ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

QUAL TESTE DE CORRELAÇÃO É MAIS ADEQUADO: PEARSON OU SPEARMAN?

MIRIAM RODRIGUES SILVESTRE[1]

Resumo É comum em climatologia a necessidade de se correlacionar elementos climáticos com outras variáveis. Por exemplo, para um determinado recorte temporal e espacial, a precipitação influenciou a produção de uma cultura? Na área da saúde, é possível afirmar que as temperaturas mais altas ocorrem conjuntamente com o aumento do número de internações (ou óbitos) de certos tipos de doença? Porém, para se afirmar que duas variáveis apresentam correlação estatisticamente significativa é necessária a comprovação através de um teste de hipóteses estatístico, tais como: o Teste de Correlação de Pearson e o Teste de Spearman. Este trabalho traz indicações de como o pesquisador ou usuário de estatística deve verificar as pressuposições necessárias para aplicar os testes corretamente e não correr o risco de apresentar conclusões incorretas em sua pesquisa.

Palavras chave: correlação estatística, teste de correlação de Pearson, teste de correlação de Spearman, teste de normalidade.

Abstract: It is common in climatology the need to relate climatic elements to other variables. For example, for a given temporal and spatial area, the precipitation influenced the production of a culture? In health, it is clear that the highest temperatures occur in conjunction with the increase in the number of hospital admissions (or deaths) of certain types of disease? However, to say that two variables have statistically significant correlation is necessary to prove by testing statistical hypotheses, such as the Pearson Correlation Test and the Spearman test. This article provides instructions on how the researcher or user should check the statistical assumptions required to apply the tests correctly and not run the risk of having incorrect conclusions in his research.

Keywords: statistical correlation, Pearson Correlation Test, Spearman Correlation Test, normality test.

1 – Introdução

A análise do comportamento de duas variáveis em conjunto é geralmente realizada por meio da medida de correlação entre ambas. Entretanto, quando o objetivo é avaliar se a correlação é significativa estatisticamente, devem ser empregados testes estatísticos de correlação, e não somente calcular a medida de correlação entre elas. É comum trabalhos científicos apresentarem os resultados do Teste de Correlação de Pearson, mas nada afirmarem sobre se as suas pressuposições foram verificadas e confirmadas como verdadeiras para os dados em estudo. Alguns trabalhos trazem apenas o valor da estatística de correlação e já fazem conclusões sobre seus resultados, sem nem ao menos informarem

[1] Doutoranda em Geografia e Profa. do Dep. Estatística, UNESP – Presidente Prudente. E-mail: miriam@fct.unesp.br

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

o nível de significância adotado ou o valor obtido no teste, nessas situações é difícil dar crédito científico às conclusões apresentadas pela pesquisa. Nesse trabalho apresentamos dois testes de correlação, de Pearson e de Spearman, porém justificando quando se pode aplicar um ou o outro teste. Isto será feito através da avaliação das pressuposições necessárias ao teste de Pearson, e indicando o de Spearman, quando as pressuposições não forem satisfeitas.

2 – Discussão

Gregory (1968) trata do problema de correlação no capítulo 11. Embora o autor desenvolva o Teste de Correlação Momento Produto de Pearson adequadamente, nem sequer comenta a necessidade da distribuição normal bivariada para os dados. Porém, o autor apresenta o Teste de Spearman como sendo útil para situações nas quais não há uma medida numérica para uma ou até mesmo as duas variáveis. É o caso de uma variável apresentar uma escala de medida ordinal, ou seja, apenas a ordem da variável é informada. Por exemplo, suponha que a variável X é apenas a ordem de importância de uma região, numa lista de n regiões, enquanto que Y é o número de itens que cada região possui. Então, para obter a variável X não foi realizada nenhuma medida, somente organizar em ordem de importância as regiões, segundo algum critério. Esse tipo de escala de medida é denominado escala ordinal, e um teste de Pearson não poderia ser aplicado em tal situação.

King (1969) apresenta uma seção de análise de correlação como parte de um capítulo sobre Análise de Regressão. O autor apresenta o coeficiente de correlação, e informa que na teoria formal é assumido que os dados representam uma amostra aleatória de uma população normal bivariada. Porém, não apresenta mais detalhes sobre como verificar a normalidade, nem tampouco outro teste não-paramétrico para correlação (como o Teste de Spearman ou de Kendall), pois sua atenção está voltada à análise de regressão.

Gerardi e Silva (1981) apresentam a seção 2.8.4 dedicada ao conceito de correlação, na qual apresentam o Teste de Correlação de Pearson e o de Spearman. As autoras alertam que:

> Devemos ainda considerar um fato importante. O coeficiente de correlação foi desenvolvido a partir de uma distribuição normal bidimensional. Se temos outras distribuições ele não é claramente interpretável e a interpretação deve ser feita com cuidado. A pressuposição de normalidade é necessária se queremos fazer a inferência estatística. Em caso de acentuado desvio da normalidade podemos transformar os dados, por exemplo, através de logaritmos, para conseguir a pressuposição da normalidade ou utilizar o coeficiente de correlação que vamos apresentar a seguir. (GERARDI e SILVA, 1982, p.104).

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

E as autoras seguem apresentando o Coeficiente de Correlação de Spearman. As autoras não apresentam nenhum teste para avaliar a normalidade individual ou bivariada.

Rogerson (2012) apresenta um capítulo sobre correlação muito bem escrito e avalia a presença de outliers (pontos discrepantes) de forma bastante cuidadosa:

> O tratamento dos valores discrepantes depende das circunstâncias. Um bom entendimento subjacente do por quê pontos específicos são discrepantes fornece alguns fundamentos para a remoção desses pontos da análise. Nesse caso, temos uma explicação razoavelmente boa para a discrepância e temos motivos para perguntar como seria a correlação sem os valores discrepantes...
> E sugere que é uma boa prática representar graficamente as variáveis, e pensar cuidadosamente sobre as razões de qualquer discrepância. (ROGERSON, 2012, p. 189).

autor apresenta o teste de correlação de Pearson, e fala que existe a suposição de que os dados de cada variável vêm de distribuições normais. Também acrescenta que as observações devem ser independentes para cada variável, ou seja, uma observação de X não afeta e não é influenciada por outros valores de X. O mesmo ocorre para Y. Essa preocupação tem toda razão de ser principalmente quando as variáveis X e Y provêm de localizações espaciais, pois nessa situação a suposição de independência não pode ser satisfeita, pois dados espaciais geralmente apresentam dependência.

Rogerson (2002) traz também o Teste de Correlação por Postos de Spearman, e o indica nas situações em que apenas os dados de posição estão disponíveis, ou quando a suposição de normalidade requerida para o teste de correlação de Pearson não é satisfeita. Embora tenha sido bastante cuidadoso na apresentação do testes de correlação, o autor não apresenta como avaliar se a suposição de normalidade é satisfeita.

Após descrever como os testes de correlação são apresentados por alguns autores selecionados pela disponibilidade no acesso ao material bibliográfico, serão discutidos nas próximas seções os testes de correlação de Pearson e de Spearman, bem como brevemente os testes de normalidade.

2.1 – Teste de Correlação Momento Produto de Pearson

Para variáveis quantitativas X e Y, o coeficiente de correlação é uma medida adequada para se avaliar o grau de associação linear entre essas duas variáveis. O verdadeiro e desconhecido coeficiente de correlação para a população é definido por $\rho = \rho(X,Y) = \dfrac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$. Porém, embora seja desconhecido, o coeficiente populacional *ρ* pode ser estimado através do coeficiente amostral *r*, calculado a partir de uma amostra retirada de uma população normal bidimensional (Bussab e Morettin, 2002).

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

Para se avaliar se a correlação populacional é nula ou não, são construídas as hipóteses nula (Ho) e alternativa (Ha), e as hipóteses são arranjadas como segue: $\begin{cases} Ho: \rho = 0 \\ Ha: \rho \neq 0 \end{cases}$.

Em Ho, as variáveis são independentes, logo a correlação entre elas é nula; e em Ha, as variáveis são dependentes, e então, existe alguma correlação não nula. O Teste de Correlação Momento Produto de Pearson, mais conhecido como Teste de Correlação de Pearson, é adequado para testar tais hipóteses. O teste é baseado na estatística *r* dada na equação (1), a qual pode apresentar valores entre -1 e 1:

$$r = \frac{\text{cov}(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i}^{n}(Y_i - \overline{Y})^2}} . \quad (1)$$

Se *r*<0 indica que a correlação entre as duas variáveis é negativa; e caso *r*>0, indica correlação positiva entre X e Y, a correlação igual a -1 ou 1 indica correlação perfeita negativa ou positiva, e *r*=0 indica ausência total de correlação.
A estatística *r* pode ser aproximada pela distribuição *t* de Student, com *n*-2 graus de liberdade, através da expressão (2):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} . \quad (2)$$

A Tabela 01 apresenta os valores de *r* para diferentes tamanhos amostrais (*n*) e mantendo um nível de significância de 5%. Os valores de *r* tabelados são obtidos através da expressão (3):

$$r^2 = \frac{t^2}{n-2+t^2} , \text{ e } r = \sqrt{r^2} \quad (3)$$

Pode-se observar que os valores de *r* diminuem conforme aumenta o tamanho da amostra *n*. Finalizando, nem sempre é possível obter valores altos para a correlação, pois esse está intimamente relacionado ao tamanho da amostra.

Caso o valor de *r* calculado na amostra pela equação (3), seja inferior a –*r* ou superior a *r* (valores apresentados na Tabela 1), pode-se rejeitar Ho em favor da hipótese alternativa (Ha) e afirmar que existe correlação não nula entre as variáveis X e Y, caso contrário, a correlação entre as variáveis é nula.

n	t_{n-2}	r	n	t_{n-2}	r
5	3,18245	0,878339	60	2,00172	0,254204
10	2,30600	0,631897	70	1,99547	0,235198
15	2,16037	0,513977	80	1,99085	0,219901
20	2,10092	0,443763	90	1,98729	0,207246
25	2,06866	0,396070	100	1,98447	0,196551
30	2,04841	0,361007	120	1,98027	0,179343
35	2,03452	0,333845	130	1,97867	0,172277

IBSN: 978-85-7846-278-9 p. 1965 – 1977

40	2,02439	0,312006	150	1,97612	0,160335
45	2,01669	0,293955	180	1,97338	0,146319
50	2,01063	0,278711	180	1,97202	0,138789

Tabela 01 – Valores mínimos de *r*, de acordo com o tamanho amostral, para um nível de significância de 5%.

Uma consideração a ser feita é que a distribuição do par de variáveis aleatórias (X,Y) deve ser normal bivariada, e para avaliar essa suposição existem testes adequados para tal, e caso a distribuição seja verdadeira para a amostra observada, então, o Teste de Correlação de Pearson é adequado. Caso contrário, existem testes não-paramétricos, que dispensam tal exigência, e também podem ser aplicados em situações onde as variáveis X e Y não tenham escala de medida intervalar ou razão, seria o caso de uma escala ordinal, na qual os dados se constituem em valores ordenados, ou até mesmo a ordem das observações segundo alguma característica. Os testes não-paramétricos são: Teste de Correlação de Spearman e Teste de Correlação de Kendall. Na próxima seção será apresentado o Teste de Spearman, cuja estatística é bem parecida com a do Teste de Pearson, o que facilita bastante seu entendimento, se comparada à estatística do Teste de Kendall.

2.2 – Teste de Correlação de Spearman

O Teste de Correlação de Spearman, é um teste não-paramétrico desenvolvido por Spearman (1904). Para um teste de hipóteses bilateral, as hipóteses nula (Ho) e alternativa (Ha) do Teste de Correlação de Spearman são as mesmas que as descritas anteriormente, porém, a estatística do teste é baseada nas ordens (postos ou ranks) de cada uma das variáveis ordenadas individualmente, R(Xi) é a ordem de Xi na variável X ordenada em ordem crescente, onde o menor valor de X recebe a ordem (ou posto) 1, e assim sucessivamente, até que o maior valor receba o posto n. O mesmo deve ser feito para Y separadamente, e R(Yi) representa o posto de Yi para os dados de Y ordenados. Se não houver empates, a estatística do Teste de Correlação de Spearman é dada por:

$$r = 1 - \frac{6 \sum_{i=1}^{n} \left[R(X_i) - R(Y_i) \right]^2}{n(n^2 - 1)} . \qquad (4)$$

No caso de empates entre os valores de Xi, a média dos postos para as observações empatadas deverá substituir os postos das mesmas. A mesma observação é válida para Yi. E para calcular a estatística de Spearman, para dados com empates, basta substituir os dados originais por seus postos e calcular a estatística de Pearson dada na equação (1). Para avaliar a significância do teste, para pequenas amostras n≤30, existe uma tabela

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

específica para o Teste de Spearman (ver Tabela A10, em Conover, 1999). No caso de grandes amostras, a estatística do teste pode ser aproximada pela distribuição Normal Padrão (ver Tabela A1, em Conover, 1999) através de $r^* = r\sqrt{n-1}$. No caso de hipótese alternativa bilateral (Ha: ρ≠0), rejeita-se a hipótese nula em favor de Ha, se ocorrer r* for menor ou igual a $z_{\alpha/2}$ ou r* for igual ou superior a $z_{1-\alpha/2}$. Por exemplo, se o nível de significância desejado do teste é α=0,05 ou 5%, então α/2=0,025 e $z_{\alpha/2}$=-1,96 e $z_{1-\alpha/2}$=1,96. E rejeita-se Ho se r*≤-1,96 ou r*≥1,96.

Outra forma de decidir sobre a rejeição ou não de Ho, é avaliando o pvalor do teste, que será dado por: $pvalor = 2P\left(Z \geq |r|\sqrt{n-1}\right)$. Geralmente, rejeita-se Ho quando pvalor≤0,05 ou 5%. Por exemplo, para n=180 e correlação *r**=-0,1089887, tem-se:

$$pvalor = 2P\left(Z \geq |-0,1089887|\sqrt{180-1}\right) =$$

$$2P(Z \geq 1,45817) = 2(0,0723969) = 0,144794 \sim 0,145 \text{ ou } 14,5\%.$$

Nesse caso, o pvalor é de aproximadamente 14,5% e nesse caso, não se rejeita Ho, e conclui-se que a correlação entre as variáveis X e Y é nula.

Uma observação a ser feita é que o coeficiente de Correlação de Spearman, não avalia apenas relação linear entre X e Y, ou seja, o teste é adequado para testar a existência de uma relação não linear também. Caso exista a correlação pode-se afirmar, que se a hipótese Ho é rejeitada em favor de Ha, isso significa que ocorre uma das duas situações: a) há uma tendência para valores grandes de X também ocorrerem de forma pareada com valores grandes de Y, ou seja, ocorre uma correlação positiva, ou b) há uma tendência para valores pequenos de X ocorrerem de forma pareada com valores grandes de Y, indicando uma correlação negativa entre as variáveis X e Y. O sinal da correlação pode ser observado na estatística do teste *r**.

2.3 – Testes de Normalidade

A aplicação do Teste de Correlação de Pearson tem como pressuposição a normalidade das variáveis X e Y que se deseja correlacionar, tanto individualmente (univariada) como em dupla (X,Y) (bivariada). Para avaliar se um conjunto de dados apresenta uma distribuição normal podem ser realizados testes estatísticos próprios para esta finalidade. Existem vários testes para normalidade, dentre eles destacam-se os Testes de Shapiro-Wilk, Anderson-Darling, Kolmogorov-Smirnov, etc. E para normalidade multivariada, podem ser citados o Teste de Shapiro-Wilk Multivariado e o Teste de Assimetria e Curtose de Mardia.

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

Nos testes de normalidade, a hipótese nula é definida como **Ho: Os dados seguem a distribuição normal**, e na hipótese alternativa **Ha: Os dados não seguem a distribuição normal**. Então, caso os dados sejam normais, a hipótese nula não deverá ser rejeitada.

Os software estatístico R apresenta diversos testes para normalidade univariada e o pacote “mvnormtest” realiza o Teste de Normalidade Multivariado de Shapiro-Wilk (SWM), o qual é uma extensão do Teste de Shapiro-Wilk (SW) para normalidade univariada.

Não será possível apresentar os testes de normalidade, porém, o Teste SW pode ser encontrado em Johnson e Wichern (2007), e o teste SWM em Royston (1982).

3 – Resultados

Para ilustrar a aplicação dos testes de correlação, foi analisado um conjunto de dados obtidos do INMET, referentes à Estação Convencional de Presidente Prudente (SP), referentes ao período Janeiro de 2006 a Dezembro de 2009, totalizando n=48 observações mensais de cinco variáveis: precipitação total, pressão média, temperatura máxima média, temperatura mínima média, umidade média. A Figura 01 apresenta gráficos de séries temporais das cinco variáveis climáticas de interesse.

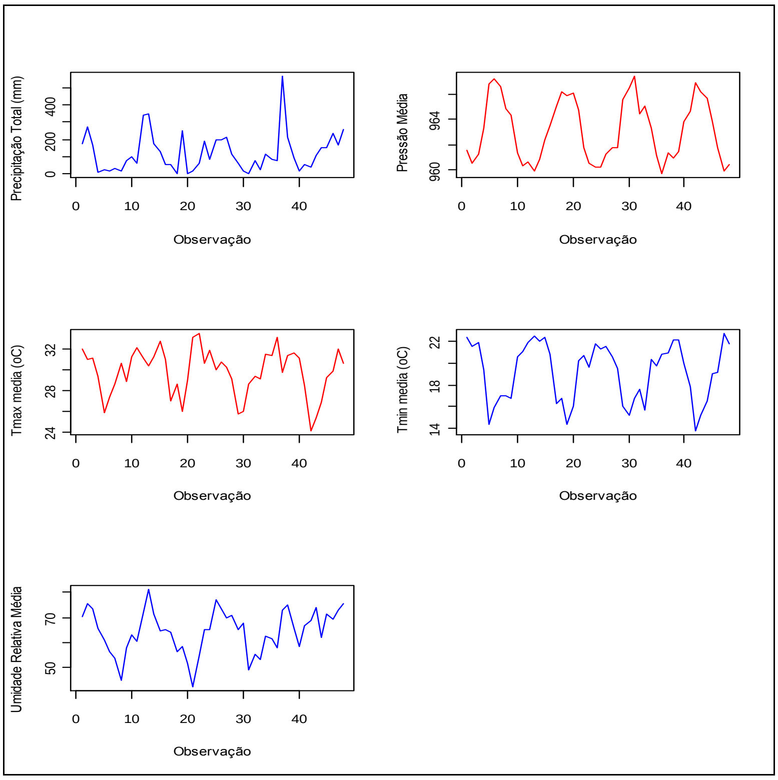

Figura 01 – Gráficos de séries temporais para as variáveis em estudo.

Para se averiguar a correlação entre duas variáveis, o gráfico de dispersão apresentado na Figura 02 é o mais indicado. Analisando a Figura 02, pode-se observar que aparentemente

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

parece haver correlação negativa entre Pressão e Precipitação, Pressão e Temp. Máxima Média, Pressão e Temp. Mínima Média, Pressão e Umidade Relativa Média. E correlação positiva entre Precipitação e Umidade Relativa Média, Temp. Máxima Média e Temp. Mínima Média.

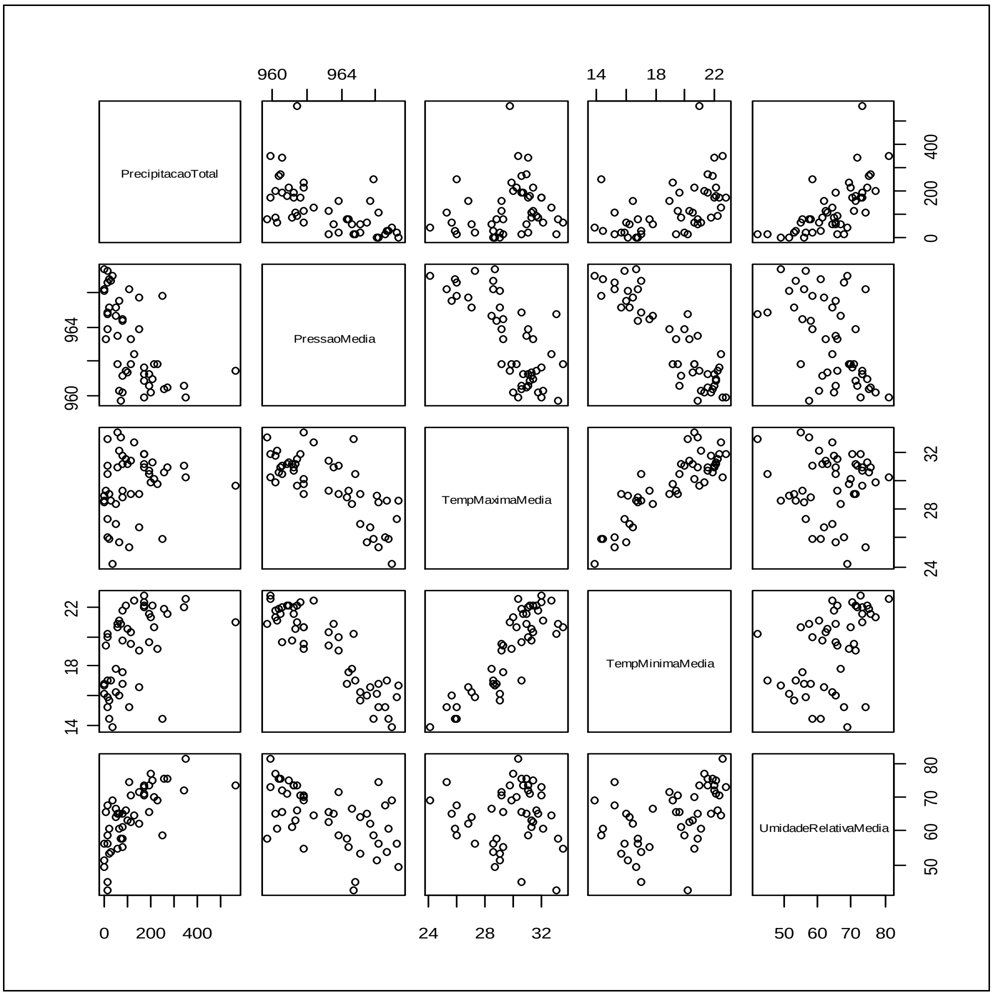

Figura 02 – Gráficos de dispersão para pares de variáveis.

Para avaliar se as correlações entre as variáveis são estatisticamente significativas serão realizados testes de correlação, porém, para decidir sobre qual teste é indicado entre os Testes de Correlação de Pearson ou de Spearman, será necessário primeiramente, analisar as variáveis individualmente quanto à normalidade de sua distribuição.

A análise visual foi feita através de gráficos Boxplot, os quais possibilitam a visualização da maneira como os dados se distribuem, se tem uma forma simétrica ou assimétrica de distribuição. A Figura 03 apresenta gráficos Boxplot para as cinco variáveis. Lembrando que uma distribuição normal deve apresentar simetria em sua forma de distribuição.

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

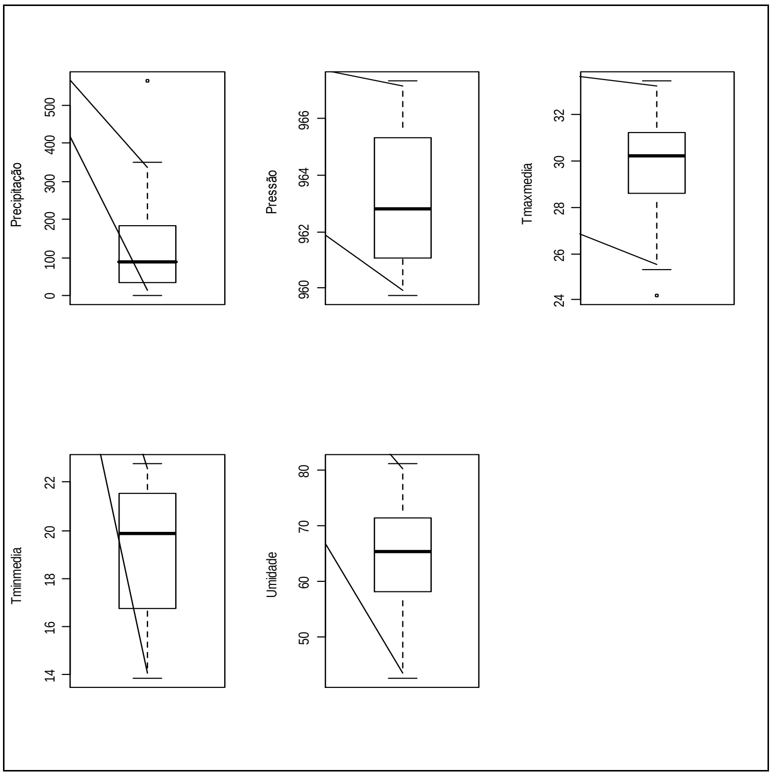

Figura 03 – Gráficos Boxplot para todas as variáveis

A análise dos Boxplots permite considerar que somente a variável umidade tem distribuição simétrica, necessária para a ocorrência de uma distribuição normal. As demais variáveis apresentam-se assimétricas, pois as caixas (Box) apresentam proporções internas desiguais, quando comparadas as linhas centrais das caixas.

Para garantir a normalidade além da análise visual, também foi aplicado o Teste de Normalidade de Shapiro-Wilk para cada variável individualmente. O Teste de Shapiro-Wilk, apresentou estatísticas e pvalor conforme a Tabela 02.

Tabela 02 - Resumo dos Testes de Normalidade Univariada de Shapiro-Wilk

Variáveis	Estatística W	Pvalor
Precipitação	0,8646	5,446e-05*
Pressão	0,9132	0,001728*
Tmaxmedia	0,9494	0,03769*
Tminmedia	0,914	0,00184*
Umidade	0,977	0,4620

Os Pvalores indicados por * são pequenos (<0,10) indicando que as variáveis não apresentam distribuição normal. Somente para Umidade Pvalor>0,10, indicando que essa variável pode ser considerada normalmente distribuída.

Para analisar a normalidade bivariada dos dados, deve ser construído o gráfico de dispersão para o par de variáveis de interesse, e verificado se aparece uma forma de elipse ou mais circular, na distribuição dos dados. A Figura 02, já apresentada anteriormente, possibilita a análise visual da distribuição dos pares de variáveis. Aparentemente, pode-se suspeitar de que haja normalidade bivariada para alguns pares de variáveis. Porém, para garantir a normalidade bivariada, foram realizados testes SWM, cujos resultados podem ser consultados na Tabela 03, a seguir.

Variáveis	Precipitação	Pressão	Tmaxmedia	Tminmedia
Pressão	0,72 3.068e-12*			
Tmaxmedia	0,6658 1,827e-13*	0,641 5,582e-14*		
Tminmedia	0,6656 1,81e-13*	0,6413 5,667e-14*	0,9195 1,910e-05*	
Umidade	0,7078 1,582e-12*	0,6478 7,69e-14*	0,8371 6,827e-09*	0,8136 1,145e-09*

Tabela 03 - Resumo dos Testes de Normalidade Bivariada de Shapiro-Wilk

Observa-se na Tabela 03 que nenhum par de variáveis pode ser considerado normal bivariado, pois os todos os Pvalores são muito pequenos, próximos de zero.

Então, após toda essa análise inicial, conclui-se que não é possível aplicar o Teste de Correlação de Pearson para nenhum par entre as variáveis em estudo, devido ao não atendimento das pressuposições desse teste. Poderiam ser tomadas outras medidas, com o objetivo de alcançar a normalidade necessária para a aplicação do Teste de Pearson. Algumas possibilidades, seriam: retirar *outliers* (valores extremos) que aparecem como “o” nos gráficos Boxplot, que estão presentes na variável Precipitação e Temperatura Máxima Média; ou então, realizar transformações nas variáveis, que são na verdade, funções matemáticas aplicadas aos valores reais, de forma que essas consigam transformar os dados em simétricos e reais. Após a retirada de observações, ou qualquer transformação executada, a normalidade deve ser novamente avaliada, e convém alertar que nem sempre é possível chegar à normalidade mesmo realizando as ações sugeridas. Uma observação a ser feita é conforme sugeriu Rogerson (2012) é verificar o valor extremo, notou-se que para precipitação, o valor obtido para o mês de janeiro de 2009 (563,8 mm) é realmente elevado para a região, e esse ano foi realmente um ano atípico de chuvas elevadas em várias regiões brasileiras. Poderia ser refeita toda a análise sem a observação de janeiro de 2009. Já com relação à temperatura máxima média de 24,17^{o}C, referente ao mês de junho de

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

2009, é um valor baixo para Presidente Prudente, porém, não se sabe se o mesmo ocorreu para as demais regiões próximas, entretanto é um fato a ser verificado.

Uma opção na falta de normalidade é aplicar o Teste de correlação não-paramétrico de Spearman, apresentado na Seção 2.2. O Teste de Spearman substitui os dados originais por postos no momento de calcular sua estatística *r*, dada na equação (4). Os resultados dos Testes de Spearman estão organizados na Tabela 4, a seguir. Em cada célula da Tabela 4 são dispostos dois valores, o superior é a correlação estimada e o inferior o Pvalor do teste de Spearman. Os valores em azul representam pares com correlação negativa estatisticamente significativa e valores em vermelho indicam pares com correlação positiva estatisticamente significativa.

Variáveis	Precipitação	Pressão	Tmaxmedia	Tminmedia
Pressão	-0,6569327 3,955435e-07*			
Tmaxmedia	0,2612967 0,07282814	-0,7344768 2,818612e-09*		
Tminmedia	0,5753135 1,890510e-05*	-0,8538009 1,243450e-14*	0,8268787 4,447553e-13*	
Umidade	0,7487177 9,404284e-10*	-0,5353886 8,818281e-05*	0,06317846 0,6696693	0,5062013 0,0002418761*

Tabela 04 - Resumo dos Testes de Correlação de Spearman

Nota-se que praticamente todos os pares de variáveis apresentam correlação significativa, exceto os pares (Precipitação,Tmaxmedia) e (Tmaxmedia, Umidade).

Uma observação a ser feita é que caso se queira avaliar a correlação unilateral, por exemplo, no caso de existir uma justificativa teórica para se acreditar que a correlação seja positiva, então, pode-se fazer um teste unilateral à direita e testar diretamente essa correlação, através das hipóteses: $\begin{cases} Ho: \rho = 0 \\ Ha: \rho > 0 \end{cases}$.

Por outro lado, se se acredita que a correlação seja negativa, pode-se fazer o teste unilateral à esquerda, e testar as hipóteses: $\begin{cases} Ho: \rho = 0 \\ Ha: \rho < 0 \end{cases}$.

O ganho é com relação ao Pvalor que será menor por considerar as probabilidades de um único lado, e poderá captar uma correlação como significativa, quando se fosse aplicado um teste bilateral $\begin{cases} Ho: \rho = 0 \\ Ha: \rho \neq 0 \end{cases}$, a hipótese nula não seria rejeitada.

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

Se os testes forem realizados em software estatísticos (no software R, por exemplo) é possível solicitar um teste de correlação unilateral, e se for manual, basta consultar as tabelas com o nível de significância α que leve em consideração um teste unilateral, pois nos testes bilaterais, geralmente, deve-se utilizar $\alpha/2$, mas isso irá depender da tabela de distribuição da estatística do teste disponível no momento.

4 – Conclusões

Nesse trabalho apresentamos dois testes de correlação: o Teste de Correlação de Pearson, um teste que exige a normalidade aos dados para que possa ser aplicado; e o Teste de Correlação de Spearman, não-paramétrico, e que por ser de distribuição livre, não exige nenhuma pressuposição com relação à distribuição dos dados.

Cabe ao usuário de estatística, realizar os testes adequados para que suas análises não sejam prejudicadas, pois caso o Teste de Pearson seja aplicado a dados que não atendam a pressuposição de normalidade, os resultados poderão ser significativos, indicando presença de correlação quando essa na verdade não existe, ou caso contrário, não indicar correlação quando essa realmente existe.

O usuário mais desavisado poderá incorrer em outro tipo de erro, como por exemplo, aplicar diretamente um teste de correlação não-paramétrico, dessa forma evitando o trabalho de se checar a normalidade, porém, também estará sujeito aos mesmos tipos de erro que acabamos de explicar, pois aplicar um teste não-paramétrico quando a distribuição dos dados é normal, pode não detectar significância estatística nos testes quando deveria. Além disso, o teste não-paramétrico tem menor poder (probabilidade de rejeitar Ho, dado que ela é realmente falsa) que os testes paramétricos quando a distribuição é normal. Portanto, a melhor maneira de um pesquisador usar a estatística a seu favor, é verificando cuidadosamente todas as pressuposições exigidas pelo teste que deseja aplicar, e não aplicá-lo sem que as mesmas sejam atendidas, pois como vimos, há testes alternativos como os testes não-paramétricos; ou meios de se alcançar a normalidade, tais como a retirada de observações *outliers* ou transformação de variáveis. Porém, o usuário não deve se esquecer de fazer a checagem novamente das pressuposições, e assegurar que as mesmas são verídicas caso opte por essas duas últimas alternativas.

ANAIS DO X SIMPÓSIO BRASILEIROS DE CLIMATOLOGIA GEOGRÁFICA
IBSN: 978-85-7846-278-9 p. 1965 – 1977

5 – Referências Bibliográficas

BUSSAB, W. de O.; MORETTIN, P. A. **Estatística básica**. 5 ed. São Paulo: Saraiva, 2002. 526 p.

CONOVER, W. J. **Practical nonparametric statistics**. 3rd ed. New York: John Wiley & Sons, 1999. 584p.

GERARDI, L. H. de O.; Silva, B-C N. **Quantificação em geografia**. São Paulo: DIFEL, 1981. 161p.

GREGORY, S. **Statistical methods and the geographer**. London: Longman, 1968. 277p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6th. ed. Upper Saddle River: Pearson, 2007. 773p.

KING, L. J. **Statistical analysis in geography**. Englewoods Cliffs: Prentice Hall, 1969. 288p.

ROGERSON, P. A. **Métodos estadísticos para geografia: um guia para o estudante**. Peter A. Rogerson; tradução: Paulo Fernando Braga Carvalho, José Irineu Rangel Rigotti. 3 ed. Porto Alegre: Bookman, 2012. 348p.

ROYSTON, P. An extension of Shapiro and Wilk’s W test for normality to large samples. **Applied Statistics**, 31, p. 115-124, 1982.

SPEARMAN, C. The proof and measurement of association between two things. **American Journal of Psychology**, 15, p. 72-101, 1904.