



Universidade Federal do Espírito Santo

Centro de Ciências Exatas

Departamento de Estatística

Disciplina de Séries Temporais

MODELOS AUTORREGRESSIVOS DE VALORES  
INTEIROS(INAR)

ESTUDANTE: Christian do Espírito Santo

Vitória - ES  
2023

# Conteúdo

1	INTRODUÇÃO	1
2	MODELOS INAR	2
2.1	Como surgiu . . . . .	2
2.2	Operador Thinning . . . . .	2
2.3	Modelo INAR . . . . .	3
2.3.1	modelo INAR(p) . . . . .	4
2.4	Estimação dos parâmetros . . . . .	4
2.4.1	Estimador de Yule-Walker . . . . .	4
2.4.2	Estimador de Mínimos quadrados condicional . . . . .	5
2.4.3	Estimador de máxima verossimilhança condicional . . . . .	5
2.5	INAR vs. AR . . . . .	6
3	RESULTADOS E DISCUSSÕES	8
3.1	Qualidade do ar em Cariacica em 2007 . . . . .	8
4	CONSIDERAÇÕES FINAIS	10
	BIBLIOGRAFIA	11

# 1 Introdução

Os modelo INAR compõem uma classe de modelos estatísticos para séries temporais de natureza discreta de números inteiros não negativos com aplicações diversas para dados que possuem esta natureza, baseados nos modelos autorregressivos AR foram adaptados para aplicações apenas usando distribuições de probabilidade discretas como Poisson, geométrica, Binomial entre outras.

O presente trabalho contará com informações mais detalhadas e aprofundadas para descrição, estimação, previsão, aplicação e análise do mesmo mostrando seus pontos úteis como um modelo alternativo para casos no qual se aplica para estudos estatísticos.

## 2 Modelos INAR

### 2.1 COMO SURTIU

A primeira apresentação de um processo autorregressivo de valores inteiros denominado por INAR(1) foi feita em 1987 por Al-Osh, modelo usado para séries de contagem, a partir da teoria dos modelos ARMA, que sob hipótese de normalidade, se adequam bem na modelagem de sequências dependentes estacionárias, acrescido dos modelos de JACOBS e LEWIS (1977), LAWRENCE (1980), LAWRENCE e LEWIS (1980) e GAVIER e LEWIS (1980) de séries temporais com distribuições marginais positivas, como exponencial e gama, também da ideia de processos discretos de médias móveis autoregressiva (DAM) desenvolvidos por JACOBS e LEWIS (1978a, banda 1983).

Em 1991 Du e Li generalizaram o processo para ordem  $p$ , juntamente com Latour(1998), a condição de estacionariedade foi aplicada ao processo INAR( $p$ ), e Latour(1997) também propôs o INAR multivariado. Usando as propriedades do operador thinning generalizado por Gauthier e Latour(1994), Silva e Oliveira (2005) obtiveram as expressões dos momentos e culminantes do processo. Desde então diversos autores passaram a apresentar teorias e procedimentos do processo de contagem, principalmente com o objetivo de criar modelos onde os erros seguem a distribuição poisson.

### 2.2 OPERADOR THINNING

Os modelos mais comumente usados no contexto de séries temporais, no geral não se adequam a muitas das séries de valores inteiros não negativos, incluindo particularmente as séries de contagem, pois o produto de uma variável aleatória(valor real) por um valor inteiro gera uma variável aleatória(valor real) não inteiro, essa necessidade de adaptação levou a apresentação de novos modelos baseados no modelo ARMA com a alteração de que a multiplicação citada foi substituída por um multiplicador inteiro, designado como operador thinning ou parâmetro thinning binomial, denotado por  $\alpha$  em

$$\alpha \circ X = \sum_{j=1}^X Y_j$$

cujo  $Y_j$  é i.i.d e independente de  $X$ (v.a inteira não negativa), seguindo uma distribuição Bernoulli( $\alpha$ ). Assim a multiplicação foi substituída pelo produto de um parâmetro thinning binomial por um escalar, dando origem a classe de modelos INARMA, que são adaptações dos modelos ARMA usando o operador thinning.

## 2.3 MODELO INAR

Um processo INAR(1) é definido com um processo estocástico discreto com valor inteiro e não negativo que satisfaz a seguinte equação

$$y_t = p \circ y_{t-1} + \epsilon_t,$$

em que  $p \in [0, 1]$ ,  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  é uma sequência de v.a i.i.d de valores inteiros não negativos com média  $\mu$  e variância  $\sigma^2$ , se  $p = 1$   $y_t$  é um passeio aleatório, se  $p = 0$  o processo será a sequência  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ .

$y_t$  é a soma de duas variáveis aleatórias com distribuições discretas, sendo  $W_t = p \circ y_{t-1}$ , com distribuição binomial, e  $\epsilon_t$  seguindo qualquer distribuição de probabilidade adequada, como Poisson, Binomial Negativa, Geométrica entre outras.  $W_t$  e  $\epsilon_t$  são independentes, assim temos:

$$W_t \mid y_{t-1} \sim \text{Binomial}(p, y_{t-1})$$

O modelo INAR(1) possui um alto fator de interpretabilidade, podendo ser aplicado naturalmente a qualquer fator de contagem cotidiano que se encaixe nas especificações, como por exemplo para o número de alunos em um curso de Estatística, o modelo pode ser descrito como:

$$\underbrace{y_t}_{\text{N alunos no tempo } t} = \underbrace{p \circ y_{t-1}}_{\text{sobreviventes no tempo } t-1} + \underbrace{\epsilon_t}_{\text{Novos alunos}}$$

No caso apresentado  $y_t$  descreve a quantidade total de alunos no curso de estatística no período  $t$  e  $y_{t-1} - p \circ y_{t-1}$  descreve a quantidades de alunos que foram perdidos no ultimo período, por fim  $\epsilon_t$  nos mostra os alunos ingressantes no curso após o período  $y_{t-1}$ .

A distribuição marginal do processo INAR(1) pode ser definida em função da sequência de avanço  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  por

$$y_t \stackrel{d}{=} p^t \mathbb{E}(y_0) + \mu \sum_{j=0}^{\infty} p^j \circ \epsilon_{t-j}.$$

Com média e variância incondicionais

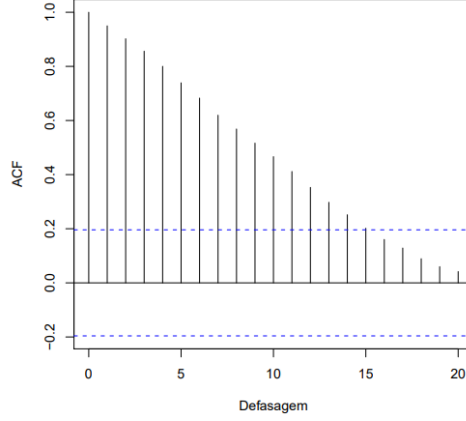
$$\mathbb{E}[y_t] = p \circ \mathbb{E}[y_{t-1}] + \mu = p^t \mathbb{E}[y_0] + \mu \sum_{j=0}^{t-1} p^j$$

$$\text{Var}[y_t] = p^2 \text{Var}[y_{t-1}] + (1-p)p \mathbb{E}[y_{t-1}] + \sigma^2 = p^{2t} \text{Var}[Z_0] + (1-p) \sum_{j=1}^t p^{2j-1} \mathbb{E}[y_{t-j}] + \sigma^2 \sum_{j=1}^t p^{2(j-1)}$$

Os processos INAR(1) mais apresentados na literaturas, eram construídos com a sequência de inovação  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  seguindo distribuição Poisson com parâmetro  $\lambda$ , denotado por  $\{\epsilon_t\}_{t \in \mathbb{Z}} \sim Po(\lambda)$ , nessas condições as métricas apresentadas acima podem ser concluídas como

$$\mathbb{E}(y_t) = \text{Var}(y_t) = \frac{\lambda}{1-p}$$

A função de autocorrelação  $\rho(k)$  do processo INAR(1) possui um comportamento de decaimento em função de  $k$  similar a um processo AR(1), com a diferença de que no processo INAR(1) são apresentados exclusivamente valores positivos. A figura 1 simula um processo com  $p=0.1$ ,  $\lambda=1$  e tamanho amostral 100.



**Figura 1:** Função de autocorrelação amostral do processo INAR(1)

### 2.3.1 modelo INAR( $p$ )

A generalização do modelo INAR de ordem superior possuem formulação similar aos modelos AR( $p$ ), aplicando a alteração dos parâmetro thinning apresentados anteriormente:

$$y_t = p_1 \circ y_{t-1} + p_2 \circ y_{t-2} + \dots + p_p \circ y_{t-p} + \epsilon_t$$

Onde  $p \in [0,1]$ ;  $j=1,2,\dots,p$  e  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  é uma sequência de v.a i.i.d de valores inteiros não negativos com média  $\mu$  e variância  $\sigma^2$  finitas.

## 2.4 ESTIMAÇÃO DOS PARÂMETROS

A estimação dos parâmetros do modelo INAR(1) possui mais complexidade que do modelo AR(1), já que a distribuição de  $X_t|X_{t-1}$  é composta do conjunto que compõe uma distribuição de  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  e uma distribuição binomial com parâmetros  $y_{t-1}$  e  $p$ . A estimação do vetor de parâmetros  $\theta = (p, \lambda)^T$  pode ser feita através de 3 métodos estatísticos.

### 2.4.1 Estimador de Yule-Walker

A partir de uma amostra de tamanho  $n$  do processo estacionário  $\{y_t\}_{t \in \mathbb{Z}}$ , onde o parâmetro  $p$  é estimado seguindo a função de autocorrelação e autocovariância amostral usando sua média aritmética  $\bar{y}$ .

$$\hat{p} = \hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

e a estimação do parâmetro  $\lambda$  é feita com base no primeiro momento do processo  $\mathbf{E}(y_t) = \frac{\lambda}{(1-p)}$  precedendo

$$\hat{\lambda} = (1 - \hat{p}) \frac{1}{n} \sum_{t=1}^n y_t$$

#### 2.4.2 Estimador de Mínimos quadrados condicional

Nesse estimador, sendo  $\{y_t\}_{t \in \mathbb{Z}}$  processo INAR(1) com  $\{\epsilon_t\}_{t \in \mathbb{Z}} \sim \text{Po}(\lambda)$  temos  $\mathbb{E}[y_t | y_{t-1}] = py_{t-1} + \lambda = g(\theta, y_{t-1})$ , onde  $\theta$  é o vetor de parâmetros a serem estimados.

com uma amostra de tamanho  $n$ , o estimador de MQC visa minimizar a soma de quadrados

$$Q_n(\theta) = \sum_{t=2}^n [y_t - g(\theta, y_{t-1})]^2$$

, que derivando e igualando a zero obtemos os estimadores

$$\hat{p} = \frac{\sum_{t=2}^n y_t y_{t-1} - \frac{1}{n-1} \sum_{t=2}^n y_{t-1}}{\sum_{t=2}^n y_{t-1}^2 - \frac{1}{n-1} (\sum_{t=2}^n y_{t-1})^2}$$

$$\hat{\lambda} = \frac{1}{n-1} \left( \sum_{t=2}^n y_t - \hat{p} \sum_{t=2}^n y_{t-1} \right)$$

#### 2.4.3 Estimador de máxima verossimilhança condicional

Dado  $\{y_t\}_{t \in \mathbb{Z}}$  o processo INAR(1) com  $\{\epsilon_t\}_{t \in \mathbb{Z}} \sim \text{Po}(\lambda)$  temos a distribuição condicional  $f(y_t | y_{t-1})$  que é uma convolução de uma distribuição binomial do operador thinning  $p \circ y_{t-1}$  e uma Poisson de  $\epsilon_t$  que pode ser descrita como

$$f(y_t | y_{t-1}) = e^{-\lambda} \sum_{i=0}^{\min(y_t, y_{t-1})} \frac{\lambda^{(y_t)-i}}{(y_t - i)!} \binom{y_{t-1}}{i} p^i (1-p)^{(y_{t-1})-i}$$

com uma amostra de tamanho  $n+1$  a função de verossimilhança incondicional é dada por

$$L(p, \lambda; y) = \frac{[\lambda/(1-p)]^{y_0}}{y_0!} \exp\left(\frac{\lambda}{1-p}\right) \prod_{t=1}^n f(y_t | y_{t-1})$$

na qual a verossimilhança condicional se aplica no produtório, com isso obtemos o log-verossimilhança e a função escore condicional

$$\frac{\partial l(p, \lambda; Y | y_0)}{\partial p} = \sum_{t=1}^n \frac{y_{t-1}}{1-p} \frac{[f(y_{t-1} - 1 | y_{t-1} - 1) - f(y_t | y_{t-1})]}{f(y_t | y_{t-1})}$$

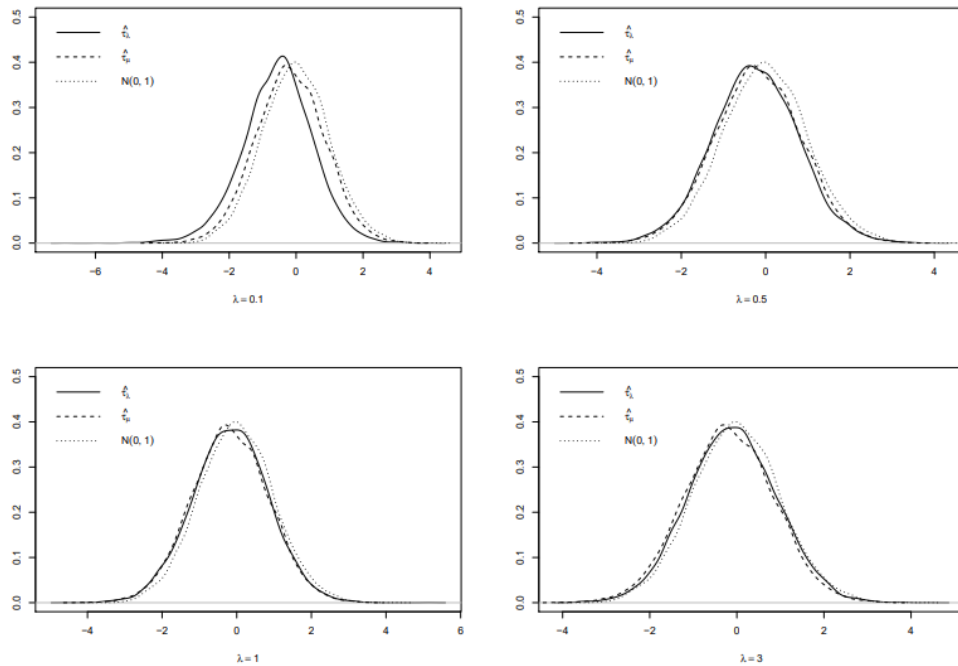
$$\frac{\partial l(p, \lambda; y | y_0)}{\partial \lambda} = \sum_{t=1}^n \frac{[f(y_{t-1} - 1 | y_{t-1} - 1) - f(y_t | y_{t-1})]}{f(y_t | y_{t-1})}$$

e a partir dela conigiremos obter os estimadores ao igualar a zero encontrando a função que maximiza o log-verossimilhança condicional do processo  $y_t$ .

## 2.5 INAR VS. AR

Os processos AR e INAR possuem diversas semelhanças, ambos sendo baseados a uma estrutura autoregressiva, o que desencadeia a similaridade em sua função de autocorrelação mostrado na figura 1, também possuem muitas diferenças sendo a principal a qual o modelo AR conta com uma dependência contínua com aplicação abundante em diversos tipos de séries contínuas como temperatura, financeira entre outras, já o modelo INAR com uma apresentação inteira não negativa, por conta disso os modelos INAR possuem aplicação mais funcional em séries de contagens ou eventos discretos, essa disposição se mostra claro no parâmetro thinning com distribuição binominal e no fator de erro, já que o mesmo assume distribuições discretas (geralmente distribuição de Poisson( $\lambda$ )) enquanto no autorregressivos assumem distribuições contínuas (geralmente um ruído branco avaliado como normal( $0, \sigma^2$ )).

O pesquisador Marcelo Bouruignon Pereira em sua tese defendida para o mestrado aplicou um estudo que a partir de um teorema que afirma que processos autorregressivos, sob hipótese de estacionariedade, tem estatísticas de teste com a mesma distribuição assintótica para AR com o fator de erro sendo um ruído branco com distribuição normal( $0, \lambda$ ) e INAR com  $\{\epsilon_t\}_{t \in \mathbb{Z}} \sim \text{Po}(\lambda)$ , uma vez que quando  $\lambda \rightarrow \infty$  a distribuição Poisson converge para Normal. O estudo consiste em usar o 1000 replicações de Monte Carlo nos processos citados acima para avaliar empiricamente a distribuição das estatísticas de teste, sendo  $\hat{\tau}_\mu = \frac{\hat{\phi}_\mu - 1}{s(\hat{\phi}_\mu)}$  a estatística de teste para o processo AR, onde  $\hat{\phi}_\mu$  é o estimador do parâmetro do modelo AR e  $s(\hat{\phi}_\mu)$  a variância do estimador, e  $\hat{\tau}_\lambda = \frac{\hat{p} - 1}{s(\hat{p})}$  a estatística de teste para o processo INAR, onde  $\hat{p}$  é o estimador do parâmetro do modelo e  $s(\hat{p})$  a variância do estimador.



**Figura 2:** Comparação das distribuições com  $n=100$  e  $\lambda = [0.1; 0.5; 1; 3]$



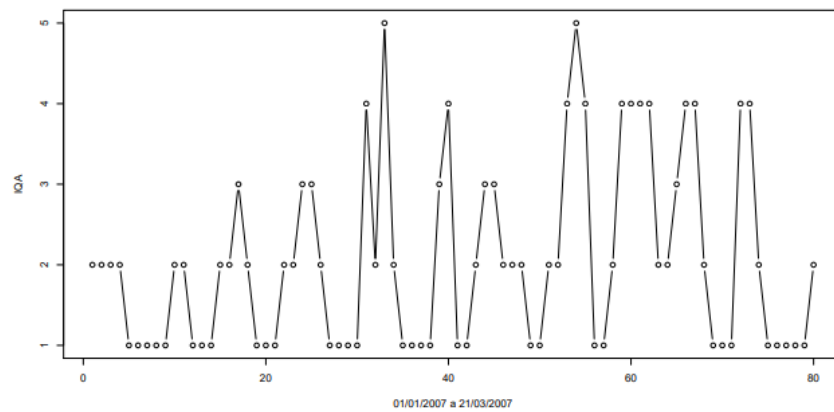
Podemos ver que com a aplicação as distribuições vão se ajustando a normal padrão no que  $\lambda$  aumenta, com a estatística do AR se aproximando logo no segundo gráfico e a estatística do processo INAR se ajustando aos valores mais altos nos levando a observar a convergência citada.

### 3 Resultados e discussões

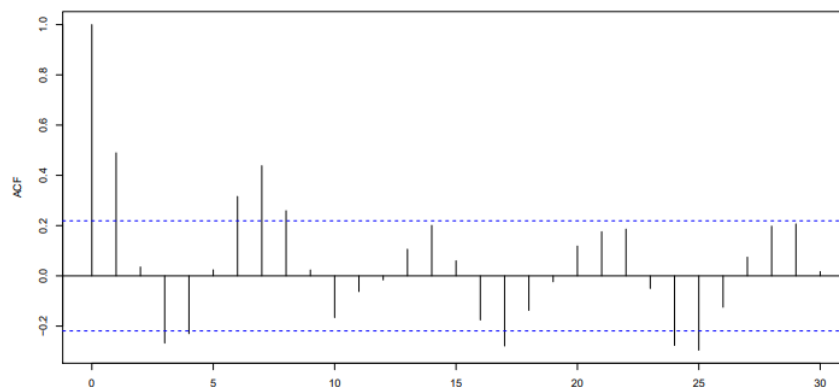
#### 3.1 QUALIDADE DO AR EM CARIACICA EM 2007

Ainda seguindo a tese de mestrado do pesquisador Marcelo Bouruignon Pereira, ele apresenta um estudo no qual ele usa dados disponibilizados pelo instituto estadual de meio ambiente e recursos hídricos do Espírito Santo (IEMA) medidos entre 01/01/2007 e 21/03/2007 avaliando a quantidade de poluente no ar, calculando o parâmetro de índice de qualidade do ar(IQA) que transforma diversas medições em um valor inteiro não negativo podendo comparar com os padrões legais de qualidade do ar.

A série conta com 80 observações no qual 41% correspondem ao valor 1 e com limite superior igual 5, características que dificultariam o uso de qualquer modelo baseado na distribuição gaussiana pela magnitude pequena e não assumir valores negativos.



**Figura 3:** IQM de Cariacica



**Figura 4:** IQM de Cariacica

As figuras 2 e 3 sugerem estacionaridade e indícios de sazonalidade com período 7.

**Tabela 1:** *Estimativas dos parâmetros do modelo.*

Método	$\hat{P}$	$\hat{\lambda}$
YW	0.4526	1.1507
MQC	0.4367	1.1849
MVC	0.6416	0.7336

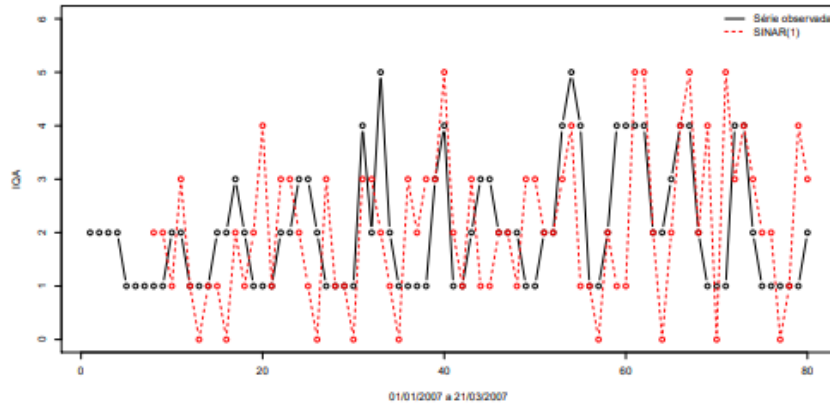
Ao avaliar as variáveis e calcular as estimativas do modelo, o autor chega a conclusão de que a melhor estimação foi feita pelo método de máxima verossimilhança condicional, para um modelo SINAR(1)<sub>7</sub> com equação

$$y_t = 0.6416 \circ y_{t-7} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{Po}(0.7336)$$

e com base nela calculou as previsões 3 passos a frente e seus intervalos estimadores pontuais

**Tabela 2:** *previsão 3 passos a frente*

$h$	$y_{n+h}$	$\hat{y}_{n+h}$	$\hat{e}_{n+h}$
1	2	2.0168	-0.0168
3	2	1.3752	0.6248
5	3	1.3752	1.6248



**Figura 5:** *dados reais x modelo ajustado*

Por fim o autor mostra o ajuste do modelo SINAR(1)<sub>7</sub> comparado aos dados originais.

## 4 Considerações finais

Ao longo do trabalho foi explorado a classe de modelos INAR destacando suas características fundamentais com sua capacidade de lidar com dados discretos de forma eficaz para modelagem de fenômenos com séries temporais de contagem capturando nuances e padrões que poderiam ser erroneamente avaliados por modelos mais gerais apresentando uma nova ferramenta que pode ser de grande utilidade para projetos e estudos futuros.

- (1987) — Al-Osh, M. A. & Alzaid, A. A. (1987). First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*.
- (1997) — DA ROSA, Joel Mauricio Corrêa. Modelos INAR (1) e estruturais para series temporais de contagens: um estudo comparativo, utilizando as distribuições poisson e geometrica. 1997. Tese de Doutorado. [sn].
- (2003) — DA SILVA, Isabel; SILVA, Maria Eduarda. Seleção de ordem em modelos INAR. *Literacia e Estatística*, 2003.
- (2011) — Bourguignon Pereira, Marcelo; Leite Pinto Vasconcellos, Klaus. Modelos inar sazonais e de raízes unitárias. 2011. Dissertação (Mestrado). Programa de Pós-Graduação em Estatística, Universidade Federal de Pernambuco, Recife, 2011.
- (2013) — LIMA, Tiago de Almeida Cerqueira. Modelos INAR e RCINAR, estimação e aplicação. 2013. Tese de Doutorado. Universidade de São Paulo.