

Métodos regressivos para estimação de coberturas de óbitos para as regiões do Nordeste

Pedro Rafael Diniz Marinho¹, Neir Antunes Paes², Julice Suares de Souza³

¹Aluno do Departamento de Estatística e bolsista PIBIC/CNPq – Universidade Federal da Paraíba (UFPB)

pedro.rafael.marinho@gmail.com

²Professor do Departamento de Estatística – Universidade Federal da Paraíba (UFPB)

antunes@de.ufpb.br

³Aluna do Departamento de Estatística – Universidade Federal da Paraíba (UFPB)

julice-suares@hotmail.com

1. Introdução

A preocupação com a qualidade dos dados no Brasil remota pelo menos desde os anos 50, quando começaram a surgir os primeiros trabalhos referentes às estatísticas vitais. Porém, somente a partir dos anos 70 começaram a aparecer trabalhos mais sistemáticos e o emprego de procedimentos mais elaborados sobre o comportamento das estatísticas vitais para os Estados brasileiros. As estatísticas vitais do Nordeste brasileiro são debilitadas devido à presença de sub-registro, principalmente em se tratando de dados de mortalidade. Este estudo se reveste de importância, uma vez que com o conhecimento do comportamento e da distribuição dos óbitos pode-se traçar medidas públicas mais efetivas para o melhoramento das condições de vida de uma população. O principal objetivo deste trabalho consiste em propor um modelo preditivo para as coberturas de óbitos, utilizando coberturas estimadas para as 42 mesorregiões do Nordeste brasileiro. [1]

2. Metodologia

2.1 Coberturas de óbitos

Coberturas de óbitos se referem ao percentual de óbitos de uma dada região que foram registrados. As coberturas utilizadas neste trabalho foram estimadas por Paes [2] para as 42 mesorregiões do Nordeste utilizando o método de Brass para cálculo de coberturas. [2] O método de Brass (1975) supõe que em qualquer grupo de idade aberto $a+$ de uma população fechada, a taxa de entrada em um dado grupo etário é igual à taxa de crescimento populacional menos a taxa de saída (mortes) do mesmo. Numa população estável, a taxa de crescimento é constante para todos os grupos de idade, de forma que a taxa de entrada e saída devem estar relacionadas linearmente.

Se a taxa de entrada é calculada exclusivamente a partir da estrutura etária da população, qualquer erro de cobertura que não varia com a idade é eliminado. No entanto, a taxa de saída calculada a partir dos óbitos por idade como da população por idade estará afetada por qualquer diferencial da cobertura entre a população e as mortes. A inclinação da reta que relaciona a taxa de entrada com a taxa de saída estimará o grau de cobertura do registro de população em relação com o registro dos óbitos, oferecendo um fator de ajuste para os óbitos. [3]

2.2 Modelos Regressivos para o cálculo preditivo das coberturas

2.2.1 Modelos Lineares Generalizados

Os Modelos Lineares Generalizados foram propostos em 1972 por Nelder e Wedderburn, sendo esses derivados dos modelos normais lineares. A idéia básica do MLG consiste em abrir um leque de possibilidades de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições. Além disso, tal modelo dá maior flexibilidade para a relação funcional entre a média da variável resposta (μ) e o preditor linear (η). A ligação entre a média e o preditor linear não é necessariamente a identidade, podendo assumir qualquer forma monótona não-linear.

Considerando n variáveis aleatórias independentes y_1, y_2, \dots, y_n de média μ_i cada uma com função densidade ou de probabilidade na família exponencial da forma apresentada abaixo:

$$f(y_i | \theta_i, \phi) = e^{\frac{y_i \theta_i - b(\theta_i) + c(y_i, \phi)}{a_i(\phi)}}, \text{ com } i = 1, \dots, n \quad (1)$$

em que $a_i(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções específicas para cada distribuição. Se Φ for conhecido tem-se uma distribuição da família exponencial com parâmetro canônico θ . Caso o parâmetro Φ seja desconhecido a distribuição pode ou não pertencer à família exponencial, e há então que considerar apenas os casos em que a distribuição pertence àquela família. [4]

A função $a_i \phi$ toma na maior parte dos casos a forma $a_i(\phi) = \frac{\phi}{\omega_i}$ onde se assume que ω_i é conhecido enquanto que ϕ denominado parâmetro de dispersão ou de escala pode sê-lo ou não. A partir da matriz de dados defini-se um preditor linear η_i com i variando de 1 à N . As componentes η_i e μ_i relacionam-se através de uma função de ligação g_i que se admite existir sendo ela monótona e diferenciável em que transforma μ_i em η_i , ou seja, $\eta_i = g(\mu_i)$ com $i = 1, \dots, N$. [5]

2.2.2 Variáveis consideradas nos modelos regressivos

Para a construção do modelo para predição das coberturas de óbitos foram utilizadas variáveis socioeconômicas e demográficas, ou seja, conhecendo os valores das variáveis que foram selecionadas pelo modelo pode-se ter uma estimativa plausível das coberturas de óbitos de uma região de onde os valores das variáveis selecionadas foram obtidos. O modelo geral partiu das seguintes variáveis:

Variáveis educacionais:

1. V1 - Percentual de crianças de 7 a 14 anos com mais de um ano de atraso escolar;
2. V2 - Percentual de pessoas de 18 a 24 anos que estão frequentando curso superior;
3. V3 - Percentual de pessoas de 15 anos ou mais com menos de 4 anos de estudo;
4. V4 - Percentual de crianças de 10 a 14 anos com menos de 4 anos de estudo;
5. V5 - Percentual de pessoas de 18 a 24 anos com acesso ao curso superior;
6. V6 - Percentual de pessoas que frequentam o ensino médio em relação à população de 15 a 17 anos;
7. V7 - Percentual de professores do ensino fundamental residentes com curso superior;
8. V8 - Percentual de crianças de 10 a 14 anos com menos de 4 anos de estudo;
9. V9 - Percentual de pessoas de 18 a 22 anos com acesso ao curso superior.

Variáveis de Renda:

1. V10 - Percentual de crianças com renda per capita menor que R\$ 37,75;
2. V11 - Renda per capita;
3. V12 - Percentual de pessoas com renda per capita abaixo de R\$ 75,55.

Variáveis de infra-estrutura:

1. V13 - Percentual de pessoas que vivem em domicílios urbanos com coleta de lixo;
2. V14 - Percentual de pessoas que vivem em domicílios com água encanada.

Variáveis populacionais:

1. V15 - Percentual de mulheres chefes de família sem cônjuge e com filhos menos de 15 anos.

3. Resultados e Discussão

Com as 42 coberturas de óbitos tentou-se propor um modelo de regressão para predição das coberturas de óbitos para o Nordeste brasileiro. Para isso foram utilizadas variáveis socioeconômicas e demográficas de modo que conhecendo as variáveis selecionadas pelo modelo pode-se ter uma estimativa plausível dentro das limitações do modelo para a estimativa das coberturas de dados de óbitos. Tal

modelo pode ser aplicado a nível regional e em nível de mesorregiões não sendo aconselhável seu uso para desagregações menores.

Partindo inicialmente do modelo com as 15 variáveis descritas na metodologia, foram utilizados Modelos Lineares Generalizados, em busca de encontrar um modelo para estimativa de níveis de cobertura. Das 15 variáveis, apenas 2 foram significativas para o modelo proposto. A equação do modelo de estimação de coberturas considerou apenas as variáveis V4 e V5, como pode ser observado no modelo apresentado:

$$Y = 4,667572 - 0,007819 V4 + 0,002207 V5$$

onde Y é a cobertura de órbitos predita.

Para a criação do modelo, foram verificadas todas as hipóteses necessárias para validar tal modelo como as hipóteses de normalidade, linearidade e homocedasticidade dos resíduos padronizados. A normalidade foi aceita com um p-valor de 0,56 a um nível de significância de 5%. As demais hipóteses foram observadas graficamente. A Figura 1 apresenta os gráficos que comprovam as hipóteses de linearidade e homocedasticidade, necessárias para validação do modelo. Não foi necessária nenhuma transformação sobre as variáveis consideradas para a satisfação das hipóteses necessárias para validação do modelo, onde o ajustamento do modelo foi 73,04%.

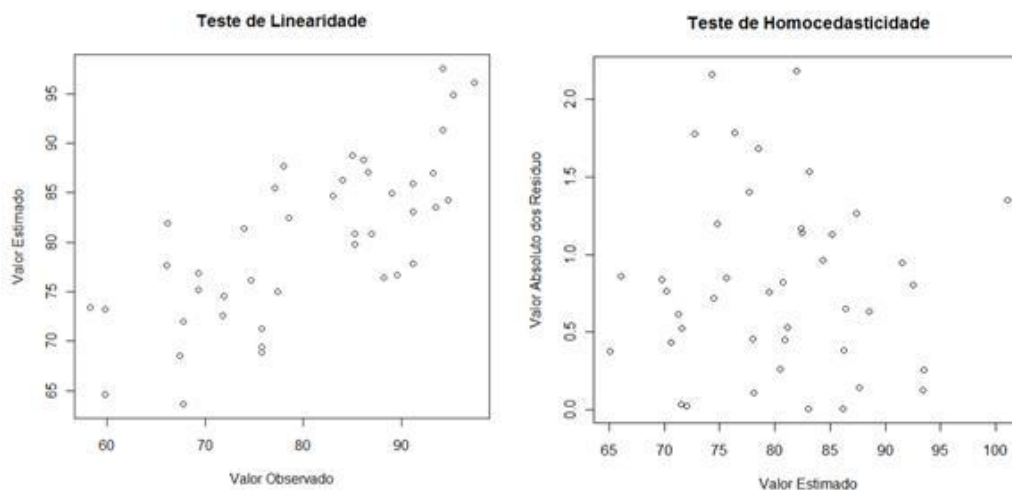


Figura 1: Gráfico das pressuposições de linearidade e homocedasticidade do modelo proposto.

O Gráfico 2 apresenta as coberturas estimadas por Paes, para as 42 mesorregiões do Nordeste brasileiro para o ano de 2000 em preto e as coberturas estimadas pelo modelo proposto nesse artigo em vermelho. Trata-se de um modelo parcimonioso onde são consideradas poucas variáveis e pode ser facilmente aplicado. Os métodos demográficos para estimativas de coberturas de óbitos são bastante eficientes, no entanto requer um pouco de subjetividade e experiência do pesquisador sobre as estatísticas vitais e o comportamento da mortalidade da região que se queira calcular a cobertura. É importante observar que as variáveis selecionadas pelo modelo regressivo pertencem à categoria das variáveis educacionais. Tal fato não implica que outras variáveis possam explicar a variável de interesse.

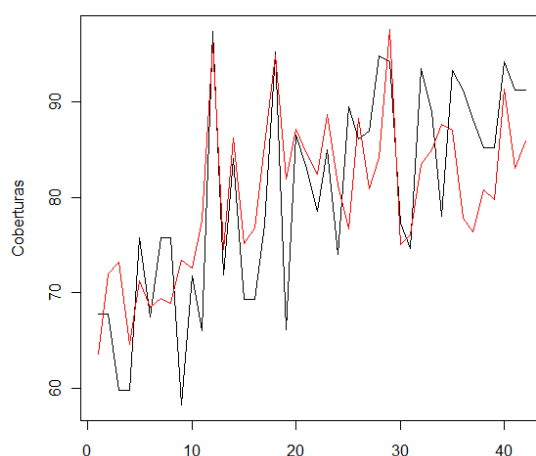


Gráfico 2: Distribuição das estimativas das coberturas de óbitos pelo método de Brass e pelo modelo Regressivo, 2000.

As coberturas preditas pelo modelo proposto apresentaram-se satisfatórias, tanto nas estimativas quanto na facilidade de uso do modelo e pelo poder de predição. Foram comparados os resultados do modelo de previsão com as estimativas obtidas pelo modelo Brass, onde as diferenças foram razoavelmente pequenas e aceitáveis. Como ilustração, na Tabela 1, são apresentados os resultados para o Estado da Paraíba.

Tabela 1: Estimativas das coberturas pelo método de Brass e Regressivo para as mesorregiões da Paraíba, 2000.

Mesorregiões	Método Brass(%)	Modelo Regressivo(%)
Mata Paraibana	94,00	94,63
Agreste	84,55	82,43
Borborema	89,60	88,73
Sertão	85,99	81,34

4. Considerações Finais

Utilizando a teoria de Modelos Lineares Generalizados foi possível propor um modelo regressivo para predição das coberturas de óbitos, para os Estados e suas respectivas mesorregiões do Nordeste brasileiro. O modelo proposto mostrou-se bastante eficiente devido as suas previsões, bem como a sua facilidade de aplicação, bastando conhecer o percentual de crianças de 10 a 14 anos com menos de 4 anos de estudo e o percentual de pessoas de 18 a 24 anos com acesso ao curso superior.

5. Referências

- [1] Paes NA, Albuquerque ME. Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. *Rev. Saúde Pública*. 1999; 33(1): 33-43.
- [2] Paes NA. Avaliação da cobertura dos registros de óbitos dos Estados brasileiros em 2000. *Rev. Saúde Pública*. 2005; 39(6): 882-90.
- [3] Brass W. Methods for estimating fertility and mortality from limited and defective data. Chapel Hill (NC): International Program of Laboratories for Population Statistics; 1975.
- [4] Paula A. Modelos de Regressão com apoio computacional. Instituto de Matemática e Estatística, USP, 2004.
- [5] Miller J, Haden P. Statistical Analysis whit the general linear model. Copyrigh, 2006.