

ANÁLISE MULTIVARIADA DAS ESTAÇÕES PLUVIOMÉTRICAS DO ANO DE 2005, DA SUB-BACIA 50 PERTENCENTE À REDE HIDROMETEOROLÓGICA DA SUREG-SA

*Denize Ribeiro da Silva*¹

RESUMO - As precipitações atmosféricas são variáveis hidrológicas geradas pelo monitoramento hidrometeorológico, que expressam informações que podem está relacionadas no tempo e no espaço. A natureza multivariada desses dados podem ser melhor interpretada através das análises multivariadas, que são técnicas estatísticas desenvolvidas para verificar as múltiplas relações existentes entre as variáveis. A interdependência características desta variável permite que seja analisada por diversas técnicas estatísticas. Neste estudo serão aplicadas as Técnicas Multivariadas - Análise de Agrupamentos e Análise de Componentes Principais, aos totais mensais de chuvas, da Sub-bacia 50, monitoradas pela Sureg-SA, no período de 2005. A primeira visa definir grupos homogêneos de estações dentro desta região, que melhore a análise preliminar dos dados e a segunda pretende determinar a significância de cada variável em seu respectivo componente. Para tanto, serão estimados as falhas dos totais mensais através de matriz de correlação e análise de regressão, para que a técnica seja eficiente e avalie a série completa.

Palavras-Chave: precipitações, multivariada, técnicas estatísticas.

RAINFALL STATIONS MULTIVARIATE ANALYSIS OF THE YEAR 2005 OF THE BASIN 50 PERTENCENT AND THE HYDROMETEOROLOGICAL NETWORK SUREG-SA.

ABSTRACT - The atmospheric precipitation are generated by hydrometeorological monitoring elements, which express information can is related in time and space. The multivariate nature of the data can be best interpreted through multivariate analyzes, which are statistical techniques developed to check the multiple relationships between the variables. The interdependence characteristics of this variable allows it to be analyzed by various statistical techniques. This study will apply the Multivariate Techniques - Cluster Analysis and Principal Component Analysis, the total monthly rainfall, the Sub-basin 50, monitored by Sureg-SA, from 2005. The first aims to define homogeneous groups of stations within this region, to improve the preliminary analysis of the data and the second aims to determine the significance of each variable on its respective component. To do so, shall be estimated monthly totals of faults through correlation and regression analysis, that the technique is efficient and evaluate the full range.

Keywords: rainfall, multivariate, statistical techniques

1) Técnica em Geociências da CPRM/SGB; Av. Ulysses Guimarães, 2862, CAB, Salvador/BA. CEP: 41213 - 000. Tel: (71) 2101 – 7313. E-mail: denize.silva@cprm.gov.br.

INTRODUÇÃO

As precipitações atmosféricas são variáveis hidrológicas geradas pelo monitoramento hidrometeorológico que possui grande variabilidade espacial e temporal. A aleatoriedade do fenômeno onde nem todos os fatores e influencias, bem com as interdependências podem ser precisamente expressadas; confere-lhe um caráter não determinístico (NAGHETTINI, 2007). O correto entendimento das regularidades e variabilidades inerentes aos processos hidrológicos podem ser mais bem alcançados, através de técnicas estatísticas, mais precisamente, técnicas de análise multivariada.

As series ininterruptas de dados de precipitação se mostram como um dos grandes problemas para a análise de dados, pois impactam na utilização desses dados. As falhas das séries temporais geram dificuldades na interpretação dos fenômenos, bem como excluem pontos de observação importante para análise. Uma maneira de solucionar este problema é utilizar técnicas estatísticas (correlação e regressão) que verifiquem o relacionamento entre as variáveis e estimem essas falhas.

As alturas diárias de chuva observadas em diversos pontos numa bacia hidrográfica possuem características multivariadas, devido à aleatoriedade e a interdependência entre as variáveis. A análise multivariada é uma técnica que simultaneamente analisa as múltiplas variáveis, cujos efeitos não podem ser significativamente interpretados de forma separada (StatSoft, 2004). Essas técnicas são baseadas em inferência estatística e tem o objetivo de capturar a essência multivariada da análise. (HAIR et. al., 2005). O resultado representa uma combinação linear do conjunto de variáveis, com pesos determinados pela técnica que melhor atinja um objetivo específico (HAIR et. al., 2005). Existem várias técnicas de análise multivariada, dentre elas podemos citar a análise de agrupamento e análise de componentes principais.

A análise de agrupamento é uma técnica matemática que visa à formação de grupos mutuamente excludentes, com base nas similaridades (HAIR et. al., 2005). Dentro dos grupos haveria maior homogeneidade, enquanto que entre os grupos maior heterogeneidade (Stat Soft, 2004). A análise de componentes principais é uma abordagem estatística que avalia as inter-relações entre um grande número de variáveis; e explica essas variáveis em termos de suas dimensões. O objetivo é resumir as informações contidas em um grande número de variáveis originais num conjunto menor de variáveis estatísticas, com uma perda mínima de informação (HAIR et. al. , 2005).

A tecnologia computacional favoreceu enormemente a análise de dados. Boa parte desta compreensão e domínio vem do estudo da estatística e inferência estatística (HAIR et. al. , 2005). Esta área, explorada por muito pesquisadores, contribuiu sobremaneira, para o desenvolvimento de Software que melhorariam o processamento de dados. As dificuldades de interpretação de variáveis múltiplas que se inter-relacionam puderam ser superadas por técnicas que analisam essas medidas simultaneamente, as Técnicas de análise multivariada.

MATERIAIS E MÉTODOS

Área de estudo

A sub-bacia 50 pertencente à rede hidrometeorológica da Superintendência Regional de Salvador – Sureg-SA, contempla 34 estações pluviométricas. A bacia localizada no estado da Bahia, entre os paralelos 10°00” e 13°00” latitude sul e 38°00” e 42°00” longitude oeste. As estações pluviométricas que serão analisadas estão representadas na Figura 1, segundo sistema de posicionamento georreferenciado (GPS), a seguir:

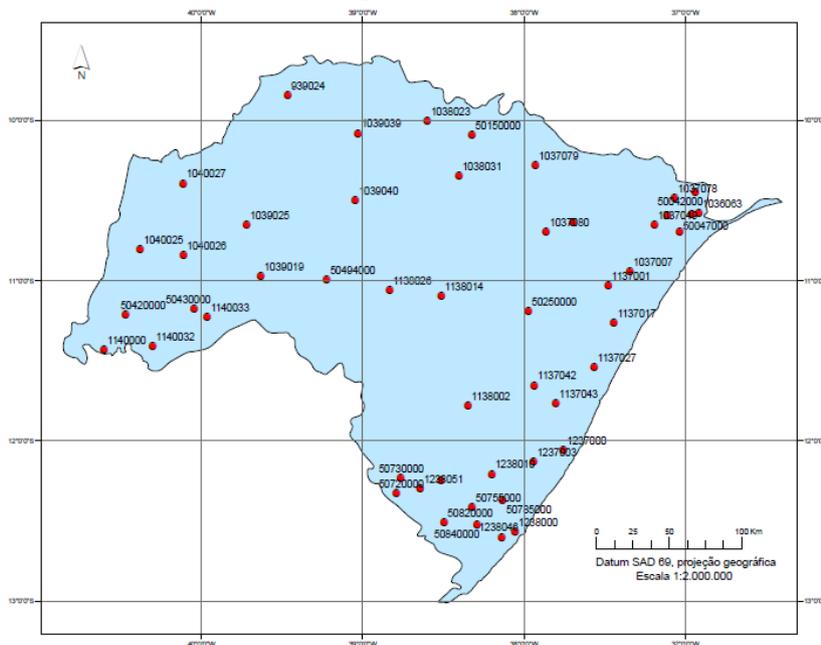


Figura 1 – Mapa de localização das estações pluviométricas da Sub-bacia 50 da Sureg-SA

Dados utilizados

Foram utilizados os totais mensais de precipitação das 34 estações pluviométricas, da sub-bacia 50 do ano de 2005. Essas informações foram obtidas do Hidro 1.2 (Sistema de Informações Hidrológicas). As estações podem ser visualizadas através da Tabela 1, cujos dados foram recodificados para simplificar a análise.

Tabela 1 – Recodificação das estações de chuva da sub-bacia 50.

Codificação	ESTAÇÃO	Codificação	ESTAÇÃO	Codificação	ESTAÇÃO
1	AC.E RIO DO PEIXE	12	CORONEL JOÃO SÁ	23	MIGUEL CALMON
2	AÇÚDE SERROTE	13	CORTE GRANDE	24	NOVO TRIUNFO
3	AÇÚDE JACURICI	14	EMBOACICA	25	PINDOBACU
4	AÇÚDE SOHÉM	15	ESTÂNCIA	26	PONTE EUCLIDES DA CUNHA
5	ÁGUA BRANCA	16	FAZENDA BELÉM	27	PONTO NOVO
6	ALECRIM	17	FAZ. CAJUEIRO	28	QUEIMADAS
7	ALTO REDONDO	18	FAZ. TRIANON	29	SALGADO
8	ARACÁS	19	INHAMBUPE	30	SANTA ROSA DE LIMA
9	BURACICA	20	ITANHI	31	TEODORO SAMPAIO
10	CAPELA	21	JANGADO	32	TIRIRICA

11	CIPO	22	LAGOA PRETA	33	UAUÁ
				34	USINA ALTAMIRA

Fonte: Próprio autor.

Métodos

Foram aplicadas Técnicas Multivariadas - Análise de Agrupamentos e Análise de Componentes Principais, aos totais mensais de chuva das estações. Para que a técnica fosse eficiente e abrangesse a avaliação sobre a série completa, buscou-se estimar as falhas através de técnicas de relacionamento entre variáveis: correlação e regressão. Para tanto, utilizou-se o módulo Análise de dados do Software Excel 2007. Como se tratava de dados de precipitação que variam no tempo e espaço, as metodologias foram aplicadas anualmente e verificada a frequência na formação dos grupos, por três métodos diferentes, frequentemente utilizados e verificadas a significância dessas variáveis.

A análise de agrupamento é uma técnica multivariada com objetivo de agregar objetos com base em suas características. Objetos semelhantes ficariam no mesmo grupo distanciando-se dos outros pelas diferenças. Dentro dos grupos haveria maior homogeneidade, enquanto que entre os grupos elevada heterogeneidade. A representação desses grupos é feita por meio de um gráfico com uma estrutura de árvore denominado Dendrograma (Sneath & Sokal *apud* StatSoft; 2004). Foram utilizados os Métodos Ligação Simples (*Single Linkage*), Ligação Completa (*Complete Linkage*) e *Método Ward* e comparadas às frequências na formação dos grupos.

Para o teste de significância das variáveis foi realizada a análise de componentes principais, técnica multivariada que busca reduzir o espaço das variáveis, criando eixos ortogonais, que são combinações lineares das variáveis originais - os componentes principais. É medido o poder de cada variável em seu componente, o que permite reduzir a quantidade de variáveis originais num conjunto menor (StatSoft, 2004).

Para realização destas duas técnicas multivariadas foi utilizado o Software STATISTICA 7.0 ® (StatSoft Inc., 2004).

RESULTADOS E DISCUSSÕES

A partir dos resultados mostrados nos gráficos da Figura 2, da análise de agrupamento, podemos perceber a frequência na formação de alguns grupos de estações que foram codificadas, a fim de facilitar a análise. Nas abscissas encontram-se os grupos, enquanto que nas ordenadas estão as distâncias entre os grupos.

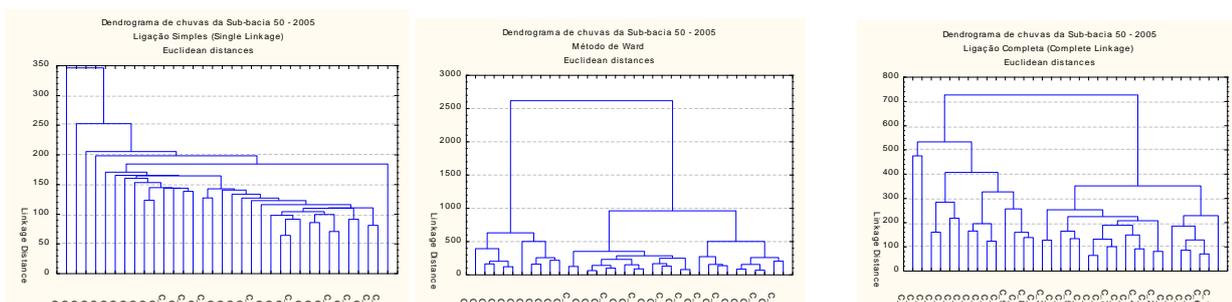


Figura 2 – Gráficos de análise de agrupamento segundo os métodos: *Single Linkage*, *Complete Linkage* e *Método Ward*.

A Figura 3 apresenta a formação dos grupos de análise, de acordo com os métodos acima citados. Com exceção das estações codificadas 14 e 17 que hora fazem partes de grupos diferentes; ora estão juntas no mesmo grupo, numa análise preliminar de chuvas haveria uma preocupação maior com essas estações, no sentido de definir em que grupo elas seriam analisadas.

Formação dos grupos de análise de chuva segundo os três métodos	
Grupo1	13-20-21-32
Grupo2	15-16-29-30
Grupo3	8-9-31-34
Grupo4	3-23
Grupo5	5-24-33
Grupo6	4-7-11-12-25-26-27
Grupo7	1-6-10-18-19-22
Grupo8	2-28

Figura 3 – Grupos de estações de análise de chuvas formados, segundo os métodos: *Single Linkage*, *Complete Linkage* e *Método Ward*.

A verificação da relevância das variáveis foi realizada com a análise de componentes principais e os resultados aparecem na Tabela 2, a seguir:

Tabela 2 – Autovalores da matriz de correlação e estatísticas das variáveis

Autovalores da matriz de correlação e estatísticas das variáveis				
Variável	Autovalor	% Total	Autovalor	% Acumulada
1	4,885905	40,71588	Autovalores	40,7159
2	2,163696	18,03080	7,04960	58,7467
3	1,485885	12,38237	8,53549	71,1290
4	1,014545	8,45454	9,55003	79,5836
5	0,716394	5,96995	10,26643	85,5535
6	0,520026	4,33355	10,78645	89,8871
7	0,471016	3,92513	11,25747	93,8122
8	0,305231	2,54360	11,56270	96,3558
9	0,225266	1,87722	11,78796	98,2330
10	0,112604	0,93837	11,90057	99,1714
11	0,066580	0,55483	11,96715	99,7262
12	0,032851	0,27376	12,00000	100,0000

Fonte : STATISTICA 7.0 ® (StatSoft Inc.,2004)

Cada componente principal, que é uma combinação linear das variáveis originais, retém uma quantidade de informação - sua variância - que é organizada em ordem decrescente. Num conjunto grande de variáveis nem todas tem quantidade de informação relevante, então a aplicação desta técnica visa selecionar quais as variáveis que mais contribuem com informações. De acordo com a

Tabela 2, os dois primeiros componentes principais são os que retêm a maior quantidade de informação; cerca de 58,7 % da variabilidade das variáveis originais.

Tabela 3 – Correlação entre as variáveis e os componentes principais.

Variável	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12
Janeiro	-0,001	0,740	0,163	-0,335	0,237	-0,357	0,347	-0,089	0,025	-0,006	-0,016	-0,013
Fevereiro	0,088	0,683	0,378	0,296	0,205	0,460	0,119	0,110	0,117	-0,012	-0,007	0,002
Março	-0,289	0,059	0,560	0,588	-0,417	-0,251	0,104	-0,001	-0,007	0,060	0,004	-0,016
Abril	-0,788	0,254	0,227	-0,055	0,249	-0,064	-0,210	0,247	-0,286	0,063	0,009	-0,021
Mai	-0,848	-0,239	-0,040	-0,137	-0,028	-0,132	-0,012	0,313	0,282	-0,057	-0,043	-0,036
Junho	-0,883	0,258	-0,138	0,124	-0,112	0,069	-0,051	-0,174	-0,105	-0,203	-0,117	-0,028
Julho	-0,854	-0,221	0,372	-0,020	0,151	-0,028	-0,027	-0,108	0,045	-0,125	0,156	0,069
Agosto	-0,706	-0,358	0,427	-0,175	0,179	0,104	-0,099	-0,251	0,096	0,158	-0,089	-0,014
Setembro	-0,823	0,138	-0,431	-0,041	-0,143	0,172	0,167	-0,083	-0,001	0,078	0,122	-0,095
Outubro	-0,855	0,116	-0,372	-0,025	-0,167	0,073	0,223	0,067	-0,039	0,095	-0,048	0,122
Novembro	-0,119	0,786	-0,326	0,152	-0,014	-0,155	-0,419	-0,093	0,156	0,057	0,026	0,025
Dezembro	0,139	0,358	0,421	-0,612	-0,514	0,144	-0,116	0,028	-0,012	-0,015	0,015	0,007

Fonte: STATISTICA 7.0 ® (StatSoft Inc.,2004)

A matriz de correlação da Tabela 3 permite inferir que boa parte dos meses apresenta poder discriminatório no primeiro componente principal, exceto para os meses de janeiro, fevereiro, março, novembro e dezembro. Esses meses, cujas correlações são altas e negativas indicam que estações localizadas mais à esquerda sofrem mais influencia desses meses. Já no segundo componente principal os meses de janeiro e novembro são os que aparecem com um alto poder discriminatório. Como ambas correlações são positivas estações localizadas mais a direita sofrem mais influencia desses meses. Essas informações sobre as estações podem ser visualizadas na Figura 4, a seguir:

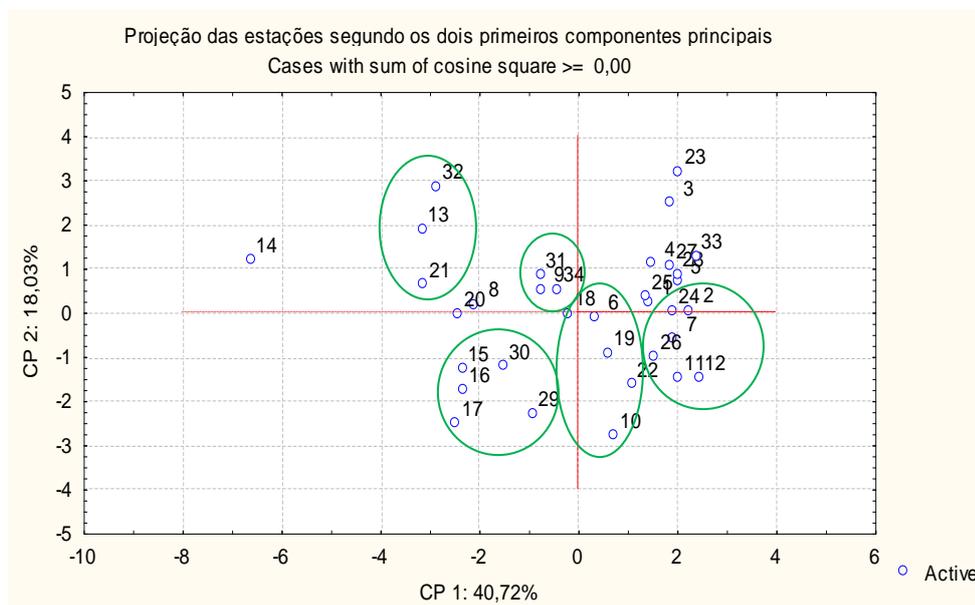


Figura 4 – Gráfico da distribuição das estações de chuva segundo os componentes principais.

Através da Figura 4, podemos observar a formação de alguns grupos e a influência dos meses para estes grupos. Os grupos de estações mais localizadas a direita do primeiro componente principal sofrem mais influência dos meses de abril a outubro, enquanto os mais a esquerda dessa mesma reta, sofre fraca influência dos demais meses. Quanto ao segundo componente principal os grupos mais a esquerda, sobretudo os formados pelas estações 13, 21, 32 sofrem forte influência dos meses de janeiro e novembro.

CONCLUSÃO

A aleatoriedade das variáveis hidrológicas, bem como sua interdependência espaço-temporal faz com que o uso de técnicas estatísticas, melhore a interpretação do fenômeno, bem como, permite tomadas de decisões mais precisas. A maioria das técnicas multivariadas é baseada em inferência estatística, onde é preciso especificar os níveis de erro aceitáveis - com exceção da análise de agrupamento. A natureza multivariada das precipitações atmosféricas se mostrou útil para a aplicação desta técnica. A análise multivariada possibilitou a seleção de grupos de estações de análise, como também identificou dentro do período de estudo os meses que mais contribuem, ou mais significantes para o total pluviométrico.

REFERÊNCIAS

- HAIR, Jr., J. F. et. al. Análise multivariada de dados. Trad. Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5 ed._Porto Alegre: Bookman, 2005.
- BARBETTA, P. A. Estatística aplicada às Ciências Sociais. 5. ed. rev _Florianópolis: Ed. da UFSC, 2005.
- ANDRIOTTI, J. L. S. Interpretação de prospecção geoquímica com o auxílio de estatística. _Porto Alegre: CPRM, 2010.
- SPIEGEL, M. R. Estatística. Trad. Pedro Consentino. 3ª. ed._São Paulo: Makron Books, 1993.
- TUCCI, C. E. M. et. al. (Org.). Hidrologia: ciência e aplicação. 4ª. ed. 4ª. reimp._ Porto Alegre: Editora da UFRGS/ABRH, 2012.