

UNIVERSIDADE FEDERAL DO PARANÁ

BRENDHA RODRIGUES DE LIMA
LEONARDO HENRIQUE BARCHE KRUGER

Aplicação de Cadeias de Markov para determinação de preços de
insumo na cidade de Curitiba

CURITIBA

2018

SUMÁRIO

1. RESUMO	03
2. INTRODUÇÃO	04
3. DEFINIÇÃO DE PREÇOS.....	05
3.1 ASPECTO MERCADOLÓGICO	05
3.2 ASPECTO FINANCEIRO.....	05
3.3 CALCULANDO PREÇOS	05
4. PROCESSOS ESTOCÁSTICOS	07
4.1 CADEIAS DE MARKOV	07
5. ESTUDO DE CASO	09
5.1 METODOLOGIA	09
5.2 RESULTADOS	10
6. CONCLUSÃO	14
7. REFERÊNCIAS	15
8. ANEXOS.....	16

1. RESUMO

Acompanhar a variação de preço de um produto é de fundamental importância para atividades econômicas. Por tanto, o objetivo do presente estudo é analisar a variação de preços em produtos na cidade de Curitiba, mais precisamente, a variação do preço do quilo do pão por meio das cadeias de Markov a partir da análise os dados históricos e identificação das probabilidades de ocorrência dos estados e o desenvolvimento de um modelo que apresente de maneira prática essa transição, observando informações no período de final março a começo de junho de 2018.

2. INTRODUÇÃO

Acompanhar a variação de preço de um produto é de fundamental importância para atividades econômicas, pois assim podemos entender e melhorar/modificar, tanto para quem compra como para quem vende. Portanto, o objetivo do presente estudo é analisar a variação de preços em produtos na cidade de Curitiba, mais precisamente, a variação do preço do quilo do pão por meio das cadeias de Markov. Atualmente, os preços são levantados em toda a cidade por meio da prefeitura, a partir de uma coleta de dados sobre o preço atual de cada produto em diferentes mercados, sendo assim, por meio da compreensão econômica e dos processos estocásticos, é possível projetar como o preço irá se comportar.

No entanto, com a atual tecnologia não é possível prever todos os desdobramentos destes valores. Isto se deve ao fato de que este apresenta um envolvimento de várias variáveis, por exemplo, o preço de mercado de seus insumos principais e inflação nacional, além das questões políticas envolvidas. Ou seja, um fator, que pode ser menor do que a margem de erro dos dados numéricos, pode desencadear em variações de preços imprevisíveis.

A partir da análise os dados históricos e identificação das probabilidades de ocorrência dos estados, o desenvolvimento de um modelo que apresente de maneira prática essa transição, observando informações no período de final março a começo de junho de 2018.

3. DEFINIÇÃO DE PREÇOS

A definição do preço adequado de venda de um produto/serviço junto ao mercado depende do equilíbrio entre o preço de mercado e o valor calculado, em função dos seus custos e despesas. o valor deve cobrir o custo direto da mercadoria/produto/serviço, somado as despesas variáveis e fixas proporcionais. Além disso, deve gerar lucro líquido. Para definir o preço de venda de um produto e/ou serviço, o empresário deve considerar dois aspectos: o mercadológico (externo) e o financeiro (interno).

3.1 ASPECTO MERCADOLÓGICO

Pelo aspecto mercadológico, o preço de venda deverá estar próximo do praticado pelos concorrentes diretos da mesma categoria de produto e qualidade. Fatores como conhecimento da marca, tempo de mercado, volume de vendas já conquistado e agressividade da concorrência também exercem influência direta sobre o valor do produto.

3.2 ASPECTO FINANCEIRO

Aqui, o preço de venda deverá cobrir o custo direto da mercadoria/produto/serviço vendido, as despesas variáveis (por exemplo, comissões de vendedores), as despesas fixas (como aluguel, água, luz, telefone, salários, pró-labore).

3.3 CALCULANDO PREÇOS

Para calcular a previsão de vendas de produtos/serviços, a empresa deve seguir algumas alternativas:

- a) Com base nas informações internas, analisar o comportamento das vendas realizadas em um determinado período e projetá-la para o mesmo período seguinte. Alguns aspectos podem interferir nessa projeção, como concorrentes, novos produtos, novos hábitos dos consumidores e eventos especiais, tais como festas, Olimpíadas, eleições, Copa do Mundo etc.;
- b) Por meio de pesquisas de mercado, realizar um estudo da

demanda de mercadorias/produtos/serviços que poderia ser atendida pela empresa. Variáveis externas, como população, atividade econômica, situação política, nível de renda e emprego, concorrência, novos produtos etc., devem ser consideradas.

4. PROCESSOS ESTOCÁSTICOS

Processo estocástico é uma coleção de variáveis aleatórias (sobre um espaço de probabilidade (Ω, F, P) , indexados por um conjunto de índices T . O conjunto T é chamado de espaço paramétrico. Os valores assumidos por $X(t)$ são chamados de estados, e o conjunto de todos os possíveis estados é chamado de espaço de estados do processo estocástico.

4.1 CADEIAS DE MARKOV

Um processo estocástico ser considerado uma Cadeia de Markov, seu comportamento dinâmico deve respeitar uma característica, chamada propriedade markoviana. Essa característica demonstra a probabilidade condicional que em um evento futuro, dado qualquer evento passado, no estado presente $X_t = i$, depende somente do estado presente do processo, não importando como o processo chegou a tal estado.

Para ser considerado uma Cadeia de Markov, um processo estocástico $\{X_t, t = 0, 1, 2, \dots\}$ com espaço de estado $S = \{1, 2, \dots, s\}$ deve satisfazer a propriedade de Markov, para todo n pertencente aos números naturais e para todo i pertencente ao conjunto S :

$$P(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_n = i_n | X_{n-1} = i_{n-1})$$

Para a cadeia ser considerada homogênea ou estacionária no tempo, a probabilidade de ir de um estado a outro deve independender do tempo em que o passo é dado. Ou seja, para quaisquer estados $i, j \in S$, temos:

$$P(X_n = j | X_{n-1} = i) = P(X_{n+k} = j | X_{n+k-1} = i), \text{ para } k = -(n-1), -(n-2), \dots, -1, 0, 1, 2, \dots$$

Num determinado conjunto de estados, o processo começa em um desses estados e move-se sucessivamente de um estado para outro. Se a cadeia está atualmente no estado s_i , então ela se move para o estado s_j no próximo passo com uma probabilidade denotada por p_{ij} , e essa probabilidade não depende dos estados ocorridos nos passos anteriores, apenas do estado atual. Este termo é chamado de probabilidade de transição.

Realizar uma matriz é mais conveniente de resumir as probabilidades de transição. Assim, a matriz P define a cadeia de Markov, sendo todas as probabilidades de transição fixas e independentes ao longo do tempo. Para determinar a matriz, eleva-se a matriz P na n -ésima potência. A matriz P , demonstrada abaixo, expressa as probabilidades de transição de uma Cadeia de Markov

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots \\ p_{21} & p_{22} & \dots & p_{2j} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ p_{i1} & p_{i2} & \dots & p_{ij} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix}$$

Sendo X_n , com $n \geq 0$ uma Cadeia de Markov com espaço de estados S e função de transição P . Seja um vetor de números $\pi(x)$, com $x \in S$, não negativo e cujo a soma resulte em 1 e se $\sum \pi(x)P(x, y) = \pi(y)$, $x, y \in S$, então π será chamado de distribuição estacionária.

Supondo que essa distribuição existe e que $\lim_{n \rightarrow \infty} P^n = \pi(y)$, $y \in S$. Então, independentemente da distribuição inicial da cadeia, a distribuição de X_n irá se aproximar de π conforme n vai para infinito.

A condição de distribuição estacionária pode ser escrita na forma matricial como $\pi.P = \pi$.

5. ESTUDO DE CASO

5.1 METODOLOGIA

A metodologia aplicada para desenvolvimento do trabalho foi dividida em quatro etapas:

Etapa 1 – Coleta de dados: os dados utilizados para este trabalho foram retirados de uma base de dados do Programa "Disque Economia". Este sistema viabiliza o serviço prestado pela Secretaria Municipal do Abastecimento de Curitiba, a qual coloca à disposição para consulta o preço de 302 itens que são coletados em 14 supermercados da cidade. A lista dos itens pesquisados é composta por gêneros alimentícios como hortifrutigranjeiros, bebidas, massas, carnes e ainda material de higiene e limpeza, levando em conta quantidades e marcas, no período de dias úteis dos últimos 3 meses;

Etapa 2 – Identificação das probabilidades: realizou-se a separação dos dados em faixas de preços para o quilo do pão, ou seja, de R\$0,00 até R\$8,50, marcou-se o produto na categoria "A", de R\$8,51 a R\$8,99, na categoria "B" e de R\$9,00 até "infinito", na categoria "C". Sendo assim, realizou-se a contagem do número de transições ocorridas entre as faixas e desta forma, obteve-se a frequência absoluta das transições entre as faixas no intervalo de tempo considerado, e com estes dados construiu-se a tabela de transição;

Etapa 3 – Montagem da cadeia de Markov: Após a montagem da tabela de transição foi obtida a matriz de transição probabilística. As entradas dessa matriz contêm as probabilidades associadas à transição entre estado.

Etapa 4 – Validação do método: Nessa etapa, buscou-se validar a

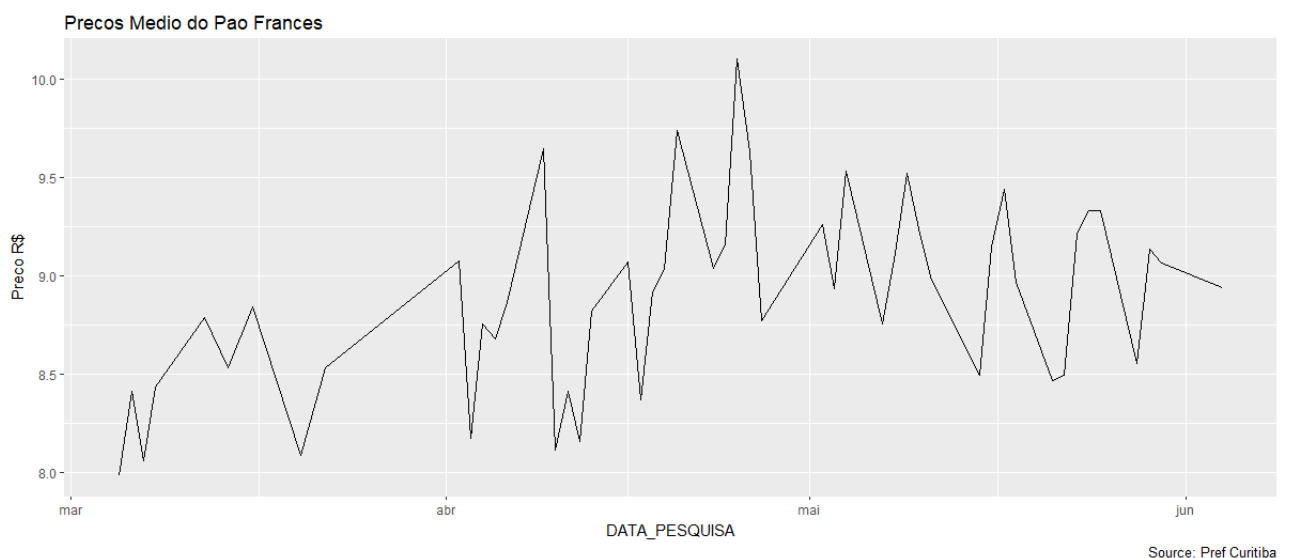
previsão gerada por meio da Cadeia de Markov.

5.2 RESULTADOS

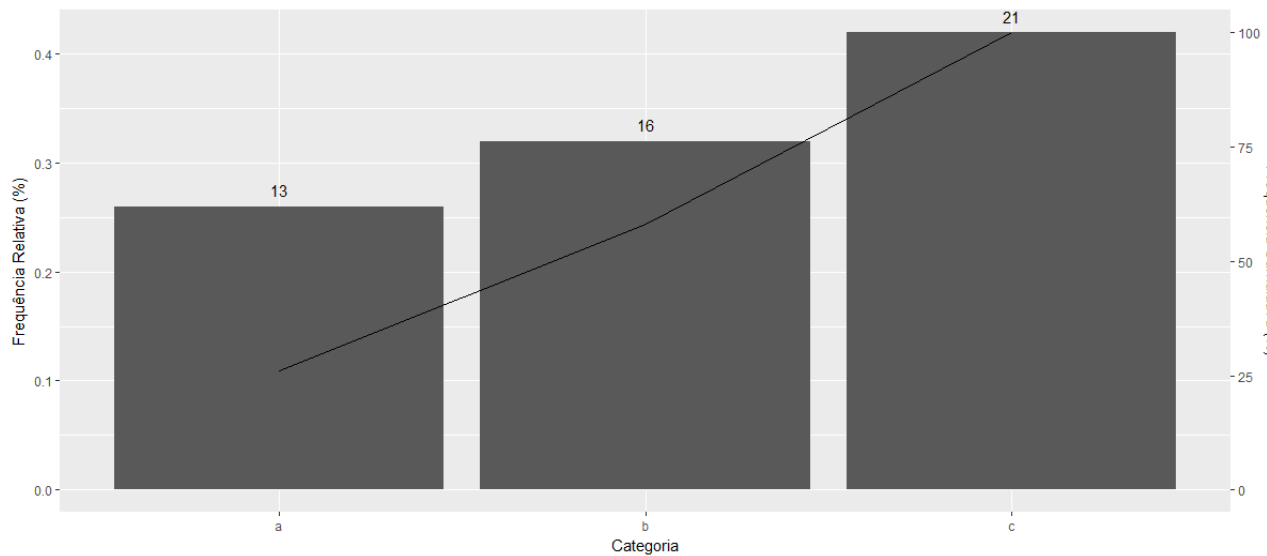
Antes da descrição das etapas realizadas, foi realizado uma análise descritiva dos dados, afim de verificar como os dados estão distribuídos.

No gráfico abaixo, é possível verificar a variação dos preços observados, nota-se que os preços rotineiramente têm bastante variação, isso deve-se ao fato de que não se estabelece uma faixa fixa de variação em muitos comércios e também ao fato de nossa base ser bastante diversificada com relação aos comércios.

O que também pode explicar estes valores é que houve, recentemente, reajustes nas tarifas de luz e água, assim como o preço da farinha e outros ingredientes.



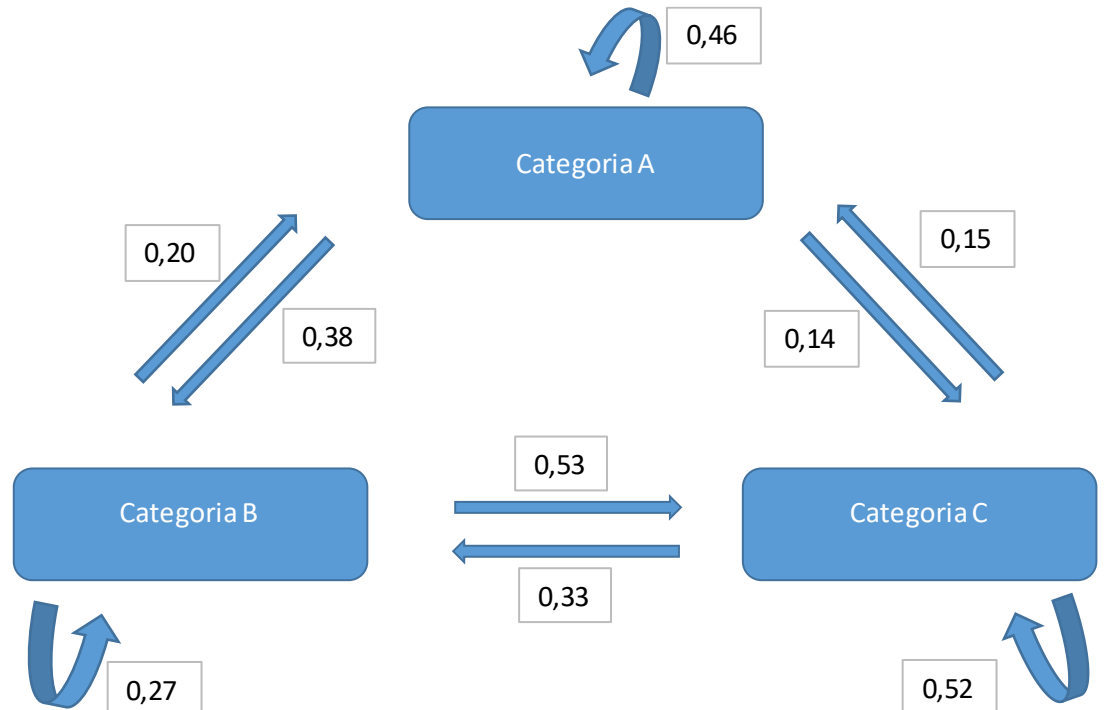
Na etapa um, foram utilizados dos dados fornecidos pela prefeitura, para o período de março a junho, foram totalizados 57 dias, com 435 observações. Na etapa dois, realizou-se a separação dos dados nas três categorias já citadas, de acordo com o preço do quilo no pão. Sendo assim, seu histograma apresentou dessa forma:



Após, foi realizada a contagem do número de transições das faixas correspondentes, obteve-se a frequência absoluta das transições entre uma faixa inicial para cada uma das possibilidades de faixas finais no intervalo de tempo considerado, e através dela, obtivemos a matriz de transição encontrada abaixo. Utilizou-se o software R, com o pacote `marcovchain` e a função `steadyStates`. Abaixo da matriz de transição, encontra-se os códigos utilizados. As matrizes de transição iteradas encontram-se no anexo 1.

	A	B	C
A	0.4428239	0.3964200	0.1607560
B	0.2007225	0.2614132	0.5378643
C	0.1548791	0.3397767	0.5053442

Analisando a matriz podemos ver que os todos os estados se interligam além disso $\mathbb{P}_{ii}^n > 0$, para $n \geq 1$, portanto o período que é o máximo divisor comum é 1, o que, por definição, torna o processo aperiódico e redutível, pois existe a possibilidade de mudra de estado em mais de um passo.



#Partindo para Matriz de Transição

```
>precos <- resumo_2$class_preco
```

```
>mcFit <- markovchainFit(data=precos,method = "bootstrap",nboot = 1000)
```

```
#matriz de transição preco
```

```
>mcFit$estimate
```

```
>is.irreducible(mcFit$estimate)
```

```
>period(mcFit$estimate)
```

```
>steadyStates(ProbT)
```

```
>show(mcFit$estimate)
```

```
#distribuica estacionária
```

```
>steadyStates(mcFit$estimate)
```

Com base na matriz de transição foi feito uma tabela para prever a categoria de preço do pão nos próximos 8 dias, foi observada que próximo da 9ª iteração a matriz alcançou sua estabilidade e obteve 56% de acerto tendo a classe como maior probabilidade a classe predita.

Data	Probabilidades				Predito
	Observado	A	B	C	
18/05/2018	B				C
21/05/2018	A	0.2007225	0.2614132	0.5378643	C
22/05/2018	A	0.3005612	0.3337949	0.3656439	C
23/05/2018	C	0.2567264	0.3306442	0.4126294	C
24/05/2018	C	0.2365341	0.3272893	0.4361766	C
25/05/2018	C	0.2379919	0.3275273	0.4344808	C
28/05/2018	B	0.2384226	0.3275912	0.4339862	C
29/05/2018	C	0.2385729	0.3276141	0.4338130	C
30/05/2018	C	0.2385869	0.3276162	0.4337968	C
04/06/2018	B	0.2385980	0.3276179	0.4337841	C
Acertos					56%

Sendo que a partir da 9ª iteração as probabilidades de transição permanecem inalteradas, demonstrando que existe uma distribuição estacionária, o vetor da distribuição estacionária encontrado foi este:

A	B	C
0.2368421	0.3238866	0.4392713

6. CONCLUSÃO

Para os dados estudados sobre o quilo do pão francês em Curitiba no últimos 3 meses de 2018 com o preço segmentado de uma forma arbitrária não foi possível prever com acurácia e precisão desejada.

Porém, as propriedades de Markov foram facilmente verificadas pelo Pacote *markovchain* e também as operações e com o conjunto de dados foram feitas sem grandes dificuldades utilizando o pacote.

Cabe para estudos futuros desenvolver uma métrica eficiente para categorizar o preço desse produto a fim de obter resultados mais significativos.

7. REFERÊNCIAS:

AGUILERA, B. O. Processos Estocásticos Aplicados – Anotações em Aula. Universidade Federal do Paraná.

HOEL, PORT & STONE. Introduction to Stochastic Processes. Houghton Mifflin Co.

SEBRAE NACIONAL. Como definir o preço de venda de um produto ou serviço. Disponível em:
<<http://www.sebrae.com.br/sites/PortalSebrae/artigos/como-definir-o-preco-de-venda-de-um-produto-ou-servico,cc9836627a963410VgnVCM1000003b74010aRCRD>>. Acesso em: 09/06/2018

PREFEITURA DE CURITIBA. Disque economia. Disponível em:
<<http://www.curitiba.pr.gov.br/dadosabertos/consulta/?grupo=0>>. Acesso em: 09/06/2018

PASSOLINI, M.; FRANCO, M. M.; CORSO, L. L. . APLICAÇÃO DAS CADEIAS DE MARKOV PARA ANALISAR A VARIABILIDADE DO PREÇO DA COMMODITY LEITE.

LUCAMBIO, F. Disponível em:
<<https://docs.ufpr.br/~lucambio/CE064/1S2018/CM.pdf>>. Acesso em: 09/06/2018

SPEDICATO, G. A. The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R.

8. ANEXOS

Segunda iteração			
	A	B	C
A	0.3005612	0.3337949	0.3656439
B	0.2246601	0.3306611	0.4446788
C	0.2150522	0.3219235	0.4630243

Terceira iteração			
	A	B	C
A	0.2567264	0.3306442	0.4126294
B	0.2347274	0.3265905	0.4386821
C	0.2315603	0.3267309	0.4417088

Quarta iteração			
	A	B	C
A	0.2439600	0.3284081	0.4276319
B	0.2374396	0.3274797	0.4350807
C	0.2365341	0.3272893	0.4361766

Quinta iteração			
	A	B	C
A	0.2401814	0.3278602	0.4319584
B	0.2382614	0.3275636	0.4341750
C	0.2379919	0.3275273	0.4344808

Sexta iteração			
	A	B	C
A	0.2390683	0.3276891	0.4332426
B	0.2385018	0.3276036	0.4338946
C	0.2384226	0.3275912	0.4339862

Sétima iteração			
	A	B	C
A	0.2387399	0.3276395	0.4336206
B	0.2385729	0.3276141	0.4338130
C	0.2385495	0.3276106	0.4338399

Oitava iteração			
	A	B	C
A	0.2386431	0.3276248	0.4337321
B	0.2385938	0.3276173	0.4337889
C	0.2385869	0.3276162	0.4337968

Nona iteração			
	A	B	C
A	0.2386145	0.3276204	0.4337650
B	0.2386000	0.3276182	0.4337818
C	0.2385980	0.3276179	0.4337841