

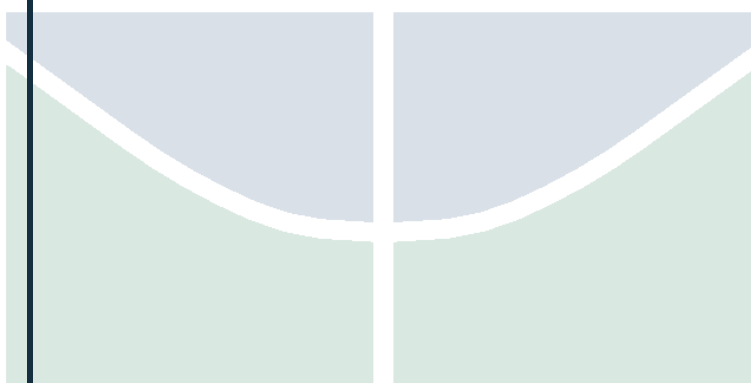


GWR para acessibilidade ao telefone fixo nos municípios de São Paulo

Um estudo de área local

Micael Egídio Papa da Silva - 211029236

Universidade de Brasília



Dezembro, 2023

Sumário

Sumário	2
1 Introdução	3
2 Metodologia	4
3 Resultados	5
3.1 Modelo Global	5
3.2 Escolha da bandwidth	6
3.3 Modelos Locais	7
3.4 Comparando os modelos	10
4 Conclusão	11
Referências Bibliográficas	12

Introdução

A estatística espacial emerge como um campo fundamental para compreender a interseção entre dados geospaciais e análises estatísticas, oferecendo insights valiosos em diversas áreas. No contexto da saúde pública e das pesquisas epidemiológicas, a avaliação da acessibilidade aos telefones fixos entre os municípios ganha destaque como um fator determinante na coleta de dados e na representatividade das pesquisas.

A acessibilidade aos telefones fixos desempenha um papel crucial na realização de inquéritos e estudos epidemiológicos. Estudos como "Home landline telephone coverage and potential bias in epidemiological surveys" (BERNAL; SILVA, 2009) enfatizam a relevância de considerar vieses potenciais que podem surgir devido à cobertura de telefones fixos domiciliares. Essas investigações ressaltam a importância de compreender e mitigar possíveis distorções nos dados resultantes da exclusão de populações sem acesso a esse meio de comunicação.

A cidade de São Paulo, como um microcosmo complexo e diversificado, oferece um cenário propício para a análise da acessibilidade aos telefones fixos nos municípios. Além disso, estudos como o realizado em Rio Branco, AC, intitulado "Inquérito por telefone: pesos de pós-estratificação para corrigir vícios de baixa cobertura em Rio Branco, AC" (BERNAL et al., 2013), destacam a necessidade de estratégias corretivas para vícios resultantes da baixa cobertura em áreas específicas, visando garantir a representatividade e validade dos dados coletados por meio de inquéritos telefônicos.

Neste contexto, este trabalho busca explorar a relação entre a acessibilidade aos telefones fixos nos municípios de São Paulo e sua influência na representatividade dos dados estatísticos. Utilizando técnicas de estatística espacial, pretende-se compreender como a distribuição desigual dessa acessibilidade pode impactar pesquisas epidemiológicas e a representação da população em estudos estatísticos, permitindo identificar áreas suscetíveis a vieses devido à baixa cobertura ou exclusão por falta de acesso a telefones fixos.

Metodologia

Será utilizada a Regressão Beta devido ao intervalo do índice estar entre 0 e 1. Além disso, foi considerado que a distribuição Beta é muito versátil, alterando seu formato dependendo dos valores de seus parâmetros.

Modelos globais e locais serão ajustados e comparados para verificar qual é mais adequado para esses dados específicos. A abordagem de regressão ponderada geograficamente será utilizada para os modelos locais. Em todos os casos, será empregada a função de ligação logit.

Será coletada e utilizada uma amostra aleatória simples de tamanho 100 para otimizar o uso computacional e permitir o teste de mais modelos.

Todos os cálculos serão realizados utilizando o software *R* e seus pacotes disponíveis, incluindo o pacote *gwbr*.

Resultados

Dado que algumas das colunas estão fortemente correlacionadas (como população total e população urbana, por exemplo), para evitar multicolinearidade, apenas três foram inicialmente selecionadas como variáveis explicativas:

- $prop_{urb}$: Proporção da população urbana;
- $prop_{rural}$: População rural;
- $prop_{poor}$: Proporção da população pobre (Considerando renda per capita domiciliar igual ou inferior a R\$140,00 por mês).

3.1 | Modelo Global

Tabela 3.1: Estimativas dos parâmetros para o modelo de regressão beta global.

	Estimate	Std. Error	t Value	Pr> t	Odds Ratio
Intercept	-4.24	0.39	-10.75	0.00	0.00
prop_urb	3.92	0.41	9.49	0.00	50.25
pop_rural	0.00	0.00	5.28	0.00	1.00
prop_poor	-2.15	1.09	-1.97	0.05	0.12
Phi	24.57	3.42	7.17	0.00	

De acordo com os p-valores da tabela 1, podemos concluir que todos os parâmetros foram significativos, com exceção do parâmetro relacionado a proporção da população em situação de pobreza a um nível de 5 %.

Temos que a proporção da população urbana e a população rural possuem uma relação positiva com a proporção de telefones fixos por domicílio, ou seja, ao passo que esses valores crescem a variável resposta tende a crescer proporcionalmente.

Por outro lado, temos que a proporção da população em situação de pobreza tem estimativas negativas, de modo a indicar que quando seu valor cresce a proporção de telefones fixos por domicílio decresce.

Dentre as variáveis, a proporção de população urbana possui a relação mais forte e evidente com a variável resposta.

Vale ressaltar que o parâmetro de precisão ϕ foi estimado sendo 24.56 .

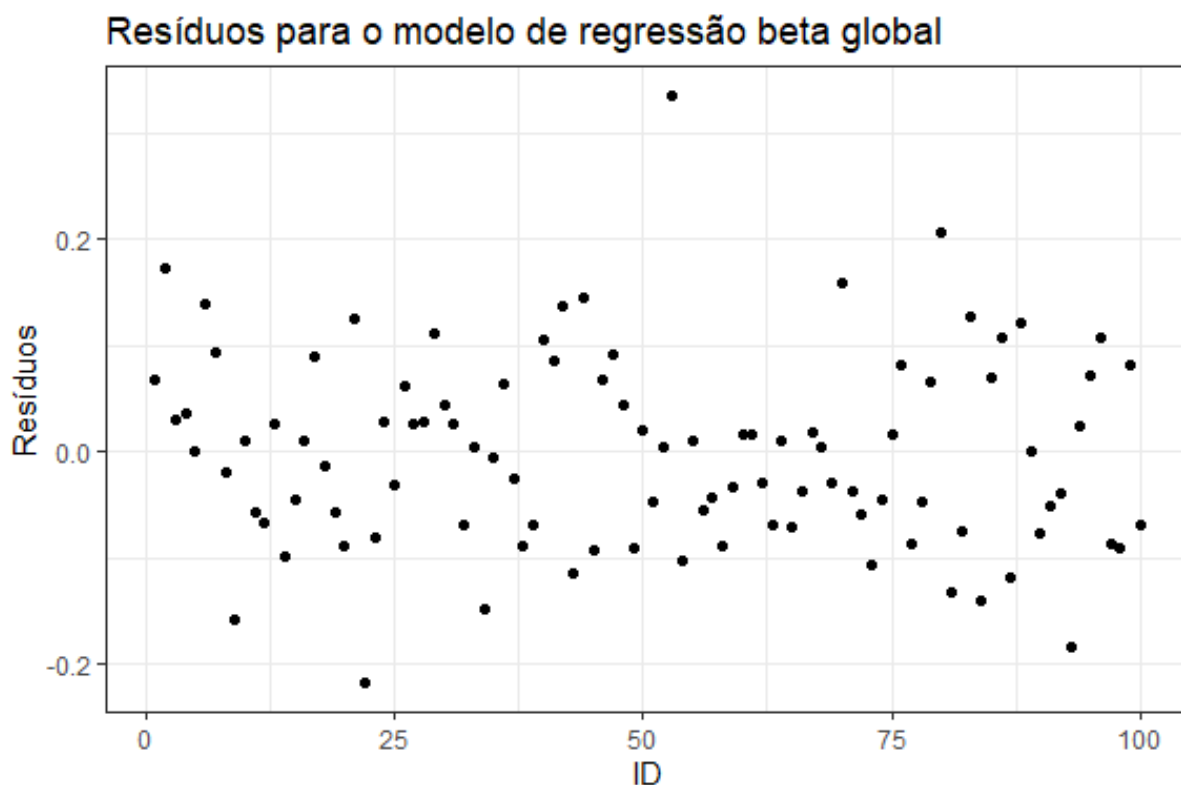


Tabela 3.2: Medidas de qualidade de ajuste para o modelo de regressão beta global.

Pseudo.R2	Adj.pseudo.R2	Log-verossimilhança	AICc
0.58	0.56	102.18	-193.72

3.2 | Escolha da bandwidth

A escolha do valor da largura de banda (bandwidth) é uma das partes mais importantes da técnica de regressão geograficamente ponderada. O valor ideal pode ser encontrado por meio do algoritmo de Busca da Seção Áurea (Golden Section Search - GSS), minimizando o Escore de Validação Cruzada (CV) ou o Critério de Informação de Akaike (AIC).

Tabela 3.3: Escolha da bandwidth via GSS.

	Fixo	Adaptado
CV	838.513	40
AIC	-	98

O algoritmo GSS foi executado utilizando as medidas AIC e CV, resultando nos valores mostrados na Tabela 3. Outra variação testada foi o método de cálculo: uma largura de banda fixa com base na distância ou uma largura de banda adaptativa com base nos vizinhos mais próximos.

O método fixo resultou em erro quando usado com o AIC, então apenas três valores de largura de banda foram encontrados. Cada um deles será utilizado para ajustar um modelo local.

3.3 | Modelos Locais

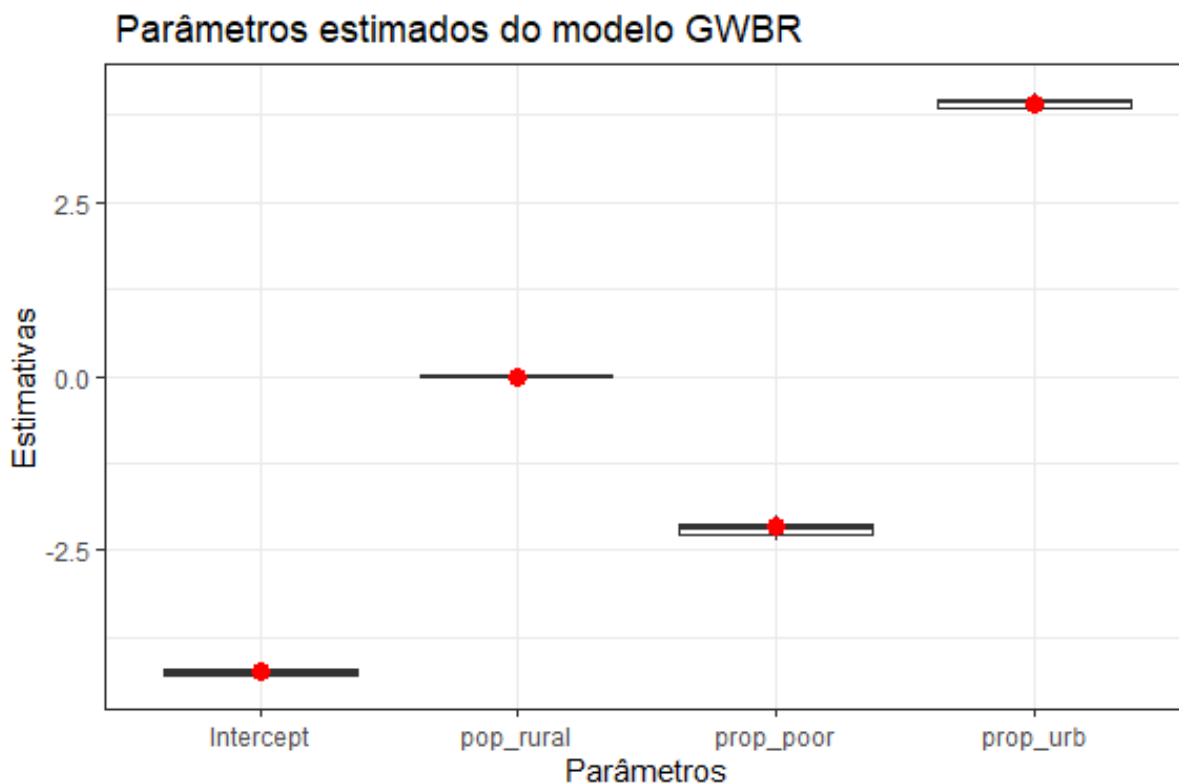
Depois de ajustar os três modelos, algumas medidas podem ser avaliadas para se optar pelo modelo que melhor adequara aos dados.

Tabela 3.4: Medidas de ajuste para modelos GWBR por largura de banda.

	Pseudo R2	Adj. Pseudo R2	Log-verossimilhança	AIC	AICc
h=838.513	0.59	0.58	103.60	-198.80	-198.34
h=40	0.58	0.62	99.73	-219.93	-218.20
h=98	0.58	0.58	102.00	-201.06	-200.98

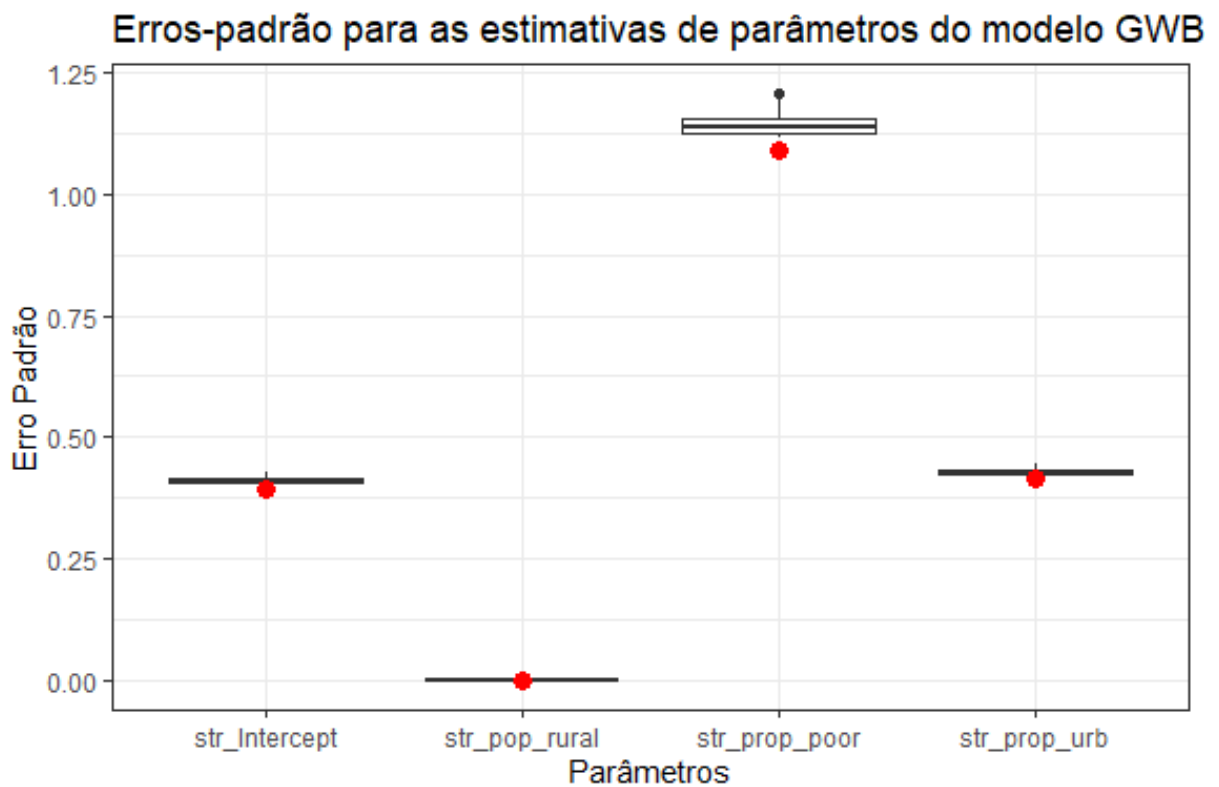
O modelo com a largura de banda fixa de 838,513 apresenta um ajuste melhor (pseudo- R^2 e log-verossimilhança mais altos) e é mais parcimonioso, conforme demonstrado pelos valores mais baixos de AIC e AICc.

Portanto, os resultados que serão apresentados a seguir são todos do modelo com largura de banda fixa de 838,513.



Todas as estimativas de parâmetros parecem estar centralizadas nos seus valores do modelo global, representados como pontos vermelhos.

Nenhum dos parâmetros apresentou uma alta variabilidade, demonstrando ser desnecessário cogitar modelos locais para cada variável tendo em mente a ausência de variações locais.

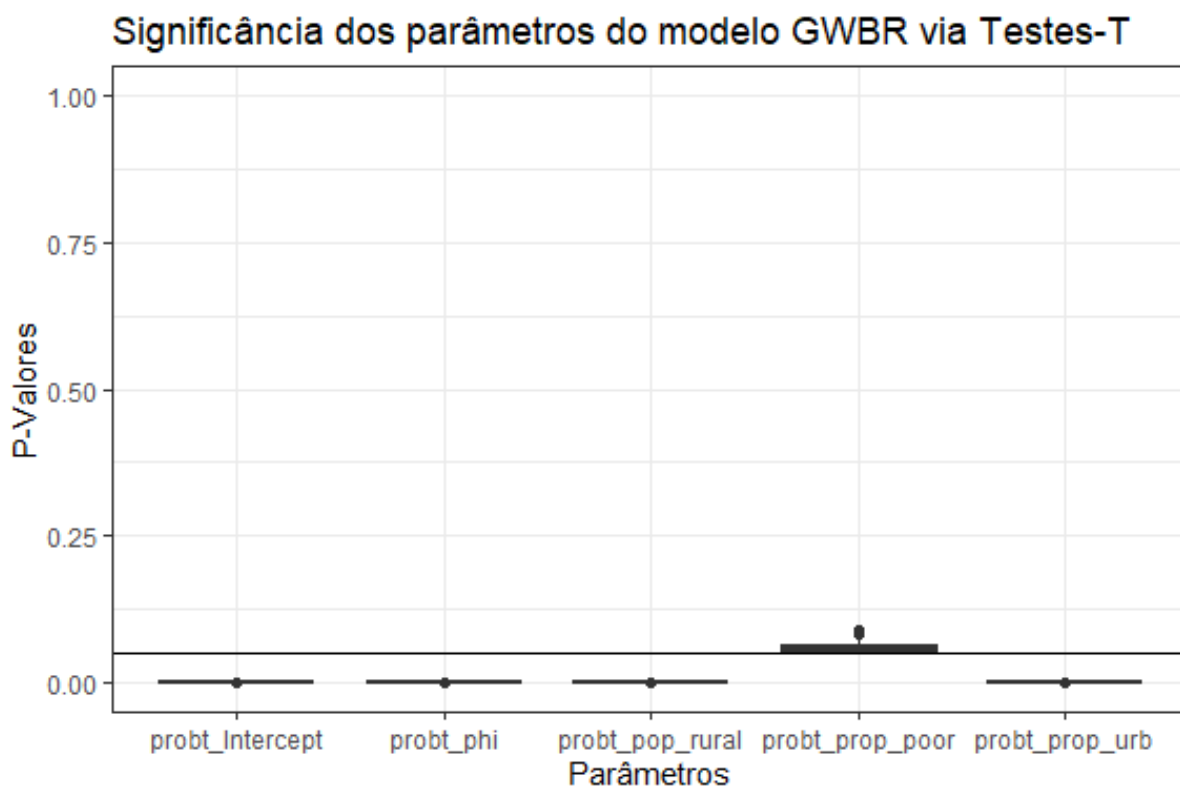


O erro padrão que se distancia um pouco da regressão beta global é apenas o que diz respeito a proporção da população em situação de pobreza.

Tabela 3.5: Quartis para estimativas de parâmetros de precisão e erros padrão.

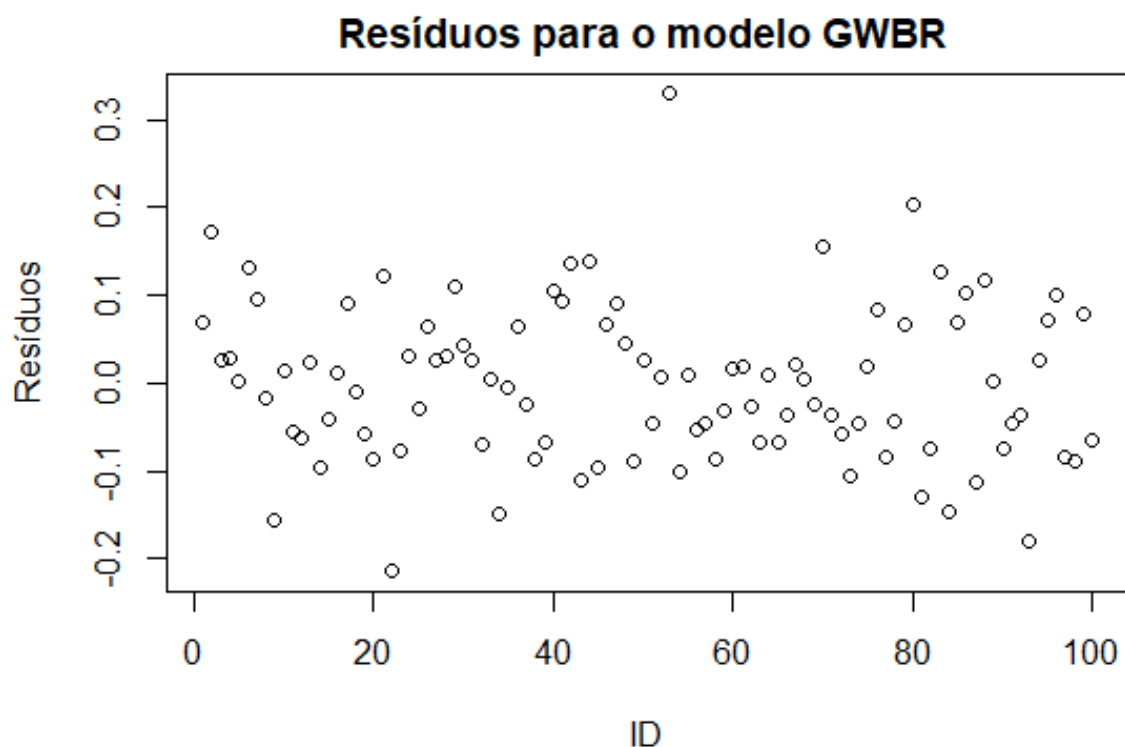
	est	std error
Q1	24.39	3.52
Q2	24.60	3.55
Q3	24.98	3.61
IQR	0.60	0.08

O parâmetro de precisão ϕ foi estimado globalmente como 24.56, sendo bem próximo da mediana das estimativas locais (24.596)



A linha traçada nas ordenadas no gráfico indica o nível de significância de 0,05, facilitando a visualização dos valores p significativos.

O interessante é que, para algumas localidades, o parâmetro da proporção da população pobre não foi significativo, indicando que a relação entre essa variável e a $prop_{baseline}$ varia localmente.



3.4 | Comparando os modelos

Por fim, os modelos global e locais serão ajustados novamente com a exclusão da variável não significativa.

Para fins de comparação, um modelo semi-paramétrico também será ajustado, permitindo que o parâmetro $prop_{urb}$ seja global.

Tabela 3.6: Modelos de regressão beta global, local e mista.

	Pseudo R2	Adj. Pseudo R2	Log-verossimilhança	AIC	AICc
Global	0.58	0.56	102.18	-193.72	-193.72
Local	0.59	0.58	103.60	-198.80	-198.34
Mista	0.20	0.19	-433.66	871.83	871.99

Para todas as medidas apresentadas, o GWBR foi o melhor modelo, seguido pelo BR global. O modelo misto GWBR teve os piores resultados.

Conclusão

A comparação de diferentes métodos para escolher a largura de banda reforçou o quão importante esse parâmetro é para modelos ponderados geograficamente e o quanto ele influencia o resultado do modelo.

Em geral, os modelos GWBR tiveram um desempenho melhor do que a Regressão Beta global, indicando que esses dados são espacialmente não estacionários. A proporção de população urbana, em particular, mostrou fortes indícios de ter uma relação variável localmente com a $prop_{landline}$

Referências Bibliográficas

BERNAL, R.; SILVA, N. N. d. Home landline telephone coverage and potential bias in epidemiological surveys. *Revista de Saúde Pública*, SciELO Brasil, v. 43, p. 421–426, 2009.

BERNAL, R. T. I. et al. Inquérito por telefone: pesos de pós-estratificação para corrigir vícios de baixa cobertura em rio branco, ac. *Revista de Saúde Pública*, SciELO Public Health, v. 47, p. 316–325, 2013.