

CONSIDERAÇÕES SOBRE TESTES DE NORMALIDADE UTILIZADOS PELO SOFTWARE GRETl

Gérson Dos Santos Nunes¹

Viviane Leite Dias de Mattos²

Resumo:

Na inferência estatística, muitas vezes deseja-se que os dados analisados apresentem uma distribuição normal, para cumprirem as suposições do modelo teórico utilizado. Para verificação desta propriedade são utilizados os chamados testes de normalidade, que possuem a função de averiguar se a distribuição de probabilidade associada a um conjunto de dados pode ser aproximada a distribuição normal. O presente estudo avaliou a qualidade dos quatro testes de normalidade existentes no software estatístico Gretl: teste de Jarque-Bera, teste de Doornik-Hansen, teste de Shapiro-Wilk e teste de Kolmogorv-Smirnov Lilliefors. Foram geradas amostras com 20, 50, 100 e 500 unidades amostrais oriundas de população normalmente distribuída e não normalmente distribuída. Ao final da análise, constatou-se que todos os testes apresentaram bom desempenho para as amostras geradas com normalidade, o mesmo não acontecendo em caso contrário. Em amostras geradas com distribuição não normal, o teste de Shapiro-Wilk apresentou o melhor desempenho, principalmente em amostras menores. Destaca-se, entretanto que, para grandes amostras, todos os testes também apresentaram bom desempenho.

Palavras-chave: Teste de Normalidade; Jarque-Bera; Doornik-Hansen; Shapiro-Wilk; Kolmogorv-Smirnov Lilliefors;

Modalidade de Participação: Pós-Graduação

CONSIDERAÇÕES SOBRE TESTES DE NORMALIDADE UTILIZADOS PELO SOFTWARE GRETl

¹ Aluno de pós-graduação. gersonnunes@ifsul.edu.br. Autor principal

² Docente. viviane.leite.mattos@gmail.com. Orientador

CONSIDERAÇÕES SOBRE TESTES DE NORMALIDADE UTILIZADOS PELO SOFTWARE GRETL

1 INTRODUÇÃO

Uma das mais importantes distribuições contínuas de probabilidade é a distribuição normal. Além de descrever o padrão de comportamento de uma série de fenômenos, ela é de grande importância na estatística inferencial. Esta distribuição influencia diretamente na qualidade e confiabilidade de resultados de algumas análises estatísticas desenvolvidas em pesquisas científicas, em que são adotados procedimentos em que a normalidade dos dados é uma exigência do processo. A não comprovação dessa suposição pode acarretar em resultados e conclusões incorretas.

Segundo Morettin (2006), no estudo das séries temporais, por exemplo, se uma série de dados for considerada normal ou gaussiana, seu comportamento poderá ser descrito por um modelo linear, como por exemplo um modelo ARMA. Sendo assim, os testes de normalidade são utilizados para verificar se a distribuição de probabilidade associada a um conjunto de dados pode ser aproximada a uma distribuição normal.

O software estatístico Gretl executa quatro testes para verificar a normalidade: o teste de Jarque-Bera, o teste de Doornik-Hansen, o teste Shapiro-Wilk e o teste de Kolmogorov-Smirnov Lilliefors.

Segundo Bueno (2011), o teste Jarque-Bera testa a hipótese nula de que os dados seguem uma distribuição normal verificando se os coeficientes de assimetria e curtose da série estudada são iguais aos respectivos coeficientes de uma distribuição normal. É importante notar que a rejeição da hipótese nula indica evidências de não normalidade, porém a não rejeição indica apenas que não foram encontradas evidências de que o terceiro e o quarto momentos da distribuição empírica coincidem com os da normal. O teste de Jarque-Bera é conhecido por ter boas propriedades para verificar normalidade. É claro e simples de calcular, sendo muito utilizado no contexto de regressão em econometria. Uma limitação é que somente pode ser utilizado na verificação de normalidade (LUCAMBIO, 2008).

Em concordância com Adkins (2009), o teste de Doornik-Hansen utiliza a estatística Qui-quadrado para averiguar a veracidade na hipótese de normalidade, já que se trata de um modelo mais complexo que o Jarque-Bera. Sua hipótese nula (H_0) considera que a amostra provem de uma distribuição normal, enquanto a hipótese alternativa (H_1) considera que a amostra não provem de uma distribuição normal. Esse teste tem bom desempenho em pequenas amostras.

O teste de Shapiro-Wilk é um teste de normalidade publicado em 1965 por Samuel Sanford Shapiro e Martin Wilk e se baseia na regressão dos valores amostrais ordenados com as correspondentes estatísticas de seis ordens normais que, para uma amostra de uma população normalmente distribuída, é linear (Royston 1982).

De acordo com ABDI & MOLIN (2007), o teste Kolmogorov-Smirnov Lilliefors é uma adaptação do teste de Kolmogorov-Smirnov usada para verificação de normalidade de um conjunto de dados. A diferença básica entre eles está no fato de que o primeiro testa a normalidade a partir de média e desvio-padrão fornecidos, enquanto a adaptação Lilliefors usa a média e o desvio-padrão, calculados a partir do próprio conjunto de dados. A hipótese nula destes testes considera que os dados seguem a distribuição normal, enquanto a hipótese alternativa indica que os dados não seguem a distribuição normal.

Sendo assim, o objetivo deste estudo é analisar o desempenho dos testes desenvolvidos pelo software estatístico Gretl para verificar normalidade, relacionando-os com algumas propriedades dos dados.

2 METODOLOGIA

O presente estudo foi desenvolvido de acordo com o seguinte protocolo:

- por simulação foram geradas quatro amostras com, respectivamente, 20, 50, 100 e 500 unidades amostrais, extraídas de uma população com distribuição normal com média e desvio-padrão iguais a 0 e 1, respectivamente;
- por simulação foram geradas quatro amostras com, respectivamente, 20, 50, 100 e 500 unidades amostrais, extraídas de uma população com distribuição não normal com média e desvio-padrão iguais a 0 e 1, respectivamente;
- nas oito amostras a normalidade foi verificada pelos testes Jarque-Bera, Doornik-Hansen, Shapiro-Wilk e Kolmogorov-Smirnov Lilliefors por meio do software estatístico Gretl;
- os resultados foram analisados, sendo identificado aquele que apresentou melhor desempenho no contexto considerado.

Além do software Gretl, estes procedimentos foram realizados com o auxílio do software Excel.

3 RESULTADOS e DISCUSSÃO

A tabela 01 mostra os resultados dos testes de normalidade para as amostras estudadas.

Tabela 01: Resultados dos testes de normalidade.

p –valor	Doornik-Hansen		Shapiro-Wilk		KS-Lilliefors		Jarque-Bera	
	Normal	Não Normal	Normal	Não Normal	Normal	Não Normal	Normal	Não Normal
n=20	0.3659	0.0778	0.7266	0.0373	0.6300	0.2500	0.6403	0.3560
n=50	0.3196	0.0004	0.8433	0.0002	0.8600	<0.0001	0.7844	0.0732
n=100	0.9673	<0.0001	0.8865	<0.0001	0.4000	<0.0001	0.9436	0.0054
n=500	0.1175	<0,0001	0.1377	<0,0001	0.3200	<0.0001	0.1238	<0,0001

Fonte: do autor, 2018.

Através dos resultados apresentados na tabela 01, considerando um nível de significância de 5%, foi possível constatar que todos os testes não encontraram evidências de que as amostras geradas a partir de uma distribuição normal não seguissem uma distribuição normal, sendo o valor p maior nas amostras de menor tamanho. Chamou a atenção o fato de que, no teste KS-Lilliefors, o maior valor p foi encontrado na amostra de tamanho 50 e não na amostra de tamanho 100 como aconteceu com os demais.

Os equívocos de julgamento foram encontrados na verificação da amostras geradas a partir de distribuições sem normalidade. Para pequenas amostras, 20 elementos, o teste de Shapiro-Wilk obteve o melhor resultado, pois foi o único que conseguiu identificar evidências de não normalidade, já o teste de Doornik-Hansen apresentou indícios fracos de normalidade. Se fosse adotado o nível de significância de 1%, evidências de não normalidade não teriam sido encontradas. Já para amostras de tamanho 50, o Jarque-Bera foi o único que apresentou indícios fracos de normalidade, os demais apresentaram evidências de não normalidade das séries. Para amostras de tamanho 100, os testes de Doornik-Hansen, Shapiro-Wilk e KS-Lilliefors apresentaram fortes indícios de não normalidade. No entanto, o teste de Jarque-Bera apresentou fracos indícios de que os dados amostrais não provém de uma distribuição são normal. No entanto, pode-se notar que, com o aumento do tamanho da amostra, o desempenho de todos os testes melhora.

4 CONSIDERAÇÕES FINAIS

Para os dados gerados a partir de distribuições normais, os quatro critérios mostraram-se equivalentes, com bons resultados e não identificando evidências de não normalidade em nenhuma das situações analisadas. Já para dados gerados a partir de distribuições não normais, pode-se observar que para amostras de tamanho maior, todos os testes apresentam bons resultados na observação de existência de não normalidade. Quando o tamanho da amostra diminuiu, entretanto, os testes começaram a apresentar problema. Destaca-se o teste de Shapiro-Wilk, que se mostrou superior aos demais, concluindo-se que este é, aparentemente, o melhor teste de aderência à normalidade para pequenas amostras.

Mais estudos poderiam ser feitos, considerando-se outras distribuições, diferentes testes de normalidade, distintos parâmetros destas distribuições e outros tamanhos de amostra.

REFERÊNCIAS

ABDI, HervéAbdi & MOLIN,Paul. **Lilliefors/Van Soest's test of normality**. Encyclopedia of Measurement and Statistics. 2007. Disponível em: <http://www.utdallas.edu/~herve/Abdi-Lillie2007-pretty.pdf>. Acesso em 23 de dezembro de 2016.

ADKINS, L. C. "Using gretl for Principles of Econometrics". 3ª Edição. Versão 1.313, 2010.

BUENO, R. D. L.S. "Econometria de séries temporais". 2 ed. São Paulo: Editora Cengage Learning, 2011. 360p.

MORETTIN, P. A. **Econometria Financeira**. 2. ed. São Paulo, SP. Blucher, 2006.

ROYSTON, J. P. Algorithm AS 181: the W test for normality. **Journal of the Royal Statistical Society: série C - applied statistics**, London, v. 31, n. 2, p. 176-180. 1982.