



**SENAI CIMATEC**

**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM  
COMPUTACIONAL E TECNOLOGIA INDUSTRIAL  
Doutorado em Modelagem Computacional e Tecnologia Industrial**

**Tese de Doutorado**

**AUTOCORRELAÇÃO E CORRELAÇÃO  
CRUZADA: TEORIAS E APLICAÇÕES**

Apresentada por: Aloísio Machado da Silva Filho  
Orientador: Prof. Dr. Gilney Figueira Zebende

Setembro de 2014

Aloísio Machado da Silva Filho

# AUTOCORRELAÇÃO E CORRELAÇÃO CRUZADA: TEORIAS E APLICAÇÕES

Tese de Doutorado apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Doutorado em Modelagem Computacional e Tecnologia Industrial do SENAI CIMATEC, como requisito parcial para a obtenção do título de **Doutor em Modelagem Computacional e Tecnologia Industrial**.

Área de conhecimento: Interdisciplinar

Área de concentração: Sistemas complexos

Orientador: Prof. Dr. Gilney Figueira Zebende  
*SENAI CIMATEC*

Salvador-BA  
SENAI CIMATEC  
2014

---

## Nota sobre o estilo do PPGMCTI

---

Esta tese de doutorado foi elaborada considerando as normas de estilo (i.e. estéticas e estruturais) propostas aprovadas pelo colegiado do Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial e estão disponíveis em formato eletrônico (*download* na Página Web [http://ead.fieb.org.br/portal\\_faculdades/dissertacoes-e-teses-mcti.html](http://ead.fieb.org.br/portal_faculdades/dissertacoes-e-teses-mcti.html) ou com solicitação via e-mail à secretaria do programa) e em formato impresso para consulta.

Ressalta-se que o formato proposto considera diversos itens das normas da Associação Brasileira de Normas Técnicas (ABNT), entretanto opta-se, em alguns aspectos, seguir um estilo próprio elaborado e amadurecido pelos professores do programa de pós-graduação supracitado.

# SENAI CIMATEC

Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial

Doutorado em Modelagem Computacional e Tecnologia Industrial

A Banca Examinadora, constituída pelos professores abaixo listados, leu e recomendou a aprovação da Tese de Doutorado, intitulada “AUTOCORRELAÇÃO E CORRELAÇÃO CRUZADA: TEORIAS E APLICAÇÕES”, apresentada no dia 26 de Setembro de 2014, como requisito parcial para a obtenção do título de **Doutor em Modelagem Computacional e Tecnologia Industrial**.

Orientador:

---

Prof. Dr. Gilney Figueira Zebende  
SENAI CIMATEC

Membro de Banca Interno:

---

Prof. Dr. Marcelo Albano Moret Simões Gonçalves  
SENAI CIMATEC

Membro de Banca Interno:

---

Prof. Dr. Renelson Ribeiro Sampaio  
SENAI CIMATEC

Membro de Banca Externo:

---

Prof. Dr. Gilênio Borges Fernandes  
Universidade Federal da Bahia-UFBA

Membro de Banca Externo:

---

Prof. Dr. José Carlos Oliveira de Jesus  
Universidade Estadual de Feira de Santana-UEFS

Dedico este trabalho às minhas duas irmãs (Aloisia e Aloidilza) e à minha mãe Dilza e a  
querida Vitória Bispo.

---

## Agradecimentos

---

- Agradeço a minha mãe Dilza, as irmãs Aloisia e Aloidilza e a querida Vitória Bispo, pelo incentivo durante a pesquisa.

- Aos professores e colegas do Doutorado e especialmente **ao meu orientador Gilney Figueira Zebende** pelos ensinamentos durante a pesquisa e parceria em alguns trabalhos.

- Aos doutorandos Arleys Pereira Nunes de Castro e Marcus Fernandes da Silva e ao professor Pedro Alves Silva pelo apoio e parceria em alguns trabalhos.

- Aos professores e amigos: Aloísio Santos Nascimento Filho, Jair Sampaio Soares Junior, Magali Teresopolis Reis Amaral, Gilson de Almeida Dantas e Wanisson Silva Santana pelo incentivo durante a pesquisa.

- Ao professor Dr. Miguel Angel Rivera-Castro pelas contribuições no exame de qualificação do Doutorado.

- A Fundação de Amparo à Pesquisa do Estado da Bahia - FAPESB pela Bolsa de Doutorado.

- A Fundação Visconde de Cairu pela licença concedida para a conclusão do Doutorado e pelo espaço para estudo nos finais de semana.

- Ao CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio financeiro.

- Ao CEDEP- Centro de Documentação e Estatística Policial do Estado da Bahia e ao CEAT - Centro Especial de Auxílio ao Taxista pelos dados fornecidos.

- Aos líderes e pesquisadores do **Grupo de Pesquisa em Sistemas Complexos do SENAI CIMATEC** pelo apoio dado durante a pesquisa e pela oportunidade de fazer parte do grupo.

E a todos, que de maneira direta ou indiretamente, vêm contribuindo para o desenvolvimento desta pesquisa.

---

## Resumo

---

O volume significativo de dados e informações que são produzidos atualmente nas diversas áreas do conhecimento torna necessária a adoção de conceitos e métodos que possibilitem uma melhor compreensão dos fenômenos estudados. Nesse contexto, a estatística, enquanto área do conhecimento, tem desempenhado papel relevante. Assim, o objetivo geral desta pesquisa é analisar os métodos da mecânica estatística: DFA-*Detrended Fluctuation Analysis*, o DCCA - *Detrended Cross-Correlation Analysis* e o coeficiente de correlação cruzada  $\rho_{DCCA}$ . O método DFA pode ser aplicado para identificar e mensurar autocorrelação de longo alcance em séries temporais não-estacionárias via seu expoente de autocorrelação. O método DCCA é um método capaz de quantificar a correlação cruzada de longo alcance entre duas séries temporais enquanto lei de potência, em regime não estacionário e, além disso, é capaz de identificar componentes sazonais ou periodicidade. O denominado coeficiente de correlação cruzada  $\rho_{DCCA}$  é capaz de quantificar o nível de correlação cruzada baseado no DFA e no DCCA. A fim de atender ao objetivo geral desta pesquisa, foram definidos os seguintes objetivos específicos: estabelecer uma relação estatística entre o DFA, DCCA e o coeficiente de correlação cruzada  $\rho_{DCCA}$ , otimizar o coeficiente de correlação cruzada  $\rho_{DCCA}$  e propor uma nova metodologia de análise de séries temporais. Para consecução de tais objetivos, foram utilizadas séries temporais da criminalidade e violência de Salvador-BA, séries do mercado financeiro internacional e séries simuladas. Os principais resultados desta pesquisa foram: a implementação de uma relação estatística entre os métodos DFA e DCCA via diferenciação do coeficiente de correlação cruzada  $\rho_{DCCA}$ , reprodução do teste estatístico do  $\rho_{DCCA}$  para diferentes escalas de tempo de tamanho  $n$  e magnitudes de confiança e, por fim, a definição de três grupos distintos de correlação cruzada por meio da função de covariância sem tendência do método DCCA.

**Palavras chaves:** DFA, DCCA,  $\rho_{DCCA}$ , Séries temporais.

---

## Abstract

---

The significant amount of data and information that are currently produced in various areas of knowledge becomes necessary the adoption of concepts and methods that enable a better understanding of the phenomena studied. In this context, statistical, while knowledge of the area, has played an important role. Thus, the objective of this research is to analyze the methods of statistical mechanics: the DFA - Detrended Fluctuation Analysis-DFA, the DCCA - Detrended Cross-Correlation Analysis and the cross-correlation coefficient  $\rho_{DCCA}$ . The DFA method can be applied to identify and measure long-range autocorrelation in non-stationary time series via its exponent autocorrelation. The method DCCA is a method to quantify the long-range cross-correlation between two time series as a power law in non-steady state, and moreover, it is able to identify components or seasonal basis. The cross-correlation coefficient called  $\rho_{DCCA}$  is able to quantify the level of cross-correlation based on DFA and DCCA. In order to reach the overall objective of this research, the following specific objectives were: to establish a statistical relationship between DFA and DCCA cross-correlation coefficient  $\rho_{DCCA}$ , optimizing cross correlation coefficient  $\rho_{DCCA}$  and propose a new methodology time series analysis. To achieve these goals, time series of crime and violence in Salvador, Bahia, series of financial international market and simulated series were used. The main results of this research were: the implementation of a statistical relationship between DFA and DCCA methods via differentiation of the cross-correlation coefficient  $\rho_{DCCA}$ , the reproduction the statistical test  $\rho_{DCCA}$  for different time scales of size  $n$  and magnitudes of confidence and, finally, the definition of three groups for cross-correlation using the detrended covariance function of DCCA method.

**Keywords:** DFA, DCCA,  $\rho_{DCCA}$  and time series.



---

# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Fundamentação teórica</b>	<b>4</b>
2.1	Revisão de séries temporais . . . . .	4
2.1.1	Conceitos básicos . . . . .	4
2.1.2	Modelos de séries temporais . . . . .	16
2.1.3	Modelos de decomposição . . . . .	18
2.1.4	Coefficiente de correlação de Pearson . . . . .	20
2.2	Fractalidade e séries temporais . . . . .	25
2.2.1	Fractais . . . . .	25
2.2.2	Auto-similaridade exata e estatística . . . . .	30
2.2.3	Persistência, correlação de longo alcance e curto alcance . . . . .	30
2.3	Métodos de análise para séries temporais estacionárias . . . . .	32
2.3.1	Função de Autocorrelação - FAC . . . . .	32
2.3.2	Estatística de Husrt . . . . .	34
2.4	Métodos de análise para séries temporais não estacionárias . . . . .	38
2.4.1	Método DFA-Detrended Fluctuation Analysis . . . . .	38
2.4.2	Método DCCA-Detrended Cross-Correlation Analysis . . . . .	44
2.4.3	Coefficiente de correlação cruzada $\rho_{DCCA}$ . . . . .	50
<b>3</b>	<b>Resultados</b>	<b>56</b>
3.1	Estudo das correlações cruzadas em séries temporais auto afins de acidentes de táxis . . . . .	58
3.1.1	Objetivos e dados . . . . .	59
3.1.2	Resultados e conclusões . . . . .	62
3.2	Diferenciação do $\rho_{DCCA}$ : Abordagens práticas e teóricas . . . . .	68
3.2.1	Implementação teórica . . . . .	70
3.2.2	Dados e resultados . . . . .	71
3.2.3	Conclusões . . . . .	78
3.3	Autocorrelação e correlação cruzada nas séries temporais dos homicídios dolosos e homicídios tentados . . . . .	79
3.3.1	Dados e resultados . . . . .	80
3.3.2	Conclusões . . . . .	87
<b>4</b>	<b>Considerações finais</b>	<b>88</b>
4.1	Conclusões . . . . .	88
4.2	Sugestões para trabalhos futuros . . . . .	91
<b>A</b>	<b>Valores críticos do <math>\rho_{DCCA}</math></b>	<b>92</b>
<b>B</b>	<b>Pedido de registro de programa de computador ao INPI</b>	<b>94</b>
B.1	Programa que calcula o DFA, DCCA e $\rho_{DCCA}$ . . . . .	94
B.2	Programa que aplica o teste estatístico do $\rho_{DCCA}$ . . . . .	97
	<b>Referências</b>	<b>100</b>

---

## Lista de Tabelas

---

3.1	Matriz de correlações cruzadas entre as séries temporais de acidentes de táxis registrados pelo CEAT. A matriz superior representa a análise DCCA, enquanto que a matriz inferior representa o coeficiente de correlação clássico (ver Figura 3.5). Aqui nessa tabela temos os seguintes códigos: LR= correlação cruzada de longo alcance, SR= correlação cruzada de curto alcance e NC sem correlação entre as séries temporais. Fonte: (ZEBENDE; SILVA; MACHADO, 2011). . . . .	67
3.2	Expoente de autocorrelação $\alpha_{DFA}$ e expoente de correlação cruzada $\lambda_{DCCA}$ .	78
3.3	Estatística descritiva dos homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de janeiro de 2004 a dezembro de 2011.	80
3.4	Expoente de autocorrelação $\alpha_{DFA}$ dos homicídios dolosos (+), homicídios tentados (*) e expoente de correlação cruzada $\lambda_{DCCA}(\circ)$ . . . . .	84
A.1	Valores positivos críticos do coeficiente de correlação cruzada $\rho_{DCCA}$ . . . .	92
A.2	Valores positivos críticos do coeficiente de correlação cruzada $\rho_{DCCA}$ . . . .	92
A.3	Valores positivos críticos do coeficiente de correlação cruzada $\rho_{DCCA}$ . . . .	92
A.4	Valores positivos críticos do coeficiente de correlação cruzada $\rho_{DCCA}$ . . . .	93
A.5	Valores positivos críticos do coeficiente de correlação cruzada $\rho_{DCCA}$ . . . .	93
A.6	Valores positivos críticos do coeficiente de correlação cruzada $\rho_{DCCA}$ . . . .	93

---

## Lista de Figuras

---

2.1	Taxa anual por 100 mil habitantes dos homicídios dolosos registrados em Salvador-BA, 2003 a 2011. Fonte: Secretaria de Segurança Pública do Estado da Bahia. . . . .	6
2.2	Homicídios tentados registrados diariamente em Salvador-BA, 2004 a 2011. Fonte: Secretaria de Segurança Pública do Estado da Bahia. . . . .	7
2.3	Acidentes de trânsito com morte registrados mensalmente em Salvador-BA, 2003 a 2010. Fonte: Transalvador. . . . .	7
2.4	Série de veículos roubados e veículos furtados em Salvador-BA, 2003 a 2011. Fonte: Secretaria de Segurança Pública do Estado da Bahia. . . . .	8
2.5	Série não estacionária quanto ao nível e inclinação. Fonte:( <a href="#">MORETTIN; TOLOI, 2004</a> ). . . . .	10
2.6	Série temporal dos homicídios dolosos registrados mensalmente em Salvador-BA, 2003 - 2007. Fonte: Secretaria de Segurança Pública da Bahia. . . . .	12
2.7	Série temporal transformada com diferença de 4 meses na série original dos homicídios dolosos registrados mensalmente em Salvador-BA, 2003 - 2007. Fonte: Cálculo do autor. . . . .	12
2.8	Um processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: ( <a href="#">MORETTIN; TOLOI, 2004</a> ). . . . .	14
2.9	Um processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: ( <a href="#">MORETTIN; TOLOI, 2004</a> ). Nota: Adaptada pelo autor. . . . .	15
2.10	Um processo estocástico e série temporal. Fonte: ( <a href="#">SOUZA; CAMARGO, 2004</a> ). . . . .	16
2.11	Observações de uma ST com previsões de origem $t$ e horizonte $h$ . Fonte:Elaboração própria. . . . .	17
2.12	Homicídios dolosos registrados mensalmente em Salvador-BA de janeiro de 2003 a dezembro de 2011. Fonte: Secretaria de Segurança Pública da Bahia. Notas: 1 - A linha vertical representa o mês de maio no ano de 2006. 2 - O cálculo da taxa foi realizado com a população estimada disponibilizada pelo IBGE. . . . .	19
2.13	Demanda de passageiros transportados diariamente pelo sistema <i>ferry boat</i> de 5 de janeiro de 1996 a 31 dezembro 2003. Fonte:( <a href="#">NASCIMENTO, 2005</a> ). . . . .	19
2.14	Diagrama de dispersão. Fonte: ( <a href="#">TRIOLA, 1999</a> ). . . . .	20
2.15	Hipóteses estatísticas do parâmetro $\rho_{xy}$ . RR - Região de rejeição de $H_0$ e RNR Região de não rejeição de $H_0$ . $f_{dp}$ denota função de densidade de probabilidade e $-t_{\alpha/2;(n-2)}$ e $t_{\alpha/2;(n-2)}$ são os valores críticos da distribuição $t$ -Student. . . . .	23
2.16	Processo de formação do Conjunto de Cantor. Fonte:( <a href="#">MOREIRA, 1999</a> ). . . . .	28
2.17	Processo de formação do Conjunto de Koch. Fonte:( <a href="#">MOREIRA, 1999</a> ). . . . .	28
2.18	Triangulo Sierpinski. . . . .	29
2.19	Correlograma dos homicídios dolosos e roubos de veículos registrados diariamente em Salvador-BA, 2003 a 2008 ( $n = 72$ meses). . . . .	33
2.20	Modelo do reservatório de Hurst (1951) cálculo de $X(t, \tau)$ . Fonte: ( <a href="#">FEDER, 1988</a> ). . . . .	35
2.21	Modelo do reservatório de Hurst (1951) cálculo de $R(\tau)$ . Fonte: ( <a href="#">FEDER, 1988</a> ). . . . .	36

2.22	Série temporal dos acidentes de trânsito ( $u_i$ ) sem lesão corporal registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. . . . .	40
2.23	Série integrada $y(k)$ dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. Nota: Cálculo do autor. . . . .	40
2.24	Série integrada $y(k)$ dividida em (box) de tamanho 365 dias dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. Nota: Cálculo do autor. . . . .	41
2.25	Série integrada $y(k)$ dividida em (box) de tamanho 365 e as tendências locais $y_n(k)$ , dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010. . . . .	42
2.26	Comportamento do expoente $\alpha$ dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010. . . . .	43
2.27	Acidentes de trânsito sem lesão e com lesão registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. Nota: Cálculo do autor. . . . .	45
2.28	Divisão dos sinais integrados $R_K$ e $R'_K$ em $N - n$ boxes (com superposição) de tamanho $n$ , cada um contendo $n + 1$ valores. Fonte: (MACHADO, 2009). . . . .	46
2.29	Autocorrelação individual (DFA) e cruzada (DCCA) em valores absolutos de mudanças de preços (volatility) e os volumes negociados (volume) tanto pelos índices da Dow Jones (DJI) e da Nasdaq (Nasd), registrados diariamente, no período de julho de 1993 à novembro de 2003. Fonte: (PODOBNIK; STANLEY, 2008). . . . .	48
2.30	Autocorrelação e correlação cruzada: a) séries temporais volatilidade do (o) açúcar ( $F_{DFA}$ ), (●) açúcar $\times$ café ( $F_{DCCA}$ ) e (□) café ( $F_{DFA}$ ). b) séries temporais volatilidade do (o) açúcar ( $F_{DFA}$ ), (●) açúcar $\times$ algodão ( $F_{DCCA}$ ) e (□) algodão ( $F_{DFA}$ ). Fonte: (SIQUEIRA et al., 2010). . . . .	49
2.31	$\rho_{DCCA}$ entre a temperatura do ar x umidade relativa: a) corresponde aproximadamente a 40° de latitude (EUA), b) corresponde a cerca de 40° de latitude de outros países e c) outras latitudes. As linhas verticais denotam 90 e 365 dias respectivamente. Fonte: (VASSOLER; ZEBENDE, 2011) . . . . .	51
2.32	Hipóteses estatísticas do teste proposto por (PODOBNIK et al., 2011) para os valores obtidos pelo $\rho_{DCCA}$ . Aqui $f_{dp}$ denota função de densidade de probabilidade e VCMIn - valor crítico mínimo e VCMax - valor crítico máximo. . . . .	52
2.33	Simulação para o teste estatístico da equação 2.32 com 10.000 pares de séries iid com $N = 16.000$ pontos e o tamanho do <i>box</i> $n = 4$ . . . . .	53
2.34	Simulação para o teste estatístico da equação 2.32 com 10.000 pares de séries iid com $N = 16.000$ pontos e o tamanho do <i>box</i> $n = 256$ . . . . .	54
3.1	Séries temporais originais dos acidentes de táxi registrados pela CEAT:(a) (o) Idade do taxista (b)(●) Tempo da licença (c) (□) Ano de fabricação do veículo (d) (◇) Tempo de reparo (e) (◆) Custo de reparo do táxi. Fonte: (ZEBENDE; SILVA; MACHADO, 2011). . . . .	60

- 3.2 Histograma a) Idade do taxista, b) Tempo da Licença, c) Ano da fabricação, d) Tempo de reparo e e) Custo do reparo do táxi. As linhas contínuas em a), b) e c) correspondem ao um ajuste Gaussiano com R quadrado, média e erro padrão. Nos casos d) (tempo de reparo do taxi) e e) (custo de reparo do táxi), não ajustam a uma curva Gaussiana. f) representa o histograma para o número de acidentes por registro de táxi (rank de acidentes), e a linha contínua denota uma queda exponencial (somente uma curiosidade para ser visualizada). Fonte: (ZEBENDE; SILVA; MACHADO, 2011). . . . . 61
- 3.3 DFA aplicado em séries de acidente de taxi a) Idade do taxista  $\alpha_a$ , b) Tempo de licença  $\alpha_b$  c) Ano de fabricação do taxi  $\alpha_c$ , d) Tempo de reparo  $\alpha_d$  e) Custo do reparo do táxi  $\alpha_e$ . Para cada caso foi definido o coeficiente angular do ajuste linear (expoente  $\alpha$  com seu respectivo erro padrão). Fonte: (ZEBENDE; SILVA; MACHADO, 2011). . . . . 63
- 3.4 DCCA aplicado nas séries de acidentes de táxi: a) Correlação cruzada de longo alcance, b) Correlação positiva de curto alcance, c) Correlação negativa de curto alcance e d) Ausência de correlação cruzada entre as séries temporais. Fonte: (ZEBENDE; SILVA; MACHADO, 2011). . . . . 64
- 3.5 Ajuste linear (linha contínua) entre as variáveis objeto de estudo. Nota: Para cada ajuste linear foi definido o coeficiente de correlação de Pearson e o p-valor. Fonte: (ZEBENDE; SILVA; MACHADO, 2011). . . . . 65
- 3.6 (a) variância sem tendência,  $F_{DFA}$  e a covariância sem tendência  $F_{DCCA}$  em função de  $n$  (*Langs*) (Amplitude de tempo) para as duas séries temporais geradas pelo processo ARFIMA com  $\rho_1 = 0, 1(\circ)$ ,  $\rho_2 = 0, 4(\blacksquare)$  e  $W = 1, 0$ . Nestes processos ARFIMA ambas séries temporais,  $\{y_1\}$  e  $\{y_2\}$  compartilham o mesmo processo Gaussiano i.i.d  $\varepsilon_{1,i} = \varepsilon_{2,i}$ . Uma autocorrelação de longo alcance de lei de potência,  $F_{DFA}(n) \sim n^\alpha$  e uma correlação cruzada de longo alcance de lei de potência,  $F_{DCCA}(n) \sim n^\lambda$  são geradas. (b) O coeficiente de correlação cruzada  $\rho_{DCCA} \times n$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0, 99$ , erro padrão = 0,02 e  $p$ -valor  $< 0, 0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013). . . . . 73
- 3.7 a) variância sem tendência  $F_{DFA}(n)$  e a covariância sem tendência  $F_{DCCA}(n)$  como função de  $n$  (amplitude de tempo) para Dow Jones  $\{y_1\}$  e NASDAQ  $\{y_2\}$  diferenças de preços de fechamento de volume ajustado para dividendos e desdobramentos registrados diariamente entre 5 de fevereiro de 1971 e 9 de Dezembro de 2010. b) Representa a relação  $\log_{10} \times \log_{10}$  de  $\rho_{DCCA}(n) \times (n)$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0, 99$ , erro padrão = 0,02 e  $p$ -valor  $< 0, 0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013). . . . . 74
- 3.8 (a) variância sem tendência,  $F_{DFA}$  e a covariância sem tendência  $F_{DCCA}$  em função de  $n$  para as duas séries temporais  $\{y_1\}$  e  $\{y_2\}$  geradas pelo processo ARFIMA: neste processo temos  $\rho_1 = 0, 1$ ,  $\rho_2 = 0, 4$ ,  $\varepsilon_i \neq \varepsilon'_i$  e  $W = 0, 85$ . Uma autocorrelação de longo alcance,  $F_{DFA}(n) \sim n^\alpha$  e uma correlação cruzada de lei de potência,  $F_{DCCA}(n) \sim n^\lambda$  são gerados. (b) Coeficiente de correlação cruzada  $\rho_{DCCA}(n) \times n$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0, 98$ , erro padrão = 0,05 e  $p$ -valor  $< 0, 0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013). . . . . 75

3.9	a) variância sem tendência $F_{DFA}(n)$ e a covariância sem tendência $F_{DCCA}(n)$ como função de $n$ (amplitude de tempo) para diferenças sucessivas nos ajustes de preços de fechamento para os índices Dow Jones $\{y_1\}$ e NASDAQ $\{y_2\}$ registrados diariamente de 4 de janeiro de 2000 a 9 de dezembro de 2010. b) Coeficiente de correlação cruzada $\rho_{DCCA}(n) \times n$ . A linha contínua representa o ajuste linear, $dy/dx$ , com $R = 0,94$ , erro padrão = 0,11 e $p - \text{valor} < 0,0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013). . . . .	77
3.10	Séries temporais de vítimas/100.000 habitantes em Salvador-BA entre Janeiro 2004 e Dezembro de 2011 para: (a) homicídios dolosos; (b) homicídios tentados. A Linha contínua (colorida) representa a média móvel para $n = 90$ dias. Fonte: (MACHADO; SILVA; ZEBENDE, 2014). . . . .	81
3.11	Processo de análise do sinal para homicídios dolosos (esquerda) e homicídios tentados (direita). Neste caso (a) e (b) representam a amplitude do espectro para Transformada de Fourier e (c) e (d) a Função de autocorrelação (FAC). As linhas verticais na figura representam sete dias de análise. Fonte: (MACHADO; SILVA; ZEBENDE, 2014). . . . .	82
3.12	Variância sem tendência $F_{DFA}(n)$ para os homicídios dolosos (+), homicídios tentados (*), e covariância sem tendência $F_{DCCA}(n)$ entre os homicídios dolosos e homicídios tentados (o) em função de $n$ . As linhas verticais denotam a escala de tempo. Em cada intervalo definimos os valores dos expoentes. Fonte: (MACHADO; SILVA; ZEBENDE, 2014). . . . .	83
3.13	Coeficiente de correlação cruzada $\rho_{DCCA}$ para os homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de janeiro de 2004 a dezembro de 2011. Fonte: (MACHADO; SILVA; ZEBENDE, 2014). . . . .	85
3.14	Teste estatístico idealizado por (PODOBNIK et al., 2011) sobre o coeficiente de correlação cruzada $\rho_{DCCA}$ com 95 por cento de confiança, nos homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de janeiro de 2004 a dezembro de 2011. . . . .	86
B.1	Primeiro pedido de registro de programa de computador ao INPI - Página 1/3 . . . . .	94
B.2	Primeiro pedido de registro de programa de computador ao INPI - Página 2/3 . . . . .	95
B.3	Primeiro pedido de registro de programa de computador ao INPI - Página 3/3 . . . . .	96
B.4	Segundo pedido de registro de programa de computador ao INPI - Página 1/3 . . . . .	97
B.5	Segundo pedido de registro de programa de computador ao INPI - Página 2/3 . . . . .	98
B.6	Segundo pedido de registro de programa de computador ao INPI - Página 3/3 . . . . .	99

---

## Lista de Siglas

---

AST .....	Análise de Séries Temporais
ARMA .....	Modelo auto-regressivo e de médias móveis
ARIMA .....	Modelo auto-regressivo integrado e de médias móveis
ARFIMA ...	Modelo auto-regressivo fracionário integrado de médias móveis
CEAT .....	Centro Especial de Auxílio ao Taxista
CEP .....	Controle Estatístico de Processo
CNPq .....	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DCCA .....	Detrended Cross-Correlation Analysis
DFA .....	Detrended Fluctuation Analysis
EQM .....	Erro quadrático médio
FAC .....	Função de autocorrelação
FAPESB ....	Fundação de Amparo à Pesquisa do Estado da Bahia
IBGE .....	Instituto Brasileiro de Geografia e Estatística
INPI .....	Instituto Nacional da Propriedade Industrial
PPGMCTI ..	Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial
SENASP ....	Secretaria Nacional de Segurança Pública
ST .....	Série Temporal
v.a .....	Variável Aleatória
iid .....	Independente e Identicamente Distribuída

---

# Introdução

---

Atualmente, com o grande volume de dados e informações que são produzidos nas diversas áreas do conhecimento, é necessária a aplicação e desenvolvimento de métodos que possam auxiliar pesquisadores e profissionais para uma melhor compreensão dos fenômenos estudados. Nesse contexto, o conhecimento e a aplicação de métodos estatísticos assumem, nos dias atuais, um papel de extrema relevância, em função da amplitude e complexidade dos fenômenos estudados na área social, econômica, engenharias, entre outras. E também por proporcionar à investigação científica meios para consecução de seus objetivos.

Dentre as áreas de estudo da Estatística, neste projeto de doutorado, daremos ênfase à Análise de Séries Temporais que, segundo (MORETTIN; TOLOI, 2004), tem como objetivo descrever e analisar os comportamentos passados da série, visando a compreensão do comportamento e conseqüente previsão de movimentos futuros. Aqui lembramos que nas últimas duas décadas, têm crescido o estudo e o desenvolvimento de métodos destinados à análise de séries temporais em regime não estacionário como o método DFA - *Detrended Fluctuation Analysis* idealizado por (PENG et al., 1994), o DCCA - *Detrended Cross-Correlation Analysis* proposto por (PODOBNIK; STANLEY, 2008) e o coeficiente de correlação cruzada  $\rho_{DCCA}$  implementado por (ZEBENDE, 2011).

O presente projeto de Doutorado é continuação da pesquisa desenvolvida no Mestrado em Modelagem Computacional e Tecnologia Industrial do SENAI CIMATEC, onde foram realizado alguns trabalhos relacionados ao tema autocorrelação e correlação cruzada em séries temporais como o estudo de séries temporais do Sistema Ferry Boat (ZEBENDE; MACHADO, 2009a), teste de aleatoriedade em constantes matemáticas (ZEBENDE; MACHADO, 2009b), análise estatística dos indicadores de criminalidade de Salvador-BA (SOARES et al., 2008), (MACHADO; ZEBENDE; SOARES, 2008), entre outros.

O objetivo geral desta pesquisa é analisar os métodos DFA, DCCA e o coeficiente de correlação cruzada  $\rho_{DCCA}$ . Como fonte de estudo, escolhemos séries temporais de criminalidade e violência de Salvador-BA, séries do mercado financeiro internacional e séries simuladas.

E, para atender ao objetivo geral desta pesquisa, definimos os seguintes objetivos específicos:

- Estabelecer uma relação estatística entre o DFA, DCCA e  $\rho_{DCCA}$ ;



- Otimizar o coeficiente de correlação cruzada  $\rho_{DCCA}$ ;
- Propor uma nova metodologia de análise de séries temporais.

Para a execução desta pesquisa de doutorado, realizamos uma revisão de literatura (Capítulo 2) sobre séries temporais e métodos da mecânica estatística: DFA, DCCA e o coeficiente de correlação cruzada  $\rho_{DCCA}$  (ver seção 2.2), onde utilizamos como dados experimentais séries temporais simuladas e reais. Os dados relativos a violência e criminalidade foram obtidos na Secretaria de Segurança Pública do Estado da Bahia e os de acidentes de táxi em uma Cooperativa de Táxi denominada CEAT- Centro Especial de Auxílio ao Taxista e os financeiros no Yahoo finanças. Após a coleta desses dados, construímos uma ferramenta computacional capaz de aplicar os métodos estatísticos desta pesquisa, aplicamos os métodos e analisamos os resultados (ver Capítulo 3).

O trabalho e o material apresentado nos capítulos subsequentes desta tese de doutorado foram publicados (em eventos científicos ou periódicos) e estão listados abaixo:

■ Correlação de longo alcance em séries de furto de veículos e roubo de veículos de Salvador-BA: Uma aplicação utilizando o DFA e DCCA (*XII Escola de Modelos de Regressão*)(MACHADO; ZEBENDE; SILVA, 2011);

■ Study of cross-correlation in a self-affine time series of taxi accidents (**Physica A: Statistical Mechanics and its Applications**) (ZEBENDE; SILVA; MACHADO, 2011);

■ Autocorrelação e correlação cruzada em séries temporais de acidentes de trânsito de Salvador-BA (*XXX Encontro de Físicos do Norte e Nordeste*) (MACHADO et al., 2012);

■ Índice de Performance Econômica e Social dos Municípios Baianos: Conceito, Revisão Metodológica e Resultados (**Conjuntura e Planejamento**) (SANTOS et al., 2012);

■ Modelagem estatística da criminalidade: O estado da arte (**Conjuntura e Planejamento**) (MACHADO; ZEBENDE; ALVES, 2013);

■ Coeficiente de correlação cruzada múltiplo: Teorias e aplicações em séries temporais (*IV Encontro de Pesquisa Operacional no Nordeste*) (MACHADO; CASTRO; ZEBENDE, 2013);

■ DCCA cross-correlation coefficient differentiation: theoretical and practical approaches (**Physica A: Statistical Mechanics and its Applications**) (ZEBENDE; SILVA; MACHADO, 2013);

■ Análise da disfagia através da autocorrelação e correlação cruzada: Modelo computacional para a deglutição (*IV Encontro de Pesquisa Operacional no Nordeste*) (CASTRO; MACHADO; ZEBENDE, 2013);

■ Autocorrelation and cross-correlation in time series of homicide and attempted homicide (**Physica A: Statistical Mechanics and its Applications**) (MACHADO; SILVA; ZEBENDE, 2014);

■ Quantificando a influência do mercado de câmbio nos preços do milho e da soja no município de Barreiras (**Conjuntura e Planejamento**) (FERNANDES; CASTRO; MACHADO, 2014);

■ Coeficiente de correlação cruzada: Uma aplicação nas séries temporais dos homicídios dolosos e homicídios tentados de Salvador-BA (*XXI Simpósio Nacional de Probabilidade e Estatística*) (MACHADO et al., 2014b);

■ Modelagem do teste estatístico do coeficiente de correlação cruzada sem tendência  $\rho_{DCCA}$ : Teoria e aplicação (*IV Workshop de Pesquisa, Tecnologia e Inovação - PTI do SENAI CIMATEC*) (MACHADO et al., 2014a).

Além dos trabalhos supracitados modelamos dois programas que foram registrados com co-autoria do SENAI CIMATEC no Instituto da Propriedade Industrial - INPI com seus respectivos registros em apêndice (ver Apêndices B.1 e B.2) e nomes registrados abaixo:

■  $\rho_{DCCAmult}$  (ZEBENDE; CASTRO; MACHADO, 2013);

■  $\rho_{DCCAsort}$  teste estatístico (ZEBENDE et al., 2014).

As demais fases desta tese de doutorado estão estruturadas em três partes: na primeira é realizada uma revisão de literatura (Capítulo 2) sobre os conceitos e procedimentos de cálculos dos métodos estatísticos abordados na pesquisa, na segunda fase são apresentados os resultados (Capítulo 3), por fim, as conclusões e trabalhos futuros (Capítulo 4).

---

## Fundamentação teórica

---

O presente capítulo versa sobre a revisão de literatura que deu suporte à execução desta pesquisa de doutorado. Na seção 2.1 abordamos os conceitos básicos e procedimentos da análise de séries temporais, em seguida na seção 2.2 abordamos sobre fractalidade e séries temporais. Nas seções 2.3 e 2.4 discutimos sobre métodos que avaliam a memória de séries temporais em dois grupos distintos, respectivamente: métodos de análise para séries temporais estacionárias e métodos de análise para séries temporais não estacionárias.

### 2.1 Revisão de séries temporais

#### 2.1.1 Conceitos básicos

Segundo (FONSECA; MARTINS; TOLEDO, 1985) as investigações empíricas dependem, em grande parte, de dados arranjados em forma cronológica e a análise de séries temporais (AST) tem como objetivo descrever e analisar o comportamento passado da série, objetivando à compreensão do comportamento da série e a consequente previsão de movimentos futuros. A previsão de uma série temporal é simplesmente o estabelecimento dos valores futuros da série (SOUZA; CAMARGO, 2004). De acordo com os mesmos autores uma previsão é uma estimativa quantitativa (ou conjunto de estimativas) acerca da verossimilhança de eventos futuros, baseados na informação atual e passada.

A análise das séries temporais é uma área da estatística que vem sendo aplicada em diversas áreas do conhecimento como, por exemplo:

- Economia (LAPA, 1998), (MARTIN, 2000) e (LATIF, 2008);
- Biofísica (GALHARDO, 2010);
- Astrofísica (TEXEIRA, 2012);
- Geologia (CASTRO, 1995);
- Ciências sociais (MACHADO, 2009);
- Controle de processos (BALESTRASSI, 2000);

- Em redes neurais (PEREIRA; RODRIGUES, 1998);
- Engenharias (BALLINI, 2000), dentre outras (BARBOSA, 2007).

Para (MORETTIN; TOLOI, 2004) série temporal é um conjunto de observações ordenadas no tempo. (BOX; JENKINS, 1976) definem uma série temporal como uma sequência de observações tomadas sequencialmente no tempo. (EHLERS, 2007) relata que a característica mais relevante desse tipo de dados é que as observações vizinhas são dependentes e geralmente estamos interessados em analisar tal dependência. E segundo o mesmo autor enquanto em modelos de regressão, por exemplo, a ordem das observações não importa para a análise em séries temporais a ordem dos dados é um fator relevante. Seja  $Z$  a variável observada e  $t$  o tempo. Uma série temporal pode ser definida pelos valores  $Z_1, Z_2, Z_3 \dots Z_n$ , nos tempos  $t_1, t_2, t_3 \dots t_n$ . Desta forma  $Z$  é uma função de  $t$  denotada por:

$$Z = f(t) \tag{2.1}$$

São exemplos de séries temporais:

- (i) Valores diários das temperaturas médias de Salvador-BA;
- (ii) Taxa anual dos homicídios dolosos por 100 mil habitantes de Salvador-BA (Figura 2.1);
- (iii) Número de vítimas de homicídio tentado registrado diariamente em Salvador-BA (Figura 2.2);
- (iv) Número mensal de vítimas fatais de acidente de trânsito em Salvador-BA (Figura 2.3);
- (v) Índice diário de uma Bolsa de valores;
- (vi) Produção mensal de soja no município de Barreiras na Bahia em toneladas;
- (vii) Número de casos de dengue por ano no estado da Bahia;
- (viii) Faturamento mensal de uma empresa em determinado exercício;
- (ix) Número de veículos furtados e veículos roubados mensalmente em Salvador-BA (Figura 2.4);

- (x) Frota anual de veículos de Salvador-BA;
- (xi) Registros de marés no porto de Santos.

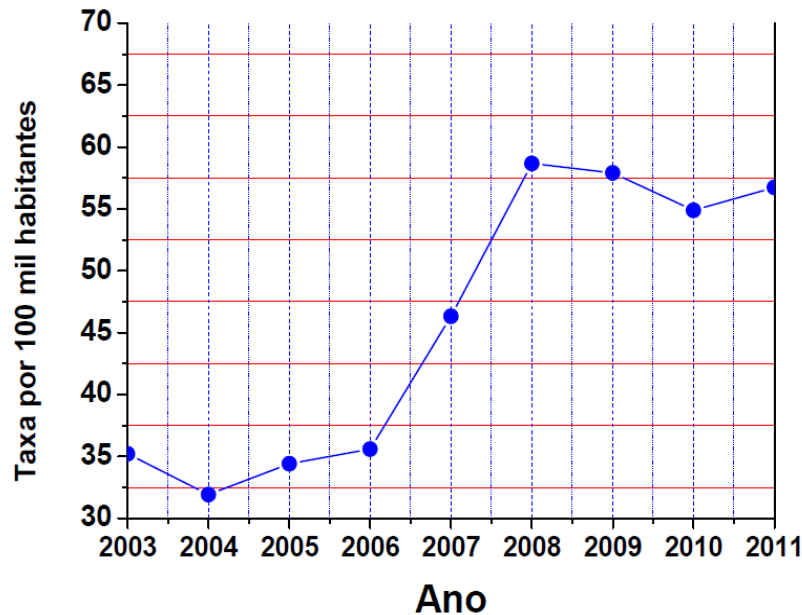


Figura 2.1: Taxa anual por 100 mil habitantes dos homicídios dolosos registrados em Salvador-BA, 2003 a 2011. Fonte: Secretaria de Segurança Pública do Estado da Bahia.

Nos exemplos (i) - (x) são exemplos de séries temporais *discretas*, enquanto (xi) é um caso de uma série *contínua*. Frequentemente, uma série temporal discreta é obtida por meio da amostragem de uma série temporal contínua em intervalos de tempos iguais,  $\Delta t$ . Desta maneira, para analisar a série (xi) será preciso amostrá-la, por exemplo, em intervalos de tempo de uma hora, transformando a série contínua, objeto de estudo no intervalo  $[0, T]$  em uma série discreta de  $N$  pontos, em que  $N = \frac{T}{\Delta t}$ . Já nas séries (i) a (x) o valor da série num dado instante é obtido agregando-se valores em intervalos de tempos iguais. É importante relatar que os termos utilizados anteriormente não se referem à variável observada, esta pode assumir valores discretos ou contínuos (MORETTIN; TOLOI, 2004) e (EHLERS, 2007).

Um série temporal pode ter também as seguintes classificações:

□ Determinística: quando uma função matemática pode ser utilizada para estabelecer exatamente os valores futuros da série;

□ Estocástica: quando os valores futuros da série somente podem ser estabelecidos em termos probabilísticos, pois o modelo compõe-se também de um termo aleatório.

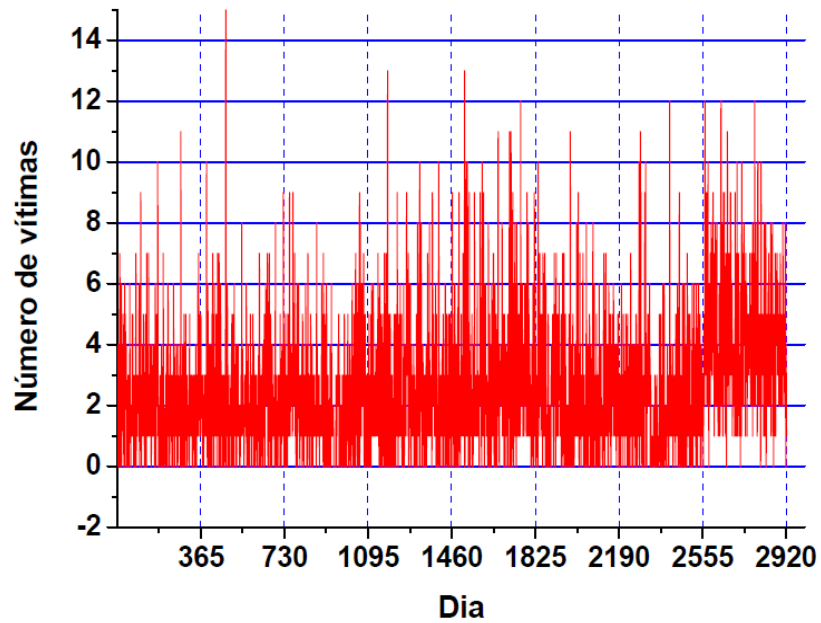


Figura 2.2: Homicídios tentados registrados diariamente em Salvador-BA, 2004 a 2011. Fonte: Secretaria de Segurança Pública do Estado da Bahia.

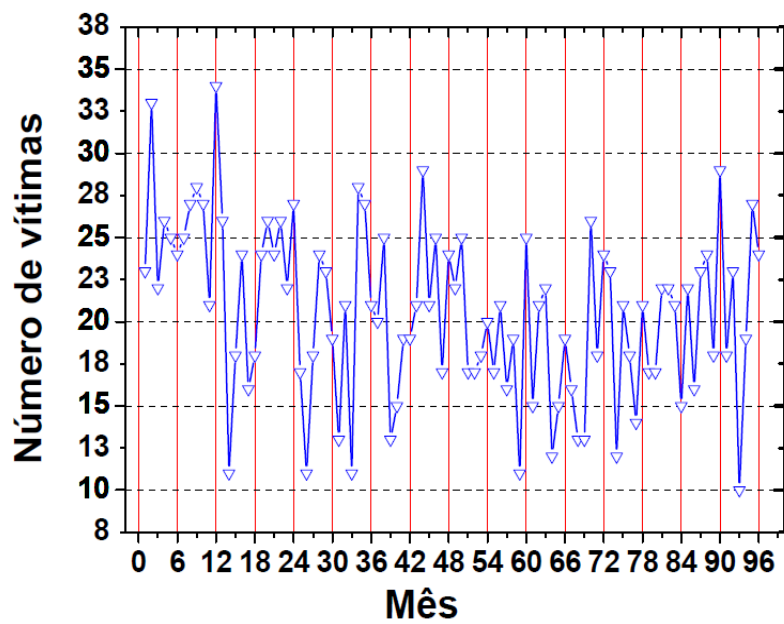


Figura 2.3: Acidentes de trânsito com morte registrados mensalmente em Salvador-BA, 2003 a 2010. Fonte: Transalvador.

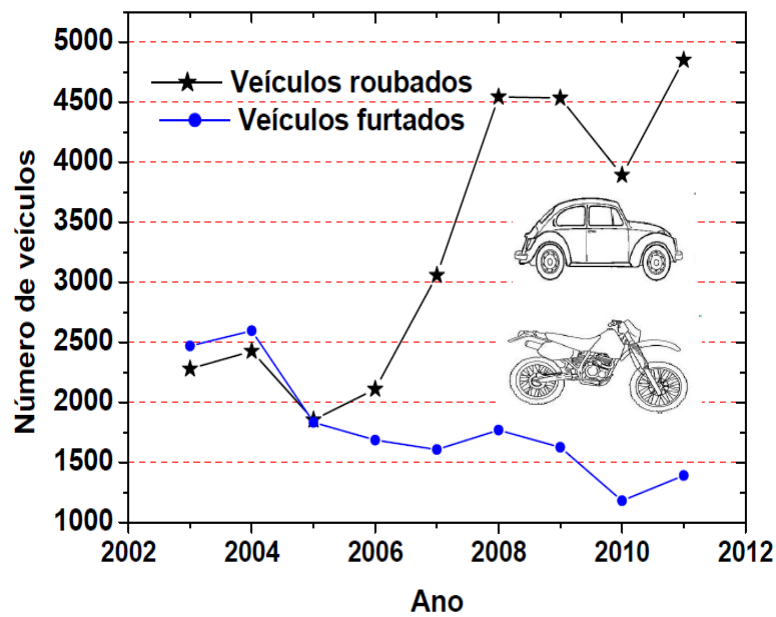


Figura 2.4: Série de veículos roubados e veículos furtados em Salvador-BA, 2003 a 2011. Fonte: Secretaria de Segurança Pública do Estado da Bahia.

Os dados provenientes de uma série temporal possuem algumas características particulares, por exemplo (EHLERS, 2007):

- ◇ Observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas;
- ◇ É necessário levar em consideração a ordem das observações;
- ◇ Fatores complicadores como presença de tendências e variação sazonal ou cíclica podem ser difíceis de se estimar ou se remover;
- ◇ A seleção de modelos pode ser complicada, e as ferramentas podem ser de difícil interpretação;
- ◇ Observações perdidas e dados discrepantes são fatores complicadores, devido à natureza sequencial.

Com base no que foi descrito até aqui é possível afirmar que a Análise de Séries Temporais (AST) é o procedimento pelo qual são identificados os segregados e fatores relacionados com o tempo, que influenciam os valores observados na série. Uma vez identificados, podem ser tomados como base, para o auxílio da interpretação e da projeção de valores da série temporal. De acordo com (MORETTIN; TOLOI, 2004) e (SOUZA; CAMARGO, 2004) existem basicamente dois enfoques utilizados nos procedimentos de análise de séries temporais, ambos com objetivos determinados. No primeiro enfoque, a análise é realizada no domínio temporal. Neste caso, é considerada a evolução temporal do processo, e os modelos propostos são modelos paramétricos (com um número finito de parâmetros), como tal, é possível citar o modelo ARFIMA. Já no segundo enfoque, a análise é concebida no domínio de frequência, e os modelos propostos são modelos não paramétricos, pode-se citar a análise espectral como exemplo.

De posse de uma série temporal  $Z(t_1), Z(t_2), Z(t_3), \dots, Z(t_n)$ , observada nos instantes  $t_1, t_2, t_3, \dots, t_n$  é possível estarmos interessados em:

Investigar o mecanismo gerador da série; por exemplo, analisando uma série de acidentes de trânsito, é possível estarmos interessados em saber como esses dados foram concebidos;

Fazer previsões de valores futuros da série de curto, médio e longo prazo;

Descrever apenas o comportamento da série; neste caso, a utilização de medidas estatísticas descritivas, a construção de gráficos, a verificação de tendências, ciclos e variações sazonais, etc podem ser ferramentas relevantes;

Procurar periodicidade relevantes nos dados; neste caso, a análise espectral pode ser de grande utilidade;

Controlar processos de qualidade.

Até agora nos referimos ao parâmetro  $t$  como tempo, todavia a série  $Z(t)$  pode ser função de algum parâmetro, como espaço, profundidade ou volume.



Um aspecto proeminente no estudo de séries temporais, segundo (ANTENEODO, 2004), é identificar se a série cronológica é estacionária, ou seja, se ela se desenvolve aleatoriamente no tempo em torno de uma média constante, caracterizando assim algum padrão de equilíbrio estável. Porém, a maioria das séries que modelamos apresentam alguma característica de não estacionariedade. Como tal, as séries econômicas e financeiras apresentam geralmente tendências, sendo o caso mais comum a série que flutua ao redor de uma reta, com inclinação positiva ou negativa. É possível também ocorrer uma forma de não estacionariedade explosiva, como o crescimento de uma colônia de bactérias. Uma série temporal pode apresentar característica de estacionariedade durante um período longo ou apenas em períodos muito curtos, alternando de nível e inclinação (Figura 2.5). Existem séries que oscilam ao redor de um nível médio durante algum tempo e em seguida saltam para outro nível temporário que é um caso típico de séries econômicas. E, segundo (KANTELHARDT, 2008), em sentido mais geral, mudanças na dinâmica do sistema também representam não estacionariedade da série temporal.

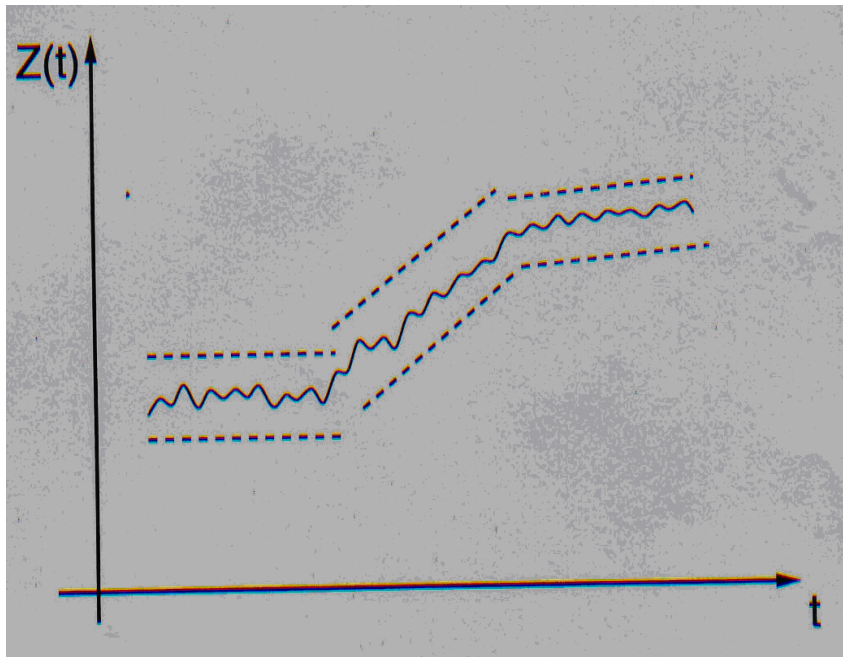


Figura 2.5: Série não estacionária quanto ao nível e inclinação. Fonte:(MORETTIN; TOLOI, 2004).

O universo de modelos aplicados em séries cronológicas tem como fundamento o conceito de estacionariedade. Nestes casos, em sua maioria, os modelos considerados são lineares estacionários fracos, como modelos autoregressivos (AR), de médias móveis (MA) e misto (ARMA). Existem também modelos que se destinam a descrever estatisticamente séries estacionárias e não estacionárias, por exemplo, o modelo ARIMA (auto regressivos integrados médias móveis) com a condição de que não apresentem comportamento explosivo (MORETTIN; TOLOI, 2004).

Como grande parte dos modelos estatísticos de AST têm a premissa das séries objeto de estudo serem estacionárias, é necessário transformar os dados originais, se estes não se comportarem como uma série estacionária. O procedimento de transformação mais comum é estabelecer diferenças sucessivas da série até obter a desejada estacionariedade (MORETTIN; TOLOI, 2004). Neste caso, a primeira diferença de  $Z(t)$  é definida pela seguinte expressão:

$$\Delta Z(t) = Z(t) - Z(t - 1), \quad (2.2)$$

a segunda diferença

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t - 1)] \quad (2.3)$$

de maneira geral, a  $n$  - ésima diferença de  $Z(t)$  é

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)] \quad (2.4)$$

Há, basicamente, dois motivos para se transformar os dados originais de uma série temporal: estabilizar a variância e tornar o efeito sazonal aditivo. É frequente, em séries econômicas, a existência de tendências, podendo ocorrer um acréscimo de variância da ST à medida que o tempo passa. Nesta situação, uma transformação logarítmica pode ser conveniente (MORETTIN; TOLOI, 2004) e (VEIGA; SAFADI, 1999). Um procedimento frequente em séries econômicas é considerar:

$$\Delta \log Z_t = \log Z_t - \log Z_{t-1} \quad (2.5)$$

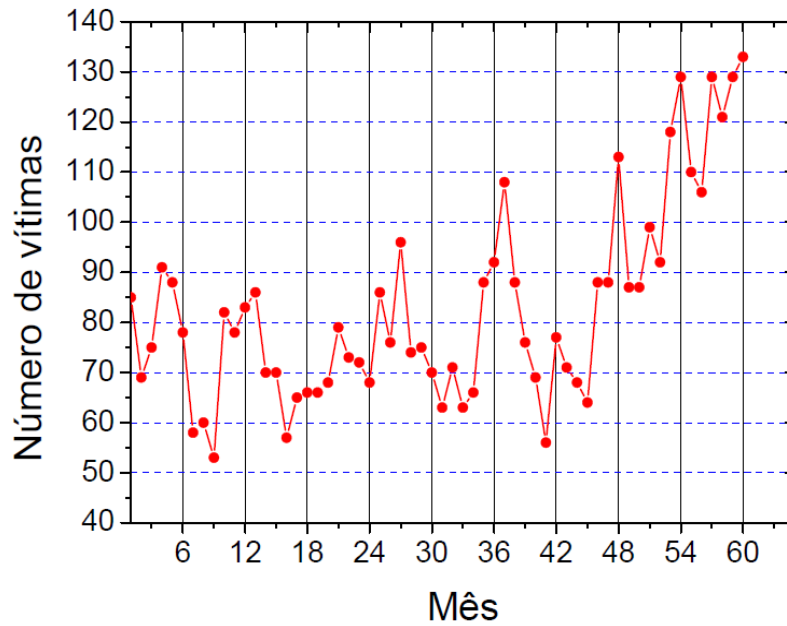


Figura 2.6: Série temporal dos homicídios dolosos registrados mensalmente em Salvador-BA, 2003 - 2007. Fonte: Secretaria de Segurança Pública da Bahia.

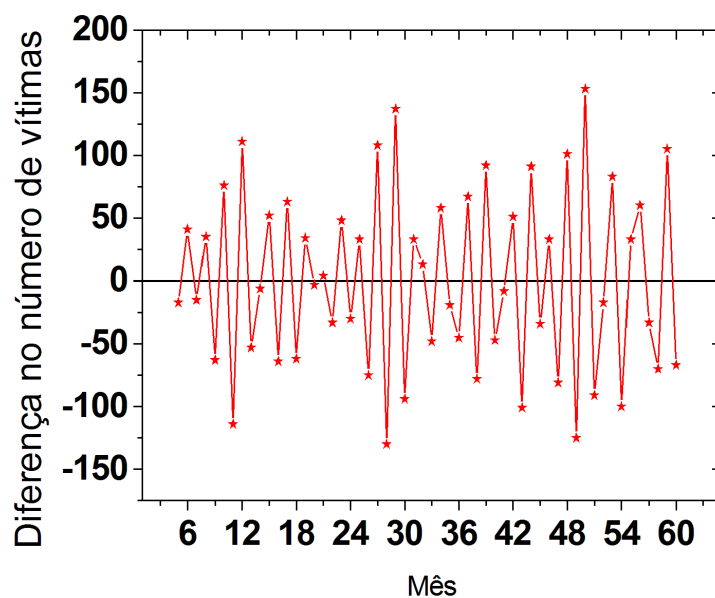


Figura 2.7: Série temporal transformada com diferença de 4 meses na série original dos homicídios dolosos registrados mensalmente em Salvador-BA, 2003 - 2007. Fonte: Cálculo do autor.

Um outro motivo de realizar transformações nos dados de séries temporais, segundo (MORETTIN; TOLOI, 2004), é obter uma série mais próxima possível de uma distribuição simétrica aproximadamente normal, ou seja,  $\mu = 0$  (média igual a 0) e  $\sigma^2 = 1$  (variância igual a 1). A Figura 2.6 representa a série original não estacionário dos homicídios dolosos registrados mensalmente em Salvador-BA de 2003 a 2007 e a (Figura 2.7) apresenta a série temporal estacionária com diferença de 4 meses na série original dos homicídios dolosos. Segundo (EHLERS, 2007), a diferenciação pode ser também uma ferramenta exploratória útil. Isso porque observações discrepantes podem ter um efeito relevante na série diferenciada e uma representação gráfica geralmente é suficiente para identificar tais discrepâncias.

Considerando  $Z_1, Z_2, \dots, Z_n$  como variáveis aleatórias, tendo uma distribuição conjunta  $f(z_1, \dots, z_n)$  e distribuições marginais  $f(z_1), (z_n)$ , a série será estacionária se, para qualquer momento de tempo  $t$  e  $n$ , tivermos  $f(z_t) = f(Z_{t+n})$ ,  $n = \pm 1, \pm 2, \dots$ , isto é, as distribuições de  $Z_1, Z_2, \dots, Z_n$  são invariantes por translações ao longo do tempo. Neste caso  $Z_t$  e  $Z_{t+n}$  terão os mesmos parâmetros, como tal,

$$E(Z_t) = E(Z_{t+n}), \quad (2.6)$$

e

$$Var(Z_t) = Var(Z_{t+n}), \quad (2.7)$$

onde  $E(Z_t)$  é o valor esperado de  $Z$  e  $Var(Z_t)$  é a variância de  $Z$ . A média e a variância de uma ST estacionária são constantes independentemente do instante de tempo  $t$ , e pode-se representar pelas seguintes expressões:

$$E(Z_t) = \mu, \forall t \quad (2.8)$$

e

$$Var(Z_t) = \sigma^2, \forall t \quad (2.9)$$

Como já foi dito, algumas séries temporais podem ser classificadas como estocásticas, isto é, séries controladas por leis probabilísticas. Se denotarmos  $T$  como um conjunto arbitrário. Um processo estocástico é uma família  $Z = \{Z(t), t \in T\}$ , tal que, para  $t \in T$ ,  $Z(t)$  é uma variável aleatória (PEREIRA, 1984). Com essas condições um processo estocástico pode ser considerado uma família de variáveis aleatórias (v.a), e é possível supor que elas sejam definidas no espaço amostral de probabilidade  $\Omega$ . Na realidade  $Z(t)$  é uma função de dois argumentos,  $Z(t, \omega), t \in T, \omega \in \Omega$ .

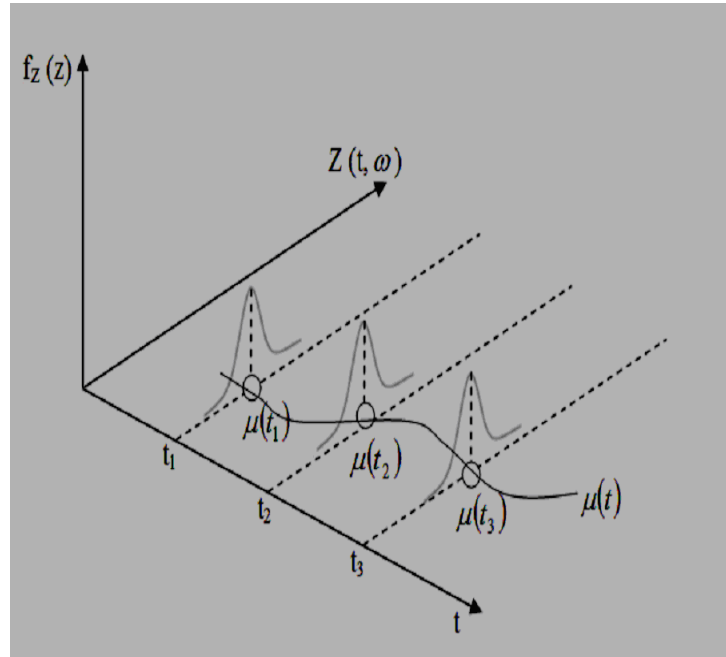


Figura 2.8: Um processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: (MORETTIN; TOLOI, 2004).

A partir da (Figura 2.8) é possível constatar que para cada  $t \in T$ , teremos uma v.a  $Z(t, \omega)$ , como uma distribuição de probabilidade, ou seja é possível que a função densidade de probabilidade de  $f_z(Z)$  no momento  $t_1$  difira da  $f_z(Z)$  no momento  $t_2$ . Se denotarmos as possíveis realizações de  $Z(t, \omega)$  por  $Z^1(t), Z^2(t), \dots, Z^n(t)$ , poderemos denominar o conjunto de todas as possíveis trajetórias como amostra. E, além disso, podemos observar que cada realização  $Z^j(t)$  é uma função do tempo  $t$  não aleatória e, para cada  $t$  fixo,  $Z^j(t)$  é um número real.

De acordo com (MORETTIN; TOLOI, 2004), uma forma de analisar uma distribuição de probabilidade  $Z(t \in \omega)$  para um  $t$  fixo é levar em consideração durante a análise a quantidade de trajetórias que passam por uma janela de amplitude  $\Delta t$  de modo que tal quantidade de trajetórias será relacionada à  $f_z(Z)\Delta t$  (Figura 2.9).

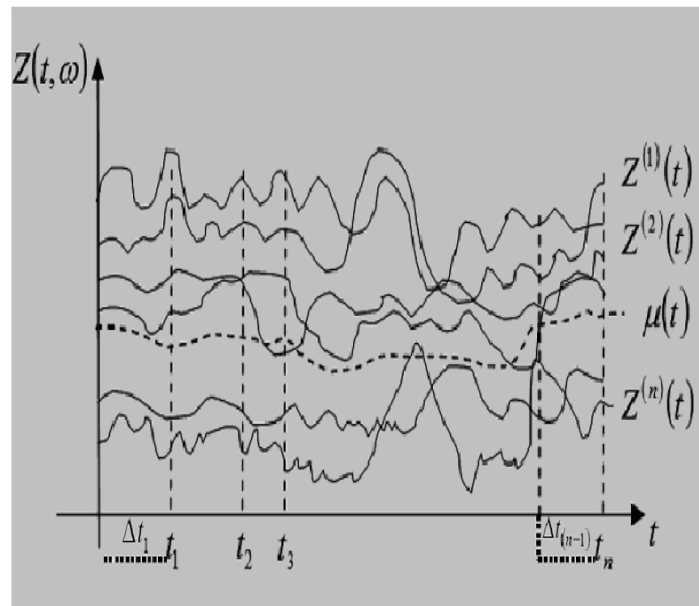


Figura 2.9: Um processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: (MORETTIN; TOLOI, 2004). Nota: Adaptada pelo autor.

Teoricamente um processo estocástico está estatisticamente determinado quando são conhecidas suas funções de distribuição até a  $N - \text{ésima}$  ordem. Como exemplo, dado um processo estocástico  $Z = \{Z(t), t \in T\}$ , o mesmo estará especificado, se para todo  $n > 1$  conhecermos as distribuições finito-dimensionais. Isto significa que, para  $n = 1$ , conhecemos as distribuições unidimensionais da v.a  $Z(t_1), t \in T$ , para  $n = 2$ , conhecemos as distribuições bidimensionais da v.a  $(Z(t_1), Z(t_2)), t_1, t_2 \in T$ , e assim por diante (PEREIRA, 1984) (MORETTIN; TOLOI, 2004). Mas na prática, segundo (SOUZA; CAMARGO, 2004), ocorrem duas situações problemáticas: não se conhece todas as funções de distribuição até a  $N - \text{ésima}$  ordem e, frequentemente, tem-se somente uma realização do processo estocástico em questão a partir da qual se deseja inferir todas as características do mecanismo gerador da série. Diante dessa realidade, é necessário realizar uma inferência do mecanismo gerador da série como na figura 2.10, que denota a sequência e o objeto do estudo. Nesta figura 2.10, retira-se de um processo estocástico uma amostra finita de observações de uma série temporal e, por meio do estudo dessa amostra (AST), identifica-se um modelo cujo objetivo é inferir sobre o comportamento da realidade (GUJARATI, 2000).

Uma importante classe de modelos estocásticos corresponde ao processo estacionário, isto é, aquele que apresenta média constante com equilíbrio em seu desenvolvimento.

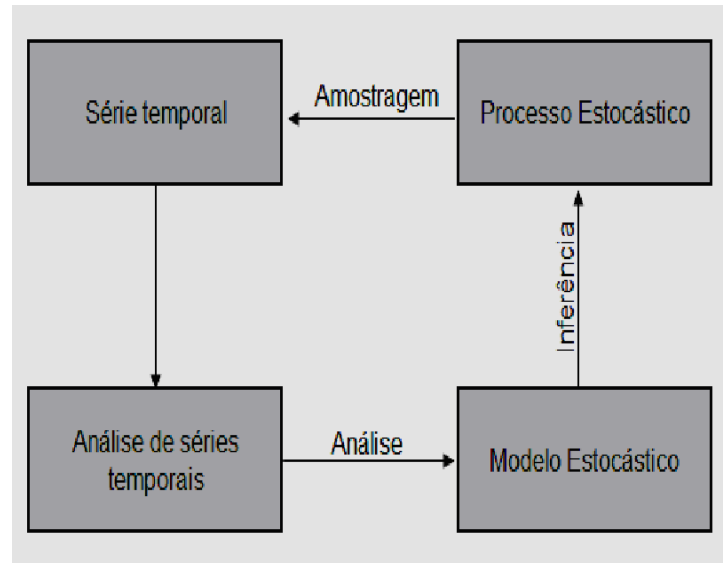


Figura 2.10: Um processo estocástico e série temporal. Fonte: (SOUZA; CAMARGO, 2004).

### 2.1.2 Modelos de séries temporais

Uma das significativas transformações ocorridas na metade dos anos 90, foi a crescente tendência no desenvolvimento de se obter modelos matemáticos a partir de dados observados e não mais exclusivamente com base nas equações que descrevem a física do processo (AGUIRRE, 2000). De acordo com o mesmo autor uma das possíveis razões dessa mudança foi o acesso a computadores com baixo custo e bom desempenho. A expressão modelo pode ser denotada como uma representação simplificada da realidade, estruturada de tal forma que possibilite a compreensão do funcionamento total ou parcial de uma dada realidade ou fenômeno (MATOS, 2000). Ao especificar uma determinada ST por um modelo, é necessário especificar uma função-perda como, por exemplo, o erro quadrático médio denotado pela seguinte expressão:  $EQM = [Z(t+h) - \hat{Z}_t(h)]^2$ . Para a partir daí obter um procedimento de previsão. Dispondo de observações de uma série temporal até o momento  $t$  e necessitando prever o valor da série no instante  $t+h$  (Figura 2.11), nesta situação  $t$  denota a origem e  $\hat{Z}_t(h)$  a previsão de  $Z(t+h)$ , de origem  $t$  e horizonte  $h$ .

Diante de determinado modelo de ST que permita descrever uma ST até o momento  $t$  e levando em consideração que o objetivo é minimizar o EQM, obteremos uma expressão para  $\hat{Z}_t(h)$ .

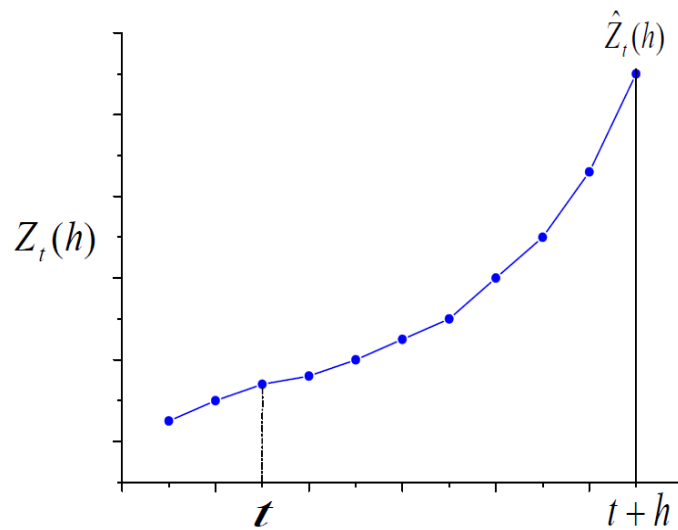


Figura 2.11: Observações de uma ST com previsões de origem  $t$  e horizonte  $h$ . Fonte:Elaboração própria.

A previsão de uma série temporal pode ser classificada como de curto, médio e longo prazo, isso dependendo do valor associado ao horizonte máximo de previsão. Porém, qualquer que seja o horizonte de previsão (curto, médio ou longo prazo), não existe um consenso entre especialistas sobre o horizonte máximo de previsão. E frequentemente os modelos de ST estão fundamentados na suposição de que observações passadas contém propriedades sobre o padrão de comportamento da ST. Neste caso, o objetivo dos métodos é estabelecer diferença entre o padrão de qualquer ruído que possa estar presente nas observações para daí então usar esse padrão para prever valores ou movimentos futuros.

Os modelos de séries temporais são classificados com base no número de parâmetros envolvidos que são: modelos paramétricos (com número de parâmetros finito) e não-paramétricos (que envolve um número infinito de parâmetros). No rol dos modelos paramétricos, cuja análise é realizada no domínio do tempo, os mais comumente utilizados são os modelos de erro (ou de regressão), os modelos auto-regressivos e de médias móveis (ARMA), os modelos auto-regressivos integrados e de médias móveis (ARIMA), modelos de memória longa (ARFIMA)(ver Seção 3.2.2), os modelos estruturais e modelos não-lineares. Os modelos não-paramétricos frequentemente utilizados são a função de autocovariância (autocorrelação) e a transformada de Fourier.

A modelagem de apenas uma série temporal denomina-se análise univariada, enquanto que análise de mais de uma série distinta trata-se de uma análise multivariada (MADDALA, 2003).



Segundo (MORETTIN; TOLOI, 2004) a construção e seleção de modelos de séries temporais dependem de vários fatores como, por exemplo, o comportamento da variável objeto de estudo, ou o conhecimento *a priori* que temos de sua natureza, do objetivo do estudo e também do acesso a computadores com *software* adequados.

### 2.1.3 Modelos de decomposição

Um modelo clássico relativo a séries temporais supõe que a série  $Z_t$  pode ser decomposta em tendência ( $T_t$ ), ciclo ( $C_t$ ), sazonalidade ( $S_t$ ) e termo errático ( $E_t$ ). Uma ST ( $Z_t$ ) pode resultar na combinação das quatro componentes listadas anteriormente ou pelo subconjunto delas, na forma aditiva teremos  $Z_t = T_t + C_t + S_t + E_t$ , já na multiplicativa denota-se por  $Z_t = T_t \times C_t \times S_t \times E_t$ , em que  $t = 1, 2, 3, \dots, n$ . O modelo aditivo é adequado, por exemplo, quando  $S_t$  não depende das outras componentes, como  $T_t$ . Porém, se as amplitudes sazonais variam ou dependem da tendência ( $T_t$ ) o modelo mais adequado é o multiplicativo.

Uma série temporal pode estar relacionada a uma mudança no nível da série em longo prazo. Em outras palavras, a tendência pode refletir a redução, o acréscimo ou a estabilidade em torno de um valor médio da ST em longo prazo (Figura 2.12). A sazonalidade é caracterizada por padrões cíclicos de flutuação que se repetem em intervalos relativamente constantes de tempo (Figura 2.13). A variação cíclica é identificada quando a ST apresenta variações ascendentes e descendentes, porém em intervalos não regulares de tempo. Caso a variação de uma série temporal não possa ser relacionada com uma das características supracitada podemos atribuir essa variação ao ruído aleatório no processo gerador da série. Para uma ST ruído é uma variação aleatória do sinal ao redor do seu valor médio. Segundo (VEIGA; SAFADI, 1999), a suposição usual é que essa variação resulte em uma série puramente aleatória ou ruído branco, com média zero ( $\mu = 0$ ) e variância constante.

A figura 2.12 representa uma ST com uma tendência de crescimento a partir de maio de 2006 nas ocorrências mensais das taxas de homicídio dolosos registrados em Salvador-BA. Já na figura 2.13 é possível constatar um comportamento sazonal, uma vez que existem movimentos semelhantes a cada 365 dias na demanda de veículos do sistema *ferry boat*.

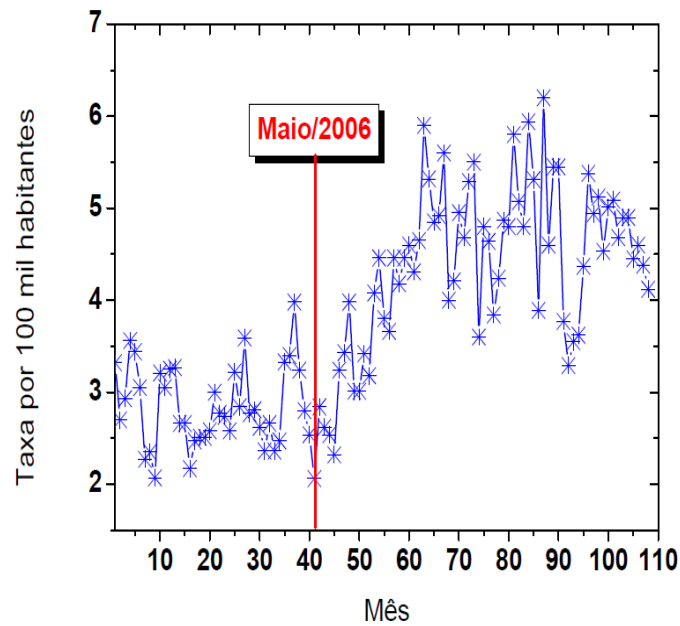


Figura 2.12: Homicídios dolosos registrados mensalmente em Salvador-BA de janeiro de 2003 a dezembro de 2011. Fonte: Secretaria de Segurança Pública da Bahia. Notas: 1 - A linha vertical representa o mês de maio no ano de 2006. 2 - O cálculo da taxa foi realizado com a população estimada disponibilizada pelo IBGE.

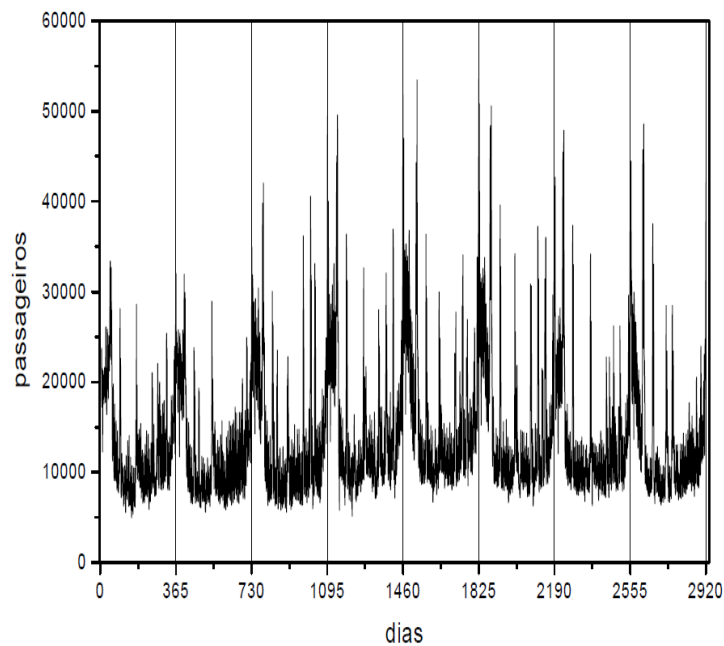


Figura 2.13: Demanda de passageiros transportados diariamente pelo sistema *ferry boat* de 5 de janeiro de 1996 a 31 de dezembro de 2003. Fonte:([NASCIMENTO, 2005](#)).

### 2.1.4 Coeficiente de correlação de Pearson

É possível identificar e mensurar relação entre duas ou mais variáveis e em estatística tal relacionamento é denominado correlação cruzada. Existe correlação entre duas ou mais variáveis quando elas estão relacionadas de alguma forma (TRIOLA, 1999). Nesta seção trataremos apenas da correlação linear. Uma ferramenta de uso frequente na detecção de correlação é o diagrama de dispersão. O diagrama de dispersão é a representação dos pares de valores em um sistema cartesiano (ver Figura 2.14).

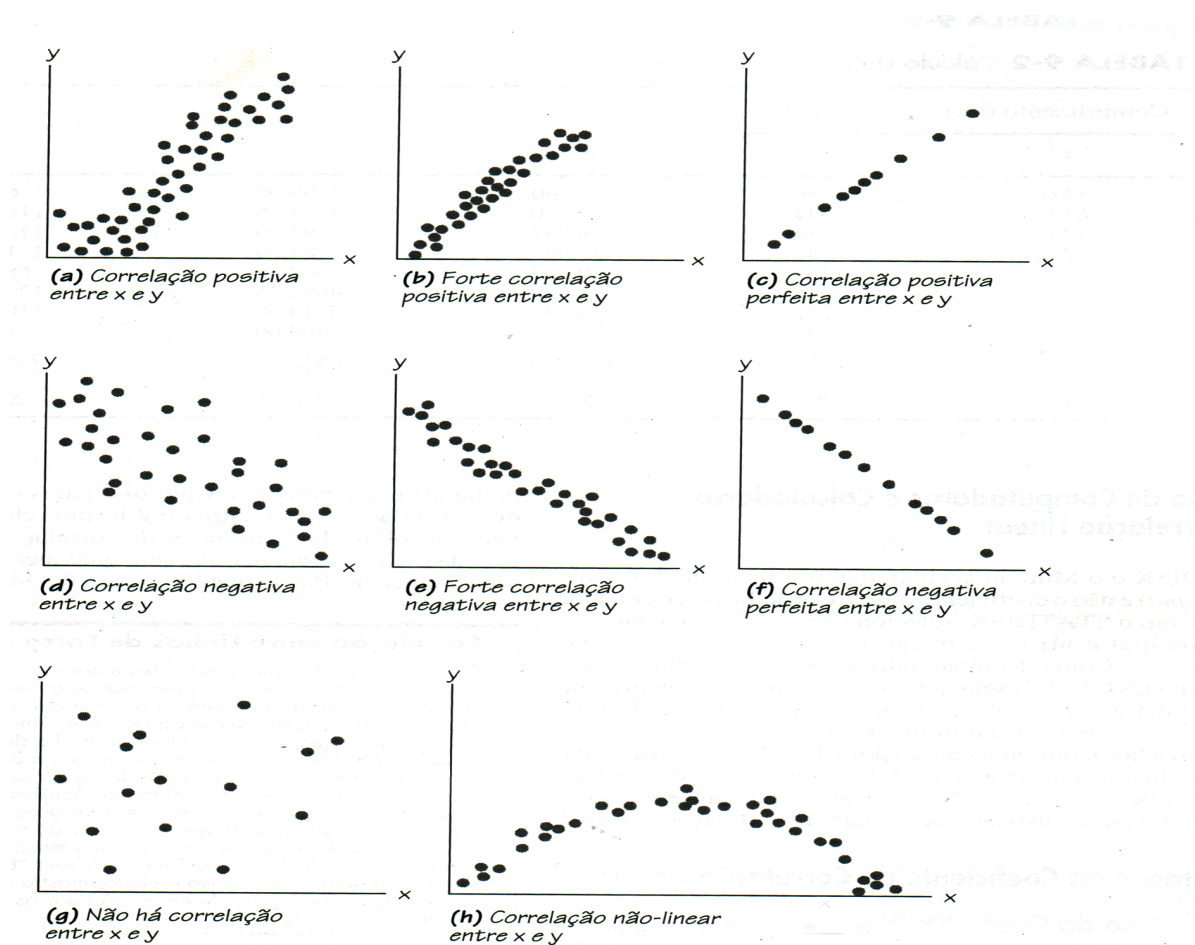


Figura 2.14: Diagrama de dispersão. Fonte: (TRIOLA, 1999).

Devido ao fato das conclusões a partir dos diagramas de dispersão tenderem a análise subjetivas, é necessário métodos mais precisos e objetivos (ou métodos científicos). Um dos métodos para esse fim é o coeficiente de correlação de Pearson, capaz de detectar padrões lineares e não padrões não-lineares como da figura 2.14 h.

O coeficiente de correlação de Pearson ou  $\rho_{xy}$ , proposto por Karl Pearson (1857-1936), é também denotado como correlação momento-produto. Na população, o coeficiente  $\rho_{xy}$  mensura a aderência ou a qualidade do ajuste à verdadeira reta por meio da qual procuramos relacionar as variáveis  $x$  e  $y$  ou ainda o grau de relação linear existente entre elas (FONSECA; MARTINS; TOLEDO, 1985).

O coeficiente de correlação  $\rho_{xy}$  é definido como a razão entre a covariância e a raiz quadrada do produto das variâncias de  $x$  e de  $y$  definido pela seguinte expressão para o caso discreto:

$$\rho_{xy} = \frac{COV_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} \quad (2.10)$$

em que a covariância entre  $x$  e  $y$  é definida por:

$$COV_{x,y} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x) \sum_{i=1}^N (y_i - \mu_y); \quad (2.11)$$

a variância de  $x$  e a  $y$  são definidas respectivamente por:

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2 \quad (2.12)$$

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu_y)^2 \quad (2.13)$$

$N$  na equação 2.11 denota o número de pares  $(x_i, y_i)$  e nas equações 2.12 e 2.13 denota o número de observações da variável objeto de estudo. Já  $\mu_x$  e  $\mu_y$  denotam respectivamente a média de  $x$  e de  $y$ .

O coeficiente de correlação de Pearson, enquanto medida estatística de relacionamento, possui algumas propriedades. Dentre elas:

- O campo de variação de  $\rho_{xy}$  esta entre -1 e +1 inclusive ( $-1 \leq \rho_{xy} \leq 1$ );
- O valor de  $\rho_{xy}$  não varia se todos os valores de qualquer uma das variáveis são convertidos para uma escala diferente; por exemplo, se os pesos de uma variável estiverem em quilos e forem convertidos em libras, o valor de  $\rho_{xy}$  não se modificará;
- O valor de  $\rho_{xy}$  não é afetado pela escolha de x ou y. Permutando todos os pares de x e y,  $\rho_{xy}$  permanecerá inalterado;
- $\rho_{xy}$  mede a intensidade, ou grau de relacionamento linear e não é adequada para medir a intensidade de um relacionamento não linear.

A interpretação do valor do coeficiente de correlação dependerá do valor numérico e do sinal que ele assume. Em relação ao valor e sinal,  $\rho_{xy}$  poderá ter as seguintes classificações:

- a) Correlação linear positiva - quando os valores crescentes de x estiverem associados a valores crescentes de y, ou valores decrescente de x estiverem associados a valores decrescentes da variável y e  $\rho_{xy}$  estiver entre 0 e 1 (ver Figura 2.14 (a) e (b));
- b) Correlação linear positiva perfeita - corresponde ao caso precedente, porém os pontos  $(x_i, y_i)$  estão perfeitamente alinhados e  $\rho_{xy}$  será igual a 1 (ver Figura 2.14 (c));
- c) Correlação linear negativa - quando valores crescente da variável x estiverem associados a valores decrescente da variável y, ou valores decrescente x associados a valores crescente da variável y. Neste caso  $\rho_{xy}$  variará entre -1 e 0 (ver Figura 2.14 (d) (e));
- d) Correlação linear negativa perfeita - quando os pontos estiverem perfeitamente alinhados, mas em sentido contrário, A correlação é denominada negativa. E o valor  $\rho_{xy}$ , neste caso, será igual a -1 (ver Figura 2.14 (f));
- e) Correlação linear nula - quando não houver relação linear entre as variáveis x e y (ver Figura 2.14 (h)).

A correlação será denotada como forte quando mais próximo estiver o resultado de  $\pm 1$ , e será fraca quanto mais próximo estiver de zero.

Assim como outros parâmetros populacionais, como tal a média populacional, é possível testar a significância para o  $\rho_{xy}$ . As hipóteses usuais são:

$$H_0 : \rho_{xy} = 0 \text{ (Correlação linear não significativa)}$$

$H_1 : \rho_{xy} \neq 0$  (Correlação linear significativa)

em que  $H_0$  denota a hipótese nula e  $H_1$  hipótese alternativa.

O teste de hipótese para o parâmetro  $\rho_{xy}$  segue em geral os seguintes passos:

1º passo: Determinar as hipóteses nula e alternativa;

2º passo: Selecionar a estatística de teste que será utilizada;

$$t_c = \hat{\rho}_{xy} \sqrt{\frac{n-2}{1-\hat{\rho}_{xy}^2}} \quad (2.14)$$

em que  $t_c \sim t_{n-2}$  ( $t_c$  tem distribuição  $t$ -student com  $n-2$  graus de liberdade). A expressão 2.14 é válida quando a distribuição de  $X$  e  $Y$  é uma normal bivariada.

3º passo: Especificar o nível de significância  $\alpha$  para o teste;

4º passo: Usar o nível de significância para estabelecer uma regra de decisão que levará a rejeição ou não de  $H_0$  (ver Figura 2.15);

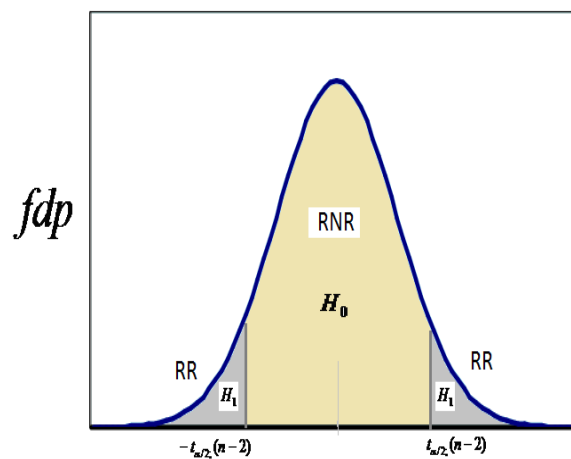


Figura 2.15: Hipóteses estatísticas do parâmetro  $\rho_{xy}$ . RR - Região de rejeição de  $H_0$  e RNR Região de não rejeição de  $H_0$ .  $fdp$  denota função de densidade de probabilidade e  $-t_{\alpha/2; (n-2)}$  e  $t_{\alpha/2; (n-2)}$  são os valores críticos da distribuição  $t$ -Student.

5º passo: A partir dos dados amostrais calcular a estatística de teste do 2º passo;

6º passo: Comparar o valor da estatística do teste com os valores específicos na regra de decisão, para determinar se  $H_0$  deve ser rejeitada ou não; ou calcular o valor p, baseado na estatística de teste. Comparar o valor p com  $\alpha$  para determinar se  $H_0$  deve ser rejeitada ou não. E os valores críticos  $t_{\alpha/2(n-2)}$  devem ser obtidos em uma tabela  $t$  de Student;

7º passo: Concluir, baseado na decisão tomada (na rejeição ou não rejeição da hipótese nula).

Mais detalhes sobre o teste estatístico do parâmetro  $\rho_{xy}$  podem ser encontrados em (SPIEGEL, 1992), (TRIOLA, 2008) e (BARBETTA; REIS; BORNIA, 2010). O coeficiente de correlação de Pearson é também utilizado para analisar correlação cruzada entre séries temporais como a pesquisa realizada por (ZEBENDE; SILVA; MACHADO, 2011) que estudaram correlação cruzada entre séries temporais de acidentes de táxi de uma cooperativa e a de (WANG et al., 2013) que avalizaram o mercado de ações dos Estados Unidos.

Apesar de existir na literatura alguns trabalhos usando o coeficiente de correlação de Pearson para análise de séries temporais, como mencionado anteriormente,  $\rho_{xy}$  não é adequado para este fim, devido ao fato de exigir que as observações das variáveis X e Y sejam independentes. Entretanto, a referida exigência contraria o princípio básico da análise de séries temporais que é a autocorrelação ou dependência temporal entre as observações das variáveis objeto de estudo. E, recentemente foi proposto por (ZEBENDE, 2011) um coeficiente capaz de mensurar correlação cruzada em séries temporais que não possui a referida limitação (ver subseção 2.4.3).

## 2.2 Fractalidade e séries temporais

Na presente seção, serão abordados conceitos relativos a fractalidade e séries temporais, que servirão de alicerce para o entendimento dos métodos DFA, DCCA e o coeficiente de correlação cruzada  $\rho_{DCCA}$ .

### 2.2.1 Fractais

Nas últimas décadas aconteceram investigações cujo o objeto central do estudo foi a construção de entidades geométricas e tais entidades (ou objetos) foram chamados de fractais pelo seu precursor o cientista polonês Benoit Mandelbrot (BARBOSA, 2005). Benoit Mandelbrot está para geometria fractal como Euclides está para a geometria clássica (GLEISER, 2002). Durante séculos, a geometria Euclidiana foi considerada a melhor forma de representar o mundo onde vivemos e a natureza era concebida pelo homem com o conceito de formas, planos, volumes, corpos, retas e curvas. E os modelos matemáticos, construídos ao longo dos últimos séculos, para descrição de fenômenos físicos e no uso tecnológico como na construção de máquinas predominavam sempre as formas regulares: retas, o círculo (roda), a esfera, o cilindro, o triângulo, a elipse, a parábola e outras curvas similares (MOREIRA, 1999). Porém, qual a geometria da natureza? Este questionamento entre outros externados por Mandelbrot que dão origem as primeiras idéias sobre fractais (MOREIRA, 1999).

Em 1975 Benoit Mandelbrot introduziu o termo *fractal* e após a publicação do seu livro denominado *The Fractal Geometry of Nature* (MANDELBROT, 1982), o termo fractal passou a caracterizar formas irregulares. E de acordo com (MOREIRA, 1999) uma frase escrita logo no primeiro parágrafo do seu livro ficou conhecida por conter uma espécie de *slogan* da nova geometria que se propunha: “Nuvens não são esferas, montanhas não são cones, linhas costeiras não são círculos, cascas de árvores não são suaves nem raio se propaga em linha reta”.

A palavra fractal, introduzida pelo cientista Mandelbrot, tem origem no radical *fractus*, proveniente do verbo latino *frangere*, que denota quebrar, produzir pedaços irregulares; e vem da mesma raiz da palavra fragmentar do português (MOREIRA, 1999). Segundo (MOREIRA, 1999), os fractais são conjuntos que apresentam uma forma irregular ou fragmentada e que têm essencialmente a mesma estrutura em todas as escalas. Para (FEDER, 1988) um fractal é uma forma cujas partes se assemelham ao seu todo em alguns aspectos.

Entre as principais propriedades que permitem definir e caracterizar os conjuntos fractais destacam-se (MOREIRA, 1999):



□ Auto-similaridade, que pode ser classificada como exata ou estatística. E neste caso o sistema é invariante (mantém a mesma forma e estrutura sob uma transformação de escala);

□ A extrema irregularidade no sentido de rugosidade (não suavidade) ou fragmentação;

□ E por fim uma dimensão fractal não-inteira ou fracionada.

Para um melhor entendimento da auto-similaridade considere a seguinte situação (JANOS, 2008): Tire uma foto de um couve-flor e de uma pequena parte de seu corpo e amplie as duas fotos do mesmo tamanho. Se a foto não tiver fundo, será geralmente impossível distinguir qual a couve-flor inteira e qual é o pedaço. Isto se comporta assim porque pequenos pedaços da couve-flor são semelhantes ao todo e, neste caso, podemos afirmar que a couve-flor é auto-similar.

O entendimento de dimensão é algo relevante para estabelecer a diferença entre Geometria Euclidiana e a Geometria Fractal. Essas duas idéias próximas, porém diversas, estão ligadas ao termo *dimensão* usualmente abordado (MOREIRA, 1999):

1) O número de informações necessárias para se localizar um ponto no espaço e neste caso é possível afirmar que o espaço possui três dimensões;

2) A noção de medida de comprimento que, neste caso, é possível afirmar que a dimensão de um objeto é de 50cm.

O termo importante de dimensão é a chamada *dimensão topológica* relacionada a primeira idéia supracitada que foi abordada por Poincaré em 1911 e por Brouwer em 1913 afirma que: um contínuo tem  $n$  dimensões quando podemos dividi-lo por meio de cortes que sejam eles próprios contínuos de  $(n - 1)$  dimensões. Assim, um ponto possui dimensão zero, a reta terá dimensão 1 (porque pode ser separada por um ponto), o plano terá dimensão 2 (por que pode ser separado por uma reta), o espaço usual terá três dimensões, e assim sucessivamente. Nesta definição não entra a noção de medida de distância (métrica), e trata-se de uma noção topológica, relacionada apenas à noção de vizinhança entre pontos de um conjunto, o que permite a definição de continuidade.

Considerando o aspecto métrico uma definição de dimensão é a denominada *capacidade*, definida por Kolmogorov, que mensura o quanto o conjunto ou objeto considerado preenche o espaço que está imerso. E tal definição, é uma das mais simples, que permite caracterizar os fractais e é usualmente chamada de dimensão fractal, porém existem outras dimensões capazes de caracterizar o grau de fractalidade de um conjunto. A definição da capacidade ( $d_{cap}$ ) de um conjunto é definida pela seguinte expressão (MOREIRA, 1999):

$$d_{cap} = \lim_{\varepsilon \rightarrow 0} \{\log[N(\varepsilon)]/\log(1/\varepsilon)\}, \quad (2.15)$$

em que  $N(\varepsilon)$  é o número mínimo de cubos elementares necessários para cobrir o conjunto considerado e  $\varepsilon$  é a dimensão linear do cubo elementar.

Na Geometria Euclidiana as dimensões são bem definidas para as formas da natureza, isto é, uma reta tem dimensão igual a 1 ( $d = 1$ ), um plano tem dimensão igual a 2 ( $d = 2$ ), o volume tem dimensão igual a 3 ( $d = 3$ ). Já a Geometria Fractal tem uma dimensão fracionária ( $d_f$ ), um número não inteiro que proporciona um melhor ajuste a realidade da natureza.

A geometria da natureza é fractal tendo em vista que em diversas formas naturais como árvores, nuvens, linhas costeiras e folhas partes menores são similares a partes maiores. E estes objetos não podem ser perfeitamente representados utilizando elementos da geometria euclidiana, porque o formato de pedras, nuvens, montanhas e árvores são mais complicados do que a geometria euclidiana poderia representar. Montanhas não são cones, um rio não é uma linha reta (GLEISER, 2002). O desenvolvimento da Geometria fractal é um dos mais importantes estudos científicos do século XX e com o uso da teoria fractal os cientistas podem descrever formas naturais com o uso de simples equações matemáticas, calculadas de forma recursiva em seus computadores (GLEISER, 2002).

A título de exemplo da determinação da dimensão fractal obtida por meio da expressão 2.15 considere os seguintes casos (MOREIRA, 1999):

### Conjunto de Cantor

O conjunto de Cantor é um conjunto modelado da seguinte forma: tomamos um segmento de reta e o dividimos em três pedaços iguais. Em seguida, o pedaço intermediário é retirado. Os dois segmentos restantes são novamente repartidos em três segmentos iguais e os segmentos intermediários são retirados. O processo de dividir os segmentos e de retirar o pedaço intermediário se repete recursivamente. E a capacidade, ou dimensão fractal desse conjunto pode ser inferido por meio da figura 2.16 e sua  $d_{cap} = \log(2)/\log(3) \cong 0,6$ . Devido ao fato de cada etapa do processo de construção do conjunto foi utilizado dois segmentos (cubos elementares) para cobrir a figura, sendo que nesta situação cada segmento elementar tinha comprimento de  $\frac{1}{3}$  e é possível observar que esse conjunto tem comprimento zero, porque a cada passo do processo o seu comprimento é reduzido por um fator de  $\frac{2}{3}$ . Assim, seu comprimento, no limite em que  $n \rightarrow \infty$  teremos  $L = (\frac{2}{3})^n \rightarrow 0$ .

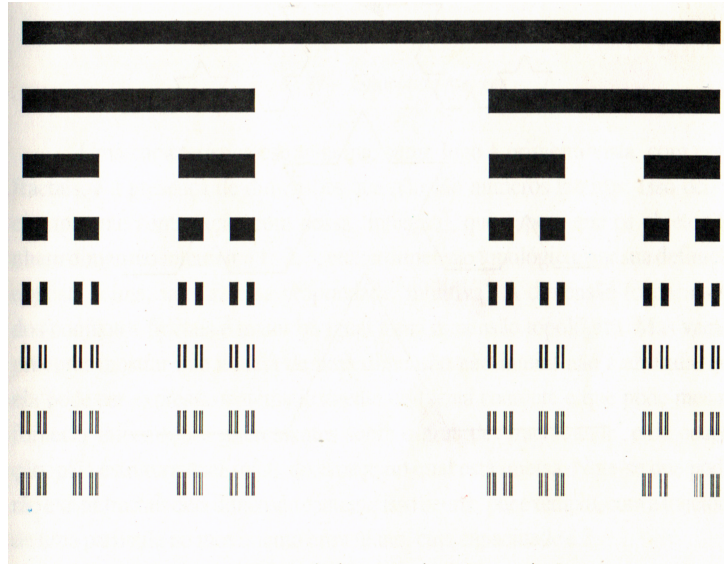


Figura 2.16: Processo de formação do Conjunto de Cantor. Fonte:([MOREIRA, 1999](#)).

### Conjunto de Koch

O conjunto de Koch é construído segundo representação da figura 2.17. Neste caso ao invés de retirarmos o pedaço intermediário do segmento inicial é realizada uma substituição por mais dois segmentos iguais, como representado na figura 2.17. A dimensão fractal do conjunto de Koch será dada por  $d_{cap} = \log(4)/\log(3) \cong 1,26$ . E o comprimento desse conjunto tende para infinito, considerando em cada etapa do processo de construção  $(\frac{4}{3})^n$ .

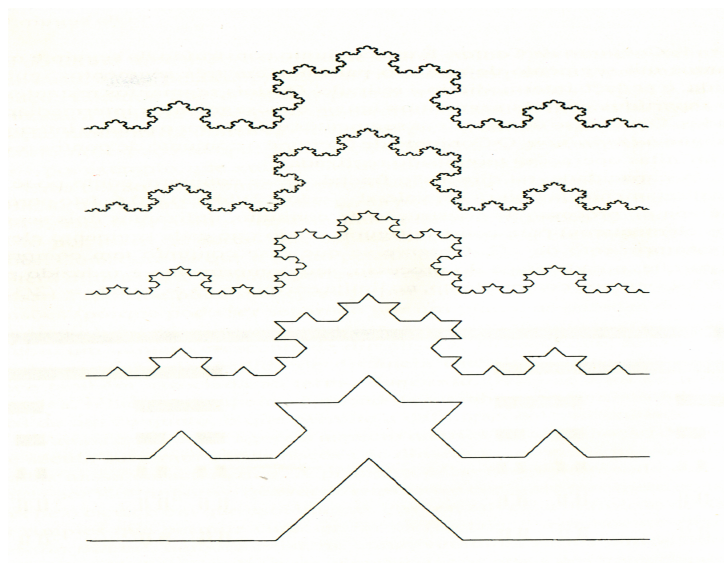


Figura 2.17: Processo de formação do Conjunto de Koch. Fonte:([MOREIRA, 1999](#)).

### Triângulo de Sierpinski

A dimensão fractal do triângulo de Sierpinski é 1,58. Para construir um triângulo de Sierpinski começa-se com um triângulo equilátero ao invés de adicionar outros triângulos remove o triângulo equilátero de dentro do triângulo anterior em seguida repita por 10.000 vezes este processo e o resultado será a figura 2.18 (GLEISER, 2002). Os fractais gerados teoricamente ou matematicamente como o triângulo de Sierpinski possui a característica de auto-similaridade em um número infinito de escala, porém os fractais naturais têm escala limitada de similaridade (COSTA; VASCONCELOS, 2003).

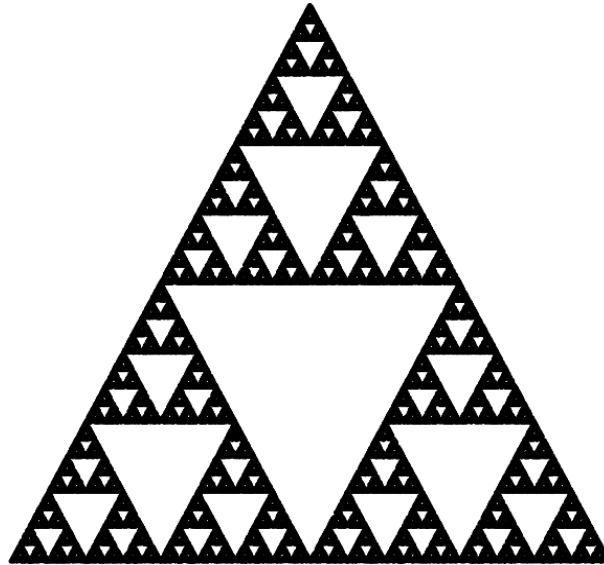


Figura 2.18: Triângulo Sierpinski.

Existem várias aplicações com modelos que possuem estrutura fractal para representar ou descrever muitos processos naturais em diversas áreas do conhecimento da ciência, da física à biologia, da geologia à astrofísica (MOREIRA, 1999). E o primeiro cientista que aplicou a análise fractal em séries temporais foi Benoit B. Mandelbrot, que inclui as primeiras abordagens por Hurst sobre sistemas hidrológicos (KANTEHARDT, 2008). E na última década o comportamento de escala fractal tem sido aplicada em várias séries temporais (KANTEHARDT, 2008).

Na próxima seção abordaremos características relevantes para compreensão das propriedades de escala contidas em algumas séries temporais que servirão de alicerce para o entendimento dos métodos DFA, DCCA e o coeficiente de correlação cruzada  $\rho_{DCCA}$ .

### 2.2.2 Auto-similaridade exata e estatística

Propriedades de escala podem ser observadas também em séries temporais. O cientista Mandelbrot estudou exaustivamente o fenômeno da auto-similaridade em escalas espaciais e temporais por meio de trabalhos relativos a frequência de palavras, periodicidade de ruídos em ligações telefônicas, turbulência, flutuações de preços de *comodities*. E no universo das flutuações dos preços de *comodities* sua área de interesse era a flutuação dos preços de algodão (GLEISER, 2002).

Como foi dito anteriormente, os fractais naturais possuem escala de similaridade limitada, ou seja, o conceito de auto-similaridade exata trata-se de uma abstração matemática porque na natureza não existem objetos com tal característica. E como já mencionamos, o caso do triângulo de Sierpinski trata-se de uma *auto-similaridade exata*.

Os objetos que não possuem a denominada auto-similaridade exata podem ser caracterizados como um fractal, porém a auto-similaridade, neste caso, será denominada *auto-similaridade estatística* ou *auto-afinidade*. E este conceito será de grande relevância para o estudo e compreensão dos métodos que serão abordados em seções subsequentes conhecidos na literatura como métodos da mecânica estatística: DFA, DCCA e  $\rho_{DCCA}$  que são métodos capazes de avaliar propriedades de escala. Para maiores esclarecimentos sobre a auto-afinidade consultar (FEDER, 1988).

### 2.2.3 Persistência, correlação de longo alcance e curto alcance

Uma propriedade relevante em uma série temporal é a autocorrelação, que determina, se a série possui memória de longa duração ou não, isto é, se a ST é persistente, antipersistente ou decorrelacionada. Considerando os incrementos  $\Delta_{t_i} = t_i - t_{i-1}$  de uma série temporal auto afim,  $(t_i)$ ,  $i = 1, 2, \dots, N$  com  $N$  valores equidistantes no tempo, neste caso,  $\Delta_{t_i}$  pode ser persistente, decorrelacionado ou antipersistente (KANTELHARDT, 2008).

Segundo (ANTENEODO, 2004), uma série temporal pode ser caracterizada como persistente se observações adjacentes estão correlacionadas positivamente. E a ST é caracterizada como antipersistente se a série apresentar correlação inversa se existir a tendência de valores grandes seguidos por valores pequenos como o batimento cardíaco. Caso nenhuma das situações mencionadas ocorra, a série é dita decorrelacionada, é o que acontece com o ruído branco.

De acordo com (KANTELHARDT, 2008), a persistência é válida para todas as escalas de tempo, em que a relação de auto afinidade também é válida. No entanto, o grau de persistência também pode variar em diferentes escalas de tempo. Condições do tempo (temperatura, umidade) é um exemplo típico: enquanto as condições de tempo de amanhã ou de uma semana é provável ser semelhante as condições do tempo de hoje (devido a uma condição estável do tempo geral), porém a persistência é muito mais difícil de ser identificada em escala de tempo mais longas.

Nas seções subsequentes abordaremos métodos capazes de identificar e mensurar o grau de persistência de uma série temporal, ou seja, se a série apresenta correlação de longo alcance ou não.

## 2.3 Métodos de análise para séries temporais estacionárias

Como já foi dito a estacionariedade é não uma característica que está presente na maioria das séries temporais reais. Os métodos abordados na presente seção, tem como característica comum a necessidade de tornar a série temporal estacionária para a sua aplicação.

### 2.3.1 Função de Autocorrelação - FAC

A Função de Autocorrelação ou correlação serial a (FAC) é uma importante ferramenta para identificação das propriedades de uma ST. A FAC é relevante quando o objetivo é caracterizar uma série temporal em relação a sua dependência temporal, ou seja, quando queremos mensurar quanta dependência existe entre observações de uma ST. A FAC tem o propósito de mensurar a correlação entre observações defasadas por períodos de uma série temporal com  $n$  observações. A FAC com defasagem  $k$ , denotada por  $\rho_k$  e definida como a razão entre a covariância com defasagem  $k$  e a variância. Como na prática temos a realização (amostra) de um processo estocástico, podemos definir a função de autocorrelação amostral  $\hat{\rho}_k$  pela seguinte expressão (SOUZA; CAMARGO, 2004):

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{X})(x_{t+k} - \bar{X})}{\sum_{t=1}^n (x_t - \bar{X})^2}, k = 0, 1, \dots \quad (2.16)$$

A expressão acima é a covariância amostral da série na defasagem  $k$  dividida pela variância amostral da série, em que  $n$  denota o comprimento da série e  $\bar{X}$  o valor médio das observações, determinadas a partir da variação temporal  $k$ . Devido ao fato que tanto a covariância como a variância são mensuradas nas mesmas unidades de medida,  $\hat{\rho}_k$  é um número adimensional e varia de -1 a +1 inclusive ( $-1 \leq \hat{\rho}_k \leq 1$ ). E para sua interpretação considere os seguintes casos:

- i) ( $0 < \hat{\rho}_k \leq 1$ ) a ST possui autocorrelação positiva;
- ii) ( $-1 \leq \hat{\rho}_k < 0$ ) a ST possui autocorrelação negativa;
- iii) ( $\hat{\rho}_k = 0$ ) a ST não tem autocorrelação ou correlação serial.

Representando graficamente  $\hat{\rho}_k \times k$  teremos o correlograma amostral, que pode ser um procedimento estatístico importante para caracterizar uma série temporal (ver Figura 2.19)

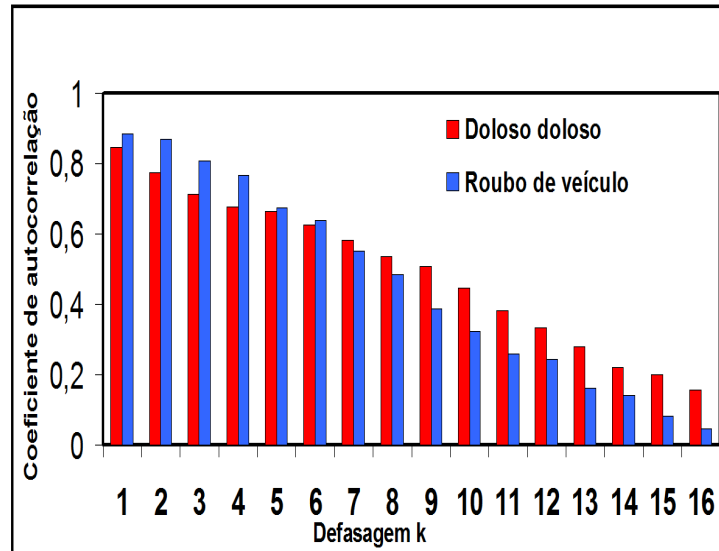


Figura 2.19: Correlograma dos homicídios dolosos e roubos de veículos registrados diariamente em Salvador-BA, 2003 a 2008 ( $n = 72$  meses).

A figura 2.19 possui 16 defasagens para séries consideradas e segundo (GUJARATI; PORTER, 2011) a extensão da defasagem é uma questão basicamente empírica e segundo o autores uma regra geral é computar  $\frac{1}{3}$  da extensão da série temporal. E, segundo (BOX; JENKINS, 1976), para obter uma boa estimativa de  $\rho_k$  é necessário pelo menos 50 observações .

Tendo como método de análise a FAC em diferentes defasagens é possível também caracterizar uma série temporal como:

- a) Estacionária, caso os valores de  $\hat{\rho}_k$  decrescerem rapidamente para zero;
- b) Não estacionárias, se as estatística de  $\hat{\rho}_k$  decrescerem lentamente para zero;
- c) Periódica, quando a série apresentar característica de sazonalidade de um dado período mudando de sinal;
- d) Ruído branco se os valores de  $\hat{\rho}_k$  forem nulos, exceto  $k = 0$ , que é igual a 1.



Analisando a figura 2.19 é possível inferir que trata-se de séries não estacionárias porque os valores de  $\hat{\rho}_k$  nas duas séries temporais tenderam a zero lentamente. Por meio da FAC é possível também caracterizar a memória da ST. Ou seja, se a série temporal possui memória curta ou longa. Caso as estatísticas de  $\hat{\rho}_k$  decrescerem rapidamente para zero a série é caracterizada como de memória curta e a sua dependência temporal permanece presente por um período curto. Mas, a FAC é adequada na detecção de memória se a série for estacionária e, como já foi dito anteriormente, a maioria das séries temporais reais são não estacionárias.

### 2.3.2 Estatística de Hurst

Harold E. Hurst (1880-1978) foi um hidrólogo que dedicou parte de sua vida descrevendo os problemas relativos aos reservatórios do rio Nilo. Seu estudo estava fundamentado na otimização dos fluxos de água do rio Nilo, de forma que a represa não excedesse o seu volume de água e nem ficasse vazia. Ao longo desse estudo, Hurst percebeu que as referidas enchentes e vazões do rio Nilo caracterizavam-se como um processo estocástico (ver Seção 2.1.1). Com o objetivo de caracterizar estatisticamente os problemas de dimensionamento das represas foi idealizado por (HURST, 1951), a análise estatística  $R/S$ , que pode ser encontrada em detalhes no livro *Long-Term Storage: An Experimental Study* (HURST; BLACK; SIMAIKA, 1965). Apesar da Estatística  $R/S$  originalmente ter sido aplicada em hidrologia, é possível encontrar na literatura um número significativo de aplicações em diversas áreas da ciência (MORETTIN, 2008). É possível citar sua aplicação em:

- Séries financeiras (SOUZA; TABAK; CAJUEIRO, 2006), (SILVA, 2009);
- Séries temporais biológicas (LIEBOVITCH; YANG, 1997), (SANTOS et al., 2008), (OLIVEIRA; GOUVEA; RIBEIRO, 2008);
- Redes de computadores (GOMES et al., 2001) e demais áreas do conhecimento.

De acordo com (HURST, 1951), a estatística  $R/S$  estava baseada na medição dos volumes máximos e mínimos do reservatório (amplitude) e no cálculo do desvio padrão dos fluxos de água em um período de tempo  $\tau$ , de forma que  $R/S$  denotaria a razão entre a amplitude ( $R_\tau$ ) e o desvio padrão ( $S_\tau$ ) das vazões de água. Trata-se de um valor sem dimensão (adimensional).

Hurst ao modelar a estatística  $R/S$  para diversos períodos de tempo observou que esta estatística apresentava uma função de relacionamento com o número de amostras ou observações utilizadas na modelagem, e isso também para outros fenômenos naturais (SOUZA; TABAK; CAJUEIRO, 2006).

(HURST, 1951) definiu os seguintes passos para modelar o expoente de Hurst (ver Figura 2.20):

i) Calcula-se a média  $\langle \xi \rangle_\tau$  do influxo de água no período  $\tau$  definida por:

$$\langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t) \quad (2.17)$$

em que  $\xi(t)$  representa o influxo de água no reservatório no momento de tempo  $t$ .

ii) Determina a variação da média para o tempo  $t$  como:

$$X(t, \tau) = \sum_{v=1}^t [\xi(v) - \langle \xi \rangle_\tau], \quad (2.18)$$

em que  $\langle \xi \rangle_\tau$  denota a média de todos os pontos da série no período em estudo.

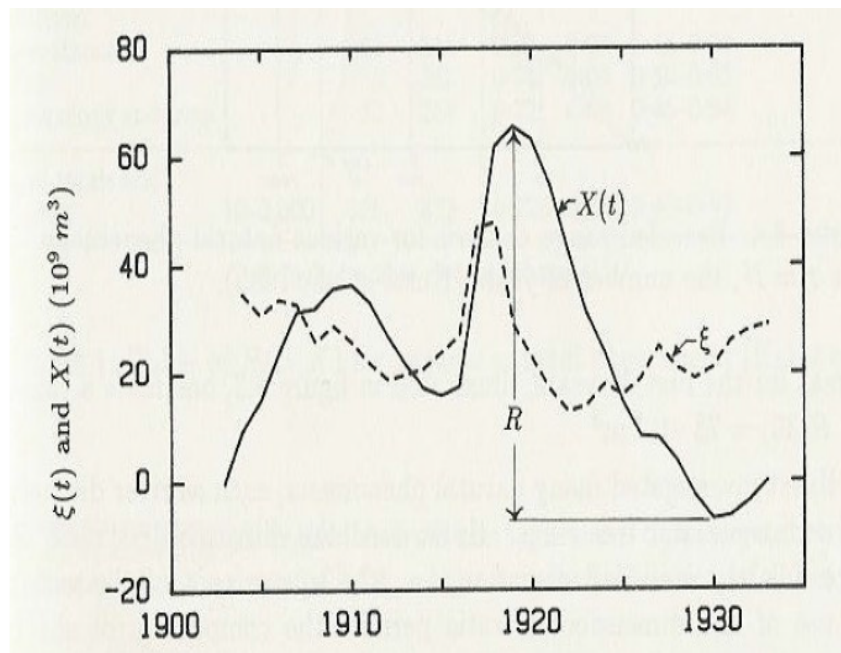


Figura 2.20: Modelo do reservatório de Hurst (1951) cálculo de  $X(t, \tau)$ . Fonte: (FEDER, 1988).

iii) Calcula-se a diferença entre o valor máximo e mínimo de  $X(t)$  no período  $\tau$  definida pela seguinte expressão (ver Figura 2.21):

$$R(\tau) = \text{Max}[X(t, \tau)] - \text{Min}[X(t, \tau)] \quad (2.19)$$

em que  $1 \leq t \leq \tau$

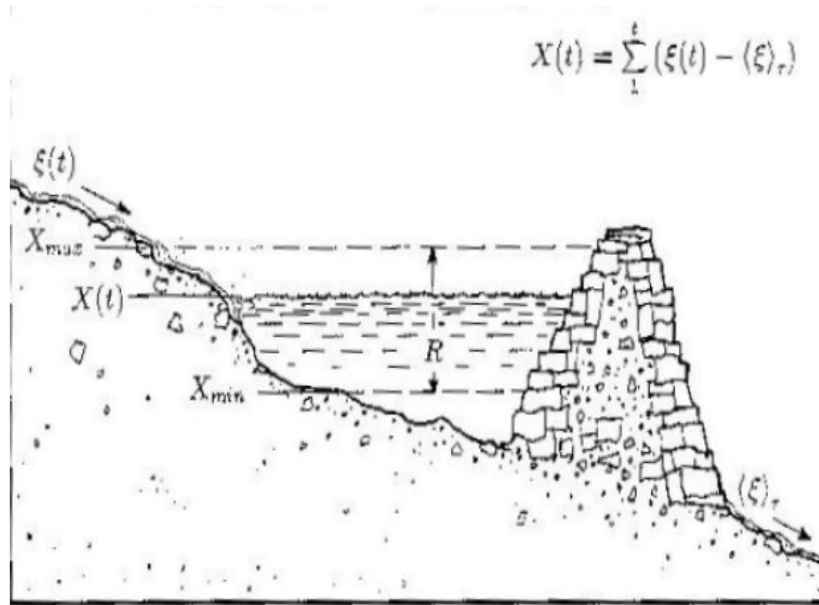


Figura 2.21: Modelo do reservatório de Hurst (1951) cálculo de  $R(\tau)$ . Fonte: (FEDER, 1988).

iv) Divide-se  $R(\tau)$  pelo desvio padrão  $S(\tau)$ ,

$$R/S = \frac{R(\tau)}{S(\tau)}, \quad (2.20)$$

em que,

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} [\xi(t) - \langle \xi \rangle_{\tau}]^2} \quad (2.21)$$

Considerando que o procedimento anterior seja repetido sistematicamente para valores distintos de  $(\tau)$  e a relação  $(R/S)$  seguir uma lei de potência, então temos:

$$R/S = (\tau/2)^H, \quad (2.22)$$

onde  $H$  denota o expoente de Hurst.

Os valores  $H$  tem campo de variação entre 0 e 1 ( $0 \leq H \leq 1$ ). Uma propriedade relevante do expoente Hurst é a possibilidade de informar a persistência ou não do sinal. Se o valor de  $H$  for igual a 0,5 a série é caracterizada como decorrelacionada. Caso  $H$  esteja compreendido entre 0,50 e 1 ( $0,50 < H \leq 1,0$ ) a série temporal é denotada como persistente. Neste caso, um aumento na tendência de uma série temporal implicará em um aumento futuro desta série, ou uma redução na série temporal tenderá a uma redução futura. Para o expoente de Hurst variando de 0 a 0,5 ( $0 \leq H < 0,5$ ) é possível inferir que a série apresenta comportamento antipersistente ([FEDER, 1988](#)).

Mesmo com sua robustez matemática a Estatística  $R/S$  necessita de filtragem ou transformação nos dados antes de sua aplicação, afim de eliminar possíveis tendências ([SOUZA; TABAK; CAJUEIRO, 2006](#)). O que não é necessário nos métodos que abordaremos na seção seguinte.

## 2.4 Métodos de análise para séries temporais não estacionárias

A caracterização de uma ST em relação a sua correlação de longo alcance tem sido realizada por várias metodologias (TAQQU; TEVEROVSKY; WILLINGER, 1995). Nas próximas subseções da presente Tese de Doutorado apresentaremos o *Detrended Fluctuation Analysis* - DFA (PENG et al., 1994), o *Detrended Cross-Correlation Analysis* - DCCA (PODOBNIK; STANLEY, 2008), e por fim, o coeficiente de Correlação Cruzada  $\rho_{DCCA}$  (ZEBENDE, 2011), métodos capazes de detectar e mensurar correlação em séries temporais não estacionárias em diferentes escalas de tempo. E são conhecidos na literatura como métodos da mecânica estatística.

O método DFA idealizado por (PENG et al., 1994) tem a função de detectar e mensurar autocorrelação (correlação do sinal com ele mesmo) de longo alcance em séries temporais em regime não estacionário, enquanto o DCCA modelado por (PODOBNIK; STANLEY, 2008) se destina a identificação e mensuração de correlação cruzada em duas séries temporais de tamanho  $N$ , também em regime não estacionário. Devido ao fato do método DCCA apenas identificar e quantificar a correlação cruzada, foi idealizado por (ZEBENDE, 2011) um coeficiente capaz de identificar e mensurar o nível da correlação cruzada em diferentes escalas de tempo, o  $\rho_{DCCA}$ .

### 2.4.1 Método DFA-*Detrended Fluctuation Analysis*

A análise das flutuações sem tendência do inglês *Detrended Fluctuation Analysis*-DFA idealizado (PENG et al., 1994), pode ser aplicado para identificar e mensurar autocorrelação de longo alcance em séries temporais não-estacionárias. O DFA é uma evolução da análise  $R/S$  clássica. A peculiaridade do DFA é permitir a identificação de auto-afinidade, e identificação de correlações de longo alcance em séries temporais com tendências. Auto-afinidade em matemática trata-se de fractal cujas partes são diferentes na direção  $x$ ,  $y$  e  $z$ . Isso denota que, para observar a auto-semelhança desse fractal, é necessário redimensioná-lo usando uma transformação anisotrópica (FEDER, 1988), (CRUZ, 2002). Já em uma série temporal existe auto-afinidade quando são mantidas as propriedades estatísticas observadas em diferentes escalas.

O DFA tem sido modelado em diversos campos da ciência, por exemplo:

■ Em sinais biológicos (PENG et al., 1994), (PENG et al., 1995), (IVANOV et al., 1998), (WALLECZEK, 2000), (OLIVEIRA et al., 2006), (TAKAKURA, 2007), (GALHARDO; BARBOSA; PENNA, 2008), (GALHARDO et al., 2009);

- Em séries temporais do controle motor ([DINIZ, 2008](#));
- Em séries temporais de manchas solares ([MORET, 2014](#));
- Em séries temporais do mercado financeiro ([MANTEGNA; STANLEY, 1995](#)), ([LIU et al., 1999](#)), ([PODOBNIK; STANLEY, 2008](#)), ([RUAN; ZHOU, 2011](#)), ([ZEBENDE; SILVA; MACHADO, 2013](#));
- No estudo de séries econômicas ([NASCIMENTO; ZEBENDE; MORET, 2008](#)), ([NASCIMENTO, 2005](#)), ([ZEBENDE; MACHADO, 2009a](#)), ([ZEBENDE; SILVA; MACHADO, 2011](#)), ([MORET; ANTONIO; PEREIRA, 2012](#));
- Na análise de fenômenos físicos ([ZEBENDE et al., 2004](#));
- Flutuações de sistemas binários de raios-x estelar ([MORET et al., 2003](#));
- No estudo de séries temporais de indicadores de criminalidade ([SOARES et al., 2008](#)), ([MACHADO; ZEBENDE; SOARES, 2008](#)), ([SOARES; MACHADO, 2008a](#)), ([SOARES; MACHADO, 2008b](#)), ([MACHADO; SOARES; ZEBENDE, 2009](#)), ([MACHADO; ZEBENDE; SOARES, 2009](#)), ([MACHADO, 2009](#)), ([MACHADO; ZEBENDE; SILVA, 2011](#)), ([MACHADO et al., 2012](#)), ([MACHADO; SILVA; ZEBENDE, 2014](#));
- Na análise de séries cronológicas de acidentes ([ZEBENDE; SILVA; MACHADO, 2011](#)), ([MACHADO et al., 2012](#));
- Em climatologia ([PEDRON, 2007](#)), ([TALKNER; WEBER, 2000](#)).

Para modelar estatisticamente o DFA é necessário seguir ao menos três passos principais. Com base em uma variável  $u_i$ , com  $i$  variando de 1 a  $N$  (número total de pontos da série) (ver Figura 2.22).

Passo 1: Primeiramente, calcula-se o desvio de cada registro em relação à incidência média da série completa (ver Figura 2.23), desta forma, integrando o sinal  $u_i$  e obtendo a série integrada  $y(k)$  pela seguinte expressão:

$$y(k) = \sum_{i=1}^k u_i - \langle u \rangle; k = 1, 2, \dots, N \quad (2.23)$$

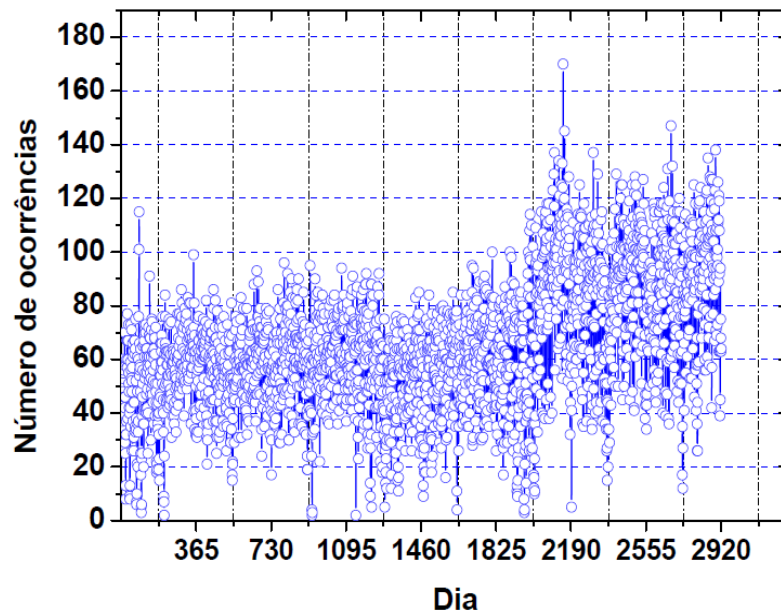


Figura 2.22: Série temporal dos acidentes de trânsito ( $u_i$ ) sem lesão corporal registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR.

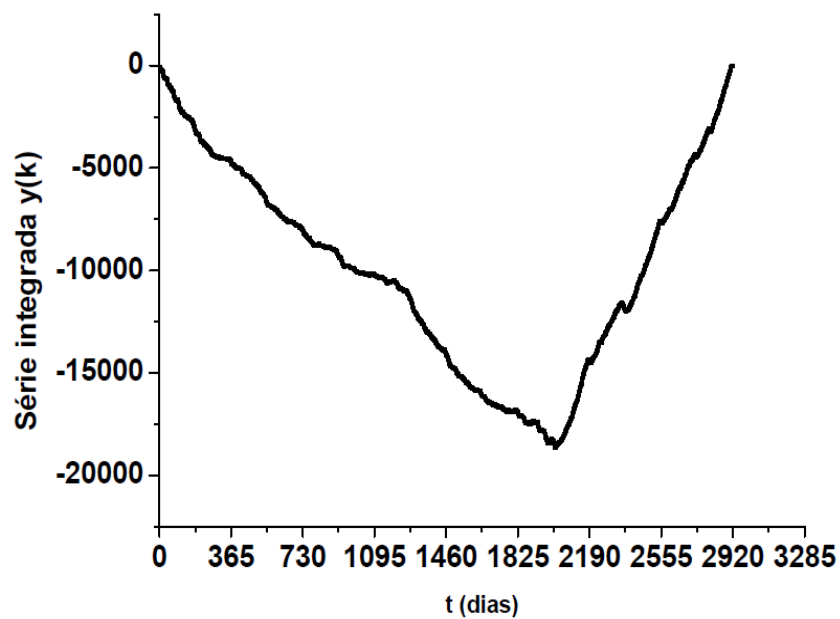


Figura 2.23: Série integrada  $y(k)$  dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. Nota: Cálculo do autor.

em que  $\langle u \rangle$  denota o valor médio de  $u_i$ .

Passo 2: Em seguida a série integrada  $y(k)$  é dividida em intervalos de igual amplitude ( $n$ ) não sobrepostos (ver Figura 2.24).

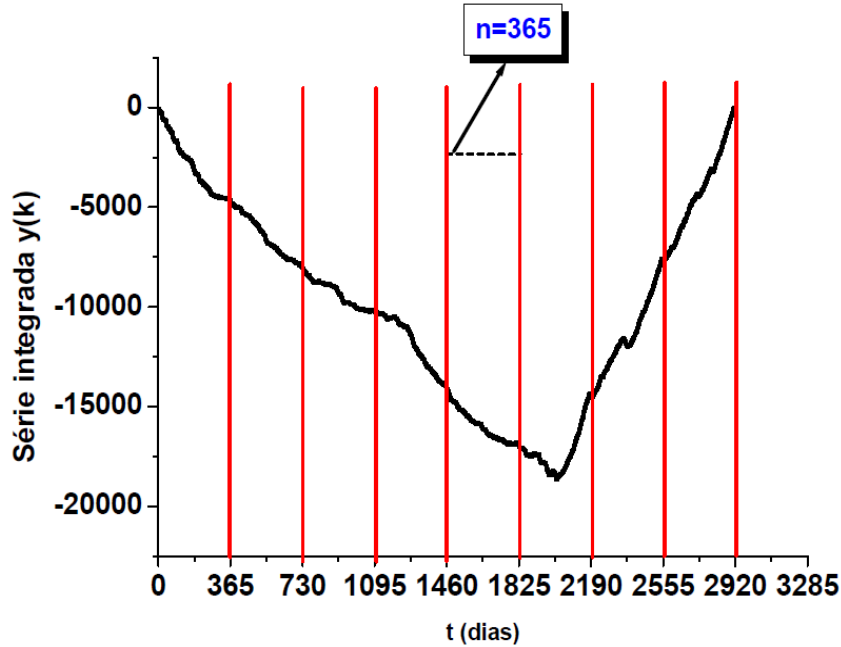


Figura 2.24: Série integrada  $y(k)$  dividida em (box) de tamanho 365 dias dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. Nota: Cálculo do autor.

Passo 3: Para cada intervalo de tamanho  $n$ , é necessário ajustar um polinômio de grau  $\geq 1$  a  $y(k)$ , denotado por  $y_n(k)$ , esta estatística representa a tendência local na amplitude de tamanho  $n$  (ver Figura 2.25).

A série transformada  $y(k)$  é subtraída de  $y_n(k)$  em cada amplitude de tamanho  $n$ . Com base na seguinte expressão:

$$F_{DFA}(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (2.24)$$

O cálculo anterior deve ser repetido sistematicamente para diferentes amplitudes de tamanho  $n$  ( $n$  mínimo igual a 4 a um máximo de  $N/4$ ). O valor do  $n$  mínimo e  $n$  máximo foi definido empiricamente.



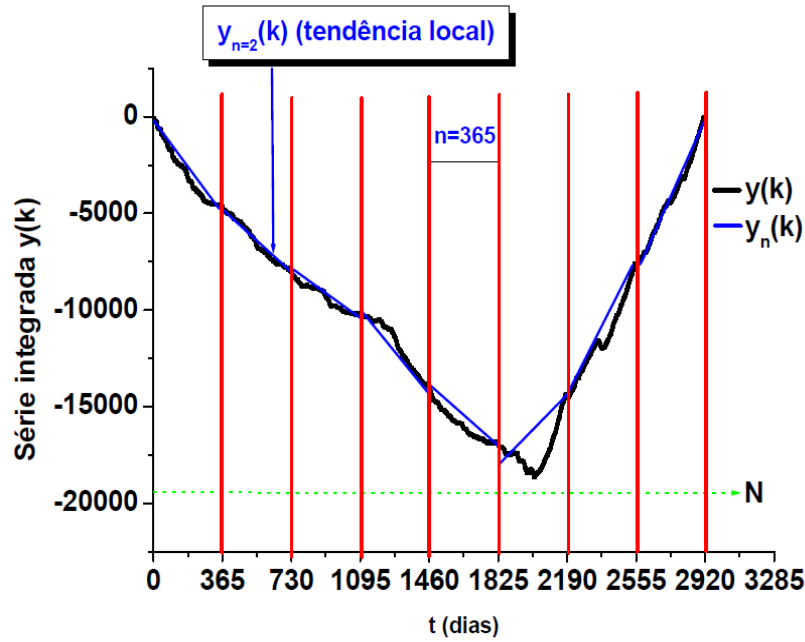


Figura 2.25: Série integrada  $y(k)$  dividida em (box) de tamanho 365 e as tendências locais  $y_n(k)$ , dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010.

Passo 4: Verifica-se se  $F_{DFA}(n)$  tem comportamento do tipo  $F_{DFA}(n) \sim n^\alpha$  neste caso  $\alpha$  será o expoente de correlação de longo alcance (ver Figura 2.26 e expressões 2.25 e 2.26).

$$F_{DFA}(n) = an^\alpha \quad (2.25)$$

Aplicado log na expressão 2.25 temos:

$$\log F_{DFA}(n) = \log a + \alpha \log n \quad (2.26)$$

Para séries não correlacionadas espera-se  $\alpha = 0,50$ , caso típico do *Random Walk*. Em contra partida, se  $\alpha > 0,50$  pode-se inferir que a série apresenta persistência de longo alcance e se o expoente  $\alpha < 0,50$ , espera-se um comportamento antipersistente de longo alcance. Através do expoente  $\alpha$  obtido por meio do DFA (ver Figura 2.26) é possível avaliar em que medida a tendência observada na série temporal passada implica em manutenção do comportamento no futuro, indicando um efeito de memória de longa duração na série.

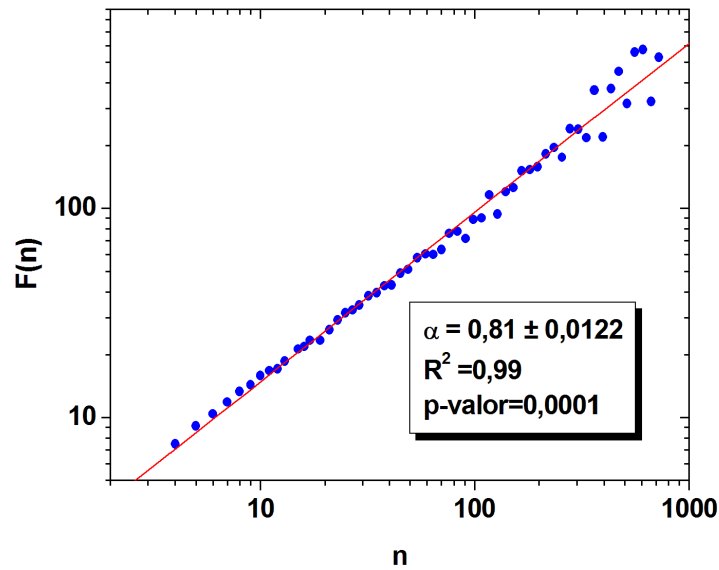


Figura 2.26: Comportamento do expoente  $\alpha$  dos acidentes de trânsito sem lesão registrados diariamente em Salvador-BA, 2003 a 2010.

O expoente  $\alpha$  de autocorrelação do método DFA apresenta campo de variação diferente do expoente  $H$  de Hurst com as seguintes características (GALHARDO et al., 2009):

- um  $\alpha > 1$  denota não estacionariedade do sinal e um comportamento sub difusivo;
- um  $\alpha = 1,5$  a série apresenta um comportamento difusivo;
- $\alpha > 1,5$  indica um comportamento super difusivo.

Devido a sua robustez e aplicações em diversas áreas do conhecimentos o DFA motivou pesquisadores a explorá-lo em relação a não-estacionariedade, tendências polinomiais e aos efeitos de tamanho finito (GALHARDO, 2010). E é possível encontrar na literatura vários trabalhos comparando o DFA com outros métodos, como o de (SHAO et al., 2012).

Na próxima subseção abordaremos o *Detrended Cross-Correlation Analysis*-DCCA que é uma generalização do DFA.

### 2.4.2 Método DCCA-Detrended Cross-Correlation Analysis

Os métodos apresentados até aqui tem como função caracterizar a autocorrelação de longo alcance de um sistema. Mas há casos que o resultado de determinada série sofre interferência não só dela mesma mais também de outras séries e é neste contexto que o estudo da correlação cruzada se faz necessário. E, como exemplo, é possível citar a área financeira, que o risco é calculado com base nas matrizes de correlação cruzada para diferentes ativos (PODOBNIK et al., 2011).

Caso exista interesse do pesquisador em identificar correlação cruzada entre duas séries temporais em regime não estacionário com mesmo número de observações  $N$ , pode-se aplicar uma generalização do método DFA, denominado *Detrended Cross-Correlation Analysis*-DCCA (PODOBNIK; STANLEY, 2008). O DCCA é um método estatístico capaz de estimar o expoente que caracteriza a correlação de longo alcance entre duas séries temporais, em regime não estacionário. E segundo (ZEBENDE; MACHADO, 2009a) o DCCA é capaz de identificar componentes sazonais ou periodicidade, ver (MACHADO, 2009), (VASSOLER; ZEBENDE, 2011), (ZHAO et al., 2011) e (MACHADO; SILVA; ZEBENDE, 2014).

O DCCA tem sido aplicado em diversas áreas das ciências como, por exemplo:

- Em séries financeiras (PODOBNIK; STANLEY, 2008), (ZHOU, 2008), (PODOBNIK et al., 2009b), (SIQUEIRA et al., 2010), (HE; CHEN, 2011a), (LIN; SHANG; ZHAO, 2012)
- Em sinais biológicos (URSULEAN; LAZAR, 2009);
- Em fluxo de tráfego (XU; SHANG; KAMAE, 2010), (ZHAO et al., 2011);
- Na dinâmica temporal de indicadores de criminalidade (MACHADO, 2009), (MACHADO; ZEBENDE; SILVA, 2011), (MACHADO; SILVA; ZEBENDE, 2014);
- No estudo de séries econômicas (ZEBENDE; MACHADO, 2009a), (ZEBENDE; SILVA; MACHADO, 2011);
- No estudo de fenômenos físicos (SHADKHO; JAFARI, 2009);
- Na análise de características climáticas (PODOBNIK; STANLEY, 2008), (HAJIAN; MOVAHED, 2010), (VASSOLER; ZEBENDE, 2011), (HORVATIC; STANLEY; PODOBNIK, 2011);
- Em testes de aleatoriedade de séries temporais simuladas (ZEBENDE; MACHADO, 2009b) e em outras áreas do conhecimento.

Segundo (PODOBNIK; STANLEY, 2008), de posse de duas séries temporais  $\{y_i\}$  e  $\{y'_i\}$  (ver Figura 2.27) com o mesmo número de pontos ( $N$ ), devemos seguir os seguintes passos:

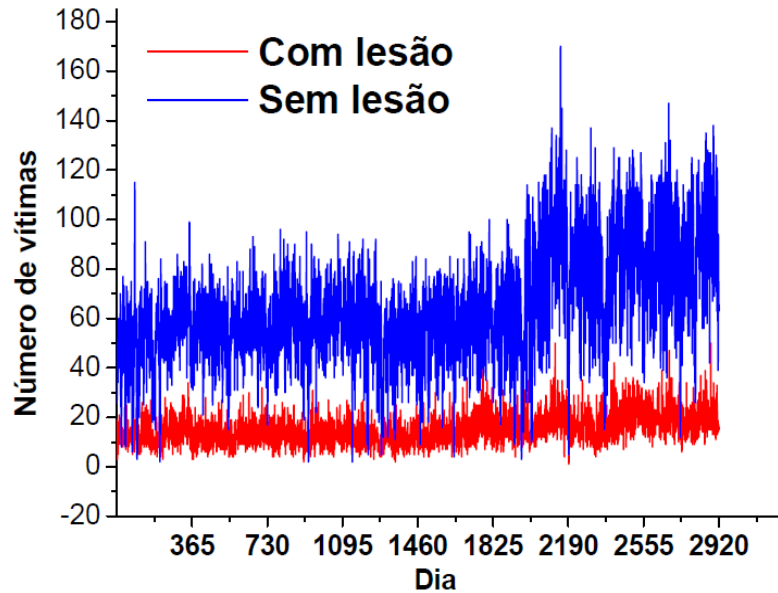


Figura 2.27: Acidentes de trânsito sem lesão e com lesão registrados diariamente em Salvador-BA, 2003 a 2010. Fonte: TRANSALVADOR. Nota: Cálculo do autor.

Passo 1: Integram-se duas séries  $\{y_i\}$  e  $\{y'_i\}$  por meio das seguintes expressões:

$$\begin{aligned} R_K &= y_1 + y_2 + y_3 + \cdots + y_k \\ R'_K &= y'_1 + y'_2 + y'_3 + \cdots + y'_k \end{aligned} \quad (2.27)$$

em que  $K = 1, 2, \dots, N$

Passo 2: Divide-se os sinais integrados  $R_K$  e  $R'_K$  em  $(N - n)$  *boxes* (com superposição) de tamanho  $n$ , cada um contendo  $n + 1$  valores (ver Figura 2.28). Para as duas séries temporais, em cada box (que inicia em  $i$  e termina em  $i + n$ ) calculamos as tendências em cada box, isto é,  $\tilde{R}_{K,i}$  e  $\tilde{R}'_{K,i}$  ( $i \leq K \leq i + n$ ) como sendo a ordenada do ajuste linear (processo dos mínimos quadrados) da série somada.

Passo 3: Calcula-se a partir do cálculo anterior a covariância dos resíduos em cada intervalo, ou seja,

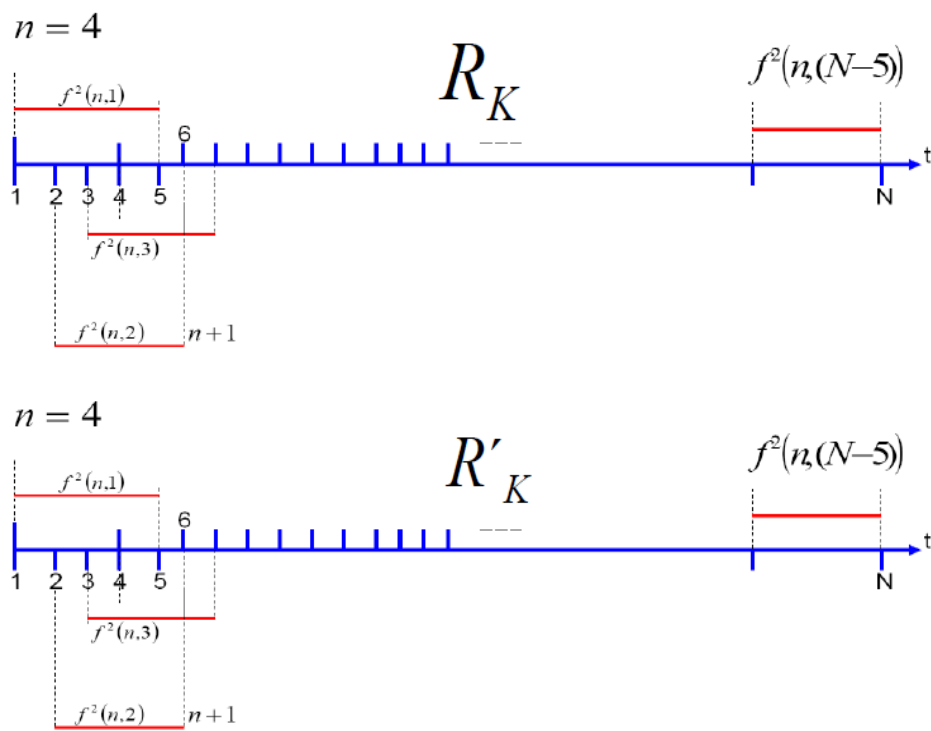


Figura 2.28: Divisão dos sinais integrados  $R_K$  e  $R'_K$  em  $N - n$  boxes (com superposição) de tamanho  $n$ , cada um contendo  $n + 1$  valores. Fonte: (MACHADO, 2009).

$$f_{DCCA}^2(n, i) = \frac{1}{(n+1)} \sum_{K=i}^{i+n} (R_K - \tilde{R}_{K,i})(R'_K - \tilde{R}'_{K,i}) \quad (2.28)$$

Passo 4: Calcula-se a função de correlação (detrended covariance) expressa como:

$$F_{DCCA}^2(n) \equiv \frac{1}{N-n} \sum_{i=1}^{N-n} f_{DCCA}^2(n, i) \quad (2.29)$$

Aplicando o procedimento anterior a uma série temporal o DCCA será uma generalização do DFA (PODOBNIK; STANLEY, 2008). Neste caso, se o procedimento anteriormente explicado for repetido para diferentes tamanhos de  $n$  sistematicamente, é possível constatar a existência ou não de uma lei de potência (ver Figura 2.29 e 2.30), ou seja,

$$F_{DCCA} \sim n^\lambda \quad (2.30)$$

E  $\lambda$  na expressão 2.30 quantifica a correlação cruzada. De acordo com (PODOBNIK; STANLEY, 2008), em geral  $\lambda$  tende a ser a média dos expoentes do DFA:

$$\lambda \cong \frac{\alpha_1 + \alpha_2}{2}. \quad (2.31)$$

Segundo (ZEBENDE; SILVA; MACHADO, 2011) uma das peculiaridades do método DCCA em relação a outros métodos de detecção de correlação cruzada, é o fato do mesmo retirar as tendências polinomiais que podem mascarar as verdadeiras correlações cruzadas. A seguir apresentaremos o coeficiente de correlação cruzada  $\rho_{DCCA}$  capaz de quantificar o nível de correlação cruzada.

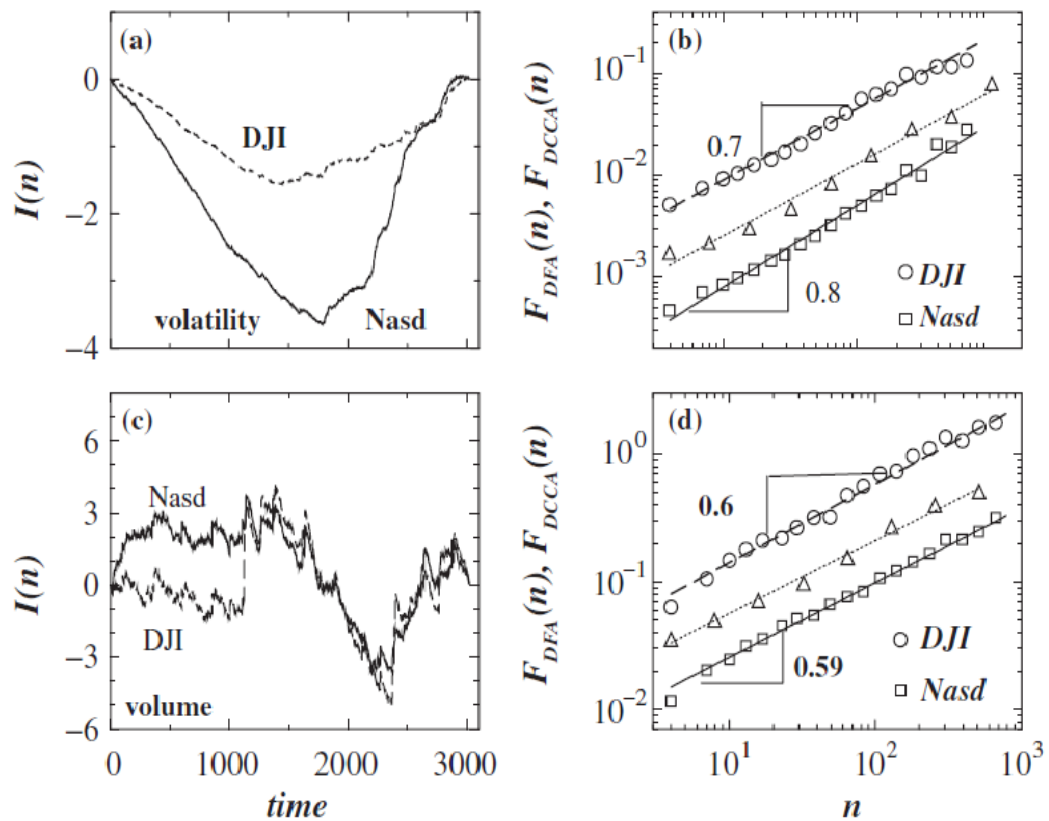


Figura 2.29: Autocorrelação individual (DFA) e cruzada (DCCA) em valores absolutos de mudanças de preços (volatility) e os volumes negociados (volume) tanto pelos índices da Dow Jones (DJI) e da Nasdaq (Nasd), registrados diariamente, no período de julho de 1993 à novembro de 2003. Fonte: (PODOBNIK; STANLEY, 2008).

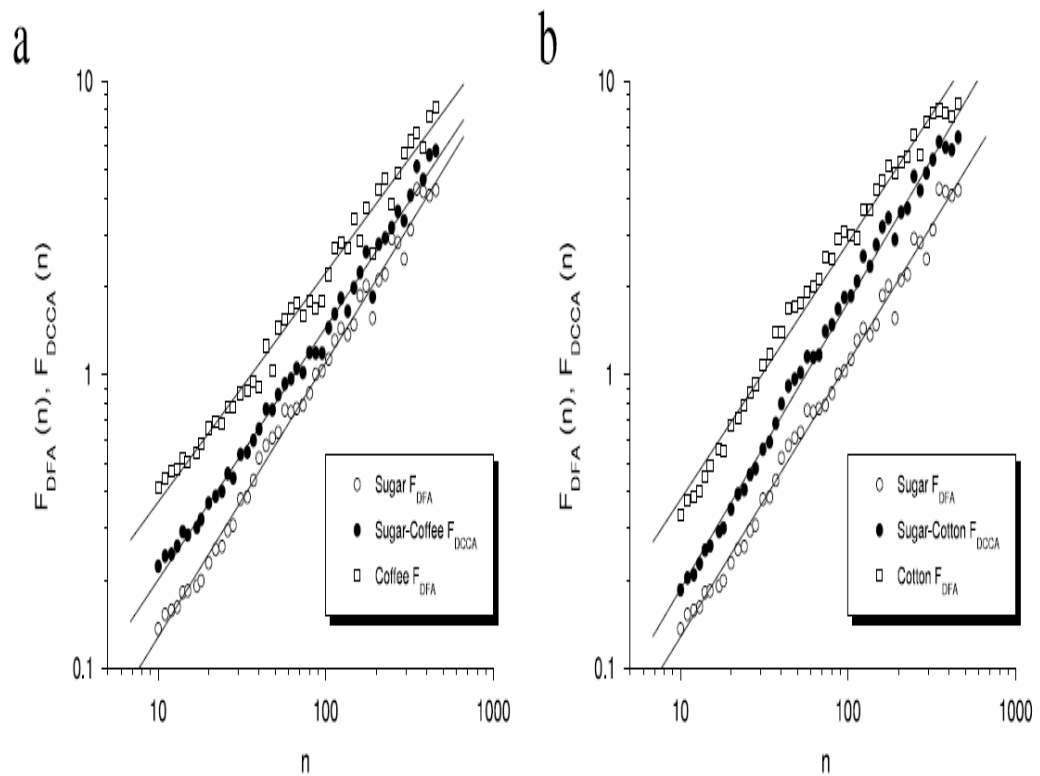


Figura 2.30: Autocorrelação e correlação cruzada: a) séries temporais volatilidade do ( $\circ$ ) açúcar ( $F_{DFA}$ ), ( $\bullet$ ) açúcar  $\times$  café ( $F_{DCCA}$ ) e ( $\square$ ) café ( $F_{DFA}$ ). b) séries temporais volatilidade do ( $\circ$ ) açúcar ( $F_{DFA}$ ), ( $\bullet$ ) açúcar  $\times$  algodão ( $F_{DCCA}$ ) e ( $\square$ ) algodão ( $F_{DFA}$ ). Fonte: (SIQUEIRA et al., 2010).



### 2.4.3 Coeficiente de correlação cruzada $\rho_{DCCA}$

O  $\rho_{DCCA}$  modelado por (ZEBENDE, 2011), é um coeficiente capaz de quantificar o nível de correlação cruzada tendo como base o DFA e o DCCA. O expoente  $\lambda$  obtido através do DCCA quantifica a correlação cruzada como lei de potência, porém não quantifica o nível de correlação cruzada. O  $\rho_{DCCA}$  destina-se à estimação do coeficiente de correlação cruzada em diferentes escalas de tamanho  $n$ , cujo coeficiente possui sua variação limitada entre -1 a +1 inclusive ( $-1 \leq \rho_{DCCA} \leq 1$ ) (ver Figura 2.31). Em que -1 representa anticorrelação perfeita e +1 correlação perfeita. E o valor do  $\rho_{DCCA}$  igual a zero significa que não existe correlação cruzada. É importante relatar que o campo de variação do  $\rho_{DCCA}$  ( $-1 \leq \rho_{DCCA} \leq 1$ ) foi provado teoricamente por (PODOBNIK et al., 2011), tendo como método a *Desigualdade de Cauchy*.

O  $\rho_{DCCA}$  é definido como a relação entre a função de covariância sem tendência  $F_{DCCA}^2$  e a função de variância sem tendência  $F_{DFA}$ , que pode ser expressa pela seguinte expressão:

$$\rho_{DCCA}(n) \equiv \frac{F_{DCCA}^2(n)}{F_{DFA1}(n)F_{DFA2}(n)} \quad (2.32)$$

Apesar de ter sido proposto recentemente, o  $\rho_{DCCA}$  já foi modelado em diversos campos da ciência. É possível citar:

■ Em fenômenos climáticos (VASSOLER; ZEBENDE, 2011) (ver Figura 2.31), (BALOCCHI; VARANINI; MACERRATA, 2013), (KANG et al., 2013), (YUAN; FU, 2014);

■ No estudo do mercado financeiro (PODOBNIK et al., 2011), (WANG; XIE, 2013), (GU; SHAO; WANG, 2013), (FENG; WEI; HUANG, 2013), (ZEBENDE; SILVA; MACHADO, 2013), (WANG et al., 2013), (ALAOUI; BENBACHIR, 2013), (LIU; CHEN; WAN, 2013), (REBOREDO; RIVERA-CASTRO; ZEBENDE, 2014), (WANG et al., 2014), (FERNANDES; CASTRO; MACHADO, 2014), (YINA; SHANG, 2014);

■ Análise de dados geofísicos (MARINHO; SOUSA; ANDRADE, 2013);

■ Na análise de séries temporais de acidentes de trânsito (MACHADO et al., 2012);

■ Na dinâmica dos movimentos de deglutição de um indivíduo (CASTRO, 2012), (CASTRO; MACHADO; ZEBENDE, 2013);

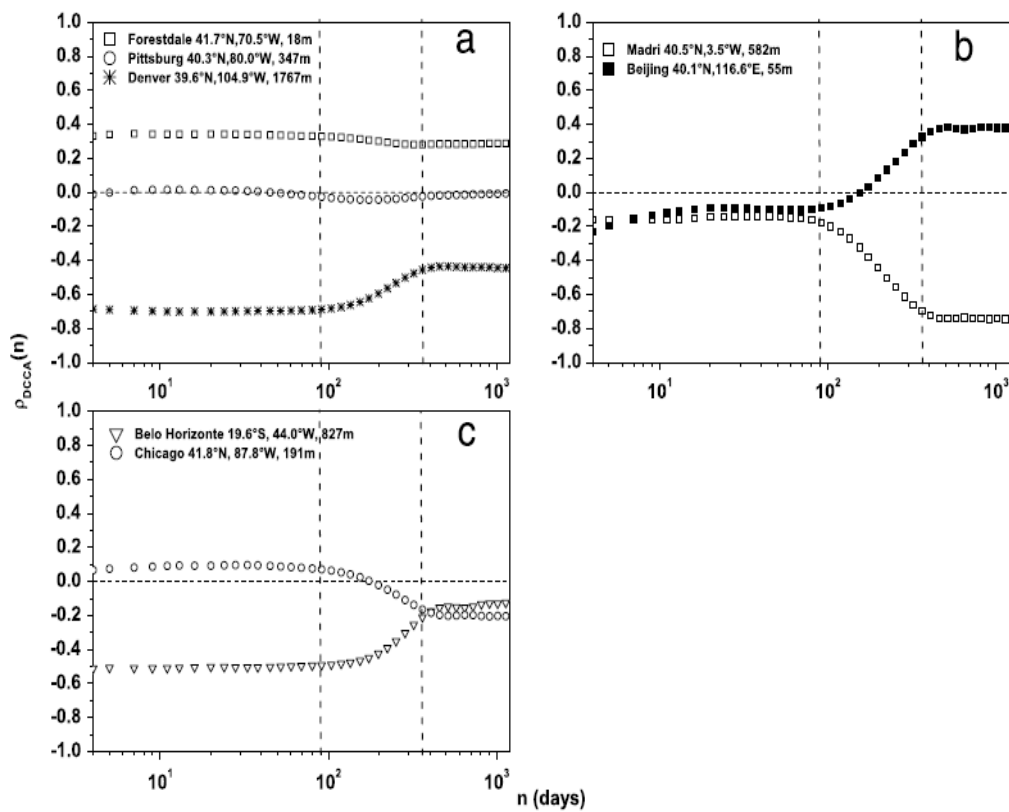


Figura 2.31:  $\rho_{DCCA}$  entre a temperatura do ar x humidade relativa: a) corresponde aproximadamente a 40° de latitude (EUA), b) corresponde a cerca de 40° de latitude de outros países e c) outras latitudes. As linhas verticais denotam 90 e 365 dias respectivamente. Fonte: (VASSOLER; ZEBENDE, 2011)

■ Em séries históricas de indicadores de criminalidade (MACHADO; CASTRO; ZEBENDE, 2013), (MACHADO; SILVA; ZEBENDE, 2014).

Com objetivo de testar a robustez estatística do  $\rho_{DCCA}$ , foi proposto por (PODOBNIK et al., 2011) um teste estatístico para avaliar a significância dos valores obtidos pelo  $\rho_{DCCA}$  (Equação 2.32). O referido teste estatístico tem como base duas séries temporais independentes e identicamente distribuídas (iid), em um processo Gaussiano com média zero ( $\mu = 0$ ) e variância igual a 1 ( $\sigma^2 = 1$ ), com  $\alpha_1 = \alpha_2 = 0,5$  (séries aleatórias) com 95 por cento de confiança. E o teste tem as seguintes hipóteses (ver Figura 2.32):

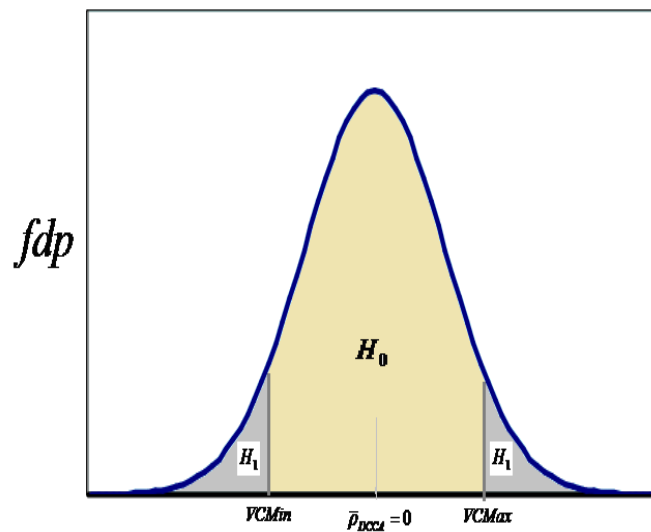


Figura 2.32: Hipóteses estatísticas do teste proposto por (PODOBNIK et al., 2011) para os valores obtidos pelo  $\rho_{DCCA}$ . Aqui  $fdp$  denota função de densidade de probabilidade e  $VCMin$  - valor crítico mínimo e  $VCMMax$  - valor crítico máximo.

- $H_0 : \rho_{DCCA} = 0$  (não existe correlação cruzada significativa);
- $H_1 : \rho_{DCCA} \neq 0$  (existe correlação cruzada significativa).

aqui  $H_0$  denota a hipótese nula ou hipótese de igualdade e  $H_1$  hipótese alternativa.

O teste estatístico supracitado determina a variação dentro da qual a correlação pode ser considerada estatisticamente significativa ou não. O teste estatístico do  $\rho_{DCCA}$  levou em consideração também o tamanho da série  $N$  e o tamanho  $n$  do *box* (janela) (ver Figura 2.33, Figura 2.34 e tabelas do apêndice A).

Os resultados das simulações do teste estatístico do  $\rho_{DCCA}$  realizados nesta pesquisa corroboraram com os de (PODOBNIK et al., 2011). E maiores detalhes sobre o teste podem ser encontrados em (PODOBNIK et al., 2011), (BLYTHE, 2013), (WANG; XIE, 2013), (KANG et al., 2013), (KRISTOUFEK, 2013) e (FERNANDES; CASTRO; MACHADO, 2014).

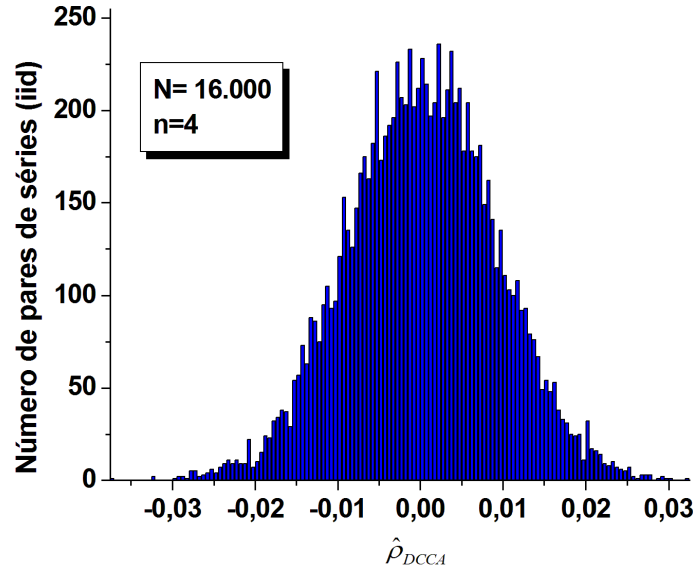


Figura 2.33: Simulação para o teste estatístico da equação 2.32 com 10.000 pares de séries iid com  $N = 16.000$  pontos e o tamanho do *box*  $n = 4$ .

Com o objetivo de contribuir com o estudo teórico do  $\rho_{DCCA}$ , encontra-se no apêndice A tabelas com os valores críticos do  $\rho_{DCCA}$  com 90 e 95 por cento de confiança para pares de séries independentes e identicamente distribuídas (iid) para *box* de diferentes tamanhos  $n$  e séries de diferentes tamanho  $N$ . Tendo como base a distribuição amostral das médias e o Teorema do Limite Central, os valores críticos do  $\rho_{DCCA}$  foram definidos pela seguinte expressão:

$$\bar{\rho}_{DCCA} \pm Z\sigma_{\rho_{DCCA}} \quad (2.33)$$

em que:

$\bar{\rho}_{DCCA}$  denota a média dos  $\rho_{DCCA}$ ;

$Z$  é o quantil que define o nível de significância (5 por cento, 10 por cento ...);

$\sigma_{\rho_{DCCA}}$  o desvio padrão das séries dos  $\rho_{DCCA}$ .

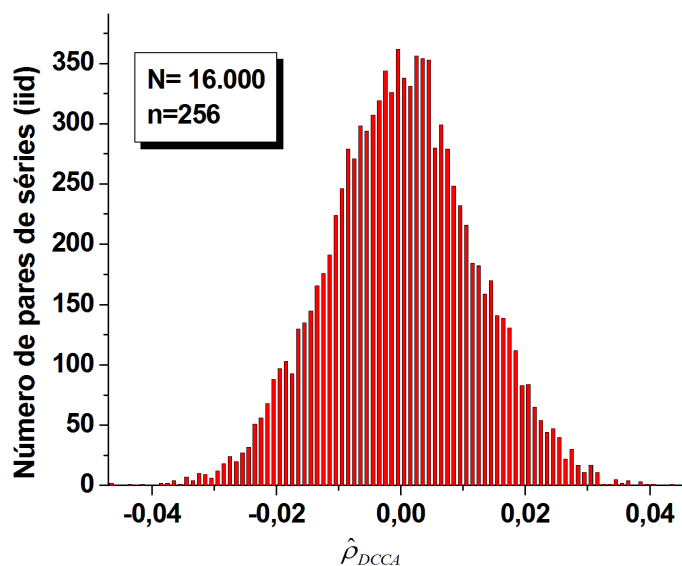


Figura 2.34: Simulação para o teste estatístico da equação 2.32 com 10.000 pares de séries iid com  $N = 16.000$  pontos e o tamanho do *box*  $n = 256$ .

Em relação aos conceitos e procedimentos de cálculo utilizados no teste estatístico do  $\rho_{DCCA}$  como Teorema do Limite Central, Distribuição amostral das médias e de outros estimadores é possível consultar (BUSSAB; MORETTIN, 2003), (BOLFARINE; BUSSAB, 2005), (TRIOLA, 2008), (BOLFARINE; SANDOVAL, 2010) e (MORETTIN, 2010).

Ainda com objetivo de contribuir com o aprimoramento do  $\rho_{DCCA}$ , modelamos um programa capaz de executar o teste estatístico do  $\rho_{DCCA}$  idealizado por (PODOBNIK et al., 2011) para diferentes pares de séries de tamanho  $N$  e diferentes escalas de tempo de tamanho  $n$  e diferentes magnitudes de intervalos de confiança (90 por cento, 95 por cento, ...). E encontra-se no apêndice B.2 o pedido de registro de programa de computador ao INPI (Instituto Nacional da Propriedade Industrial) do referido teste estatístico com co-autoria do SENAI CIMATEC.

Uma das vantagens do coeficiente de correlação cruzada  $\rho_{DCCA}$  em relação a outros coeficientes de correlação cruzada como o coeficiente de correlação de Pearson, é a possibilidade de mensurar correlação entre dois sinais em diferentes escalas de tempo  $n$  (WANG; XIE, 2013). E, além disso, segundo (MACHADO; SILVA; ZEBENDE, 2014), a análise da correlação entre os sinais é realizada sem a componente tendência que tende a mascarar as verdadeiras correlações. Outra vantagem em relação a o coeficiente de Pearson é a capacidade de analisar séries não lineares. E recentemente os resultados estatísticos obtidos pelo  $\rho_{DCCA}$  foram comparados com o coeficiente de Pearson nas séries temporais do mercado de ações dos Estados Unidos por (WANG et al., 2013). E os resultados obtidos pelo  $\rho_{DCCA}$  apresentaram diversas propriedades interessantes em diferentes escalas de tempo  $n$  no mercado de ações dos Estados Unidos que segundo (WANG et al., 2013), são úteis para o gerenciamento de riscos e seleção ótima de portfólio, especialmente para diversidade do portfólio de ativos.

Como já foi mencionado anteriormente (ver seção 2.1.4), o  $\rho_{DCCA}$  em sua construção leva em consideração a ordem dos pares de valores das séries temporais o que não contraria o princípio básico da análise das séries temporais que é a dependência temporal entre os valores da série fato que não estar presente em alguns coeficiente de correlação cruzada como o coeficiente de correlação de Pearson.

No próximo capítulo serão apresentados os resultados desta pesquisa de Doutorado, e na subseção 3.2.1 é provado que existe uma relação estatística por diferenciação entre o coeficiente de correlação cruzada  $\rho_{DCCA}$  e os expoentes dos métodos DFA e DCCA.

## Resultados

Neste capítulo, são apresentados os resultados contemplados por esta tese de doutorado relacionados com os objetivos listados na mesma (ver Capítulo 1). Fizemos a opção metodológica de expor os resultados em formato de artigo. Nas seções 3.1, 3.2 e 3.3 são apresentados três artigos científicos que foram publicados no periódico *Physica A: Statistical Mechanics and its Applications*.

Os artigos apresentados no presente capítulo têm como característica comum o estudo teórico e a aplicação de métodos capazes de detectar e mensurar autocorrelação e correlação cruzada em séries temporais não estacionárias em diferentes escalas de tempo enquanto lei de potência.

O primeiro artigo (seção 3.1) tem como propósito o estudo da autocorrelação e correlação cruzada nas séries temporais de acidentes de táxi registrados diariamente por uma cooperativa de táxi denominada Centro Especial e Auxílio ao Taxista - CEAT no período de 18 de agosto de 2002 a 21 de maio de 2010. Para este fim, foi aplicado, principalmente, métodos conhecidos na literatura como métodos da mecânica estatística o DFA-*Detrended Fluctuation Analysis* (PENG et al., 1994), o DCCA - *Detrended Cross-Correlation Analysis* (PODOBNIK; STANLEY, 2008). Entre os três artigos apresentados neste capítulo o primeiro artigo (seção 3.1) é o único que não aplicou o  $\rho_{DCCA}$  implementado por (ZEBENDE, 2011) devido ao fato dele ter dado origem ao denominado coeficiente de correlação cruzada  $\rho_{DCCA}$ . Ao aplicar a função de covariância sem tendência nos dados de acidentes de táxi registrados pela CEAT, foi constatado que o método DCCA identifica e mensura a correlação cruzada enquanto lei de potência, porém não quantifica o nível dessa correlação cruzada como o coeficiente de correlação Pearson. E também a partir do estudo realizado no primeiro artigo foi possível definir via função de covariância sem tendência do método DCCA três grupos distintos de correlação cruzada: de longo alcance, de curta duração e um grupo que não há correlações cruzadas.

O segundo artigo (seção 3.2) é teórico e tem como objetivo estabelecer uma relação estatística entre o expoente de autocorrelação do método DFA e o expoente de correlação cruzada do método DCCA por diferenciação do  $\rho_{DCCA}$ . Segundo (PODOBNIK; STANLEY, 2008), existe a seguinte relação estatística entre o expoente  $\alpha$  de autocorrelação do método DFA e expoente  $\lambda$  de correlação cruzada do método DCCA:

$$\lambda \cong \frac{\alpha_1 + \alpha_2}{2}. \quad (3.1)$$

Porém, outros pesquisadores identificaram resultados diversos da (Equação 3.1) como (ZHOU, 2008), (CAO; XU; CAO, 2012), (HE; CHEN, 2011a) e (HE; CHEN, 2011b). Entretanto, não foi encontrado na literatura um trabalho que defina uma relação estatística bem estabelecida entre os expoente da equação 3.1. Diante dessa realidade, no segundo artigo (seção 3.2) é definida uma relação estatística entre os supracitados expoentes via diferenciação do  $\rho_{DCCA}$ . Assim como o trabalho (PODOBNIK; STANLEY, 2008) para validar a teoria foram escolhidas séries simuladas de um processo ARFIMA e séries reais do mercado financeiro internacional.

O terceiro artigo (seção 3.3) tem como objetivo estabelecer uma relação entre os homicídios dolosos e homicídios tentados do ponto de vista das autocorrelações e correlações cruzadas. Esta relação foi estabelecida tendo como métodos de análise o *Detrended Fluctuation Analysis* (DFA) e *Detrended Cross-Correlation Analysis* (DCCA), e pelo coeficiente de correlação cruzada  $\rho_{DCCA}$ . Como dados experimentais escolhemos os homicídios dolosos e homicídios tentados registrados diariamente na capital baiana no período de 2004 a 2011. Os resultados obtidos com a modelagem do  $\rho_{DCCA}$  nos dados foram testados estatisticamente com o teste estatístico proposto por (PODOBNIK et al., 2011) para todas as escalas de tempo consideradas. O artigo apresentado na seção 3.3 não tem a pretensão de discutir e modelar as variáveis socioeconômicas associadas a criminalidade como educação, saúde e renda. Após a aplicação dos supracitados métodos no terceiro artigo (seção 3.3) identificamos propriedades interessantes em diferentes escalas de tempo nos delitos escolhidos que poderão servir como objeto de análise para os gestores e pesquisadores da área criminal.

Nas seções subsequentes deste capítulo serão apresentados os resultados desta pesquisa de doutorado que proporcionaram implementações teóricas nos métodos da mecânica estatística: DFA, DCCA e  $\rho_{DCCA}$ .



### 3.1 *Estudo das correlações cruzadas em séries temporais auto afins de acidentes de táxis*

#### Resumo

Estudamos neste artigo a correlação cruzada em séries temporais auto afins de variáveis reais registradas simultaneamente em casos de acidentes de táxi. Com esse propósito, aplicamos o método DCCA e foi mostrado que a correlação cruzada pode ser dividida em três grupos distintos, se analisarmos as funções de covariância sem tendência, isto é as correlações cruzadas de longo alcance, curto alcance e sem correlação cruzada. Nesta pesquisa encontramos fortes evidências que o DCCA é um método robusto se comparado com outros métodos que tem como objetivo identificar estes tipos de correlações cruzadas

1 2 3

Frequentemente as séries temporais exibem comportamento complexo caracterizados pela correlação de lei de potência de longo alcance (PENG et al., 1992), (VOSS, 1992), (STANLEY et al., 1994), (ZEBENDE; OLIVEIRA; PENNA, 1998). Estas séries temporais podem ser observadas usando registros de tempo ou série de observações. Com o objetivo de estudar as séries de tempo, a função de autocorrelação é um possível método para análise de séries temporais. A autocorrelação é uma medida que informa quanto o valor de uma realização ou observações é capaz de influenciar valores ou observações vizinhas. Entretanto, a maioria das séries temporais reais não são estacionárias, assim o desvio padrão médio, momentos de pico, ou as funções da autocorrelação não são constantes em função do tempo (STRATONOVICH, 1981), (WITT; KURTHS; PIKOVSKY, 1998), (CHEN et al., 2002).

---

<sup>1</sup>O presente artigo foi publicado no periódico Physica A no volume 390 em 2011.

<sup>2</sup>Agradecemos a CEAT pelos dados fornecidos e a FAPESB pelo apoio financeiro.

<sup>3</sup>Foi realizado um pedido de registro de programa de computador ao INPI (Instituto Nacional da Propriedade Industrial) em 22 de julho de 2013 do programa que aplica os métodos DFA e DCCA utilizados neste artigo com co-autoria do SENAI CIMATEC (ver Apêndice B.1).

A estacionariedade é um aspecto relevante da variabilidade complexa e pode ser frequentemente associada a diferentes tendências no sinal ou segmentos heterogêneos com propriedades estatísticas locais diferentes. Para lidar com este problema o DFA e o DCCA foram implementados para séries temporais não estacionárias.

### 3.1.1 *Objetivos e dados*

Sabemos que o táxi é um veículo de grande uso diário, se comparado aos veículos particulares, e este fato é uma das razões por que as seguradoras cobram valor mais alto pelo seguro, que denomina-se premium, tornando-os mais caros do que os veículos particulares. No caso de cooperativas seguradas o custo deste serviço é menor pelo fato de não ter como objetivo o lucro. O objetivo deste artigo é estudar a correlação cruzada entre as séries temporais de algumas variáveis relacionadas a acidentes de táxi registrados pela CEAT - Centro Especial de Auxílio ao Taxista (uma cooperativa com 575 associados que representa aproximadamente oito por cento da frota de Salvador-BA), localizada na cidade de Salvador, Bahia, Brasil. Nesta pesquisa, analisamos cinco variáveis a partir do banco de dados da CEAT (ver Figura 3.1).

Os dados foram coletados entre 18 de agosto de 2002 a 21 de maio de 2010. Para uma visualização detalhada das variáveis objeto de estudo desta pesquisa, construímos o histograma para as séries não estacionárias (Figura 3.2) (nem todas as variáveis seguem uma distribuição normal).

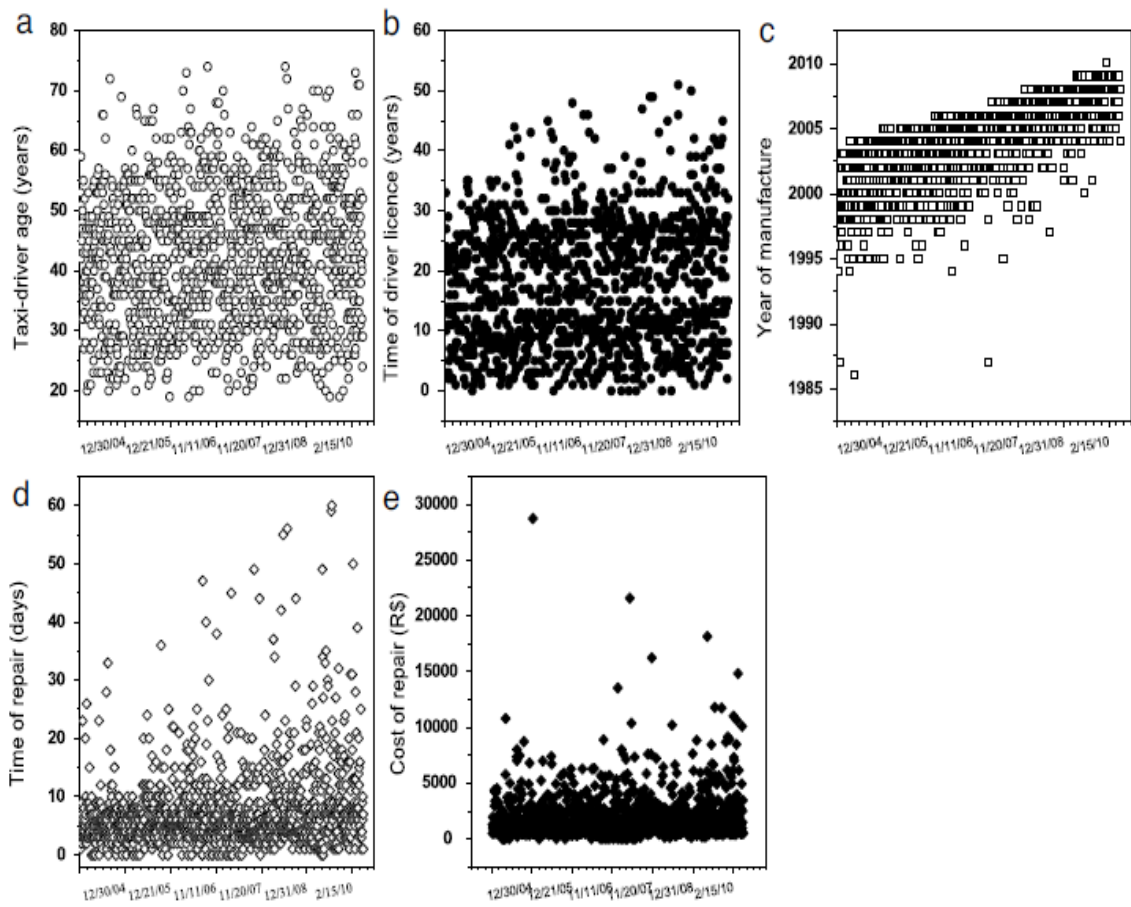


Figura 3.1: Séries temporais originais dos acidentes de táxi registrados pela CEAT:(a) (○) Idade do taxista (b)(●) Tempo da licença (c) (□) Ano de fabricação do veículo (d) (◇) Tempo de reparo (e) (◆) Custo de reparo do táxi. Fonte: (ZEBENDE; SILVA; MACHADO, 2011).

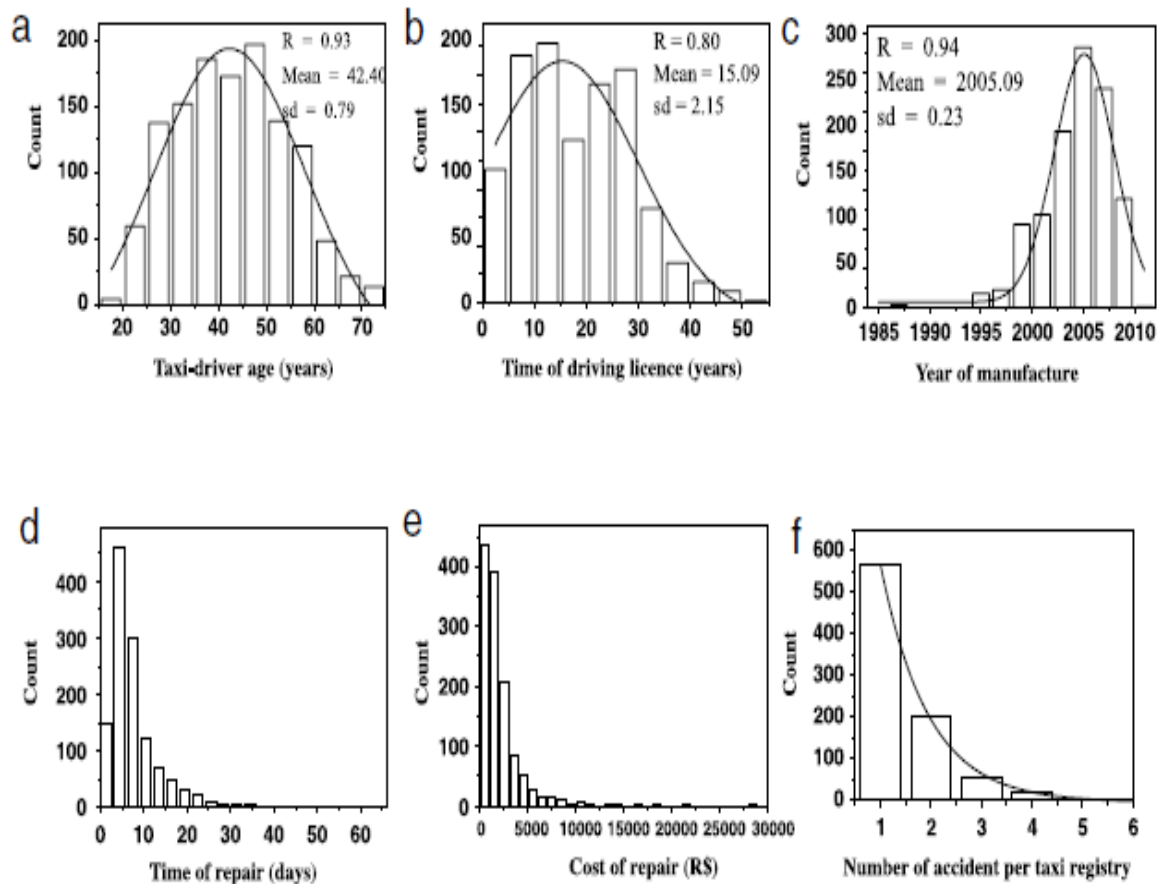


Figura 3.2: Histograma a) Idade do taxista, b) Tempo da Licença, c) Ano da fabricação, d) Tempo de reparo e e) Custo do reparo do táxi. As linhas contínuas em a), b) e c) correspondem ao um ajuste Gaussiano com R quadrado, média e erro padrão. Nos casos d) (tempo de reparo do taxi) e e) (custo de reparo do táxi), não ajustam a uma curva Gaussiana. f) representa o histograma para o número de acidentes por registro de táxi (rank de acidentes), e a linha contínua denota uma queda exponencial (somente uma curiosidade para ser visualizada). Fonte: (ZEBENDE; SILVA; MACHADO, 2011).

Um acidente de carro é um acontecimento desagradável e inesperado, que pode ser gerado por diferentes razões, tais como a negligência e imprudência do motorista, problemas nas estradas, e muitos outros fatores. Baseado nessa realidade, pode-se assumir que o acidente de carro é uma ocorrência aleatória. Com intuito de verificar se esta afirmação é verdadeira ou falsa, nesta pesquisa foi estudada a autocorrelação e a correlação cruzada pelos métodos DFA e DCCA.

### 3.1.2 Resultados e conclusões

Com o objetivo de identificar autocorrelação de longo alcance aplicamos o método DFA para as variáveis simultaneamente registradas nos acidentes de taxi da CEAT (Figura 3.3). Por meio da figura 3.3 é possível ver que  $F_{DFA}(n) \propto n^\alpha$ , com valores  $\alpha$  próximos de 0,50 (séries temporais descorrelacionadas) para a idade do taxista, o tempo da habilitação do motorista, e custo do concerto do veículo. Enquanto que para o ano da fabricação do táxi o valor de  $\alpha = 0,67$  (estes dados denotam taxa de renovação da frota de táxi). No sentido do método DFA, um acidente de táxi é um acontecimento quase aleatório. Fatores tais como cultura, economia, clima, entre outros, não geram efeitos de memória no sistema.

Se compararmos as séries objeto de estudo dessa pesquisa é possível constatar que algumas dessas séries apresentam correlação cruzada. Com a finalidade de comprovar essas correlações foi aplicado o método DCCA. Os resultados encontrados a partir do DCCA podem ser visualizados na (Figura 3.4) para as correlações cruzadas entre:

1. Idade do taxista  $\times$  tempo de licença ( $\square$ );
2. Idade do taxista  $\times$  ano de fabricação do táxi ( $\circ$ );
3. Idade do taxista  $\times$  tempo de manutenção do táxi ( $\nabla$ );
4. Idade do taxista  $\times$  custo da manutenção do táxi ( $\diamond$ );
5. Tempo da habilitação do taxista  $\times$  ano de fabricação do táxi ( $\times$ );
6. Tempo da habilitação do taxista  $\times$  época da manutenção do táxi ( $\otimes$ );
7. Tempo da habilitação do taxista  $\times$  custo da manutenção do taxi ( $\blacksquare$ );
8. Ano de fabricação do táxi  $\times$  tempo de manutenção do taxi ( $\blacktriangledown$ );
9. Ano de fabricação do táxi  $\times$  custo de reparo do táxi ( $\blacklozenge$ );

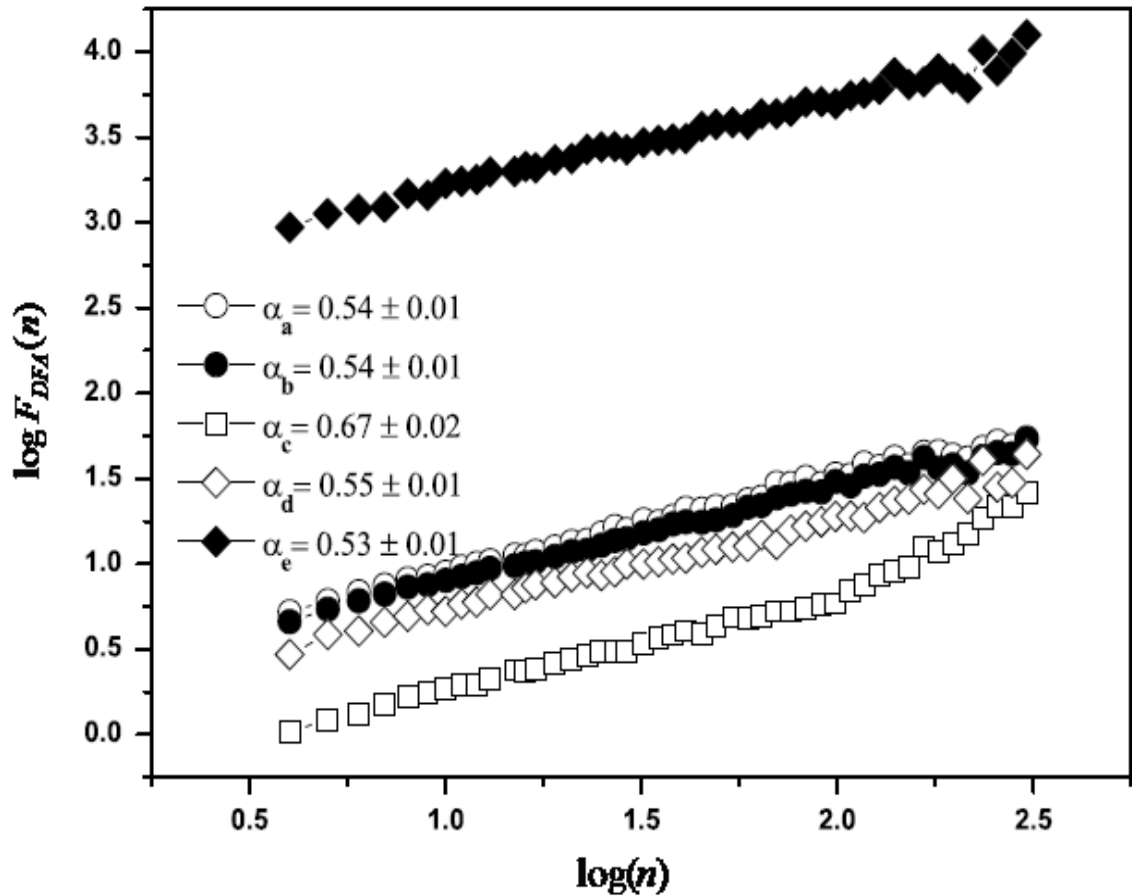


Figura 3.3: DFA aplicado em séries de acidente de taxi a) Idade do taxista  $\alpha_a$ , b) Tempo de licença  $\alpha_b$  c) Ano de fabricação do táxi  $\alpha_c$ , d) Tempo de reparo  $\alpha_d$  e) Custo do reparo do táxi  $\alpha_e$ . Para cada caso foi definido o coeficiente angular do ajuste linear (expoente  $\alpha$  com seu respectivo erro padrão). Fonte: (ZEBENDE; SILVA; MACHADO, 2011).

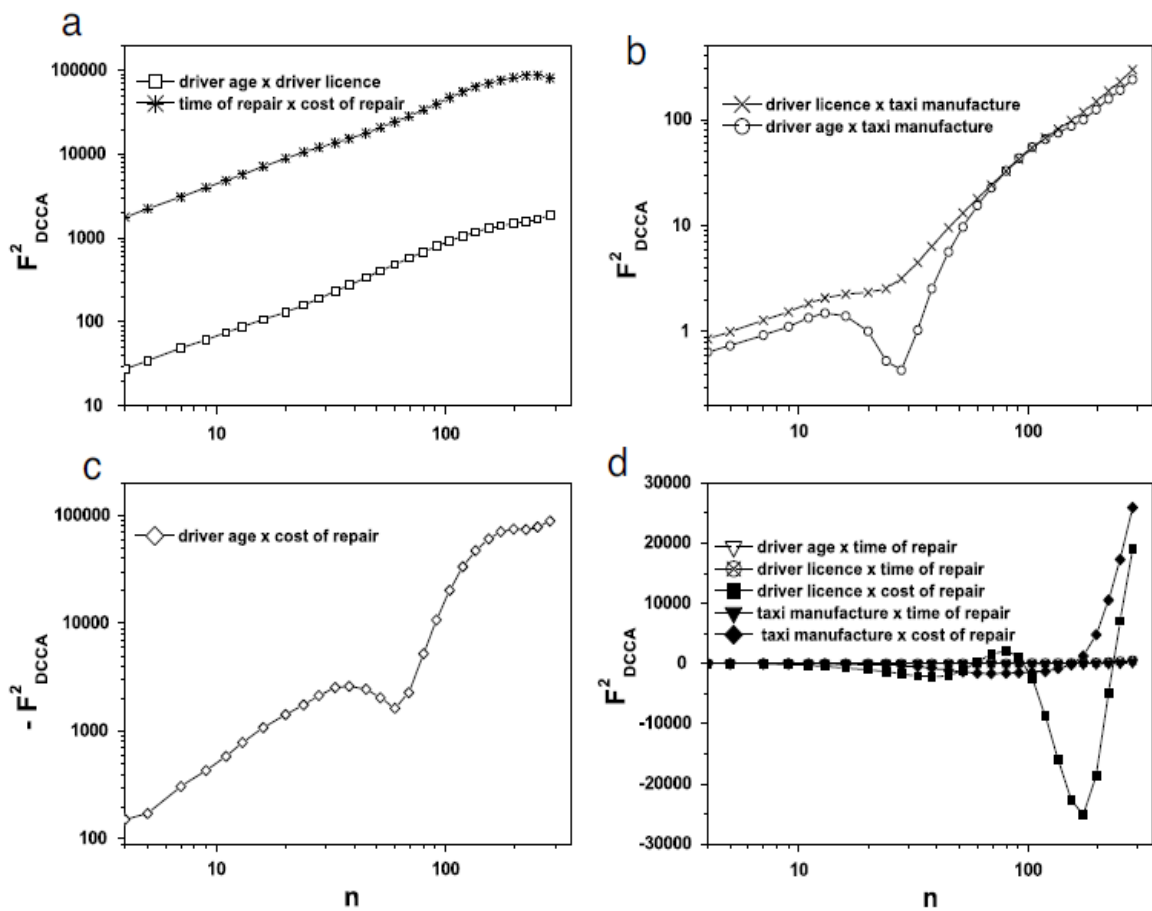


Figura 3.4: DCCA aplicado nas séries de acidentes de táxi: a) Correlação cruzada de longo alcance, b) Correlação positiva de curto alcance, c) Correlação negativa de curto alcance e d) Ausência de correlação cruzada entre as séries temporais. Fonte: (ZEBENDE; SILVA; MACHADO, 2011).

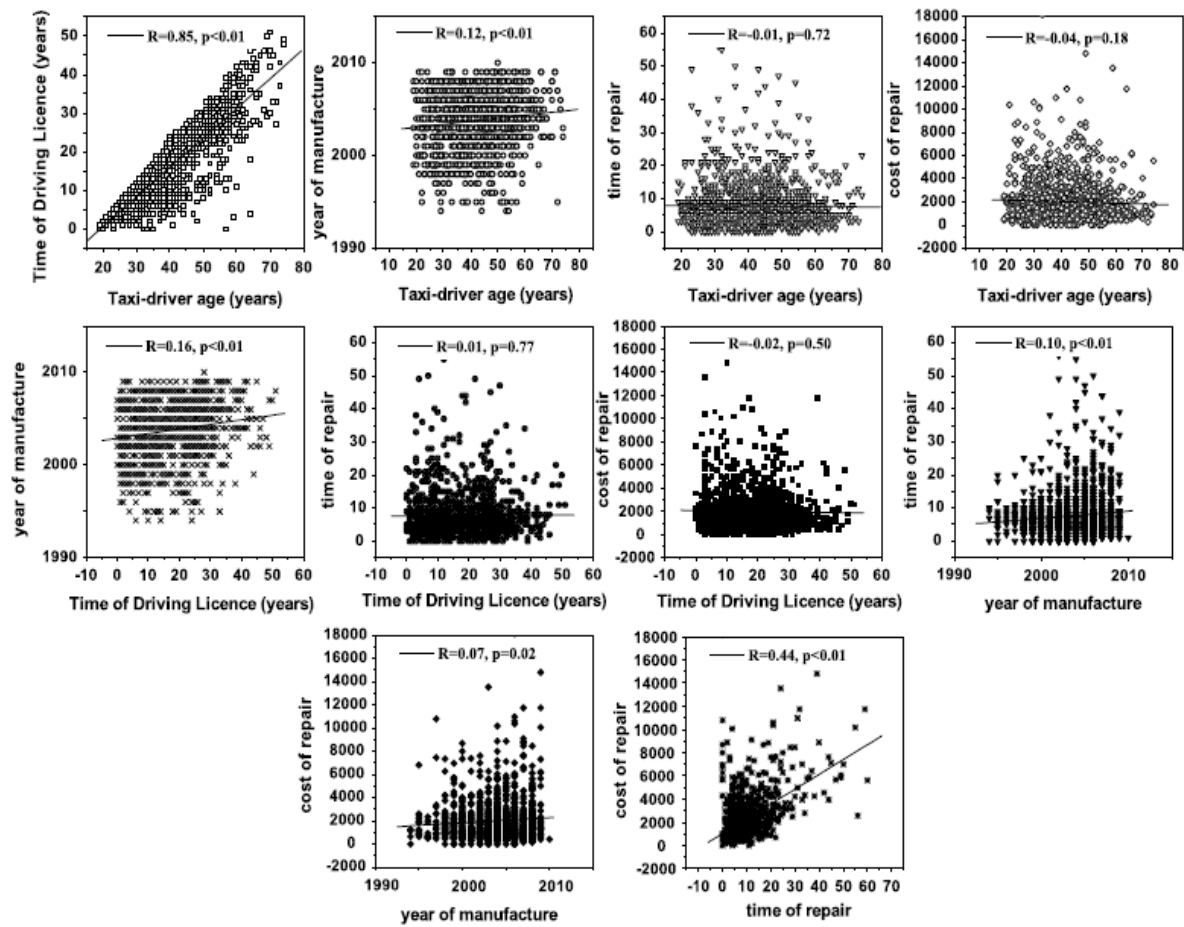


Figura 3.5: Ajuste linear (linha contínua) entre as variáveis objeto de estudo. Nota: Para cada ajuste linear foi definido o coeficiente de correlação de Pearson e o p-valor. Fonte: (ZEBENDE; SILVA; MACHADO, 2011).



10. Tempo da manutenção do táxi  $\times$  custo de manutenção do taxi (\*).

Ao analisar o comportamento das funções de covariância sem tendência na (Figura 3.4) é possível dividir os resultados em três grupos distintos:

(i) Em que as correlações de longo alcance foram apresentadas ( $F_{DCCA}^2 \propto n^{2\lambda}$ ) (Figura 3.4 (a));

(ii) Em que as correlação cruzada de curto alcance, aqui com correlações cruzadas positivas (Figura 3.4(b)) e a correlação cruzada negativa (Figura 3.4(c));

(iii) Onde a função de covariância sem tendência oscila em torno de zero e neste caso não foi identificado correlação cruzada entre as séries temporais (Figura 3.4(d)).

Assim, especificamente as correlações cruzadas entre a idade do taxista  $\times$  o tempo da licença ( $\square$ ), tempo de manutenção do táxi  $\times$  custo da manutenção do táxi (\*) (Figura 3.4(a)), mostra uma evidência de correlação cruzada (grupo i) com  $\lambda_{\square} = 0,52$  e  $\lambda_{*} = 0,48$  (a lei de potência foi medida por um ajuste linear com  $\chi^2 > 0,98$  e erro padrão  $< 0,02$ ). Entretanto, através do método DCCA foi constatado que existe outros tipos de comportamento nos acidentes de táxi. Por exemplo, na (Figura 3.4(b)) as correlações de curto alcance aparece entre a idade do taxista  $\times$  o ano de fabricação do táxi ( $\circ$ ) e também a licença do motorista  $\times$  ano de fabricação do táxi ( $\times$ ). A partir desta figura é possível constatar uma leve tendência dos taxista com mais experiência permanecer por um longo período com seu táxi. A (Figura 3.4(c)) denota uma correlação cruzada de curto alcance entre a idade do taxista  $\times$  custo da manutenção do carro ( $\diamond$ ), mas pode-se provar pelo método DCCA que as correlações cruzadas são negativas. Esta figura apresenta uma tendência para os taxistas mais velhos gastar menos dinheiro na época da manutenção do carro, mas como na (Figura 3.4(b)), esta correlação cruzada não é de longo alcance. Também pelo método DCCA, foi possível constatar que não existe correlações cruzadas entre a idade dos taxista  $\times$  tempo de manutenção do carro ( $\nabla$ ), tempo da licença do taxista  $\times$  tempo de manutenção do carro ( $\otimes$ ), tempo de licença  $\times$  custo da manutenção ( $\blacksquare$ ), ano de fabricação  $\times$  tempo da manutenção ( $\blacktriangledown$ ), ou ano de fabricação  $\times$  custo da manutenção ( $\blacklozenge$ ).

Sucintamente, as análises das correlações cruzadas por meio do método DCCA nas séries objeto de estudo dessa pesquisa foram sumarizadas na tabela (Tabela 3.1). Com o propósito de fazer uma comparação entre os métodos foi relacionado uma variável em relação a outra (Figura 3.5 e Tabela 3.1). Nesta figura, propomos um ajuste linear (linha contínua) com os valores do coeficiente de correlação R bem como o p-valor. Os resultados são equivalentes àqueles obtidos pelo método DCCA (Figura 3.4), mas o DCCA é mais eficiente em identificar estes tipos de correlação cruzada, especialmente porque podemos ver  $F_{DCCA}^2$  em diferentes escalas de tempo, e identificar uma auto afinidade nestas séries temporais. E, além disso, o método DCCA pode remover as tendências polinomiais.

Tabela 3.1: Matriz de correlações cruzadas entre as séries temporais de acidentes de táxis registrados pelo CEAT. A matriz superior representa a análise DCCA, enquanto que a matriz inferior representa o coeficiente de correlação clássico (ver Figura 3.5). Aqui nessa tabela temos os seguintes códigos: LR= correlação cruzada de longo alcance, SR= correlação cruzada de curto alcance e NC sem correlação entre as séries temporais. Fonte: (ZEBENDE; SILVA; MACHADO, 2011).

	Idade	Licença	Fabricação	Reparo	Custo
Idade	#	LR( $\lambda_{DCCA} = 0,52$ )	SR <sub>+</sub>	NC	SR <sub>-</sub>
Licença	0,85	#	SR <sub>+</sub>	NC	NC
Fabricação	0,12	0,16	#	NC	NC
Reparo	-0,01	0,01	0,10	#	LR( $\lambda_{DCCA} = 0,48$ )
Custo	-0,04	-0,02	0,07	0,44	#

Concluindo, o método DFA indica a presença de auto afinidade nas séries temporais dos acidentes de táxi, e o valor do expoente  $\alpha$  está muito próximo de 0,50 (séries de tempos não relacionadas), exceto para o ano de fabricação do táxi, onde o valor de alfa é 0,67 (este dado pode ser identificado com a taxa renovação da frota de táxi). No sentido do método DFA, nossas análises mostram que um acidente de táxi é essencialmente um acontecimento aleatório. Por outro lado, com o método DCCA podemos classificar as correlações cruzadas, em três grupos distintos: de longo alcance, de curta duração e um grupo que não há correlações cruzadas. Finalmente, a análise dos dados usados no método DCCA podem nos conduzir a efeitos econômicos e sociais se, por exemplo, lembramos que o CEAT (companhia) adiciona o valor “premium” para um táxi, levando em conta estas variáveis. O método DCCA mostrou aqui ser muito consistente para as análises dos dados e novas observações podem ser encontradas, como neste documento para os acidentes de táxi.

## 3.2 Diferenciação do $\rho_{DCCA}$ : Abordagens práticas e teóricas

### Resumo

Propomos neste trabalho estabelecer uma relação entre  $\alpha_{DFA}$  (expoente de autocorrelação de longo alcance) e  $\lambda_{DCCA}$  (o expoente de correlação cruzada de longo alcance), respectivamente descritos pelos métodos DFA e DCCA. Esta relação será implementada teoricamente pela diferenciação do coeficiente de correlação cruzada  $\rho_{DCCA}(n)$ . Também, para algumas séries temporais específicas aplicamos esta teoria com o objetivo estabelecer sua validade <sup>1 2 3</sup>.

Nas últimas décadas o estudo de sistemas complexos tem se tornado cada vez mais relevante. Estes sistemas são estudados em muitas áreas das ciências naturais, matemática e ciências naturais (SCIENCE... , 2009), (SETHNA; DAHMEN; MYERS, 2001), (MANDELBROT, 1982). Sistemas complexos não possuem comportamento linear, e podem ser estudados levando em conta as propriedades fractais (MANDELBROT, 1982), tais como a auto afinidade. Se, por exemplo, em uma série temporal fornecida  $\{y_i\}$  a auto afinidade aparece, então as correlações da lei de potência de longo alcance são apresentadas. Isto faz com que o estudo de sistemas complexos seja muito interessante, porque é possível identificar universalidade em uma variedade de problemas (FEDER, 1988), (BAK, 1997). Sabe-se que no mundo real, os dados são altamente não estacionários (DEVLIN; GNANADESIKAN; KETTENRING, 1975) e muitos métodos convencionais de análise não se adequam para séries não estacionárias. Para solucionar este problema e quantificar as correlações de longo alcance presentes em séries temporais não estacionárias o método DFA fornece uma relação entre  $F_{DFA}(n)$  e  $n$ , caracterizada por uma lei de potência:

$$F_{DFA}(n) \propto n^\alpha \quad (3.2)$$

O expoente de escala  $\alpha$  é um parâmetro de auto afinidade, representando as propriedades da autocorrelação de longo alcance da lei de potência em séries temporais. Se o sinal tem somente correlações de curto alcance (ou nenhuma correlação),  $\alpha = 0,50$ . Por outro lado, se  $\alpha < 0,50$ , a correlação no sinal é antipersistente, se  $\alpha > 0,50$ , a correlação no sinal é persistente.

<sup>1</sup>O presente artigo foi publicado no periódico Physica A no volume 392 em 2013.

<sup>2</sup>Foi feito um pedido de registro de programa de computador ao INPI (Instituto Nacional da Propriedade Industrial) em 22 de julho de 2013 do programa que aplica o método DFA, DCCA e o coeficiente de correlação cruzada sem tendência  $\rho_{DCCA}(n)$  utilizados neste artigo com co-autoria do SENAI CIMATEC (ver Apêndice B.1).

<sup>3</sup>Esta pesquisa recebeu auxílio financeiro da FAPESB e do CNPq.

Neste sentido, muitas observações podem ser mensuradas e registradas simultaneamente, em intervalos de tempo sucessivos, formando séries temporais com o mesmo comprimento  $N$ . Por exemplo, se temos duas séries temporais  $\{y_1\}$  e  $\{y_2\}$ , a análise da correlação cruzada entre estas séries temporais podem ser aplicadas. Neste caso, é possível aplicar a generalização do método DFA denominado DCCA para estudar as correlações cruzadas de longo alcance na presença de não estacionariedade. Assim, dado duas séries temporais com correlação cruzada de longo alcance, calculamos os sinais integrados  $R_1(k) \equiv \sum_{i=1}^k y_1(i)$  e  $R_2(k) \equiv \sum_{i=1}^k y_2(i)$  em que  $k = 1, 2, \dots, N$ . Em seguida divide-se as séries temporais integradas em  $(N - n)$  *boxes* sobrepostos, cada um contendo  $(n + 1)$  valores. Para ambas as séries temporais, em cada box que começa em  $i$  e termina com  $i + n$  defini-se a tendência local  $\tilde{R}_{1,i}(k)$  e  $\tilde{R}_{2,i}(k)$  ( $i \leq k \leq i + n$ ) para ser a coordenada do ajuste linear definimos o caminho sem tendência como uma diferença entre o caminho original e a tendência local. Depois calculamos a covariância dos resíduos em cada *box*  $f_{DCCA}^2(n, i) \equiv 1/(n + 1) \sum_{k=i}^{i+n} (R_1(k) - \tilde{R}_{1,i}(k))(R_2(k) - \tilde{R}_{2,i}(k))$ . Finalmente, a função de covariância sem tendência é calculada pela soma de todos box sobrepostos  $(N - n)$  de tamanho  $n$ .

$$F_{DCCA}^2(n) \equiv (N - n)^{-1} \sum_{i=1}^{N-n} f_{DCCA}^2(n, i). \quad (3.3)$$

Se a auto afinidade está presente, então uma lei de potência existe nas correlações cruzadas, ou seja:

$$F_{DCCA}^2(n) \sim n^{2\lambda}, \quad (3.4)$$

em que  $\lambda$  denota o expoente de correlação cruzada de longo alcance da lei de potência. Supondo  $R_1(k) = R_2(k)$  a covariância sem tendência  $F_{DCCA}^2(n)$  reduz a variância sem tendência  $F_{DFA}^2(n)$  usada no método DFA. Segundo (PODOBNIK; STANLEY, 2008), em geral,  $\lambda$  tende a ser a média dos expoentes do DFA isto é,

$$\lambda \cong \frac{\alpha_1 + \alpha_2}{2}. \quad (3.5)$$

Neste trabalho, mostraremos que é possível estabelecer uma relação clara entre  $\alpha$  e  $\lambda$  se empregarmos o coeficiente de correlação cruzada  $\rho_{DCCA}(n)$ , como veremos a seguir.

### 3.2.1 Implementação teórica

O coeficiente de correlação cruzada é definido como uma relação entre a covariância sem tendência  $F_{DCCA}^2$  e a variância sem tendência  $F_{DFA}$  de  $\{y_1(i)\}$  e  $\{y_2(i)\}$ , isto é,

$$\rho_{DCCA}(n) \equiv \frac{F_{DCCA}^2(n)}{F_{DFA1}(n)F_{DFA2}(n)} \quad (3.6)$$

A seguir, vamos estabelecer uma relação bem definida entre  $\lambda$  e  $\alpha$ . E esta relação será aplicada para suas séries temporais não estacionárias com autocorrelação de longo alcance e com correlações cruzadas de longo alcance, isto é,

$$F_{DFA1}(n) = K_1 n^{\alpha_1}, F_{DFA2}(n) = K_2 n^{\alpha_2}, \quad (3.7)$$

e

$$F_{DCCA}^2(n) = K_3 n^{2\lambda} \quad (3.8)$$

$K_1$ ,  $K_2$  e  $K_3$  são constantes diferentes de zero. Assim a partir das equações 3.6 e 3.8, temos:

$$\rho_{DCCA}(n) = K n^{2\lambda - \alpha_1 - \alpha_2}, \quad (3.9)$$

em que  $K \equiv \frac{K_3}{K_1 K_2}$

Estabelecendo  $y \equiv \log \rho_{DCCA}(n)$  e  $x \equiv \log(n)$ , então a partir da equação 3.9,

$$\frac{dy}{dx} = 2\lambda - \alpha_1 - \alpha_2. \quad (3.10)$$

Olhando a equação 3.10 podemos ver três possibilidades:

$$\lambda = (\alpha_1 + \alpha_2)/2;$$

$$\lambda < (\alpha_1 + \alpha_2)/2;$$

e

$$\lambda > (\alpha_1 + \alpha_2)/2.$$

É relevante relatar que quando  $F_{DCCA}^2(n)$  é negativa para todo  $n$ , a constante  $K_3$  é negativa. Neste caso podemos apresentar  $-F_{DCCA}^2(n)$  (ou simplesmente fazer  $K_3$  positiva) em um gráfico log – log com o propósito de calcular o valor de  $\lambda$ . No caso de não existir lei de potência na correlação cruzada  $\rho_{DCCA}(n) = 0$ , ou  $F_{DCCA}^2(n)$  oscilar em torno de zero, não há qualquer relação entre os expoentes  $\alpha$  e  $\lambda$ .

Analisando a equação 3.9 é possível ver que, no limite de  $n \mapsto \infty$ , então  $1/\log(n) \mapsto 0$  (uma queda muito lenta). Consequentemente  $\lambda \cong (\alpha_1 + \alpha_2)/2$ , porque  $(\log \rho_{DCCA} - \log K)/\log(n) \rightarrow 0$ .

Visando verificar essas possibilidades foi considerado dois casos relevantes:

(a)  $\frac{dy}{dx} = 0$  com  $\lambda = (\alpha_1 + \alpha_2)/2$ , que é um caso típico;

(b)  $\frac{dy}{dx} = \xi (\xi \in R)$ . Neste caso,  $\lambda = (\xi + \alpha_1 + \alpha_2)/2$  ou  $\lambda \neq (\alpha_1 + \alpha_2)/2$ .

Para validar as suposições dos itens (a) e (b) analisamos algumas séries temporais não estacionárias. As referidas análises serão estabelecidas em séries temporais reais e simuladas.

### 3.2.2 Dados e resultados

Começamos tratando duas séries temporais com correlações cruzadas de longo alcance da lei de potência, por um o processo ARFIMA (HOSKING, 1981), (GRANGER, 1980) e (PODOBNIK et al., 2005). Estas séries temporais são independentes e não estacionárias, com os expoentes  $\alpha_1 = \alpha_2 = 0,50$ . Foram geradas duas séries auto afins com número aleatórios entre 0 e 1,  $\{y_{rand_1}\}$  e  $\{y_{rand_2}\}$  com  $N = 20.000$  pontos. Em um processo ARFIMA cada variável depende não somente do seu próprio passado, mas também dos valores passados de outras variáveis, seja,

$$\begin{aligned}
y_{1,i} &= W \sum_{n=i}^{\infty} a_n(\rho_1) y_{1,i-n} + (1-W) \sum_{n=1}^{\infty} a_n(\rho_2) y_{2,i-n} + \varepsilon_{1,i}, \\
y_{2,i} &= (1-W) \sum_{n=i}^{\infty} a_n(\rho_1) y_{1,i-n} + W \sum_{n=1}^{\infty} a_n(\rho_2) y_{2,i-n} + \varepsilon_{2,i},
\end{aligned} \tag{3.11}$$

Neste processo  $\varepsilon_{1,i}$  e  $\varepsilon_{2,i}$  denotam duas séries independentes e identicamente distribuídas (i.i.d) variáveis Gaussianas com média zero ( $\mu = 0$ ) e variância igual a 1. Aqui  $a_n(\rho)$  são pesos estatisticamente definidos por  $a_n(\rho) = \Gamma(n - \rho) / (\Gamma(-\rho)\Gamma(1 + n))$  em  $\Gamma$  denota a função Gamma. E  $\rho$  são os parâmetros que variam de -0,50 até 0,50, relacionado com o expoente DFA,  $\alpha = 0,50 + \rho$ .  $W$  é um parâmetro livre variando de 0,50 a 1,0 e da intensidade das correlações cruzadas entre  $\{y_1(i)\}$  e  $\{y_2(i)\}$ .

Como nosso primeiro teste, propomos analisar o processo ARFIMA em que ambas as séries temporais compartilham o mesmo processo Gaussiano i.i.d  $\varepsilon_{1,1} = \varepsilon_{2,1}$  e  $W = 1$  (ver (PODOBNIK et al., 2009a) apêndice A). Segundo (PODOBNIK et al., 2009a) estas séries temporais possuem correlações cruzadas de longo alcance, em que o expoente de escala das correlações cruzadas  $\lambda$  é igual a média dos expoentes  $\alpha$ . E este resultado é também encontrado numericamente em (PODOBNIK; STANLEY, 2008).

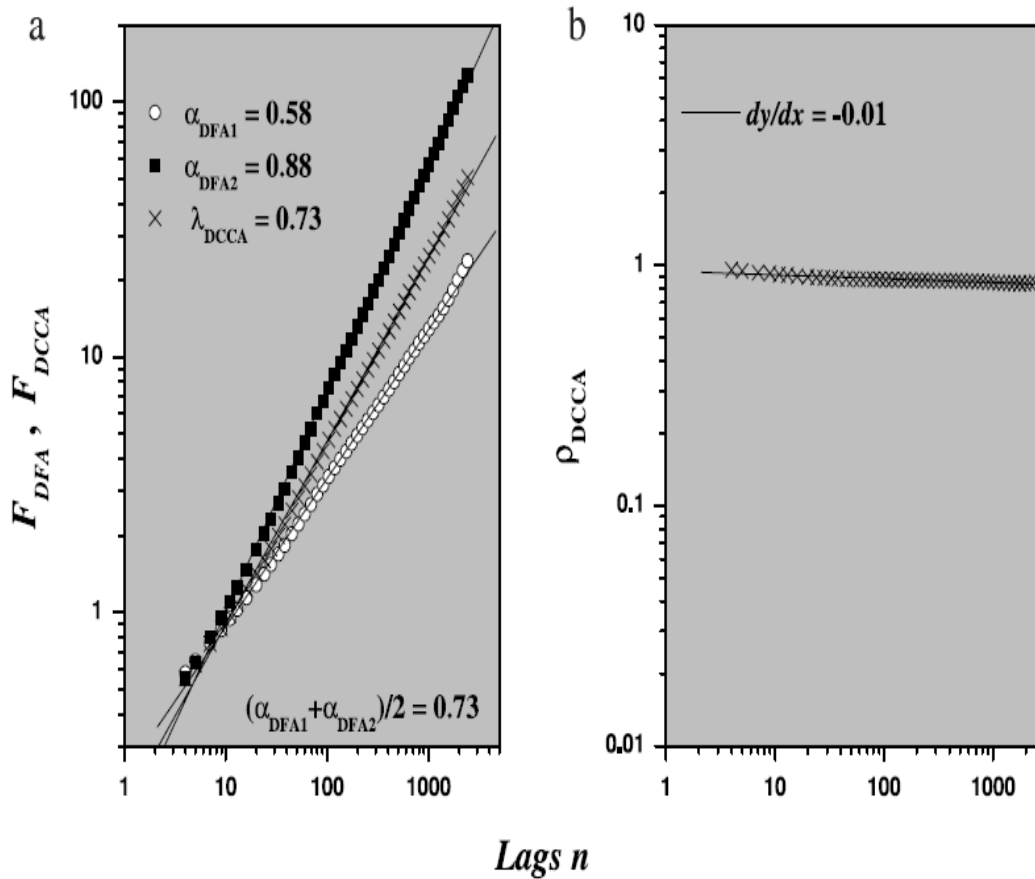


Figura 3.6: (a) variância sem tendência,  $F_{DFA}$  e a covariância sem tendência  $F_{DCCA}$  em função de  $n$  (*Lags*) (Amplitude de tempo) para as duas séries temporais geradas pelo processo AR-FIMA com  $\rho_1 = 0,1(\circ)$ ,  $\rho_2 = 0,4(\blacksquare)$  e  $W = 1,0$ . Nestes processos ARFIMA ambas séries temporais,  $\{y_1\}$  e  $\{y_2\}$  compartilham o mesmo processo Gaussiano i.i.d  $\varepsilon_{1,i} = \varepsilon_{2,i}$ . Uma autocorrelação de longo alcance de lei de potência,  $F_{DFA}(n) \sim n^\alpha$  e uma correlação cruzada de longo alcance de lei de potência,  $F_{DCCA}(n) \sim n^\lambda$  são geradas. (b) O coeficiente de correlação cruzada  $\rho_{DCCA} \times n$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0,99$ , erro padrão = 0,02 e  $p$ -valor  $< 0,0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013).



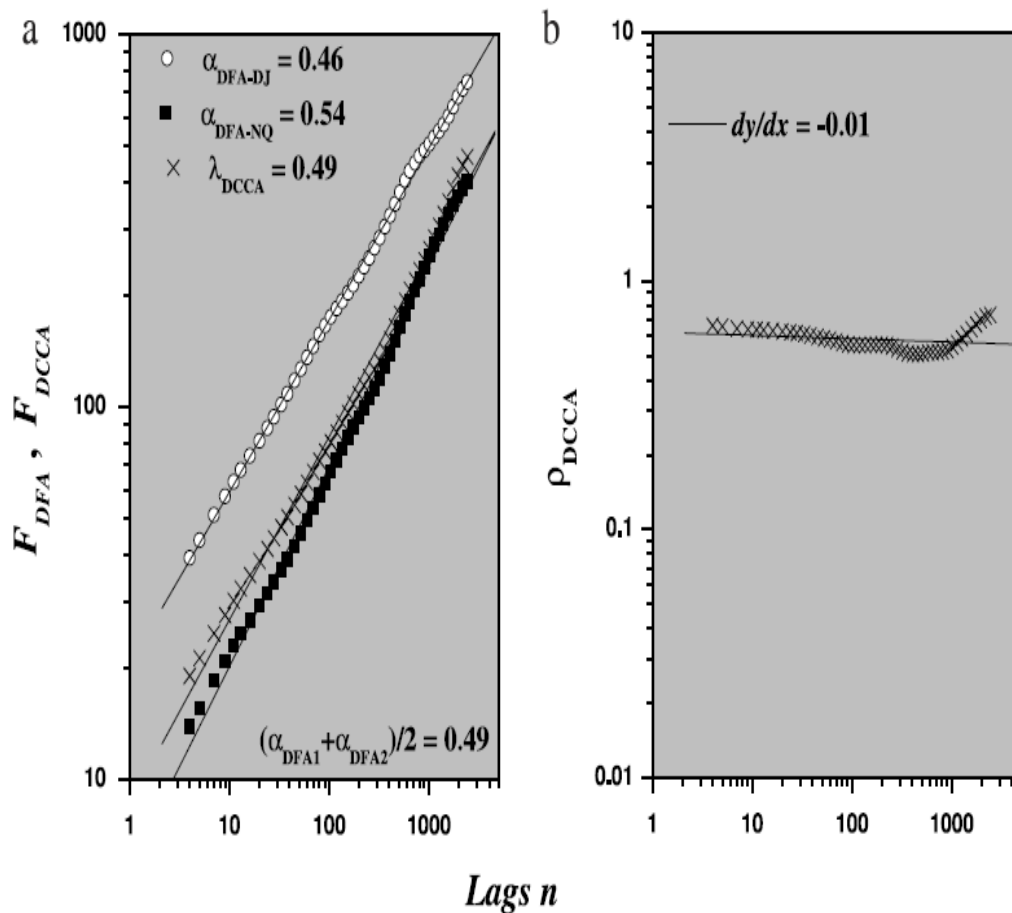


Figura 3.7: a) variância sem tendência  $F_{DFA}(n)$  e a covariância sem tendência  $F_{DCCA}(n)$  como função de  $n$  (amplitude de tempo) para Dow Jones  $\{y_1\}$  e NASDAQ  $\{y_2\}$  diferenças de preços de fechamento de volume ajustado para dividendos e desdobramentos registrados diariamente entre 5 de fevereiro de 1971 e 9 de Dezembro de 2010. b) Representa a relação  $\log_{10} \times \log_{10}$  de  $\rho_{DCCA}(n) \times (n)$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0,99$ , erro padrão = 0,02 e  $p - valor < 0,0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013).

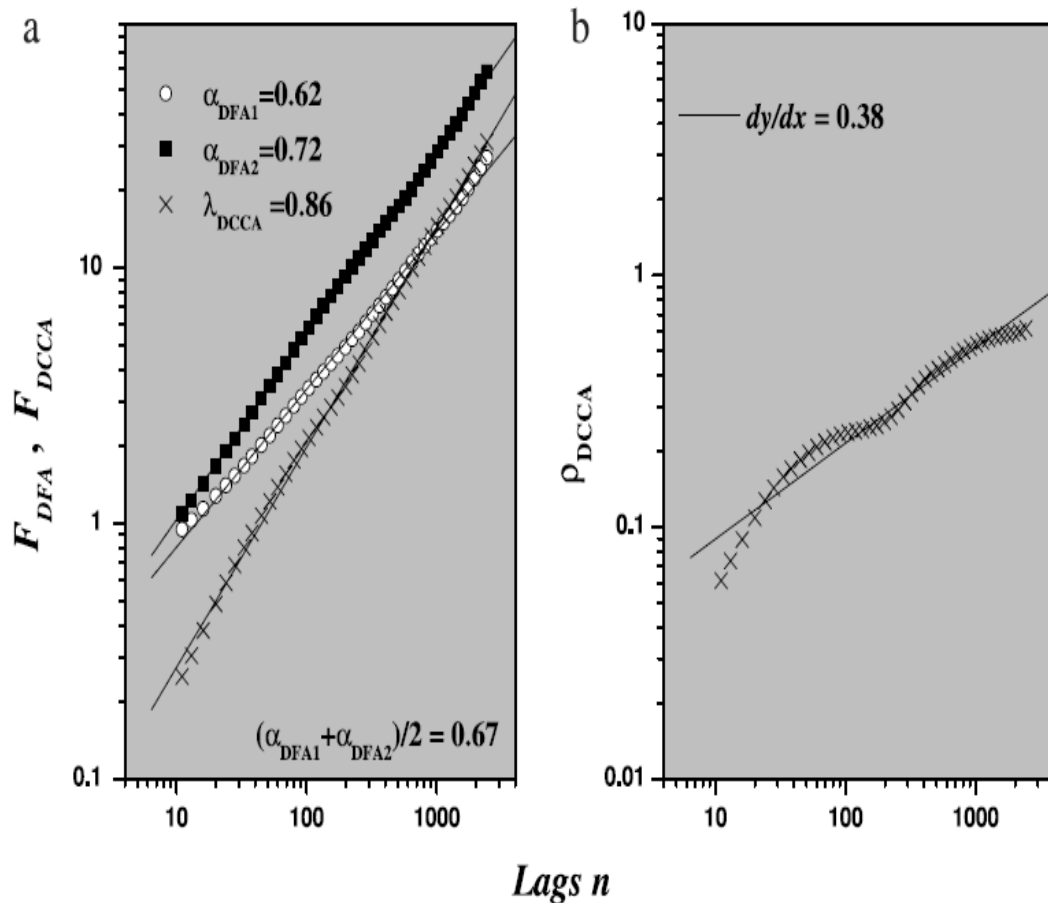


Figura 3.8: (a) variância sem tendência,  $F_{DFA}$  e a covariância sem tendência  $F_{DCCA}$  em função de  $n$  para as duas séries temporais  $\{y_1\}$  e  $\{y_2\}$  geradas pelo processo ARFIMA: neste processo temos  $\rho_1 = 0,1$ ,  $\rho_2 = 0,4$ ,  $\varepsilon_i \neq \varepsilon'_i$  e  $W = 0,85$ . Uma autocorrelação de longo alcance,  $F_{DFA}(n) \sim n^\alpha$  e uma correlação cruzada de lei de potência,  $F_{DCCA}(n) \sim n^\lambda$  são gerados. (b) Coeficiente de correlação cruzada  $\rho_{DCCA}(n) \times n$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0,98$ , erro padrão= 0,05 e  $p$ -valor  $< 0,0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013).

Usando o processo ARFIMA com duas componentes com  $\rho_1 = 0,1$ ,  $\rho_2 = 0,4$  e  $\varepsilon_{1,i} = \varepsilon_{2,i}$  as correlações cruzadas de longo alcance foram geradas (ver Figura 3.6). Como resultado desse processo a (Figura 3.6 a)) apresenta a variância sem tendência,  $F_{DFA}(n)$  e a covariância sem tendência,  $F_{DCCA}(n)$  como funções de  $n$ . Este processo pode ser ajustado pelas leis de potência,  $F_{DFA}(n) \propto n^\alpha$  e  $F_{DCCA}(n) \propto n^\lambda$  com expoente de escala  $\alpha_1 = 0,58$ ,  $\alpha_2 = 0,88$  e  $\lambda = 0,73$ . Na Figura 3.6 é possível ver que  $\lambda = \frac{\alpha_1 + \alpha_2}{2}$  pois  $\frac{dy}{dx} \simeq 0$  (linha contínua), e isto ilustra a relação esperada entre os expoentes.

Ainda com o objetivo de contemplar um caso real de correlações cruzadas com  $\frac{dy}{dx} = 0$  e  $\lambda = \frac{\alpha_1 + \alpha_2}{2}$  escolhemos séries temporais de alguns índices de ações de mercados, porque estes têm sido extensivamente estudadas. A nossa escolha, neste caso, foi estudar a correlação cruzada entre Dow Jones e NASDAQ para as diferenças ajustadas nos preços de fechamento. Os dados foram coletados diariamente ao longo do período entre 5 de fevereiro de 1971 a 9 de dezembro de 2010. Como esperado foi identificada autocorrelação de lei de potência  $F_{DFA}(n) \propto n^\alpha$ , com  $\alpha_{DJ} = 0,46(\circ)$  e  $\alpha_{NQ} = 0,54(\blacksquare)$ , assim como a correlação cruzada de lei de potência  $F_{DCCA}(n) \propto n^\lambda$ , com  $\lambda = 0,49(\times)$  (ver Figura 3.7 a)). Como é possível constatar a média de  $\alpha = 0,49$  e o mesmo valor de  $\lambda = 0,49(\times)$ . Já a figura 3.7 b) ilustra a relação esperada entre os expoentes porque  $\frac{dy}{dx} \cong 0$ .

A fim de testar o caso  $\lambda \neq (\alpha_1 + \alpha_2)/2$  um processo ARFIMA foi gerado com  $W = 0,85$ ,  $\rho_1 = 0,1$  e  $\rho_2 = 0,4$  e  $\varepsilon_{1,i} \neq \varepsilon_{2,i}$ . Os resultados para este caso simulado estão na 3.8. É possível visualizar na figura 3.8 a) que  $F_{DFA}(n) \propto n^\alpha$  com  $\alpha_1 = 0,62(\circ)$  e  $\alpha_2 = 0,72(\blacksquare)$  para as séries temporais 1 e 2. Também, podemos ver que  $F_{DCCA}(n) \propto n^\lambda$  com  $\lambda = 0,86$ . Porém, a correlação cruzada não é perfeita para este processo ARFIMA e  $\lambda \neq (\alpha_1 + \alpha_2)/2$ . Entretanto, como é possível ver  $dy/dx \cong 0,38$  e evidentemente esta simulação confirma a teoria sugerida na equação 3.10, simplesmente porque  $\lambda = (0,38 + \alpha_1 + \alpha_2)/2$ .

Finalmente, com um exemplo empírico de  $\lambda \neq (\alpha_1 + \alpha_2)/2$  propomos estudar a correlação cruzada entre os índices Dow Jones e o 00062.SS (a companhia local SSE Shanghai). Com o propósito de atingir esta meta, estudaremos os valores de diferenças sucessivas nos preços de fechamento ajustados desde 4 de Janeiro de 2000 até 9 de Dezembro de 2010, registrados diariamente (FINANCE, 2013). Novamente, como no caso anterior,  $F_{DFA}(n)$  e  $F_{DCCA}(n)$  podem ser ajustados para correlação de lei de potência com  $\alpha_{DJ} = 0,47(\circ)$ ,  $\alpha_{SSE} = 0,59(\blacksquare)$  e  $\lambda_{DJ \times SSE} = 0,77(\times)$  (ver Figura 3.9) A figura 3.9 b) apresenta o valor do coeficiente de correlação cruzada sem tendência,  $\rho_{DCCA}(n)$ , com  $dy/dx \cong 0,48$  (linha contínua). Como esperávamos teoricamente da equação 3.10,  $\lambda = (0,48 + \alpha_{DJ} + \alpha_{SSE})/2$ .

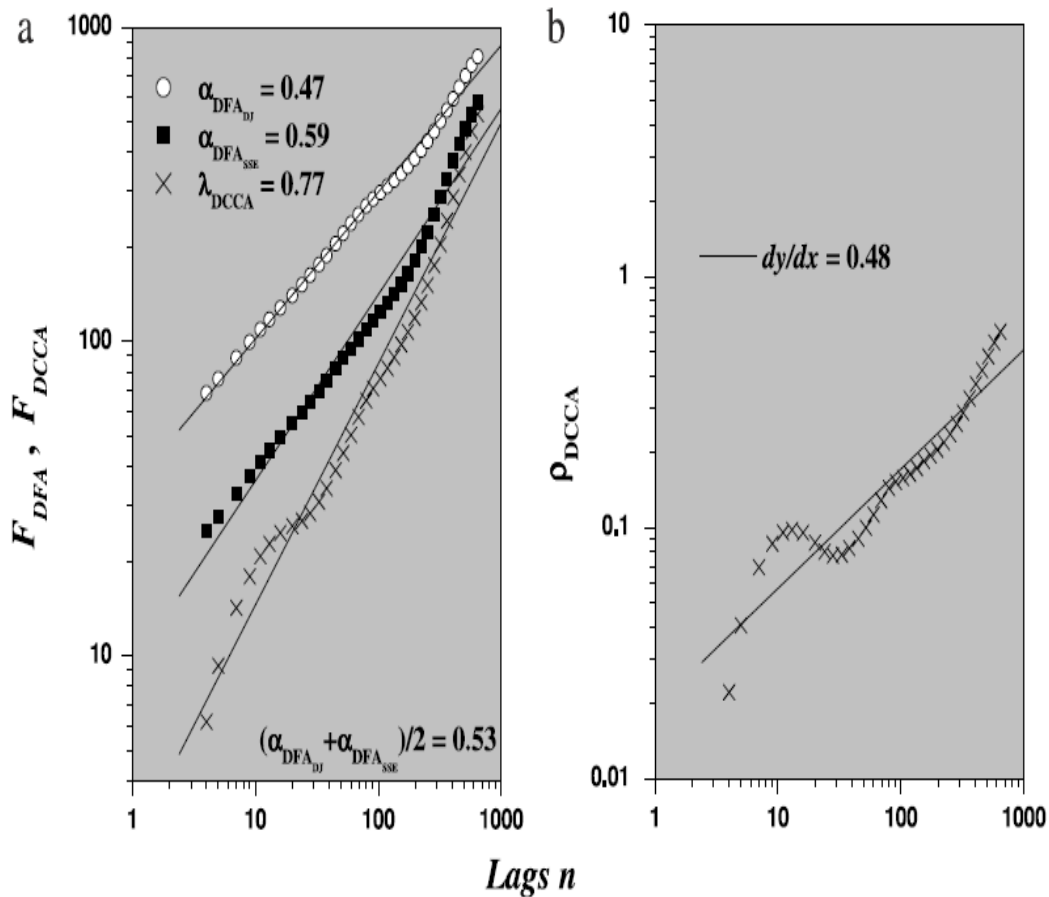


Figura 3.9: a) variância sem tendência  $F_{DFA}(n)$  e a covariância sem tendência  $F_{DCCA}(n)$  como função de  $n$  (amplitude de tempo) para diferenças sucessivas nos ajustes de preços de fechamento para os índices Dow Jones  $\{y_1\}$  e NASDAQ  $\{y_2\}$  registrados diariamente de 4 de janeiro de 2000 a 9 de dezembro de 2010. b) Coeficiente de correlação cruzada  $\rho_{DCCA}(n) \times n$ . A linha contínua representa o ajuste linear,  $dy/dx$ , com  $R = 0,94$ , erro padrão= 0,11 e  $p$ -valor  $< 0,0001$ . Fonte: (ZEBENDE; SILVA; MACHADO, 2013).

Para uma melhor análise dos resultados obtidos nas figuras 3.6 a, 3.7 a, 3.8 a, 3.9 a, é possível consultar a tabela 3.2. E os resultados de  $dy/dx$  obtidos nas figuras 3.6 b, 3.7 b, 3.8 b, 3.9 b denotam a média das médias das derivadas ponto a ponto e nestes casos, tendo como base o teorema do limite central, admitimos  $dy/dx$  como o valor aproximado para derivada.

Tabela 3.2: Expoente de autocorrelação  $\alpha_{DFA}$  e expoente de correlação cruzada  $\lambda_{DCCA}$ .

Figura	$\alpha\{y_1\}$	$\alpha\{y_2\}$	$\lambda$
3.6	0,58 $\pm$ 0,01	0,88 $\pm$ 0,01	0,73 $\pm$ 0,01
3.7	0,46 $\pm$ 0,01	0,54 $\pm$ 0,02	0,49 $\pm$ 0,01
3.8	0,62 $\pm$ 0,01	0,72 $\pm$ 0,01	0,86 $\pm$ 0,02
3.9	0,47 $\pm$ 0,01	0,59 $\pm$ 0,01	0,77 $\pm$ 0,01

### 3.2.3 Conclusões

Aqui apresentamos uma nova abordagem teórica sobre o coeficiente de correlação cruzada sem tendência  $\rho_{DCCA}(n)$ . A meta nesta teoria foi utilizar o referido coeficiente para estabelecer uma relação entre  $\alpha_{DFA}$  (expoente de autocorrelação de longo alcance) e  $\lambda_{DCCA}$  (expoente de correlação cruzada de longo alcance) uma vez que ainda não existia tal relação. É importante enfatizar que esta teoria foi testada em séries reais e simuladas, e os resultados estão inteiramente em concordância com a equação 3.10. Finalmente, esta teoria pode ajudar a compreender as correlações cruzadas um pouco melhor na presença de não estacionariedade porque no futuro a diferenciação do  $\rho_{DCCA}(n)$ , pode ser um indicador interessante para este propósito.

### **3.3 Autocorrelação e correlação cruzada nas séries temporais dos homicídios dolosos e homicídios tentados**

#### **Resumo**

Propomos aqui neste trabalho estabelecer a relação entre homicídios dolosos e homicídios tentados na cidade de Salvador (BA) através de uma análise de séries temporais. Esta análise será realizada pelo DFA e DCCA, e pelo coeficiente de correlação cruzada  $\rho_{DCCA}$ . Através desta análise foi possível identificar uma correlação cruzada positiva entre homicídios dolosos e homicídios tentados. Ao mesmo tempo, olhando do ponto de vista da autocorrelação (DFA), esta análise pode ser mais informativa dependendo da escala de tempo analisada. Para pequenas escalas de tempo (dias), não identificamos autocorrelação, na escala de semanas o DFA apresentou comportamento antipersistente, e para grandes escala de tempo ( $n > 90$  dias) o DFA apresentou comportamento persistente. Finalmente, a aplicação deste novo tipo de análise estatística provou ser eficiente e, neste sentido, este artigo pode contribuir para uma estatística descritiva mais precisa do crime.<sup>1</sup>

2 3 4

Neste artigo, o crime é estudado em termos de homicídios dolosos e homicídios tentados porque estes são crimes contra pessoa e são amplamente usados nos estudos empíricos sobre os determinantes da criminalidade. A Figura 3.10 mostra as séries temporais dos homicídios dolosos e homicídios tentado por 100.000 habitantes. Nesta figura, podemos ver grandes irregularidades (imprevisíveis), características de sistemas não lineares. Tais sistemas têm sido estudados do ponto de vista dos sistemas complexos.

---

<sup>1</sup>Uma versão do presente artigo foi submetida ao periódico Physica A em agosto de 2013 revisada em novembro de 2013 e publicada no volume 400 em janeiro de 2014.

<sup>2</sup>Foi realizado um pedido de registro de programa de computador ao INPI (Instituto Nacional da Propriedade Industrial) em 22 de julho de 2013 do programa que aplica o  $\rho_{DCCA}$  utilizado neste artigo com co-autoria do SENAI CIMATEC (ver Apêndice B.1).

<sup>3</sup>Foi realizado um pedido de registro de programa de computador ao INPI (Instituto Nacional da Propriedade Industrial) em 2 de abril de 2014 do programa que aplica o teste estatístico  $\rho_{DCCA}$  idealizado por (PODOBNIK et al., 2011) utilizado neste artigo com co-autoria do SENAI CIMATEC (ver Apêndice B.2).

<sup>4</sup>Agradecemos a FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro e também a SSP-BA (Secretaria de Segurança Pública do Estado da Bahia) pelos dados fornecidos.

### 3.3.1 Dados e resultados

Nossos dados foram obtidos junto à Secretaria de Segurança Pública do Estado da Bahia (SSP-BA, 2013). Estes dados são baseados nos registros diários da polícia civil, registrados pelos cidadãos de janeiro de 2004 a dezembro de 2011. Como uma primeira análise, vemos a Figura 3.10, a taxa diária de vítimas por 100 mil habitantes. Há um crescimento evidente na taxa de homicídios (+), porém não é possível observar o mesma evidência nas taxas de homicídios tentados (\*). Ainda analisando a figura 3.10 não é possível diagnosticar visualmente uma relação entre os delitos em questão. Com o propósito de conhecer o comportamento dos delitos, calculamos a sua Estatística Descritiva (ver Tabela 3.3).

Tabela 3.3: Estatística descritiva dos homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de janeiro de 2004 a dezembro de 2011.

Medida	Homicídios dolosos	Homicídios tentados
Média (Desvio padrão)	3,60 (2,52)	2,34 (2,05)
Soma	10.529	6.853
Assimetria	0,90	1,37
Curtose	0,99	2,71
Mínimo (máximo)	0 (16)	0 (15)

A média diária do homicídio doloso é de 3,60, enquanto a do homicídio tentado de 2,34 vítimas por dia. Ambas séries temporais apresentaram assimetria positiva, o que indica concentração de pontos da série abaixo da média e as ocorrências mais elevadas homicídios dolosos e homicídios tentados é de, respectivamente, 16 e 15 vítimas por dia (ver Tabela 3.3). Com intuito de avaliar o comportamento destas séries temporais não estacionárias, calculamos a média móvel das taxas de vítimas por 100 mil habitantes (a linha contínua na Figura 3.10). Esse procedimento suaviza as oscilações e torna mais fácil compreender o comportamento da ST. Dessa forma, podemos identificar as tendências de crescimento ou de decréscimo das séries. Já a Figura 3.11 mostra uma análise de sinais clássicos com a Função de Autocorrelação (FAC) e a transformada de Fourier (Amplitude). Por meio da Figura 3.11, é possível identificar claramente uma periodicidade de sete dias, mas para uma escala de longo prazo não é possível identificar outras periodicidades. Nossa proposta é analisar estas séries temporais com novas características, e se analisamos estas séries por meio do DFA, DCCA e  $\rho_{DCCA}$ , então podemos obter novas conclusões, como serão mostradas a seguir.

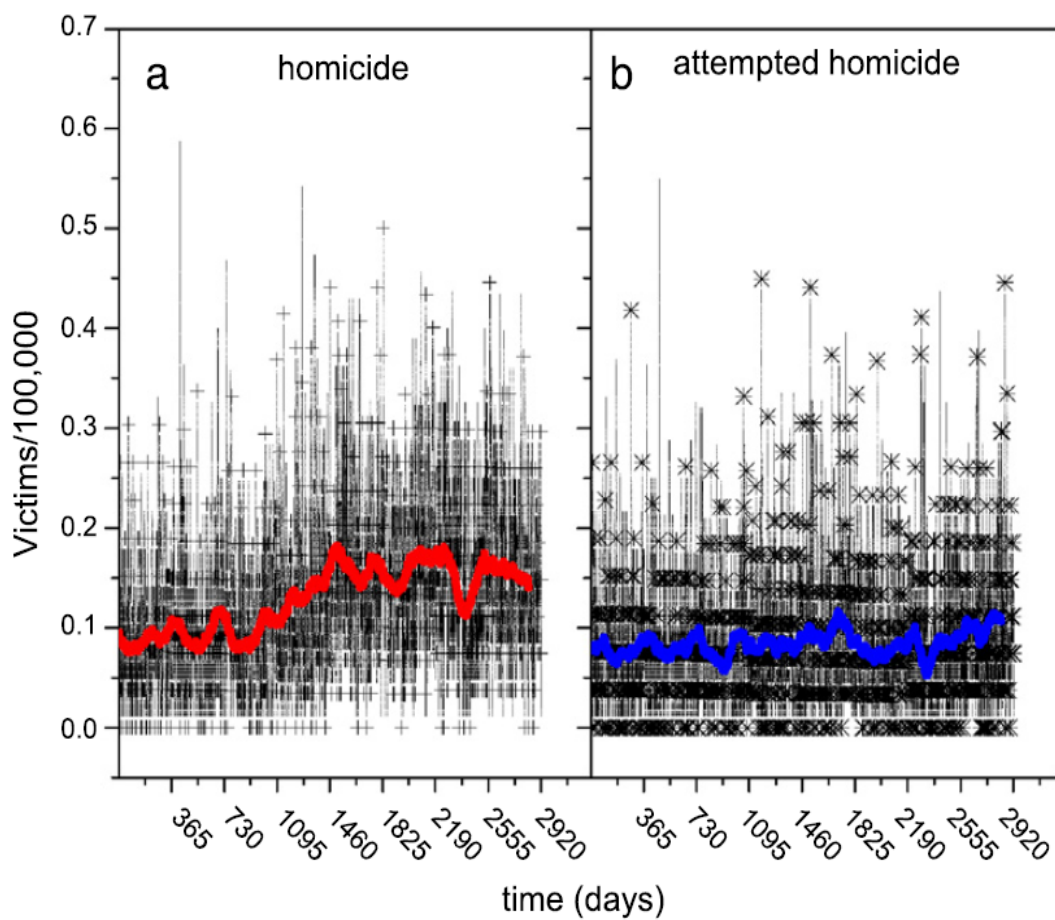


Figura 3.10: Séries temporais de vítimas/100.000 habitantes em Salvador-BA entre Janeiro 2004 e Dezembro de 2011 para: (a) homicídios dolosos; (b) homicídios tentados. A Linha contínua (colorida) representa a média móvel para  $n = 90$  dias. Fonte: (MACHADO; SILVA; ZEBENDE, 2014).



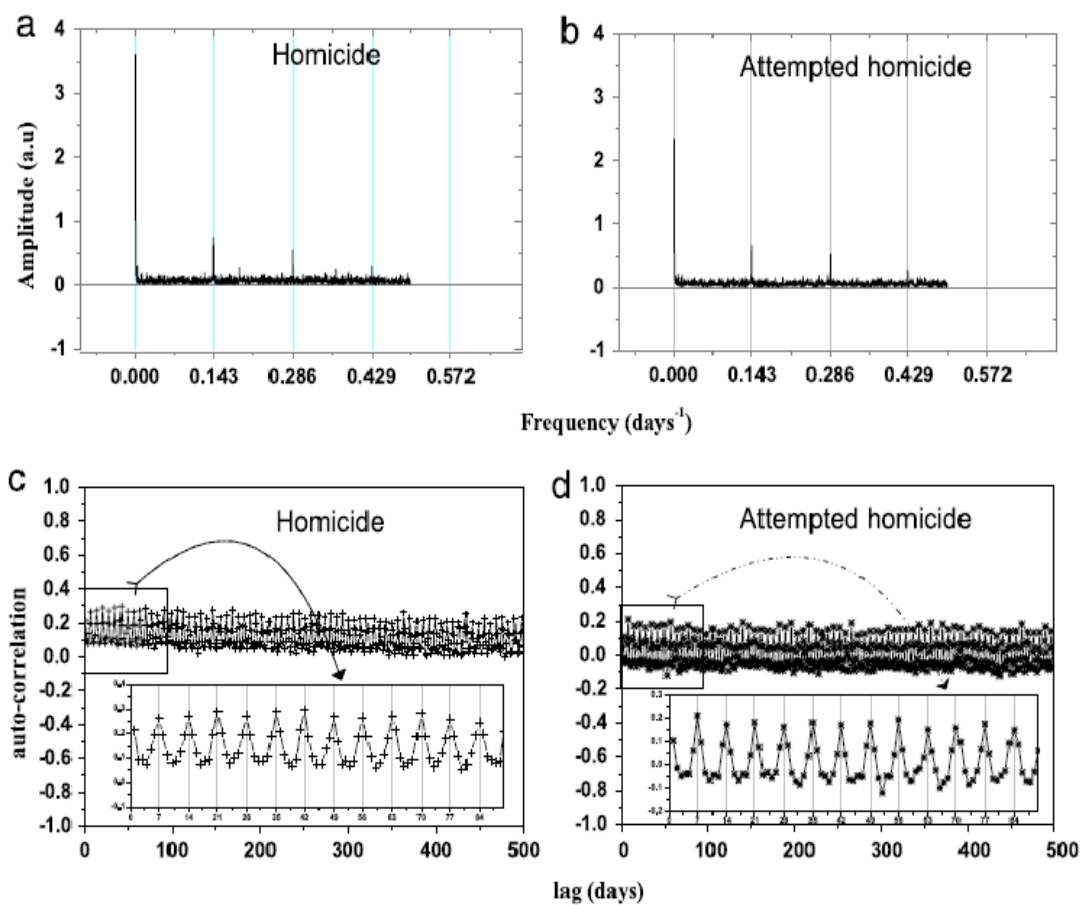


Figura 3.11: Processo de análise do sinal para homicídios dolosos (esquerda) e homicídios tentados (direita). Neste caso (a) e (b) representam a amplitude do espectro para Transformada de Fourier e (c) e (d) a Função de autocorrelação (FAC). As linhas verticais na figura representam sete dias de análise. Fonte: (MACHADO; SILVA; ZEBENDE, 2014).

A figura 3.12 e tabela 3.4 mostram os valores do expoentes de autocorrelação DFA para os homicídios dolosos  $\alpha_+$ , homicídios tentados  $\alpha_*$  e o expoente de correlação cruzada DCCA  $\lambda$ , (homicídio doloso  $\times$  homicídio tentado). Por meio da figura 3.12 e tabela 3.4 é possível estimar se as séries temporais exibem comportamento persistente, antipersistente ou decorrelacionado.

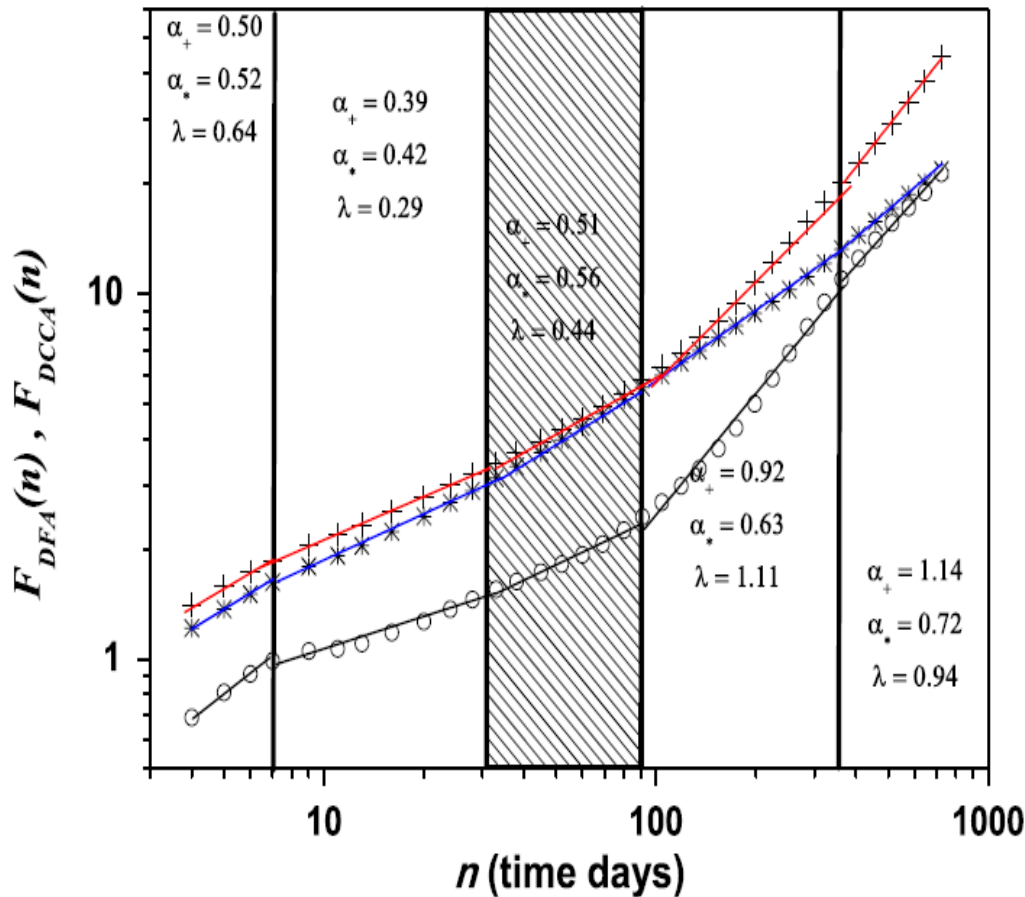


Figura 3.12: Variância sem tendência  $F_{DFA}(n)$  para os homicídios dolosos (+), homicídios tentados (\*), e covariância sem tendência  $F_{DCCA}(n)$  entre os homicídios dolosos e homicídios tentados (o) em função de  $n$ . As linhas verticais denotam a escala de tempo. Em cada intervalo definimos os valores dos expoentes. Fonte: (MACHADO; SILVA; ZEBENDE, 2014).

Para ambas séries temporais (delitos) podemos identificar situações interessantes. Há alguma correlação cruzada entre as séries, porque  $F_{DCCA}(n) \neq 0$ . É possível identificar componentes sazonais (linhas verticais), isto é, para  $n \leq 7$ , para  $7 \leq n \leq 30$ , para  $30 \leq n \leq 90$ , para  $90 \leq n \leq 365$  dias e finalmente para  $n \geq 365$  dias. Também, existe uma transição de comportamento entre as séries, dependendo da escala de tempo. Para  $n \leq 7$ , o crime ocorre de forma aleatória,  $\alpha_{DFA} \cong 0,50$ . Para  $7 \leq n \leq 30$ , os delitos em estudo são antipersistente, isto porque os homicídios dolosos apresentaram  $\alpha_+ = 0,39$  e  $\alpha_* = 0,42$  para os homicídios tentados.

Tabela 3.4: Expoente de autocorrelação  $\alpha_{DFA}$  dos homicídios dolosos (+), homicídios tentados (\*) e expoente de correlação cruzada  $\lambda_{DCCA}(\circ)$ .

Escala de tempo (dias)	$\alpha_+$	$\alpha_*$	$\lambda$
$n \leq 7$	$0,50 \pm 0,02$	$0,52 \pm 0,01$	$0,64 \pm 0,03$
$7 \leq n \leq 30$	$0,39 \pm 0,01$	$0,42 \pm 0,01$	$0,29 \pm 0,02$
$30 \leq n \leq 90$	$0,51 \pm 0,01$	$0,56 \pm 0,01$	$0,44 \pm 0,02$
$90 \leq n \leq 365$	$0,92 \pm 0,02$	$0,63 \pm 0,01$	$1,11 \pm 0,04$
$n \geq 365$	$1,14 \pm 0,02$	$0,72 \pm 0,01$	$0,94 \pm 0,02$

Se olharmos a coluna hachurada na figura 3.12, entre  $30 \leq n \leq 90$  dias, há uma transição clara antipersistente/persistente, e neste caso  $n = 90$  dias denota o ponto principal para esta transição. Esta observação nos levou a definir  $n = 90$  para o cálculo da média móvel (Ver Figura 3.10).

Vale a pena mencionar aqui que o efeito de tendências foram estudadas via DFA em (HU et al., 2001). Neste artigo, mostrou-se que o método DFA possui melhor desempenho do que a análise R/S para quantificar o comportamento de escala de sinais ruidosos para uma ampla variedade de correlações, e estimaram a variação das escalas onde o desempenho do método DFA é otimizado.

Levando em conta que identificamos uma correlação cruzada entre as séries temporais de homicídios dolosos e homicídios tentados, por meio da função de covariância sem tendência  $F_{DCCA}$  (ver Figura 3.12), então usamos o coeficiente de correlação cruzada  $\rho_{DCCA}(n)$ , com o objetivo de quantificar o nível das correlações cruzadas entre esses crimes (nova metodologia). A análise estatística da relação entre os homicídios dolosos e homicídios tentados identificou correlações cruzadas persistentes e positivas em diferentes escalas de tempo (ver Figura 3.13). Especificamente, os resultados mostram que as correlações cruzadas oscilaram em torno de 0,29, com  $\rho_{DCCA}(7) = 0,32$ ,  $\rho_{DCCA}(90) = 0,18$  (valor mínimo), e  $\rho_{DCCA>(> 365) = 0,47$  (valor máximo).

O conhecimento que o coeficiente de correlação cruzada  $\rho_{DCCA}(n)$  é positivo para qualquer escala de tempo  $n$ , nos permite declarar que se temos um aumento/decréscimo nos homicídios dolosos, nós esperamos um aumento/decréscimo nos homicídios tentados. Este valor positivo não diz nada sobre o relacionamento das séries defasadas (no sentido da correlação cruzada), mas é melhor dizer que se movem provavelmente juntas até mesmo durante os períodos seguintes, para mais informações veja (KRISTOUFEK, 2011).

Com o objetivo de testar estatisticamente os resultados do coeficiente  $\rho_{DCCA}$  obtidos na figura 3.13, aplicamos o teste de significância modelado por (PODOBNIK et al., 2011) (ver 3.14) e todos os valores do  $\rho_{DCCA}$  foram significativos com 95 por cento de confiança. E neste caso rejeitamos a hipótese nula ( $H_0: \rho_{DCCA} = 0$ ).

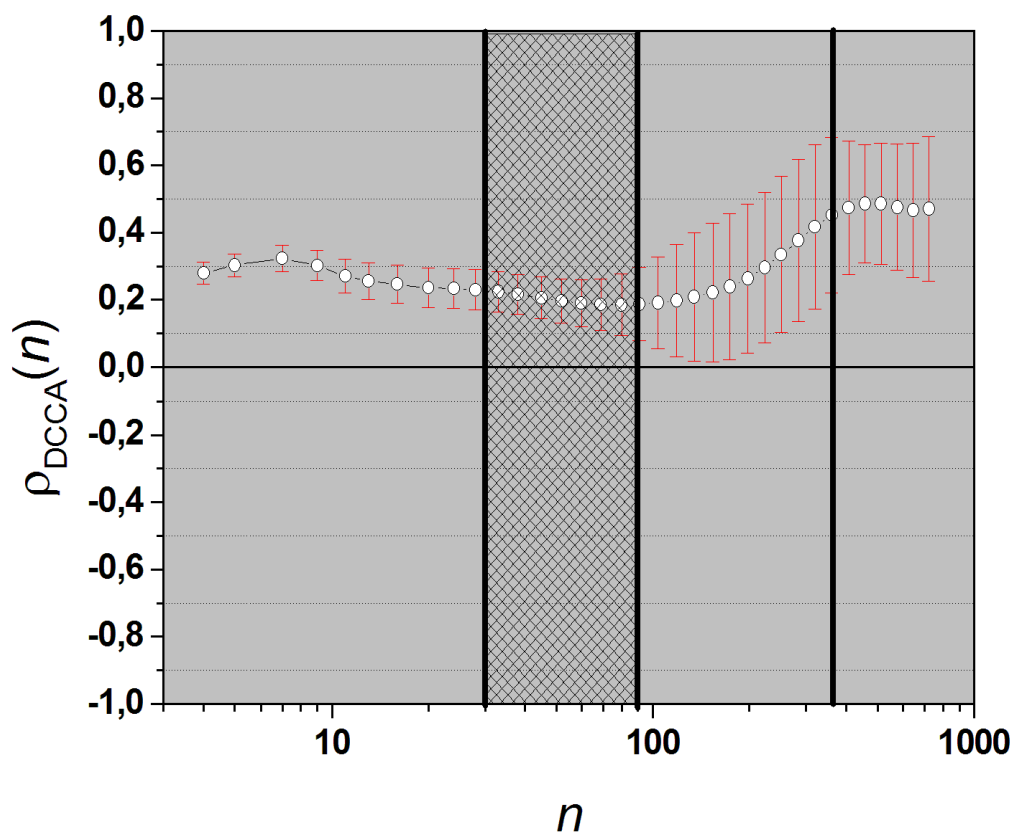


Figura 3.13: Coeficiente de correlação cruzada  $\rho_{DCCA}$  para os homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de janeiro de 2004 a dezembro de 2011. Fonte: (MACHADO; SILVA; ZEBENDE, 2014).

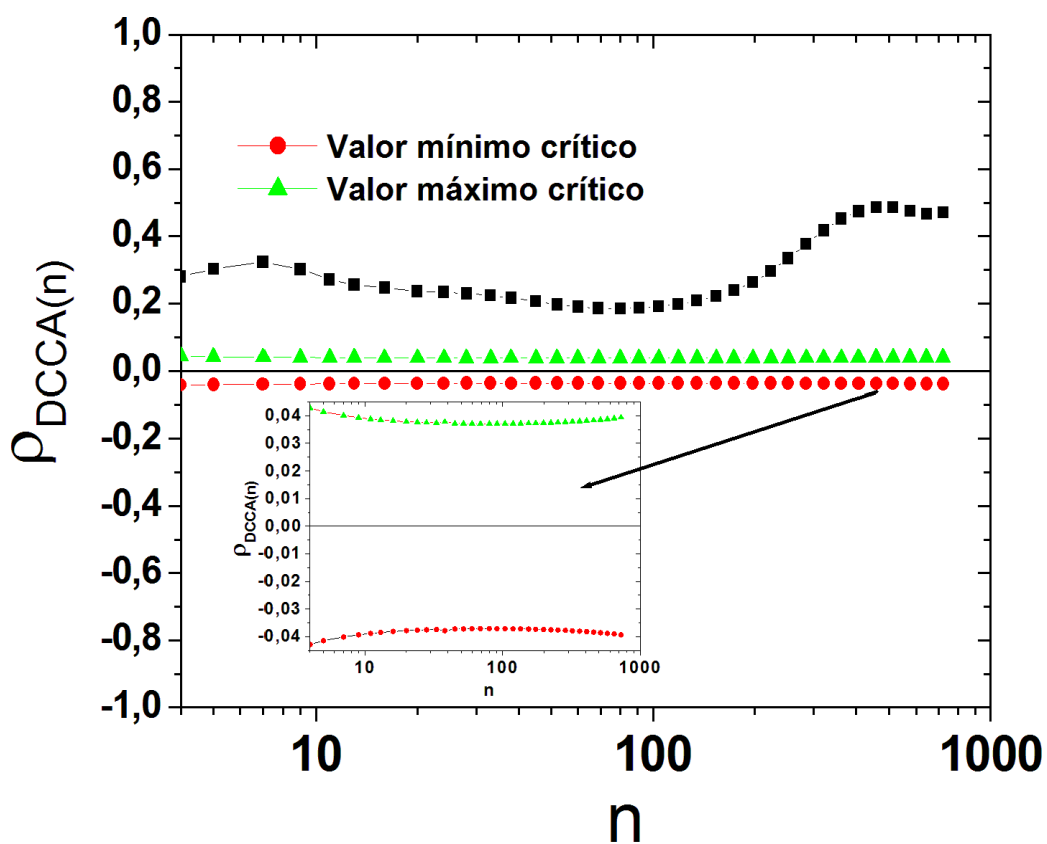


Figura 3.14: Teste estatístico idealizado por (PODOBNIK et al., 2011) sobre o coeficiente de correlação cruzada  $\rho_{DCCA}$  com 95 por cento de confiança, nos homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de janeiro de 2004 a dezembro de 2011.

### 3.3.2 Conclusões

Por meio do DFA, e DCCA e do coeficiente de correlação cruzada  $\rho_{DCCA}$ , identificamos nestas séries temporais autocorrelação, correlação cruzada (quantificando o seu nível), e também componentes sazonais. A autocorrelação oscilou entre persistente, anti-persistente, e sem memória, isso a depender da escala de tempo  $n$  em questão. A correlação cruzada (DCCA) mostrou que as séries temporais são relacionadas, e usando o  $\rho_{DCCA}$ , foi possível constatar que essa correlação é positiva para todas as escalas de tempo consideradas e estatisticamente significativas com 95 por cento de confiança. Assim, se a taxa de homicídio aumentar, então a taxa de homicídios tentados tende a aumentar, e/ou vice-versa.

Usando o DFA, DCCA e o  $\rho_{DCCA}$  foi possível constatar várias propriedades interessantes para diferentes escalas de tempo, principalmente nas colunas hachuradas nas figuras 3.12 e 3.13, onde há uma transição de antipersistente para comportamento persistente. O  $\rho_{DCCA}$  tem sido apropriado para este tipo de análise, porque é possível mensurar a correlação cruzada em diferentes escalas de tempo, e também remover as tendências polinomiais.

Poderíamos também pensar como a temperatura do ar influencia nos homicídios dolosos ou nos homicídios tentados, e medindo desta forma a influência do clima no crime. Os resultados preliminares, não apresentados neste artigo, mostram que a temperatura do ar não tem influência no homicídio doloso ou homicídio tentado, neste caso  $\rho_{DCCA} = 0$ . Finalmente, esta pesquisa apresentou uma possível metodologia para análise de dados criminais. E tal metodologia provou ser robusta para os dados em questão pelo fato de ser possível identificar componentes sazonais e quantificar o nível das correlações cruzadas entre as séries temporais. Neste sentido, este artigo explorou novas direções para estudo criminal, e nós podemos aplicar estes métodos para estudar outros tipos de crimes, a fim de ajudar o gerenciamento da Segurança Pública.

## Considerações finais

### 4.1 Conclusões

Neste projeto de Doutorado, realizamos uma breve revisão bibliográfica de séries temporais com ênfase nos métodos da mecânica estatística: DFA, DCCA e o coeficiente de correlação cruzada  $\rho_{DCCA}$ . Com o intuito de atender aos objetivos desta tese de doutorado (ver Capítulo 1) modelamos os supracitados métodos da mecânica estatística em dados reais e simulados (ver seção 3.1, 3.2 e 3.3).

Nas seções (3.1 e 3.3) foi comprovado que é possível extrair informações de séries temporais não estacionárias ao analisar conjuntamente do ponto de vista das autocorrelações (DFA) e correlações cruzadas (DCCA e  $\rho_{DCCA}$ ). Corroborando com os trabalhos de (ZEBENDE; MACHADO, 2009a) (MACHADO, 2009), (VASSOLER; ZEBENDE, 2011) e (ZHAO et al., 2011) os resultados obtidos na seções 3.1 e 3.3 mostraram que é possível identificar componentes sazonais com o método DCCA. Na seção 3.1 constatamos que a correlação cruzada mensurada com a covariância sem tendência (ver equação 2.29) do método DCCA pode ser dividida em três grupos distintos: correlação cruzada de longo alcance, curto alcance e sem correlação cruzada (ver Figura 3.4).

Implementamos neste projeto uma relação estatística entre os expoente  $\alpha_{DFA}$  e  $\lambda_{DCCA}$  obtida por diferenciação do coeficiente de correlação cruzada  $\rho_{DCCA}$ . Tal relação foi comprovada teoricamente e com aplicações em séries reais e simuladas. De acordo com (PODOBNIK; STANLEY, 2008)  $\lambda$  em geral seria a média dos expoentes do DFA, porém nossos resultados provaram que este é somente um caso particular, pois  $\lambda = \frac{C + \alpha_1 + \alpha_2}{2}$  com  $C$  igual ao coeficiente angular da relação linear entre o  $\log n$  e o  $\log \rho_{DCCA}$ . Constatamos também, neste trabalho, que no caso de não existir correlação cruzada ( $\rho_{DCCA} = 0$ ), ou a função de covariância sem tendência  $F_{DCCA}^2$  variar em torno de zero, não existe nenhuma relação entre  $\alpha_{DFA}$  e  $\lambda_{DCCA}$  (ver seção 3.2). Analisando  $\rho_{DCCA}(n) = Kn^{2\lambda - \alpha_1 - \alpha_2}$  (equação 3.9) foi possível constatar que no limite de  $n \mapsto \infty$ , então  $1/\log(n) \mapsto 0$  e neste caso o expoente de correlação cruzada obtido pelo método DCCA é aproximadamente a média do expoente de autocorrelação obtido pelo DFA, ou seja  $\lambda \cong (\alpha_1 + \alpha_2)/2$ , isso porque  $(\log \rho_{DCCA} - \log K)/\log(n) \rightarrow 0$ .

O estudo realizado na seção 3.1 nas variáveis registradas simultaneamente na série temporal de acidentes de táxi do banco de dados do Centro Especial de Auxílio ao Taxista localizada em Salvador-BA de 18 de agosto de 2002 a 21 de maio de 2010, utilizando como método de análise o DFA, com exceção da variável ano de fabricação com  $\alpha = 0,67 \pm 0,02$  caracterizou o acidente de táxi como um evento aleatório ( $\alpha \cong 0,50$ ) (ver Figura 3.3) e deduzimos que variáveis como economia, clima, entre outras, não geram efeito de memória nos acidentes. O estudo da correlação cruzada pelo método DCCA provou ser robusto para esse tipo de análise, porque foi possível avaliar a correlação cruzada em diferentes escalas de tamanho  $n$  e retirar a tendência (ver Figura 3.4).

Ao analisar as séries dos homicídios dolosos e homicídios tentados de Salvador-BA de 2004 a 2011 do ponto de vista da autocorrelação (DFA), constatamos comportamento diferente a depender da escala de tempo  $n$  considerada, ou seja comportamento persistente, antipersistente e sem memória (ver Figura 3.12). Para pequenas escalas de tempo  $n \leq 7$  dias, os homicídios dolosos e homicídios tentados são considerados eventos aleatórios com  $\alpha_{DFA} \cong 0,50$ . Já para escalas de semanas  $7 \leq n \leq 30$ , os delitos (crimes) apresentaram comportamento anti persistente ( $\alpha < 0,50$ ). Por fim, para escalas de tempo entre  $30 \leq n \leq 90$  dias, há uma transição de comportamento anti persistente/persistente. Identificamos correlação cruzada entre os homicídios dolosos e homicídios tentados para todas as escalas de tempo de tamanho  $n$ , por meio da função de covariância sem tendência ( $F_{DCCA}$ ) (ver Figura 3.12) e mensuramos o nível dessa correlação com o coeficiente de correlação cruzada  $\rho_{DCCA}$  (ver Figura 3.13). O coeficiente de correlação cruzada  $\rho_{DCCA}$  foi positivo para todas as escalas consideradas. Neste caso, podemos afirmar que se temos um aumento/decréscimo nos homicídios dolosos, podemos esperar um aumento/decréscimo nos homicídios tentados. Como o DFA, DCCA e  $\rho_{DCCA}$  oferece a possibilidade de elaborar cenários futuros, é possível pensar em aplicar esta metodologia em outros indicadores de criminalidade com o propósito de auxiliar o gerenciamento da segurança pública.

Reproduzimos o teste estatístico do  $\rho_{DCCA}$ , modelado por (PODOBNIK et al., 2011), cujos resultados encontram-se no apêndice A. Registramos no INPI o programa que calcula em diferentes escalas de tamanho  $n$  o DFA, DCCA e o  $\rho_{DCCA}$  (ver Apêndice B.1) e, após os resultados desta pesquisa, é possível testar estatisticamente os valores obtidos pelo coeficiente de correlação cruzada  $\rho_{DCCA}$ , considerando todas as escalas de tempo  $n$  e também para diferentes tamanhos de séries e magnitudes de confiança.

Entre as contribuições deste projeto de doutorado pode-se destacar: a revisão de literatura dos métodos DFA, DCCA e do coeficiente de correlação cruzada  $\rho_{DCCA}$ , as implementações teóricas realizadas nos métodos estatísticos supracitados e uma nova metodologia para análise de séries temporais.



Em relação aos objetivos desta tese de Doutorado, durante a pesquisa realizamos uma análise dos métodos da mecânica estatística: DFA, DCCA e  $\rho_{DCCA}$  e tal análise resultou em uma nova proposta teórica relativa a esses métodos, a saber:

□ Definimos via função da covariância sem tendência do método DCCA três grupos de correlação cruzada: sem correlação cruzada, correlação cruzada de longo alcance e correlação cruzada de curto alcance (ver seção 3.2);

□ Implementamos uma relação estatística entre os expoentes de autocorrelação do DFA e o expoente de correlação cruzada do DCCA, via diferenciação do coeficiente de correlação cruzada  $\rho_{DCCA}$ ; (ver seção 3.1)

□ Reproduzimos o teste estatístico do  $\rho_{DCCA}$  implementado por (PODOBNIK *et al.*, 2011), considerando todas as escalas temporais de tamanho  $n$  e diferentes magnitudes de confiança (ver seção 3.3 e Apêndice A);

□ Baseado nos itens supracitados, definimos uma nova metodologia de análise de séries temporais.

## 4.2 Sugestões para trabalhos futuros

A Estatística exerce papel fundamental enquanto área do conhecimento e diante do grande volume de dados e informações que são produzidos atualmente é relevante a apropriação desse conhecimento para aplicá-lo e se possível implementar novas metodologias. Assim, como uma possível continuidade da pesquisa, pretendemos estudar o significado estatístico da diferenciação do coeficiente de correlação cruzada  $\rho_{DCCA}$  implementado por (ZEBENDE; SILVA; MACHADO, 2013). E realizar outros estudos relativos ao teste estatístico do  $\rho_{DCCA}$  (PODOBNIK et al., 2011).

Pelos resultados encontrados na aplicação do DFA e do DCCA na série temporal de acidentes de táxi do banco de dados do Centro Especial de Auxílio ao Taxista localizada em Salvador-BA (ver seção 3.1); e do DFA, DCCA e  $\rho_{DCCA}$  nas séries de homicídios dolosos e homicídios tentados de Salvador-BA (ver seção 3.3) pretendemos dar continuidade ao estudo com outros indicadores de violência e criminalidade em diferentes localidades do Brasil. Se possível, identificar um padrão de comportamento entre esses indicadores. Além disso, produzir outros trabalhos que relacione os indicadores de violência e criminalidade com indicadores de outras categoria como saúde, educação e renda.

O estudo de séries temporais de processo de produção é algo relevante para a industria, devido a presença de variabilidade contida em qualquer processo de produção, o que juntamente com a pesquisa realizada por (BALESTRASSI, 2000), que identificou padrões em gráficos de controle estatístico de processos com análise de séries temporais e redes neurais nos motiva a desenvolver um trabalho que utilize os métodos aplicados na presente tese em séries temporais do Controle Estatístico de Processo - CEP e tal estudo será desenvolvido com o grupo de Pesquisa em Sistemas Complexos do SENAI CIMATEC.

Pretendemos, ainda, propor um indicador de insegurança a partir séries temporais dos indicadores de criminalidade disponibilizados pela Secretaria Nacional de Segurança Pública - SENASP.

Pensamos também, como continuidade da pesquisa, desenvolver uma variante do  $\rho_{DCCA}$  idealizado por (ZEBENDE, 2011).

Valores críticos do  $\rho_{DCCA}$

Tabela A.1: Valores positivos críticos do coeficiente de correlação cruzada  $\rho_{DCCA}$  (Equação 2.33) com 90 por cento de confiança tendo como base 10.000 pares de séries temporais (iid) com média zero e variância unitária com ajuste polinomial de grau 1 no *box*.

(iid)	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$N = 250$	0,11895	0,15574	0,16172	0,16566	0,17675	-	-
$N = 500$	0,08354	0,10946	0,11295	0,11400	0,11644	0,12461	-
$N = 1.000$	0,05910	0,07939	0,08147	0,08073	0,08116	0,08377	0,08978
$N = 2.000$	0,04152	0,05466	0,05641	0,05587	0,05550	0,05587	0,05750
$N = 4.000$	0,02938	0,03836	0,03951	0,03918	0,03878	0,03880	0,03923
$N = 8.000$	0,02051	0,02737	0,02831	0,02790	0,02762	0,02761	0,02771
$N = 16.000$	0,01504	0,01971	0,02000	0,01976	0,01958	0,01950	0,01948

Tabela A.2: Valores positivos críticos do coeficiente de correlação cruzada  $\rho_{DCCA}$  (Equação 2.33) com 90 por cento de confiança tendo como base 10.000 pares de séries temporais (iid) com média zero e variância unitária com ajuste polinomial de grau 2 no *box*.

(iid)	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$N = 250$	0,13222	0,13134	0,13218	0,13725	0,14895	-	-
$N = 500$	0,09322	0,09275	0,09252	0,09393	0,09724	0,10544	-
$N = 1.000$	0,07864	0,06495	0,06435	0,06491	0,06606	0,06863	0,07417
$N = 2.000$	0,08070	0,04554	0,04511	0,04522	0,04561	0,04647	0,04823
$N = 4.000$	0,02572	0,03235	0,03191	0,03196	0,03207	0,03236	0,03286
$N = 8.000$	0,01730	0,02264	0,02237	0,02239	0,02244	0,02250	0,02270
$N = 16.000$	0,00558	0,01648	0,01627	0,01625	0,01628	0,01631	0,01637

Tabela A.3: Valores positivos críticos do coeficiente de correlação cruzada  $\rho_{DCCA}$  (Equação 2.33) com 90 por cento de confiança tendo como base 10.000 pares de séries temporais (iid) com média zero e variância unitária com ajuste polinomial de grau 3 no *box*.

(iid)	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$N = 250$	0,15992	0,14878	0,15609	0,16982	0,16982	-	-
$N = 500$	0,11091	0,10274	0,10493	0,10865	0,10865	0,11725	-
$N = 1.000$	0,07874	0,07274	0,07313	0,07433	0,07433	0,07710	0,08387
$N = 2.000$	0,05596	0,05166	0,05176	0,05221	0,05221	0,05318	0,05502
$N = 4.000$	0,04012	0,03707	0,03703	0,03720	0,03720	0,03748	0,03802
$N = 8.000$	0,02761	0,02548	0,02542	0,02547	0,02547	0,02553	0,02572
$N = 16.000$	0,02001	0,01848	0,01841	0,01843	0,01843	0,01848	0,01852

Tabela A.4: Valores positivos críticos do coeficiente de correlação cruzada  $\rho_{DCCA}$  (Equação 2.33) com 95 por cento de confiança tendo como base 10.000 pares de séries temporais (iid) com média zero e variância unitária com ajuste polinomial de grau 1 no *box*.

(iid)	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$N = 250$	0,14128	0,18506	0,19222	0,19681	0,20998	-	-
$N = 500$	0,09921	0,13014	0,13432	0,13554	0,13845	0,14814	-
$N = 1.000$	0,07017	0,09420	0,09667	0,09582	0,09630	0,09936	0,10655
$N = 2.000$	0,04937	0,06498	0,06703	0,06640	0,06599	0,06645	0,06839
$N = 4.000$	0,03488	0,04562	0,04700	0,04658	0,04609	0,04611	0,04664
$N = 8.000$	0,02439	0,03251	0,03363	0,03317	0,03282	0,03280	0,03293
$N = 16.000$	0,01782	0,02338	0,02373	0,02346	0,02324	0,02314	0,02312

Tabela A.5: Valores positivos críticos do coeficiente de correlação cruzada  $\rho_{DCCA}$  (Equação 2.33) com 95 por cento de confiança tendo como base 10.000 pares de séries temporais (iid) com média zero e variância unitária com ajuste polinomial de grau 2 no *box*.

(iid)	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$N = 250$	0,15707	0,15598	0,15700	0,16295	0,17685	-	-
$N = 500$	0,11069	0,11010	0,10982	0,11153	0,11547	0,12519	-
$N = 1.000$	0,07772	0,07714	0,07643	0,07711	0,07848	0,08148	0,08809
$N = 2.000$	0,05510	0,05414	0,05363	0,05377	0,05423	0,05524	0,05734
$N = 4.000$	0,03941	0,03840	0,03789	0,03794	0,03807	0,02841	0,03902
$N = 8.000$	0,02740	0,02691	0,02659	0,02661	0,02668	0,02675	0,02700
$N = 16.000$	0,01978	0,01953	0,01928	0,01926	0,01929	0,01933	0,01940

Tabela A.6: Valores positivos críticos do coeficiente de correlação cruzada  $\rho_{DCCA}$  (Equação 2.33) com 95 por cento de confiança tendo como base 10.000 pares de séries temporais (iid) com média zero e variância unitária com ajuste polinomial de grau 3 no *box*.


(iid)	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$N = 250$	0,18993	0,17674	0,17939	0,18547	0,20180	-	-
$N = 500$	0,13186	0,12212	0,12257	0,12472	0,12918	0,13940	-
$N = 1.000$	0,09357	0,08644	0,08647	0,08692	0,08836	0,09160	0,09958
$N = 2.000$	0,06639	0,06129	0,06122	0,06141	0,06194	0,06309	0,06530
$N = 4.000$	0,04754	0,04393	0,04383	0,04388	0,04408	0,04441	0,04507
$N = 8.000$	0,03283	0,03029	0,03020	0,03023	0,03029	0,03037	0,03059
$N = 16.000$	0,02376	0,02193	0,02186	0,02184	0,02187	0,02190	0,02198

# Pedido de registro de programa de computador ao INPI

## B.1 Programa que calcula o DFA, DCCA e $\rho_{DCCA}$

**INPI**

**PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR**

protocolo  
**INPI** INSTITUTO BRASILEIRO DE PROPRIEDADE INDUSTRIAL  
 22/07/2013 01113000445  
 12:08 REBR  
  
 BR 51 2013 000758 5

**IDENTIFICAÇÃO DO PEDIDO** (Para uso do INPI)

Número do Pedido \_\_\_\_\_ Protocolo, Data e Hora \_\_\_\_\_

**DADOS DO AUTOR DO PROGRAMA**

Nº de Autores **3** Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assiné.

CPF\* \_\_\_\_\_

Nome **GILNEY FIGUEIRA ZEBENDE**

Nome Abreviado, pseudônimo ou sinal convencional (se houver) **GILNEY ZEBENDE**

Data de Nascimento \_\_\_\_\_ Nacionalidade **BRASILEIRO**

Endereço \_\_\_\_\_

Cidade **FEIRA DE SANTANA** UF **BA** País **BRASIL**

CEP \_\_\_\_\_ Telefone \_\_\_\_\_ FAX \_\_\_\_\_

E-mail **gilneyz@fieb.org.br**

**DADOS DO TITULAR DOS DIREITOS PATRIMONIAIS**

Nº de Titulares **1** Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assiné.

CPF/CNPJ\* **03795071000116**

Nome/Razão Social **SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL**

Nome abreviado, pseudônimo ou sinal convencional (se houver) \_\_\_\_\_

Data de Nascimento \_\_\_\_\_ Nacionalidade/Origem **BRASIL**

Endereço **RUA EDISTIO PONDÉ,342, STIEP**

Cidade **SALVADOR** UF **BA** País **BRASIL**

CEP **41.770-395** Telefone **7134629586** FAX **7134629599**

E-mail **cimatec-npi@fieb.org.br**

**SIM**, este Titular é Pessoa Jurídica. Caso afirmativo, assinale a melhor classificação:

Órgão Público  Sociedade com Intuito não Econômico  Microempresa  Software House

Instituição Pública de Ensino ou Pesquisa  Instituição Privada de Ensino ou Pesquisa  Outras

**ENDEREÇO PARA CORRESPONDÊNCIA E CONTATO** (Preencha apenas o necessário)

Toda correspondência será enviada para:  O Procurador ou  O Titular acima ou

Escaninho nº \_\_\_\_\_  Representação INPI em: \_\_\_\_\_  O Endereço abaixo:

Nome **SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL**

Endereço **RUA EDISTIO PONDÉ,342, STIEP**

Cidade **SALVADOR** UF **BA** País **BRASIL**

CEP **41.770-395** Telefone **7134629586** FAX **7134629599**

E-mail **cimatec-npi@fieb.org.br**

Modelo 1 (folha 1/2) E

Figura B.1: Primeiro pedido de registro de programa de computador ao INPI - Página 1/3

DADOS DO PROGRAMA			
Título	rhoDCCAmult		
Data de Criação do Programa	15/03/2013		
Regime de Guarda	<input checked="" type="checkbox"/> COM SIGILO <input type="checkbox"/> SEM SIGILO		
Linguagens	C++		
Classificação do Campo de Aplicação	MT - 06   MT - 01   -   -   -		
Classificação do Tipo de Programa	SM - 01   IT - 04   AP - 01   -   -		
<input type="checkbox"/> SIM, este Programa é Modificação Tecnológica ou Derivação. Caso afirmativo, informe Título do Programa Original e (se houver) Número de Registro:			
Título do Programa Original			
<input type="checkbox"/> SIM, este Registro é composto por obra(s) de outra(s) natureza(s) de ordem intelectual. Caso afirmativo assinala-a(s) abaixo:			
<input type="checkbox"/> Literária <input type="checkbox"/> Musical <input type="checkbox"/> Artes Plásticas <input type="checkbox"/> Audio-Visual <input type="checkbox"/> Arquitetura <input type="checkbox"/> Engenharia			
DOCUMENTOS ANEXADOS (Informe as quantidades de documentos, não o número de páginas)			
Quant	Nome	Quant	Nome
01	Guia de Recolhimento		Contrato de Trabalho/Prestação de Serviço
	Procuração	01	Involucros/mídia eletrônica Utilizados
03	Termo de Cessão		Contrato/Estatuto Social e Alterações (ou equivalente)
	Termo de Autorização para Modificações Tecnológicas ou Derivações	01	Autorização para Cópia do CD
		01	Outros(especificar) <i>fontes CNJ SENAI mº 50/2011</i>
DECLARAÇÕES			
<p>DECLARO, PARA TODOS OS FINS DE DIREITO:</p> <p>A) que estou ciente de TODAS AS RECOMENDAÇÕES constantes do "Manual do Usuário de Registro de Programas de Computador", ESPECIALMENTE NO QUE TANGE AO TÍTULO E AOS DOCUMENTOS DO PROGRAMA, bem como da legislação pertinente ao assunto, constante dos anexos "A"; "B"; "C"; "E" e "F", do referido Manual;</p> <p>B) que se deixar de solicitar a prorrogação do sigilo, nos casos necessários, estarei desistindo desse caráter de guarda dos documentos de programa do presente depósito, na forma do art. 3º, § 2º, da Lei 9.609, de 12 de fevereiro de 1998;</p> <p>C) que, se devido à qualidade do papel ou à qualidade gráfica dos documentos sigilosos anexos ao presente, houver deterioração ou perda de seu conteúdo, nenhuma responsabilidade caberá ao INPI, desde que mantida a inviolabilidade dos involucros (ressalvadas as hipóteses de serem abertos por ordem judicial ou motivo de força maior);</p> <p>D) que em caso de perda do SIGILO ou dos documentos, por culpa exclusiva do INPI, a indenização por perdas e danos, porventura cabível, estará limitada a 20 (vinte) salários mínimos;</p> <p>E) que devo manter guardado, em segurança e inviolado, o COMPARTIMENTO "3" do Involucro especial para depósito, que é restituído pelo INPI, para fins de recomposição do arquivo do Instituto, no caso de sua destruição total ou parcial por algum tipo de sinistro;</p> <p>F) que deverei manter endereço atualizado junto à Divisão de Registro de Programa de Computador, a fim de garantir o recebimento das comunicações relativas ao andamento do meu pedido/registro, ressalvando o INPI de qualquer responsabilidade decorrente da não observação deste preceito.</p>			
DADOS DO PROCURADOR			
CPF/CNPJ*	Código do Procurador (se houver)		
Nome			
Endereço			
Cidade	UF	País	
CEP	Telefone	FAX	
E-mail			
DECLARO, SOB AS PENAS DA LEI, SEREM VERDADEIRAS AS INFORMAÇÕES PRESTADAS			
Salvador, 22/07/2013		Assinatura/Carimbo	
Local/Data		 Leona Peter C. S. Andrade Diretor Regional do SENAI-BF Modelo 1 (folha 3/2) E	

Figura B.2: Primeiro pedido de registro de programa de computador ao INPI - Página 2/3

REGISTRO DE PROGRAMA DE COMPUTADOR - CONTINUAÇÃO					
Utilize este ANEXO, em quantas folhas forem necessárias, para complementar as informações dos formulários "Pedido de Registro de Programa de Computador" e "Folha de Petição" (DIRTEC).					
<b>DADOS DOS DEMAIS AUTORES DO PROGRAMA</b>					
CPF*					
Nome	ARLEYS PEREIRA NUNES DE CASTRO				
Nome Abreviado, pseudônimo ou sinal convencional (se houver)	ARLEYS DE CASTRO				
Data de Nascimento		Nacionalidade	BRASILEIRO		
Endereço					
Cidade	LAURO DE FREITAS	UF	BA	Pais	BRASIL
CEP	42700-000	Telefone		FAX	
E-mail	arleys.castro@gmail.com				
<b>DADOS DOS DEMAIS AUTORES DO PROGRAMA</b>					
CPF*					
Nome	ALOISIO MACHADO DA SILVA FILHO				
Nome Abreviado, pseudônimo ou sinal convencional (se houver)	ALOISIO FILHO				
Data de Nascimento		Nacionalidade	BRASILEIRO		
Endereço					
Cidade	SALVADOR	UF	BA	Pais	BRASIL
CEP	40352-055	Telefone		FAX	71 3462-9599
E-mail	aloisioestatistico@yahoo.com.br				

Figura B.3: Primeiro pedido de registro de programa de computador ao INPI - Página 3/3

**B.2 Programa que aplica o teste estatístico do  $\rho_{DCCA}$**



**PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR**

protocolo

**INPI** INSTITUTO NACIONAL DE PROPRIEDADE INTELECTUAL

02/04/2014 01114000084 14:36 REBR



BR 51 2014 000339 6

**IDENTIFICAÇÃO DO PEDIDO** (Para uso do INPI)

Número do Pedido \_\_\_\_\_ Protocolo, Data e Hora \_\_\_\_\_

**DADOS DO AUTOR DO PROGRAMA**

Nº de Autores  Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.

CPF\* \_\_\_\_\_

Nome

Nome Abreviado, pseudônimo ou sinal convencional (se houver)

Data de Nascimento \_\_\_\_\_ Nacionalidade

Endereço \_\_\_\_\_

Cidade  UF  País

CEP \_\_\_\_\_ Telefone \_\_\_\_\_ FAX \_\_\_\_\_

E-mail

**DADOS DO TITULAR DOS DIREITOS PATRIMONIAIS**

Nº de Titulares  Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.

CPF/CNPJ\*

Nome/Razão Social

Nome abreviado, pseudônimo ou sinal convencional (se houver) \_\_\_\_\_

Data de Nascimento \_\_\_\_\_ Nacionalidade/Origem

Endereço

Cidade  UF  País

CEP  Telefone  FAX

E-mail

**SIM**, este Titular é Pessoa Jurídica. Caso afirmativo, assinale a melhor classificação:

Órgão Público     Sociedade com Intuito não Econômico     Microempresa     Software House

Instituição Pública de Ensino ou Pesquisa     Instituição Privada de Ensino ou Pesquisa     Outras

**ENDEREÇO PARA CORRESPONDÊNCIA E CONTATO** (Preencha apenas o necessário)

Toda correspondência será enviada para:  O Procurador ou  O Titular acima ou

Escaninho nº \_\_\_\_\_  Representação INPI em: \_\_\_\_\_  O Endereço abaixo:

Nome

Endereço

Cidade  UF  País

CEP  Telefone  FAX

E-mail

Figura B.4: Segundo pedido de registro de programa de computador ao INPI - Página 1/3



DADOS DO PROGRAMA						
Título <b>rhoDCCAsort teste estatístico</b>						
Data de Criação do Programa <b>28/03/2013</b>		Regime de Guarda <input checked="" type="checkbox"/> COM SIGILO <input type="checkbox"/> SEM SIGILO				
Linguagens <b>C++</b>						
Classificação do Campo de Aplicação <b>MT - 06</b>		<b>MT - 01</b>		<b>-</b>		<b>-</b>
Classificação do Tipo de Programa <b>SM - 01</b>		<b>IT - 04</b>		<b>AP - 01</b>		<b>-</b>
<input type="checkbox"/> <b>SIM</b> , este Programa é Modificação Tecnológica ou Derivação. Caso afirmativo, informe Título do Programa Original e (se houver) Número de Registro:						
Título do Programa Original						
<input type="checkbox"/> <b>SIM</b> , este Registro é composto por obra(s) de outra(s) natureza(s) de ordem intelectual. Caso afirmativo assinala-a(s) abaixo:						
<input type="checkbox"/> Literária <input type="checkbox"/> Musical <input type="checkbox"/> Artes Plásticas <input type="checkbox"/> Áudio-Visual <input type="checkbox"/> Arquitetura <input type="checkbox"/> Engenharia						
DOCUMENTOS ANEXADOS (Informe as quantidades de documentos, não o número de páginas)						
<u>Quant</u>	<u>Nome</u>	<u>Quant</u>	<u>Nome</u>			
<b>01</b>	Guia de Recolhimento		Contrato de Trabalho/Prestação de Serviço			
	Procuração	<b>01</b>	Involúcos/mídia eletrônica Utilizados <b>CD</b>			
<b>04</b>	Termo de Cessão		Contrato/Estatuto Social e Alterações (ou equivalente)			
	Termo de Autorização para Modificações Tecnológicas ou Derivações	<b>01</b>	Autorização para Cópia do CD			
		<b>01</b>	Outros(especificar) <b>Biblioteca CNI SENAI nº 30/2011</b>			
DECLARAÇÕES						
DECLARO, PARA TODOS OS FINS DE DIREITO: A) que estou ciente de <b>TODAS AS RECOMENDAÇÕES</b> constantes do "Manual do Usuário de Registro de Programas de Computador", <b>ESPECIALMENTE NO QUE TANGE AO TÍTULO E AOS DOCUMENTOS DO PROGRAMA</b> , bem como da legislação pertinente ao assunto, constante dos anexos "A", "B", "C", "E" e "F", do referido Manual; B) que se deixar de solicitar a prorrogação do sigilo, nos casos necessários, estarei desistindo desse caráter de guarda dos documentos do programa do presente depósito, na forma do art. 3º, § 2º, da Lei 9.809, de 12 de fevereiro de 1998; C) que, se devido à qualidade do papel ou à qualidade gráfica dos documentos sigilosos anexos ao presente, houver deterioração ou perda de seu conteúdo, nenhuma responsabilidade caberá ao INPI, desde que mantida a inviolabilidade dos invólucros (ressalvadas as hipóteses de serem abertos por ordem judicial ou motivo de força maior); D) que em caso de perda do SIGILO ou dos documentos, por culpa exclusiva do INPI, a indenização por perdas e danos, porventura cabível, estará limitada a 20 (vinte) salários mínimos; E) que devo manter guardado, em segurança e inviolado, o COMPARTIMENTO "3" do invólucro especial para depósito, que é restituído pelo INPI, para fins de recomposição do arquivo do Instituto, no caso de sua destruição total ou parcial por algum tipo de sinistro; F) que deverei manter endereço atualizado junto à Divisão de Registro de Programa de Computador, a fim de garantir o recebimento das comunicações relativas ao andamento do meu pedido/registro, ressalvando o INPI de qualquer responsabilidade decorrente da não observação deste preceito.						
DADOS DO PROCURADOR						
CPF/CNPJ*		Código do Procurador (se houver)				
Nome						
Endereço						
Cidade		UF	País			
CEP		Telefone	FAX			
E-mail						
DECLARO, SOB AS PENAS DA LEI, SEREM VERDADEIRAS AS INFORMAÇÕES PRESTADAS						
<b>Salvador, 02/04/14</b>						
Local/Data			Assinatura/Carimbo			

Figura B.5: Segundo pedido de registro de programa de computador ao INPI - Página 2/3

**REGISTRO DE PROGRAMA DE COMPUTADOR - CONTINUAÇÃO**

Utilize este ANEXO, em quantas folhas forem necessárias, para complementar as informações dos formulários "Pedido de Registro de Programa de Computador" e "Folha de Petição" (DIRTEC).

---

**DADOS DOS DEMAIS AUTORES DO PROGRAMA**

---

CPF\*

Nome **ARLEYS PEREIRA NUNES DE CASTRO**

Nome Abreviado, pseudônimo ou sinal convencional (se houver) **ARLEYS DE CASTRO**

Data de Nascimento  Nacionalidade **BRASILEIRO**

Endereço

---

Cidade **LAURO DE FREITAS** UF **BA** País **BRASIL**

CEP **42700-000** Telefone  FAX **71 3462-9599**

E-mail **arleys.castro@gmail.com**

---

CPF\*

Nome **ALOISIO MACHADO DA SILVA FILHO**

Nome Abreviado, pseudônimo ou sinal convencional (se houver) **ALOISIO FILHO**

Data de Nascimento  Nacionalidade **BRASILEIRO**

Endereço

---

Cidade **SALVADOR** UF **BA** País **BRASIL**

CEP **40352-055** Telefone  FAX **71 3462-9599**

E-mail **aloisioestatistico@yahoo.com.br**

---

CPF\*

Nome **MARCUS FERNANDES DA SILVA**

Nome Abreviado, pseudônimo ou sinal convencional (se houver)

Data de Nascimento  Nacionalidade **BRASILEIRO**

Endereço

---

Cidade **SALVADOR** UF **BA** País **BRASIL**

CEP **41150-000** Telefone  FAX **71 3462-9599**

E-mail **marcus.fisico@gmail.com**

Figura B.6: Segundo pedido de registro de programa de computador ao INPI - Página 3/3

---

## Referências

---

- AGUIRRE, L. A. *Introdução À Identificação de Sistemas*. 2<sup>a</sup>. ed. Minas Gerais: UFMG, 2000.
- ALAOUI, M. E.; BENBACHIR, S. Multifractal detrended cross-correlation analysis in the mena area. *Physica A*, p. 5985–5993, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378437113007139>>.
- ANTENEODO, C. Processos estocásticos. In: *V Escola do Centro Brasileiro de Pesquisas Físicas*. Rio de Janeiro: [s.n.], 2004. p. 1–34. Disponível em: <<http://mesonpi.cat.cbpf.br/e2004/docs/PG2-cban.pdf>>.
- BAK, P. *How Nature Works: The Science of Self-Organized Criticality*. Oxford: Oxford University Press, 1997.
- BALESTRASSI, P. P. *Identificação de padrões em gráficos de controle estatístico de processos, em tempo real, utilizando séries temporais e redes neurais artificiais*. Tese (Doutorado em Engenharia de Produção) — Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2000.
- BALLINI, R. *Análise e previsão de vazões utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas*. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, São Paulo, 2000.
- BALOCCHI, R.; VARANINI, M.; MACERRATA, A. Quantifying different degrees of coupling in detrended cross-correlation analysis. *Europhysics Letters*, v. 101, p. 20011, 2013.
- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. *Estatística para cursos de engenharia e informática*. 3<sup>a</sup>. ed. São Paulo: Atlas, 2010.
- BARBOSA, E. B. de M. *Estudo da variabilidade de séries temporais ambientais através de espectros de assimetria implementados em software científicos livre*. Dissertação (Mestrado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, 2007.
- BARBOSA, R. M. *Descobrendo a Geometria Fractal*. 2<sup>a</sup>. ed. Belo Horizonte: Coleção tendências em educação matemática, 2005.
- BLYTHE, D. A. J. A rigorous and efficient asymptotic test for power-law cross-correlation. *Cornell University*, p. 1–23, 2013. Disponível em: <<http://arxiv-web3.library.cornell.edu/abs/1309.4073>>.

BOLFARINE, H.; BUSSAB, W. O. *Elementos de amostragem*. São Paulo: ABE- Projeto Fischer Editora Edgard Blucher, 2005.

BOLFARINE, H.; SANDOVAL, M. *Introdução à Inferência Estatística*. [S.l.]: Coleção de matemática Aplicada - Sociedade Brasileira de Matemática, 2010.

BOX, G. P.; JENKINS, G. M. *Time series analysis: forecasting and control*. [S.l.]: New York: Holden Day, 1976.

BUSSAB, W.; MORETTIN, P. *Estatística Básica*. São Paulo: Saraiva, 2003.

CAO, G.; XU, L.; CAO, J. Multifractal detrended cross-correlations between the chinese exchange market and stock market. *Physica A*, v. 391, p. 4855–4866, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2012.05.035>>.

CASTRO, A. P. N. *Método de detecção de padrões na dinâmica dos movimentos Complexos na deglutição de um indivíduo*. Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) — Faculdade de Tecnologia SENAI Cimatec, Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Salvador-BA, 2012.

CASTRO, A. P. N.; MACHADO, A.; ZEBENDE, G. F. Análise da disfagia através da autocorrelação e correlação cruzada: Modelo computacional para a deglutição. In: *IV Encontro de Pesquisa Operacional no Nordeste IV ERPO NE*. Salvador-BA: [s.n.], 2013.

CASTRO, F. C. C. *Estudo de séries temporais aplicado a perfis de poços de petróleo*. Dissertação (Mestrado em geoengenharia de reservatórios) — Universidade Estadual de Campinas, Campinas, São Paulo, 1995.

CHEN, Z.; IVANOV, P. C.; HU, K.; STANLEY, H. Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E*, v. 65, p. 041107, 2002. Disponível em: <<http://polymer.bu.edu/hes/articles/cihs02.pdf>>.

COSTA, R. L.; VASCONCELOS, G. Long-range correlations and nonstationarity in the brazilian stock market. *Physica A*, v. 239, p. 231 – 248, 2003. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378437103006071>>.

CRUZ, T. S. *Leis de escala e dimensão fractal em filmes: microscopia de força atômica e técnicas eletroquímicas*. Tese (Doutorado em Física) — Universidade estadual de campinas, Campinas-SP, 2002. Disponível em: <<http://webbif.ifi.unicamp.br/tesesOnline/teses/IF1352.pdf>>.

DEVLIN, S.; GNANADESIKAN, R.; KETTENRING, J. Robust estimation and outliers detection with correlation coefficients. *Biometrika*, v. 62, p. 531–545, 1975. Disponível em: <<http://biomet.oxfordjournals.org/content/62/3/531>>.

- DINIZ, A. M. F. A. *Séries temporais de memória longa com aplicações ao controle motor estudo de tarefas de tapping repetido*. Tese (Em motricidade humana especialidade de métodos quantitativos) — Universidade Técnica de Lisboa, Lisboa, 2008. Disponível em: <<https://www.repository.utl.pt/handle/10400.5/489>>.
- EHLERS, R. S. *Análise de Séries Temporais*. Departamento de Estatística, UFPR. Disponível em <http://leg.est.ufpr.br/ehlers/notas>. Acesso em: 23 abril de 2013. 2007.
- FEDER, J. *Fractals*. New York,: Plenum Press, New York, 1988.
- FENG, M.; WEI, Y.; HUANG, D. Multifractal detrended cross-correlation analysis between the chinese stock market and surrounding stock markets. *Physica A*, v. 392, p. 1659–1670, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2012.12.010>>.
- FERNANDES, M.; CASTRO, A.; MACHADO, A. Quantificando a influência do mercado de câmbio nos preços do milho e da soja no município de barreiras. *Conjuntura & Planejamento*, v. 182, p. 42 – 49, 2014.
- FINANCE. 2013. Disponível em: <<http://finance.yahoo.com/>>.
- FONSECA, J. S. da; MARTINS, G. de A.; TOLEDO, G. L. *Estatística Aplicada*. 2<sup>a</sup>. ed. São Paulo: Atlas, 1985.
- GALHARDO, C.; BARBOSA, L.; PENNA, T. Análise de sinais não estacionários de pressão arterial de ratos sujeitos à deservação. In: *XXXI Encontro Nacional de Física da Matéria Condensada*. Águas de Lindóia: [s.n.], 2008. Disponível em: <[http://complex.if.uff.br/\\_media/ratos.pdf](http://complex.if.uff.br/_media/ratos.pdf)>.
- GALHARDO, C.; PENNA, T.; MENEZES, M.; SOARES, P. Detrended fluctuation analysis of a systolic blood pressure control loop. *New Journal of Physics*, v. 11, p. 103005, 2009. Disponível em: <<http://iopscience.iop.org/1367-2630/11/10/103005/>>.
- GALHARDO, C. E. C. *Análise e modelagem estocástica do barorreflexo através de séries temporais de pressão arterial sistólica*. Tese (Doutorado em Física) — Universidade Federal Fluminense, Instituto de Física, 2010.
- GLEISER, I. *Caos e complexidade a evolução do pensamento econômico*. Rio de Janeiro: [s.n.], 2002.
- GOMES, D.; AGOULMINE, N.; FONSECA, N.; SOUZA, J. N. Neurocomputação na estimativa do parâmetro de hurst. In: *19º Simpósio brasileiro de redes de computadores*. Florianópolis, Santa Catarina: [s.n.], 2001. p. 306–321.
- GRANGER, C. Long memory relationship and the aggregation of dynamic models. *Journal of econometrics*, v. 14, p. 227–238, 1980.
- GU, R.; SHAO, Y.; WANG, Q. Is the efficiency of stock market correlated with multifractality? an evidence from the shanghai stock market. *Physica A*, v. 392, p. 316–370, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2012.09.008>>.

- GUJARATI, D. N. *Econometria básica*. 3<sup>a</sup>. ed. São Paulo: Makron Books, 2000.
- GUJARATI, D. N.; PORTER, D. *Econometria Básica*. 5<sup>a</sup>. ed. Porto Alegre: Bookman, 2011.
- HAIJIAN, S. S.; MOVAHED, M. S. Multifractal detrended cross-correlation analysis of sunspot numbers and river flow fluctuations. *Physica A*, v. 389, p. 4942–4957, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2010.06.025>>.
- HE, L.; CHEN, S. Nonlinear bivariate dependency of price-volume relationships in agricultural commodity futures markets: A perspective from multifractal detrended cross-correlation analysis. *Physica A*, v. 390, p. 297–308, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2010.09.018>>.
- HE, L.-Y.; CHEN, S.-P. A new approach to quantify power-law cross-correlation and its application to commodity markets. *Physica A*, v. 390, p. 3806–3814, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2011.06.013>>.
- HORVATIC, D.; STANLEY, H. E.; PODOBNIK, B. Detrended cross-correlation analysis for non-stationary time series with periodic trends. *Eur. Phys. Lett*, v. 94, p. 18007, 2011. Disponível em: <<http://polymer.bu.edu/hes/articles/hsp11.pdf>>.
- HOSKING, J. Fractional differencing. *Biometrika*, v. 28, p. 165–176, 1981. Disponível em: <<http://www2.imperial.ac.uk/~ejm/M3S8/Problems/hosking81.pdf>>.
- HU, K.; IVANOV, P. C.; CHEN, Z.; CARPENA, P.; STANLEY, H. Effect of trends on detrended fluctuation analysis. *Physical Review E*, v. 64, n. 19, p. 011114, 2001. Disponível em: <<http://link.aps.org/doi/10.1103/PhysRevE.64.011114>>.
- HURST, E. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, v. 116, p. 770–799, 1951.
- HURST, H.; BLACK, R.; SIMAIKA, Y. *Long-Term Storage: an experimental Study*. 1<sup>a</sup>. ed. Constable: London, 1965.
- IVANOV, P. et al. Scaling and universality in heart rate variability distributions. *Physica A*, v. 249, p. 587–593, 1998. Disponível em: <[http://dx.doi.org/10.1016/S0378-4371\(97\)00522-0](http://dx.doi.org/10.1016/S0378-4371(97)00522-0)>.
- JANOS, M. *Geometria fractal*. Rio de Janeiro: Editora Ciência Moderna, 2008.
- KANG, D. D.; LEE, D. I.; JUNG, J.-W.; KIM, K. Dynamical characteristics in time series between pm10 and wind speed. In: *Proceedings of the 2013 International Conference on Environment, Energy, Ecosystems and Development*. [s.n.], 2013. Disponível em: <<http://www.europment.org/library/2013/venice/bypaper/EEEAD/EEEAD-09.pdf>>.
- KANTELHARDT, J. W. Fractal and multifractal time series. *Institute of Physics, Martin-Luther-University*, p. 1–59, 2008. Disponível em: <<http://arxiv.org/abs/0804.0747>>.

KRISTOUFEK, L. Multifractal height cross-correlation analysis: A new method for analyzing long-range cross-correlations. *Europhys. Lett.*, v. 95, p. 68001, 2011. Disponível em: <http://iopscience.iop.org/0295-5075/95/6/68001>.

KRISTOUFEK, L. *Long - Range Cross-Correlations: Tests, estimators and applications*. Tese (Doutorado) — Charles University in Prague Faculty of Social Sciences Institute of Economic Studies, 2013. Disponível em: <http://ies.fsv.cuni.cz/work/index/show/id/1950/lang/en>.

LAPA, T. *Modelagem dinâmica para a previsão da demanda de automóveis no Brasil*. Dissertação (Mestrado em Engenharia de Produção) — Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 1998.

LATIF, S. A. *Medidas de dependência local para séries temporais*. Tese (Doutorado em Ciências) — Universidade de São Paulo, São Paulo, 2008.

LIEBOVITCH, L.; YANG, W. Transition from persistent to antipersistent correlation in biological systems. *PHYSICAL REVIEW E*, v. 56, p. 4557–4566, 1997. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.56.4557>.

LIN, A.; SHANG, P.; ZHAO, X. The cross-correlations of stock markets based on dcca and time-delay dcca. *Nonlinear Dynamics*, v. 67, p. 425–435, 2012. Disponível em: <http://link.springer.com/article/10.1007%2Fs11071-011-9991-8>.

LIU, L.; CHEN, C.-C.; WAN, J. Is world oil market one great pool?: An example from china's and international oil markets. *Economic Modelling*, v. 35, p. 364 – 373, 2013.

LIU, Y. et al. The statistical properties of the volatility of price fluctuations. *Physical Review E*, v. 60, p. 1390–1400, 1999. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.60.1390>.

MACHADO, A. *Correlação de longo alcance em indicadores de criminalidade de Salvador-BA: DFA e DCCA*. Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) — Faculdade de Tecnologia SENAI Cimatec, Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Salvador-BA, 2009.

MACHADO, A.; CASTRO, A. P. N.; SILVA, M. F.; ZEBENDE, G. F. Autocorrelação e correlação cruzada em séries temporais de acidentes de trânsito de salvador-ba. In: *XXX Encontro de Físicos do Norte e Nordeste*. Salvador-BA: [s.n.], 2012.

MACHADO, A.; CASTRO, A. P. N.; ZEBENDE, G. F. Coeficiente de correlação cruzada múltiplo: Teorias e aplicações em séries temporais. In: *IV Encontro de Pesquisa Operacional no Nordeste IV ERPO NE*. Salvador-BA: [s.n.], 2013. Disponível em: <http://ictios.org/iv-erpon/programacao-e-resumos-das-apresentacoes/>.

- MACHADO, A.; SILVA, M. da; ZEBENDE, G. F. Autocorrelation and cross-correlation in time series of homicide and attempted homicide. *Physica A*, v. 400, p. 12–19, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378437114000193>>.
- MACHADO, A.; SOARES, J. S.; ZEBENDE, G. F. Dfa and dcca applied in a time series of thefts and robbery of vehicles (salvador-ba). In: *XXXII Encontro Nacional de Física da Matéria Condensada*. Águas de Lindóia: [s.n.], 2009.
- MACHADO, A.; ZEBENDE, G. F.; ALVES, L. A. Modelagem estatística da criminalidade: o estado da arte. *Conjuntura & Planejamento*, v. 178, p. 52–61, 2013. Disponível em: <<http://pt.scribd.com/doc/145845132/c-p-178>>.
- MACHADO, A.; ZEBENDE, G. F.; CASTRO, A.; FERNANDES, M. Modelagem do teste estatístico do coeficiente de correlação cruzada sem aplicação sem tendência rhodcca: Teoria e aplicação. In: *IV Workshop de Pesquisa, Tecnologia e Inovação - PTI do SENAI CIMATEC*. Salvador-BA: [s.n.], 2014.
- MACHADO, A.; ZEBENDE, G. F.; FERNANDES, M.; TERESOPOLIS, M. Coeficiente de correlação cruzada: Uma aplicação nas séries temporais dos homicídios dolosos e homicídios tentados de salvador-ba. In: *21º Simpósio Nacional de Probabilidade e Estatística*. Natal: [s.n.], 2014.
- MACHADO, A.; ZEBENDE, G. F.; SILVA, P. D. Correlação de longo alcance em séries de furto de veículos e roubo de veículos de salvador-ba: Uma aplicação utilizando o dfa e dcca. In: *XII Escola de Modelos de Regressão*. Fortaleza-CE: [s.n.], 2011.
- MACHADO, A.; ZEBENDE, G. F.; SOARES, J. S. Estatística policial de salvador-ba: Uma proposta de modelagem utilizando dfa. In: *XL Simpósio Brasileiro de Pesquisa Operacional*. João Pessoa: [s.n.], 2008.
- MACHADO, A.; ZEBENDE, G. F.; SOARES, J. S. Detrended fluctuation analysis e detrended cross-correlation analysis: Uma aplicação em indicadores sociais de criminalidade. In: *XLI Simpósio Brasileiro de Pesquisa Operacional*. Porto Seguro-BA: [s.n.], 2009. p. 3228–3229. Disponível em: <[http://www.din.uem.br/sbpo/sbpo2009/posters/arquivos\\_posteres/59113.pdf](http://www.din.uem.br/sbpo/sbpo2009/posters/arquivos_posteres/59113.pdf)>.
- MADDALA, G. S. *Introdução à econometria. Tradução e revisão técnica: Leonardo Weller e Paulo Tafner*. 3ª. ed. Rio de Janeiro: LTC, 2003.
- MANDELBROT, B. *The Fractal Geometry of Nature*. [S.l.]: W. H. Freeman and Company, New York, 1982.
- MANTEGNA, R.; STANLEY, H. Scaling behavior in the dynamics of an economic index. *Nature*, v. 367, p. 46–49, 1995. Disponível em: <<http://www.nature.com/nature/journal/v376/n6535/abs/376046a0.html>>.



- MARINHO, E.; SOUSA, A.; ANDRADE, R. Using detrended cross-correlation analysis in geophysical data. *Physica A*, v. 392, p. 2195–2201, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2012.12.038>>.
- MARTIN, C. *Aplicações de redes neurais para prognóstico com base em séries temporais*. Dissertação (Mestrado em Engenharia de Produção) — Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2000.
- MATOS, O. C. de. *Econometria Básica: teoria e aplicação*. São Paulo: Atlas, 2000.
- MOREIRA, I. C. Complexidade e caos. In: \_\_\_\_\_. Rio de Janeiro: UFRJ/COPEA, 1999. cap. Fractais, p. 51–82.
- MORET, M. Self-affinity and nonextensivity of sunspots. *Physics Letters A*, v. 378, p. 494 – 496, 2014.
- MORET, M.; ANTONIO, Q. L.; PEREIRA, H. P. P. Classical and fractal analysis of vehicle demand on the ferry-boat system. *Physica A*, v. 391, p. 1667–1661, 2012.
- MORET, M. A.; ZEBENDE, G. F.; NOGUEIRA, J. E.; PEREIRA, M. Fluctuation analysis of stellar x-ray binary systems. *Physical Review E*, v. 68, p. 041104, 2003. Disponível em: <[http://www.aticenter.com.br/mestrdomcti/mcti/uploads/Profiles/pR041104\\_2003.pdf](http://www.aticenter.com.br/mestrdomcti/mcti/uploads/Profiles/pR041104_2003.pdf)>.
- MORETTIN, L. G. *Estatística Básica Volume único Probabilidade e Estatística*. São Paulo: Pearson, 2010.
- MORETTIN, P. *Econometria Financeira*. 1ª. ed. São Paulo: Edgard Blucher, 2008.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. São Paulo: ABE - Projeto Fisher. Editora Edgard Blucher LTDA, 2004.
- NASCIMENTO, A. *Correlação de longo alcance em demandas de passageiros e veículos do sistema Ferry Boat*. Dissertação (Mestrado Interdisciplinar em Modelagem Computacional) — Centro de Pós-Graduação e Pesquisa da Fundação Visconde de Cairu, Salvador, Bahia, 2005.
- NASCIMENTO, A.; ZEBENDE, G. F.; MORET, M. Self- affinity of vehiche demand. *International Journal of Modern*, v. 19, p. 665–669, 2008.
- OLIVEIRA, D.; GOMES, M.; GUIMARÃES, H.; AGUIRRE, L. Síntese de sinais de variabilidade da frequência cardíaca baseada em modelagem não-linear. *Revista Brasileira de Engenharia Biomédica*, v. 22, p. 51–62, 2006. Disponível em: <<http://www.cpdee.ufmg.br/~MACSIN/services/papers/oliveira93.pdf>>.
- OLIVEIRA, E. C.; GOUVEA, D. S. A.; RIBEIRO, A. L. P. Investigação da lei de potência em sinais simulados e sinais rr de indivíduos normais e chagásticos através do expoente de hurst. In: *21º Congresso de engenharia biomédica*. Salvador-BA: [s.n.], 2008. p. 1655–1658.

- PEDRON, I. T. Correlação em séries temporais de temperatura na região de cascavel estado do paran . *Technology Acta Scientiarum.*, v. 29, p. 205–208, 2007. Dispon vel em: <http://periodicos.uem.br/ojs/index.php/ActaSciTechnol/article/view/710>.
- PENG, C. K.; HAVLIN, H. E. S.; STANLEY; GOLDBERGER, A. L. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, v. 5, p. 82–87, 1995.
- PENG, C.-K. et al. Long-range correlations in nucleotide sequences. *Nature*, v. 356, p. 168–171, 1992. Dispon vel em: <http://www.nature.com/nature/journal/v356/n6365/abs/356168a0.html>.
- PENG, C. K. et al. On the mosaic organization of dna sequences. *Physical Review E*, v. 49, p. 1685–1689, 1994. Dispon vel em: <http://polymer.bu.edu/hes/articles/pbhssg94.pdf>.
- PEREIRA, B. B. S ries temporais multivariadas. In: *VI Simp sio de Probabilidade e Estat stica*. Rio de Janeiro: [s.n.], 1984.
- PEREIRA, B. B.; RODRIGUES, C. V. S. Redes neurais em estat stica. In: *XIII Simp sio Brasileiro de Probabilidade e Estat stica*. Caxambu: [s.n.], 1998. p. 1 – 140.
- PODOBNIK, B.; GROSSE, I.; HORVATIC, D.; ILIC, S.; IVANOV, P. Quantifying cross-correlations using local and global detrending approaches. *European Physical Journal B*, v. 71, p. 243–250, 2009.
- PODOBNIK, B.; HORVATIC, D.; PETERSEN, A. M.; STANLEY, H. E. Cross-correlations between volume change and price change. *PNAS*, v. 106, p. 22079–22084, 2009. Dispon vel em: <http://www.pnas.org/content/106/52/22079.full.pdf>.
- PODOBNIK, B. et al. Fractionally integrated process with power-law correlations in variables and magnitudes. *Phys. Rev. E*, v. 72, p. 026121, 2005. Dispon vel em: <http://link.aps.org/doi/10.1103/PhysRevE.72.026121>.
- PODOBNIK, B.; JIANG, Z.-Q.; ZHOU, W.-X.; STANLEY, H. Statistical tests for power-law cross-correlated processes. *Physical Review E*, v. 84, p. 066118, 2011. Dispon vel em: <http://polymer.bu.edu/hes/articles/pjzs11.pdf>.
- PODOBNIK, B.; STANLEY, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters*, v. 100, p. 084102, 2008. Dispon vel em: <http://link.aps.org/doi/10.1103/PhysRevLett.100.084102>.
- REBOREDO, J. C.; RIVERA-CASTRO, M. A.; ZEBENDE, G. F. Oil and us dollar exchange rate dependence: A detrended cross-correlation approach. *Energy Economics*, v. 42, p. 132–139, 2014. Dispon vel em: <http://www.sciencedirect.com/science/article/pii/S0140988313002910>.

- RUAN, Y. P.; ZHOU, W. X. Long-term correlations and multifractal nature in the intertrade durations of a liquid chinese stock and its warrant. *Physica A*, v. 390, p. 1646–1654, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2011.01.001>>.
- SANTOS, A. G. Q.; MACHADO, A.; PESSOTI, G. C.; ALVES, L. A. A.; FREITAS, U. R. P. Índice de performance econômica e social dos municípios baianos: Conceito, revisão metodológica e resultados. *Conjuntura & Planejamento*, v. 177, p. 66–81, 2012.
- SANTOS, W. et al. Correlação de longo alcance no fenômeno da depressão alastrante em ratos nutridos e desnutridos. In: *8º Encontro Regional de Matemática Aplicada e Computacional*. Natal: [s.n.], 2008. Disponível em: <[http://www.dimap.ufrn.br/~sbmac/ermac2008/Anais/Resumos%20Estendidos-/Correla%E7%E3o%20de%20longo%20alcance\\_Santos.pdf](http://www.dimap.ufrn.br/~sbmac/ermac2008/Anais/Resumos%20Estendidos-/Correla%E7%E3o%20de%20longo%20alcance_Santos.pdf)>.
- SCIENCE - Complex Systems. *Science*, 2009. 357-504 p. Disponível em: <<https://www.sciencemag.org/content/325/5939.toc>>.
- SETHNA, J. P.; DAHMEN, K. A.; MYERS, C. R. Crackling noise. *Nature*, v. 410 (6825), p. 242–284, 2001. Disponível em: <<http://www.nature.com/nature/journal/v410/n6825/pdf/410242a0.pdf>>.
- SHADKHOV, S.; JAFARI, G. Multifractal detrended cross-correlation analysis of temporal and spatial seismic data. *The European Physical Journal*, v. 72, p. 679–683, 2009. Disponível em: <<http://link.springer.com/article/10.1140/epj/b/2009-00402-2>>.
- SHAO, Y.-H.; GU, G.-F.; JIANG, Z.-Q.; ZHOU, W.-X.; SORNETTE, D. Comparing the performance of fa, dfa and dma using different synthetic long-range correlated time series. *Nature*, v. 835, p. 1–5, 2012. Disponível em: <<http://arxiv.org/abs/1208.4158>>.
- SILVA, M. D. *Estudo da persistência no mercado de câmbio: análise do expoente de Hurst à mudança de câmbio fixo para câmbio flutuante*. Dissertação (Mestrado em Física) — Universidade Federal da Bahia, Salvador-BA, 2009. Disponível em: <<https://repositorio.ufba.br/ri/bitstream/ri/10683/1/Dissertacao%20Marcus%20da%20Silva.pdf>>.
- SIQUEIRA, E. L. J.; STOSIC, J.; BEJAN, L.; STOSIC, B. Correlations and cross-correlations in the brazilian agrarian commodities and stocks. *Physica A*, v. 389, p. 2739–2743, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2010.01.040>>.
- SOARES, J. S. J.; MACHADO, A. Indicadores de criminalidade e políticas públicas: knowledge discovery in databases como instrumento de apoio à decisão na segurança pública. In: *XI Encontro de Modelagem Computacional*. Volta Redonda: [s.n.], 2008.
- SOARES, J. S. J.; MACHADO, A. Séries temporais de indicadores sociais: uma abordagem detrended fluctuation analysis. In: *XI Escola de Modelos de Regressão*. Recife: [s.n.], 2008.

- SOARES, J. S. J.; MACHADO, A.; QUINTELLA, R. H.; ZEBENDE, G. F. Uma análise estatística dos indicadores de criminalidade de salvador. *Conjuntura & Planejamento*, v. 161, p. 40–49, 2008.
- SOUZA, R. C.; CAMARGO, M. E. *Análise e previsão de séries temporais: Os modelos ARIMA*. 2ª. ed. Rio de Janeiro: Reinaldo Castro & Maria Emilia Camargo, 2004.
- SOUZA, S. R.; TABAK, B. M.; CAJUEIRO, D. O. Investigação da memória de longo prazo na taxa de câmbio no brasil. *Revista Brasileira de Economia*, v. 60, p. 193–209, 2006. Disponível em: <http://repositorio.ucb.br/jspui/handle/10869/307>.
- SPIEGEL, M. *Theory and Problems of Probability and Statistics*. New York: McGraw-Hill, 1992.
- SSP-BA. *Secretaria de Segurança do Estado da Bahia*. 2013. Disponível em: [www.ssp.ba.gov.br](http://www.ssp.ba.gov.br).
- STANLEY, H. et al. Statistical mechanics in biology: How ubiquitous are long-range correlations? *Physica A*, v. 205, p. 214–253, 1994.
- STRATONOVICH, R. L. *Topics in the Theory of Random Noise*. New York: Gordon and Breach, 1981.
- TAKAKURA, I. T. *Variabilidade da frequência cardíaca no domínio do caos preditora de morbimortalidade em pacientes submetidos à cirurgia de revascularização do miocárdio*. Dissertação (Mestrado em Ciências da Saúde) — Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, 2007.
- TALKNER, P.; WEBER, R. Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Phys. Rev. E*, p. 150–160, 2000. Disponível em: [http://www.physik.uni-augsburg.de/theo1/Talkner/Papers/T\\_Weber\\_PRE\\_2000.pdf](http://www.physik.uni-augsburg.de/theo1/Talkner/Papers/T_Weber_PRE_2000.pdf).
- TAQQU, M. S.; TEVEROVSKY, V.; WILLINGER, W. Estimators for long-range dependence: an empirical study. *Fractals, London*, v. 3, p. 785 – 798, 1995. Disponível em: <http://www.worldscientific.com/doi/abs/10.1142/S0218348X95000692>.
- TEXEIRA, A. J. *Detecção de correlação e causalidades em séries temporais não categóricas*. Dissertação (Mestrado em engenharia informática) — Universidade Nova de Lisboa - Faculdade de ciências e tecnologia - Departamento de informática, Lisboa, 2012. Disponível em: <http://www.rcaap.pt/detail.jsp?id=oai:run.unl.pt:10362/7814>.
- TRIOLA, M. *Introdução à Estatística*. 7ª. ed. Rio de Janeiro: LTC - Livros Técnicos e Científicos, 1999.
- TRIOLA, M. *Introdução à Estatística*. 10ª. ed. Rio de Janeiro: LTC - Livros Técnicos e Científicos, 2008.

- URSULEAN, R.; LAZAR, A. Detrended cross-correlation analysis of biometric signals used in a new authentication method. *ELECTRONICS AND ELECTRICAL ENGINEERING*, v. 89, p. 55–58, 2009. Disponível em: [http://www.ee.ktu.lt/journal/2009/1/11\\_ISSN\\_1392-1215\\_Detrended%20Cross-Correlation%20Analysis%20of%20Biometric%20Signals%20used%20in%20a%20new%20Authenticati.pdf](http://www.ee.ktu.lt/journal/2009/1/11_ISSN_1392-1215_Detrended%20Cross-Correlation%20Analysis%20of%20Biometric%20Signals%20used%20in%20a%20new%20Authenticati.pdf).
- VASSOLER, R.; ZEBENDE, G. F. Dcca cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A*, v. 391, p. 2438–2443, 2011. Disponível em: <http://dx.doi.org/10.1016/j.physa.2011.12.015>.
- VEIGA, R. D.; SAFADI, T. *Análise de regressão e séries temporais*. 1ª. ed. Lavras: UFLA - Universidade Federal de Lavras. FAEPE - Fundação de Apoio ao Ensino, Pesquisa e Extensão Lavras - MG, 1999.
- VOSS, R. F. Evolution of long-range fractal correlations and 1/f noise in dna base sequences. *Physical Review E*, v. 68, p. 3805–3808, 1992.
- WALLECZEK, J. *Self-organized Biological Dynamics and Nonlinear Control*. Cambridge University Press, 2000. Disponível em: [www.cambridge.org/9780521624367](http://www.cambridge.org/9780521624367).
- WANG, G.-J.; XIE, C. Cross-correlations between renminbi and four major currencies in the renminbi currency basket. *Physica A*, v. 392, p. 1418–1428, 2013. Disponível em: <http://dx.doi.org/10.1016/j.physa.2012.11.035>.
- WANG, G.-J.; XIE, C.; CHEN, S.; YANG, J.-J.; YANG, M.-Y. Random matrix theory analysis of cross-correlations in the us stock market: Evidence from pearsons correlation coefficient and detrended cross-correlation coefficient. *Physica A*, v. 392, p. 3715–3730, 2013. Disponível em: <http://dx.doi.org/10.1016/j.physa.2013.04.027>.
- WANG, G.-J.; XIE, C.; CHEN, S.; HAN, F. Cross-correlations between energy and emission markets: New evidence from fractal and multifractal analysis. *Mathematical Problems in Engineering*, v. 2014, p. 1–13, 2014.
- WANG, G.-J.; XIE, C.; CHEN, Y.; CHEN, S. Statistical properties of the foreign exchange network at different time scales: Evidence from detrended cross-correlation coefficient and minimum spanning tree. *Entropy*, v. 15, p. 1643–1662, 2013. Disponível em: <http://www.mdpi.com/1099-4300/15/5/1643>.
- WITT, A.; KURTHS, J.; PIKOVSKY, A. Testing stationarity in time series. *Physical Review E*, v. 58, p. 1800–1810, 1998.
- XU, N.; SHANG, P.; KAMAE, S. Modeling traffic flow correlation using dfa and dcca. *Nonlinear Dynamics*, v. 61, p. 207–216, 2010. Disponível em: <http://link.springer.com/article/10.1007%2Fs11071-009-9642-5>.

- YINA, Y.; SHANG, P. Modified multidimensional scaling approach to analyze financial markets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, v. 24, p. 022102, 2014.
- YUAN, N.; FU, Z. Different spatial cross-correlation patterns of temperature records over china: A dcca study on different time scales. *Physica A*, v. 400, p. 71–79, 2014.
- ZEBENDE, G. F. Dcca cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A*, v. 390, n. 2, p. 614–618, 2011. Disponível em: <http://dx.doi.org/10.1016/j.physa.2010.10.022>.
- G. F Zebende, A. P. N CASTRO e A Machado. *Rho dcca mult.* 2013. 011130000445/BR 51 2013 0007585.
- G. F Zebende, A. P. N CASTRO, A Machado e M Fernandes. *Rhodccasort teste estatístico.* 2014. 011140000 84 BR 51 2014 000339 6.
- ZEBENDE, G. F.; MACHADO, A. Cross-correlation between time series of vehicles and passengers. *Physica A*, v. 388, p. 4863–4866, 2009. Disponível em: <http://dx.doi.org/10.1016/j.physa.2009.07.046>.
- ZEBENDE, G. F.; MACHADO, A. Detrended cross-correlation analysis: Measuring cross-correlation between mathematical constants. In: *XXXII Encontro Nacional de Física da Matéria Condensada*. Águas de Lindóia: [s.n.], 2009. p. 1 – 9.
- ZEBENDE, G. F.; OLIVEIRA, P. D.; PENNA, T. Long-range correlations in computer diskettes. *Physical Review E*, v. 57, p. 3311–3314, 1998. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.57.3311>.
- ZEBENDE, G. F. et al. Studying long-range correlations in a liquidvapor-phase transition. *Physical A*, v. 342, p. 322–328, 2004. Disponível em: <http://dx.doi.org/10.1016/j.physa.2004.01.071>.
- ZEBENDE, G. F.; SILVA, M. da; MACHADO, A. Dcca cross-correlation coefficient differentiation: Theoretical and practical approaches. *Physica A*, v. 392, p. 1756–1761, 2013. Disponível em: <http://dx.doi.org/10.1016/j.physa.2013.01.011>.
- ZEBENDE, G. F.; SILVA, P. A. D.; MACHADO, A. Study of cross-correlation in a self-affine time series of taxi accidents. *Physica A*, v. 390, p. 1677–1683, 2011. Disponível em: <http://dx.doi.org/10.1016/j.physa.2010.12.038>.
- ZHAO, X.; SHANG, P.; LIN, A.; CHEN, G. Multifractal fourier detrended cross-correlation analysis of traffic signals. *Physica A*, v. 390, p. 3670–3678, 2011. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378437111004651>.
- ZHOU, W.-X. Multifractal detrended cross-correlation analysis for two nonstationary signals. *Physical Review E*, v. 77, p. 066211, 2008. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.77.066211>.

*AUTOCORRELAÇÃO E CORRELAÇÃO CRUZADA: TEORIAS E APLICAÇÕES*

Aloísio Machado da Silva Filho

Salvador, Setembro de 2014.