de  $j$  e  $i$ ).

A matriz das probabilidades de transição (matriz estocástica) é dada por:

$$T = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1r} \\ P_{21} & P_{22} & \dots & P_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ P_{r1} & P_{r2} & \dots & P_{rr} \end{bmatrix}$$

(Observe que  $P_{ij} \geq 0$  e que a soma de cada coluna deve ser 1.) O vetor de probabilidades é aquele cuja  $i$ -ésima linha dá a probabilidade de ocorrência do estado  $a_i$  após  $n$  iterações:

$$\begin{bmatrix} P_1^n \\ \vdots \\ P_r^n \end{bmatrix}$$

Seguindo o raciocínio do exemplo anterior vemos que, após  $n$  passos,

$$\begin{bmatrix} P_1^n \\ \vdots \\ P_r^n \end{bmatrix} = T^{(n)} \cdot \begin{bmatrix} P_1^1 \\ \vdots \\ P_r^1 \end{bmatrix}$$

### 2.3. Previsões em longo prazo

Para podermos fazer previsões a longo prazo, a matriz  $T$  deve cumprir certas condições. Desse modo, introduzimos a definição a seguir.

*Definição 1.* Uma matriz das probabilidades de transição é regular se alguma de suas potências tem todos os elementos não nulos.

A importância da matriz regular para as previsões em longo prazo é dada pelo teorema abaixo:

**Teorema 2.2.2** Se a matriz  $T_{r \times r}$  das probabilidades de transição é regular, então:

- i) As potências  $T^{(n)}$  aproximam-se de uma matriz  $P$ , no sentido de que cada elemento de  $T^{(n)}$  aproxima-se do elemento correspondente  $P$ .
- ii) Todas as colunas de  $P$  são iguais, sendo dadas por um vetor-coluna

$$V = \begin{bmatrix} P_1^1 \\ \vdots \\ P_r^1 \end{bmatrix}$$

com  $P_1 > 0, P_2 > 0, \dots, P_r > 0$ .

- iii) Para qualquer vetor de probabilidades inicial

$$V_1 = \begin{bmatrix} p_1^1 \\ \vdots \\ p_r^1 \end{bmatrix}$$

o vetor de probabilidades  $T^{(n)} \cdot V_1$  aproxima-se de  $V$  (dado no item anterior).

- iv) O vetor  $V$  é o único vetor que satisfaz  $V = T \cdot V$ .

O que esse teorema nos informa é que se a matriz das probabilidades de transição é regular, então é possível fazer previsões em longo prazo, e essa não depende das probabilidades iniciais  $V_1$ . Além disso, o item iv) indicará como achar as probabilidades depois de um longo prazo. O processo

utilizado para se encontrar o vetor “final” de probabilidades, usando o item iv), corresponde à procura de autovetor associado ao autovalor um da matriz  $T$ , que por definição seria  $T : V \rightarrow V$  um operador linear, onde se existirem  $v \in V, v \neq 0$  e  $\lambda \in \mathfrak{R}$ , tais que  $T(v) = \lambda.v$ ,  $\lambda$  é um autovalor de  $T$  e  $v$  um autovetor de  $T$  associado a  $\lambda$  [2].

EXEMPLO: Em uma determinada região, observa-se que se chover bastante durante um ano, a probabilidade de que chova bastante no ano seguinte é  $\frac{1}{4}$ , e que a probabilidade de que faça seca é de  $\frac{3}{4}$ . Ainda, se houver seca em um ano, no ano seguinte a probabilidade de haver seca ou chuva suficiente será a mesma, e igual a  $\frac{1}{2}$ . Suponhamos também, para simplificar (o que não ocorre na prática, embora possamos usar como recurso para ter um indicador da situação), que tais probabilidades não mudem no decorrer do tempo. Os estados possíveis para esse processo são: chuva e seca. Podemos ter, então, as seguintes seqüências de acontecimentos (árvore de probabilidades). Assim, sabendo que no primeiro ano houve seca, a probabilidade de que chova bastante no terceiro ano é  $\frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{8}$ . Conforme o número de anos aumenta, as contas tornam-se mais complicadas e, se estivermos interessados em previsões em longo prazo sobre o clima da região, temos que procurar um outro procedimento. Isso pode ser feito se introduzirmos a noção de matriz das probabilidades de transição e a de vetor de probabilidades onde o elemento na  $i$ -ésima linha e  $j$ -ésima coluna indica a probabilidade de transição do  $j$ -ésimo para o  $i$ -ésimo estado.

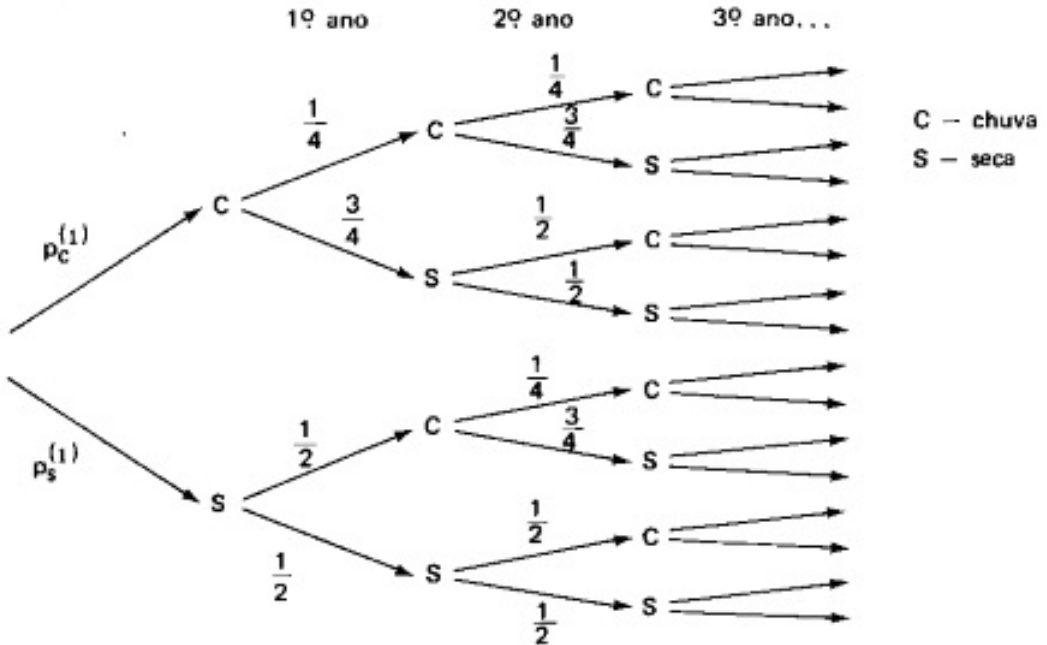


Figura 1: Probabilidade de Chuva e Seca (Fonte: [1])

$$\begin{array}{cc}
 C & S \\
 C & \frac{1}{4} \quad \frac{1}{2} \\
 S & \frac{3}{4} \quad \frac{1}{2}
 \end{array}
 \quad T = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix}$$

Onde o vetor de probabilidades é a matriz

$$\begin{bmatrix} p_c^n \\ \vdots \\ p_s^n \end{bmatrix}$$

Cuja primeira linha dá a probabilidade de que haja chuva no n-ésimo ano e a segunda linha dá a probabilidade de que haja seca no n-ésimo ano.

Analisando a árvore de probabilidades obtém-se

$$\begin{aligned}
 P_C^{(2)} &= \frac{1}{4}P_C^{(1)}\frac{1}{2}P_S^{(1)} \\
 P_S^{(2)} &= \frac{3}{4}P_C^{(1)}\frac{1}{2}P_S^{(1)}
 \end{aligned}$$

Observa-se que

$$T \cdot \begin{bmatrix} P_c^1 \\ P_s^1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} P_c^1 \\ P_s^1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4}P_c^{(1)} & \frac{1}{2}P_s^{(1)} \\ \frac{3}{4}P_c^{(1)} & \frac{1}{2}P_s^{(1)} \end{bmatrix}$$

e portanto

$$\begin{bmatrix} P_c^2 \\ P_s^2 \end{bmatrix} = T \cdot \begin{bmatrix} P_c^1 \\ P_s^1 \end{bmatrix}$$

O mesmo ocorre do segundo para o terceiro ano, deste para o quarto etc. Temos então a seqüência:

$$\begin{array}{ccccccc}
 1^\circ \text{ ano} & & 2^\circ \text{ ano} & & 3^\circ \text{ ano} & & \\
 \left[ \begin{array}{c} P_c^1 \\ P_s^1 \end{array} \right] & \rightarrow & \left[ \begin{array}{c} P_c^2 \\ P_s^2 \end{array} \right] & = T \cdot & \left[ \begin{array}{c} P_c^1 \\ P_s^1 \end{array} \right] & \rightarrow & \left[ \begin{array}{c} P_c^3 \\ P_s^3 \end{array} \right] = T \cdot \left[ \begin{array}{c} P_c^2 \\ P_s^2 \end{array} \right] = T^2 \cdot \left[ \begin{array}{c} P_c^1 \\ P_s^1 \end{array} \right] \rightarrow \dots \rightarrow
 \end{array}$$

n-ésimo ano

$$\begin{bmatrix} P_c^n \\ P_s^n \end{bmatrix} = T^{n-1} \cdot \begin{bmatrix} P_c^1 \\ P_s^1 \end{bmatrix}$$

Portanto, o comportamento do clima dessa região em longo prazo (i.e., quando n aumenta) poderá ser previsto se soubermos que os elementos das matrizes  $T^n$ ,  $n = 1, 2, \dots$  aproximam-se dos

elementos de uma matriz fixa  $P$  pois, neste caso,  $P_c^n \rightarrow P_1$  e  $P_s^n \rightarrow P_2$  quando  $n \rightarrow \infty$ .

Assim:

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = P \cdot \begin{bmatrix} P_c^1 \\ P_s^1 \end{bmatrix}$$

Tal previsão é importante, pois se chegarmos, por exemplo, à conclusão que  $P_s^n \rightarrow 1$  quando  $n \rightarrow \infty$  (Comportamento Limite de Vetores-estado: os vetores-estado convergem a um vetor fixo à medida que o número de observações cresce), assim, longo prazo, a região tenderá a se tornar um deserto.

Se  $T^n$  não se aproxima de uma matriz  $P$ , então não poderemos fazer nenhuma previsão em longo prazo, pois o processo acabará por se modificar bastante a cada passo. Assim, um dos problemas que devemos resolver é que são as condições sobre a matriz  $T$  das probabilidades de transição, para que suas potências se aproximem de uma determinada matriz.

Continuando a resolução sobre previsão do clima que estávamos estudando:

$$T = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix}$$

é regular, pois sua primeira potência, isto é, ela mesma, já tem todos os elementos estritamente positivos, e assim podemos concluir, usando o item (iv), que quaisquer que sejam as probabilidades iniciais, as probabilidades a longo prazo são dadas por:

$$\begin{bmatrix} P_c \\ P_s \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} P_c \\ P_s \end{bmatrix} \Rightarrow \begin{cases} P_c = \frac{1}{4} \cdot P_c + \frac{1}{2} \cdot P_s \\ P_s = \frac{3}{4} \cdot P_c + \frac{1}{2} \cdot P_s \end{cases} \Rightarrow \begin{cases} 2P_s = 3P_c \\ 2P_s = 3P_c \end{cases} \Rightarrow$$

$$P_s = \frac{3}{2}P_c$$

Como devemos ter  $P_c + P_s = 1$  (que é a probabilidade total), temos  $P_c + \frac{3}{2}P_c = 1$  ou  $P_c = \frac{2}{5}$  e, portanto,  $P_s = \frac{3}{5}$ . Assim, em longo prazo a probabilidade de um ano de chuva é  $\frac{2}{5}$ , enquanto que a probabilidade de um ano de seca é  $\frac{3}{5}$  (dentro das hipóteses simplificadoras), e portanto, a região tenderá a uma ligeira aridez.

### 3. Procedimentos metodológicos

Este trabalho baseia-se numa pesquisa de campo, com dados obtidos num estudo realizado em algumas escolas públicas e privadas (Escola Alfredo Agostinho de Deus, da rede pública estadual, Colégio Ramo da Videira, rede particular, Escola Francisco Pereira Franco, da rede pública estadual, e no Colégio Modelo Luís Eduardo Magalhães) localizadas na cidade de Salvador e região metropolitana. A pesquisa foi realizada presencialmente em quatro momentos diferentes. Em cada escola uma turma, no mês de maio de 2016, três escolas, e, em fevereiro de 2017, uma escola. Isso, em acordo com a coordenação de cada escola e com o(a) professor(a) plantonista que cedia



um horário para que pudéssemos explicar para os estudantes presentes o que estava envolvido em escolher uma formação técnica ou graduação.

Foram passadas, em aula anterior ao preenchimento do questionário, informações sobre as áreas a serem escolhidas; por exemplo, a área de Exatas envolvia profissões que trabalham com cálculos e números, sendo algumas citadas: computação, matemática, química, engenharia e afins. Na área de Humanas: administração, direito, economia, pedagogia, dentre outras, e por fim Biológicas: medicina, biologia, fisioterapia, educação física, farmácia e as demais. Foi um questionário com quatro perguntas objetivas, no qual os estudantes do terceiro ano do ensino médio assinalavam em quais áreas (Humanas, Biológicas ou Exatas, de nível técnico e/ou superior) os mesmos tinham interesse em adquirir formação de graduação. As perguntas foram:

1. Ao concluir o ensino médio, você pretende cursar, seja um curso de nível superior e/ou técnico, em que área?
2. O que te levou a escolher, em termos de estudos, essa área?
3. Como você acredita que está o mercado de trabalho para a área que você demonstra interesse?
4. Se fosse migrar para outra área, qual seria sua escolha?

Não havia questão discursiva, porém alguns expressaram, por escrito, o porquê de sua migração. Esses resultados foram aplicados para obtermos uma sequência de probabilidades coerentes para aplicação num processo de Cadeia de Markov, avaliando assim a tendência de escolha de cursos para formação acadêmica ou técnica, e comparando essa tendência com a oferta de mercado pelo Sistema Nacional de Emprego da Bahia (SineBahia), órgão público que disponibiliza uma equipe de profissionais da área de Recursos Humanos para atender o empregador parceiro, possibilitando encontrar mais rapidamente um profissional com o perfil desejado para preenchimento das vagas disponíveis.

Essa pesquisa foi realizada entre os períodos de 2016 e 2017, com início na disciplina *Iniciação à Vida Universitária*, com colegas de disciplinas variadas e retomada para a disciplina de TCC. Com a última pesquisa realizada, na escola pública no bairro da avenida San Martins, obteve-se um espaço amostral de 120 alunos. As respostas aos questionários, depois de colhidas, foram alocadas em suas respectivas áreas. As informações estão expressas sob a forma de gráfico de barras e o categórico da pesquisa: Matrizes, Sistemas de Equações, Vetores de Probabilidades (grafos) e Cálculos de Porcentagens, com pesquisas em livros de álgebra Linear 2. Os gráficos de barras foram gerados no Excel, o grafo e gráfico de autovalores foram gerados no Matlab.

A partir dessa etapa, buscou-se representar os modelos no Processo Estocástico, relacionado à Cadeia de Markov, por um conjunto de valores de probabilidades, calculados por um sistema de equações. Tais valores representam em percentual a fidelidade das escolhas dos estudantes do terceiro ano do ensino médio, nas áreas por eles almejadas. Com isso, estimam-se as probabilidades da Cadeia iniciando com os dados observados, gerando um sistema percentual recorrente para obter valores de probabilidades futuras, conforme procedimento característico da Cadeia de Markov. Os gráficos de probabilidades de transição (Figura 2) e de autovalores da matriz de probabilidades (Figura 3) foram gerados no Matlab, versão 2017b. Já os gráficos das Figuras 4 e 5 foram gerados no Excel.

#### 4. Resultados da tendência para a formação acadêmica ou técnica obtidos pela cadeia

De acordo com pesquisa realizada entre os anos de 2016 a 2017, para o teste vocacional em quatro escolas, sendo três públicas e uma privada, situadas nas cidades de Salvador e região metropolitana. Contatamos num total de 120 alunos. A questão frisava o seguinte:

— Em qual área: Exatas (E), Humanas (H) ou Biológicas, os alunos têm interesse em cursar, seja a nível técnico e/ou superior. Os resultados coletados foram que 26% escolheram a área de Exatas, 29% escolheram a área de Biológicas e 45% escolheram a área de Humanas. No entanto, dos alunos que escolheram área de Exatas, 24% migraram para Humanas e 50% para a área de Biológicas. Dos estudantes que escolheram a área de Biológicas, 22% migraram para Humanas e 49% para a área de Exatas, e por fim, dos estudantes que escolheram a área de Humanas, 43% migraram para Exatas e 12% para Biológicas. Com base nesses dados, determina-se o estado de Equilíbrio (Probabilidade de Estado Limite) das probabilidades de escolha no Momento 2 ( $M_2$ ) averiguando quais seriam as tendências de migração, ou seja, mudança para outras áreas.

Procedimentos:

Foram feitos cálculos, por meio das matrizes das probabilidades de transição, e verificando os vetores (Grafos Dirigidos <sup>1</sup>) de transição. Sendo que na matriz  $T$  das probabilidades de transição obtém-se na  $i$ -ésima linha e  $j$ -ésima coluna a probabilidade de transição do  $j$ -ésimo para o  $i$ -ésimo estado, respectivamente.

Todavia, calcula-se no tempo discreto, ou seja:  $T$  é finito ou enumerável.

Seguem-se os cálculos:

O vetor de Probabilidades é a matriz

$$P^{(n)} = \begin{bmatrix} P_E^{(n)} \\ P_B^{(n)} \\ P_H^{(n)} \end{bmatrix}$$

A Matriz  $T$  das Probabilidades de Transição é:

$$T = \begin{array}{c} \begin{matrix} & E & B & H \end{matrix} \\ \begin{bmatrix} 0,26 & 0,5 & 0,24 \\ 0,49 & 0,29 & 0,22 \\ 0,43 & 0,12 & 0,45 \end{bmatrix} \end{array}$$

Observa-se que:

$$T \cdot \begin{bmatrix} P_E^{(n)} \\ P_B^{(n)} \\ P_H^{(n)} \end{bmatrix} = \begin{bmatrix} P_E^{(n+1)} \\ P_B^{(n+1)} \\ P_H^{(n+1)} \end{bmatrix}$$

Realizando a operação de multiplicação de matrizes:

<sup>1</sup>Grafo Dirigido é um conjunto finito de elementos  $P_1, P_2, \dots, P_n$  junto com uma coleção finita de pares ordenados  $(P_i, P_j)$  de elementos distintos desse conjunto, sem repetição de pares ordenados, onde os elementos do conjunto são chamados vértices e os pares ordenados arestas dirigidas do grafo dirigido [4]

$$\begin{bmatrix} 0,26 & 0,5 & 0,24 \\ 0,49 & 0,29 & 0,22 \\ 0,43 & 0,12 & 0,45 \end{bmatrix} \times \begin{bmatrix} P_E \\ P_B \\ P_H \end{bmatrix} = \begin{bmatrix} P_E \\ P_B \\ P_H \end{bmatrix} \rightarrow \begin{bmatrix} 0,26P_E + 0,5P_B + 0,24P_H \\ 0,49P_E + 0,29P_B + 0,22P_H \\ 0,43P_E + 0,12P_B + 0,45P_H \end{bmatrix} = \begin{bmatrix} P_E \\ P_B \\ P_H \end{bmatrix}$$

*Observação 1.* Podemos adotar como verdade que  $P_E + P_B + P_H = 1$ , porque a probabilidade de escolha de Exatas ( $E$ ), Biológicas ( $B$ ) ou Humanas ( $H$ ) tem que ter como resultado 1 (o que também é chamado de probabilidade total <sup>2</sup>).

Retomando aos cálculos:

$$\begin{cases} 0,26P_E + 0,5P_B + 0,24P_H = P_E(I) \\ 0,49P_E + 0,29P_B + 0,22P_H = P_B(II) \\ 0,43P_E + 0,12P_B + 0,45P_H = P_H(III) \end{cases} \rightarrow \begin{cases} 0,26P_E + 0,5P_B + 0,24P_H = 1(I) \\ 0,49P_E + 0,29P_B + 0,22P_H = 1(II) \\ 0,43P_E + 0,12P_B + 0,45P_H = 1(III) \end{cases}$$

Por conveniência, foi utilizando método da substituição para um procedimento mais prático. Calculando na equação  $I$ :

$$\begin{aligned} 0,26P_E + 0,5P_B + 0,24P_H &= 1 \\ P_B &= 2 - 0,52P_E + 0,48P_H(I) \end{aligned}$$

Substituindo na equação  $II$ :

$$\begin{aligned} 0,49P_E + 0,29P_B + 0,22P_H &= P_B \\ 0,49P_E + 0,29(2 - 0,52P_E + 0,48P_H) + 0,22P_H &= 2 - 0,52P_E + 0,48P_H \\ 0,86P_E + 0,56P_H &= 1,42(II) \end{aligned}$$

Substituindo a equação  $I$  na equação  $III$ :

$$\begin{aligned} 0,43P_E + 0,12(2 - 0,52P_E + 0,48P_H) + 0,45P_H &= P_H \\ 0,37P_E - 0,61P_H &= -0,24(III) \end{aligned}$$

Realizando as substituições nas equações  $II$  e  $III$ :

$$-0,357P_E = -1,79 \Rightarrow P_E = \frac{1,79}{0,57} = 3,14$$

Substituindo na equação  $II$ :

$$0,86P_E + 0,56P_H = 1,42$$

<sup>2</sup>Tal previsão é importante, pois se chegarmos, por exemplo, à conclusão de que  $P_S^{(n)} \rightarrow 1$  quando  $n \rightarrow \infty$  (Comportamento Limite de Vetores-Estado: os vetores-estado convergem a um vetor fixo à medida que o número de observações cresce) - [4]

$$0,86(3,14) + 0,56P_H = 1,42 \Rightarrow P_H = \frac{-1,28}{0,56} = -2,29$$

Substituindo os valores de  $P_H = -2,29$  e  $P_E = 3,14$  na equação  $I$  obtemos:

$$P_B = 2 - 0,52P_E + 0,48P_H \Rightarrow P_B = 2 - 0,52(3,14) + 0,48(-2,29) \Rightarrow P_B = 1,5$$

Podemos então calcular os valores de probabilidades no Momento 2 ( $M_2$ ):

$$T \cdot \begin{bmatrix} P_E^1 \\ P_B^1 \\ P_H^1 \end{bmatrix} = \begin{bmatrix} P_E^2 \\ P_B^2 \\ P_H^2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0,26 & 0,5 & 0,24 \\ 0,49 & 0,29 & 0,22 \\ 0,43 & 0,12 & 0,45 \end{bmatrix} \cdot \begin{bmatrix} 3,14 \\ -2,29 \\ 1,5 \end{bmatrix} = \begin{bmatrix} P_E^2 \\ P_B^2 \\ P_H^2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0,26.(3,14) + 0,5.(-2,29) + 0,24.(1,5) \\ 0,49.(3,14) + 0,29.(-2,29) + 0,22.(1,5) \\ 0,43.(3,14) + 0,12.(-2,29) + 0,45.(1,5) \end{bmatrix} = \begin{bmatrix} P_E^2 \\ P_B^2 \\ P_H^2 \end{bmatrix} = \begin{bmatrix} 0,03 \\ 1,21 \\ 1,76 \end{bmatrix}$$

Retirando os valores inteiros e multiplicando por 100, encontra-se a probabilidade dos alunos que escolheram Exatas ( $P_E$ ), Biológicas ( $P_B$ ) e/ou Humanas ( $P_H$ ) mudarem a sua escolha:

$$\begin{aligned} \text{Exatas } P_E^{(2)} &= 3\% \\ \text{Biológicas } P_B^{(2)} &= 21\% \\ \text{Humanas } P_H^{(2)} &= 76\% \end{aligned}$$

Ou seja, a fidelização dos estudantes do terceiro ano do Ensino Médio, em longo prazo, se não houver modificações nas tendências de migração, teremos que apenas 3% dos alunos que escolheram a área de Ciências Exatas pensariam em migrar para outra área e **97% permanecem em Exatas**. Os 21% de alunos que escolheram a área de Ciências Biológicas migrariam para outras áreas e **79% permaneceriam**. E 76% dos estudantes que escolheram a área de Ciências Humanas migrariam, enquanto **24% permaneceriam**.

Tomando como base os mesmos percentuais de migração do momento  $M_1$  para cada área, acrescentando sobre os percentuais dos que permaneceram em  $M_2$ , obtemos que: **em Exatas totalizam 41,09%, em Biológicas 36,82% e em Humanas 21,89%**, ou seja, de fato a área de Exatas ainda sai na frente em relação às outras áreas.

Fazendo uma análise via grafos, ou diagramas de transição de estado (ou momentos), sobre os dados obtidos via cadeia de Markov, ainda se confirmam os resultados obtidos. Com uma cadeia de Markov por produtos consecutivos de  $T$  para obter os momentos  $M_1$ ,  $M_2$  e  $M_3$ , pode-se modelar a probabilidade de transição entre esses momentos. O gráfico a seguir (Figura 2) evidencia que é justamente mais provável ocorrer uma transição do momento  $M_1$  para o momento  $M_2$ , cujo valor é aproximadamente 55%.

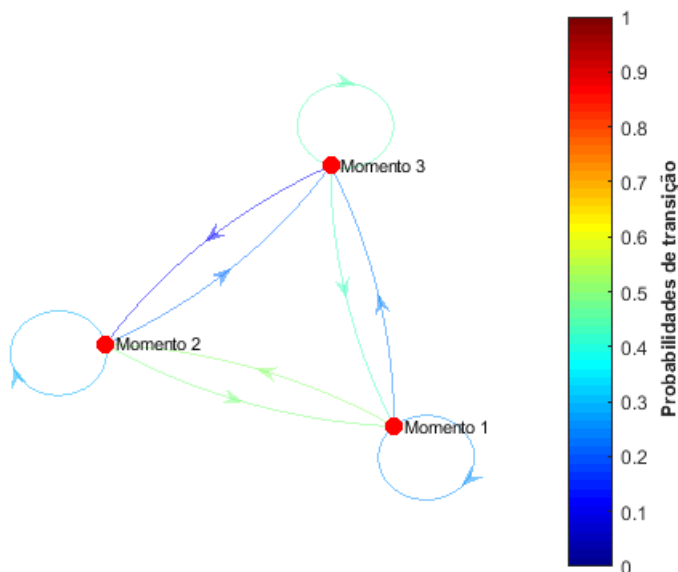


Figura 2: Probabilidades de transição da Cadeia de Markov da matriz  $T$  (Fonte própria)

Além disso, essa matriz  $T$  tem um comportamento ergódico nesse Processo Markoviano. Isso nos garante ser um processo recorrente não periódico [3]. Os autovalores de  $T$  indicam isso, pela espessura ou intervalo de afastamento entre seus valores, que foram:  $\lambda_1 = 1$ ,  $\lambda_2 = 0,2274$  e  $\lambda_3 = -0,2274$ , como pode-se observar na Figura 3.

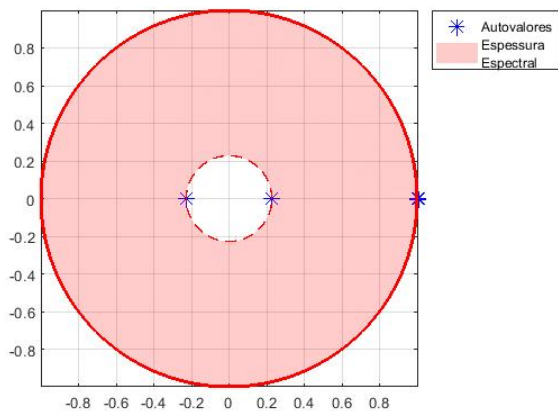


Figura 3: Autovalores da matriz de probabilidades  $T$  (Fonte própria)

#### 4.1. Comparação entre oferta do mercado de trabalho e a tendência de formação

Esse trabalho baseou-se numa pesquisa realizada nas escolas Alfredo Agostinho de Deus, da rede pública estadual, Ramo da Videira, rede particular, Francisco Pereira Franco, da rede pública estadual, e no Colégio Modelo Luís Eduardo Magalhães. Em um total de 120 alunos observa-se que no momento  $M_1$  os estudantes estão mais voltados às Ciências Humanas, como demonstra a Figura 4.

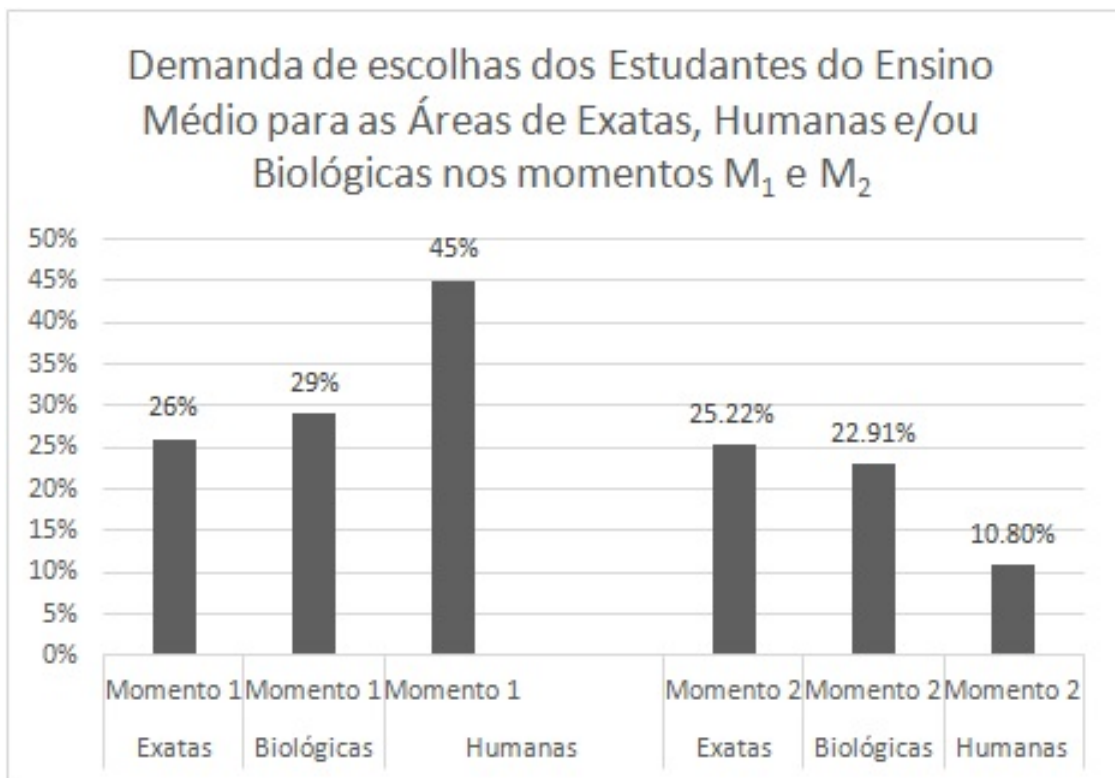


Figura 4: Percentuais por área escolhida nos momentos 1 e 2 (Fonte própria)

Ainda pela Figura 4 percebe-se pelos percentuais do momento  $M_2$  que os estudantes que escolhem área de Exatas ou Biológicas permanecem praticamente fidedignos às suas respectivas escolhas. No que tange a área de Humanas, um relevante percentual tem perspectiva de mudar para outra área, o que se pode se tornar ainda maior se fosse levado adiante o Processo Estocástico, visto que o total das probabilidades  $P_E + P_B + P_H = 1$  quando de  $n \rightarrow \infty$ .

Isso, segundo dados coletados na Unidade Central do SineBahia, com informações cedidas do número de vagas para trabalho divulgadas (por ocupação) entre os meses de Setembro/2017, Outubro/2017, Novembro/2017, Dezembro/2017, Janeiro/2018 e Fevereiro de 2018, sendo que no

interim de 2017 foram 5.923 vagas e no período de 2018 2.395 vagas, totalizando 8.318 vagas. Dessas, foram extraídas as ocupações que necessitavam de algum curso voltado à área de pesquisa desse trabalho: Exatas, Biológicas ou Humanas. Do total dessas 8.318 vagas foram retiradas 548, sendo 235 da área de Exatas, 185 da área de Humanas e 128 da área de Biológicas, o que equivale em respectivos percentuais (E) = 43%, (H) = 34% e (B) = 23%, ou seja, a maior oferta tem sido na área de Exatas (E) com 43% das vagas. Se colocarmos em evidência o resultado da Cadeia de Markov, verificado no momento  $M_2$ , pode-se contrastar com as ocupações disponíveis no mercado de trabalho, de acordo informações cedidas pelo SineBahia, conforme a Figura 5:

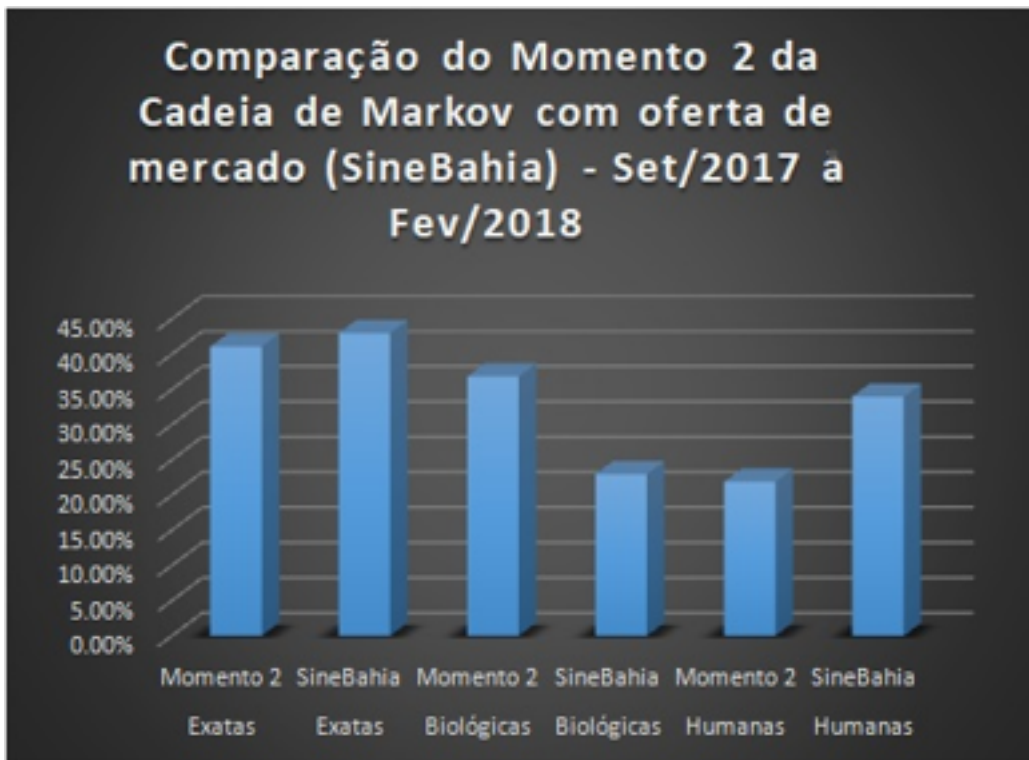


Figura 5: Percentuais de escolha no momento  $M_2 \times$  oferta de vagas/SineBahia - (Fonte própria)

Diante dos dados coletados, pode-se observar que os estudantes que optam pela área de Exatas estão propensos a manter o caminho escolhido. E, com isso, estão inclinados a garantir o sucesso para angariar um emprego, devido ao número de oportunidades ofertadas pelo mercado. O resultado apresentado na matriz de transição de probabilidades 1, do momento  $M_1$ , isto é, momento em que a maioria dos estudantes tinham decidido pela área de Humanas, percebeu-se que isso provavelmente se deu por estar relacionado a aspectos sociais, políticos ou até mesmo religiosos. Ao questionar a motivação sobre suas escolhas, os estudantes responderam que tiveram influência dos pais ou por conta de princípios religiosos adquiridos. Isso ocorreu principalmente entre os entrevistados que optaram pelos cursos de Filosofia, Serviço Social e Teologia.

## 5. Conclusão

Hoje, ao observarmos as escolhas de cursos para uma formação acadêmica ou técnica, pretendidos pelos estudantes do ensino fundamental II e médio, percebe-se que a vocação deixou de ser a única influência na decisão desses jovens que fluirão para o mercado de trabalho. Muitos se deixam seduzir pelo poder salarial que algumas profissões oferecem, alguns outros ainda são influenciados por tradição ou tendência familiar e outros por aspectos religiosos e sociais. Assim, buscou-se analisar quais são as atuais tendências ao mercado de trabalho, levantando dados sobre possibilidades de escolhas de áreas para uma formação acadêmica ou técnica, de alunos do ensino médio.

Para tal análise de dados estatísticos foi utilizado o processo de Markov, o qual nem sempre é um processo periódico cíclico, podendo ser não periódico. No caso do presente trabalho, o processo é recorrente, ou seja, uma vez que se entrou no estado  $M_i$  do processo estocástico da Cadeia de Markov, gerado pelas matrizes de transição de probabilidades  $T$ , um retorno a algum momento  $M_i$  é assegurado dentro de um número finito de passos, porém ele não é periódico e pode voltar antes de qualquer passo  $n$ . Essa característica o identifica como um estado ergódico.

Analisando os resultados obtidos, percebe-se que as preferências dos alunos do ensino médio no terceiro ano, quando tomadas em virtude das áreas por eles escolhidas, contribuem para um caminho que se harmoniza com o número de vagas disponíveis no mercado de trabalho, na área de Exatas, que segundo os dados do SineBahia, estava em alta nos períodos de setembro de 2017 a fevereiro de 2018, pois no momento  $M_2$  a permanência total em Exatas foi de 41,09% com 43% de vagas ofertadas nessa área. De fato, os estudantes que têm preferência pela área de Exatas dificilmente costumam ter aptidão para Biológicas e tendem ainda menos para Humanas.

Para o estudo aqui desenvolvido, foi utilizada a teoria de processos estocásticos aplicados de modo probabilístico estatístico, e pretende-se ampliar esse estudo, futuramente, através da Teoria Ergódica, numa linha de pesquisa de Sistemas Dinâmicos.

## Referências

- [1] Boldrini, José L.; Costa, Suely I. R.; Figueiredo, Vera Lúcia e Wetzler, Henry G. Álgebra Linear. 3ª Edição, São Paulo: Harbra, (1980).
- [2] Kolman, Bernard. Introdução à álgebra Linear com Aplicações, 6ª Edição, Rio de Janeiro: LTC Livros Técnicos e Científicos Editora, (1999).
- [3] Nogueira, Fernando. Modelagem e Simulação – Cadeias de Markov. Notas de aula. UFJF - Universidade Federal de Juiz de Fora, (2009).
- [4] Rorres, Anton. A álgebra Linear com Aplicações, 8ª Edição, Porto Alegre: Bookman, (2004).



George Mamede  
Universidade Católica do Salvador – UCSAL  
<[georgemamede@gmail.com](mailto:georgemamede@gmail.com)>

Vânia Santos  
Universidade Católica do Salvador – UCSAL  
Universidade do Estado da Bahia – UNEB  
<[vania.santos@pro.ucsal.br](mailto:vania.santos@pro.ucsal.br)>

Recebido: 03/07/2018