
DISTRIBUIÇÃO LOGNORMAL BOOTSTRAP COM APLICAÇÃO EM RADIOECOLOGIA

SILVA, Cleomacio Miguel da¹
VIEIRA, José Wilson¹

Recebido em: 2019.05.12

Aprovado em: 2019.08.20

ISSUE DOI: 10.3738/1982.2278.3605

RESUMO: Para avaliar a distribuição de um elemento de ocorrência natural ou contaminante no meio ambiente, é necessária uma análise estatística robusta. Para tanto, são usados diferentes métodos estatísticos paramétricos e não-paramétricos. Atualmente, o método bootstrap vem sendo eficientemente utilizado em pesquisas ambientais com grande vantagem, devido à sua robustez estatística. Em estudos radioecológicos, o método bootstrap vem apresentando desempenho satisfatório na análise estatística de valores *outliers* (anômalos). A distribuição lognormal é bastante utilizada na análise estatística de dados radioecológicos, por assumir sempre valores positivos. Em muitos casos de padronização, as concentrações de contaminantes ambientais têm sido analisadas usando o limite superior do intervalo de confiança da média da distribuição lognormal. O método bootstrap consiste de um procedimento estatístico computacionalmente intensivo que permite avaliar diversas estatísticas, com base nos dados obtidos da amostra. Então, o objetivo do presente trabalho foi utilizar o método bootstrap para determinar o limite superior do intervalo de confiança para dados radioecológicos com valores anômalos. Os resultados obtidos mostraram que o método bootstrap diminuiu significativamente a dispersão em torno da média da distribuição lognormal, resultando assim num limite superior robusto.

Palavras-chave: Meio Ambiente. Modelo Estatístico. Monte Carlo. Radionuclídeo.

LOGNORMAL BOOTSTRAP DISTRIBUTION WITH APPLICATION IN RADIOECOLOGY

SUMMARY: To evaluate the distribution of a naturally occurring or contaminating element in the environment, robust statistical analysis is required. For this, different parametric and non-parametric statistical methods are used. Currently, the bootstrap method has been used efficiently in environmental research with great advantage, due to its statistical robustness. In radioecological studies, the bootstrap method has shown satisfactory performance in the statistical analysis of outliers (anomalous) values. The lognormal distribution is widely used in the statistical analysis of radioecological data, since it always takes positive values. In many cases of standardization, concentrations of environmental contaminants have been analyzed using the upper limit of the lognormal distribution mean confidence interval. The bootstrap method consists of a computationally intensive statistical procedure that allows the evaluation of several statistics, based on the data obtained from the sample. Therefore, the objective of the present work was to use the bootstrap method to determine the upper limit of the confidence interval for radioecological data with anomalous values. The results showed that the bootstrap method significantly reduced the dispersion around the lognormal distribution average, resulting in a robust upper limit.

Keywords: Environment. Statistical Model. Monte Carlo. Radionuclide.

INTRODUÇÃO

Em radioecologia, na análise estatística dos dados, principalmente na construção de intervalo de confiança, considera-se que os dados são normalmente distribuídos. Porém, os valores das concentrações de radionuclídeos em locais tipicamente anômalos, possuem assimetria à direita, e geralmente são modeladas pela distribuição lognormal. Entretanto, isto pode ser devido a amostragem tendenciosa,

¹ Escola Politécnica da Universidade de Pernambuco

populações múltiplas ou outliers, e não necessariamente aos dados lognormalmente distribuídos (SINGH *et al.*, 1997). Os fenômenos radioativos naturais, em cuja análise intervém o método estatístico, bem como os dados estatísticos a eles referentes, caracterizam-se tanto pela sua semelhança, quanto pela sua variabilidade. Desde modo, o cálculo de uma determinada medida de tendência central só se justifica em razão da variabilidade presente no meio ambiente. Entretanto, caso a variação seja grande, a utilização da média aritmética, como valor mais representativo, seria totalmente inadequada, devido à discrepância entre os valores.

Devido às flutuações estatísticas causadas pelos valores *outliers* (anômalos), os radioecologistas utilizam a média geométrica ou a mediana como valor mais representativo do conjunto de dados obtidos da amostra. No caso da média geométrica, utilizam-se os logaritmos dos valores obtidos da amostra para se obter uma curva simétrica. Nesse caso, a média geométrica revela-se mais adequada que a média aritmética (TOLEDO; OVALLE, 1983). A mediana não é afetada pelos valores anômalos, sendo a medida de tendência central mais utilizada em análises estatísticas de dados discrepantes (TOLEDO; OVALLE, 1983). O grande problema na utilização da média geométrica e mediana como valor mais representativo do conjunto de dados de distribuição de probabilidade assimétrica positiva, é que tais medidas de tendência central, são sempre menores do que a média aritmética (SINGH *et al.*, 1997). Sendo assim, possivelmente, a média geométrica e a mediana não converge adequadamente para o valor mais representativo do conjunto de dados.

O limite de confiança superior baseado na estatística-H para a média de uma população lognormal foi recomendado por documentos de orientação da Agência de Proteção Ambiental dos Estados Unidos (SINGH *et al.*, 1997), e foi amplamente utilizado para tomar decisões de remediação em locais tipicamente anômalos. Porém, a estatística-H mostrou-se insuficiente para calcular um limite de confiança superior para a média de uma distribuição lognormal. Porém, foram testados também os métodos jackknife, bootstrap e a desigualdade de Chebychev (SINGH *et al.*, 1997).

O termo bootstrap surgiu da frase “*to pull oneself up one's bootstrap*” retirada de “*Adventures of Baron Munchausen*” by Rodolph Erich Raspe, XVII century: “*The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought of picking himself up by his own bootstraps*” (EFRON; TIBSHIRANI, 1993). Este texto relata uma situação em que o Barão Munchausen está afundando em um lago e vendo que tudo estava perdido, pensa que conseguirá emergir puxando os cadarços dos próprios sapatos. O sentido estatístico do termo é passar a ideia de que, em situações difíceis, devem-se tentar as mais variadas soluções possíveis a partir dos dados originais. Em estatística, situações difíceis podem ser vistas como os problemas de soluções analíticas complexas. As variadas soluções possíveis seria a utilização de uma metodologia com grande quantidade de cálculos, objetivando extrapolar os resultados a partir de um pequeno conjunto de dados. Com o uso sistematizado de ferramentas computacionais, a solução para esses casos, é obtida substituindo-se a resolução analítica pelo poder de processamento dos computadores através do método de reamostragem bootstrap. Geralmente, as situações difíceis encontram-se relacionadas, principalmente, com os dados obtidos das concentrações de radionuclídeos provenientes de regiões anômalas. Pois, a distribuição de contaminantes em locais tipicamente anômalos, possui elevada assimetria à direita, causada pelos efeitos dos “outliers” (SINGH *et al.*, 1997). Por outro lado, o método de reamostragem bootstrap, quando aplicado na reamostragem dos dados originais obtidos da amostra, fornece uma média aritmética resistente às flutuações causadas pelos efeitos dos valores anômalos. Nesse caso, a reamostragem é utilizada para diminuir a assimetria, acomodando os valores de tal maneira, que a discrepância em torno da média aritmética passa a ser a menor possível (EFRON; TIBSHIRANI, 1993). Em estudos radioecológicos, não existem procedimentos estatísticos utilizados para reduzir os efeitos causados pelos valores anômalos sobre a distribuição

lognormal, o que resulta em assimetria acentuada à direita. Sendo assim, dentro deste contexto, o objetivo do presente estudo foi determinar o intervalo de confiança para a distribuição lognormal usando o método bootstrap em estudos radioecológicos. Para tanto, foram utilizadas as concentrações de ^{228}Ra em amostras de águas coletadas numa região anômala da região agreste do estado de Pernambuco.

MATERIAL E MÉTODO

O método bootstrap

Efron e Tibshirani (1993) apresentaram as ideias básicas subjacentes ao método bootstrap, no âmbito da inferência clássica da estatística, como se segue. Com $x = (x_1, x_2, \dots, x_n)$ amostra aleatória obtida a partir de uma população com função de distribuição desconhecida, F , seja, $\hat{\Theta}(x_1, x_2, \dots, x_n)$, um estimador do parâmetro $\Theta(F)$ que, como se indica, depende naturalmente de F . Seja \hat{F} a função de distribuição empírica associada à amostra obtida, tal que a cada valor observado x_i , onde $(i=1, 2, \dots, n)$, atribui massa probabilística $1/n$. Então, o valor de \hat{F} é dado pela Equação 1.

$$\hat{F}_{(n)}(x) = \frac{\left[\sum_{i=1}^n I(x_i \leq x) \right]}{n} \quad (1)$$

Onde:

$\hat{F}_{(n)}(x)$ - o estimador não-paramétrico de máxima verossimilhança de F ;

$I(x_i \leq x)$ - função indicadora.

Uma amostra bootstrap é uma amostra $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ obtida de forma aleatória e com reposição a partir da amostra original $x = (x_1, x_2, \dots, x_n)$, também designada população bootstrap. A notação com asterisco indica que x^* não é um novo conjunto de dados reais x , mas sim uma versão randomizada, ou reamostrada de x . A amostra bootstrap consiste dos correspondentes membros de x , onde: $x_1^* = x_{i1}$, $x_2^* = x_{i2}$, \dots , $x_n^* = x_{in}$. O conjunto $(x_{i1}^*, x_{i2}^*, \dots, x_{in}^*)$ representa a i -ésima amostra de tamanho n com reposição dos dados originais do conjunto $x = (x_1, x_2, \dots, x_n)$.

No método bootstrap, a média amostral calculada é denominada por \bar{x}_i . A cada procedimento de reamostragem do conjunto original $x = (x_1, x_2, \dots, x_n)$, correspondem estimadores, dados por $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. Neste caso, o estimador bootstrap da média da população é a média aritmética, \bar{x}_B , dos n estimadores \bar{x}_i . Sendo este, um método de reamostragem com reposição, pode-se ter, por exemplo: $x_1^* = x_7$, $x_2^* = x_{10}$, $x_3^* = x_2, \dots$, $x_B^* = x_3$. Portanto, o conjunto de dados reamostrados é constituído de elementos do conjunto dos dados originais $x = (x_1, x_2, \dots, x_n)$, onde alguns não aparecem nenhuma vez, outros aparecem uma vez, outros aparecem duas vezes, etc. Da distribuição $\hat{F}_{(n)}(x)$ obtêm-se B amostras bootstrap de mesmo tamanho n , como apresentada na sequência abaixo:

$$\begin{aligned} X_1^* &= [X_{11}^*, X_{12}^*, \dots, X_{1n}^*] \\ X_2^* &= [X_{21}^*, X_{22}^*, \dots, X_{2n}^*] \\ &\vdots \\ X_B^* &= [X_{B1}^*, X_{B2}^*, \dots, X_{Bn}^*] \end{aligned}$$

Neste caso, o estimador do desvio padrão da população é dado pela Equação 2.

$$s_B = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{x}_i - \bar{x}_B)^2} \quad (2)$$

Especificamente, \bar{x}_i , pode ser substituído pelo estimador $\hat{\Theta}_i$, para cada procedimento de reamostragem. A média \bar{x}_B pode também ser substituída por $\hat{\Theta}_B$, que é a média aritmética dos n estimadores bootstrap. A diferença $\hat{\Theta}_B - \hat{\Theta}_i$ é o estimador do enviesamento de $\hat{\Theta}$. Deste modo, o estimador do erro padrão de $\hat{\Theta}$ é dado pela Equação 3.

$$s_B = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\hat{\Theta}_i - \hat{\Theta}_B)^2} \quad (3)$$

onde: $\hat{\Theta}_i = \sum_{i=1}^n \frac{X_i^*}{n}$ e $\hat{\Theta}_B = \sum_{i=1}^B \frac{\hat{\Theta}_i}{B}$

Para testar o método bootstrap utilizaram-se os dados das concentrações de ^{228}Ra em amostras de águas coletadas no Rio Ipanema, mais especificamente no trecho localizado na anomalia radioativa de urânio e tório existente nas cidades de Pedra e Venturosa, em nove fazendas produtoras de leite na região agreste do estado de Pernambuco, conforme trabalho desenvolvido por Silva (2006). Por questão de anonimato, as fazendas foram descritas em códigos (F-1 a F-9).

RESULTADO E DISCUSSÃO

Na Tabela 1 estão apresentados os valores das concentrações de ^{228}Ra nas amostras de água do rio Ipanema obtidas das fazendas F-1 a F-9 localizadas na região anômala das cidades de Pedra e Venturosa. Observa-se na Tabela 1 grande variabilidade nos valores das concentrações de ^{228}Ra . As amostras de água foram provenientes de poços do tipo amazonas e tubular construídos no leito do rio Ipanema. Em particular, os valores de 430 e 2.667 mBq.L^{-1} foram provenientes de água de poço tubular, que possui profundidade maior do que o tipo amazonas. A literatura mostra que poços profundos possuem, geralmente, concentração elevada de ^{228}Ra (MICHEL; COTHEN, 1986; FISHER *et al.*, 2000; OLECKNO *et al.*, 2001).

Tabela 1. Concentrações de ^{228}Ra nas amostras de água das fazendas F-1 a F-9.

Código da fazenda	Concentração (mBq.L ⁻¹)
F-1	116 ± 6
F-1	436 ± 22
F-2	82 ± 4
F-2	47 ± 5
F-3	130 ± 7
F-3	390 ± 20
F-4	82 ± 9
F-5	132 ± 12
F-6	2.667 ± 150
F-7	189 ± 15
F-8	71 ± 10
F-9	104 ± 8

De acordo com Singh *et al.* (1997), as concentrações de elementos químicos em locais tipicamente anômalos podem variar desde o *background* até valores extremamente elevados. Isto provavelmente explica a grande variação nos resultados apresentados na Tabela 1. A variabilidade existente na Tabela 1 pode ser decorrente também de diferentes pontos de afloramento de ^{232}Th no aquífero subterrâneo do rio Ipanema. Devido à elevada variação nos valores observados na Tabela 1, certamente, diferentes fatores naturais afetaram a migração do ^{228}Ra nas águas do aquífero das fazendas F-1 a F-9. Segundo Baker e Toque (2005), diversos fatores naturais interferem na distribuição do rádio em águas. Assim, tentar explicar a migração de ^{228}Ra para as águas, não é uma tarefa fácil, principalmente, num aquífero como o do rio Ipanema que, devido à elevada porosidade do solo tem suas águas totalmente transportadas para o lençol subterrâneo durante a estação seca que ocorre na região agreste de Pernambuco. Como as águas desse rio são superficiais no inverno e subterrâneas no verão, existem mudanças extremas no comportamento de migração do ^{228}Ra . Daí, provavelmente, as variações elevadas nos resultados das concentrações apresentadas na Tabela 1. Assim, as complexidades ambientais relacionadas com a migração do ^{228}Ra para as águas do rio Ipanema causaram elevadas flutuações nos valores das concentrações, como mostra a Figura 1. Aplicando o teste de Kolmogorov-Smirnov no conjunto de dados da Tabela 1, verificou-se que os valores das concentrações não possuem distribuição normal (Figuras 2 e 3). Por outro lado, as Figuras 4, 5 e 6 mostram que os dados da Tabela 1 não se ajustaram perfeitamente à distribuição lognormal, em decorrência da influência dos valores outliers.

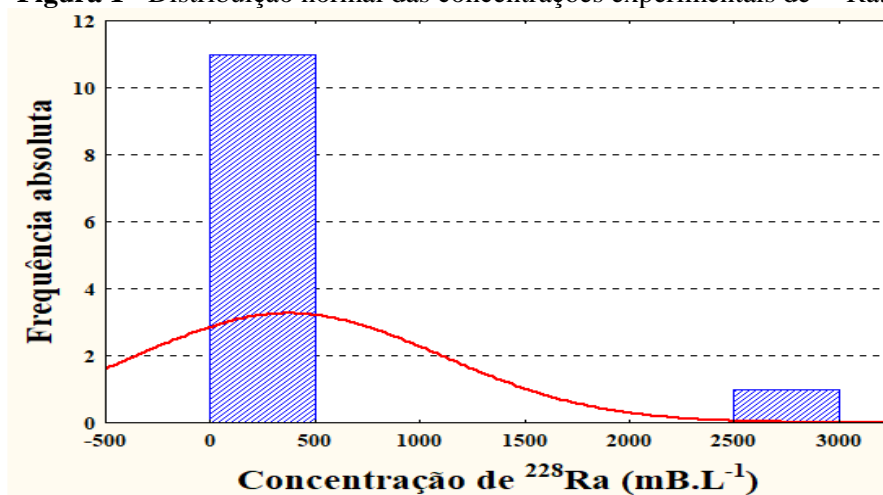
Figura 1 - Distribuição normal das concentrações experimentais de ^{228}Ra .

Figura 2 - Gráfico probabilidade-probabilidade da distribuição normal das concentrações experimentais de ^{228}Ra .

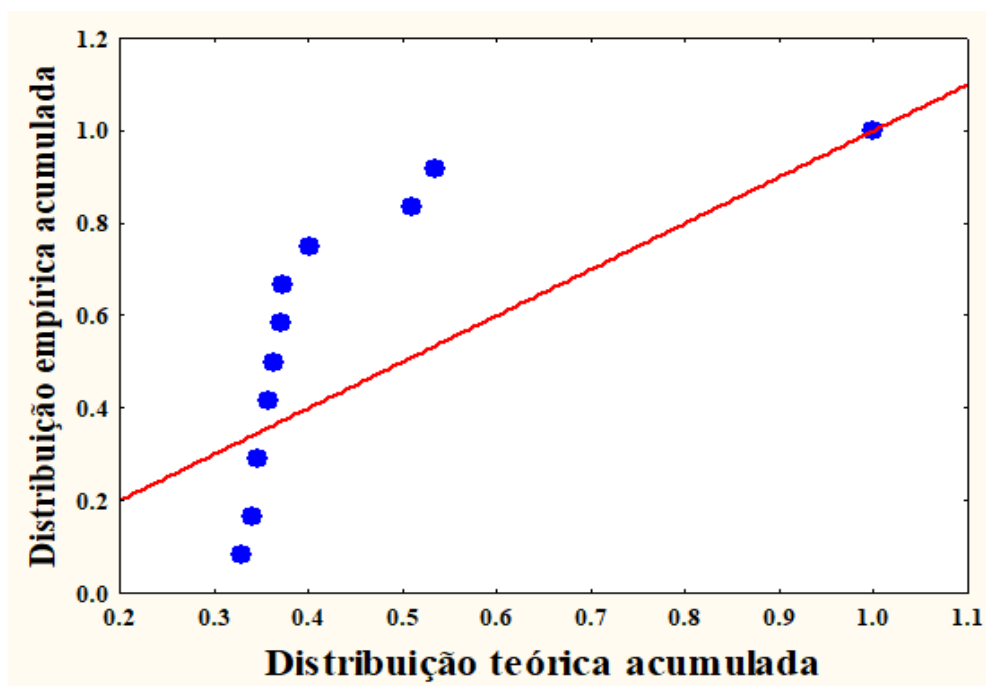


Figura 3 - Gráfico quartil-quartil da distribuição normal das concentrações experimentais de ^{228}Ra .

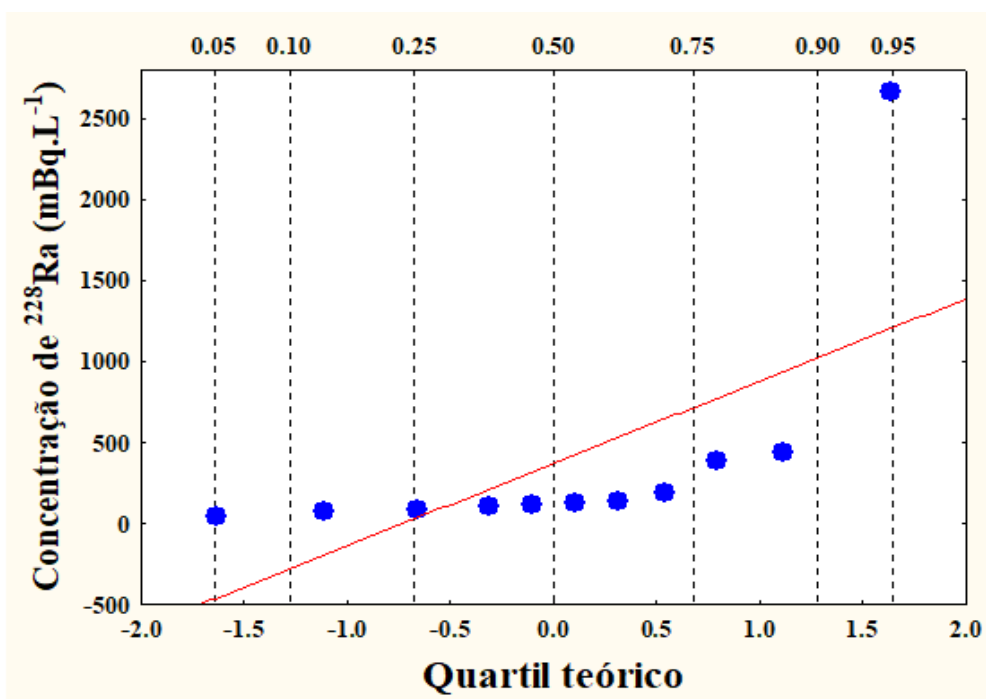


Figura 4 - Distribuição lognormal das concentrações experimentais de ^{228}Ra .

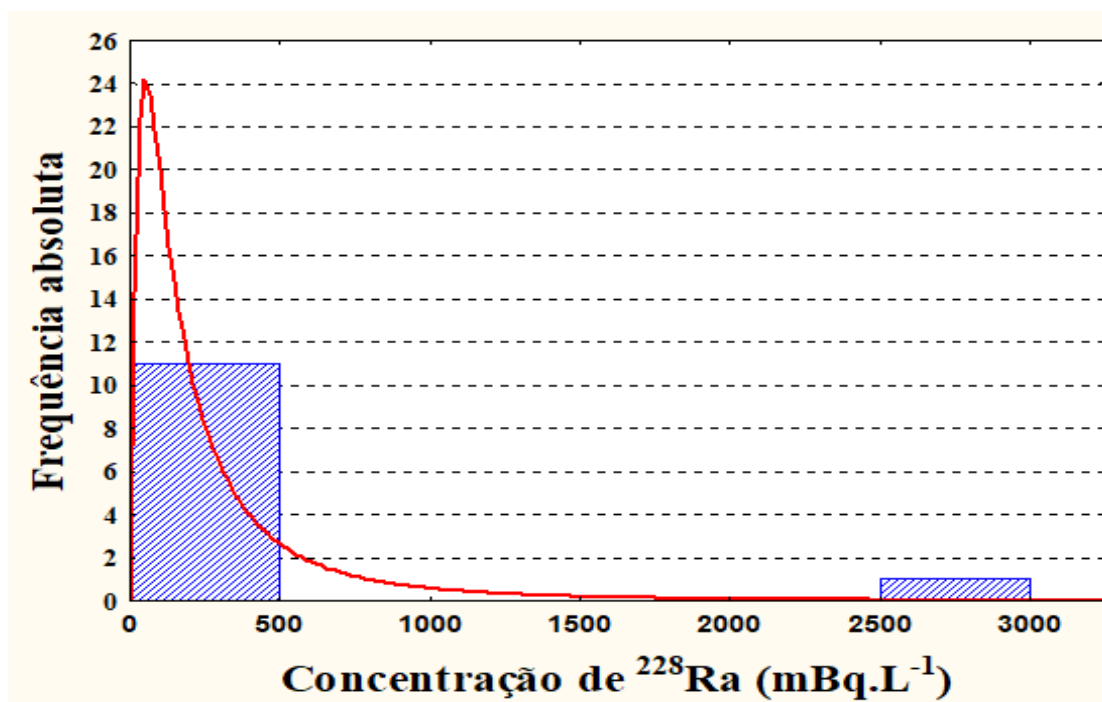


Figura 5 - Gráfico probabilidade-probabilidade da distribuição lognormal das concentrações experimentais de ^{228}Ra .

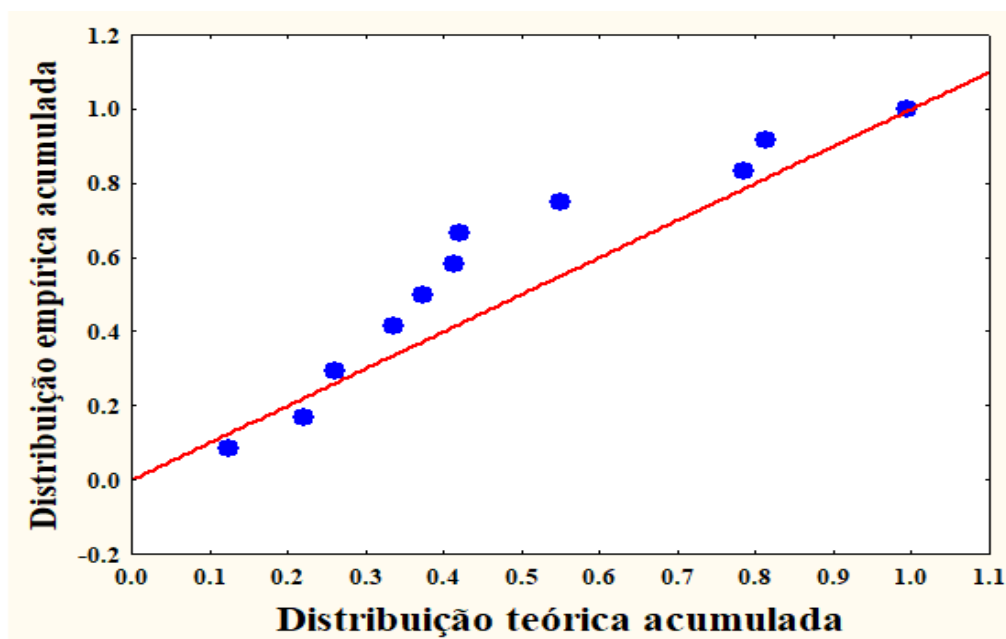
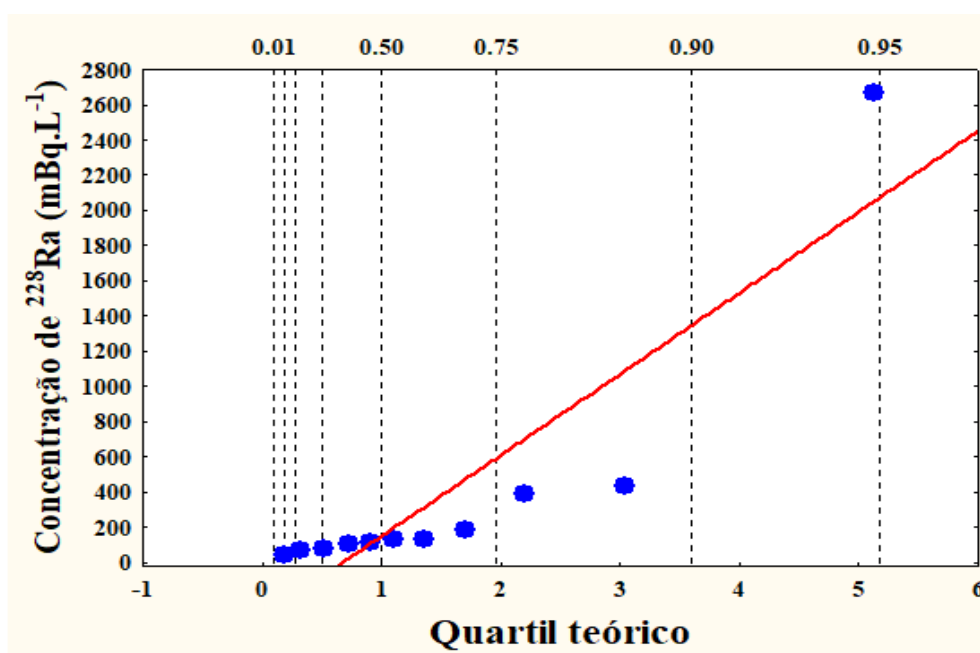


Figura 6 - Gráfico quartil-quartil da distribuição lognormal das concentrações experimentais de ^{228}Ra .



Entretanto, muitas avaliações na teoria da amostragem são de uso limitado quando o conjunto dos dados é pequeno. Por exemplo, a estimativa do desvio padrão para n valores normalmente distribuídos, é muito sensível às flutuações quando n é pequeno do que quando ele é grande (HELENE; VANIN, 2002). Entretanto, para resolver tal problema, é necessário coletar um número grande de amostras, com a finalidade de diminuir as flutuações estatísticas no conjunto de dados. Por outro lado, muitas vezes por questões financeiras e limitações das condições ambientais, nem sempre é possível fazer coletas de um número grande de amostras. A lei dos Grandes Números garante que, quando $n = 30$, existe a probabilidade de que os dados possuam distribuição normal (SPIEGEL, 1978). Porém, Silva *et al.* (2007) verificaram que, mesmo para $n > 30$, a condição de normalidade não é garantida quando existe valores anômalos, sendo a média aritmética inadequada para representar o conjunto de dados. No caso dos dados apresentados na Tabela 1, observa-se que, além da existência de valores anômalos, a quantidade de amostra foi menor do que 30. Isto dificulta muito uma análise estatística adequada, muitas vezes limitada apenas à discussão de faixa de variação dos valores, ou no muito, uma estimativa usando a mediana. Em estudos radioecológicos (RIBEIRO *et al.*, 2018; YAWODOYI *et al.*, 2018), a mediana é bastante utilizada, em decorrência de não ser afetada pelos valores anômalos, sendo, portanto, a medida de tendência central preferida. Por outro lado, existe um grande problema estatístico que precisa ser resolvido. A mediana não representa adequadamente um conjunto de dados, pois, ela é uma medida de tendência central menor do que a média aritmética (SINGH *et al.*, 1997). Isto causaria uma subestimativa, pois os valores extremos são discriminados pela mediana. Nesse caso, a mediana não representa adequadamente os valores do conjunto de dados. O método bootstrap é uma alternativa bastante viável para resolver este problema, pois, o Teorema do Limite Central garante que, quando “ n ” for suficientemente grande, os estimadores da média e do desvio padrão são normalmente distribuídos e convergem para os seus valores verdadeiros (HELENE; VANIN, 2002). Porém, usando o método bootstrap, é possível obter uma média aritmética resistente às flutuações dos valores anômalos. A Figura 7 mostra a distribuição normal das concentrações bootstrap do ^{228}Ra para 2000 interações. O teste de Kolmogorov-Smirnov mostrou que as concentrações bootstrap possuem distribuição normal. As Figuras 8

e 9 mostram que as influências dos valores anômalos no conjunto de dados foram significativamente atenuadas, não afetando a forma da distribuição (Figura 7). Assim, as análises estatísticas baseadas na distribuição bootstrap, são notoriamente robusta, conforme observou Efron e Tibshirani (1993). Na Tabela 2 estão apresentados os valores de medidas de tendência central para a distribuição normal das concentrações experimentais e bootstrap. Observa-se na Tabela 2, grande dispersão nas medidas de tendência central experimental. Isto mostra que, estas grandezas foram inadequadas para representar o conjunto de dados da Tabela 1. Por outro lado, a dispersão em torno das medidas de tendência central bootstrap foram significativamente reduzidas. Entretanto, a média aritmética bootstrap foi quase três vezes maior do que a mediana bootstrap, e muito próxima da média aritmética experimental. Assim, dentre as medidas de tendência central analisadas, a média aritmética bootstrap foi a mais adequada para representar o conjunto de dados da Tabela 1.

Tabela 2. Medidas de tendência central experimental e bootstrap.

Estatística	Concentração de ^{228}Ra (mBq.L$^{-1}$)
Média aritmética experimental \pm desvio padrão	370,50 \pm 733,80
Mediana experimental \pm desvio quartílico	123,00 \pm 157,25
Média aritmética bootstrap \pm desvio padrão	369,76 \pm 161,9
Mediana bootstrap \pm desvio quartílico	131,77 \pm 29,62

Figura 7 - Distribuição normal das concentrações bootstrap de ^{228}Ra

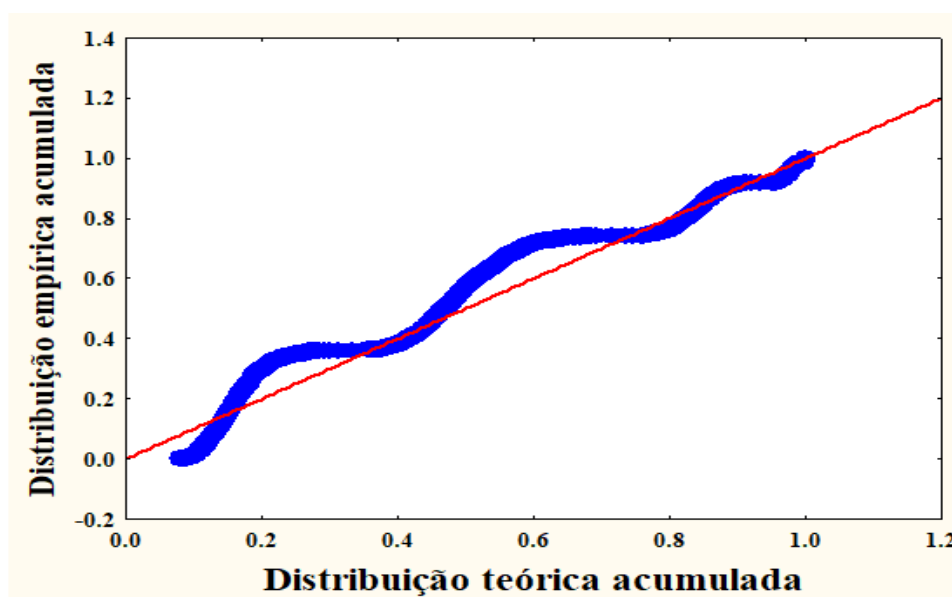


Figura 8 - Gráfico probabilidade-probabilidade da distribuição normal das concentrações bootstrap de ^{228}Ra .

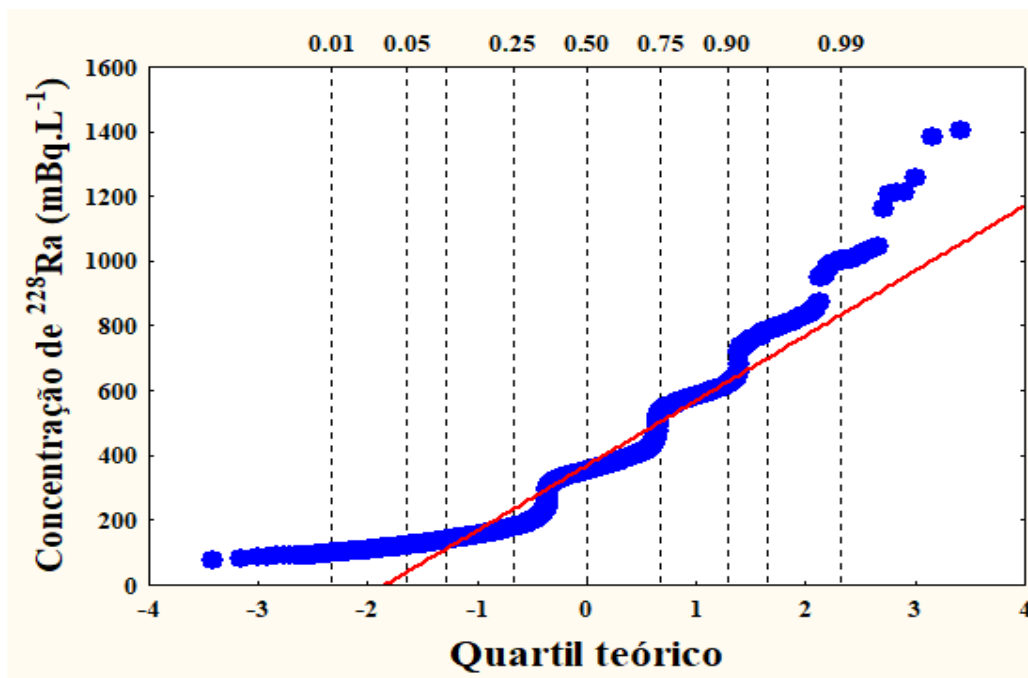
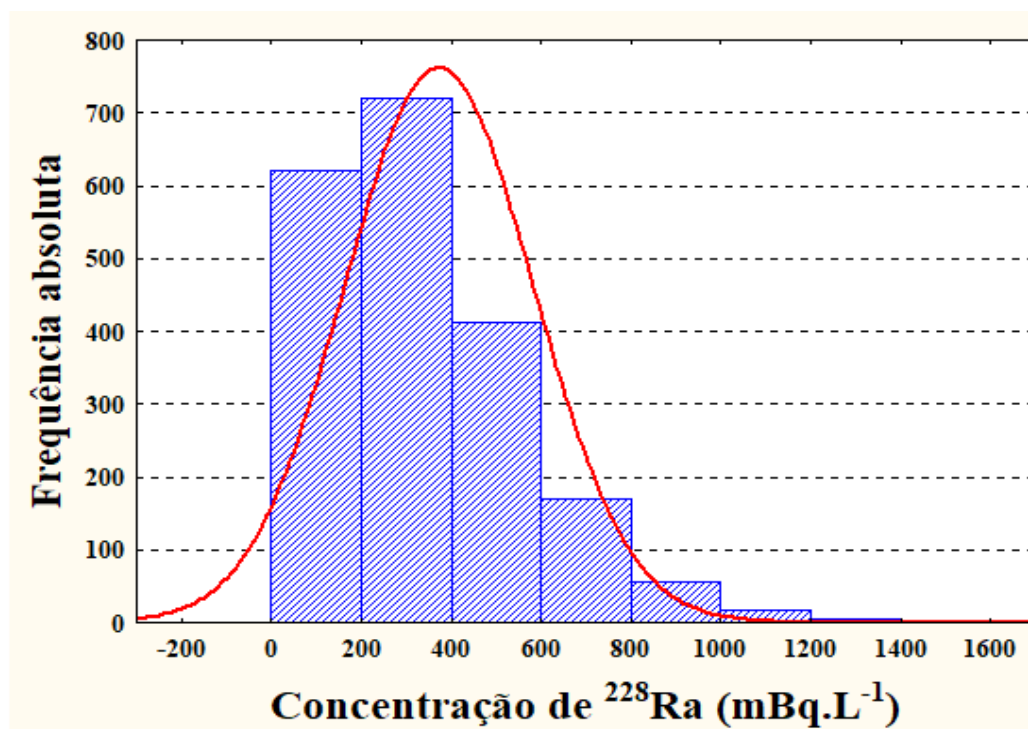


Figura 9 - Gráfico quartil-quartil da distribuição normal das concentrações bootstrap de ^{228}Ra .



A Figura 10 mostra a distribuição bootstrap dos dados apresentados na Tabela 1, em relação à distribuição lognormal com 2000 iterações. Comparando as Figuras 4 e 10, observa-se grande diferença na forma da distribuição. Isto mostra um ajuste bastante significativo dos dados bootstrap à distribuição lognormal, podendo assim, ser confirmado pelas Figuras 11 e 12. Apesar do ajuste significativo dos dados bootstrap à distribuição normal e lognormal, observou-se grande diferença entre as Figuras 9 e 12. Nesse

caso, a análise quartílica mostrou que, mesmo para um número grande de amostras, a mediana é uma medida de tendência central inadequada para representar um conjunto de dados com valores anômalos. Assim, tanto na Figura 9, como na Figura 12, observa-se que a mediana discrimina os valores extremos, sendo, portanto, uma medida de tendência central inadequada para representar conjuntos de dados provenientes de estudos radioecológicos em locais tipicamente anômalos.

Figura 10 - Distribuição lognormal das concentrações bootstrap de ^{228}Ra

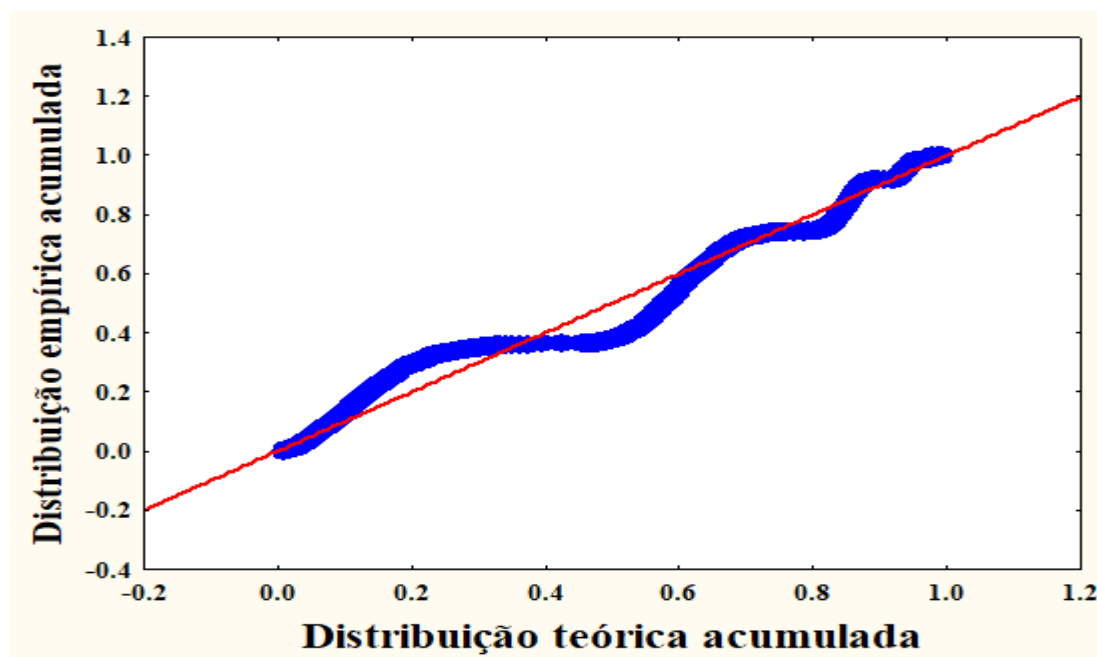


Figura 11 - Gráfico probabilidade-probabilidade da distribuição lognormal das concentrações bootstrap de ^{228}Ra .

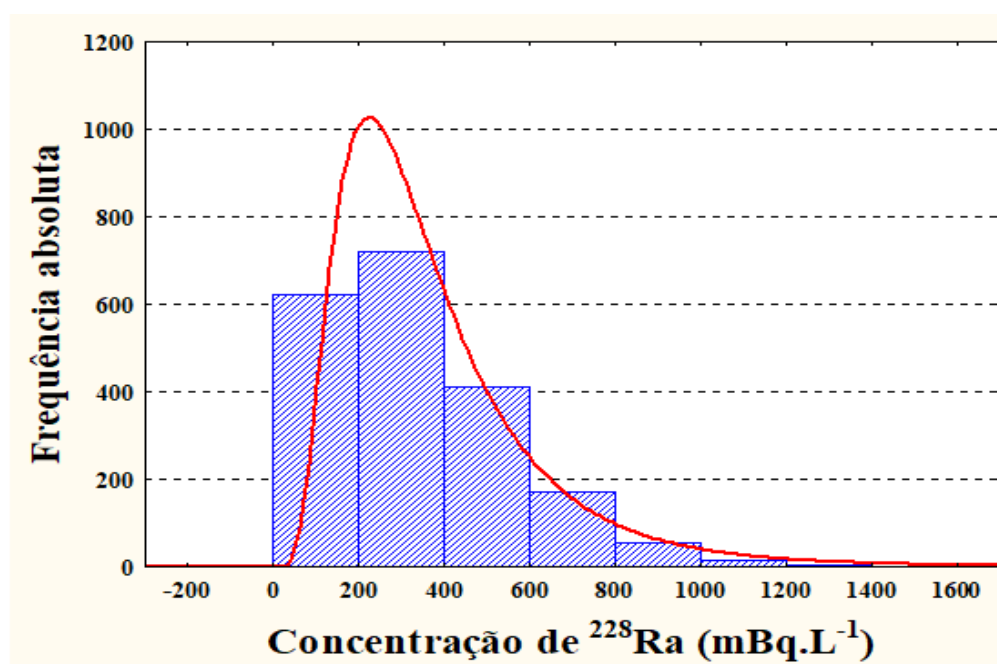
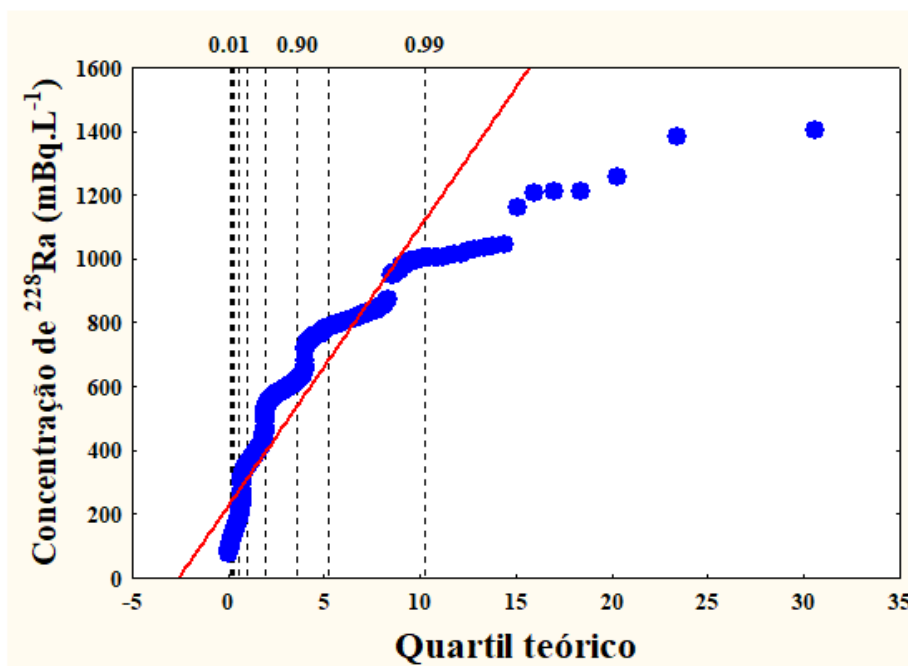


Figura 12 - Gráfico quartil-quartil da distribuição lognormal das concentrações bootstrap de ^{228}Ra 

Singh et al. (1997) mostram a importância de se determinar o limite superior do intervalo de confiança da distribuição lognormal, em estudos estatísticos sobre contaminantes ambientais. Porém, com o método bootstrap, foi possível verificar a aplicabilidade da distribuição lognormal na determinação do limite superior do intervalo de confiança para os dados apresentados na Tabela 1. Tomando $H_{1-\alpha} = H_{0,95}$ na tabela da estatística-H com $\alpha = 0,05$, temos para $\theta = 0,6$ (desvio padrão amostral) e $n = 12$ (número de amostra da Tabela 1), $H_{0,95} = 2,271$ (GILBERT, 1987). Com isto calculou-se o limite superior do intervalo de confiança da distribuição lognormal. Os valores da estatística da distribuição lognormal bootstrap estão apresentados na Tabela 3. Nesse caso, a mediana foi menor do que a média, confirmando assim que, mesmo para um número elevado de amostra, a mediana foi uma medida de tendência central inadequada para representar o conjunto de dados da Tabela 1. Por outro lado, a média da distribuição lognormal bootstrap foi maior (Tabela 3) do que a média da distribuição normal bootstrap (Tabela 2). Além disso, a dispersão em torno da média da distribuição lognormal bootstrap (Tabela 3) foi menor do que na média da distribuição normal bootstrap (Tabela 2). Assim, o limite superior do intervalo de confiança da média da distribuição lognormal bootstrap, pode ser eficientemente utilizado em estudos radioecológicos com dados anômalos, como aqueles apresentados na Tabela 1.

Tabela 3. Estatística da distribuição lognormal bootstrap.

Estatística	Concentração de ^{228}Ra (mBq.L⁻¹)
Limite superior	665,18
Média \pm desvio padrão	444,08 \pm 25,6
Mediana	313,15

CONCLUSÃO

O método de reamostragem bootstrap foi bastante eficiente para determinar medidas de tendência

central em análise estatística de dados radioecológicos anômalos. A aplicação do método bootstrap no conjunto de dados permitiu calcular o limite superior robusto para o intervalo de confiança da distribuição lognormal.

REFERÊNCIAS

BAKER, A.C; TOQUE, C.A review of the potential for radium from luminising activities to migrate in the environment. **Journal of Radiological Protection**, v. 25, p. 127-140, 2005.

EFRON, B., TIBSHIRANI, R.J. **An introduction to the bootstrap**, New York: Chapman e Hall, 1993.

FISHER, E. L.et al. Dissolution of ^{226}Ra from pipe-scale deposits in a public water supply. **Environment International**, v. 26, p. 69-73, 2000.

GILBERT, R.O. **Statistical Methods for Environmental Pollution Monitoring**, Washington: John Wiley & Sans, 1987.

HELENE, O.; VANIN, V.R. Analysis of discrepant data using a bootstrap procedure. **Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment**, v. 481, p. 626-631, 2002.

MICHEL, J.; COTHEN, C.R. Predicting the occurrence of ^{228}Ra in ground water. **Health Physics**, v. 51(6), p. 715 - 721, 1986.

OLECKNO, W.A.; NATHAN, L.M.; ANDERSON, R.B. Achieving compliance with the radium standards for drinking water in Midwestern community: a case study. **Journal Environment Health**, v. 63, n. 8, p. 9-15, 2001.

RIBEIRO, F.C.A. *et al.* Natural radioactivity in soils of the state of Rio de Janeiro (Brazil): Radiological characterization and relationships to geological formation, soil types and soil properties. **Journal of Environmental Radioactivity**, v. 182, p. 34-43, 2018.

SILVA, C.M. **Ra-226 e Ra-228 na dieta de bovinos leiteiros do Agreste Semiárido de Pernambuco e avaliação de risco decorrente do consumo de leite por uma população potencialmente exposta**. Tese de Doutorado do Programa de Pós-graduação em Tecnologias Energéticas e Nucleares da Universidade Federal de Pernambuco, 2006. 152p.

SILVA, C.M.et al. Application of bootstrap method for evaluating discrepant levels of radium-226 in forage palm (*Opuntia spp*). **Revista Brasileira de Biometria**, v.25, n.3, p.109-114, 2007.

SINGH, A.K., SINGH, A., ENGELHARDT, M. The lognormal distribution in environmental applications. **Technology Support Center Issue**, EPA/600/R-97/006, 1997.

SPIEGEL, M. R. **Probabilidade e estatística**, São Paulo: McGraw-Hill, 1978.

TOLEDO, G.L.; OVALLE, I.I. **Estatística básica**, 2 ed. São Paulo: Atlas, 1983.

YAWODOYI, I.N. *et al.* Soil-to-cassava transfer of naturally occurring radionuclides from communities along Ghana's oil and gas rich Tano Basin. **Journal of Environmental Radioactivity**, v. 182, p. 138-141, 2018.