

## CRITÉRIOS ESTATÍSTICOS PARA DEFINIR ANOS PADRÃO: UMA CONTRIBUIÇÃO À CLIMATOLOGIA GEOGRÁFICA\*

Miriam Rodrigues SILVESTRE\*\*

João Lima SANT'ANNA NETO\*\*\*

Edilson Ferreira FLORES\*\*\*\*

### Resumo

Um problema muito comum na área de Climatologia diz respeito a classificar um conjunto de dados em um determinado número de classes, como por exemplo, para variáveis de temperatura e precipitação no momento de se definir valores como sendo habituais ou extremos. Entretanto, existem diversas técnicas que podem ser empregadas para essa finalidade e o objetivo deste artigo é abordá-las, apontando alguns pontos favoráveis e desfavoráveis em sua utilização. As técnicas apresentadas são: a Fórmula de Sturges, empregada para definir o número de classes e a amplitude dos intervalos das classes; a técnica baseada no gráfico *Box Plot* e que utiliza os percentis de 5%, 25%, 75% e 95%; Quantis, que considera os percentis de 15%, 35%, 65%, 85%; a técnica de um Desvio-padrão em relação à média; técnicas de Agrupamentos, para a qual serão considerados seis métodos de Agrupamentos Hierárquicos denominados Ligação Simples, Ligação Completa, Ligação Média, Ligação Mediana, Ligação de Centróides e Ward. As técnicas apresentadas são aplicadas aos dados médios anuais de temperatura mínima e máxima e precipitação total anual de Presidente Prudente (SP), obtidos junto à Estação Meteorológica da FCT/UNESP, referente ao período de 1961 a 2010.

**Palavras-chave:** construção de classes, gráfico *Box plot*, percentis, análise de agrupamentos, climatologia.

## CRITERIOS DE ESTADÍSTICA PARA DEFINIR AÑOS ESTÁNDAR: UNA CONTRIBUCIÓN A LA CLIMATOLOGÍA GEOGRÁFICA

### Resumen

Un problema muy común en el campo de la climatología se refiere a ordenar un conjunto de datos en un número de clases, como por ejemplo, variando la temperatura y la precipitación en el momento de la definición de valores como normal o extrema. Sin embargo, hay varias técnicas que se pueden emplear para este propósito y objetivo de este artículo es para hacer frente a tales técnicas, señalando algunos puntos favorables y desfavorables en su uso. Las técnicas presentadas son: Fórmula Sturges, utilizado para definir el ancho de los rangos de las clases; una técnica basada en el gráfico Diagrama del Caja, que utiliza los percentiles 5%, 25%, 75 % y 95%; la técnica de los cuantiles, que considera a los percentiles 15 %, 35 %, 65 % y 85 %; la técnica de una desviación estándar de la media, y seis técnicas de agrupamiento jerárquico: Single Link, Complete Link, Media Link, Median Link, Centroide y Ward. Las técnicas presentadas se aplican a los datos sobre temperatura mínima promedio y media máxima y la precipitación de Presidente Prudente (SP), obtenido de la Estación Meteorológica de la FCT / UNESP, para el período 1961-2010.

**Palabras clave:** construcción de clases, gráfico diagrama de caja, percentiles, análisis cluster, climatología.

## STATISTICAL CRITERIA FOR DEFINING YEARS STANDARD: A CONTRIBUTION TO THE GEOGRAPHICAL CLIMATOLOGY

### Abstract

A very common problem in climatology field relates to classify a data set into a number of classes, as for example, to the variables temperature and precipitation at the time of defining values as normal or extreme. However, there are several techniques that can be employed for this purpose and aim of this article is to address them, pointing out some favorable and unfavorable points in their use. The techniques presented are: Formula Sturges, used to define the number of classes and the width of the ranges of classes; the techniques based on the Box Plot chart, which uses percentiles of

\* Trabalho de pesquisa de Doutorado e da disciplina Teoria e Método em Climatologia Aplicada e Análise Ambiental.

\*\* Doutoranda Programa de Pós-Graduação em Geografia da FCT/UNESP/Presidente Prudente, Profa. Dep. Estatística da FCT/UNESP, miriam@fct.unesp.br.

\*\*\* Prof. Dr. Departamento de Geografia da FCT/UNESP – Campus de Presidente Prudente, joalima@fct.unesp.br.

\*\*\*\* Prof. Dr. Departamento de Estatística da FCT/UNESP – Campus de Presidente Prudente, efflores@fct.unesp.br

5%, 25%, 75% and 95%; quantiles, which considers the percentiles 15%, 35%, 65% and 85%; the technique of one standard deviation from the average, and six techniques of Hierarchical Cluster: Single Linkage, Complete Linkage, Media Linkage, Median Linkage, Centroid and Ward Method. The techniques presented are applied to average annual data on minimum and maximum temperature and to total annual precipitation of Presidente Prudente (SP), obtained from the Meteorological Station of São Paulo State University, for the period 1961-2010.

**Key-words:** building classes, box plot, percentiles, cluster analysis, climatology.

## **1 Introdução**

A técnica de definir os anos do período de estudo em anos padrão é comumente empregada em Climatologia Geográfica. O principal objetivo da técnica é separar anos que apresentem características semelhantes quanto a uma determinada variável, de forma que esses possam ser considerados como um grupo com características semelhantes; e posteriormente eleger um ano padrão para representar cada grupo.

Tavares (1976) apresentou uma técnica de agrupamento para definir anos padrão, para analisar os dados de precipitação pluvial de Campinas, referentes ao período de 1961 a 1970. Inicialmente, o autor agrupa os dados mensais de precipitação em períodos trimestrais, correspondendo às estações do “ano civil”: verão (janeiro, fevereiro e março), outono (abril, maio e junho), inverno (julho, agosto e setembro) e primavera (outubro, novembro e dezembro) e calcula a média para cada uma das estações do ano, para todos os anos da série. Após, considera o desvio percentual de cada estação em relação aos dados médios obtidos, e calcula o coeficiente de similaridade multidimensional entre os desvios percentuais e organiza os resultados em formato de matriz à qual aplica uma técnica de agrupamento hierárquica por pares recíprocos, detalhada em Diniz (1971). O resultado final da técnica é apresentado num gráfico denominado árvore de ligação, também conhecida por dendrograma, a partir do qual pode-se observar o agrupamento dos anos considerados, e o nível de perda de detalhe (%) em que eles foram realizados. O autor classifica os anos padrão em: secos, normais e chuvosos. E aponta os anos que poderiam representar cada uma dessas categorias. Assim, ao invés de se realizar um estudo mais detalhado de todos os anos, seleciona-se um ano para cada um dos grupos de interesse para dar prosseguimento em outras análises, como por exemplo, a análise rítmica.

Sant’Anna Neto (1990), um estudo sobre “O ritmo climático e a gênese das chuvas na zona costeira paulista”, utilizou as estações do “ano civil” conforme adotado por Tavares (1976), numa análise sazonal. Para a definição dos anos padrão baseou-se na análise dos dendrogramas e no coeficiente de variação. Primeiramente o autor classificou todos os 20 anos de estudo referente ao período de 1967 a 1986 em cinco classes: H para anos cujos índices sazonais foram habituais ou normais; I para anos cujos índices sazonais variaram próximos ao coeficiente de variação, portanto intermediário; NC para anos com índices sazonais superiores as médias e que apresentaram tendências chuvosas; NS idem ao anterior porém apresentando tendências secas; C para índices sazonais excepcionalmente superiores ao coeficiente de variação, portanto, chuvosos e S para anos com índices excepcionalmente superiores ao coeficiente de variação, porém negativos, definidos como secos. Após as análises, o autor selecionou um ano para cada uma das três

categorias de interesse como ano padrão, estes foram: 1981 como o ano padrão habitual, 1983 como excepcionalmente chuvoso e 1985 o ano padrão excepcionalmente seco.

Flores (1995) apresentou o Sistema de Informação Climatológica (SIC), um sistema computacional que realiza diversas análises estatísticas voltadas as variações do ritmo pluvial, entre elas os cálculos dos desvios mensais sazonais e a construção da árvore de ligação por pares recíprocos segundo Diniz (1971), Sanches (1972) e Tavares (1976). Vários trabalhos utilizaram programas computacionais que posteriormente foram incorporados ao SIC, entre eles: Sant'Anna Neto (1990), Zavattini (1990), Boin (2000) e Flores (2001).

Sant'Anna Neto (1995) faz a classificação do Estado de São Paulo no período de 1941 a 1993, dividindo-o em oito unidades pluviais e vinte e cinco subunidades homogêneas. Nesse estudo o autor analisou a variabilidade temporal organizando os dados em cada subunidade a partir da classificação da precipitação em cinco grupos ou classes de intensidade, utilizando a técnica de um Desvio-padrão em relação à média. O autor classificou os anos em: seco, tendente a seco, habitual, tendente a chuvoso e chuvoso.

Xavier (2001) emprega a técnica dos Quantis aos dados quadrimestrais de precipitação para as regiões pluviométricas do Ceará, considerando os percentis: P15, P35, P65 e P85. Em seu trabalho utiliza a seguinte classificação: muito seco (0 a P15), seco (P15 a P35), normal (P35 a P65), chuvoso (P65 a P85) e muito chuvoso (P85 a P100).

A técnica de agrupamento utilizando o método de Ward foi empregada em Jácomo (2011, p. 55-9) para a região oeste do Estado de São Paulo, no período de 1970 a 2000, para 108 estações pluviométricas do Departamento de Águas e Energia Elétrica (DAEE). O autor classificou o período de 31 anos em 3 grupos de anos padrão (seco, habitual e chuvoso), porém, considerando os dados de todas as estações conjuntamente numa única análise, e dessa forma os anos padrão foram construídos para toda a região oeste paulista.

Galvani (2012) apresenta a técnica de classificação baseada nos quartis do gráfico *Box Plot*, e introduz os percentis P5 e P95, os quais separam os anos padrão super-secos (0 a P5), dos 5% menores valores de precipitação, e super-úmidos (P95 a P100), para os 5% maiores valores de precipitação para Piracicaba (SP), no período de 1917 a 1994. As demais categorias são definidas pelos intervalos entre os 1º. e 3º. quartis, ou seja, os percentis P25 e P75, respectivamente.

Neste artigo são apresentadas algumas técnicas selecionadas pelos autores e que podem vir a ser empregadas para a classificação de anos padrão seja de temperatura, precipitação ou outro elemento climático de interesse. As técnicas estão definidas para classificar dados univariados de somente um local ou estação, embora, as técnicas de agrupamentos possam também ser aplicadas em várias estações concomitantemente para classificar os anos padrão de uma determinada região usando dados multivariados.

## **2 Técnicas para definição de anos padrão**

Nesta seção serão abordadas cinco técnicas que podem ser utilizadas para classificar os anos de

estudo em anos padrão: a Fórmula de Sturges, *Box Plot*, Quantis, um Desvio-padrão em relação à média e Análise de Agrupamentos, na qual serão avaliados seis métodos de agrupamentos hierárquicos.

## 2.1 Fórmula de Sturges

Sturges (1926, apud GERARDI, 1981, p.34) apresenta uma maneira de determinar o número de classes adequado  $k$  para um conjunto de dados, a partir da equação (1), conhecida por Fórmula de Sturges:

$$k = 1 + 3,33 \log_{10} N. \quad (1)$$

Por exemplo, para  $N=50$  observações de precipitação total anual, substituindo o valor de  $N$  na equação (1), tem-se:  $k = 1 + 3,33 \log_{10} 50 \Rightarrow k = 1 + 3,33(1,69897000 4) \Rightarrow k = 1 + 5,61 = 6,61 \cong 7$  classes.

Portanto, para  $N=50$ , seria adequado dividir o conjunto de dados em 7 classes de intervalos iguais. A amplitude dos intervalos de cada classe é encontrada utilizando-se a amplitude total dos dados ( $At$ ), a qual é a diferença entre o maior e o menor valor da série de dados, através da equação (2):

$$AI = \text{Amplitude do Intervalo de Classe} = At/k = (\text{Max}-\text{Min})/k \quad (2)$$

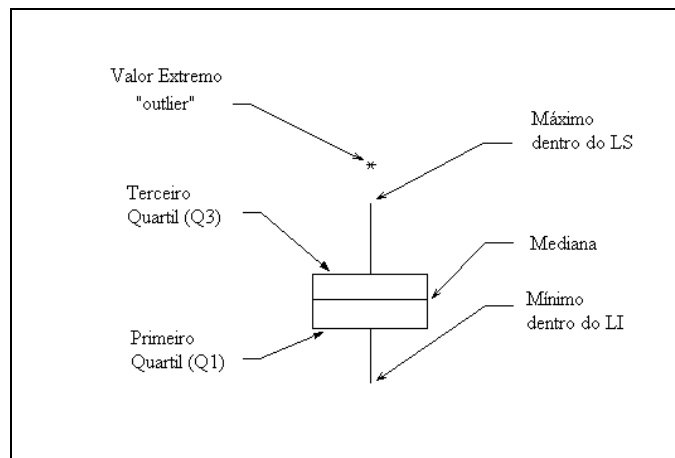
E para definir cada classe inicia-se do valor  $Min$  e adiciona-se o valor de  $AI$ , constituindo a primeira classe. A segunda classe deve iniciar do valor anterior e novamente deve-se adicionar o valor de  $AI$ , e assim sucessivamente até que se tenham construídas todas as  $k$  classes. Deve-se definir os intervalos fechados à esquerda e abertos a direita, conforme será ilustrado nos exemplos da Seção 3, dessa forma, as classes formada são mutuamente exclusivas, ou seja, um mesmo número não pode ser classificado em duas classes diferentes.

Entretanto, é comum em climatologia o interesse em definir o conjunto de dados num número fixo de classes, como por exemplo para 5 classes de anos padrão. Nesse caso, define-se  $k=5$  e considera-se a Amplitude do Intervalo de Classe =  $At/5$ .

## 2.2 Box Plot

O gráfico *Box Plot*, ou diagrama de caixa, é bastante conhecido na área de Estatística, e a maioria dos softwares estatísticos apresenta este gráfico. Para construí-lo, os dados devem ser ordenados, do menor para o maior valor, e devem ser calculadas estatísticas resumo conhecidas como quartis, que dividem o conjunto de dados em quatro partes, sendo o primeiro quartil ( $Q1$ ) para o qual 25% das observações situam-se iguais ou abaixo de seu valor, o segundo quartil ( $Q2$ ), também conhecido por mediana para a qual 50% das observações são iguais ou situam-se abaixo dela, e o terceiro quartil ( $Q3$ ), que separa 75% das observações iguais ou abaixo de seu valor. A Figura 1 apresenta um modelo de *Box Plot*.

Figura 1 – Esquema de um gráfico *Box Plot*



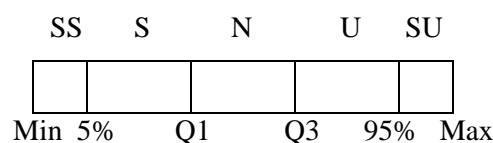
Fonte: Autoria própria.

A linha desenhada no centro do retângulo, na Figura 1, representa o valor da mediana, a linha inferior do retângulo o primeiro quartil (Q1), e linha no topo o terceiro quartil (Q3). A partir do terceiro quartil, é apresentada uma linha que se estende até o maior valor observado para variável, desde que este não exceda o limite superior,  $LS=Q3+1,5(Q3-Q1)$ . Abaixo do primeiro quartil, há outra linha que se estende até o menor valor observado da variável, desde que este não exceda o limite inferior  $LI=Q1-1,5(Q3-Q1)$ . Os asteriscos representam casos na amostra considerados valores extremos também chamados anômalos/*outliers*, os quais excedem os limites inferior ou superior. Maiores detalhes sobre este podem ser encontrados em Bussab e Morettin (2002, p. 48).

A vantagem do *Box Plot* é fornecer uma visualização rápida da distribuição dos dados, e se a distribuição é simétrica a caixa encontra-se equilibrada com a mediana se posicionando no centro da mesma. Já para distribuições assimétricas, há um desequilíbrio na caixa, com relação à mediana.

Os quartis dividem o conjunto de dados em quatro partes. Entretanto, para fins de classificação em climatologia é bastante comum o uso de cinco classes diferentes, e para isso, é preciso redefinir os limites entre as classes para que seja satisfeita essa necessidade. Aproveitando os valores já definidos pelos quartis, Galvani e Luchiari (2012), fazem uma nova divisão para o elemento climático chuva, com o objetivo de definir anos padrão considerando as classes: super seco (S) para os 5% menores valores, seco (S) de 5% a 25%, normal (N) de 25% a 75%, úmido (U) de 75% a 95% e super úmido (SU) para os 5% maiores valores, conforme ilustrado na Figura 2.

Figura 2 – Ilustração da divisão de ano-padrão em 5 classes usando a técnica dos quartis.



Fonte: Autoria própria.

Para calcular os 5% menores valores e os 5% maiores, é necessário definir a medida denominada percentil. Os percentis dividem o conjunto de dados em 100 partes, cada uma delas representa um percentual acumulado das observações até aquele limite. Dessa forma, Galvani e Luchiari (2012) sugerem utilizar o 5º. Percentil (P5) e o 95º. Percentil (P95) para a definição das classes extremas. Reescrevendo os quartis como percentis, tem-se que: Q1=P25, Q2=P50, Q3=P75. A Figura 3 é uma reorganização da Figura 2 para a notação dos percentis.

Figura 3 – Ilustração da divisão de ano-padrão em 5 classes com a técnica *Box Plot*

SS	S	N	U	SU	
5%	20%	50%	20%	5%	
Min	P5	P25	P75	P95	Max

Fonte: Autoria própria.

Segundo Bussab e Morettin (2002), a vantagem de se utilizar as medidas de percentis definidas na Figura 2, é que os percentis são medidas robustas, ou seja, seus valores são pouco alterados por mudanças de uma pequena porção dos dados. A mediana é uma medida resistente, enquanto que a média e o desvio-padrão não são.

### 2.3 Quantis

A técnica dos Quantis utiliza valores de percentis diferentes dos aplicados na técnica *Box Plot*. O primeiro pesquisador a utilizá-la foi Pinkayan (1966, apud Xavier e Xavier, 1999), para avaliar a ocorrência de anos secos e chuvosos sobre extensas áreas continentais, em especial na área continental dos Estados Unidos, utilizando os percentis P15, P35, P65 e P85 para definir cinco classes de precipitação: muito seco (MS) para os 15% menores valores, seco (S) entre 15% e 35% das observações, normal (N) entre 35% e 65%, chuvoso (C) entre 65% a 85% e muito chuvoso (MC) para valores acima de 85%. A Figura 4 ilustra essa divisão considerando a notação dos percentis.

Figura 4 – Ilustração da divisão de ano-padrão em 5 classes usando a técnica dos Quantis

MS	S	N	C	MC	
15%	20%	30%	20%	15%	
Min	P15	P35	P65	P85	Max

Fonte: Autoria própria.

Foram encontrados vários trabalhos envolvendo esta técnica, tais como: Xavier e Xavier (1999) utilizaram o esquema de classificação apresentado na Figura 4 para classificar a chuva acumulada em quadras chuvosas de regiões pluviométricas do Estado do Ceará; Xavier (2001, p. 170) comenta que em

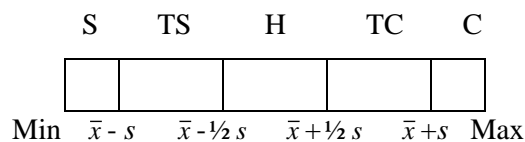
trabalhos anteriores realizados entre 1984 e 1990, a autora verificou que esses níveis de quantis permitiram classificar corretamente o grau de severidade das secas ocorridas no passado; Xavier (2007, apud Barbosa, 2008) utilizou a técnica para caracterizar períodos secos e chuvosos, com o objetivo de instituir um sistema de “alarme de seca”; Barbosa (2008), desenvolveu sua dissertação de mestrado identificando extremos de precipitação para o litoral paulista e entorno; Monteiro (2011) para classificar os anos de 1980 a 2009 para municípios do Estado do Ceará.

Com a definição dos Percentis P15, P35, P65 e P85, as classes extremas conterão mais observações, 15% em cada extremo, do que se forem utilizados os Percentis P5, P25, P75 e P95, cujos limites deixam apenas 5% das observações em cada extremo.

## 2.4 Desvio-padrão em relação à média

Um critério bastante utilizado para classificar anos padrão é através do desvio-padrão em relação à média. Entretanto, essa técnica é mais adequada para conjuntos de dados que apresentem distribuição simétrica. Sant’Anna Neto (1995) utilizou a técnica de um Desvio-padrão em sua tese de doutorado para representar a evolução temporal das chuvas, bem como sua distribuição no Estado de São Paulo, no período de 1941 a 1993. O autor definiu cinco classificações para os anos com relação à chuva: S=ano seco, TS=ano tendente a seco, H=ano habitual, TC=ano tendente a chuvoso e C=ano chuvoso. Para a definição das classes de chuva são utilizados a média ( $\bar{x}$ ) e o desvio-padrão ( $s$ ), conforme ilustra a Figura 5.

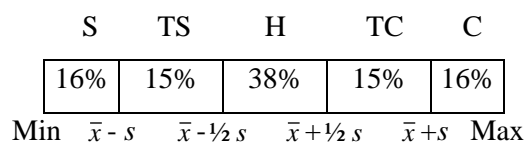
Figura 5 – Ilustração da divisão de ano-padrão em 5 classes usando a técnica de um Desvio-padrão.



Fonte: Autoria própria.

Quando os dados seguem uma distribuição normal, a qual é simétrica e tem uma forma de sino, é possível calcular qual seriam as frequências esperadas em cada classe. Na Figura 6 são apresentadas as probabilidades que seriam aproximadamente esperadas nas cinco classes sob a normalidade dos dados.

Figura 6 – Ilustração da divisão de ano-padrão em 5 classes usando a técnica do desvio-padrão.



Fonte: Autoria própria.

## 2.5 Análise de Agrupamentos

A Análise de Agrupamentos (AA) faz parte de um conjunto maior de técnicas estatísticas denominado Análise Multivariada, e pode ser encontrada em Johnson e Wichern (2002) e Mingoti (2005). A técnica AA é utilizada para agrupar objetos, indivíduos, anos, etc., formando grupos homogêneos internamente e heterogêneos entre si. Neste artigo ela foi empregada para agrupar anos que apresentaram características semelhantes. A técnica permite que sejam utilizadas mais de uma variável ( $p$ ) no agrupamento ao mesmo tempo, entretanto, para a Climatologia, é mais interessante a análise individual de cada uma das variáveis, como por exemplo: temperatura máxima média, temperatura mínima média e precipitação total anual.

Antes de aplicar uma técnica de agrupamento, primeiramente, é necessário definir qual medida de proximidade será utilizada. Existem medidas de similaridade, para a qual, quanto maior seu valor, mais os objetos se parecem; e a medida de dissimilaridade, tal que quanto menor for seu valor, mais parecidos são os objetos.

Considere o vetor de observações do objeto  $i$ , sobre o qual foram anotadas  $p$  variáveis:

$$x_i = (X_{i1} \quad X_{i2} \quad \dots \quad X_{ip}).$$

A definição da distância euclidiana entre os objetos  $i$  e  $k$  é derivada da idéia de distância existente

entre dois pontos no espaço, e é dada por:  $d_{ik} = \left[ \sum_{j=1}^p (X_{ij} - X_{kj})^2 \right]^{1/2}$ .

Existem outras medidas de distância, tais como:

- Distância Euclidiana Média:  $d_{ik} = \left[ \sum_{j=1}^p (X_{ij} - X_{kj})^2 / p \right]^{1/2}$ .
- Distância Euclidiana ao Quadrado:  $d_{ik} = \sum_{j=1}^p (X_{ij} - X_{kj})^2$ .
- Distância *city-block* ou Distância de Manhattan:  $d_{ik} = \sum_{j=1}^p w_j |X_{ij} - X_{kj}|$ , onde  $w_j$  representam pesos de ponderação para as variáveis. Os valores mais usados são: a equiponderação em que  $w_j=1$  ou da média para a qual  $w_j=1/p$ .
- Distância de Minkowsky:  $d_{ik} = \left[ \sum_{j=1}^p w_j |X_{ij} - X_{kj}|^\lambda \right]^{1/\lambda}$ , onde  $w_j$  representam pesos de ponderação para as variáveis. No caso de  $\lambda=1$  esta distância transforma-se na distância de *city-block* ou Manhattan, se  $\lambda=2$  tem-se a distância euclidiana. A métrica de Minkowsky tem a vantagem de ser menos afetada pela presença de *outliers* na amostra do que a distância euclidiana.

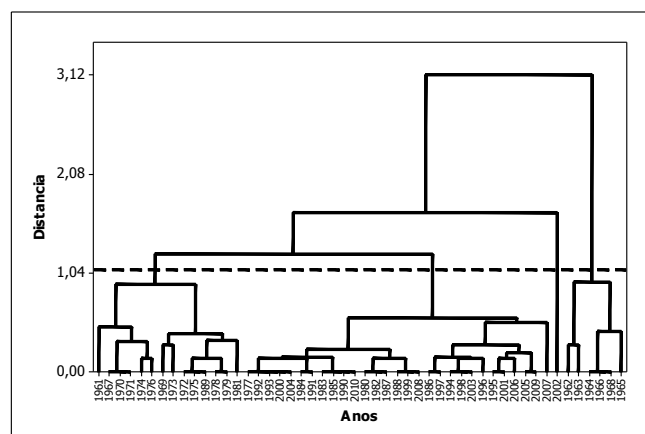
Após escolhida a medida de distância, deve-se definir qual será o método de agrupamento. Existem métodos hierárquicos e não-hierárquicos. Neste artigo serão abordados os métodos hierárquicos de agrupamento:



- Método de Ligação Simples (*Single Linkage*): considera a menor distância entre um elemento de um grupo  $G_1$  e um elemento de outro grupo  $G_2$ , e utiliza a expressão:  $d[G_1, G_2] = \min_{i \in G_1, k \in G_2} d_{ik}$ .
- Método de Ligação Completa (*Complete Linkage*): considera a maior distância entre um elemento de  $G_1$  e um elemento de  $G_2$ , através da expressão:  $d[G_1, G_2] = \max_{i \in G_1, k \in G_2} d_{ik}$ .
- Método de Ligação Média (*Average Linkage*): considera a média das distâncias entre os elementos de  $G_1$  e os elementos de  $G_2$ , dada pela expressão:  $d[G_1, G_2] = \frac{\sum_{i \in G_1} \sum_{k \in G_2} d_{ik}}{g_1 g_2}$ .
- Método de Ligação Mediana (*Median Linkage*): a distância entre dois grupos é a distância mediana entre uma observação em um grupo e uma observação no outro grupo. A vantagem é ponderar a influência de *outliers* (pontos extremos).
- Método de Ligação entre Centróides (*Centroid Linkage*): a coordenada de cada grupo, denominado centróide, é a média das coordenadas de seus objetos. Após calcular os centróides, a distância entre os grupos é obtida através do cálculo das distâncias entre os mesmos.
- Método de Ward: a distância entre dois grupos é a soma de desvios ao quadrado dos pontos aos centróides. A atribuição de um elemento a um grupo é feita de modo a minimizar a soma de quadrados dentro dos grupos.

A análise de agrupamentos produz um gráfico denominado dendrograma, ou árvore de ligação, o qual apresenta um resumo do método aplicado passo a passo, considerando desde o passo inicial, unindo os primeiros dois objetos (anos, por exemplo), até o passo final no qual todos os elementos foram agrupados constituindo um único grupo. O usuário da análise deve decidir sobre o ponto de corte, que define o número de grupos desejado. A Figura 7 apresenta um exemplo de dendrograma produzido pelo software Minitab, e a linha pontilhada, o ponto de corte que define cinco grupos (*clusters*) de anos padrão.

Figura 7 – Exemplo de árvore de ligação (dendrograma)



Fonte: Autoria própria.

A linha pontilhada na Figura 7 serve para dividir o conjunto completo de observações em vários grupos, de forma que aqueles que se situam abaixo da linha constituam grupos homogêneos de observações (anos). Observa-se a formação de cinco grupos. O grupo 1 é constituído por 14 anos 1961, 1967, 1970, 1971, 1974, 1976, 1969, 1973, 1972, 1975, 1989, 1978, 1979, 1981. O grupo 2 contém 29 anos: 1977, 1992, 1993, 2000, 2004, 1984, 1991, 1983, 1985, 1990, 2010, 1980, 1982, 1987, 1988, 1999, 2008, 1986, 1997, 1994, 1998, 2003, 1996, 1995, 2001, 2006, 2005, 2009 e 2007. No terceiro grupo há somente o ano 2002. O grupo 4 contém os anos 1962 e 1963. O quinto e último grupo é formado pelos anos: 1964, 1966, 1968, 1965. Se a linha que indica o ponto de corte for movida para cima o número de grupos diminui, porém, a distância entre os elementos que formam os grupos tornam-se maiores, ou seja, quanto maior é a distância adotada menor é a semelhança interna entre os objetos que formam os grupos.

O número de grupos pode ser pré-definido, por exemplo, quando se adota as mesmas classificações e portanto, o mesmo número de grupos já definidos por um determinado autor, e também pode-se procurar o número de classes mais adequado ao conjunto de dados, através da análise do dendrograma, observando a distância obtida entre as ligações realizadas ao se considerar um determinado número de grupos. Mingoti (2005) apresenta outras formas de se definir o número ideal de grupos, são elas: a análise do comportamento do nível de fusão (distância), análise do comportamento do nível de similaridade, análise da soma de quadrados entre grupos (coeficiente  $R^2$ ), a estatística pseudo F, correlação semiparcial, estatística pseudo  $T^2$  e estatística CCC (*Cubic Clustering Criterium*).

### **3 Aplicação a dados de temperatura e precipitação de Presidente Prudente**

As técnicas para definição de anos padrão foram aplicadas aos dados anuais médios de Temperatura mínima e máxima (°C) e Precipitação total anual (mm) de Presidente Prudente (SP), referentes ao período de 1961 a 2010, obtidos junto à Estação Meteorológica de Presidente Prudente, da FCT/UNESP.

Os dados observados no período são apresentados na Tabela 1 dada a seguir.

Tabela 1 – Dados anuais de Temperatura mínima (Tmin),  
Temperatura máxima (Tmax) e Precipitação (Precip).

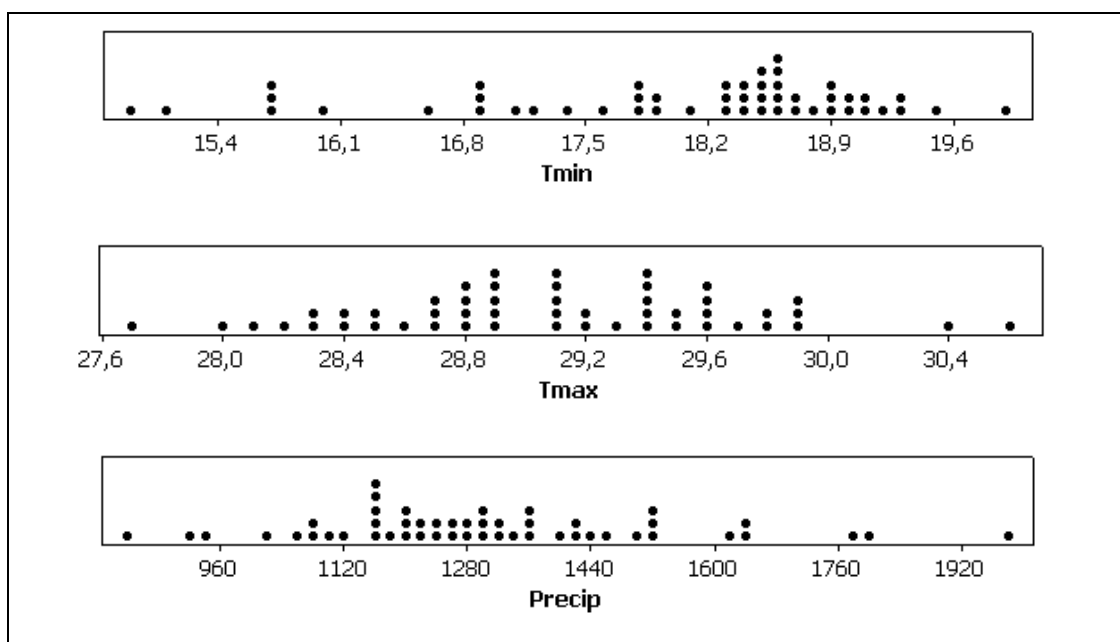
Obs	Ano	Tmin	Tmax	Precip	Obs	Ano	Tmin	Tmax	Precip
1	1961	16,6	29,9	1206,2	26	1986	19,0	29,2	1169,5
2	1962	14,9	28,4	1254,3	27	1987	18,4	28,7	1415,8
3	1963	15,1	29,9	1097,9	28	1988	18,3	29,4	1196,4
4	1964	15,7	29,4	1520,7	29	1989	17,8	28,0	1802,4
5	1965	16,0	29,4	1624,3	30	1990	18,5	28,8	1302,8
6	1966	15,7	28,8	926,0	31	1991	18,7	29,4	938,1

7	1967	16,9	29,3	1152,5	32	1992	18,6	28,2	1493,5
8	1968	15,7	28,6	1396,7	33	1993	18,6	28,9	1196,4
9	1969	17,6	29,5	1065,4	34	1994	18,9	29,8	1334,9
10	1970	16,9	29,1	1298,8	35	1995	19,1	29,6	1185,8
11	1971	16,9	28,7	1268,4	36	1996	18,8	28,9	1522,7
12	1972	17,8	28,1	1783,2	37	1997	19,0	29,1	1462,0
13	1973	17,4	28,8	1303,6	38	1998	18,9	28,9	1649,0
14	1974	17,2	28,3	1647,4	39	1999	18,3	29,7	1227,5
15	1975	17,8	29,1	1160,8	40	2000	18,6	29,2	1317,8
16	1976	17,1	27,7	1424,7	41	2001	19,1	29,4	1239,6
17	1977	18,6	29,5	1276,5	42	2002	19,9	30,6	1166,1
18	1978	17,9	29,1	1012,6	43	2003	18,9	29,6	1279,0
19	1979	17,9	28,5	1089,2	44	2004	18,6	28,9	1232,0
20	1980	18,4	28,7	1369,9	45	2005	19,3	29,6	1081,7
21	1981	18,1	28,9	1114,3	46	2006	19,2	29,9	1313,4
22	1982	18,4	28,5	1521,0	47	2007	19,5	30,4	1355,9
23	1983	18,5	28,4	1350,7	48	2008	18,3	28,8	1153,6
24	1984	18,7	29,8	1216,5	49	2009	19,3	29,1	1973,0
25	1985	18,5	29,6	841,9	50	2010	18,5	28,3	1432,4

Fonte: Autoria própria.

O *Dot Plot* é um diagrama pontual que permite a visualização horizontal de como as observações da variável se distribuem na reta. Cada ponto representa uma observação, ou valor obtido em um determinado ano para cada variável. Na Figura 8 estão dispostos os *Dot Plots* das três variáveis apresentadas na Tabela 1.

Figura 8 – Diagrama pontual de Tmin, Tmax e Precip.



Fonte: Autoria própria.

Foi realizada uma análise estatística descritiva para os dados apresentados na Tabela 1, envolvendo as medidas de tendência central: média, mediana; além das medidas de dispersão: coeficiente de variação e desvio-padrão, os quartis, valores máximos e mínimos observados, amplitude total (At), coeficiente de assimetria (As) e curtose (K). As estatísticas citadas podem ser encontradas em Bussab e Morettin (2002). Um resumo das estatísticas obtidas é apresentado na Tabela 2.

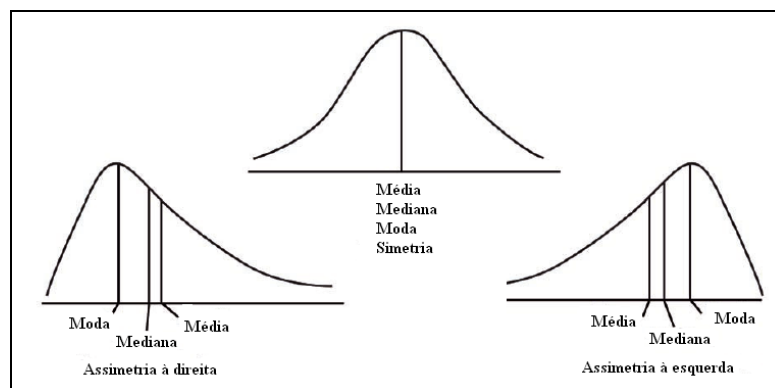
Tabela 2 – Resumo estatístico das variáveis: Temperatura Mínima Média (Tmin), Temperatura Máxima Média (Tmax) e Precipitação (Precip)

Variável	Media	Desvio Padrão (DP)	CV (%)	Min	Q1 (25%)	Md Q2 (50%)	Q3 (75%)	Max	Amplitude Total (At)	Assimetria (As)	Curtose (K)
Tmin	18,008	1,18059826	6,55596544	14,9	17,4	18,4	18,8	19,9	5	-1,0638364	0,48535019
Tmax	29,088	0,61066708	2,09937803	27,7	28,7	29,1	29,5	30,6	2,9	0,09614291	-0,0686119
Precip	1307,296	226,507417	17,3264063	841,9	1166,1	1277,75	1424,7	1973,0	1131,1	0,68171719	0,86066174

Fonte: Autoria própria.

A medida de assimetria (As) é utilizada para afirmar se a distribuição dos dados é aproximadamente igual à esquerda e à direita da média, nesse caso  $As=0$ ; ou se a assimetria é positiva ( $As>0$ ) para a qual os valores localizados à direita da média, em direção aos maiores valores observados, têm um maior espalhamento do que os valores localizados à esquerda; ou se a assimetria é negativa ( $As<0$ ) e ocorre maior espalhamento à esquerda da média, em direção aos menores valores observados. Na Figura 9 estão representados os três tipos de curvas possíveis.

Figura 9 – Relações entre a média, mediana e a moda



Fonte: Adaptado de WMO-n.100 (2011).

Analisando a Tabela 2, pode-se dizer que:

- Para a variável Tmin,  $As = -1,0638364 < 0$  indicando que a distribuição dos dados é assimétrica à esquerda;
- Tmax apresenta  $As = 0,09614291 \sim 0$  é aproximadamente simétrica e;
- Precip apresenta assimetria positiva, pois  $As = 0,68171719 > 0$ .

A medida de curtose (K) é utilizada para avaliar o achatamento de uma curva, comparada à curva com distribuição normal: uma curva normalmente achatada é denominada mesocúrtica se  $K=0$ ; uma curva excessivamente achatada é denominada platicúrtica se  $K>0,263$ ; e uma curva muito afilada é denominada leptocúrtica se  $K<0,263$ .

Para os dados de Presidente Prudente, tem-se:

- Para a variável Tmin,  $K=0,48535019 > 0$  indicando que a curva é platicúrtica;
- Tmax apresenta  $K=-0,0686119 \sim 0$  é aproximadamente mesocúrtica e;
- Precip também é platicúrtica, pois  $K=0,86066174 > 0$ .

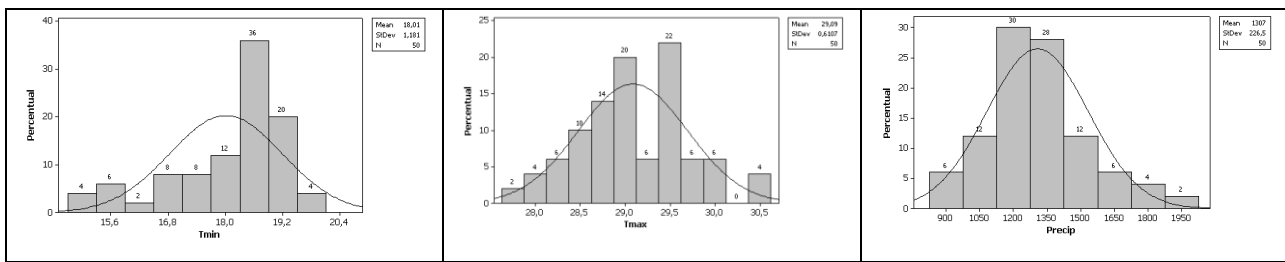
Analisando ambas as medidas de assimetria e curtose tem-se:

- Tmin apresenta assimetria negativa e curva platicúrtica;
- Tmax é simétrica e mesocúrtica;
- Precip possui assimetria positiva e curva platicúrtica.

É importante salientar que as medidas de assimetria e curtose dizem respeito à curva ou forma de distribuição dos dados, e isso influencia os resultados obtidos com as técnicas de construção de classes, que são o objeto de estudo deste artigo. Na Figura 10 são apresentados os histogramas com o ajuste da curva de distribuição normal, nos quais é possível observar a forma das curvas para as três variáveis: Tmin é assimétrica a esquerda (em direção aos valores menores valores de temperatura), Tmax é aproximadamente simétrica, e Precip é assimétrica a direita (em direção aos valores maiores de precipitação).

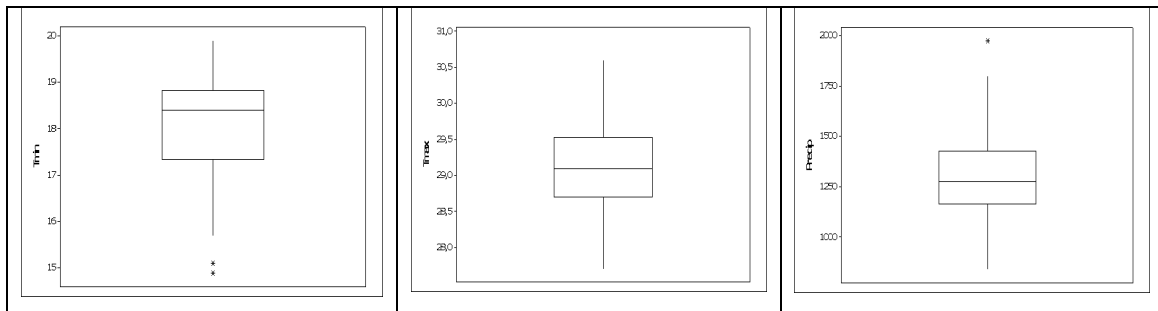
Analisando a Figura 11, que apresenta os *Box Plots* de cada variável, nota-se que para Tmin (a esquerda), e a mediana não está no centro da caixa, indicando forma assimétrica da curva de distribuição dos dados, em direção aos menores valores de temperatura e a presença de dois valores muito baixos de temperatura, indicados por \*, e que representam *outliers* em relação as observações restantes. O gráfico *Box Plot* no centro, representa Tmax, e a caixa equilibrada em relação à mediana, indicando simetria. A variável Precip, no *Box Plot* à direita na Figura 11, apresenta um *outlier* e a mediana não está centralizada indicando assimetria (em direção aos maiores valores da variável).

Figura 10 – Histogramas e ajuste de curva para a distribuição normal, para as variáveis Tmin, Tmax e Precip



Fonte: Autoria própria.

Figura 11 – Box Plots, para as variáveis Tmin, Tmax e Precip



Fonte: Autoria própria.

### 3.1 Tmin

Foram construídas classes para Tmin, utilizando-se todos os métodos apresentados na Seção 2, partindo-se dos dados organizados em uma lista crescente, do menor para o maior valor.

As classes de temperatura construídas foram intituladas como segue: F=ano frio, MF=ano moderadamente frio, H=ano habitual, MQ=ano moderadamente quente e Q=ano quente.

Inicialmente foi aplicada a técnica de construção de classes baseada na Fórmula de Sturges para  $k=5$  classes, e os resultados foram organizados na Tabela 3. Devido ao fato da distribuição de Tmin ser assimétrica à esquerda, as frequências relativas observadas também apresentam valores assimétricos, pois nas classes mais baixas de temperatura, F e MF, as frequências são baixas, com valores de 10% e 4%, respectivamente. As frequências aumentam para 20% na classe central H, e tem a maioria das observações 42% na classe MQ, voltando a diminuir para 20% para a última classe dos maiores valores de temperatura Q. Portanto, a técnica de construção de classes utilizando a Fórmula de Sturges é capaz de refletir a forma assimétrica de distribuição dos dados para Tmin, conforme pode ser observado na última coluna da Tabela 3.

Tabela 3 – Distribuição de frequências para Tmin, utilizando a Fórmula de Sturges

Classes	Intervalos de Classes	Freq. Observada $f_i$	Freq. Relativa %	Gráfico de Barras para $f_i$
F	[14,9; 15,9)	5	10	IIII
MF	[15,9; 16,9)	2	4	II











Na Tabela 10 são apresentadas as classes construídas segundo o critério dos Quantis. Observa-se que os valores observados ficaram bastante próximos dos esperados. Conforme pode ser visto no gráfico de barras para as frequências observadas, a curva produzida é simétrica e bastante achatada.

Tabela 10 – Distribuição de frequências para Tmax, utilizando Quantis e os Percentis P15, P35, P65 e P85.

Definição das Classes de Frequências	Intervalos de Classes	Freq. Esp. %	Freq. Obs. $f_i$	Freq. Rel. %	Gráfico de Barras para $f_i$
$\text{Min} \leq F < P15 = [\text{Min a P15}]$	[27,7; 28,4)	15	6	12	
$P15 \leq TF < P35 = [P15 a P35]$	[28,4; 28,8)	20	8	16	
$P35 \leq H < P65 = [P35 a P65]$	[28,8; 29,4)	30	17	34	
$P65 \leq TQ < P85 = [P65 a P85]$	[29,4; 29,7)	20	11	22	
$Q \geq P85 = [P85 a \text{max}]$	[29,7; 30,7]	15	8	16	
Total		100%	50	100%	

Fonte: Autoria própria.

A Tabela 11 apresenta a distribuição dos anos segundo as classes formadas utilizando-se a técnica do Desvio-padrão. Nota-se um valor acima do esperado para a classe MQ.

Tabela 11 – Distribuição de frequências para Tmax, utilizando a técnica do Desvio-padrão

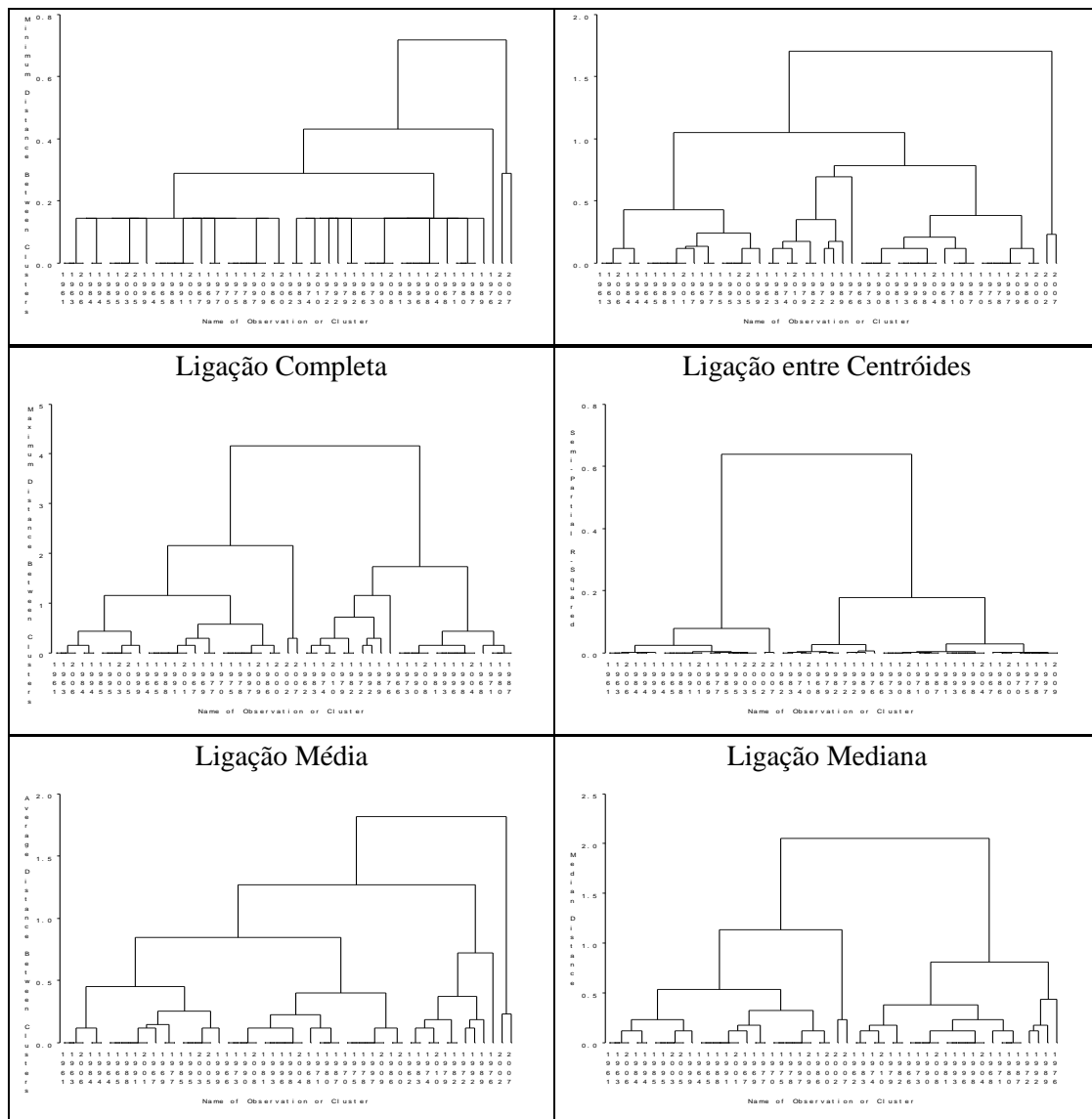
Definição das Classes de Frequências	Intervalos de Classes	Freq. Esp. % normal	Freq. Obs. $f_i$	Freq. Rel. %	Gráfico de Barras para $f_i$
$\text{Min} \leq F < \text{média} - dp$	[27,7; 28,5)	16	8	16	
$\text{Média} - dp \leq MF < \text{média} - 1/2dp$	[28,5; 28,8)	15	6	12	
$\text{Média} - 1/2dp \leq H < \text{média} + 1/2dp$	[28,8; 29,4)	38	17	34	
$\text{Média} + 1/2dp \leq MQ < \text{média} + dp$	[29,4; 29,7)	15	11	22	
$Q \geq \text{média} + dp$	[29,7; 30,7]	16	8	16	
Total		100 %	50	100%	

Fonte: Autoria própria.

Na Figura 13 são apresentados os dendrogramas para os métodos AA aplicados a Tmax.

Figura 13 - Dendrogramas para os seis métodos de agrupamentos hierárquicos aplicados a Tmax.

Ligação Simples	Ligação Ward
-----------------	--------------



Fonte: Autoria própria.

Analisando a Tabela 12, nota-se que os métodos de agrupamentos construíram classes com diferentes formas de distribuição da curva. Com o método de Ligação Simples a distribuição em cinco classes apresentou uma forma assimétrica a direita, e duas últimas classes foram construídas com uma única observação em cada uma delas. Nos demais métodos a última classe sempre agrupou os dois últimos valores, justamente os maiores valores de temperatura máxima (ver Figura 8 - *Dot Plots*). Ward classificou muitas observações na primeira classe (F), com frequência 11 enquanto todos os demais métodos apresentaram bem menos observações: 4 para a Ligação Mediana, e 1 para os demais métodos. Os métodos de Ligação Média e entre Centróides formaram exatamente os mesmos grupos.

Tabela 12 – Distribuição de frequências observada ( $f_i$ ) e acumulada (%) para Tmax aplicando a técnica AA

L.Simples	Ward	L.Completa	L.Média	L.Centróide	L.Mediana
-----------	------	------------	---------	-------------	-----------







Note que na Tabela 17 que o método de Ligação Simples, atribui um grande percentual a classe MS, 86% muito discrepante da maioria dos métodos. No método Centróide também há um grande percentual na classe MS, 64%, enquanto nos demais varia de 34 a 44%. Porém, todos os métodos captam a assimetria da curva original dos dados.

Tabela 17 – Distribuição de frequências observada ( $f_i$ ) e acumulada (%) para Precip, com a técnica AA

Classes	L.Simples		Ward		L.Completa		L.Média		L.Centróide		L.Mediana	
	$f_i$	%	$f_i$	%	$f_i$	%	$f_i$	%	$f_i$	%	$f_i$	%
S	1	2,0	4	8,0	4	8,0	3	6,0	3	6,0	9	18,0
MS	43	86,0	18	36,0	22	44,0	19	38,0	32	64,0	17	34,0
H	3	6,0	17	34,0	18	36,0	22	44,0	9	18,0	18	36,0
MC	2	4,0	8	16,0	5	10,0	5	10,0	5	10,0	5	10,0
C	1	2,0	3	6,0	1	2,0	1	2,0	1	2,0	1	2,0
Total	50	100,0	50	100,0	50	100,0	50	100,0	50	100,0	50	100,0

Fonte: Autoria própria.

A Figura 17 contendo um quadro resumo geral permitirá verificar as diferentes classificações obtidas para Precip com cada uma das técnicas. Um resumo das técnicas será apresentado na Seção 3.4 a seguir.

### 3.4 Resumo geral para as técnicas de classificação aplicadas

Para a classificação em cinco classes de Tmin foram aplicadas as seguintes técnicas: Fórmula de Sturges, *Box Plot*, Quantil, um Desvio-padrão (Desvio), além de vários métodos de agrupamentos: Ligação Simples (LSim), Ward, Ligação Completa (LCom), Ligação Média (LMedia), Centróide (Centroide) e Mediana (Mediana). O método Ligação Simples apresentou resultados muito discrepantes dos demais métodos. Após a aplicação de todas as técnicas foi construída a Figura 15, tal que a coluna Maioria representa a maioria das classificações para cada ano. Analisando a Figura 15, pode-se notar que os anos que apresentaram a mesma classificação de Tmin em praticamente todos os métodos foram:

- anos frios (F): 1962 a 1966 e 1968.
- anos moderadamente frios (MF): 1961, 1967, 1970, 1971, 1974 e 1976.
- anos habituais (H): 1969, 1972, 1973, 1975, 1978, 1979, 1981 e 1989.
- anos moderadamente quentes (MQ): 1977, 1980, 1982 a 1988, 1990 a 2001, 2003, 2004, 2008 e 2010.
- anos quentes (Q): 2002, 2005 a 2007 e 2009.



Para a classificação em cinco classes de Tmax, foram aplicadas as mesmas técnicas já apresentadas em Tmin. Os resultados são apresentados na Figura 16, na qual nota-se que os anos que apresentaram a mesma classificação de Tmax na maioria das técnicas foram:

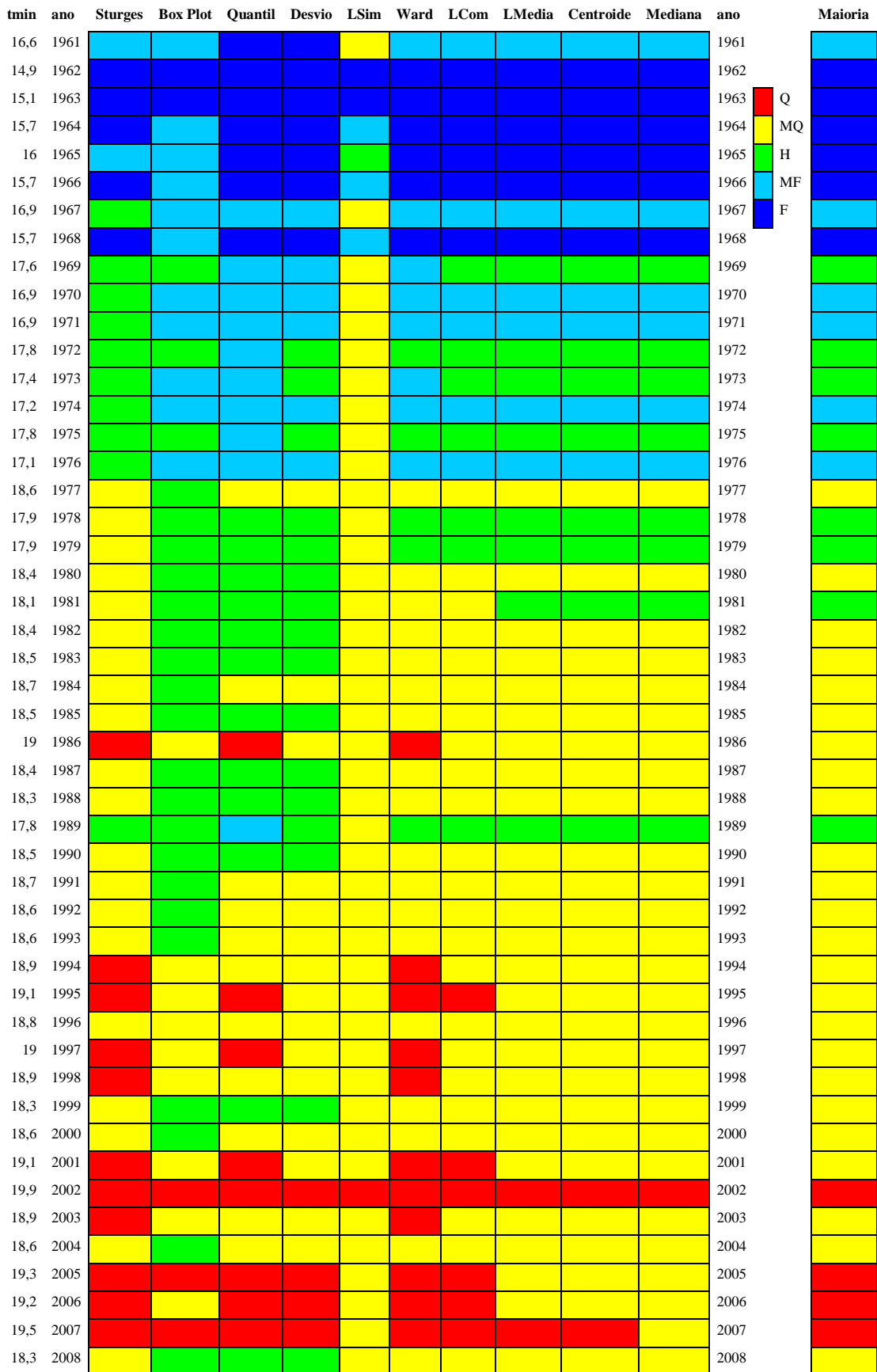
- anos frios (F): 1972, 1976, 1989 e 1992.
- anos moderadamente frios (MF): 1962, 1968, 1971, 1974, 1975, 1980, 1982, 1983, 1987 e 2010
- anos habituais (H): 1966, 1967, 1970, 1973, 1975, 1978, 1981, 1986, 1990, 1993, 1996 a 1998, 2000, 2004, 2008 e 2009.
- anos moderadamente quentes (MQ): 1961, 1963, 1964, 1969, 1977, 1984, 1985, 1988, 1991, 1994, 1995, 1999, 2001, 2003, 2005 e 2006.
- anos quentes (Q): 2002 e 2007.

Para a classificação em cinco classes de Precip, foram aplicados os mesmos métodos utilizados em Tmin e Tmax. Os resultados são apresentados na Figura 17, na qual, pode-se notar que os anos que apresentaram a mesma classificação de Precipitação na maioria das técnicas foram:

- anos secos (S): 1966, 1978, 1985 e 1991.
- anos moderadamente secos (MS): 1961, 1963, 1967, 1969, 1975, 1979, 1981, 1984, 1986, 1988, 1993, 1995, 1999, 2001, 2002, 2004, 2005 e 2008.
- anos habituais (H): 1962, 1968, 1970, 1971, 1973, 1976, 1977, 1980, 1983, 1987, 1990, 1992, 1994, 1997, 2000, 2003, 2006, 2007 e 2010.
- anos moderadamente chuvosos (MC): 1964, 1965, 1974, 1982, 1996, 1998,
- anos chuvosos (C): 1972, 1989, 2009.

Alguns anos apresentaram um consenso total entre todos os métodos, entre eles: para Tmin, 1962 e 1963 foram anos frios (F), 2002 foi quente (Q); para Tmax, 1976 foi frio (F), e 2002 e 2007 foram quentes (Q); e para Precipitação, os anos 1966, 1985 e 1991 foram extremamente secos (S) e 2009 o ano mais chuvoso (C). Houve também classificações em consenso para classes intermediárias: para Tmax, 1985, 1995, 2003 e 2005 foram moderadamente quentes (MQ); para Precipitação, 1967, 1975, 2002 e 2008 foram moderadamente chuvosos (MC).

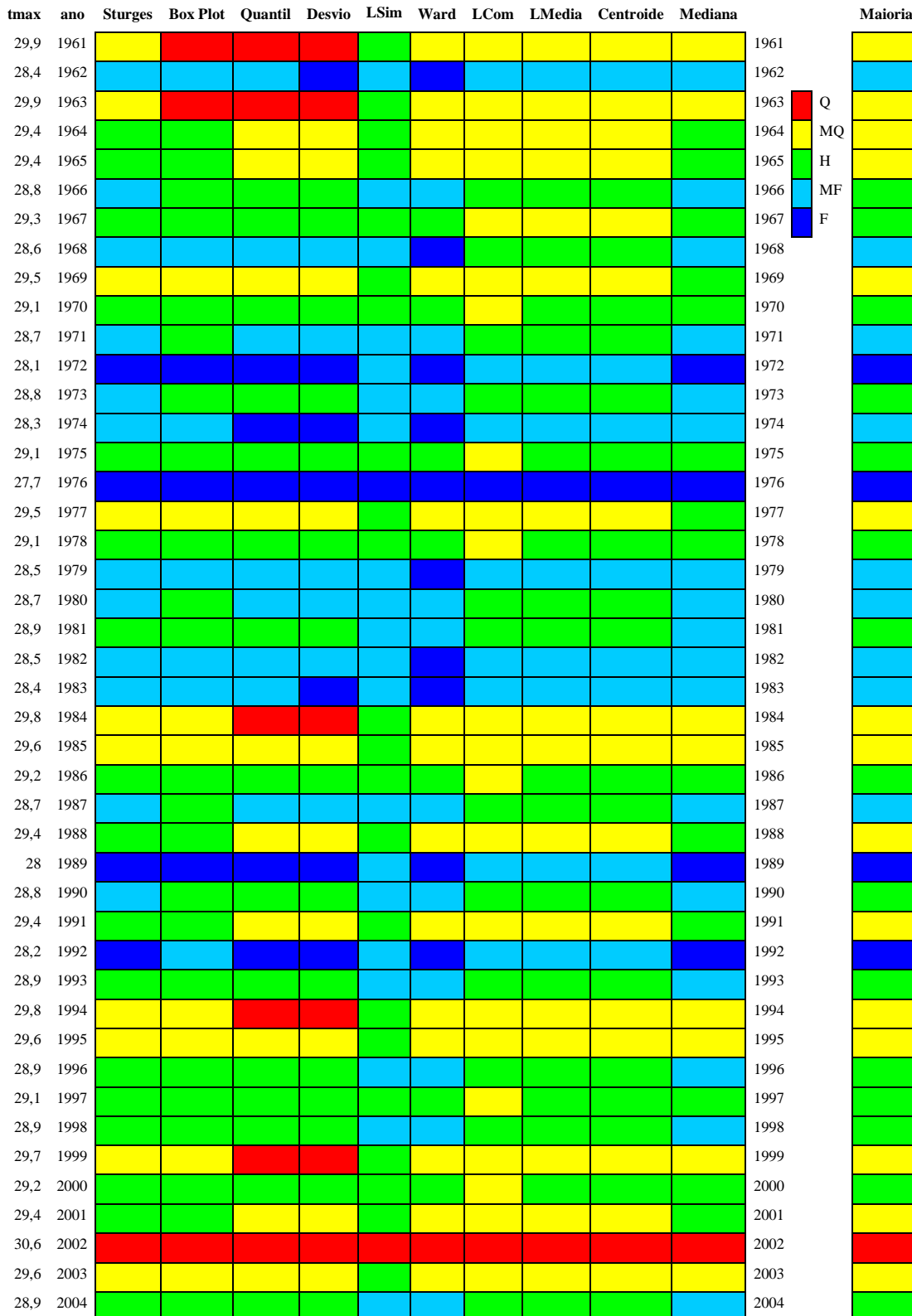
Figura 15 – Classificação em cinco classes distintas, dos anos de 1961 a 2010 de Temperaturas mínimas médias (Tmin) de Presidente Prudente, através de diferentes métodos de classificação.

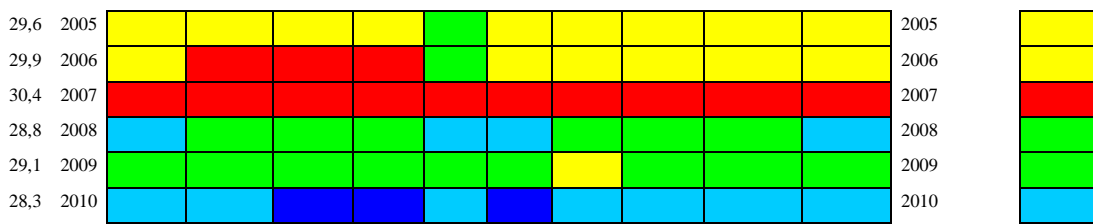




Fonte: Autoria própria.

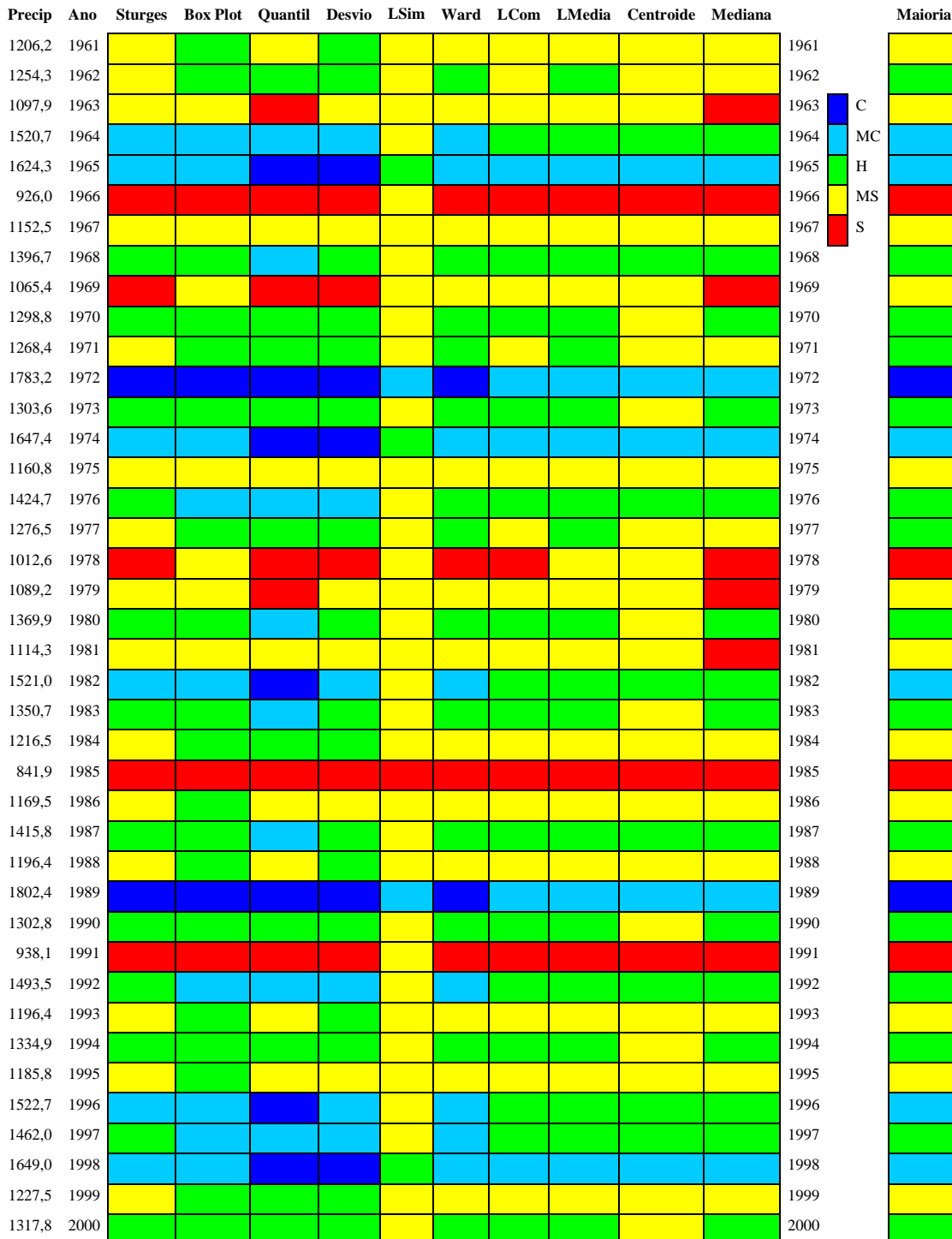
Figura 16 – Classificação em cinco classes distintas, dos anos de 1961 a 2010 de Temperaturas máximas (Tmax) de Presidente Prudente, através de diferentes técnicas de classificação.

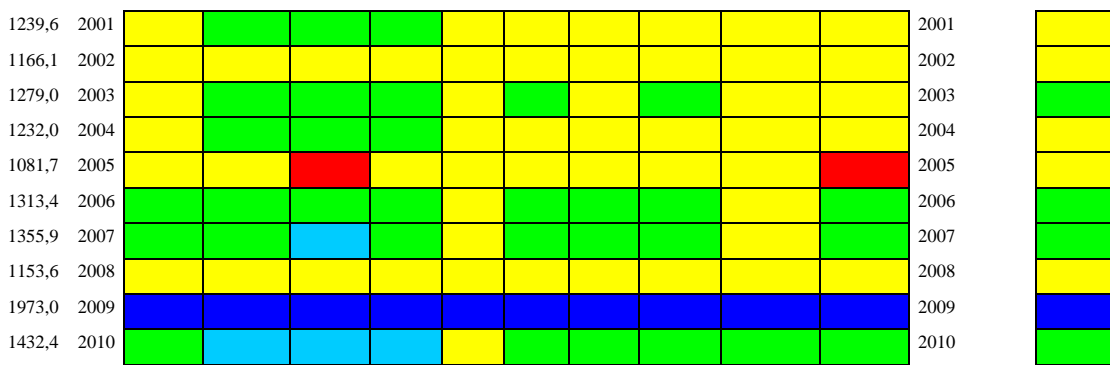




Fonte: Autoria própria.

Figura 17 - Classificação em cinco classes distintas, dos anos de 1961 a 2010 de Precipitação de Presidente Prudente, através de diferentes métodos de classificação.





Fonte: Autoria própria.

#### 4 Conclusão

Neste artigo foram apresentadas várias técnicas para construção de classes a partir da análise da série histórica de temperatura e precipitação de Presidente Prudente, referente ao período de 1961 a 2010. Observou-se que não há um consenso entre todas as técnicas, devido às características inerentes as definições de cada um. Os agrupamentos produzidos pelas técnicas de Sturges e de um Desvio-padrão e os métodos de análise de agrupamentos hierárquicos (AA) tendem a gerar percentuais de classes com frequências que reproduzem a forma simétrica ou assimétrica presente na distribuição original dos dados. Já as técnicas *Box Plot* e *Quantis* produzem agrupamentos cujas frequências observadas e relativas formem uma distribuição simétrica, mesmo que a distribuição original dos dados não seja, como ocorreu com T<sub>min</sub> e Precip. A técnica *Box Plot* gera classes que apresentam curvas mais afiladas, pois os percentuais de 5% para as classes mais extremas são muito baixos, e são destinados 50% dos dados à classe central; enquanto que, a técnica *Quantis* produz curvas bastante achatadas, justamente pelos altos percentuais nos extremos da distribuição que os percentis P15 e P85 impõem as classes, e os dados restantes precisam ser divididos ainda em três classes.

Com relação aos anos extremos, todos os métodos conseguiram detectar quais são os anos que estão nos limites extremos das observações, ou seja, os anos com temperatura extremamente frios (F) ou quentes (Q), e anos extremamente secos (S) ou chuvosos (C).

Dentre os métodos de agrupamento notou-se que o método de Ligação Simples apresentou uma classificação diferente dos demais métodos, com elevados percentuais em uma classe mais interna, além de produzir classes nos extremos com poucos elementos.

Quanto aos softwares utilizados no trabalho, o Excel<sup>®</sup> tem várias fórmulas estatísticas implementadas, dentre as que foram utilizadas podemos citar: mínimo, máximo, média, desvio-padrão, mediana e percentis. Porém, para o cálculo do coeficiente de assimetria, curtose, construção dos gráficos *Box Plot*, histogramas, dendrogramas e aplicação dos métodos de agrupamentos hierárquicos, é preciso um software estatístico para executá-los, como por exemplo o software R<sup>®</sup> gratuito, ou outros softwares pagos tais como: Minitab<sup>®</sup>, Statistica<sup>®</sup>, SPSS<sup>®</sup>, SAS<sup>®</sup>, etc. Neste trabalho foram utilizados: Minitab<sup>®</sup> para a construção das medidas de assimetria e curtose e dos gráficos *Box Plots* e histogramas e SAS<sup>®</sup> para a análise de agrupamentos e dos dendrogramas.

Para uma análise final na definição dos anos padrão para a série histórica estudada, uma solução encontrada foi construir uma classificação levando-se em conta a maioria das classificações realizadas por todos os métodos para cada ano.

Ao invés de utilizar uma única série de médias ou totais anuais para a construção dos anos padrão, como foi realizado neste artigo, outra maneira de aplicar as técnicas de agrupamento hierárquico (AA) poderia incluir as séries de médias mensais, ou até mesmo sazonais (primavera, verão, outono, inverno), pois tais técnicas permitem a incorporação de mais de uma variável ao estudo sendo necessário a aplicação de AA somente uma única vez para cada método de ligação. Dessa forma, os resultados poderiam ser mais precisos por levar em conta informações mais qualitativas, como a sazonalidade das estações, por exemplo, no momento de classificar os anos padrão. E nesta situação, todos os demais métodos (Sturges, *Box Plot*, Quantis, Desvio-padrão) teriam que ser aplicados separadamente para cada mês ou estação climática, o que tornaria muito mais trabalhosa a síntese dos resultados com estes métodos do que com as técnicas de AA.

Outra possibilidade de aplicação das técnicas AA relaciona-se ao estudo de séries de várias estações meteorológicas, de localidades distintas, simultaneamente de forma bastante rápida e realizada numa única análise, permitindo a construção de regiões com comportamentos similares, como apresentado em Jácomo (2011), que encontrou regiões de precipitação para a região oeste paulista bem próximas às obtidas por Sant'Anna Neto (1995).

Uma comparação entre os métodos com limites de classes rígidos como é o caso das técnicas Sturges, *Box Plot*, Quantis e Desvio, os quais definem valores pré-determinados para os limites das classes tem desvantagem sobre os métodos AA, que tem maior flexibilidade neste aspecto, permitindo que pontos próximos (em relação à distância euclidiana) entre si permaneçam na mesma classe, ao invés de provocar uma separação abrupta entre eles em duas classes distintas e vizinhas.

## **Referências Bibliográficas**

- BARBOSA, J. P. M. **Avaliação de técnicas empíricas e estatísticas de identificação de extremos de precipitação para o litoral paulista e entorno**. 2008. 99 f. Dissertação (Mestrado em Geografia) – Instituto de Geociências, Universidade Estadual de Campinas. Campinas.
- BOIN, M. N. **Chuvas e erosões no oeste paulista: uma análise climatológica aplicada**. 2000. 264 f. Tese (Doutorado em Geociências e Meio Ambiente) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista Júlio de Mesquita Filho, Rio Claro.
- BUSSAB, W. O.; MORETTIN, P. A.; **Estatística básica**. 5. ed. São Paulo: Saraiva, 2002. 526p.
- DINIZ, J. A. F. Classificação de uma variável e sua aplicação na Geografia. **Boletim de Geografia Teórica**. Rio Claro, n. 1, p. 25-39, 1971.
- FLORES, E. F. **Sistema de Informação Climatológica: desenvolvimento e inserção no sistema de informação geográfica GEO-INF+MAP**. 1995. 86 f. Dissertação (Mestrado em Geociências e Meio Ambiente) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista Júlio de Mesquita Filho, Rio Claro.
- FLORES, E. F. **Modelagem em climatologia geográfica: um ensaio metodológico aplicado ao oeste paulista**. 2001. Tese (Doutorado em Geociências e Meio Ambiente) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista Júlio de Mesquita Filho, Rio Claro.
- GALVANI, E.; LUCHIARI, A. Critérios para classificação de anos com regime pluviométrico normal, seco

- e úmido. In: GALVANI, E.; LIMA, N. G. B. **Climatologia aplicada: resgate aos estudos de caso**. 1. Ed. Curitiba: CRV, 2012. p. 4-29.
- GERARDI, L. H. O.; SILVA, B. C. M. N. **Quantificação em geografia**. São Paulo: Difel, 1981. 161p.
- JÁCOMO, C. A. **Modelos geoestatísticos espaço-tempo aplicados a dados pluviométricos no oeste do Estado de São Paulo**. 2011. 109 f. Dissertação (Mestrado em Engenharia Cartográfica) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente.
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 5.ed. New Jersey: Prentice Hall, 2002. 767p.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005. 295p.
- MONTEIRO, J. B. **Chover, mas chover de mansinho: desastres naturais e chuvas extremas no Estado do Ceará**. 2011. 198 f. Dissertação (Mestrado em Geografia) – Centro de Ciências e Tecnologia, Universidade Estadual do Ceará. Fortaleza.
- SANCHES, M. C. A. A problemática dos intervalos de classes na elaboração de cartogramas. **Boletim de Geografia Teórica**, Rio Claro, n. 1, p. 53-65, 1972.
- SANT'ANNA NETO, J. L. **Ritmo climático e a gênese das chuvas na zona costeira paulista**. 1990. 156f. Dissertação (Mestrado em Geografia) Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.
- SANT'ANNA NETO, J. L. **As chuvas no estado de São Paulo: contribuição ao estudo da variabilidade e tendência da pluviosidade na perspectiva da análise geográfica**. 1995. 202f. Tese (Doutorado em Geografia) Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.
- STURGES, H.A. The Choice of a Class Interval: **Jour. Am. Stat. Assoc.**, n. 21, p. 65-66, 1926 apud GERARDI, L. H. O.; SILVA, B. C. M. N. **Quantificação em geografia**. São Paulo: Difel, 1981. 161p.
- TAVARES, A. C. Critérios de escolha de anos padrão para análise rítmica. **Geografia**, Rio Claro, v. 1, n. 1, p. 79-87, 1976.
- WMO (World Meteorological Organization). **Guide to climatological practices**. WMO n.100. Geneva: WMO, 2011, 117p.
- XAVIER, T. M. B. S.; XAVIER, A. S. F. Caracterização de períodos secos ou excessivamente chuvosos no Estado do Ceará através da técnica dos quantis: 1964-1998. **Rev. Bras. Meteorologia**, v. 14, n.2, 63-78, 1999.
- XAVIER, T. M. B. S. **Tempo de chuva: estudos climáticos e de Previsão para o Ceará e Nordeste Setentrional**. Fortaleza: ABC, 2001, 478p.
- ZAVATINI, J. A. **A dinâmica atmosférica e a distribuição das chuvas no Mato Grosso do Sul**. 1990. 223 f. Tese (Doutorado em Geografia Física) – Faculdade de Filosofia Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.