

Desenvolvimento de uma Função Através de Regressão Linear Múltipla para Estimativa da Precipitação Média Anual na Região Metropolitana do Recife

Development of a Function Through Multiple Linear Regression to Estimate the Average Annual Rainfall In the Metropolitan Region of Recife

Desarrollo de una Función Mediante Regresión Lineal Múltiple para Estimar la Precipitación Promedio Anual en La Región Metropolitana de Recife

Gabriela Torres Gonçalves

Mestranda em Engenharia Civil, UPE, Brasil
gtg@poli.br

José Mário Torres Antonino

Mestrando em Engenharia Civil, UPE, Brasil
jmta@poli.br

Williames de Albuquerque Soares

Professor Doutor, UPE, Brasil
was@poli.br

Micaella Raissa Falcão de Moura

Professora Doutora, UNICAP, Brasil
micaella.moura@unicap.br

Ana Carolina Silva Wanderley

Graduanda em Engenharia Civil, UNICAP, Brasil
ana.2020107292@unicap.br

RESUMO

Nos grandes centros urbanos é primordial aprender a conviver e se adaptar aos frequentes eventos climáticos extremos. Considerando o exposto, este artigo teve como objetivo utilizar um modelo probabilístico, através da regressão linear múltipla, para determinar uma função correlacionando a precipitação anual média de vinte postos pluviométricos localizados na Região Metropolitana do Recife com a Latitude, Longitude e Altitude desses postos. A função obtida mostrou uma alta capacidade preditiva, apresentando coeficiente de determinação da regressão de 0,713 e a média do erro percentual, entre os dados observados e estimados dos postos pluviométricos estudados, foi 6,27%.

PALAVRAS-CHAVE: precipitação, estatística, clima.

ABSTRACT

In large urban centers, it is essential to learn to live together and adapt to frequent extreme weather events. For this reason, this article aimed to use a probabilistic model, through multiple linear regression, which can determine a function correlating the average annual rainfall of twenty pluviometric stations located in the Metropolitan Region of Recife and the Latitude, Longitude and Altitude of these stations. The function obtained showed a high predictive capacity, presenting a coefficient of determination of the regression of 0.713 and the average of the percentage error, between the observed and estimated data of the pluviometric stations studied, was 6.27%

KEYWORDS: rainfall, statistic, climate.

RESUMEN

En los grandes centros urbanos es fundamental aprender a convivir y adaptarse a los frecuentes fenómenos meteorológicos extremos. Por esta razón, este artículo tuvo como objetivo utilizar un modelo probabilístico, a través de regresión lineal múltiple, que puede determinar una función que correlacione la precipitación media anual de veinte estaciones pluviométricas ubicadas en la Región Metropolitana de Recife con la Latitud, Longitud y Altitud de estas estaciones. La función obtenida mostró una alta capacidad predictiva, presentando un coeficiente de determinación de la regresión de 0,713 y la media del porcentaje de error, entre los datos observados y estimados de las estaciones pluviométricas estudiadas, fue de 6,27%.

PALABRAS CLAVE: precipitación, estadísticas, clima.

1 INTRODUÇÃO

Com o crescente aumento de eventos climáticos extremos, resultado da ação antrópica, associados a fenômenos climáticos como “La Niña” e Distúrbios Ondulatórios de Leste, fica cada vez mais urgente a utilização de métodos que possam estimar a precipitação pluviométrica em determinado local. Entre as cidades mais vulneráveis às consequências desses fenômenos está a cidade do Recife, capital de Pernambuco (IPCC, 2022).

Nesse sentido, destaca-se o último evento climático extremo que atingiu fortemente a Região Metropolitana do Recife (RMR). Foi registrado, no período de 20/05/22 a 10/06/22, um acumulado de 926,53 mm de chuva na estação pluviométrica do Córrego do Jenipapo, por exemplo, e uma média de aproximadamente 829 mm de chuva entre as estações pluviométricas da Cidade do Recife, segundo o CEMADEN - Centro Nacional de Monitoramento e Alertas de Desastres Naturais.

De acordo com a presente pesquisa, nos últimos nove anos, a Cidade do Recife apresentou uma precipitação média anual de cerca de 1.850 milímetros de chuva, sendo a precipitação dos 20 dias citados anteriormente (20/05/22 a 10/06/22) equivalente a 50% da precipitação média dos últimos nove anos. Não diferente do cenário nacional, a Região Metropolitana do Recife apresenta um crescimento desordenado. Somado a isto, a impermeabilização do solo, a canalização e estreitamento das seções de rios e riachos, bem como os mencionados eventos climáticos extremos, resultam em enchentes e deslizamentos de terras (SILVA JUNIOR *et al.*, 2020). Esses problemas urbanos geram grandes consequências econômicas, ambientais e sociais, e a cada vez que acontecem, pessoas são desabrigadas e vidas são perdidas.

Diante do exposto, o presente artigo objetivou utilizar a regressão linear múltipla para avaliação da eficácia de uma função aplicável para estimar a precipitação média anual, fazendo uso de variáveis de caracterização espacial, como latitude, longitude e altitude na RMR. Este tipo de pesquisa se faz de extrema relevância considerando as exposições supracitadas e, conforme estudo de Alemu e Bawoke (2020), contribui para afirmar que compreender a distribuição das chuvas no espaço e no tempo é crucial para a gestão sustentável dos recursos hídricos e a produtividade agrícola, aspecto também bastante importante para a região em estudo.

2 MATERIAIS E MÉTODOS

2.1 DESCRIÇÃO DA ÁREA DE ESTUDO

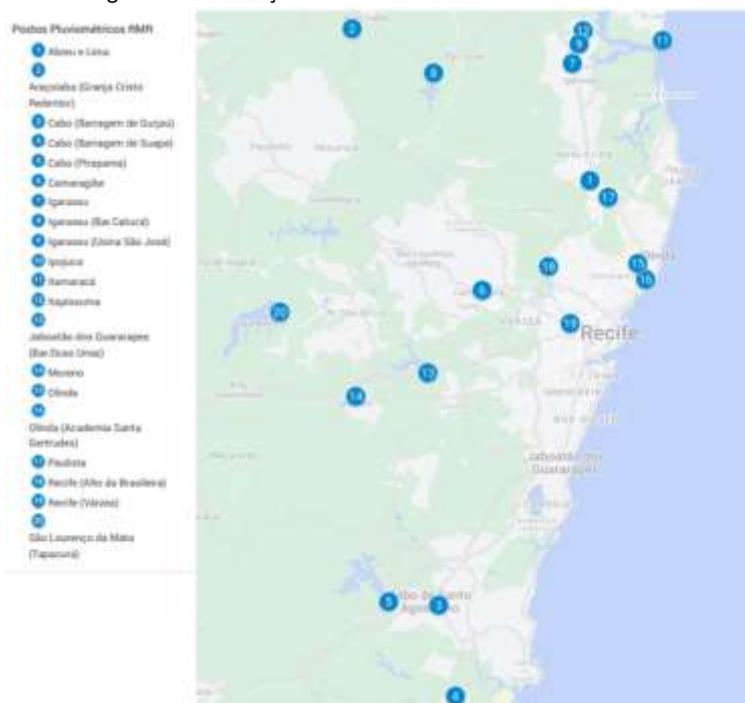
De acordo com IPEA (2015), a Região Metropolitana do Recife tem população de 3,69 milhões de pessoas (IBGE, 2010) e quatorze municípios a compõem: Araçoiaba, Igarassu, Itapissuma, Ilha de Itamaracá, Abreu e Lima, Paulista, Olinda, Camaragibe, Recife, Jaboatão dos Guararapes, São Lourenço da Mata, Moreno, Cabo de Santo Agostinho e Ipojuca.

O Serviço Geológico do Brasil - CPRM acrescenta ainda que os totais anuais médios de precipitação nos municípios da RMR variam de mais de 2.200 mm, nas áreas litorâneas, decrescendo à medida que se afastam do litoral até valores próximos a 1.200 mm, na parte oeste do município de São Lourenço da Mata. As variações interanuais são também

significativas, oscilando entre 1.200 a 3.500 mm no litoral e entre 500 mm a 2.000 mm nas áreas mais afastadas do oceano. O trimestre mais chuvoso, que engloba os meses de maio, junho e julho, concentra 47% dos totais anuais. No mais seco, que ocorre nos meses de outubro, novembro e dezembro, as chuvas representam apenas 7,5% dos totais anuais (CPRM, 2003).

Diante do exposto acima, esta pesquisa utilizou dados a partir de 20 postos pluviométricos da APAC – Agência Pernambucanas de Águas e Clima, na RMR, que apresentavam os dados de Latitude, Longitude, Altitude e Precipitação anual completos. Vale salientar que todos os municípios da RMR foram contemplados, com pelo menos um posto pluviométrico no estudo, assim representados na Figura 1 a seguir.

Figura 1 – Localização dos Pontos Pluviométricos Estudados



Fonte: elaborado pelos autores

2.2 REGRESSÃO LINEAR MÚLTIPLA

O objetivo da regressão linear múltipla é elaborar um modelo probabilístico que determine uma variável dependente (y) através de mais de uma variável independente ou de previsão (x). A equação que representa a regressão linear múltipla é:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

Os β s são os coeficientes de regressão reais (ou populacionais). O coeficiente de regressão β_1 é interpretado como a mudança esperada em Y associada com um aumento de uma unidade em x_1 , enquanto x_2, \dots, x_k são mantidos fixos (DEVORE, 2006).

Grande parte dos problemas práticos envolve mais de uma variável para fins de previsão. Por exemplo, caso alguém desejasse prever uma variável (y) por meio de outras variáveis (x_1 e x_2), o problema seria descobrir o plano que melhor se ajustasse, na questão dos mínimos quadrados, a um diagrama de dispersão de pontos em três dimensões (FILHO, 2002).

Assim, segundo Cruz et al. (2013), a partir de um conjunto de n valores da variável Y , associados às n observações correspondentes das p variáveis independentes pode-se definir os coeficientes da regressão pelo método dos mínimos quadrados.

Na regressão linear, para quantificar a força da relação entre as variáveis, analisa-se:

- a covariância: mede o grau de interdependência entre duas variáveis (variáveis independentes possuem covariância zero);
- o coeficiente de correlação: identifica a direção e a intensidade da relação linear entre duas variáveis quantitativas; e
- o diagrama de dispersão: representa graficamente a associação entre pares de dados.

Para analisar a “força” do modelo estimado por meio da regressão linear, faz-se uso do chamado “R-quadrado”, ou seja, do coeficiente de determinação da regressão (r^2). O R-quadrado permite notar se o relacionamento entre as variáveis é forte ou fraco. Especificamente na regressão linear múltipla, a medida utilizada é o R-quadrado ajustado.

$$R^2_{ajustado} = \frac{(n-1)R^2 - k}{n - (k+1)} \quad (2)$$

No presente estudo, foi proposta a realização de uma análise por regressão linear múltipla, considerando a precipitação média anual para os postos disponíveis como a variável dependente e sugeridas três variáveis independentes, levando em conta a variabilidade espacial na Região Metropolitana do Recife. As variáveis independentes analisadas foram: latitude, longitude e altitude, dados relativamente fáceis de se obter para qualquer usuário.

As variáveis “latitude” e “longitude” foram consideradas em decimais e a altitude em metros. Assim, inicialmente, foi testado o seguinte formato de função para estimativa da precipitação média anual para a RMR:

$$\text{Prec. média} = \beta_0 + \beta_1 * \text{Latitude} + \beta_2 * \text{Longitude} + \beta_3 * \text{Altitude} \quad (3)$$

3 RESULTADOS E DISCUSSÃO

Com base nas séries históricas dos 20 postos pluviométricos distribuídos na RMR, consolidou-se o Quadro 1 a seguir, para então partir para a verificação da relação entre a variável dependente (precipitação média anual) e as variáveis independentes (latitude, longitude e altitude). Realizaram-se procedimentos de regressão linear múltipla e cálculo do coeficiente de determinação.

Quadro 1- Postos Pluviométricos na Região Metropolitana do Recife com série histórica de nove anos (2013-2021)

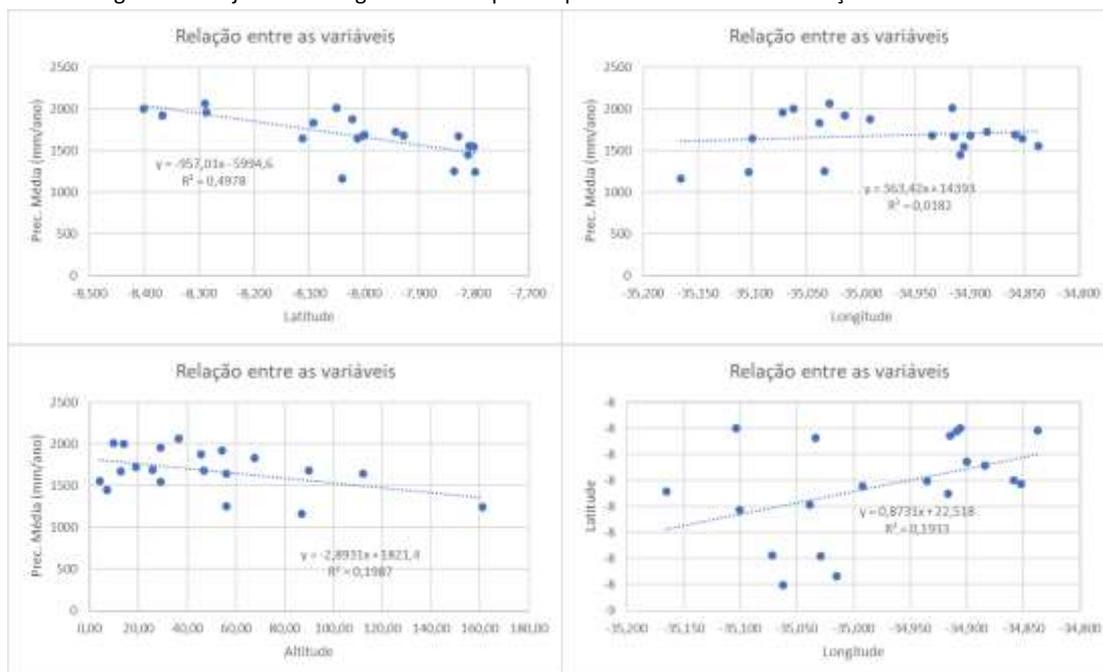
ID	Postos Pluviométricos	Município	Lat	Long	Alt (m)	Prec. Média (mm/ano)
1	Abreu e Lima	Abreu e Lima	-7,928	-34,900	46,84	1682,52
2	Araçoiaba (Granja Cristo Redentor)	Araçoiaba	-7,799	-35,104	161,08	1246,06
3	Cabo (Barragem de Gurjaú)	Cabo	-8,290	-35,029	36,36	2062,52
4	Cabo (Barragem de Suape)	Ipojuca	-8,368	-35,015	54,24	1920,52
5	Cabo (Pirapama)	Cabo	-8,287	-35,072	29,13	1955,37
6	Camaragibe	Camaragibe	-8,022	-34,992	45,80	1874,71
7	Igarassu	Igarassu	-7,828	-34,915	12,84	1673,51
8	Igarassu (Bar.Catucá)	Igarassu	-7,836	-35,034	55,96	1255,86
9	Igarassu (Usina São José)	Igarassu	-7,812	-34,909	6,99	1453,77
10	Ipojuca	Ipojuca	-8,401	-35,062	13,88	2004,13
11	Itamaracá	Itamaracá	-7,809	-34,838	4,08	1552,57
12	Itapissuma	Itapissuma	-7,800	-34,906	29,00	1545,88
13	Jaboatão dos Guararapes (Bar.Duas Unas)	Jaboatão dos Guararapes	-8,092	-35,039	67,78	1837,53
14	Moreno	Moreno	-8,112	-35,100	112,24	1647,39
15	Olinda	Olinda	-7,999	-34,859	25,72	1695,09
16	Olinda (Academia Santa Gertrudes)	Olinda	-8,013	-34,852	56,08	1650,20
17	Paulista	Paulista	-7,943	-34,884	18,84	1727,14
18	Recife (Alto da Brasileira)	Recife	-8,001	-34,935	90,01	1682,12
19	Recife (Várzea)	Recife	-8,050	-34,917	9,88	2010,19
20	São Lourenço da Mata (Tapacurá)	São Lourenço da Mata	-8,040	-35,165	87,00	1163,01

Fonte: Elaborado pelos autores.

A partir dos dados amostrais consolidados, inicialmente foram elaborados diagramas de dispersão para visualizar possíveis relações entre as variáveis. Como é possível observar através dos gráficos apresentados na Figura 2, o maior R-quadrado das regressões lineares

simples é 0,4978. Logo, na regressão simples nem 50% variações de precipitação média podem ser explicadas por alguma das variáveis estudadas.

Figura 3 – Conjunto de Diagramas de Dispersão para Analisar Possíveis Relações entre as Variáveis



Fonte: Elaborado pelos autores.

Dessa forma, optou-se pelo uso da regressão linear múltipla com as três variáveis independentes citadas, por meio do Excel, especificamente a ferramenta Análise de Dados – Regressão. O resultado desta regressão múltipla gerou diversos parâmetros, dentre eles o R-quadrado ajustado de 0,706 e o coeficiente de correlação (R múltiplo) de 0,867. Essas informações induzem a pensar que as variáveis possuem boa correlação entre si, e que 70,6% das variações de precipitação média podem ser explicadas pela variabilidade da latitude, longitude e altitude.

No entanto, é preciso analisar a tabela ANOVA, também resultado da regressão. Neste sentido, tem-se que o valor-p referente ao coeficiente da variável altitude é igual a aproximadamente 0,465, um valor superior a 5% (alfa considerado), indicando que não há relação linear entre as variáveis. Dada esta constatação, retirou-se a altitude das variáveis, e utilizando o Excel, a regressão múltipla foi gerada novamente (apenas com duas variáveis independentes: latitude e longitude).

O resultado do novo modelo de regressão apresentou o R-quadrado ajustado de 0,713, ligeiramente superior ao anterior, e o coeficiente de correlação (R múltiplo) de 0,862. Pode-se afirmar que 71,3% das variações nas precipitações médias são explicadas por variações nas

latitudes e longitudes, um índice muito superior ao encontrado através de regressão linear simples.

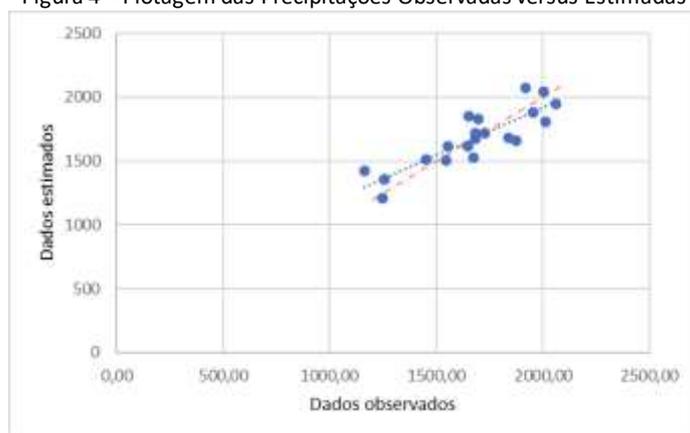
Analisando a tabela ANOVA, observou-se que os valores-p dos coeficientes referentes às duas variáveis independentes são inferiores a 5%, indicando que há relação entre as variáveis, e por isso devem permanecer na função resultado da regressão, conforme segue:

$$\text{Prec. Média} = 43.352,16708 - 1.286,058655 * \text{Latitude} + 1.486,325298 * \text{Longitude} \quad (4)$$

A equação (4) fornece uma estimativa da precipitação média anual por meio dos parâmetros de latitude e longitude. Sabendo que os dados de latitude e longitude da região estudada são negativos, quanto maior o valor da latitude (em módulo), maior é a precipitação média anual estimada; de forma contrária, a variável dependente diminui quanto maior for o valor da longitude (em módulo).

Substituindo os dados levantados dos 20 postos pluviométricos da RMR, é possível estimar a precipitação média através desta função. A qualidade da estimativa realizada pela equação (4) é representada graficamente através da Figura 3 abaixo.

Figura 4 – Plotagem das Precipitações Observadas versus Estimadas



Fonte: Elaborado pelos autores.

Considerando a reta vermelha como a reta com ângulo de 45 graus, nota-se que há uma aproximação da reta formada pelos dados estimados e observados (reta azul) com esta reta de ajuste linear (vermelha), o que caracteriza a estimativa realizada por meio do modelo de regressão linear múltipla como bastante assertiva e próxima aos dados reais.

4 CONCLUSÃO

As latitudes e longitudes foram identificadas como componentes significativos para a estimativa da precipitação média anual da Região Metropolitana do Recife. O Modelo Proposto demonstrou uma excelente performance nesta estimativa, principalmente quando se compara com os dados observados de precipitação da área.

Por meio da aplicação da análise de regressão linear múltipla, foi possível obter a Equação (4) para a estimativa da precipitação anual média em função de variáveis independentes relacionadas à variabilidade espacial da chuva, as coordenadas geográficas, com séries históricas de nove anos (2013-2021).

Essa equação apresentou o coeficiente de determinação da regressão de 0,713 e a média do erro percentual, entre os dados observados e estimados dos postos pluviométricos estudados, foi 6,27%. Dessa forma, é possível afirmar que a equação (4) obtida, através da regressão linear múltipla, teve alta capacidade preditiva.

Sendo assim, nota-se a potencialidade da presente pesquisa em desenvolver uma função para estimar a precipitação média anual considerando a variabilidade espacial na Região Metropolitana do Recife.

REFERÊNCIAS

ALEMU, Melkamu Meseret; BAWOKE, Getnet Taye. Analysis of spatial variability and temporal trends of rainfall in Amhara region, Ethiopia. **Journal of Water and Climate Change**, v. 11, n. 4, p. 1505-1520, 2020.

APAC. Agência Pernambucana de Águas e Clima. **Monitoramento Pluviométrico**. Disponível em: <<http://old.apac.pe.gov.br/meteorologia/monitoramento-pluvio.php>>. Acesso em: 25 mai. 2022.

CEMADEN. Centro Nacional de Monitoramento e Alertas de Desastres Naturais. **Monitoramento Pluviométrico**. Disponível em: <<http://www2.cemaden.gov.br/mapainterativo/#>>. Acesso em: 11 jun. 2022.

CRUZ, Marcus *et al.* Aplicação de regressão linear múltipla para estimativa da precipitação média anual considerando a variabilidade espacial no estado de Sergipe. **XX Simpósio Brasileiro de Recursos Hídricos**, Bento Gonçalves, nov. 2013.

DEVORE, Jay L. **Probabilidade e estatística: para engenharia e ciências**. 6 ed. São Paulo: Cengage Learning, 2006.

FILHO, M. L. O. A utilização da regressão linear como ferramenta estratégica para a projeção dos custos produção. **IX Congresso Brasileiro de Custos**, São Paulo, out. 2002.

IPEA- INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Caracterização e Quadros de Análise Comparativa da Governança Metropolitana no Brasil**: análise comparativa das funções públicas de interesse comum (Componente 2)-RM do Rio de Janeiro (Relatório de Pesquisa). COSTA, M. A. et al. (Org). Rio de Janeiro: IPEA, 2015.

PFALTZGRAFF, Pedro Augusto dos Santos et al. Sistema de Informações Geoambientais da Região Metropolitana do Recife. Companhia de Pesquisa de Recursos Minerais – CPRM, Recife, 2003.

SILVA JUNIOR, Marcos Antonio Barbosa et al. Desafios para a adaptação da infraestrutura de drenagem urbana em cenário de mudança do clima no Recife-PE. **Journal of Environmental Analysis and Progress**, v. 5, n. 3, p. 302-318, 2020.