

# Uso do Teste de Aleatorização na Análise das Séries Temporais

Aline Palafoz Pereira\*, Denise Nunes Viola\*\* e Gilênio Borges Fernandes\*\*

\*Aluna do curso de Estatística da UFBA

\*\*Professor(a) do Departamento de Estatística da UFBA

**Resumo** Muitas vezes o pesquisador tem interesse em saber se existe tendência em uma série temporal. Uma maneira de verificar essa tendência é através dos mínimos quadrados, mas nem sempre os dados apresentam os pressupostos para utilizar esta técnica. Quando os pressupostos não são atendidos, uma alternativa é verificar a tendência através do teste de aleatorização, que indica se existe ou não algum padrão nos os dados. Para rejeitar a hipótese nula usamos o p-valor que é calculado a partir da proporção de vezes que a estatística de teste após a aleatorização é maior que a estatística obtida com os dados originais. Se o p-valor for menor que o nível de significância, rejeita-se a hipótese nula. Para ilustrar este teste foi feito um experimento ao longo de um mês com o objetivo de verificar se existe tendência no crescimento da planta. Após 10.000 aleatorizações, verificou-se que  $p\text{-valor}=0,023$ , logo, rejeita-se a hipótese nula, portanto, existe tendência na série.

## 1. Introdução

Muitas vezes, quando o pesquisador está estudando uma série temporal, ele tem interesse em saber se existe tendência entre as observações. Uma maneira de verificar essa tendência, quando os pressupostos não são atendidos, é através do teste de aleatorização, que indica se existe ou não algum padrão nos os dados. Este teste tem sido utilizado em diversos estudos, principalmente na área de Biologia (MANLY, 2006). Para rejeitar a hipótese nula usamos o p-valor que é calculado a partir da proporção de vezes que a estatística de teste após a aleatorização é maior que a estatística obtida com os dados originais. Se o p-valor for menor que o nível de significância, rejeita-se a hipótese nula.

## 2. Metodologia

### 2.1 Tendência:

É um componente da Sucessão Cronológica e compreende os movimentos que se manifestam suavemente e consistentemente durante períodos longos. Porém existe um problema ao definir tendência, pois um “período longo” para uns pode ser pequeno para outros, ou seja, o número de anos em que deve considerar a tendência varia de caso para caso (MURTEIRA, B. J. F. 1993).

No exemplo a seguir, podemos ver um banco de dados do IPCA – 15 (Índice Nacional de Preços ao Consumidor Amplo – 15), no qual podemos perceber uma tendência nos dados através de gráficos.

Quadro 1: Banco de dados do IPCA – 15 do ano de 2001 à 2009.

| Anos | Jan  | Fev  | Mar  | Abr  | Mai  | Jun   | Jul   | Ago  | Set  | Out  | Nov  | Dez  |
|------|------|------|------|------|------|-------|-------|------|------|------|------|------|
| 2001 | 0,63 | 0,5  | 0,36 | 0,5  | 0,49 | 0,38  | 0,94  | 1,18 | 0,38 | 0,37 | 0,99 | 0,55 |
| 2002 | 0,62 | 0,44 | 0,4  | 0,78 | 0,42 | 0,33  | 0,77  | 1    | 0,62 | 0,9  | 2,08 | 3,05 |
| 2003 | 1,98 | 2,19 | 1,14 | 1,14 | 0,85 | 0,22  | -0,18 | 0,27 | 0,57 | 0,66 | 0,17 | 0,46 |
| 2004 | 0,68 | 0,9  | 0,4  | 0,21 | 0,54 | 0,56  | 0,93  | 0,79 | 0,49 | 0,32 | 0,63 | 0,84 |
| 2005 | 0,68 | 0,74 | 0,35 | 0,74 | 0,83 | 0,12  | 0,11  | 0,28 | 0,16 | 0,56 | 0,78 | 0,38 |
| 2006 | 0,51 | 0,52 | 0,37 | 0,17 | 0,27 | -0,15 | -0,02 | 0,19 | 0,05 | 0,29 | 0,37 | 0,35 |
| 2007 | 0,52 | 0,46 | 0,41 | 0,22 | 0,26 | 0,29  | 0,24  | 0,42 | 0,29 | 0,24 | 0,23 | 0,7  |
| 2008 | 0,7  | 0,64 | 0,23 | 0,59 | 0,56 | 0,9   | 0,63  | 0,35 | 0,26 | 0,3  | 0,49 | 0,29 |
| 2009 | 0,4  | 0,63 | 0,11 | 0,36 | 0,59 | 0,38  | 0,22  | 0,23 | 0,19 | 0,18 | -    | -    |

Fonte: IBGE

Após análise do Quadro 1, podemos elaborar um gráfico e perceber que, aparentemente existe uma tendência decrescente, ou seja, a medida que passa o tempo, o IPCA diminui.

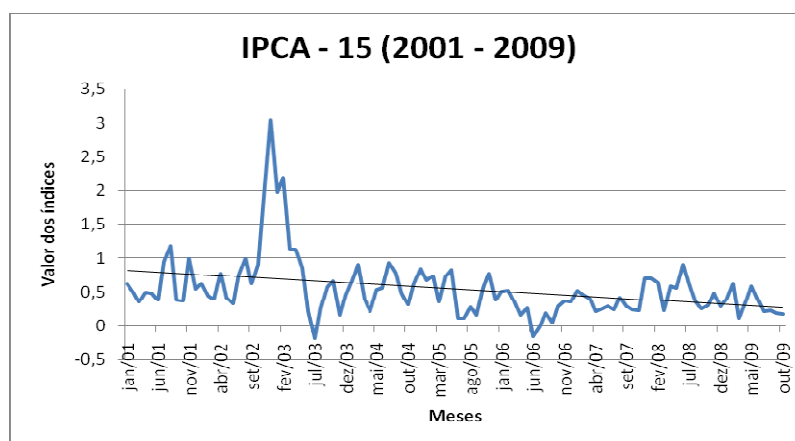


Figura 1: IPCA – 15 (dados de 2001 a 2009).

Porém, se reduzirmos esse mesmo banco de dados, pegando a observação de dois anos, a tendência fica mais fácil de ser visualizada, como podemos ver no exemplo a seguir.

Quadro 2: Banco de dados do IPCA – 15 do ano de 2008 e 2009.

| Anos | Jan | Fev  | Mar  | Abr  | Mai  | Jun  | Jul  | Ago  | Set  | Out  | Nov  | Dez  |
|------|-----|------|------|------|------|------|------|------|------|------|------|------|
| 2008 | 0,7 | 0,64 | 0,23 | 0,59 | 0,56 | 0,9  | 0,63 | 0,35 | 0,26 | 0,3  | 0,49 | 0,29 |
| 2009 | 0,4 | 0,63 | 0,11 | 0,36 | 0,59 | 0,38 | 0,22 | 0,23 | 0,19 | 0,18 | -    | -    |

Fonte: IBGE

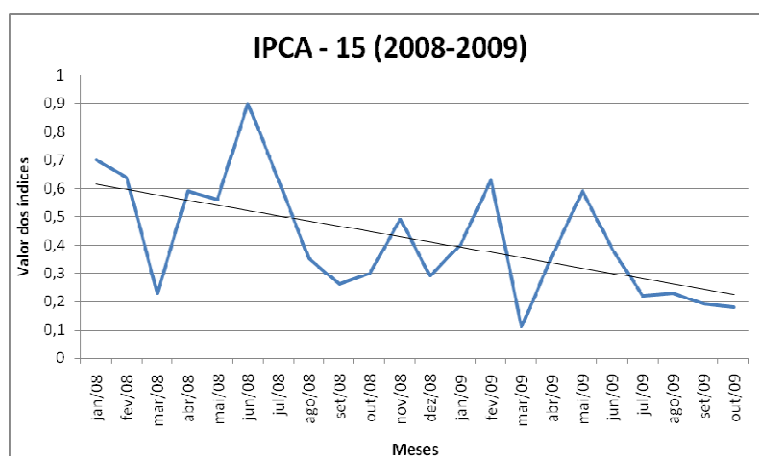


Figura 2: IPCA – 15 (dados de 2008 e 2009).

## 2.2 – Teste de Aleatorização:

O teste de aleatorização é um procedimento em que se comparam valores de uma estatística observada nos dados com os valores desta estatística após a aleatorização das observações. A principal vantagem deste teste é que ele pode ser utilizado para pequenas amostras, aleatórias ou não, e, como se trata de um teste não-paramétrico, prescinde da distribuição da população da qual a amostra foi extraída. Os testes estatísticos clássicos exigem amostras aleatórias para qualquer estudo, porém, algumas vezes, amostras não-aleatórias são mais convenientes ao pesquisador (PRAZESES FILHO, VIOLA e LIMA, 2009). Dentre as vantagens do teste de aleatorização, uma delas é que este pode ser usado para um banco de dados grande ou pequeno, além de não apresentar os pressupostos dos testes convencionais. A principal desvantagem do teste é que não é possível generalizar conclusões para a população de interesse (MANLY, 2006).

A utilização deste teste tem se tornado mais freqüente com o avanço computacional. Atualmente existem diversos *softwares* estatísticos, como por exemplo, o R, através do qual é possível aplicar o teste de aleatorização. Este teste é baseado na pressuposição de que, se a hipótese nula é verdadeira, todas as possíveis ordens dos dados são igualmente prováveis. A hipótese nula diz que não existe padrão nos dados, ou se existe este padrão, isto é efeito do acaso enquanto a hipótese alternativa afirma que os dados apresentam certo padrão (MANLY, 2006).

Para a realização deste teste, calcula-se o valor de uma estatística “E” de um conjunto de dados ( $e_o$ ). A seguir, aleatoriza-se os dados um grande número de vezes e calculam-se as estatísticas para cada uma das aleatorizações ( $e_{ai}$ , em que  $i$  é a  $i$ -ésima aleatorização). A proporção de vezes que a estatística aleatorizada foi maior que a observada é comparada com o nível de significância adotado. Rejeita-se a hipótese nula quando esta proporção for menor que o nível de significância (VIOLA, 2007).

### **2.3 – Experimento:**

Para verificar uma tendência com o teste de aleatorização no dia-a-dia, foi feito um experimento ao longo de um mês com o objetivo de verificar as diferenças de crescimento de sementes de milho plantadas de formas iguais e cultivadas de maneiras diferentes.

Chamamos de experimento A as sementes que foram deixadas em um local seco e arejado onde pudesse pegar sol e sombra e experimento B as sementes que foram deixadas em um local fechado onde não pudesse pegar luz solar em nenhum momento do dia. Ambos os experimentos foram plantados no dia 22/01/2010, regados uma vez ao dia e as medidas tiradas a cada três dias, a partir do momento em que estas germinaram. Para esta análise será considerado apenas o experimento B. A seguir veremos os resultados resumidamente.

Quadro 3: Resultado do experimento realizado em 2010.

| Experimentos |        |       |        |
|--------------|--------|-------|--------|
| A            |        | B     |        |
| Dia          | Altura | Dia   | Altura |
| 25/01        | --     | 25/01 | --     |
| 28/01        | 1,2cm  | 28/01 | 0,5cm  |
| 31/01        | 9cm    | 31/01 | 8,0cm  |
| 03/02        | 14cm   | 03/02 | 13,5cm |
| 06/02        | 20cm   | 06/02 | 19,5cm |
| 09/02        | 28cm   | 09/02 | 19,5cm |
| 12/02        | 30cm   | 12/02 | --     |
| 15/02        | 31cm   | 15/02 | --     |
| 18/02        | 33cm   | 18/02 | --     |
| 21/02        | 34cm   | 21/02 | --     |
| 24/02        | 37cm   | 24/02 | --     |
| 27/02        | --     | 27/02 | --     |

### 3. Resultados:

Sabemos que uma equação de tendência tem a seguinte forma:  $y = \beta x + \alpha$

As hipóteses testadas são:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

O nível de significância adotado foi 5% ( $\alpha=0,05$ ).

Para rejeitar a hipótese nula, o p-valor deve ser menor que  $\alpha$ .

Para os dados originais temos a seguinte equação:  $y = 4,95x + 2,65$  e, portanto, a estimativa de  $\beta$  igual a 4,95.

As 10.000 aleatorizações foram feitas utilizando o *software* Excel e o p-valor encontrado foi 0,023, logo, rejeita-se a hipótese nula, portanto,  $\beta \neq 0$ , ou seja, existe tendência na série.

### 4. Conclusão

Como o p-valor encontrado no teste de aleatorização foi menor do que 0,05 (valor de  $\alpha$ ), então rejeitamos a hipótese nula, logo, existe tendência na série. Estudos adicionais serão feitos para confirmar os resultados obtidos.

## **Referências:**

- MURTEIRA, B. J. F. 1993. **Análise Exploratória de Dados – Estatística Descritiva**. Ed. McGRAW-HILL, 329p.
- MANLY, B. F. J. **Randomization, Bootstrap and Monte Carlo Methods in Biology**. 2 ed. Florida: 1997. Chapman &Hall/CRC.
- PRAZERES FILHO, VIOLA e LIMA 2009. **Uso do Teste de Aleatorização Para Verificar Existência de Correlação entre Duas Variáveis**.
- VIOLA, D. N. **Tese de Doutorado: Detecção e modelagem de Padrão especial em dados binários e de contagem**. USP: Escola Superior de Agricultura “Luiz de Queiroz”. Piracicaba, 2007.