**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**

ANGELA ACHCAR

Padrões temporais de casos de tuberculose
em municípios do estado de São Paulo,
no período de 2009 a 2013

ANGELA ACHCAR

**Padrões temporais de casos de tuberculose
em municípios do estado de São Paulo,
no período de 2009 a 2013**

Dissertação apresentada à Faculdade de Medicina de
Ribeirão Preto da Universidade de São Paulo, como parte
dos requisitos para obtenção do título de
Mestre em Saúde na Comunidade.

Área de Concentração: Saúde na Comunidade

Orientador: Prof. Dr. Antonio Ruffino-Netto

Ribeirão Preto - SP
2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha Catalográfica

Achcar, Angela.

Padrões temporais de casos de tuberculose em municípios do estado de São Paulo, no período de 2009 a 2013.

Dissertação (Mestrado)- Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto/SP, 2016.

Orientador: Prof. Dr. Antonio Ruffino-Netto

1. Bioestatística 2. Padrões temporais 3. Tuberculose

Nome: ACHCAR, Angela

Título: Padrões temporais de casos de tuberculose em municípios do estado de São Paulo, no período de 2009 a 2013

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Saúde na Comunidade.

Aprovado em:

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

DEDICATÓRIA

À minha família.

AGRADECIMENTOS

- Aos meus pais Alina Baptista Pizzato e Jorge Alberto Achcar e ao meu irmão Fabio Achcar que se fazem presentes na minha vida e sempre orientam, apoiam e incentivam minhas decisões.
- Ao Prof. Dr. Antônio Ruffino-Netto, pela atenção e orientação durante o processo de construção desse trabalho.
- À minha amiga, Prof^a. Dr^a. Elisangela Ap. da Silva Lizzi, por ter me incentivado a fazer o mestrado no programa Saúde na Comunidade.
- Às minhas colegas de mestrado.
- À CAPES pelo apoio financeiro e concessão da bolsa de mestrado para realização dessa pesquisa.
- Ao Prof. Dr. Edson Zangiacomi Martinez, por me fazer enxergar a estatística de um novo e melhor modo.
- Aos que contribuíram com trabalho na banca de qualificação: Prof^a. Dr^a. Claudia Benedita dos Santos, Prof. Dr. Jair L. F. Santos e Prof. Dr. Roberto Molina de Souza.

RESUMO

ACHCAR, A. **Padrões temporais de casos de tuberculose em municípios do estado de São Paulo, no período de 2009 a 2013.** 2016. 106 p. Tese (Mestrado) - Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto/SP, 2016.

A tuberculose continua sendo um problema de saúde pública, apesar dos avanços científicos, da evolução do tratamento e da descoberta de vacina. O presente estudo tem como intuito descrever a distribuição temporal da doença para os casos novos notificados nos municípios com mais de 200.000 habitantes no Estado de São Paulo (centros regionais), no período de janeiro de 2009 a dezembro de 2013, além de identificar algumas covariáveis socioeconômicas, demográficas e de saúde, que podem estar associadas com a doença e utilizar estas variáveis em diferentes modelos estatísticos para serem usados na previsão mensal de novos casos reportados no SINAN (Sistema de Informação de Agravos de Notificação). Alguns modelos estatísticos especiais foram utilizados na análise dos dados: modelos de regressão linear múltipla para os dados agrupados de contagem para todas as cidades ou cidades individuais transformados para uma escala logarítmica ou para as taxas de novos casos de tuberculose para todas as cidades nos 60 meses de acompanhamento e modelos de regressão de Poisson para dados de contagem. As inferências de interesse foram obtidas sob uma abordagem clássica e bayesiana e sob diferentes estruturas para as covariáveis e efeitos aleatórios que capturam a possível dependência das contagens de tuberculose nos 60 meses de acompanhamento. As previsões obtidas pelos modelos são comparadas com os dados reais para a procura de melhores modelos para serem ajustados pelos dados.

Palavras-chave: Estudo ecológico. Análise temporal. Modelos estatísticos. Tuberculose.

ABSTRACT

ACHCAR, A. **Padrões temporais de casos de tuberculose em municípios do estado de São Paulo, no período de 2009 a 2013.** 2016. 106 p. Tese (Mestrado) - Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto/SP, 2016.

Tuberculosis remains a public health problem, despite scientific advances, the development of treatment and vaccine discovery. This study is intended to describe the temporal distribution of the new cases of the disease reported in cities with more than 200,000 inhabitants in the state of São Paulo (regional centers), from January 2009 to December 2013, and to identify some socio-economic, demographic and health covariates that may be associated with the disease and use these variables in different statistical models for use in the monthly forecast of new cases reported in the SINAN (Diseases Information notification System). Some special statistical models were used in the data analysis: multiple linear regression models for the pooled data count for all cities or individual cities transformed to a logarithmic scale or the rates of new cases of tuberculosis for all cities in the 60 months monitoring time and Poisson regression models for count data. Inferences of interest were obtained under a classical and a Bayesian approach and under different structures for covariates and random effects that capture the possible dependence of tuberculosis counts in the 60 months follow-up. The forecasts obtained by the models are compared with the actual data for the search of better models to be fitted by the data.

Keywords: Ecological study. Temporal analysis. Statistical models. Tuberculosis.

LISTA DE TABELAS

Tabela 1 - Taxas de Tb ordenadas para os 40 municípios	31
Tabela 2 - Números de meses fora de controle para os 40 municípios	41
Tabela 3 - Estimadores de mínimos quadrados, estatística T de Student e valor-p	43
Tabela 4 - Sumários a posteriori (modelo 1)	46
Tabela 5 - Sumários a posteriori (modelo 2)	47
Tabela 6 - Sumários a posteriori (modelo 3)	49
Tabela 7 - Sumários a posteriori (modelo 5).....	51
Tabela 8 - Sumários a posteriori (modelo 6)	54

LISTA DE FIGURAS

Figura 1 - Exemplo de rede de causalidade para a tuberculose	16
Figura 2 - Exemplo de carta controle.	22
Figura 3 - Distribuição normal.	23
Figura 4 - Análise geral das notificações de tuberculose para os dados dos 40 municípios agrupados.	30
Figura 5 - Carta-controlre para a taxa de Tb para os dados dos 40 municípios agrupados	30
Figura 6 - Séries temporais mensais das taxas médias de incidência de novos casos de Tb para cada um dos quatro grupos de cidades	33
Figura 7 - Mapa do Estado de São Paulo com algumas das principais cidades.....	34
Figura 8 - Gráficos dos resíduos - modelo de regressão dado por (8)	43
Figura 9 - Gráficos dos valores observados e médias ajustadas - modelo de regressão (8)	44
Figura 10 - Gráficos dos valores observados e médias ajustadas - “modelo 1”	46
Figura 11 - Gráficos dos valores observados e médias ajustadas - modelo de regressão, modelo 1 e modelo 2	48
Figura 12 - Gráficos dos valores observados e valores estimados pelo “modelo 3”.....	49
Figura 13 - a) Gráfico de linhas para a Taxa notificadas segundo o tempo. b-f) Gráficos de dispersão para as covariáveis segundo a taxa notificada.....	52
Figura 14 - Gráficos dos valores observados e médias ajustadas - “modelo de regressão, modelo 4”	52
Figura 15 - Gráficos dos valores observados e médias ajustadas- “modelo de regressão, modelo 5”	54

SUMÁRIO

1	INTRODUÇÃO	13
1.1	ALGUMAS CONSIDERAÇÕES SOBRE O DIAGNÓSTICO	14
1.2	ALGUMAS CONSIDERAÇÕES SOBRE O TRATAMENTO	15
1.3	ALGUMAS CONSIDERAÇÕES SOBRE VACINAS	15
1.4	CLASSIFICAÇÃO DAS PESSOAS QUANTO À DOENÇA	16
1.5	ALGUNS FATORES QUE PODEM INFLUENCIAR A INCIDÊNCIA DE TB	16
1.6	TUBERCULOSE NO BRASIL	17
2	OBJETIVOS	18
2.1	OBJETIVO GERAL	18
2.2	OBJETIVOS ESPECÍFICOS	18
3	MATERIAL E MÉTODOS	19
3.1	FONTE DOS DADOS	19
4	ASPECTOS ÉTICOS	20
5	CONSIDERAÇÕES GERAIS	21
6	ANÁLISE DOS DADOS	21
7	MÉTODOS	22
7.1	CARTA CONTROLE	22
7.2	MODELOS DE REGRESSÃO LINEAR	23
7.3	INTRODUÇÃO AOS MÉTODOS BAYESIANOS	24
7.3.1	Fórmula de Bayes	25
7.3.2	Distribuição a Priori	26
7.3.3	Intervalos de credibilidade	26
7.3.4	Métodos de simulação para obtenção dos sumários a posteriori	27
7.3.5	Método de Monte Carlo em cadeias de Markov	28
7.3.6	Amostrador de Gibbs	28
7.3.7	Algoritmo de Metropolis-Hastings	29
8	RESULTADOS	30
8.1	ANÁLISE PRELIMINAR DOS DADOS	30
8.2	USO DE MODELOS DE REGRESSÃO LINEAR COM ERROS NORMAIS CONSIDERANDO AS COVARIÁVEIS MÊS E ANO	42
8.2.1	Modelo de regressão sob o enfoque clássico para as contagens de Tb	42
8.2.2	Modelo de regressão sob o enfoque bayesiano para as contagens de Tb	44

8.2.3	Modelo de regressão linear sob o enfoque bayesiano para as taxas de Tb	48
8.2.4	Modelo na taxa notificada de casos de tuberculose na presença de efeitos aleatórios	50
8.2.5	Modelo de Contagem notificada de casos de tuberculose na presença de efeitos aleatórios	53
9	CONCLUSÕES E PERSPECTIVAS FUTURAS.	54
	REFERÊNCIAS	58
	APÊNDICE A - GRÁFICOS DE CONTROLE PARA CADA MUNICÍPIO	63
	APÊNDICE B - ALGUMAS INFORMAÇÕES DE CADA MUNICÍPIO.	73
	APÊNDICE C - ALGUNS PROGRAMAS OPENBUGS	80
	MODELO 1 - USO DO OPENBUGS	80
	MODELO 2 – EFEITO DEFASADO (CONTAGEM MÊS ANTERIOR)	80
	MODELO 3 – EFEITOS DEFASADOS (CONTAGEM SEIS MESES ANTERIORES)	81
	APÊNDICE D	84
	AUTOCORRELAÇÃO	84
	O CORRELOGRAMA	85
	ANEXO - DECLARAÇÃO DO COMITÊ DE ÉTICA	106

1 INTRODUÇÃO

Evidências históricas da tuberculose (Tb) datam de milhares de anos: já as colunas vertebrais de múmias egípcias documentam a sua existência, e a doença também era comum tanto na Grécia Antiga como no Império Romano, segundo o National Institute of Allergy and Infectious Diseases (NIAID).

A tuberculose (Tb) é uma doença infecciosa, causada pelo *Mycobacterium tuberculosis*. Apresenta-se principalmente como Tb pulmonar, mas pode afetar outros órgãos (extrapulmonar Tb). A tuberculose normalmente progride lentamente desde a fase latente (infecção sem doença ativa, LTBI-Latent Tuberculosis Infection) para Tb ativa, exceto em pacientes infectados pelo HIV onde a progressão pode ser rápida e fatal. Apenas uma pequena proporção (5% a 10%) de pessoas infectadas com *M. tuberculosis* irão evoluir para Tb ativa (WORLD HEALTH ORGANIZATION, 2012).

Apesar dos avanços científicos, a evolução do tratamento e a descoberta de vacina tenham contribuído para o controle da Tb, ela nunca deixou de existir e representar um perigo, sendo ainda hoje uma das doenças infecciosas de grande letalidade no mundo.

De acordo com as análises das tendências globais da saúde desde 2000 publicadas em 2015 pela Organização Mundial de Saúde (WORLD HEALTH ORGANIZATION, 2015),

“As doenças infecciosas como a tuberculose ainda são um grande problema mundial de saúde pública.”

As tendências globais desde então mostram que a incidência de tuberculose (Tb), taxas de prevalência e mortalidade estão gradualmente diminuindo em todo o mundo. No entanto, o aumento de casos de Tb resistentes à drogas é preocupante. A tuberculose continua sendo uma doença da pobreza com uma alta carga de doenças nos países de renda baixa e média (LMIC) e em países com alta incidência de HIV. Os 22 países com maiores incidências são responsáveis por 82% do número de casos de Tb.

Em 2013, foi estimado que 9 milhões de pessoas desenvolveram Tb e 1,5 milhão morreram da doença, dos quais 360.000 eram portadores do vírus HIV. A Tb vem declinando de forma lenta, a cada ano e estima-se que 37 milhões de vidas foram salvas entre 2000 e 2013 por causa do diagnóstico e tratamentos efetivos. Globalmente, a taxa de sucesso do tratamento chegou

a 85% em 2009. No entanto dado que a maioria das mortes por Tb são evitáveis, o número de mortos pela doença ainda é inaceitavelmente elevado e os esforços para combatê-la devem ser acelerados.

Em geral, as ferramentas para combater Tb existentes não são suficientes para erradicar a tuberculose. Um número substancial de novas ferramentas de diagnóstico, foram endossadas pela Organização Mundial da Saúde (OMS), mas estes testes são caros e portanto não são adequados para uso nos países mais pobres. Os regimes atuais de tratamento são longos, têm uma carga de comprimidos alto, requerem um compromisso do paciente, e podem ter efeitos colaterais graves. Não há atualmente apenas uma vacina disponível para a prevenção da Tb, no entanto, os seus efeitos protetores são altamente variáveis. Novas ferramentas de Tb são necessárias e devem ser adequadas para todas as formas de Tb em pessoas de todas as idades, incluindo toda a população e as crianças HIV-positivo. Essas ferramentas devem levar a resultados rápidos e serem adequadas para utilização nos países mais pobres.

Um progresso considerável tem ocorrido desde 2004, no desenvolvimento de novos tratamentos, diagnósticos e vacinas, muitas vezes com o apoio da Comissão Europeia.

1.1 ALGUMAS CONSIDERAÇÕES SOBRE O DIAGNÓSTICO

Um diagnóstico rápido e preciso de tuberculose (Tb) é fundamental para o início precoce do tratamento, e em última análise, o controle da doença. O teste diagnóstico mais comumente utilizado para Tb pulmonar, baciloscopia, tem mais de 100 anos e apresenta baixa sensibilidade, embora muito específico. O atual padrão ouro é a cultura microbiológica. No entanto, o mesmo envolve tempo para apresentar resultados e mão de obra intensiva e exige uma adequada infraestrutura laboratorial que não está prontamente disponível em muitos dos países mais pobres com uma elevada carga da doença. Os resultados dos testes fármaco-sensibilidade pode levar semanas. Idealmente um teste de diagnóstico para a Tb seria de baixo custo, simples, rápido, adequado para uso nas clínicas de saúde (por exemplo, centros de saúde), e daria resultados da sensibilidade aos medicamentos (WORLD HEALTH ORGANIZATION, 2012; UNITAID, 2012). Novos diagnósticos tem sido pesquisados por muitos pesquisadores em todo o mundo (ver por exemplo, WEYER; CARAI; NUNN, 2011).

1.2 ALGUMAS CONSIDERAÇÕES SOBRE O TRATAMENTO

A OMS recomenda o tratamento supervisionado, (e.g., tratamento diretamente observado - DOT), para garantir maior adesão ao tratamento. O regime de tratamento atual de novos casos de tuberculose pulmonar consiste em um regime de tratamento de seis meses, sendo os dois primeiros meses, com quatro drogas diferentes (isoniazida, rifampicina, pirazinamida e etambutol), com a dose diária ou três vezes por semana sob DOT. Nos dois meses seguinte, com apenas dois medicamentos. Esses medicamentos diferentes podem ser combinados numa única combinação de dose fixa (FDC) para reduzir a carga de comprimidos embora contenham os mesmos princípios ativos. A taxa de sucesso do tratamento para pacientes sensíveis à droga com este regime é alta. Os pacientes com Tb tratados anteriormente tem uma grande chance de uma estirpe resistente aos medicamentos de Tb e exigem uma abordagem diferente, dependendo da disponibilidade de fármaco-teste de sensibilidade (DST). Onde os DST são rotineiramente disponíveis, um novo regime de tratamento deve ser iniciado de acordo com os resultados. O tratamento pode durar de 18 meses a um máximo de 30 meses, e contém pelo menos quatro drogas diferentes, incluindo a utilização de pelo menos um medicamento injetável mas tem uma taxa de sucesso relativamente baixa (LEMOS, 2008). Novos tratamentos tem sido pesquisados (ver por exemplo, ZHANG et al., 2011; LIENHARDT et al., 2011; KANEKO; COOPER; MDLULI, 2011; MA et al., 2010; KOUL et al., 2011; MCNERNEY et al., 2012). Além disso, várias indústrias farmacêuticas tem muitos estudos em andamento com novas drogas (NOVARTIS, 2013, SANDOZ TB, 2013; TAG & I-BASE, 2014; SANOFIAVENTIS, 2014; LILLY TB DRUG DISCOVERY INITIATIVE 2014; NOVARTIS INSTITUTE FOR TROPICAL DISEASES, 2013).

1.3 ALGUMAS CONSIDERAÇÕES SOBRE VACINAS

A única vacina disponível contra Tb (Bacilo Calmette-Guérin-BCG) tem quase 100 anos. Esta vacina protege contra a tuberculose progressiva grave em crianças, mas os seus efeitos protetores em adolescentes e adultos são variáveis. A vacina não é adequada para uso em crianças HIV-positivas. Novas vacinas (ver por exemplo, BRENNAN, THOLE, 2012; TAMERIS et al., 2013) que são seguras e eficazes contra todas as formas de TB em todas as faixas etárias, incluindo a população HIV-positiva, são necessárias (WORLD HEALTH ORGANIZATION, 2009; WORLD HEALTH ORGANIZATION, 2012).

1.4 CLASSIFICAÇÃO DAS PESSOAS QUANTO À DOENÇA

Cinco categorias mutuamente excludentes classificam as pessoas quanto à doença: virgens de infecção (V), infectados (I), doentes não bacilíferos (D), doentes bacilíferos (B), e os que evoluíram para óbito (O). (RUFFINO-NETTO, 2000).

1.5 ALGUNS FATORES QUE PODEM INFLUENCIAR A INCIDÊNCIA DE TB

Com o advento do HIV no início da década de 1980, houve aumento da incidência e prevalência de TB, bem como da mortalidade relacionada ao TB no mundo, isto porque o vírus HIV ataca o sistema imunitário do organismo do indivíduo, oferecendo assim condições favoráveis ao bacilo de Koch para se multiplicar e provocar as manifestações clínicas da doença (MINISTÉRIO DA SAÚDE, 2013a), tais como tosse persistente, febre, suores noturnos, dores no peito, falta de apetite e cansaço. Crianças e idosos também têm mais propensão a desenvolver a tuberculose (HIJJAR, 2006). Há diferentes interpretações sociológicas do processo Saúde-Doença (SD), tais como: funcionalista, compreensivo e materialismo histórico.

A Figura 1 mostra um exemplo de interpretação funcionalista, onde diversos fatores podem influenciar a incidência de Tb, tais como: etnia, classe social, sexo, diabetes, etc. (SILVA, 2004).

Figura 1 - Exemplo de rede de causalidade para a tuberculose (MINISTÉRIO DA SAÚDE, 2013b)

1.6 TUBERCULOSE NO BRASIL

No Brasil, o bacilo possivelmente foi introduzido pelos portugueses, a partir de 1500 (RUFFINO-NETTO, 2002), tendo a sua incidência preocupado as autoridades desde o século XIX. Exemplos dessa preocupação são a existência de organizações filantrópicas de assistência à doença, a Reforma Carlos Chagas, que criou a Inspetoria de Profilaxia da Tuberculose, na década de 1910, o Plano Federal de Construção de Sanatórios, de 1940, o Serviço Nacional de Tuberculose, de 1941, e a Campanha Nacional Contra a Tuberculose, de 1946 (HIJJAR; PROCÓPIO, 2006).

O uso de tuberculostáticos a partir da década de 1940 no Brasil alterou a tendência da mortalidade por Tb no país, continuando a existir, no entanto, esforços para controle da doença, como a Divisão Nacional de Tuberculose, de 1970, a implantação da vacinação com BCG intradérmica em 1973, o Programa de Controle da Tuberculose, de 1975, e a obrigatoriedade da vacinação em menores de um ano de idade a partir de 1976 (HIJJAR; PROCÓPIO, 2006).

Em outubro de 1998, o governo lançou o Plano Nacional de Controle de Tuberculose, motivado por notícias de março daquele ano que chamavam a atenção para a situação calamitosa da tuberculose no mundo (HIJJAR; PROCÓPIO, 2006) e em 1999 o Programa Nacional de Combate à Tuberculose (PNCT) foi formalmente oficializado (em decorrência da recomendação, pelo Ministério da Saúde, da implantação do tratamento supervisionado). O PNCT está integrado ao Sistema Único de Saúde (SUS) e à rede de Serviços de Saúde.

Atualmente a tuberculose no Brasil continua sendo um problema de saúde pública (ver por exemplo, DUVATI et al., 2006; GUIMARÃES et al., 2012; HINO et al., 2011; HIJJAR et al., 2005; NORONHA, 2007; PAIXÃO; GONTIJO, 2007; PINHEIRO et al., 2013; RUFFINO-NETTO, 2000, 2002a, 2002b; SAN PEDRO, 2013; SANTOS et al., 2004, 2007; VENDRAMINI et al., 2005, 2010; VIEIRA, 2008), onde observa-se que algumas metas estabelecidas pelo PNCT ainda não foram atingidas. A meta “tratar corretamente 100% dos casos de tuberculose diagnosticados e curar pelo menos 85% dos mesmos” ainda não foi atingida, isto porque a taxa de abandono ao tratamento é elevada (em torno de 12%) (HIJJAR; PROCÓPIO, 2006).

No estado de São Paulo, o número de casos novos tem se mantido estável ao longo da última década. Dos cerca de 18 mil novos casos anuais no Estado, aproximadamente um terço ocorrem no município de São Paulo (SOLAR; IRWIN, 2005).

O estado de São Paulo é composto por 594 municípios; neste estudo serão considerados os 40 municípios com populações iguais ou superiores a 200.000 mil habitantes, que são: Americana, Araraquara, Barueri, Bauru, Campinas, Carapicuíba, Cotia, Diadema, Embu das Artes, Franca, Guarujá, Guarulhos, Hortolândia, Indaiatuba, Itapevi, Itaquaquecetuba, Jacareí, Jundiaí, Limeira, Marília, Mauá, Mogi das Cruzes, Osasco, Piracicaba, Praia Grande, Presidente Prudente, Ribeirão Preto, Santo André, Santos, São Bernardo do Campo, São Carlos, São José do Rio Preto, São José dos Campos, São Paulo, São Vicente, Sorocaba, Sumaré, Suzano, Taboão da Serra e Taubaté (MINISTÉRIO DA SAÚDE, 2013b). A motivação para a escolha destes municípios foi por serem de grande porte e portando, terem elevados números de casos de tuberculose, bem como, possuírem hospitais e centros de saúde envolvidos com o programa de controle da doença.

2 OBJETIVOS

2.1 OBJETIVO GERAL

- Descrever a distribuição temporal dos casos novos mensais de tuberculose notificados nos municípios com mais de 200.000 habitantes no estado de São Paulo no período 2009 a 2013.

2.2 OBJETIVOS ESPECÍFICOS

- Identificar covariáveis que estejam correlacionadas à incidência de Tb, assumindo diferentes modelos estatísticos de regressão:
 - Modelos de regressão linear para dados transformados ($\log(\text{contagens})$) com erros normais.
 - Modelos de regressão de Poisson para dados de contagem.
 - Modelos com efeitos defasados (por exemplo, aumento da incidência de casos de AIDS nos meses anteriores) na previsão dos números de casos de tuberculose.
- Assumindo os modelos de regressão acima, serão realizadas previsões para o número de casos de Tb (sob o enfoque clássico e Bayesiano).

Para análise Bayesiana (ver por exemplo, BOX; TIAO, 1973; PAULINO; TURKMAN; MURTEIRA, 2003) serão considerados métodos MCMC (Monte Carlo em Cadeias de Markov) para simular amostras da distribuição a posteriori de interesse (ver por exemplo GELFAND; SMITH, 1990; GAMERMAN, 1997; CHIB; GREENBERG, 1995) com o uso do software Open-

Bugs uma versão mais recente do free software WinBugs (SPIEGELHALTER et al., 2001). Para a análise Bayesiana serão testadas diferentes distribuições a priori não-informativas e informativas considerando opinião de especialistas na área de tuberculose.

3 MATERIAL E MÉTODOS

O modelo para este estudo será um modelo ecológico. Serão analisados os dados de pessoas diagnosticadas e notificadas com TB nos municípios com mais de 200.000 habitantes do estado de São Paulo no período de 01/01/2009 a 31/12/2013. A informação dos números de casos novos de Tb notificados será obtida do banco de dados originário do SINAN. Será verificado para este estudo, se as seguintes covariáveis possuem associação com a incidência de Tb: IDHM, renda média, PIB real per capita, quantidade de médicos por 1000 habitantes, quantidade de leitos do SUS por 1000 habitantes, grau de urbanização (em percentual), número de casos de AIDS (taxas por 1000 habitantes), percentual de consumidores de energia elétrica, número de homicídios dolosos por 10000 habitantes, participação da indústria no PIB, número médio de óbitos e quantidade de unidades prisionais. Estas covariáveis foram selecionadas, por indicarem o nível sócio econômico dos municípios e que podem possuir associação com o aumento ou diminuição da incidência de Tb.

3.1 FONTE DOS DADOS

Os dados usados nesse trabalho foram coletados de várias fontes disponíveis na Web.

- **SINAN:** <http://dtr2004.saude.gov.br/sinanweb>

Este banco de dados contém as seguintes covariáveis: município de notificação, município de residência, faixa etária, raça, sexo, tipo de entrada (se é caso novo, recidiva, etc), institucionalizado (situação do paciente que se encontra ou não em cárcere), forma da doença (pulmonar, extra-pulmonar ou ambos), se tem Aids/HIV, resultado da baciloscopia de escarro para BAAR (se foi positiva ou negativa) e se foi indicado tratamento supervisionado por ocasião do diagnóstico.

- **DATASUS:** <http://www2.datasus.gov.br>

Este banco de dados contém informações demográficas e sócio econômicas que serão necessárias para este estudo, além de possuir cadastro das redes hospitalares, dados de morbidade, incapacidade, acesso a serviços, qualidade da atenção, condições de vida, fatores ambientais e entre outros dados que serão relevantes para a avaliação das condições de saúde nos municípios

estudados. Os dados que serão analisados são: Indicadores de Saúde, Assistência à Saúde, Epidemiológicas e Morbidade, Rede Assistencial, Estatísticas Vitais, Demográficas e Socioeconômicas, Inquéritos e Pesquisas e Saúde Suplementar.

- **PNUD** (Programa das Nações Unidas para o Desenvolvimento): <http://www.pnud.org.br>

O PNUD disponibiliza dados do IDHM (Índice de desenvolvimento humano municipal), um indicador composto originado de informações econômicas, sociais e de educação e saúde. O IDHM é um número que varia entre 0 e 1. Quanto mais próximo de 1, maior o desenvolvimento humano de um município. Estes dados serão úteis para estudar se em municípios com níveis diferentes de desenvolvimento a evolução mensal dos casos novos de TB apresentam padrões específicos.

- **SEADE** (Fundação Sistema Estadual de Análise de Dados): <http://www.seade.gov.br>

Possui dados sociais, econômicos e demográficos, provenientes de levantamentos primários feitos pela Fundação Seade, assim como de registros administrativos de outras instituições, com diferentes abrangências temporais e espaciais, sobretudo para os municípios e regiões do Estado de São Paulo.

- **IBGE** (Instituto Brasileiro de Geografia e Estatística): <http://www.ibge.gov.br/home>

Disponibiliza pesquisas que contém informações históricas sobre a realidade brasileira em relação à educação, habitação, trabalho, saúde, organização familiar, características da população, dinâmica demográfica, indicadores demográficos, sistema de contas nacionais, índices de preços, de produção e de comércio; agropecuária.

- **FGV IBRE** (Instituto Brasileiro de Economia): <http://portalibre.fgv.br/>

Possui indicadores históricos de preços, taxas de câmbio, IGP-M, IGP-DI, etc.

- **SAP** (Secretaria de Administração Penitenciária): Possui informações das unidades prisionais do estado de São Paulo: <http://www.sap.sp.gov.br/uni-prisionais/pen.html>

4 ASPECTOS ÉTICOS

O presente projeto foi aprovado pelo Comitê de Ética em pesquisa da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, sob o número de parecer 1.037.788, em 04 de Março de 2015, a fim de garantir o cumprimento da Resolução 466/2012 do Conselho Nacional de Saúde (Anexo no final da dissertação).

5 CONSIDERAÇÕES GERAIS

Segundo o Caderno de Análise dos Indicadores do Programa Nacional de Controle da Tuberculose, a notificação da doença é definida como “a comunicação da ocorrência de determinada doença ou agravo à saúde, feita à autoridade sanitária por profissionais de saúde ou qualquer cidadão, para fins de adoção de medidas de intervenção pertinentes”, e os casos de tuberculose são então notificados somente após confirmação diagnóstica (conforme normas estabelecidas do PNCT), ou seja, não existe notificação de casos suspeitos. São notificados os casos novos, recidivas, outros reingressos e transferências, utilizando a ficha de notificação/investigação de tuberculose do SINAN.

6 ANÁLISE DOS DADOS

Foram considerados os casos novos notificados de tuberculose (todas as formas) no estado de São Paulo para os municípios com mais de 200.000 habitantes (40 municípios), classificados por sexo masculino e feminino, no período de janeiro de 2009 a dezembro de 2013 (mensal). Para este estudo foram excluídos os casos provenientes de presídios, asilos, orfanatos e hospitais psiquiátricos (institucionalizado).

Considerou-se como variável resposta para análise estatística tanto o número de casos novos como o coeficiente de incidência de Tb. O coeficiente de incidência é o número de casos novos da doença em uma população durante o período de tempo especificado dividido pelo número de pessoas sob o risco de desenvolver a doença durante o período de tempo específico, expresso por uma base que pode ser 100, 1 mil, 10 mil ou 100 mil, dependendo da frequência de casos (PAGANO; GAUVREAU, 2004). Os dados sobre a população foram obtidos do Instituto Brasileiro de Geografia e Estatística (IBGE) e considerou-se como base da equação 100 mil.

Para identificar se ocorreu acentuada elevação no número de casos novos de Tb em um ou mais meses do período de tempo, para cada município analisado, utilizou-se carta-controle com 2 desvios.

Foram aplicados modelos estatísticos clássicos e bayesianos para comparação e definição de melhor ajuste. A carta-controle, análises descritivas e modelagem clássica foram realizadas no software MINITAB® versão 17. A modelagem bayesiana foi realizada via software Open Bugs (versão 3.2.3 ver 1012) e via software R (Library QCC).

7 MÉTODOS

7.1 CARTA CONTROLE

A carta controle é uma das diversas ferramentas que foram desenvolvidas nos Estados Unidos nos anos 20 para medir, monitorar, controlar e melhorar a qualidade da produção industrial. A carta controle é um gráfico que auxilia na detecção de uma mudança na média ou na variabilidade em alguma característica na qualidade de um produto ou processo estudado, para que ações corretivas possam ser realizadas antecipadamente. Atualmente tem sido uma ferramenta útil para diversas áreas além do setor industrial, como na área da saúde por exemplo.

Uma carta controle típica é mostrada na figura 2 que é uma representação gráfica de uma característica de qualidade que foi medida ou calculada a partir de uma amostra em relação ao número da amostra ou ao tempo. O quadro contém uma linha central (LC) que representa o valor médio da característica de qualidade medida e duas outras linhas horizontais, chamadas de limite superior de controle (LSC) e limite inferior de controle (LIC). O LSC é o valor de variação máximo aceitável da média da característica de qualidade medida e o LIC é o valor de variação mínimo aceitável da média da característica de qualidade medida. Estes limites de controle são escolhidos de modo que se o processo estiver sob controle a grande maioria dos pontos estarão entre estas duas linhas. Na área industrial em geral existe uma variabilidade fixada e aceitável pela indústria relacionada à capacidade do processo, já na área de saúde esses limites de controle são estimados a partir dos dados.

Figura 2 - Exemplo de carta controle

O limite superior de controle (LSC) e limite inferior de controle (LIC) numa carta controle usualmente são estabelecidos como iguais a 3 desvios-padrões da média. Se assumirmos que os dados possuem uma distribuição normal, esses limites de controle capturam 99,74% da

variabilidade normal, caso os limites de controle sejam dados por 2 desvios padrões, os limites capturam 95,44% dos valores e para 1 desvio padrão os limites capturam 68,26% (ver figura 3).

Figura 3 - Distribuição normal

Existem diversos tipos de cartas de controle como por exemplo, cartas de controle de proporções (cartas-P) baseada na distribuição binomial ou cartas de contagem (cartas-C) baseadas na distribuição de Poisson. Para este estudo foi utilizado cartas de controle das médias de taxas mensais de Tb ocorridas obtidas via software MINITAB®. Para este estudo os limites foram calculados da seguinte forma:

$$\begin{aligned} \text{LCS} &= \text{média da amostra} + 2 \times \text{desvios padrão da amostra} \\ \text{LIC} &= \text{média da amostra} - 2 \times \text{desvios padrão da amostra} \end{aligned} \quad (1)$$

7.2 MODELOS DE REGRESSÃO LINEAR

Na análise de regressão linear, os dados são modelados utilizando-se funções de previsão linear, e os parâmetros do modelo são desconhecidos e estimados a partir dos dados. Tais modelos são chamados modelos lineares. Mas, geralmente, refere-se a regressão linear para um modelo em que a média condicional de Y, dado o valor de X, é uma função de X.

O método de regressão linear tem muitos usos práticos. A maioria das aplicações se enquadram em uma das duas seguintes grandes categorias:

- Se o objetivo é a predição ou previsão, modelos de regressão linear podem ser utilizados para ajustar um modelo preditivo para um conjunto de dados observados de valores Y e X. Após o ajuste de um modelo desse tipo, se um valor adicional de X é dado, então o modelo ajustado pode ser usado para fazer uma previsão do valor de y.

- Dadas as variáveis ou covariáveis $X_1, \dots, X_p$, que podem estar relacionadas com a resposta ou variável dependente Y , a análise de regressão linear, pode ser utilizada para quantificar a magnitude da relação entre Y e X_j , $j = 1, \dots, p$. Isso é dado por testes de hipóteses nos parâmetros de regressão.

Modelos de regressão linear são frequentemente ajustados usando-se a abordagem ajustamento pelo método dos mínimos quadrados.

Quando utilizamos mais de uma variável explanatória para prever o comportamento de uma variável resposta, passamos a nomeá-lo como modelo de regressão múltipla (DRAPER; SMITH, 1981).

Na análise de regressão linear verifica-se o efeito conjunto das covariáveis na resposta Y (ver por exemplo, DRAPER; SMITH, 1981; SEBER; LEE, 2003; MONTGOMERY; RUNGER, 2011).

Um exemplo de modelo de regressão com duas variáveis independentes é dado por:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (2)$$

para $i = 1, \dots, n$ e ε_i é um termo relacionado ao erro (variável não-observada) suposto como uma quantidade aleatória assumido com uma distribuição normal com média igual à zero e variância desconhecida σ^2 . Este erro aleatório inclui todos os outros fatores que poderiam influenciar a variável dependente Y não incluídos no modelo de regressão.

7.3 INTRODUÇÃO AOS MÉTODOS BAYESIANOS

Os métodos Bayesianos têm sido considerados alternativas muito eficazes e poderosas na análise de dados. Esse método possui uma filosofia muito diferente do método frequentista. No método clássico, os parâmetros do modelo são considerados constantes desconhecidas, e no método Bayesiano todos os parâmetros são considerados quantidades aleatórias. Uma característica muito importante é o fato desta análise permitir a incorporação da informação de um especialista junto à informação dos dados.

A inferência Bayesiana tem como fundamentação a Fórmula de Bayes, a qual combina os dados com a informação a priori e, então, se obtém a distribuição a posteriori (priori já complementada pela informação dos dados), onde é realizado todo o processo inferencial (ver, por exemplo, BOX; TIAO, 1973; PAULINO; TURKMAN; MURTEIRA, 2003).

7.3.1 Fórmula de Bayes

Considere uma partição do espaço amostral Ω , onde os eventos $A_1, A_2, \dots, A_k$ formam uma sequência de eventos mutuamente exclusivos e exaustivos, isto é, $\cup_{j=1}^k A_j = \Omega$ e $A_i \cap A_j = \emptyset$ (conjunto vazio) para $i \neq j$ tal que $P(\cup_{j=1}^k A_j) = \sum_{j=1}^k P(A_j) = 1$. Sendo assim, para qualquer outro evento B ($B \subset \Omega$), temos:

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^k P(B | A_j)P(A_j)} \quad (3)$$

para todo i variando de 1 até k .

A fórmula de Bayes (3) pode ser interpretada da seguinte forma: antes do conhecimento de qualquer informação sobre o evento A_i , atribui-se uma probabilidade a priori para A_i dada por P , atualizada a partir da ocorrência do evento B . Essa probabilidade atualizada, ou probabilidade condicional do evento A_i dado a ocorrência do evento B (probabilidade a posteriori), ou seja, $P(A_i | B)$, é dada pela fórmula de Bayes (3).

Considere agora $y = (y_1, \dots, y_n)$ uma amostra aleatória, onde os dados são independentes e identicamente distribuídos. Supor que o vetor aleatório Y possui uma distribuição conjunta dada pela função densidade de probabilidade assumindo dados contínuos (ou função de probabilidade assumindo dados discretos) $f(y|\theta)$ e seja $\pi(\theta)$ uma distribuição a priori para o vetor de parâmetros θ . Observar que também podemos ter distribuições de probabilidade com um parâmetro θ em vez de um vetor de parâmetros θ .

Suponha que o parâmetro θ assume valores contínuos num dado intervalo. Considerando uma amostra aleatória $y = (y_1, \dots, y_n)$ (os dados são independentes e identicamente distribuídos com uma distribuição conjunta dada pela densidade $f(y|\theta)$, também definida como função de verossimilhança para θ), e uma distribuição a priori para θ (dada por $\pi(\theta)$), temos de (3), a distribuição a posteriori para θ dado y ,

$$\pi(\theta | y) = \frac{f(y | \theta) \pi(\theta)}{\int f(y | \theta) \pi(\theta) d\theta} \quad (4)$$

em que a integral no denominador de (4) é definida no intervalo de variação de θ .

Em geral, não é necessário o cálculo da integral no denominador e, assim, a posteriori é proporcional à priori multiplicada pela verossimilhança. No caso discreto, como da distribuição Poisson, substitui-se a integral do denominador por $\sum_{\theta}$.

7.3.2 Distribuição a Priori

Na análise Bayesiana, a distribuição a priori é utilizada a fim de representar o que já é conhecido sobre parâmetros desconhecidos, antes de se avaliar os dados. Deve-se ter muita cautela ao definir uma distribuição a priori, afinal se esta informação não for bem definida pode-se chegar a interpretações errôneas. Uma distribuição a priori para um parâmetro pode se dar de várias formas, sendo possível ocorrerem a partir de procedimentos subjetivos ou objetivos.

A distribuição a priori conjugada é uma priori informativa, onde a distribuição a priori e a posteriori pertencem à mesma classe de distribuições. A passagem da priori para a posteriori envolve apenas uma simples mudança nos parâmetros, sem a necessidade de cálculos adicionais (ver, por exemplo, PAULINO; TURKMAN; MURTEIRA, 2003).

A distribuição a priori também pode ser a incorporação do conhecimento de um pesquisador, ou seja, o pesquisador se baseia na sua prática e no seu *feeling* para definir a priori. Em muitas situações práticas esse conhecimento do especialista não existe ou, se existe, não é fidedigno. Nesse caso, caracteriza-se uma “ignorância a priori” (ver, por exemplo, PAULINO; TURKMAN; MURTEIRA, 2003). Nestas situações são utilizadas distribuições a priori não informativas, ou seja, prioris de referência “neutras” (ver, por exemplo, BOX; TIAO, 1973). A utilização deste tipo de priori permite a comparação com os resultados obtidos pela inferência clássica, haja vista que através de uma priori não informativa, o modelo é baseado apenas na informação dos dados amostrais. Existem vários métodos para se definir a priori não informativa, como por exemplo: Método de Bayes- Laplace, Método de Jeffreys, entre outros (ver, por exemplo, BOX; TIAO, 1973; PAULINO; TURKMAN; MURTEIRA, 2003).

7.3.3 Intervalos de credibilidade

Na prática, muitas vezes é mais apropriado estimar com um certo grau de confiança um intervalo de valores no qual se espera que o parâmetro esteja contido, como complementação à estimadores pontuais que pode ser a média, a moda ou a mediana da distribuição a posteriori. Na inferência clássica esse intervalo é definido como intervalo de confiança. Considerando um coeficiente de confiança $(1-\alpha)$, por exemplo, igual à 95%, a interpretação para um parâmetro é a seguinte: se retirássemos da população um número grande de amostras de tamanho n , 95%

destas amostras iriam gerar intervalos de confiança que contém o verdadeiro valor de θ . Assim, em termos frequentistas, o parâmetro é considerado fixo e o intervalo de confiança é aleatório.

Diferente da inferência clássica, a construção e interpretação de um intervalo para um parâmetro de interesse sob enfoque Bayesiano é feita de maneira mais intuitiva, levando muitos cientistas e não-estatísticos a interpretar intervalos de confiança no sentido dos intervalos de credibilidade bayesianos. Seja θ um parâmetro unidimensional e sua distribuição *a posteriori*, unimodal. Uma forma de estimador por intervalo Bayesiano com probabilidade $(1-\alpha)$ será dado por (θ_*, θ^*) para

$$\int_{-\infty}^{\theta_*} \pi(\theta | y) d\theta = \frac{\alpha}{2} \quad (5)$$

$$\int_{\theta^*}^{\infty} \pi(\theta | y) d\theta = \frac{\alpha}{2}$$

O intervalo (θ_*, θ^*) é chamado intervalo de credibilidade para θ com probabilidade $(1-\alpha)$ ou seja, a probabilidade de que θ esteja no intervalo (θ_*, θ^*) é igual à $1-\alpha$.

Nos casos em que a distribuição *a posteriori* é assimétrica, uma outra forma de se encontrar intervalos de credibilidade para o parâmetro, é encontrar intervalos que minimizem $\theta_* - \theta^*$, mas com uma cobertura $1-\alpha$, como o intervalo HPD, por exemplo.

7.3.4 Métodos de simulação para obtenção dos sumários *a posteriori*

Na obtenção de sumários *a posteriori* de interesse, geralmente não é possível encontrar uma solução analítica, fato comum quando se tem um vetor θ de parâmetros. Como solução deste problema, várias alternativas foram introduzidas na literatura. Um caso especial é dado pela aproximação de Laplace que utiliza uma aproximação normal para a distribuição *a posteriori* de interesse mas esse método muitas vezes pode não ser o método mais indicado para situações onde o vetor de parâmetros é de dimensão maior ou igual à 5. Uma alternativa que tornou o uso de métodos bayesianos muito popular foi o aparecimento de várias técnicas de simulação de Monte Carlo. De forma geral, o desenvolvimento computacional propiciou que os métodos de simulação pudessem ser utilizados de forma intensiva para a obtenção dos sumários *a posteriori* de interesse, sendo o método de Monte Carlo via função por importância e o de Monte Carlo via Cadeias de Markov (métodos MCMC) muito utilizados pelos pesquisadores das mais diversas áreas.

7.3.5 Método de Monte Carlo em cadeias de Markov

Supor que não podemos gerar diretamente uma amostra de uma distribuição a posteriori $\pi(\theta | y)$, $\theta \in \mathbb{R}^k$. Entretanto, supor que podemos construir uma cadeia de Markov com espaço de estados no espaço paramétrico Θ (conjunto de todos os valores possíveis de θ) que é simples para simular e cuja distribuição de equilíbrio seja dada por $\pi(\theta | y)$. Com a geração de muitas amostras dessa cadeia, os valores simulados podem ser usados para sumarizar características da posteriori $\pi(\theta | y)$.

Se a distribuição conjunta a posteriori $\pi(\theta | y)$ for positiva em $\Theta_1 \times \Theta_2 \times \dots \times \Theta_k$, com Θ_i sendo suporte para a distribuição de θ_i , $i = 1, \dots, k$, então a distribuição a posteriori $\pi(\theta | y)$ é unicamente determinada pelas distribuições condicionais completas $\pi(\theta_i | y, \theta_{(i)})$ para $i = 1, \dots, k$ em que $\theta = \theta_1, \dots, \theta_k$ e $\theta_{(i)}$ é o vetor de todos os componentes de θ exceto θ_i , isto é, $\theta_{(i)} = \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k$.

Sob algumas condições de regularidade é fácil ver que os resultados simulados da cadeia com distribuição de equilíbrio $\pi(\theta | y)$ podem ser supostos com uma amostra aleatória de $\pi(\theta | y)$ (ver por exemplo, GAMERMAN, 1997).

7.3.6 Amostrador de Gibbs

Supor que estamos interessados em obter inferências da distribuição a posteriori conjunta, $\pi(\theta | y)$, $\theta = (\theta_1, \dots, \theta_k)$. Para isso simulamos quantidades aleatórias de distribuições condicionais completas $\pi(\theta_i | y, \theta_{(i)})$, $i = 1, \dots, k$, onde $\theta_{(i)}$ denota o vetor de todos os componentes θ_j , exceto para j igual a θ_i que produzem uma cadeia de Markov.

Observar que $\pi(\theta_i | y, \theta_{(i)})$ são facilmente identificadas como funções de θ_i por inspeção da forma de $\pi(\theta | y)$, a distribuição a posteriori para θ dado y . (GAMERMAN, 1997).

Atribuindo um conjunto arbitrário de valores iniciais $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}$ para o vetor de parâmetros θ , podemos escrever o algoritmo na forma:

- (i) Gerar $\theta_1^{(1)}$ de $\pi(\theta_1 | y, \theta_2^{(0)}, \dots, \theta_k^{(0)})$;
 - (ii) Gerar $\theta_2^{(1)}$ de $\pi(\theta_2 | y, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$;
 - (iii) Gerar $\theta_3^{(1)}$ de $\pi(\theta_3 | y, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$;
 - .
 - .
 - .
 - (k) Gerar $\theta_k^{(1)}$ de $\pi(\theta_k | y, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)})$
- (6)

Então, o próximo passo será substituir os valores iniciais com uma nova realização $\theta^{(1)} = \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)}$ de θ e repetir o processo acima.

Para t suficientemente grande, o valor $\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}$ converge para um valor da quantidade aleatória com distribuição $\pi(\theta | y)$ (ver GELFAND; SMITH, 1990). Além disso, $\theta_j^{(t)}$ pode ser considerado como uma observação simulada da distribuição a posteriori marginal $\pi(\theta_j | y)$, $j = 1, 2, \dots, k$.

7.3.7 Algoritmo de Metropolis-Hastings

O amostrador de Gibbs seleciona amostras de todas as densidades a posteriori condicionais de $\pi(\theta | y)$. Quando usamos uma priori conjugada, as distribuições condicionais em geral são de forma conhecida, como, por exemplo, as distribuições normal, gama, Poisson, Beta, etc, e a simulação de amostras dessas distribuições são disponíveis em diversos softwares, como o R e o OpenBugs.

Quando as distribuições condicionais não são facilmente identificadas, devemos usar o algoritmo de Metropolis-Hastings ou métodos de amostragem por importância. Para detalhes do algoritmo de Metropolis-Hastings ver por exemplo Chib e Greenberg (1995).

É importante salientar que nas aplicações podemos usar softwares existentes que facilitam o procedimento de simulação de amostras da distribuição a posteriori conjunta de interesse. Um software muito popular usado extensivamente na estatística aplicada é o software OpenBugs uma versão mais recente do software WinBugs (SPIEGELHALTER et al., 2001). Usando o software OpenBugs só é necessário definir a verossimilhança e a distribuição a priori para os parâmetros do modelo não tendo necessidade de especificar todas as distribuições condicionais $\pi(\theta_i | y, \theta_{(i)})$ requeridas pelo amostrador de Gibbs.

8 RESULTADOS

8.1 ANÁLISE PRELIMINAR DOS DADOS

Inicialmente foi realizada análise descritiva dos dados para avaliação dos comportamentos das variáveis selecionadas para o estudo em relação ao tempo e por cidade. O primeiro passo foi observar se havia alguma tendência cíclica no período de tempo analisado, para os 40 municípios agrupados (ver Figuras 4 e 5) e observados nos 60 meses de acompanhamento.

Figura 4 - Análise geral das notificações de tuberculose para os dados dos 40 municípios agrupados

Figura 5 - Carta-control para a taxa de Tb para os dados dos 40 municípios agrupados

Por ambas as Figuras 4 e 5, é observado um aumento no coeficiente de incidência de Tb para alguns meses do ano, como por exemplo: no mês de março em relação a fevereiro. Pela carta-control identificou-se 3 pontos fora dos limites: março de 2009 (incidência acima da média), junho de 2010 (incidência abaixo da média) e dezembro de 2010 (incidência acima da média). Assim, observa-se grande volatilidade para as incidências de novos casos de Tb no ano de 2010 considerando-se os dados de todos os 40 municípios agrupados. A partir do ano 2011 essa grande volatilidade

não é verificada. Esses resultados indicam que algum fator não observado pode estar relacionado com a incidência de Tb que poderiam ser discutidos por epidemiologistas da área de Tb.

Também é interessante observar que quando as taxas de incidência de novos casos de tuberculose são observados separadamente para cada município com mais de 200.000 habitantes do Estado de São Paulo observa-se que há quatro grupos distintos de municípios (ver Tabela 1 abaixo, onde as taxas de incidência são ordenadas):

Tabela 1 - Taxas de Tb ordenadas para os 40 municípios

Obs	cidades	taxas Tb ordenadas
1	19 - São Vicente	6,99140
2	12 - Santos	6,86468
3	22 - Guarujá	6,54189
4	25 - Praia Grande	6,13279
5	1 - São Paulo	4,36088
6	39 - Itapevi	3,84299
7	29 - Barueri	3,49577
8	30 - Embu	3,13261
9	17 - Piracicaba	3,07674
10	27 - Taboão da Serra	3,05784
11	16 - Carapicuíba	3,04557
12	18 - Bauru	2,95892
13	8 - Ribeirão Preto	2,93682
14	3 - Campinas	2,89907
15	6 - Osasco	2,86327
16	13 - Moji das Cruzes	2,77281
17	14 - Diadema	2,71501
18	20 - Itaquaquecetuba	2,66134
19	11 - São José do Rio Preto	2,65097
20	15 - Jundiaí	2,63087
21	23 - Taubaté	2,61033
22	38 - Presidente Prudente	2,56168
23	9 - Sorocaba	2,54980
24	26 - Suzano	2,49242
25	34 - Jacaré	2,35175
26	10 - Mauá	2,30852
27	2 - Guarulhos	2,28706
28	5 - Santo André	2,26031
29	24 - Limeira	2,16368
30	32 - Marília	2,12398
31	33 - Americana	2,05674
32	37 - Cotia	2,01537
33	7 - São José dos Campos	2,00117
34	4 - São Bernardo do Campo	1,92218
35	31 - São Carlos	1,84792
36	35 - Indaiatuba	1,62984
37	21 - Franca	1,45351
38	28 - Sumaré	1,42208
39	36 - Araraquara	1,33096
40	40 - Hortolândia	1,30118

- Um grupo de quatro cidades com taxas médias mensais de novos casos de Tb com valores maiores do que 6 observados no período de 5 anos que corresponde a 60 meses (grupo 1): São Vicente, Santos, Guarujá e Praia Grande.
- Um grupo de sete cidades com taxas médias mensais de novos casos de Tb com valores maiores do que 3 e menores do que 5 observados no período de 5 anos que corresponde a 60 meses (grupo 2): São Paulo, Itapevi, Barueri, Embu, Piracicaba, Taboão da Serra e Carapicuíba.
- Um grupo de vinte e duas cidades com taxas médias mensais de novos casos de Tb com valores maiores do que 2 e menores do que 3 observados no período de 5 anos que corresponde a 60 meses (grupo 3): Bauru, Ribeirão Preto, Campinas, Osasco, Mogi das Cruzes, Diadema, Itaquaquecetuba, São José do Rio Preto , Jundiaí, Taubaté, Presidente Prudente, Sorocaba, Suzano, Jacarei, Mauá, Guarulhos, Santo André, Limeira, Marília, American, Cotia e São José dos Campos.
- Um grupo de sete cidades com taxas médias mensais de novos casos de Tb com valores maiores do que 1 e menores do que 2 observados no período de 5 anos que corresponde a 60 meses (grupo 4): São Bernardo do Campo, São Carlos, Indaiatuba, France, Sumaré, Araraquara e Hortolândia.

Na Figura 6, observa-se as séries temporais mensais das taxas médias de incidência de novos casos de Tb para cada um dos quatro grupos de cidades considerando o período de 60 meses de seguimento. E nota-se picos que se repetem em alguns meses do ano nas incidências de novos casos de Tb para os quatro grupos de cidades.

Figura 6 - Séries temporais mensais das taxas médias de incidência de novos casos de Tb para cada um dos quatro grupos de cidades

Também é observado que somente as cidades do grupo 1 (São Vicente, Santos, Guarujá e Praia Grande) estão agrupadas em uma mesma região geográfica (região litorânea do Estado de São Paulo), isto poderia ter um efeito espacial na incidência de novos casos de Tb; os outros três grupos de cidades com taxas de incidência de novos casos de Tb similares estão localizadas aleatoriamente em vários pontos geográficos do Estado de São Paulo (ver mapa do estado de São Paulo na Figura 7), isso mostra que o uso de um modelo espaço-temporal, comumente assumido em análise de dados epidemiológicos (ver por exemplo, KELSALL and WAKEFIELD, 2002; WALLER et al., 1997; MARSHALL, 1991), não seria apropriado para a análise dos dados de incidência de Tb considerados neste estudo. Deste modo, serão assumidos modelos de regressão na presença de outras covariáveis que podem estar explicando essa grande variabilidade nas incidências de novos casos de Tb.

Figura 7 - Mapa do Estado de São Paulo com algumas das principais cidades

Para cada município também foi feito uma carta-controle (ver Apêndice 1) onde interpretações temporais para as ocorrências de Tb para cada município podem ser obtidas desses gráficos e para cada covariável definida para este estudo verificou-se sua distribuição no período de tempo estudado para cada município (ver Apêndice 2). A partir dos gráficos apresentados no Apêndice A para as incidências de Tb separadamente para cada cidade, observa-se em resumo que:

- Município de São Paulo: para esse município observa-se uma maior variabilidade (desvio-padrão) para a incidência de Tb no período analisado (janeiro/2009 a dezembro de 2010). É importante observar que em março de 2009 e dezembro de 2010 a incidência de Tb ultrapassa o limite superior da carta controle. A partir de janeiro de 2012 grande parte dos pontos analisados (incidência de Tb) ficam abaixo da média.
- Município de Guarulhos: nos meses iniciais de acompanhamento (janeiro/2009 a dezembro de 2010) observa-se certa estabilidade na incidência de Tb com uma volatilidade dentro de uma variação aceitável (distribuição normal); a partir de janeiro de 2011, observa-se uma volatilidade maior, com cinco meses com taxas fora da margem de controle epidemiológico (junho e julho de 2011, novembro de 2011, julho de 2012 e

outubro de 2013). E há uma tendência de aumento de taxas de incidência de Tb (mais valores acima da média) para o final de período de acompanhamento

- Município de Campinas: nos meses iniciais de acompanhamento (janeiro/2009 a setembro de 2010) observa-se certa estabilidade na incidência de Tb com uma volatilidade dentro de uma variação aceitável (distribuição normal); a partir novembro de 2010, observa-se uma volatilidade maior, com quatro meses com taxas fora da margem de controle epidemiológico (novembro de 2010, julho de 2011, setembro e outubro de 2012). Também observa-se uma tendência de aumento de taxas de incidência de Tb (mais valores acima da média) para o final de período de acompanhamento).
- Município de São Bernardo: para esse município observa-se uma certa estabilidade na incidência de Tb com uma volatilidade dentro de uma variação aceitável (distribuição normal) na maioria dos meses de acompanhamento; observa-se três meses com taxas fora da margem de controle epidemiológico (abril de 2010, julho de 2011 e maio de 2013). Possivelmente esse município apresenta níveis de algumas covariáveis socioeconômicas e de saúde que podem estar dando estabilidade nas incidências de novos casos de Tb no período de acompanhamento. A variabilidade (desvio-padrão) também não é muito grande.
- Município de Santo André: para esse município observa-se um comportamento similar ao município de São Bernardo, isto é, com uma certa estabilidade na incidência de Tb com uma volatilidade dentro de uma variação aceitável (distribuição normal) na maioria dos meses de acompanhamento; observa-se dois meses com taxas fora da margem de controle epidemiológico (janeiro de 2009 agosto de 2011). Possivelmente esse município também apresenta níveis de algumas covariáveis socioeconômicas e de saúde que podem estar dando estabilidade nas incidências de novos casos de Tb no período de acompanhamento. A variabilidade (desvio-padrão) também não é muito grande.
- Município de Osasco: para esse município observa-se uma maior variabilidade (desvio-padrão) para a incidência de Tb no período de acompanhamento com maiores volatilidades no início e no final do período de acompanhamento; observa-se quatro meses com taxas fora da margem de controle epidemiológico (abril de 2009, agosto de 2009, novembro de 2012 e janeiro de 2013).

- Município de São José dos Campos: para esse município observa-se uma menor variabilidade (desvio-padrão) para a incidência de Tb no período de acompanhamento com grande estabilidade durante todo o período de acompanhamento; observa-se apenas um mês com taxas fora da margem de controle epidemiológico (março de 2012). Possivelmente esse município também apresenta níveis de algumas covariáveis socioeconômicas e de saúde que podem estar dando estabilidade nas incidências de novos casos de Tb no período de acompanhamento. A variabilidade (desvio-padrão) também não é muito grande.
- Município de Ribeirão Preto: para esse município observa-se uma maior variabilidade (desvio-padrão) para a incidência de Tb no período de acompanhamento; observa-se três meses com taxas fora da margem de controle epidemiológico (abril de 2011, fevereiro de 2012 e março de 2013). Observa-se uma pequena volatilidade no período de julho de 2009 a janeiro de 2011; a partir desse período há uma grande volatilidade nas taxas mensais de incidência de novos casos de Tb.
- Município de Sorocaba: para esse município observa-se uma certa estabilidade na incidência de Tb com uma volatilidade dentro de uma variação aceitável (distribuição normal) na maioria dos meses de acompanhamento; observa-se apenas um mês com taxas fora da margem de controle epidemiológico (setembro de 2011). Possivelmente esse município também apresenta níveis de algumas covariáveis socioeconômicas e de saúde que podem estar dando estabilidade nas incidências de novos casos de Tb no período de acompanhamento.
- Município de Mauá: para esse município observa-se apenas dois meses com taxas fora da margem de controle epidemiológico (julho de 2011 e agosto de 2013). Possivelmente esse município também apresenta níveis de algumas covariáveis socioeconômicas e de saúde que podem estar dando estabilidade nas incidências de novos casos de Tb no período de acompanhamento.
- Município de São José do Rio Preto: para esse município observa-se uma maior variabilidade (desvio-padrão) para a incidência de Tb no período de acompanhamento; observa-se dois meses com taxas fora da margem de controle (fevereiro de 2009 e março

de 2012). Também observa-se uma menor volatilidade no final do período de acompanhamento (a partir de janeiro de 2013).

- Município de Santos: para esse município observa-se uma grande variabilidade (desvio-padrão) para a incidência de Tb no período de acompanhamento; observa-se três meses com taxas fora da margem de controle (abril de 2009, agosto de 2011 e abril de 2013). A taxa média de incidência de Tb é muito maior do que para os outros municípios.
- Município de Mogi das Cruzes: para esse município observa-se apenas dois meses com taxas fora da margem de controle epidemiológico (junho de 2010 e dezembro de 2012). Possivelmente esse município também apresenta níveis de algumas covariáveis socioeconômicas e de saúde que podem estar dando estabilidade nas incidências de novos casos de Tb no período de acompanhamento.
- Município de Diadema: para esse município observa-se apenas três meses com taxas fora da margem de controle epidemiológico (agosto de 2009, março de 2011 e agosto de 2012).
- Município de Jundiaí: para esse município observa-se que as taxas de incidência de Tb decrescem a partir de janeiro de 2011; o município apresenta três meses com taxas fora da margem de controle epidemiológico (agosto de 2009, agosto de 2010 e dezembro de 2010).
- Município de Carapicuíba: para esse município observa-se que as taxas de incidência de Tb apresentam volatilidades similares no período de acompanhamento; o município apresenta quatro meses com taxas fora da margem de controle epidemiológico (fevereiro de 2010, junho de 2010, agosto de 2012, e dezembro de 2012).
- Município de Piracicaba: para esse município observa-se que as taxas de incidência de Tb tendem a um aumento a partir de janeiro de 2012 (mais taxas acima da média); o município apresenta nove meses com taxas fora da margem de controle epidemiológico (maio de 2010, junho de 2010, dezembro de 2010, agosto de 2011, janeiro de 2012, agosto de 2012, setembro de 2012, outubro de 2012 e março de 2013).
- Município de Bauru: para esse município observa-se que as taxas de incidência de Tb tendem a um aumento a partir de janeiro de 2013 (mais taxas acima da média); o município apresenta três meses com taxas fora da margem de controle epidemiológico (fevereiro de 2010, setembro de 2011 e junho de 2013).

- Município de São Vicente: para esse município observa-se que as taxas de incidência de Tb são bem altas em todo o período de acompanhamento de 5 anos com grande volatilidade; o município apresenta um mês com taxa fora da margem de controle epidemiológico (março de 2009).
- Município de Itaquacetuba: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas: o município apresenta dois meses com taxas fora da margem de controle epidemiológico (abril de 2009 e dezembro de 2013).
- Município de Franca: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas e baixas: o município apresenta três meses com taxas fora da margem de controle epidemiológico (outubro de 2011, março de 2012 e fevereiro de 2013).
- Município de Guarujá: para esse município observa-se que as taxas de incidência de Tb são razoavelmente altas, onde no início do período de acompanhamento (ano de 2009) as taxas de incidência de Tb são menores: o município apresenta dois meses com taxas fora da margem de controle epidemiológico (julho de 2012 e dezembro de 2012).
- Município de Taubaté: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas e baixas; o município não apresenta nenhum mês com taxas fora da margem de controle epidemiológico.
- Município de Limeira: para esse município observa-se que as taxas de incidência de Tb são razoavelmente baixas; o município apresenta três meses com taxas fora da margem de controle epidemiológico (julho de 2010, fevereiro de 2012 e maio de 2012).
- Município de Praia Grande: para esse município observa-se que as taxas de incidência de Tb são razoavelmente altas (similares ao município de Guarujá), onde no início do período de acompanhamento (ano de 2009) as taxas de incidência de Tb são menores: o município apresenta seis meses com taxas fora da margem de controle epidemiológico (janeiro de 2009, abril de 2009, outubro de 2010, dezembro de 2011, abril de 2012 e junho de 2013).
- Município de Suzano: para esse município observa-se que as taxas de incidência de Tb são razoavelmente baixas; o município apresenta dois meses com taxas fora da margem de controle epidemiológico (janeiro de 2010 e janeiro de 2012).

- Município de Taboão da Serra: para esse município observa-se que as taxas de incidência de Tb são razoavelmente baixas; o município apresenta dois meses com taxas fora da margem de controle epidemiológico (agosto de 2009 e agosto de 2013).
- Município de Sumaré: para esse município observa-se que as taxas de incidência de Tb são bem baixas; o município apresenta cinco meses com taxas fora da margem de controle epidemiológico (fevereiro de 2009, novembro de 2009, agosto de 2011 e junho de 2013).
- Município de Barueri: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas: o município apresenta cinco meses com taxas fora da margem de controle epidemiológico (fevereiro de 2010, outubro de 2010, agosto de 2011, agosto de 2012 e junho de 2013).
- Município de Embu das Artes: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas: o município apresenta dois meses com taxas fora da margem de controle epidemiológico (janeiro de 2013 e julho de 2013).
- Município de São Carlos: para esse município observa-se que as taxas de incidência de Tb são razoavelmente baixas; o município apresenta um mês com taxa fora da margem de controle epidemiológico (abril de 2012).
- Município de Marília: no início do período de acompanhamento (ano de 2009) e no final do período de acompanhamento (final de 2013) as taxas de incidência de Tb são menores: o município apresenta oito meses com taxas fora da margem de controle epidemiológico (outubro de 2009, outubro de 2010, janeiro de 2011, fevereiro de 2011, outubro de 2011, novembro de 2011, dezembro de 2011 e dezembro de 2012).
- Município de Americana: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas: o município apresenta dois meses com taxas fora da margem de controle epidemiológico (fevereiro de 2009, janeiro de 2013 e junho de 2013).
- Município de Jacareí: para esse município observa-se que as taxas de incidência de Tb são razoavelmente homogêneas: o município apresenta um mês com taxa fora da margem de controle epidemiológico (março de 2011).
- Município de Indaiatuba: para esse município observa-se que as taxas de incidência de Tb são bem baixas; o município apresenta um mês com taxa fora da margem de controle epidemiológico (junho de 2011).

- Município de Araraquara: para esse município observa-se que as taxas de incidência de Tb são bem baixas; o município apresenta dois meses com taxa fora da margem de controle epidemiológico (maio de 2012 e setembro de 2013).
- Município de Cotia: para esse município observa-se que as taxas de incidência de Tb são baixas; o município apresenta quatro meses com taxa fora da margem de controle epidemiológico (junho de 2009, setembro de 2010, outubro de 2010 e novembro de 2011).
- Município Presidente Prudente: para esse município observa-se uma homogeneidade para a incidência de Tb no período de acompanhamento; observa-se quatro meses com taxas fora da margem de controle (setembro de 2009, fevereiro de 2010, dezembro de 2010, e novembro de 2011).
- Município de Hortolândia: para esse município observa-se que as taxas de incidência de Tb são bem baixas; o município apresenta um mês com taxa fora da margem de controle epidemiológico (novembro de 2010).
- Município Itapevi: para esse município observa-se uma homogeneidade para a incidência de Tb no período de acompanhamento; observa-se dois meses com taxas fora da margem de controle (novembro de 2012 e agosto de 2013).

Na Tabela 2 temos os números de meses fora de controle epidemiológico na contagem de novos casos de Tb para cada cidade (ver gráficos de controle no Apêndice A, no final desta dissertação).

A partir dos resultados da Tabela 2, observa-se que para alguns municípios há grande estabilidade para os novos casos de Tb no período de acompanhamento, independentemente dessa taxa ser pequena ou bem elevada, como por exemplo, o município de Taubaté; outros municípios, como por exemplo, Piracicaba, Marília, Praia Grande entre vários outros, aparecem vários meses com taxas fora da faixa de controle em torno da média observada para o município.

Para alguns municípios, claramente há um decaimento para meses atípicos no decorrer dos anos (por exemplo, o município de Jundiaí). Para outros municípios é observado o contrário, um aumento de meses atípicos (fora de controle) nos últimos anos de acompanhamento (por exemplo, o município de Americana).

Esses resultados claramente indicam o efeito de alguns fatores e covariáveis que podem estar afetando as taxas de incidência de Tb para os municípios no período de acompanhamento (estudo a ser realizado nessa dissertação de mestrado).

Tabela 2 - Números de meses fora de controle para os 40 municípios

Cidade	2009	2010	2011	2012	2013
São Paulo	1	1	0	0	0
Guarulhos	0	0	3	1	1
Campinas	0	1	1	2	0
S. Bernardo	0	1	1	0	1
S. André	1	0	1	0	0
Osasco	2	0	0	1	1
S. J. Campos	0	0	0	1	0
Rib. Preto	0	0	1	1	1
Sorocaba	0	0	1	0	0
Mauá	0	0	1	0	1
S. J. Rio Preto	1	0	0	1	0
Santos	1	0	1	0	1
Mogi das Cruzes	0	1	0	1	0
Diadema	1	0	1	1	0
Jundiaí	1	2	0	0	0
Carapicuíba	0	2	0	2	0
Piracicaba	0	3	1	4	1
Bauru	0	1	1	0	1
S. Vicente	1	0	0	0	0
Itaquacetuba	1	0	0	0	1
Franca	0	0	1	1	1
Guarujá	0	0	0	2	0
Taubaté	0	0	0	0	0
Limeira	0	1	0	2	0
Praia Grande	2	1	1	1	1
Suzano	1	0	0	0	1
Taboão da Serra	1	0	0	0	1
Sumaré	2	0	1	0	1
Barueri	0	2	1	1	1
Embu das Artes	0	0	0	0	2
S. Carlos	0	0	0	1	0
Marília	1	1	5	1	0
Americana	1	0	0	0	2
Jacareí	0	0	1	0	0
Indaiatuba	0	0	1	0	0
Araraquara	0	0	0	1	1
Cotia	1	2	1	0	0
Pres. Prudente	1	2	1	0	0
Hortolândia	0	1	0	0	0
Itapevi	0	0	0	1	1

As covariáveis selecionadas para o estudo foram variáveis sócio econômicas e de saúde que possam ter alguma associação com o aumento ou diminuição da taxa de Tb nos municípios

estudados e são elas: taxa de mortalidade (por mil habitantes), salário médio (em salários mínimos), PIB per Capita (PIB real¹¹), unidades prisionais, % de mães adolescentes (com menos de 18 anos), número de médicos registrados no CRM/SP por mil habitantes, quantidade de leitos SUS por mil habitantes, IDHM 2010, grau de urbanização (em percentual), número de óbito por AIDS, número de casos de AIDS, percentual de consumidores energia elétrica, número de homicídios dolosos por mil habitantes, índice de envelhecimento (em percentual), participação (em percentual) da agropecuária no PIB e participação (em percentual) da indústria no PIB.

8.2 USO DE MODELOS DE REGRESSÃO LINEAR COM ERROS NORMAIS CONSIDERANDO AS COVARIÁVEIS MÊS E ANO

8.2.1 Modelo de regressão sob o enfoque clássico para as contagens de Tb

A seguir introduzimos uma análise de regressão linear múltipla para verificar o efeito do tempo (ano e mês) na variável resposta de interesse “quantidade de casos novos de tuberculose” na escala logarítmica para os 40 municípios agrupados (isto é, o total para todos os municípios observados nos 60 meses de acompanhamento).

As covariáveis estudadas são consideradas, para a modelagem clássica, como “significantes se o valor $p < 0.05$ ”, isto significa que se o valor-p (nível de significância observado) for menor do que 5%, a hipótese nula de que o coeficiente seja igual à zero deve ser rejeitada. As tabelas resumizando os resultados das regressões lineares que serão apresentadas a seguir conterão os valores dos coeficientes relacionados aos parâmetros estudados (Coef), o erro padrão da estimativa (SE Coef), o valor da estatística t (T) e o valor p.

Assim consideramos inicialmente o seguinte modelo de regressão na resposta $y_i = \log$ (contagem Tb):

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (7)$$

onde X_1 e X_2 denotam respectivamente meses que foram assumidos como variáveis qualitativas contínuas não categóricas e anos e ε_i são erros aleatórios supostos como independentes, com uma distribuição normal com média zero, e variância constante σ^2 . O índice i denota tempo de observação (os 60 meses de acompanhamento) (aqui $i = 1, 2, \dots, 60$).

Estimadores de mínimos quadrados dos coeficientes de regressão do modelo (1) são obtidos usando o software MINITAB® versão 17.

Assumindo o modelo de regressão (7), temos o seguinte modelo ajustado, obtido usando o software MINITAB®, versão 17:

$$\log(\text{total notif}) = 6,82 + 0,00369 \text{ meses} - 0,0334 \text{ anos} \quad (8)$$

Na Tabela 3, temos os sumários das inferências (estimadores de mínimos quadrados e valor-p para testar se os parâmetros de regressão são iguais a zero) obtidas para esse modelo.

Tabela 3 - Estimadores de mínimos quadrados, estatística T de Student e valor-p

Variável	Coef	EP Coef	T	Valor p
Meses	0,0037	0,0026	1,41	0,163
Anos	-0,0334	0,0319	-1,05	0,300

Dos resultados na Tabela 3, observa-se que as covariáveis mês e ano não mostram efeitos significativos na resposta transformada total notificado (valor-p > 0,05).

A partir dos gráficos dos resíduos (ver Figura 8) verificamos que as suposições necessárias para a validade do modelo estatístico (normalidade dos resíduos, variância constante dos erros) são verificadas para o modelo assumindo a resposta $\log(\text{total notificado})$.

Figura 8 - Gráficos dos resíduos - modelo de regressão dado por (8)

Na Figura 9, temos os gráficos dos valores (totais notificados) e preditos pelo modelo.

Figura 9 - Gráficos dos valores observados e médias ajustadas - modelo de regressão (8)

Dessa forma, observa-se que apesar do gráfico de resíduos mostrar que as suposições do modelo são verificadas, os gráficos dos valores (totais notificados) e preditos pelo modelo dados na Figura 2 indicam que o modelo é pobre em termos de ajuste.

Esses resultados nos motivam a procurar melhores modelos.

8.2.2 Modelo de regressão sob o enfoque bayesiano para as contagens de Tb

Como alternativa à regressão linear clássica considerando a resposta transformada para a escala logarítmica dos dados de contagem, de modo a encontrar um modelo estatístico com melhor ajuste, utilizou-se o modelo de regressão de Poisson bayesiano considerando os dados de contagem. Para a análise Bayesiana, todas as tabelas referentes aos resultados dos modelos bayesianos contém a média estimada, o desvio padrão e o intervalo de credibilidade 95% e as covariáveis estudadas serão consideradas “significativas” se o intervalo de credibilidade para o parâmetro de regressão correspondente a cada covariável não contiver o valor zero. Para todos os modelos estatísticos bayesianos considerou-se distribuições a priori aproximadamente não-informativas. O uso de métodos bayesianos tem se tornado muito popular na análise de dados epidemiológicos (ver por exemplo, Achcar et al, 2008; 2011).

Seja Y_i uma variável aleatória com uma distribuição de Poisson dada por,

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (9)$$

onde $y_i = 0, 1, 2, \dots$ denota a contagem de Tb notificados no tempo de observação (os 60 meses de acompanhamento) para os dados totais de Tb para os 40 municípios agrupados. Observar que a média e a variância da distribuição de Poisson (9) são iguais à λ_i .

Para relacionar o parâmetro λ_i com as covariáveis X_1 e X_2 definidas em (7), considerou-se inicialmente o seguinte modelo de regressão:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad (10)$$

A formulação (10) garante que λ_i seja positivo, para $i = 1, 2, \dots, n$ onde $n = 60$ (número de meses de acompanhamento para os dados totais de Tb para os 40 municípios agrupados.). Esse modelo é denotado como “modelo 1”.

Assumindo o modelo definido acima, a função de verossimilhança para o vetor θ de parâmetros associados a cada modelo é dada por,

$$L(\theta) = \prod_{i=1}^n f(\text{dados}/\theta) \quad (11)$$

onde $\theta = (\beta_0, \beta_1, \beta_2)$ e $f(\text{dados}/\theta)$ é a função de probabilidade de Poisson (9) para os dados.

Para a análise Bayesiana, as seguintes distribuições a priori foram assumidas para os parâmetros de regressão: $\beta_0 \sim N(7; 10)$, $\beta_r \sim N(0; 10)$, $r = 1, 2$, onde $N(a, b^2)$ denota uma distribuição normal com média igual ao valor a e variância igual ao valor b^2 . Observar que foi utilizada alguma informação na escolha da média da distribuição a priori para o parâmetro β_0 (igual ao valor 7) baseado na informação da análise clássica (estimador de mínimos quadrados igual à 6,82) feita anteriormente. Da mesma forma foram assumidas variâncias grandes nas distribuições a priori para os parâmetros de regressão (distribuições a priori aproximadamente não-informativas).

Combinando-se a distribuição a priori conjunta para θ (um produto de distribuições normais) com a função de verossimilhança $L(\theta)$, dada em (11), determina-se, a partir da fórmula de Bayes, a distribuição a posteriori para θ (BOX; TIAO, 1973).

Os sumários *a posteriori* de interesse foram obtidos usando métodos de Monte Carlo em Cadeias de Markov (GELFAND; SMITH, 1990; CHIB; GREENBERG, 1995) e o software *OpenBugs* (SPIEGELHALTER et al., 2003), que só requer a especificação da distribuição para os dados e as distribuições *a priori* para os parâmetros.

Assumindo a regressão de Poisson definidas por (9) e (10), e usando o software *OpenBugs* com uma amostra simulada de aquecimento (“*burn-in-sample*”) de tamanho 1.000, descartada para eliminar o efeito dos valores iniciais usados no algoritmo Gibbs Sampling, e simulando outras 10.000 amostras escolhidas de 10 em 10 para se ter amostras aproximadamente não-correlacionadas, foi obtida uma amostra final de tamanho 1.000 de valores gerados para β_0 e β_r , $r = 1, 2$.

Os sumários obtidos (média *a posteriori*, desvio-padrão *a posteriori* e intervalos de credibilidade com probabilidade igual à 0,95) são dados na Tabela 4. A convergência do algoritmo foi monitorada usando métodos gráficos (PAULINO; TURKMAN; MURTEIRA, 2003, GAMERMAN, 1997) e obtidas diretamente do software *OpenBugs*.

Tabela 4 - Sumários a posteriori (modelo 1)

Variável	Média	Desvio Padrão	Intervalo de Credibilidade 95%
Intercepto (β_0)	6,826	0,01208	(6,800; 6,849)
Meses (β_1)	0,0038	0,0012	(0,0013; 0,0061)
Anos (β_2)	-0,0344	0,0150	(-0,0637; -0,0040)

A partir dos resultados da Tabela 2 observa-se que as covariáveis meses e anos possuem significância estatística dado que os intervalos de credibilidade não incluem o valor zero, isto é, o modelo de regressão de Poisson sob o enfoque Bayesiano é mais sensível para detectar as covariáveis importantes na variabilidade do coeficiente de incidência de Tb. Também observa-se que o estimador bayesiano do coeficiente de regressão associado à covariável meses tem sinal positivo o que indica que nos últimos meses de cada ano há um aumento da incidência de Tb e o estimador associado à covariável anos apresenta sinal negativo o que indica que nos últimos anos de acompanhamento há um declínio na incidência de Tb.

Na Figura 10, são apresentados os gráficos dos valores (totais notificados) e preditos pelo modelo 1. A partir da Figura 10, observa-se resultados similares em termos de ajuste aos obtidos usando o mesmo modelo de regressão sob o enfoque clássico.

Figura 10 - Gráficos dos valores observados e médias ajustadas - “modelo 1”

Um segundo modelo de regressão Poisson com efeitos defasados também foi assumido considerando defasagens de até 6 meses, ou seja, se há efeito do mês ou meses anteriores anteriores na variável resposta “quantidade de casos novos de tuberculose” e houve significado estatístico para as seguintes variáveis: ano, mês, efeito da contagem de Tb notificados do mês anterior, efei-

to dos 2 meses anteriores, efeito dos 3 meses anteriores, efeito dos 4 meses anteriores, efeito dos 5 meses anteriores e efeito dos 6 meses anteriores. Esse modelo é definido por (3) e o parâmetro da distribuição de Poisson (3) é dado pelo modelo de regressão,

$$\lambda_i = \exp(\beta_0 + \beta_1 * \text{meses}_i + \beta_2 * \text{anos}_i + \beta_3 * \text{total.notificado}_{i-1} + \beta_4 * \text{total.notificado}_{i-2} + \beta_5 * \text{total.notificado}_{i-3} + \beta_6 * \text{total.notificado}_{i-4} + \beta_7 * \text{total.notificado}_{i-5} + \beta_8 * \text{total.notificado}_{i-6}) \quad (12)$$

onde $i=1,2,\dots,60$.

O modelo definido por (9) e (12) é denotado por “modelo 2”.

Assumindo a mesma distribuição a priori $N(7; 10)$ para β_0 e distribuições a priori normais com médias iguais a zero aproximadamente não-informativas para os outros parâmetros de regressão e usando o mesmo esquema de simulação que foi usado para o “modelo 1”, temos na Tabela 5, os sumários a posteriori de interesse.

Tabela 5 - Sumários a posteriori (modelo 2)

Variável	Média	Desvio Padrão	Intervalo de Credibilidade 95%
Intercepto (β_0)	6,824	0,0333	(6,7600; 6,8800)
Meses (β_1)	0,0044	0,0014	(0,0018 ; 0,0072)
Anos (β_2)	-0,0390	0,0159	(-0,0711; -0,0084)
Efeito do mês anterior (β_3)	0,0006	0,0414	(-0,0822; 0,0001)
Ef de 2 meses anteriores(β_4)	0,1647	0,0382	(0,0886; 0,2378)
Ef de 3 meses anteriores(β_5)	-0,0945	0,0385	(-0,1679; -0,0182)
Ef de 4 meses anteriores(β_6)	-0,1793	0,0399	(-0,2583; -0,0994)
Ef de 5 meses anteriores(β_7)	0,0878	0,0426	(0,0043; 0,1712)
Ef de 6 meses anteriores(β_8)	0,0149	0,0343	(-0,0534; 0,0789)

A partir dos resultados na Tabela 5, observa-se que as covariáveis meses e anos dia são significativas (zero não está incluído no intervalo de credibilidade 95% para β_1 e β_2); da mesma forma os efeitos defasados relacionados com os parâmetros de regressão β_4 (total notificado nos 2 meses anteriores), β_5 (total notificado nos 3 meses anteriores), β_6 (total notificado nos 4 meses anteriores), e β_7 (total notificado nos 5 meses anteriores) também são significativos, pois o valor zero não está incluído nos intervalos de credibilidade 95% correspondentes.

Comparando as estimativas dos 3 modelos ajustados (de regressão linear com covariáveis anos e meses sob o enfoque clássico), regressão de Poisson bayesiano (“modelo 1”) e regressão de Poisson com efeitos defasados (“modelo 2”), conclui-se que o “modelo 2” sob o enfoque Bayesiano possui o melhor ajuste aos dados reais (ver Figura 11), pois acompanha melhor a curva os dados reais.

Figura 11 - Gráficos dos valores observados e médias ajustadas - modelo de regressão, modelo 1 e modelo 2

8.2.3 Modelo de regressão linear sob o enfoque bayesiano para as taxas de Tb

Nessa seção foi ajustado um modelo de regressão linear Bayesiano com erros normais para as taxas notificadas de Tb para os dados de cada um dos 40 municípios durante o período de acompanhamento de 60 meses, o que totaliza uma amostra de $n = 2400$ observações com todas as covariáveis de interesses sem a inclusão de efeitos defasados. Esse modelo é denotado por “modelo 3”. O modelo é definido por,

$$y_i = tx \text{ total notif}_i = \beta_0 + \beta_1 * meses_i + \beta_2 * anos_i + \beta_3 * \text{salário.médio}_i + \beta_4 * \text{PIB.médio}_i + \beta_5 * \text{número.médicos}_i + \beta_6 * \text{número.leitos}_i + \beta_7 * \text{IDHM}_i + \beta_8 * \text{grau.urbanização}_i + \beta_9 * \text{taxa.AIDS}_i + \beta_{10} * \text{\%energia.elétrica}_i + \beta_{11} * \text{número.homicídios}_i + \beta_{12} * \text{indústria.PIB}_i + \beta_{13} * \text{média.óbitos}_i + \varepsilon_i. \quad (13)$$

Assumindo uma distribuição a priori $N(0; 10)$ para β_0 e distribuições a priori normais $N(0; 1)$ para os outros parâmetros de regressão e usando o mesmo esquema de simulação que foi usado para os modelos 1 e 2, temos na Tabela 6, os sumários a posteriori de interesse.

A partir dos resultados na Tabela 6, observa-se que as covariáveis meses, anos e número de médicos não possuem significância estatística (zero está incluído nos intervalos de credibilidade 95% para β_1 (meses), β_2 (anos) e β_5 (número de médicos)); as demais covariáveis mostraram significância estatística pois o valor zero não está incluído nos intervalos de credibilidade 95% para os parâmetros de regressão correspondentes. Alguns resultados são bem interessantes, observa-se que quanto menor o salário médio maior, menor o coeficiente de Tb, quanto maior o PIB

maior o coeficiente de TB, quanto menor a quantidade de leitos SUS por mil habitantes maior o coeficiente de Tb. Na Figura 12 temos o gráfico dos valores observados e ajustados pelo “modelo 3”. Observa-se bom ajuste.

Figura 12 - Gráficos dos valores observados e valores estimados pelo “modelo 3”

Tabela 6 - Sumários a posteriori (modelo 3)

Variável	Média	Desvio Padrão	Intervalo de Credibilidade 95%
Intercepto (β_0)	-9,613	1,305	(-11,3200;-7,1500)
Meses (β_1)	0,0042	0,0075	(-0,0119;0,0174)
Anos (β_2)	-0,0036	0,0910	(-0,1606;0,1913)
Salário médio (β_3)	-0,2862	0,0722	(-0,4385;-0,1472)
PIB médio (β_4)	0,0070	0,0025	(0,0023;0,0118)
# Médicos (β_5)	0,0718	0,0423	(-0,0056;0,1595)
# Leitos (β_6)	-0,3427	0,0483	(-0,4333;-0,2445)
IDHM (β_7)	-2,5640	0,8342	(-4,0720;-0,8903)
Grau de urbanização (β_8)	0,1316	0,0157	(0,1020;0,1515)
# AIDS (β_9)	0,0011	0,1318	(0,8096;0,0013)
% Energia elétrica (β_{10})	0,0662	0,0056	(0,0547;0,0767)
# Homicídios (β_{11})	0,2351	0,0274	(0,1846;0,2910)
Particip. Indústria no PIB (β_{12})	-4,0850	0,3676	(-4,8280;-3,3860)
Média de óbitos (β_{13})	0,0941	0,0331	(0,0206;0,1520)

8.2.4 Modelo na taxa notificada de casos de tuberculose na presença de efeitos aleatórios

Nesta seção, introduzimos um modelo de regressão para as taxas notificadas de casos de tuberculose de uma forma um pouco diferente da considerada na seção anterior considerando os dados dos 40 municípios e 60 meses (2400 observações) assumindo a presença de um efeito aleatório que captura a possível dependência entre as medidas para cada município.

Seja Y_{ij} a taxa notificada de casos de tuberculose no município i ($i = 1, 2, \dots, 40$) e período j ($j = 1, 2, \dots, 60$), o modelo proposto para explicar esta taxa segundo algumas variáveis explicativas (covariáveis) é dado por:

$$Y_{ij} = \beta_0 + \gamma_j \sum_{k=1}^5 \beta_k X_{ijk} + w_i + e_{ij} \quad (14)$$

em que:

- β_0 é a média geral da taxa notificada de casos de tuberculose;
- γ_j é um efeito aleatório referente as variações de período, normalmente distribuído com média 0 e variância σ_γ^2 ;
- β_1 é um efeito fixo referente a covariável X_{ij1} (óbitos por 1000 habitantes);
- β_2 é um efeito fixo referente a covariável X_{ij2} (salário médio entre os anos de 2009 a 2012);
- β_3 é um efeito fixo referente a covariável X_{ij3} (PIB médio entre os anos de 2009 a 2012);
- β_4 é um efeito fixo referente a covariável X_{ij4} (número médio de médicos por 100 habitantes entre os anos de 2009 a 2013);
- β_5 é um efeito fixo referente a covariável X_{ij5} (grau de urbanização médio entre os anos de 2009 a 2013);
- w_i é um efeito aleatório que captura a possível dependência entre os municípios, normalmente distribuído com média 0 e variância σ_w^2 ;
- ε_{ij} é o erro deste modelo, suposto aleatório, independente e identicamente distribuído com uma distribuição normal e média 0 e variância constante σ^2 .

Assume-se também que γ_j , w_i e ε_{ij} são independentes. Denotar esse modelo por “modelo 4”.

Na Figura 13 são apresentadas medidas sumários sendo, (a) um gráfico de linhas que representa a taxa notificada de casos de tuberculose de janeiro/2009 a dezembro/2013 e (b-f) gráficos de dispersão das covariáveis em estudo segundo a taxa notificada. Após a observação da Figura 13

(e), sugere-se a introdução de um efeito fixo no modelo (14) que represente a relação quadrática observada na variável X_{ij4} (número médio de médicos por 100 habitantes entre os anos de 2009 a 2013). Dessa forma, assume-se um novo modelo (denotado como “modelo 5”), definido por,

$$Y_{ij} = \beta_o + \tau_j + \sum_{k=1}^5 \beta_k X_{ijk} + \beta_{42} X_{ij4}^2 + \omega_i + e_{ij} \quad (15)$$

e os resultados são apresentados na Tabela 7.

Na tabela 7 observa-se evidências, de significância estatística, para a variável número médio de médicos por 100 habitantes entre os anos de 2009 a 2013. Como o efeito fixo quadrático estimado ($\beta_{42} = 0,1718$) é positivo, isto implica que nos limites mínimos e máximos do número médio de médicos no período de estudo as taxas notificadas dos casos de tuberculose foram maiores, o que foi sugerido na Figura 13 (e) e capturado no modelo (13). As demais variáveis não mostraram significado estatístico, porém são importantes para o bom ajuste do modelo apresentado na Figura 14.

Tabela 7 - Sumários a posteriori (modelo 5)

Variável	Média (Desvio Padrão)	Intervalo de Credibilidade 95%
$\hat{\beta}_0$	1,8552 (2,8993)	(-3,5958; 7,4122)
$\hat{\beta}_1$	0,0445 (0,0391)	(-0,0294; 0,1202)
$\hat{\beta}_2$	-0,5642 (0,3194)	(-1,2019; 0,0587)
$\hat{\beta}_3$	0,0154 (0,0138)	(-0,0130; 0,0413)
$\hat{\beta}_4$	-1,0079 (0,3830)	(-1,7829; 0,2582)
$\hat{\beta}_{42}$	0,1718 (0,0635)	(0,0453; 0,3018)
$\hat{\beta}_5$	0,0315 (0,0282)	(-0,0234; 0,0843)
$\hat{\sigma}_\gamma^2$	0,2931 (0,0307)	(0,2354; 0,03576)
$\hat{\sigma}_\omega^2$	1,2718 (0,1564)	(1,0117; 1,6024)
$\hat{\sigma}^2$	1,0423 (0,0157)	(1,0130; 1,0756)

Figura 13 - a) Gráfico de linhas para a Taxa notificadas segundo o tempo. b-f) Gráficos de dispersão para as covariáveis segundo a taxa notificada

Figura 14 - Gráficos dos valores observados e médias ajustadas - “modelo de regressão, modelo 4”

8.2.5 Modelo de Contagem notificada de casos de tuberculose na presença de efeitos aleatórios

Com as mesmas considerações introduzidas na apresentação do modelo para taxas de Tb definido na presença de efeitos aleatórios (modelos 4 e 5), vamos agora introduzir nesta seção, um modelo para a contagem de casos de tuberculose na presença de efeitos aleatórios.

Seja Y_{ij} a contagem notificada de casos de tuberculose no município i ($i = 1; 2; \dots; 40$) e período j ($j = 1; 2; \dots; 60$) e assume-se que Y_{ij} segue aproximadamente distribuição de Poisson, por se tratar de contagem. Logo,

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij})$$

$$\text{com } E(Y_{ij}) = V_{ar}(Y_{ij}) = \lambda_{ij} \text{ e} \quad (16)$$

$$\lambda_{ij} = \exp(\beta_0 + \tau_j + \sum_{k=1}^5 \beta_k X_{ijk} + \beta_{42} X_{ij4}^2 + \omega_i)$$

em que:

- β_0 é o intercepto do modelo;
- γ_j é um efeito fixo referente as variações de período;
- β_1 é um efeito fixo referente a covariável X_{ij1} (óbitos por 1000 habitantes);
- β_2 é um efeito fixo referente a covariável X_{ij2} (salário médio entre os anos de 2009 a 2012);
- β_3 é um efeito fixo referente a covariável X_{ij3} (PIB médio entre os anos de 2009 a 2012);
- β_4 é um efeito fixo referente a covariável X_{ij4} (número médio de médicos por 100 habitantes entre os anos de 2009 a 2013);
- β_5 é um efeito fixo referente a covariável X_{ij5} (grau de urbanização médio entre os anos de 2009 a 2013);
- ω_i é um efeito aleatório que captura a possível dependência entre os municípios, normalmente distribuído com média 0 e variância σ_ω^2 .

Denotar esse modelo por “modelo 6”.

Adaptando o modelo (15), ou seja, introduzindo um efeito fixo neste modelo que represente a relação quadrática da X_{ij4} (número médio de médicos por 100 habitantes entre os anos de 2009 a 2013) tem-se:

$$Y_{ij} = \beta_0 + \tau_j + \sum_{k=1}^5 \beta_k X_{ijk} + \beta_{42} X_{ij4}^2 + \omega_i + e_{ij} \quad (17)$$

e os resultados são apresentados na Tabela 6.

Na tabela 8, observa-se que as covariáveis não apresentaram significância estatística, porém são importantes para o excelente ajuste do modelo apresentado na Figura 15.

Tabela 8 - Sumários a posteriori (modelo 6)

Variável	Média (Desvio Padrão)	Intervalo de Credibilidade 95%
$\hat{\beta}_0$	2,4725 (1,0034)	(0,4568; 4,4229)
$\hat{\beta}_1$	0,01504 (0,0151)	(-0,0137; 0,04455)
$\hat{\beta}_2$	0,0376 (0,2233)	(-0,4077; 0,4963)
$\hat{\beta}_3$	0,0027 (0,0104)	(-0,0183; 0,0235)
$\hat{\beta}_4$	-0,1865 (0,2626)	(-0,7001; 0,3256)
$\hat{\beta}_{12}$	0,05445 (0,0445)	(0,0341; 0,1411)
$\hat{\beta}_5$	-0,0049 (0,0118)	(-0,0283; 0,0171)
$\hat{\alpha}_2$	0,1989 (0,0186)	(0,1657; 0,2402)
$\hat{\alpha}_6$	0,9273 (0,1217)	(0,7299; 1,2009)

Figura 15 - Gráficos dos valores observados e médias ajustadas- “modelo de regressão, modelo 5”

9 CONCLUSÕES E PERSPECTIVAS FUTURAS

Um dos grandes objetivos do uso de métodos quantitativos em epidemiologia é relacionado com o uso de modelos estatísticos que consigam identificar diferentes fatores que são importantes na variabilidade da incidência de qualquer doença e também conseguir bons modelos de previsão que podem auxiliar os administradores da área da saúde na locação de recursos, leitos hospitalares e pessoal (médicos, enfermeiros, etc). Sob esse foco, consideramos nessa dissertação de mestrado, um estudo sobre uma doença que ainda pode ser de grande preocupação para

as autoridades da área de saúde em todo o mundo, especialmente em países do terceiro mundo, apesar dos grandes avanços observados nas últimas décadas: a tuberculose. Assim, considerando dados de contagem de novos casos de Tb (transformados para a escala logarítmica) ou para os coeficientes de incidência de Tb para 40 grandes municípios do estado de São Paulo no período entre 2009 e 2013 exploramos o uso de diferentes modelos de regressão linear múltipla com erros normais com variância constante sob um enfoque clássico ou Bayesiano e também modelos de Poisson sob um enfoque Bayesiano diretamente para as contagens de novos casos Tb em cada um dos 40 municípios mais populosos do estado de São Paulo (mais de 200.000 habitantes). Nesses modelos foram consideradas várias covariáveis que podem estar correlacionadas com a incidência de novos casos de Tb variáveis temporais (mês e ano), variáveis sócio-econômicas e saúde relacionadas a cada município, e também efeitos defasados.

Alguns pontos importantes obtidos com o estudo:

- A partir de gráficos de séries temporais e de controle estatístico foi observado que considerando os dados mensais (60 meses de acompanhamento) agrupados para os 40 municípios foi observado que em alguns meses houve a presença de alguns pontos discordantes (contagens de Tb acima do esperado) e também a presença de sazonalidade. Isso também foi observado considerando os dados individuais dos 40 municípios (gráficos no Apêndice da dissertação).
- Inicialmente foi considerado uma análise estatística sob um enfoque clássico para modelos de regressão linear múltiplo considerando os dados de contagem agrupados de todos os 40 municípios e observados nos 60 meses de acompanhamento (01/01/2009 a 31/12/2013) e tendo apenas duas covariáveis de interesse (mês e ano) para verificar possíveis efeitos temporais e/ou sazonais para as contagens de novos casos de Tb. Esse modelo com 60 observações foi considerado com os dados transformados para a escala logarítmica para satisfazer algumas suposições necessárias para as inferências de interesse (normalidade e variância constante dos resíduos) considerando um modelo de regressão linear múltiplo. Além desse modelo, também foi considerado um modelo de regressão de Poisson para os dados de contagem na escala original com as covariáveis mês e ano sob um enfoque Bayesiano.

- As inferências obtidas para o modelo de regressão linear com os dados agrupados (transformados para a escala logarítmica) dos 40 municípios durante o período de observação de 60 meses não mostram efeitos significativos de mês e ano na resposta de interesse (incidência de novos casos de Tb). O modelo de regressão de Poisson sob um enfoque Bayesiano foi muito mais sensível para detectar os efeitos temporais (mês e ano) sob um enfoque Bayesiano usando alguma informação a priori para o intercepto, mas com grande variância (distribuição a priori aproximadamente não-informativa). Dessa forma foi observado que nos meses do final de cada ano há um pequeno aumento de novos casos de Tb; da mesma forma foi observado que há um declínio nas contagens de novos casos de Tb para os últimos anos de acompanhamento. Esses resultados podem ser de grande interesse para a área de epidemiologia.
- Um terceiro modelo considerado para os dados de contagem foi a inclusão de efeitos defasados no modelo de regressão de Poisson com as covariáveis mês e ano sob um enfoque Bayesiano. Nesse modelo também foi observado que as covariáveis mês e ano apresentam efeitos significativos na incidência de novos casos de Tb. Além disso houve um melhor ajuste do modelo aos dados.
- Na seção 8.2.3 (Modelo de regressão linear sob o enfoque bayesiano para as taxas de Tb) foi ajustado um modelo (modelo 3) de regressão linear Bayesiano com erros normais para as taxas notificadas de Tb para os dados de cada um dos 40 municípios durante o período de acompanhamento de 60 meses, o que totaliza uma amostra de $n = 2400$ observações com todas as covariáveis de interesses sem a inclusão de efeitos defasados, mas considerando várias covariáveis de interesse associadas a cada município/mês para a epidemiologia da tuberculose (salário médio, PIB médio, número de médicos, número de leitos, IDHM, grau de urbanização, taxa de AIDS, %energia de residências com energia elétrica, taxa de homicídios, PIB associado à indústria, media de óbitos) sob um enfoque Bayesiano foi observado alguns resultados que podem ser de grande interesse para epidemiologistas: todas essas covariáveis, exceto as covariáveis, mês, ano e número de médicos) mostraram algum efeito significativo pois os intervalos de credibilidade

para os parâmetros de regressão associados a cada covariável não incluem o valor zero. Esse modelo levou a um ajuste razoável (ver Figura 12).

- É interessante observar, que temos contagens para cada município durante 60 meses (de 01/01/2009 a 31/12/2013). Dessa forma é possível a existência de uma dependência entre as contagens para cada município (dados longitudinais). Assim, com a mesma formulação do modelo 3 (2400 observações associadas aos 40 municípios e 60 meses), foi considerado um modelo na presença de um efeito aleatório que captura essa possível dependência. Na seção 8.2.4, foram considerados dois modelos de regressão para as taxas de caos notificados de Tb sob um enfoque Bayesiano. Assumindo em especial, o modelo 5, foi observado um excelente ajuste (bem melhor do que o ajuste obtido para o modelo 3), além da detecção de uma covariável significativa: número médio de médicos. As demais covariáveis consideradas (óbitos por 100 habitantes; salário médio, PIB médio) não mostraram significado estatístico, mas são importantes no ajuste do modelo.
- Na seção 8.2.5 (Modelo de Contagem notificada de casos de tuberculose na presença de efeitos aleatórios) observa-se um excelente ajuste do modelo, apesar de só a covariável (efeito quadrático do número médio de médicos) mostrar significância (zero não incluído no intervalo de credibilidade) do parâmetro de regressão correspondente.

O uso de diferentes modelos estatísticos para dados epidemiológicos, como o caso especial de notificação de novos casos de tuberculose, pode ser de grande interesse para a área de epidemiologia, como foi observado nesse estudo.

É importante salientar que os modelos estatísticos propostos nessa dissertação de mestrado, assumindo dados de contagem de notificações de Tb para outras regiões é possível o uso dos modelos propostos nessa dissertação de mestrado. O uso de modelos de regressão de Poisson sob um enfoque Bayesiano com métodos de simulação de amostras da distribuição a posteriori conjunta de interesse, pode ser uma alternativa promissora para analisar dados de contagem de novos casos de diferentes doenças com diferentes estruturas de dependência e covariáveis contínuas, binárias ou categóricas, especialmente usando softwares livres disponíveis na Internet, como o software OpenBugs que não necessita grande esforço computacional na obtenção das estimativas de Monte Carlo para os parâmetros de interesse.

Para um trabalho futuro, também poderia ser usados modelos de regressão de Poisson com estruturas (variáveis latentes) capturando efeitos de cidades geograficamente próximas ou similares de alguma forma). Isso pode levar cada vez mais a melhores modelos epidemiológicos.

REFERÊNCIAS

ACHCAR, J. A. et al. A statistical model investigating the prevalence of tuberculosis in New York City using counting processes with two change-points. **Epidemiology and Infection**, v. 136, p. 1599-1605, 2008.

ACHCAR, J. A. et al. Uso de modelos Poisson espaço-temporal de regressão para a Floresta Amazônica Brasileira: contagem de dados de malária. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, n. 6, p. 749-754, nov-dez, 2011.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**, New York: Addison-Wesley, 1973.

BRENNAN, M. J.; THOLE J. Tuberculosis Vaccines: A Strategic Blueprint for the Next Decade. **Tuberculosis** (Edinburgh), v. 92, p. S6-S13, 2012. Supplement 1.

CADERNO DE ANÁLISE DOS INDICADORES DO PROGRAMA NACIONAL DE CONTROLE DA TUBERCULOSE - INDICADORES EPIDEMIOLÓGICOS E OPERACIONAIS DA TUBERCULOSE. Disponível em: <<http://dtr2004.saude.gov.br/sinanweb/novo/Documentos/SinanNet/analise/cadernoanalisetuberculose.pdf>>. Acesso em: 21 mai. 2014.

CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. **The American Statistician**, v. 49, p. 327-335, 1995.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**, Wiley series in probability and mathematical statistics, 1981.

DUCATI, R. G.; RUFFINO NETTO, A.; BASSO, L.; SANTOS, D. S. The resumption of consumption – A review in tuberculosis. **Mem. Inst. Oswaldo Cruz**, v. 101, n. 7, p. 697-714, Rio de Janeiro, November, 2006.

GAMERMAN, D. **Markov Chain Monte Carlo: stochastic simulation for Bayesian inference**. London: Chapman and Hall, 1997.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal distributions, **Journal of the American Statistical Association**, v. 85, n. 410, p. 398- 409, 1990.

GUIMARÃES, R. M. et al. Tuberculose, HIV e pobreza: tendência temporal no Brasil, Américas e mundo. **J. Bras. Pneumol**, v. 38, n. 4, p. 511-517, 2012.

HINO, P. et al. **Padrões espaciais da Tuberculose e sua associação à condição de vida no município de Ribeirão Preto** (Spatial patterns of Tuberculosis and its association with living conditions in the city of Ribeirão Preto in the State of São Paulo), 2011.

HIJJAR, M. A. et al. Epidemiologia da tuberculose: importância no mundo, no Brasil e no Rio de Janeiro. **Pulmão RJ**, v. 14, n. 4, p. 310-4, 2005.

HIJJAR, M. A.; PROCÓPIO, M. J. Tuberculose - epidemiologia e controle no Brasil. **Revista Hospital Universitário Pedro Ernesto**. v. 5, n. 2, p. 15-23. 2006.

KANEKO, T.; COOPER, C.; MDLULI, K. Challenges and opportunities in developing novel drugs for TB. **Future Medicinal Chemistry**, v. 3, n. 11, p. 1373-1400, 2011.

KELSALL, J.; WAKEFIELD J. Modeling spatial variation in disease risk: a geostatistical approach. **Journal of the American Statistical Association**, v. 97, n. 459, p. 692-701. 2002.

KOUL, A. et al. The challenge of new drug discovery for tuberculosis. **Nature**, v. 469, edição 7331, p. 483-490, 2011. <http://dx.doi.org/10.1038/nature09657>

LEMOES, A. C. M. Co-infecção tuberculose/HIV. **J. Bras. Pneumol**, v. 34, n. 10, p. 753-755, 2018. <http://dx.doi.org/10.1590/S1806-37132008001000001>.

LIENHARDT, C. et al. Efficacy and safety of a 4-drug fixed-dose combination regimen compared with separate drugs for treatment of pulmonary tuberculosis: the Study C randomized controlled trial. **The Journal of the American Medical Association**, v. 305, n. 14, p. 1415-1423, 2011.

LILLY TB DRUG DISCOVERY INITIATIVE. Disponível em: <<http://www.idri.org/products/drugs/lilly-tb-drug-discovery-initiative/>>. Acesso em: 30 mai. 2014.

LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility, **Statistics and Computing**, v. 10 n. 4, p. 325-337, 2000.

MARSHALL, R. J. A review of methods for the statistical analysis of spatial patterns of disease. **Journal of the Royal Statistical Society**, series A, v. 154, n. 3, p. 421-441. 1991.

MA, Z. et al. Global tuberculosis drug development pipeline: the need and the reality. **The Lancet**, v. 375, n. 9731, p. 2100-2109, 2010.

MCNERNEY, R. et al. Tuberculosis diagnostics and biomarkers: needs, challenges, recent advances, and opportunities. **The Journal of Infectious Diseases**, v. 205, p. S147-58, 2012. Supplement 2.

MINISTÉRIO DA SAÚDE. **Tuberculose | Conheça os sintomas da doença que tem tratamento gratuito pelo SUS**. 2013a. Disponível em: <<http://www.blog.saude.gov.br/tuberculose-conheca-os-sintomas-da-doenca-que-tem-tratamento-gratuito-pelo-sus/>> Acesso em: 16 nov. 2013.

MINISTÉRIO DA SAÚDE. **Estimativas populacionais para os municípios brasileiros em 01.07.2013**. 2013b. Disponível em:<<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/poptsp.def>>. Acesso em: 09 Dez. 2014.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied statistics and probability for engineers**, third edition, New York: Wiley, 2011.

NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES (NIAID). Disponível em:<<http://www.niaid.nih.gov/topics/tuberculosis/understanding/history/pages/default.aspx>>. Acesso em: 16 nov. 2013.

NORONHA, E. F. Incidência de tuberculose e taxa de cura, Brasil, 2000 a 2004. **Rev. Saude Publica**, v. 41, p. 24-33, 2007, Supplement 1.

NOVARTIS INSTITUTE FOR TROPICAL DISEASES. Disponível em:<<https://www.nibr.com/our-research/institutes/novartis-institute-tropical-diseases>>. Acesso em: 24 nov. 2013

NOVARTIS. **Novartis in the fight against tuberculosis**. Disponível em:<<http://www.novartis.com/newsroom/feature-stories/2009/03/world-tuberculosis-day.shtml>> Acesso em: 16 abr. 2013.

PAGANO, M.; GAUVREAU, K., **Princípios de Bioestatística**, São Paulo: Thomson Pioneira, 2004.

PAIXÃO, L. M. M.; GONTIJO, E. D. Perfil de casos de tuberculose notificados e fatores associados ao abandono, Belo Horizonte, MG. **Revista de Saúde Pública**, v. 41, n. 2, p. 205-213, 2007.

PASSOS, A. D. C.; FRANCO, L. J. (Org.) **Fundamentos da Epidemiologia**. Barueri, SP: Manole, 2005.

PAULINO, C. D.; TURKMAN, M.; MURTEIRA, B. **Estatística Bayesiana**, Lisboa: Fundação Calouste Gulbenkian, 2003.

PEREIRA, J. C. M. **A explicação sociológica na medicina social**. São Paulo: Editora UNESP, 2005.

PINHEIRO, R. S. et al. Determinantes sociais e autorrelato de tuberculose nas regiões metropolitanas conforme a Pesquisa Nacional por Amostra de Domicílios, Brasil. **Revista Panamericana de Salud Publica**, v. 34, p. 446-51, 2013.

RUFFINO NETTO, A. Controle da tuberculose no Brasil: dificuldades na implantação do programa. **J. Pneumologia**, São Paulo, v. 26, n. 4, p. 159-162, Aug. 2000.

_____. Tuberculose: a calamidade negligenciada. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 35, n. 1, p. 51-58, jan-fev, 2002.

SANDOZ TB. Disponível em:<<http://www.tbdots.com/index.shtml>> Acesso em: 16 out. 2013.

SANOFIAVENTIS, **Tuberculosis Research & Development**. Disponível em: <<http://www.predict-tb.eu/author/sanofi-aventis-research-and-development/>>. Acesso em: 16 abr. 2014.

SAN PEDRO, A.; OLIVEIRA, R. M. Tuberculose e indicadores socioeconômicos: revisão sistemática da literatura. **Revista Panamericana de Salud Publica**, v. 33, n. 4, p. 294-301, 2013.

SANTOS, C. B. et al. Utilização de um Sistema de Informação Geográfica para descrição dos casos de tuberculose. **Bol. Pneumol. Sanit.**, v. 12, n. 1, Rio de Janeiro, 2004

SANTOS, M. L. S. G. et al. Pobreza: caracterização socioeconômica da tuberculose. **Rev. Latino-am Enfermagem**, v. 15, n. especial, 2007.

SEBER, G. A. F.; LEE, A. J. **Linear regression analysis**, second edition, Wiley series in probability and mathematical statistics, 2003.

SILVA, L. J. Tuberculose. **A epidemia através dos anos: o ledão engano de Koch**. Edição 27 - Abril/Maio/Junho de 2004. Disponível em: <<http://www.cremesp.com.br/?siteAcao=Revista&id=136>>. Acesso em: 16 Set. 2013.

SOLAR, O.; IRWIN, A. **Rumo a um modelo conceitual para análise e ação sobre os determinantes sociais de saúde**: ensaio para apreciação da Comissão de Determinantes Sociais de Saúde (Rascunho). Genebra: OMS, 2005.

SPIEGELHALTER, D. J. et al. **Winbugs user manual**. Cambridge: MRC Biostatistics Unit, Cambridge, U.K., 2001.

_____. WinBugs Version 1.4: **User Manual**. Cambridge: Medical Research Council Biostatistics Unit; 2003. Disponível em: <<http://www.mrc-bsu.cam.ac.uk/bugs>>. Acesso em: 16 mar. 2015

TAG & I-BASE. **Pipeline report Rifapentine**. July 2012. Disponível em: <<http://www.pipeline-report.org/browse/tb-treatments/rifapentine>>. Acesso em: 16 abr. 2014.

TAMERIS, M. D. et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. **The Lancet**. v. 381, n. 9871, p. 1021-28, 2013.

UNITAID. **Tuberculosis Diagnostic Technology Landscape 2012**. 2012, Geneva: World Health Organization. Disponível em: <<http://www.unitaid.eu/resources-2/news/974-tuberculosis-diagnostic-landscape>>. Acesso em: 16 Set. 2014.

VENDRAMINI, S. H. F. et al. Tuberculose em município de porte médio do sudeste do Brasil: indicadores de morbidade e mortalidade, de 1985 a 2003. **J. Bras. Pneumol.**, v. 31, n. 3, p. 237-43, 2005.

VIEIRA, R. C. A. et al. Distribuição espacial dos casos novos de tuberculose em Vitória, Estado do Espírito Santo, no período entre 2000 e 2005. **Rev. Soc. Bras. Med. Trop.** [online]. 2008, v. 41, n. 1, p. 82-86. ISSN 0037-8682. <http://dx.doi.org/10.1590/S0037-86822008000100017>

WALLER, L. A., CARLIN, B. P.; XIA H.; GELFAND, A. E. Hierarchical spatio-temporal mapping of disease rates. **Journal of the American Statistical Association**, v. 92, n. 438, p. 607-617. 1997.

WEYER, K.; CARAI, S.; NUNN, P. Viewpoint of TB Diagnostics: What Does the World Really Need? **The Journal of Infectious Diseases**, v. 204, p. S1196-s1202, 2011. Supplement 4.

WORLD HEALTH ORGANIZATION. **Global Tuberculosis Report 2012**. Disponível em: <http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf>. Acesso em: 16 set. 2014.

WORLD HEALTH ORGANIZATION. **Infectious diseases** [Chapter 5]. In: Health in 2015: from MDGs (millennium development goals) to SDGs (sustainable development goals). Geneva, Switzerland: World Health Organization; 2015. Disponível em: <http://www.who.int/gho/publications/mdgs-sdgs/MDGs-SDGs2015_chapter5.pdf?ua=1>. Acesso em: 16 Set. 2014.

WORLD HEALTH ORGANIZATION. **Global Tuberculosis Report 2014**. WHO Library Cataloguing-in-Publication Data. WHO/HTM/TB/2014.08. Disponível em: <http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf?ua=1>. Acesso em: 16 Set. 2014.

WORLD HEALTH ORGANIZATION. **Treatment of tuberculosis: guidelines – 4th ed.** 2009. Disponível em: <http://whqlibdoc.who.int/publications/2010/9789241547833_eng.pdf>. Acesso em: 16 abr. 2014.

ZHANG, T. et al. Short-Course Chemotherapy with TMC207 and Rifapentine in a Murine Model of Latent Tuberculosis Infection. **Am J Respir Crit Care Med.**, v. 184, n. 6, p. 732-7, 2011. <http://dx.doi.org/10.1164/rccm.201103-0397OC>

APÊNDICE A - GRÁFICOS DE CONTROLE PARA CADA MUNICÍPIO

APÊNDICE B - ALGUMAS INFORMAÇÕES DE CADA MUNICÍPIO

Geral40 - Casamentos - registrados - lugar do registro (Em mil habitantes)

Geral40 - Óbitos - registrados - lugar do registro (Em mil habitantes)

Óbitos - registrados - lugar do registro (por Mil habitantes)

Geral40 - Salário médio (em salário mínimo)

Enfermeiros Registrados no COREN/SP (Coeficiente por mil habitantes)

IDH 2010 por Município

Número de agências bancárias por 100 Mil Habitantes

APÊNDICE C - ALGUNS PROGRAMAS OPENBUGS

MODELO 1 - USO DO OPENBUGS

```

model
{
  for( i in 1: n ) {
    total.notificado[i] ~ dpois(lambda[i])
    log(lambda[i]) <- betao+ beta1*meses[i] +beta2*anos[i]
  }
  betao ~ dnorm(7,0.1)
  beta1 ~ dnorm(0,0.1)
  beta2 ~ dnorm(0,0.1)
}

list(total.notificado=c(874,880,1064,956,898,797,850,901,905,943,938,869,900,841,954,938,
856,776,900,928,981,968,904,1094,919,897,1005,938,1013,840,874,1039,987,964,1001,947,967,
893,993,866,930,867,874,1011,932,1029,943,855,953,848,922,1005,952,966,891,1089,926,1013,
949,894),meses=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,
31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60),
anos=c(1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4,4,5,5,
5,5,5,5,5,5,5,5),n=60)

```

MODELO 2 – EFEITO DEFASADO (CONTAGEM MÊS ANTERIOR)

```

model
{
  total.notificado[1] ~ dpois(lambda[1])
  log(lambda[1]) <- betao+ beta1*meses[1] +beta2*anos[1]
  for( i in 2: n ) {
    total.notificado[i] ~ dpois(lambda[i])

```



```

log(lambda[5]) <- betao+ beta1*meses[5] +beta2*anos[5]+beta3*total.notificado[4] +be-
ta4*total.notificado[3] +beta5*total.notificado[2] +beta6*total.notificado[1]

total.notificado[6] ~ dpois(lambda[6])

log(lambda[6]) <- betao+ beta1*meses[6] +beta2*anos[6]+beta3*total.notificado[5]
+beta4*total.notificado[4] +beta5*total.notificado[3] +beta6*total.notificado[2] +beta7*total.
notificado[1]

for( i in 7: n ) {

total.notificado[i] ~ dpois(lambda[i])

log(lambda[i]) <- betao+ beta1*meses[i] +beta2*anos[i]+beta3*total.notificado[i-1] +be-
ta4*total.notificado[i-2] +beta5*total.notificado[i-3] +beta6*total.notificado[i-4] +beta7*total.
notificado[i-5] +beta8*total.notificado[i-6]

}

betao ~ dnorm(7,1)
beta1 ~ dnorm(0,1)
beta2 ~ dnorm(0,1)
beta3 ~ dnorm(0,1000)
beta4 ~ dnorm(0,10000)
beta5 ~ dnorm(0,100000)
beta6 ~ dnorm(0,100000)
beta7 ~ dnorm(0,100000)
beta8 ~ dnorm(0,100000)

}

list(total.notificado=c(874,880,1064,956,898,797,850,901,905,943,938,869,900,841,954,938,
856,776,900,928,981,968,904,1094,919,897,1005,938,1013,840,874,1039,987,964,1001,947,967,
893,993,866,930,867,874,1011,932,1029,943,855,953,848,922,1005,952,966,891,1089,926,1013,
949,894),meses=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,
31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60),

```


APÊNDICE D

Correlograma: ferramenta gráfica estatística utilizada para verificar se uma série temporal é aleatória ou não. Os correlogramas para este estudo foram construídos utilizando o software R (R CORE TEAM, 2014), usando a função ACF.

AUTOCORRELAÇÃO

Uma importante ferramenta para se identificar as propriedades de uma série temporal consiste de uma série de quantidades chamadas coeficientes de *autocorrelação amostral*. A ideia é similar ao coeficiente de correlação usual, i.e. para n pares de observações das variáveis x e y o coeficiente de correlação amostral é dado por

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Considerando $x_1, \dots, x_{n-1}$ e $x_2, \dots, x_n$ como duas variáveis o coeficiente de correlação entre elas é dado por

$$r^1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)^2 \sum_{t=1}^{n-1} (x_{t+1} - \bar{x}_2)^2}}$$

onde as médias amostrais são

$$\bar{x}_1 = \sum_{t=1}^{n-1} x_t / (n-1) \text{ e } \bar{x}_2 = \sum_{t=2}^n x_t / (n-1)$$

Como o coeficiente r_1 mede as correlações entre observações sucessivas ele é chamado de coeficiente de autocorrelação ou coeficiente de correlação serial.

Uma versão simplificada é dada por,

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(n-1) \sum_{t=1}^n (x_t - \bar{x})^2 / n}$$

Para observações defasadas para k períodos de tempo, temos:

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

O CORRELOGRAMA

Um gráfico com os k primeiros coeficientes de autocorrelação como função de k é chamado de correlograma e pode ser uma ferramenta poderosa para identificar características da série temporal.

Como ilustração, temos na figura abaixo o correlograma de 100 observações independentes (simuladas):

Interpretação (São Paulo): alguma autocorrelação positiva (maior do que 0,2) da incidência atual com a incidência de 12 meses anteriores. As demais autocorrelações (menores do que 0,2) são pequenas (a suposição de independência é razoável).

Interpretação (Guarulhos): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Campinas): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (São Bernardo): alguma autocorrelação negativa da incidência atual com a incidência de 7 meses anteriores. As demais autocorrelações (menores do que 0,2) são pequenas (a suposição de independência é razoável).

Interpretação (S. André): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Osasco): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Sorocaba): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Mauá): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (S. J Rio Preto): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (S. J Campos): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Rib Preto): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Santos): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável). Alguma correlação positiva com 12 meses anteriores.

Interpretação (Mogi Cruzes): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Diadema): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Jundiaí): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Carapicuíba): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Piracicaba): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável). Alguma autocorrelação positiva com mês anterior (valor próximo de 0,4)

Interpretação (Bauru): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (S. Vicente): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Itaquacetuba): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Franca): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Guarujá): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Taubaté): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável). Autocorrelações negativas para meses 9 e 11 anteriores.

Interpretação (Limeira): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Praia Grande): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Suzano): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Taboão da Serra): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Sumaré): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Barueri): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável). Alguma autocorrelação positiva para mês 11.

Interpretação (Embu): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (S. André): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável). Alguma correlação (positiva e negativa) para meses 2 e 3.

Interpretação (S. André): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável). Alguma correlação (positiva e negativa) para meses 1 e 5.

Interpretação (Americana): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Jacareí): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Indaiatuba): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Araraquara): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Cotia): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (P. Prudente): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Itapevi): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

Interpretação (Hortolândia): autocorrelações são pequenas-menores do que 0,2 em valor absoluto (a suposição de independência é razoável).

ANEXO - DECLARAÇÃO DO COMITÊ DE ÉTICA

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Padrões temporais de casos de tuberculose em municípios do estado de São Paulo, no período de 2009 a 2013 **Pesquisador:** Angela Achcar **Área Temática:**

Versão: 1

CAAE: 42441615.7.0000.5440

Instituição Proponente: Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.037.788

Data da Relatoria: 04/03/2015

Apresentação do Projeto:

Trata-se de estudo observacional, em uma única instituição para estudar a distribuição temporal da 200.000 habitantes no Estado do São Paulo. Serão levantados dados junto ao Sistema de Informação de Agravos de Notificação (SINAN NET).

Objetivo da Pesquisa:

Como trata-se de estudo observacional, não há objetivo específico para o estudo exceto o de descrever a distribuição temporal da doença para os casos novos notificados nos municípios com mais de 200.000 habitantes no Estado do São Paulo (centros, no período de janeiro de 2009 a dezembro de 2013, além de identificar covariáveis socioeconômicas, demográficas e de saúde, que estejam associadas à doença e mensal de novos casos no SINAN.

Avaliação dos Riscos e Benefícios:

O risco embutido neste tipo de projeto de pesquisa é o de violação de dados dos pacientes incluídos. Não há benefícios aos pacientes estudados por se tratar de estudo retrospectivo, porém as informações obtidas poderão auxiliar pacientes e médicos no futuro.

Comentários e Considerações sobre a Pesquisa:

Estudo ecológico para avaliação da distribuição temporal da tuberculose no estado de São Paulo.

Endereço: CAMPUS UNIVERSITÁRIO

Bairro: MONTE ALEGRE

CEP: 14.048-900

UF: SP

Município: RIBEIRÃO PRETO

Telefone: (16)3602-2228

Fax: (16)3633-1144

E-mail: cep@hcrp.usp.br

Continuação do Parecer: 1.037.788

Considerações sobre os Termos de apresentação obrigatória:

Os termos apresentados estão corretos e adequados ao estudo. A solicitação de dispensa de TCLE está redigida em documento à parte.

Recomendações:

Não há.

Conclusões ou Pendências e Lista de Inadequações:

Diante do exposto e à luz da Resolução CNS 466/2012, o projeto de pesquisa, assim como a dispensa do Termo de Consentimento Livre e Esclarecido, podem ser enquadrados na categoria APROVADO.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Projeto Aprovado: Tendo em vista a legislação vigente, devem ser encaminhados ao CEP, relatórios parciais anuais referentes ao andamento da pesquisa e relatório final ao término do trabalho. Qualquer modificação do projeto original deve ser apresentada a este CEP em nova versão, de forma objetiva e com justificativas, para nova apreciação.

RIBEIRAO PRETO, 27 de Abril de 2015

Assinado por:
MARCIA GUIMARÃES VILLANOVA
 (Coordenador)

Endereço: CAMPUS UNIVERSITÁRIO

Bairro: MONTE ALEGRE

CEP: 14.048-900

UF: SP

Município: RIBEIRAO PRETO

Telefone: (16)3602-2228

Fax: (16)3633-1144

E-mail: cep@hcrp.usp.br