Avaliação do desempenho dos testes de normalidade do pacote nortest do R

Elisa Henning[1,2]
[1]Universidade Aberta; [2]Universidade do Estado de Santa Catarina
[1]Campus Do Taguspark, Av. Jacques Delors 211, 2740-122 - Porto Salvo - Portugal
[2]Campus Universitário Prof. Avelino Marcante, Joinville - 89219-710 - Brasil
elisa.henning@udesc.com

Amilcar Oliveira[1,3]
[1]Universidade Aberta; [3]Centro de Estatística e Aplicações da Universidade de Lisboa
[1]Campus Do Taguspark, Av. Jacques Delors 211, 2740-122 - Porto Salvo - Portugal
[3]Campo Grande, Lisboa -1749-016 - Portugal
aoliveira@uab.pt

RESUMO

A suposição de normalidade é uma condição exigida para a realização de muitas inferências válidas a respeito de parâmetros populacionais. Portanto, é fundamental testar a hipótese de normalidade, quando da avaliação dos pressupostos para a validade da inferência que será realizada. É possível verificar a normalidade a partir de alguns gráficos. E, existem testes estatísticos específicos para verificar a aderência de conjunto de dados a uma distribuição normal. Citam-se os testes Shapiro-Wilk, Shapiro-Francia, qui quadrado de Pearson, Kolmogorov-Smirnov, Lillefors, Anderson-Darling e Cramer-von Mises. A existência, no entanto, de um teste adequado, com propriedades ótimas é uma questão em aberto. Assim, o objetivo principal deste trabalho é avaliar o desempenho dos testes de normalidade do pacote nortest do R. O pacote GNU R tem relativo impacto no meio científico por ter código aberto, com contribuições de pesquisadores do mundo inteiro. Um dos pacotes direcionados a testes de normalidade é o pacote nortest, e tem funções para os testes anteriormente citados. O desempenho será avaliado a partir da mensuração das taxas de erro tipo I, em condições da hipótese nula de normalidade, e do poder do teste, simulando amostras sob a hipótese alternativa de não normalidade. Portanto, procura-se um teste com taxas de erro tipo I menores ou iguais a um valor nominal de probabilidade escolhido, e que possua o maior poder possível. Tendo em mente este objetivo, a partir de simulação são avaliados os testes de normalidade deste pacote. Para complementar, incluiu-se o teste Shapiro Wilk da instalação usual do R. Foram geradas amostras com distribuição normal padrão e considerando casos de distribuições fortemente e levemente assimétricas, além de aproximadamente simétricas, para tamanhos de amostra variados. Foram aplicados os testes e, selecionados os valores de probabilidade de significância (valor-p) inferiores a 1%, entre 1% e 5% e, entre 5% e 10%. Os procedimentos foram replicados mil vezes. Nas análises realizadas concluiu-se que quando a distribuição subjacente é a normal, os testes têm resultados bons, com exceção do teste qui quadrado. De modo geral, quando a distribuição não é normal, os testes só produzem resultados aceitáveis para amostras maiores. Para amostras menores os resultados apontam para um baixo poder, mesmo com distribuições fortemente assimétricas. Contudo, mais análises são sugeridas, compreendendo um delineamento mais abrangente das distribuições a serem geradas além da investigação dos algoritmos de origem dos testes.

PALAVRAS CHAVE. Testes de normalidade, Simulação, Avaliação do desempenho.

Área principal (Estatística)

ABSTRACT

The normality assumption is a precondition for the realization of many valid inferences about population parameters. Therefore, it is essential to test the hypothesis of normality, when assessing the conditions for the validity of the inference. The normality can be checked with some charts and, there are specific statistical tests to verify the adherence to a normal distribution. We can cite some tests as Shapiro-Wilk, Shapiro-Francia, chi square test, Kolmogorov-Smirnov, Lillefors, Anderson-Darling and Cramer-von Mises. However, the presence of a suitable test, with optimum properties, is an open question. The aim of this study is to evaluate the performance of tests of normality from the R package nortest. R is a free software environment for statistical computing and has a relative impact in the scientific community, because its open source, with contributions from researchers worldwide. A package used for testing of normality is the nortest package, and has functions for the tests mentioned above. The performance will be evaluated through the measurement of type I error, under the null hypothesis of normality and the power of the test, simulating samples under the alternative hypothesis of non-normality. Therefore, we try to test with a type I error rates less than or equal to a nominal value of probability chosen, and it has the most power possible. The normality tests of this package are evaluated with simulation procedures. To complement, we included in this study, the Shapiro Wilk test, of the usual installation of R. Samples were generated with standard normal distribution and considering the cases of strongly and slightly asymmetric distributions, as well as approximately symmetric, with different sample sizes. Values of the probability of significance (p-value) less than 1% and 5% and between 5% and 10% were tested and selected. The procedures were replicated a thousand times. Our analysis found that when the underlying distribution is normal, the tests have good results, with the exception of the chi square test. In general, when the distribution is not normal, the tests produce acceptable results only for large samples. For smaller samples the results indicate a low power, even with strongly asymmetric distributions. However, further analyzes are suggested, including a more comprehensive delineation of the distributions to be generated, in addition to investigating the origin of the testing algorithms.

KEYWORDS. Normality tests, Simulation, Performance Evaluation.

Main Area (Statistics)