

RELAÇÃO ENTRE DOENÇAS E NATUREZA JURÍDICA DE EMPRESAS DE ABASTECIMENTO DE  
ÁGUA E ESGOTAMENTO SANITÁRIO NO BRASIL SOB A ÓTICA DE REGRESSÕES LINEARES.

Leonardo Vitta

Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA-USP)

EAC0355 - Modelos de Regressão

Prof. Dr. Joelson Oliveira Úbida Sampaio

05 de julho, 2024.

## 1 INTRODUÇÃO

O acesso à água potável é essencial para a qualidade de vida da população e, relaciona-se diretamente com o princípio constitucional da dignidade da pessoa humana, assim como o saneamento básico é estratégia essencial para garanti-la.

Em tempo, o crescimento desenfreado de metrópoles como São Paulo e a própria constituição econômica e social brasileira não ausentam os homens de doenças associadas à pobreza e a falta de saneamento básico, particularmente em um país em desenvolvimento, como é o caso brasileiro, onde essa relação entre doenças, a disponibilidade de água e saneamento básico caminham em via contrária aos direitos à saúde e ao saneamento.

O direito à saúde está atrelado ao artigo 196 da Constituição Federal, que estabelece que “a saúde é um direito de todos e dever do Estado”, ou seja, deve ser garantido através de políticas públicas a partir da esfera Federal e atingindo as demais esferas do poder público.

O saneamento básico e as políticas públicas de saneamento são norteadas pela Lei do Saneamento (Lei Nº 11.445/2007), que estabelece suas diretrizes base e modelos de atuação a serem seguidos, compreendidos holisticamente pelo abastecimento de água, esgotamento sanitário, limpeza urbana, manejo dos resíduos sólidos, disponibilidade de drenagem, manejo de águas pluviais em áreas urbanas e fiscalização preventiva das redes de saneamento, que se alinham tecnicamente a articulação de políticas públicas de melhoria de qualidade de vida, a exemplo do desenvolvimento urbano e regional, de combate à pobreza e sua erradicação e da promoção da saúde e de recursos hídricos.

No Brasil, a oferta dos serviços de água abrange 5.191 municípios brasileiros, o equivalente a 98,2% da população urbana, e os serviços de esgotamento sanitário estão presentes em 4.226 municípios, que representam 75,9% do total, abrangendo 92,9% da população com acesso a esse serviço (AMOSTRA SNIS, 2019).

Com base em estudos empíricos compilados entre 2007 e 2018 em artigo do “*Brazilian Journal of Development*”, “*Parasitoses intestinais e saneamento básico no Brasil: estudo de revisão integrativa*”, trata-se de um estudo que se auto define como exploratório e descritivo, fonte de informações importantíssimas sobre os 29 artigos selecionados. A maior parte dos estudos sobre enteroparasitos se concentram na região sudeste, com evidente escassez de estudos sobre giardíase, infecção parasitária comum, diretamente associada a condições sanitárias precárias (Ministério da Saúde), nas regiões Norte, Nordeste e Centro-Oeste do país, onde, apesar da *Giardia intestinalis* ser o protozoário mais relatado infectando a população nos estudos que realizaram o método de diagnóstico para o parasito (10/25, 40,0%), com prevalência variante de 4,9% a 96,6%, dado corroborado pela literatura a respeito (IGNACIO et al. 2018 ; FANTINATTI et al. 2016) a giardíase ainda permanece como uma doença negligenciada no país.

Com esse exemplo e outros trazidos pelo artigo, é possível inferirmos a discrepância na quantidade de estudos atrelada a desigualdade entre as regiões brasileiras e, de forma demasiadamente complexa, intriga até mesmo especialistas.

Portanto, o objetivo desse estudo é de atestar as condições e níveis relacionais entre as ocorrências de doenças e a natureza jurídica da empresa de abastecimento de água e fornecimento de esgotamento sanitário

dos municípios brasileiros, ao invés de identificar níveis de ocorrência por tipo de parasitas, protozoário, amebas, etc., e sua participação no desenvolvimento de doenças, pois demandaria maiores estudos, por região e por espécie, presente em grande parte na documentação de literatura revisada e referenciada.

## 2 METODOLOGIA

Esse estudo trata-se de um trabalho descritivo a partir da análise e manipulação de dados empíricos, divulgados por órgãos e autarquias governamentais, embasado em literatura a respeito dos casos de doenças associadas à pobreza, relacionadas à falta de saneamento básico.

A realização deste trabalho se desenvolveu através das seguintes etapas: 1) A identificação do tema e estabelecimento da problemática a ser abordada e pesquisada. 2) pesquisa de literatura a respeito. 3) coleta e avaliação dos dados públicos dispostos. 4) análise dos dados (sua síntese, exibição, realizar comparação dos dados, conclusão) e 5) apresentação dos resultados obtidos através de programação em linguagem R.

A busca de literatura e documentação foi realizada nos bancos de dados informatizados presentes nas plataformas do governo Brasileiro e do google academics, a busca das obras indexadas foi realizada por meio das palavras chave “Saneamento Básico”, “Saneamento e Parasitoses”, em Biblioteca Virtual em do Ministério da Integração e do Desenvolvimento Regional (<https://www.gov.br/mdr/pt-br>), e do sistema SIDRA do IBGE (<https://sidra.ibge.gov.br/pesquisa/pnsb/pnsb-2017>).

Partindo das tabelas com referências a nível Brasil e Estaduais, foram encontrados dados referentes a doenças relacionadas à oferta de água e saneamento e, ao mesmo tempo, o número de municípios que possuem esgotamento sanitário, separados por doença e por natureza jurídica das unidades de redes de distribuição. Há uma dificuldade inicial nesse tipo de caso pois os dados não são, necessariamente, todos coletados e provenientes de pesquisas realizadas no mesmo ano, sendo os dados referente às doenças de 2016 e referentes a oferta de esgotamento sanitário de 2017.

Ao amalgamarmos em ordem, ambas as tabelas, podemos realizar os procedimentos de regressão linear, os modelos utilizados foram para regressões múltiplas, utilizando como a variável dependente, as natureza jurídicas das empresas de saneamento mais presentes no país, ou seja, aquelas sob administração direta do poder público e as sociedades de economia mista e como variáveis explicativas, foram utilizadas as seguintes doenças : Dengue, leptospirose, cólera, tifo, verminoses, diarreia, difteria, malária, hepatite, dermatite e febre amarela. O objetivo é ter capacidade de comparar os dois modelos e inferir sob as diferenças e particularidades de cada um.

### 3 RESULTADOS

Os resultados para os dois modelos, utilizando Sociedades de Economia Mista(modelo1) e Administração direta do poder público(modelo2), foram os seguintes:

```
Residuals:
    Min       1Q   Median       3Q      Max
-80.825 -21.345  -0.492  21.369 111.996

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -9.0906    24.5293  -0.371  0.71612
Dengue       -0.6735     0.7069  -0.953  0.35585
Leptospirose  2.4456     3.2003   0.764  0.45662
Cólera      -15.6733     6.3587  -2.465  0.02626 *
Tifo        -17.9662    19.3309  -0.929  0.36740
Verminoses   -5.6834     1.8357  -3.096  0.00738 **
Diarréia     4.6245     1.5788   2.929  0.01037 *
Difteria     0.5613     9.5514   0.059  0.95391
Malária     -2.7365     2.7305  -1.002  0.33214
Hepatite     4.0380     2.4563   1.644  0.12098
Dermatite    3.1706     3.0797   1.030  0.31954
Febre.amarela 25.2322     8.2369   3.063  0.00789 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 52.99 on 15 degrees of freedom
Multiple R-squared:  0.776,    Adjusted R-squared:  0.6117
F-statistic: 4.724 on 11 and 15 DF,  p-value: 0.003266
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-83.030 -23.810   5.824  25.188  70.464

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.8237    24.4912  -0.238  0.8153
Dengue       -1.0594     0.7058  -1.501  0.1541
Leptospirose  4.0628     3.1954   1.271  0.2229
Cólera      -8.7180     6.3488  -1.373  0.1899
Tifo       -13.2760    19.3009  -0.688  0.5020
Verminoses   -0.8785     1.8329  -0.479  0.6386
Diarréia     2.8950     1.5764   1.836  0.0862 .
Difteria     8.6854     9.5366   0.911  0.3768
Malária     -2.3849     2.7263  -0.875  0.3955
Hepatite    -3.2159     2.4525  -1.311  0.2095
Dermatite    2.4211     3.0749   0.787  0.4433
Febre.amarela 13.6242     8.2241   1.657  0.1184
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 52.91 on 15 degrees of freedom
Multiple R-squared:  0.8459,    Adjusted R-squared:  0.7329
F-statistic: 7.486 on 11 and 15 DF,  p-value: 0.0002744
```

O que indica que há muitas variáveis insignificantes se queremos um nível de significância de  $\alpha = .05$ , portanto, foram realizadas diversas simulações, cada vez retirando 1 variável com alto p-valor, até chegar nos seguintes modelos:

```
lm(formula = Administração.direta.do.poder.público ~ Verminoses +
    Malária, data = regressão)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-144.30  -39.12    7.79   21.52  153.17
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -7.7900    16.8066  -0.464  0.6472
Verminoses    1.5381     0.2191   7.021 2.94e-07 ***
Malária     -2.3070     1.0426  -2.213  0.0367 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 60.25 on 24 degrees of freedom
Multiple R-squared:  0.6803,    Adjusted R-squared:  0.6536
F-statistic: 25.53 on 2 and 24 DF,  p-value: 1.141e-06
```

```
call:
lm(formula = Sociedade.de.economia.mista ~ Cólera + Verminoses +
    Malária + Hepatite + Febre.amarela + Verminoses, data = regressão)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-81.119 -30.108  -7.276  22.945 174.383
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  34.1350    17.6914   1.929 0.067294 .
Cólera      -14.3788     6.0772  -2.366 0.027674 *
Verminoses   -1.4170     0.6435  -2.202 0.038982 *
Malária     -4.9924     1.2811  -3.897 0.000831 ***
Hepatite     5.7072     1.5934   3.582 0.001758 **
Febre.amarela 19.8525     6.6392   2.990 0.006976 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 58.46 on 21 degrees of freedom
Multiple R-squared:  0.6183,    Adjusted R-squared:  0.5274
F-statistic: 6.803 on 5 and 21 DF,  p-value: 0.0006466
```

É possível observar que, para um nível de significância de  $\alpha = .05$ , há grande diferença de variáveis relevantes com poder explicativo em relação aos níveis e a natureza jurídica da empresa que oferece o serviço de esgotamento sanitário, onde, com  $R^2$  ajustados de 65,36% e 52,74%, respectivamente, o que significa que em municípios onde a administração é realizada diretamente pelo poder público, verminoses e malária são responsáveis por explicar 65,36% dos eventos de doenças, enquanto que no caso dos municípios onde a oferta do serviço de esgotamento sanitário é realizada por Sociedades de Economia Mista, há mais doenças significantes, mas responsáveis por explicar 52,74% dos casos.

Agora, aplicando testes de normalidade sob os resíduos e de heterocedasticidade:

```
> u <- modelo1$residuals
> u2 <- modelo2$residuals
> dgof::ks.test(u, 'pnorm', mean(u), sd(u))

One-sample Kolmogorov-Smirnov test

data: u
D = 0.10108, p-value = 0.9196
alternative hypothesis: two-sided

> dgof::ks.test(u2, 'pnorm', mean(u2), sd(u2))

One-sample Kolmogorov-Smirnov test

data: u2
D = 0.12326, p-value = 0.7615
alternative hypothesis: two-sided

> shapiro.test(u)

Shapiro-Wilk normality test

data: u
W = 0.91846, p-value = 0.03626

> shapiro.test(u2)

Shapiro-Wilk normality test

data: u2
W = 0.96126, p-value = 0.3947
```

Ambos os testes de Kolmogorov-Smirno atestam normalidade dos resíduos, com p-valor  $> 0.05$  e o teste de Shapiro-Wilk do modelo1 atesta uma rejeição da hipótese nula, de que a amostra do modelo 1 não segue uma distribuição normal, enquanto a do modelo 2 sim, pois o p-valor está acima do nível de significância.

Agora, aplicando os testes de White e de Breusch Pagan para atestar se a variância dos resíduos é constante :

```
> white(modelo1)
# A tibble: 1 x 5
  statistic p.value parameter method      alternative
  <dbl>    <dbl>    <dbl> <chr>      <chr>
1    13.9   0.179      10 white's Test greater
> white(modelo2)
# A tibble: 1 x 5
  statistic p.value parameter method      alternative
  <dbl>    <dbl>    <dbl> <chr>      <chr>
1    23.9 0.0000818      4 white's Test greater
> breusch_pagan(modelo1)
# A tibble: 1 x 5
  statistic p.value parameter method      alternative
  <dbl>    <dbl>    <dbl> <chr>      <chr>
1     9.87 0.0791      5 Koenker (studentised) greater
> breusch_pagan(modelo2)
# A tibble: 1 x 5
  statistic p.value parameter method      alternative
  <dbl>    <dbl>    <dbl> <chr>      <chr>
1    22.1 0.0000157      2 Koenker (studentised) greater
> |
```

Para o modelo 1, ambos os testes verificaram a ausência de heterocedasticidade, pois p-valor é maior do que o nível de significância e para o modelo 2, ambos os testes verificaram a presença de heterocedasticidade, pois p-valor é menor do que o nível de significância, ou seja, a dispersão dos dados não é uniforme.

E, por último, foi realizado o teste VIF, com o fim de atestar se há multicolinearidade entre variáveis.

```
> car::vif(modelo1)
      Cólera      Verminoses      Malária      Hepatite Febre.amarela
      2.204383      9.313890      1.630155      7.204845      3.374224
> car::vif(modelo2)
Verminoses      Malária
      1.016383      1.016383
```

Com altos graus de multicolinearidade ( $VIF > 5$ ), podemos retirar as variáveis “Verminoses” e “Hepatite” do modelo1, e, aplicando novamente :

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    57.836    19.060    3.034  0.00589 **
Cólica         -10.844     7.259   -1.494  0.14881
Malária        -4.279     1.534   -2.788  0.01044 *
Febre.amarela  21.768     7.248    3.003  0.00634 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.96 on 23 degrees of freedom
Multiple R-squared:  0.3309,    Adjusted R-squared:  0.2437
F-statistic: 3.792 on 3 and 23 DF,  p-value: 0.02405
```

É possível observar que há grande perda de poder explicativo em  $R^2$  e, após realizar regressão sem a variável Cólera, a tendência é cair mais ainda, assim como os demais parâmetros de outros testes. Por isso, apesar da problemática multicolinearidade, acredita-se que o modelo1 possua melhor capacidade de explicação desse fenômeno.

## 4 CONCLUSÕES

Foi possível estudar e praticar o uso de regressões lineares a partir de dados reais, realizando a coleta, sua tratativa, busca e definindo o tema a ser estudado e explorado. O tema de saneamento básico é vasto e complexo, com oportunidades de realização de diversas tratativas por muitos ângulos diferentes.

Com as técnicas aplicadas, foi possível inferir que a Malária, ao menos nas duas maiores naturezas jurídicas dos prestadores de serviço de abastecimento de água e esgotamento sanitário, é uma doença relevante e de alta relação com doenças relacionadas à pobreza, uma das premissas pelas quais passamos na introdução e confecção deste trabalho, em complemento aos modelos, podemos concluir que a ocorrência dessas doenças se dá por um caso multifatorial, já que a ausência de saneamento está relacionada a existência de moradias precárias e falta de acesso a métodos de prevenção e todos esses fatores estão relacionados à pobreza.

Dito isto, a partir desse trabalho, é possível pensar em maiores aplicações das regressões nesse tema, expandindo os estudos aos municípios e não apenas aos Estados da Federação.

## REFERÊNCIAS

AUSTRIACO, Phelipe; FANTINATTI; Maria; PINTO, Monique; SANTOS; Joziane. Parasitoses intestinais e saneamento básico no Brasil: estudo de revisão integrativa. Braz. J. of Develop., Curitiba, v. 6, n. 5, p.22867-22890, may. 2020.

BRASIL. Ministério das Cidades. Secretaria Nacional de Saneamento Ambiental –SNSA. Sistema Nacional de Informações sobre Saneamento: Diagnóstico dos Serviços de Água e Esgotos – 2019. Brasília: [SNSA/MCIDADES, 2018a. 218 p.: il.]

Constituição Federal. LEI Nº 11.445, DE 5 DE JANEIRO DE 2007. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2007/lei/l11445.htm](https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/lei/l11445.htm) Acesso em: 02 de julho de 2024.

ENEGEP. RESENDE, Juliano; SOUZA; Agenor, PAULA DE, Layara; HENRIQUE, Paulo. Utilização da regressão linear múltipla para avaliação da qualidade da água : estudo de caso da estação de tratamento meia ponte, Goiânia – Goiás.

IGNACIO, Caroline Ferraz; LIMA-BARATA, Martha Macedo de; MORAES-NETO, Antonio Henrique Almeida de. The Brazilian Family Health Strategy and the management of intestinal parasitic infections. Prim Health Care Res Dev. v. 19, n. 4, p.333-343, 2018.

Ministério da Saúde. Giardíase. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/giardiae#:~:text=A%20infec%C3%A7%C3%A3o%20da%20giard%C3%ADase%20ocorre,como%20rios%2C%20piscinas%20e%20lagos>. Acesso em: 03 de julho de 2024.

Ministério da Integração e Desenvolvimento Regional. Diagnóstico dos Serviços de Água e Esgotos – 2019. Disponível em: <https://www.gov.br/mdr/pt-br/assuntos/saneamento/snis/diagnosticos-anteriores-do-snis/agua-e-esgotos-1/2019>. Acesso em: 03 de julho de 2024.

PNMA. Índice e Indicadores de Qualidade da Água – Revisão de Literatura, 2002. Disponível em: <http://www.cprh.pe.gov.br/downloads/indice-agua-volume1.pdf> Acesso em: 29 de junho de 2024.

IFSC USP. Malaria uma doença negligenciada. Disponível em : <https://www2.ifsc.usp.br/portal-ifsc/malaria-uma-doenca-negligenciada/> Acesso em : 04 de julho de 2024.

WHITTEMORE, Robin; KNAFL, Kathleen. The integrative review: updated methodology. J Adv Nurs., vol. 52, n. 5, p. 546-53, 2005.