

Uso do teste de aleatorização para comparar dois grupos considerando teste não paramétrico

Jurandir Prazeres Filho*, Denise Nunes Viola**, Gilênio Borges**

*Aluno de graduação em Estatística, UFBA / Bolsista de Iniciação Científica - CNPq

**Professor(a) do Departamento de Estatística, UFBA

viola@ufba.br

Resumo Muitas vezes o pesquisador está interessado em comparar médias ou a forma da distribuição de dois grupos. Uma maneira para compará-los seria aplicando testes paramétricos, tais como o Teste T ou Teste Z (no caso de duas amostras independentes) ou o Teste T pareado. Porém, tais testes apresentam certas exigências que freqüentemente podem não ser atendidas. Neste caso, é indicada a utilização de testes não paramétricos ou o teste de aleatorização. Este teste é baseado na suposição de que, se a hipótese nula é verdadeira, todas as possíveis ordens dos dados são igualmente prováveis. O teste de aleatorização é um procedimento em que se comparam valores de uma estatística observada para os dados no arranjo original com os valores desta estatística após a aleatorização das observações. A regra de decisão é baseada no p-valor - proporção de vezes em que a estatística de teste com os aleatorizados é maior ou igual a estatística de teste com os dados do arranjo original. Se o p-valor for menor que o nível de significância, rejeita-se H_0 . É importante escolher adequadamente a estatística de teste e como neste estudo foram comparadas as médias de duas amostras independentes e pequenas e as exigências para o uso de testes paramétricos não foram atendidas, a estatística utilizada foi a do teste não paramétrico Wilcoxon-Mann-Whitney. Dentre as vantagens em se utilizar o teste de aleatorização, destaca-se o uso em amostras não aleatórias e/ou amostras pequenas, porém seu resultado não pode ser generalizado para a população. Observa-se ainda que o teste de aleatorização não apresenta tantas exigências quanto os métodos convencionais. Para ilustração deste teste foi utilizado um conjunto de dados de plantas de milho, em que as variáveis estudadas foram as alturas da plantas. Essas alturas foram medidas no vigésimo dia após sua germinação. Foram cultivadas quatro plantas à sombra e cinco ao sol e o objetivo foi verificar se o ambiente à sombra ou ao sol influencia em seu crescimento. Após a aplicação do teste de aleatorização considerando a estatística do teste de Wilcoxon-Mann-Whitney e 10.000 aleatorizações obteve-se p-valor=0,9666. Como este valor é maior que o nível de significância ($\alpha = 0,05$), então não há evidências suficientes para rejeitar H_0 , ou seja, as amostras são provenientes da mesma população, o que equivale a afirmar que há evidências de que o ambiente não influencia no crescimento das plantas.

1 Introdução

Os testes paramétricos e não-paramétricos são técnicas de inferência estatística. A primeira é mais antiga que a segunda, que teve seu início nos primeiros anos do século XX, quando aparecem os primeiros testes não paramétricos (CAMPOS, H. 1979). Em inferência estatística tem-se dois aspectos: a estimação de parâmetros e o teste de hipóteses, sendo o último o foco deste estudo. Para utilizar os testes paramétricos é necessário o atendimento a alguns pressupostos, dentre eles: os dados devem ser provenientes de uma

população com distribuição normal, porém, como nem sempre isto é possível, o pesquisador pode optar por transformar as variáveis estudadas a fim de identificar normalidade ou utilizar testes não-paramétricos. Existem vantagens e desvantagens do uso de testes não-paramétricos em relação aos paramétricos, dentre elas destaca-se como principal vantagem o fato dos testes não paramétricos não dependerem da distribuição populacional dos dados e como desvantagem principal a sua menor eficiência (TRIOLLA, 2005). É importante ressaltar que, além do pressuposto de normalidade, é necessário para a aplicação dos testes paramétricos e dos testes não paramétricos (com exceção o teste de aleatorização) que a amostra seja aleatória. Hoje existe um número grande de testes paramétricos e não paramétricos, porém aqui serão abordados apenas o segundo tipo, no caso de teste para duas populações independentes, pois o estudo utiliza pequenas amostras. O objetivo deste trabalho é verificar, através do uso do teste de aleatorização, se os ambientes de sol e sombra diferem estatisticamente com relação ao crescimento de plantas de milho. Este trabalho está organizado como segue: na Seção 2 está apresentada a metodologia, com uma revisão bibliográfica de testes não-paramétricos, teste de aleatorização e teste de Wilcoxon-Mann-Whitney. Na Seção 3 encontram-se os resultados e discussões e, finalmente, na Seção 4 é apresentada a conclusão. Foi estudado o caso de duas amostras independentes utilizando o teste de aleatorização considerando a estatística do teste de Wilcoxon-Mann-Whitney e 10.000 aleatorizações. As análises dos dados foram feitas no *software* R.

2 Material e Métodos

2.1 Material

Para ilustrar o uso dos testes de aleatorização foram cultivadas nove plantas de milho no período de 25 de janeiro de 2010 a 11 de fevereiro de 2010, e suas alturas foram medidas em centímetros no dia 11 de fevereiro. Destas plantas, após a germinação (ao sol), quatro foram cultivadas a sombra e cinco ao sol. O objetivo foi avaliar se o cultivo à sombra difere estatisticamente do cultivo ao sol em relação ao crescimento da planta.

2.2 Metodologia

2.2.1 Testes não paramétricos

Os testes estatísticos usuais são aplicáveis quando a população tem uma distribuição normal ou aproximadamente normal. Quando isto ocorre e os dados são mensurados em escalas intervalares utilizam-se os testes paramétricos, que por sua vez dependem da distribuição de probabilidade da população. Como exemplo para dados quantitativos, no caso de testes envolvendo duas amostras, tem-se o Teste Z, Teste T para duas amostras independentes e o Teste T pareado. Porém, muitas vezes, não é adequado aplicar tais testes pois a distribuição da população da qual a amostra foi retirada é desconhecida; quando isso ocorre, utilizam-se os testes não-paramétricos. A principal vantagem dos testes não-paramétricos é que prescindem da distribuição original dos dados, sendo por isso chamados de testes livres de distribuição. Dentre outras vantagens estão: os cálculos são mais simples; podem ser aplicados para dados categóricos ou dados que podem somente ser medidos em escala ordinal; têm menor sensibilidade a erros de medida quando comparados com os métodos tradicionais; é possível utilizar dados de diferentes populações (em alguns testes não-paramétricos) e têm maior eficiência que os paramétricos se os dados populacionais não seguem distribuição normal. Além disso, os testes paramétricos apresentam certas exigências como os dados amostrais devem provir uma população com distribuição normal ou aproximadamente normal;

enquanto que para os testes não-paramétricos as pressuposições são mais flexíveis (TRIOLOLA, 2005). A desvantagem dos métodos não-paramétricos é que estes, na maioria das vezes, não consideram os valores efetivos das observações, pois freqüentemente reduzem os dados a uma forma qualitativa como no caso do teste dos sinais, ou seja, perde-se informações (TRIOLOLA, 2005). Apesar das vantagens do uso dos testes não-paramétricos, se as exigências para um método paramétrico forem satisfeitas é mais vantajosa a utilização do último devido a sua maior eficiência relativa ou poder. O poder de um teste é a probabilidade de rejeitar a hipótese nula, quando esta é realmente falsa. Neste estudo foi realizado teste de hipótese para duas populações independentes.

2.2.2 Teste de aleatorização

Uma forma de investigar se certo padrão presente nos dados é ou não efeito do acaso é através do teste de aleatorização (MANLY, 2006 e VIOLA, 2007). Este teste é um procedimento em que se comparam valores de uma estatística observada nos dados com os valores desta estatística após a aleatorização das observações. A principal vantagem da utilização do teste é que ele pode ser utilizado para pequenas amostras, aleatórias ou não, e, como se trata de um teste não-paramétrico, prescinde da distribuição da população da qual a amostra foi extraída. Os testes clássicos (paramétricos ou não-paramétricos) exigem amostras aleatórias, porém, algumas vezes, amostras não-aleatórias são mais convenientes ao pesquisador. A principal desvantagem é que não é possível generalizar as conclusões dos testes de aleatorização para a população de interesse (MANLY, 2006). A utilização deste teste tem se tornado mais freqüente com o avanço computacional. Atualmente existem diversos *softwares* estatísticos, como por exemplo, o R, através do qual é possível aplicar o teste de aleatorização. Este teste é baseado na pressuposição de que, se a hipótese nula é verdadeira, todas as possíveis ordens dos dados são igualmente prováveis. A hipótese nula diz que não existe padrão nos dados, ou se existe este padrão, isto é efeito do acaso quando as observações estão aleatorizadas; enquanto a hipótese alternativa afirma que os dados apresentam certo padrão (MANLY, 2006). Para a realização deste teste, calcula-se o valor e_o de uma estatística "E" de um conjunto de dados. A seguir, aleatorizam-se um grande número de vezes estes dados e calculam-se os valores e_{ai} desta estatística, em que i refere-se a i -ésima aleatorização. A proporção de vezes que a estatística aleatorizada foi maior ou igual que a observada é comparada com o nível de significância adotado. Rejeita-se a hipótese nula quando esta proporção for menor que o nível de significância (VIOLA, 2007). A estatística "E" deve ser escolhida adequadamente pelo pesquisador. Neste estudo foi utilizada a estatística do teste de Wilcoxon-Mann-Whitney para avaliar se as duas amostras são provenientes da mesma população, ou seja, se os cultivos a sombra ou ao sol não têm influência no crescimento das plantas de milho.

2.2.3 Teste de Wilcoxon-Mann-Whitney

Este teste não paramétrico é indicado para testar se duas amostras independentes são provenientes da mesma população ou de populações idênticas. A utilização deste teste ocorre quando as variáveis estudadas são mensuradas em escala pelo menos em nível ordinal e é uma alternativa ao Teste T (SIEGEL & CASTELLAN, 2006). O procedimento para aplicação do teste é o seguinte: Seja m o número de casos na amostra do grupo X e n o número de casos na amostra Y. Neste caso, m é o número de casos na amostra do grupo de plantas cultivadas à sombra e n o número de casos na amostra do grupo de plantas cultivadas ao sol. Devem-se ordenar os dados amostrais combinados, atribuir postos e encontrar a soma dos postos, por exemplo, do primeiro grupo (poderia ser o segundo). A estatística de teste W_x para pequenas amostras é dada pela soma dos postos de um dos grupos. É esperado sob H_0 verdadeira que a média dos postos

de cada um dos grupos seja próxima (SIEGEL & CASTELLAN, 2006). A distribuição amostral de W_x é conhecida se H_0 é verdadeira. Com o auxílio de uma tabela da distribuição nula da estatística de teste verifica-se se H_0 é verdadeira. Para grandes amostras ($m > 10$ ou $n > 10$) a estatística de teste é dada por:

$$z = \frac{W_x \pm 0,5 - \mu_{W_x}}{\sigma_{W_x}}$$

em que $\mu_{W_x} = \frac{m(N+1)}{2}$, $\sigma_{W_x}^2 = \frac{mn(N+1)}{12}$ e sob H_0 verdadeira z segue distribuição normal padrão. É adicionado +0,5 se é desejado encontrar probabilidades na cauda esquerda da distribuição e é adicionado -0,5 se é desejado encontrar probabilidades na cauda direita da distribuição (SIEGEL & CASTELLAN, 2006). Se ocorrem empates nas observações, deve-se utilizar um fator de correção dado por:

$$\sigma_{W_x}^2 = \frac{mn}{N(N-1)} \left(\frac{N^3 - N}{12} - \sum_{j=1}^g \frac{t_j^3 - t_j}{12} \right)$$

em que $N = m + n$, g é o número de agrupamentos de postos empatados diferentes e t_j é o número de postos empatados no j -ésimo grupo (SIEGEL & CASTELLAN, 2006).

3 Resultados e Discussão

Foram cultivadas plantas de milho de período de 25 de janeiro a 11 de fevereiro de 2010, em que suas alturas foram medidas em centímetros no dia 11 de fevereiro. Destas plantas, quatro foram cultivadas a sombra e cinco ao sol. As alturas das plantas foram medidas e o objetivo foi avaliar se o cultivo a sombra difere estatisticamente do cultivo ao sol em relação ao crescimento da planta. Como se trata de amostras pequenas e não é possível afirmar se a população segue distribuição normal optou-se pela utilização do teste de aleatorização considerando a estatística do teste de Wilcoxon-Mann-Whitney (teste não-paramétrico). As hipóteses testadas foram: H_0 : as duas amostras são provenientes de populações idênticas ou da mesma população versus H_1 : as duas amostras são provenientes de populações diferentes. O teste de aleatorização foi realizado considerando a estatística do teste de Wilcoxon-Mann-Whitney. Foram realizadas 10.000 aleatorizações através do R e o p-valor (do teste de aleatorização) foi 0,9666. Como este valor é maior que o nível de significância ($\alpha=0,05$) então não se rejeita H_0 ao nível de 5% de significância, ou seja, as amostras são provenientes de populações idênticas ou da mesma população, o que equivale afirmar que a altura média das plantas de milho cultivadas a sombra ou ao sol não diferem estatisticamente.

4 Conclusão

Como não foi possível detectar normalidade, dado que foram utilizadas pequenas amostras, optou-se pelo uso do teste de aleatorização considerando o teste não paramétrico de Wilcoxon-Maan-Whitney. Para ilustrar o teste utilizado, foi considerado um conjunto de dados e verificado se existe diferença no crescimento das plantas em dois ambientes: sol e sombra. Após a aplicação do teste, verificou-se que o ambiente não influencia no crescimento das plantas de milho. O experimento foi conduzido com o objetivo específico de obter dados reais para ilustrar uma aplicação do teste de aleatorizacao. Portanto, a não rejeição da hipótese nula pode ser questionada por um especialista da área, já que a expectativa é de que plantas cultivadas ao sol tenham uma maior taxa de crescimento.

5 Referências

1. CAMPOS, H. **Estatística Não-Paramétrica**. 1979.
2. MANLY, B. F. J. **Randomization, Bootstrap and Monte Carlo Methods in Biology**. 2 ed. Florida: 2006. Chapman & Hall/CRC.
3. R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, ISBN: 3-900051-07-0, url = <http://www.R-project.org>, 2009.
4. SIEGEL, S. CASTELLAN JR, N.J. **Estatística não-paramétrica para ciências do comportamento**. Edição: Artimed, 2006. 448p.
5. TRIOLLA, M. F. **Introdução à Estatística**. 9 ed. Cidade: Editora, 2005.
6. VIOLA, D. N. **Tese de Doutorado: Detecção e modelagem de padrão espacial em dados binários e de contagem**. USP: Escola Superior de Agricultura "Luiz de Queiroz". Piracicaba, 2007.