Assessing the efficiency of data normality verification tests

Sachiko Araki Lira, Anselmo Chaves Neto

Abstract - Applying parametric statistical tests requires checking some suppositions about their data nature, including data normality, since such tests are based on Normal distribution (Gaussian). For this purpose, several tests are available and among them the following are the most used: Kolmogorov-Smirnov with Lilliefors correction, Shapiro-Wilk, Pearson's Chi-squared, Anderson-Darling, Shapiro-Francia, Cramér-von Mises, Jarque-Bera and D'Agostino-Pearson. The present work assesses test efficiency through the Monte Carlo simulation, including 10,000 samples with sizes equal to 10, 20, 30, 40, 50, 70, 90, 100, 150, 200 and 500, and four different distributions: Gamma with parameters 5 and $1/5$; Double Exponential with parameters 0 and 1; Chi-squared with parameter 5; Exponential with parameter 3 and Contaminated Normal distribution, where 50% of the values were generated through Standard Normal distribution and 50% through two parameters Exponential distribution (3 and 1). We inferred that normality test performances are affected by the significance level used in the hypothesis test, the shape of data distribution and the sample size. Concerning non-normal distribution, the present study allowed us to infer that when applying normality tests the more asymmetric is data distribution, the smaller the sample size could be. On the other hand, the closer to normal (symmetric) the distribution is, the larger the sample size must be. Thus, we verify that some tests are more efficient than others depending on the shape of distribution.

Keywords – Normality test, Monte Carlo simulation, data distribution, test power.

I. Introduction

The parametric tests applied to many different areas, such as engineering, economic and social sciences, health, safety and others, it is necessary to check some suppositions, including data normality.

There are several tests available and among them the most popular are: Kolmogorov-Smirnov with Lilliefors correction, Shapiro-Wilk, Pearson's Chi-squared, Anderson-Darling, Shapiro-Francia, Cramér-von Mises, Jarque-Bera and D'Agostino-Pearson. We can also find graphic methods as Frequency Histogram, Normal Q-Q plot (*quantile-quantile plot*) and Normal P-P plot (*probability-probability plot)* that are frequently used to check if a data set behaves according to normal distribution.

When choosing the test, the question is: which one is the most adequate? Some authors discuss the efficiency of some tests aforementioned, see [4], [5], [6], [8], and [1].

The Chi-squared test is based on the comparison between sample data distribution and theoretical distribution of the sample. The Kolmogorov-Smirnov (K-S) test is based on proximity analysis or adjustment between the empirical distribution or sample distribution $S(x)$, and the population distribution function (theoretical), $F_o(x)$, that is admitted in the null hypothesis H_0. Lilliefors corrected the K-S test standardized by the use of μ and σ estimations. On the other hand, the Anderson-Darling test, which is a measure that shows how the points are approaching the straight line estimated in a probability graphic, is an alternative to the two aforementioned tests. Its great advantage is to be more sensitive since it gives more weight to tail observations. The Cramér-von Mises test is an alternative to the Kolmogorov-Smirnov test and checks the hypothesis that a data set results from a specific continuous distribution estimating the minimum distance between theoretical probability distribution and empirical distribution. The D'Agostino-Pearson test statistics combines asymmetry and kurtosis measures thus producing the normality *omnibus test*[1]. The Shapiro-Wilk test, known as powerful *omnibus test* [5] was proposed to check if a certain random sample came from a normally distributed population. Finally, the Jarque-Bera is an asymptotic normality test based on the Ordinary Least Squares (OLS) residuals by firstly calculating the asymmetry and kurtosis measures.

Sachiko A. Lira is with the Departament of Mecanic, Federal University of Paraná – DEMEC/PG-MEC, Centro Politécnico - CEP 81531-980 – Curitiba – Paraná, Brazil (e-mail: sachiko.lira@ufpr.br).
Anselmo C. Neto is with the Department of Statistic, Federal University of Paraná – DEST/PPGMNE, Centro Politécnico - CEP 81531-990 – Curitiba – Paraná, Brazil (e-mail: anselmo@ufpr.br).

[1] Allows to check if in a data set the explained-variance is superior to the non-explained variance.

II. NORMALITY TEST DESCRIPTION

A. Pearson's Chi-squared Test

The Chi-squared test (χ^2) allows to assess adherence between the frequency distribution of a sample composed of n observations expressed in any scale and a theoretical distribution. Performing this test requires a random sample size $n \geq 30$.

The n observations composing the sample are grouped in k classes (or categories) mutually exclusive (with $k \geq 2$). Chi-squared adherence test is calculated by the following expression:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

where:

O_i is the i class observed frequency;

E_i is the i[th] class expected frequency.

If the null hypothesis (H_0) is true, the aforementioned statistics has Chi-squared asymptotic distribution with degree of freedom equal to $df = k - p - 1$, where k represents number of classes and p is the number of distribution parameters estimated from the sample.

B. Kolmogorov-Smirnov Test with Lilliefors Correction

For a random sample with size n, the empirical distribution or sample distribution function $S(x)$, expresses data relative frequencies with values smaller or equal to *x*, which is a particular value of the random variable X. Thus, considering $(X_1, X_2, \ldots, X_n)$, a random sample of a continuous random variable X and the respective order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, we have that $S(x)$ empirical distribution function is obtained by:

$$S(x) = \begin{cases} 0, & x < X_{(1)} \\ \frac{k}{n}, & X_{(k)} \leq x \leq X_{(k+1)} \quad (k = 1, 2, ..., n-1) \\ 1, & x \geq X_{(n)} \end{cases} \tag{2}$$

Thus, the empirical distribution function $S(x)$ is a step function that increases by $1/n$ in the jump points (sample ordinal statistics) and the Kolmogorov-Smirnov test statistics, denoted by $D_{obs.}$ (that is a random variable), correspond to the maximum difference (*supreme*), in absolute value between $S(x)$ and $F_o(x)$, when all possible X values are considered. According to symbolic notation we have:

$$D_{obs.} = \max_x \left| S(x) - F_o(x) \right| \tag{3}$$

The test performed with Lilliefors correction is processed in the same way as the Kolmogorov-Smirnov test, the only difference is that original data are standardized by using estimates of μ and σ:

$$Z_i = \frac{X_i - \overline{X}}{S}, \quad i = 1, 2, ..., n \tag{4}$$

C. Anderson-Darling Test

This test is based on the empirical distribution function. It supports the idea that once the distribution functions under H_0 is given; data can be transformed into uniform distribution and then tested for uniformity.

The Anderson-Darling test statistic is given by:

$$A = -n - \frac{1}{n} \sum_{i=1}^{n} [2i - 1]\left[\ln(p_{(i)}) + \ln(1 - p_{(n-i+1)})\right] \tag{5}$$

where:

$$p_{(i)} = \Phi\left(\left[y_{(i)} - \overline{y}\right] / S \right) \tag{6}$$

is the percentiles of the Standard Normal distribution.

D. Cramér-von Mises Test

This test is also based on cumulative distribution function. The Cramér-von Mises test statistic is given by:

$$W = \frac{1}{12n} + \sum_{i=1}^{n} \left(p_{(i)} - \frac{2i - 1}{2n} \right) \tag{7}$$

where:

$$p_{(i)} = \Phi\left(\left[x_{(i)} - \overline{x}\right] / S \right) \tag{8}$$

is the percentiles of the Standard Normal distribution.

E. D'Agostino-Pearson Test

This procedure is also known as D test and the statistic has the following expression:

$$D = \frac{T}{n^2 S} \tag{9}$$

where:

$$T = \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) y_{(i)}; \tag{10}$$

and S is the sample standard deviation.

F. Jarque-Bera Test

It is based on the difference between the asymmetry and kurtosis coefficients of observed data and theoretical normal distribution. The test statistic is given by:

$$JB = n \left(\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24} \right) \tag{11}$$

where:

S^2, $\sqrt{b_1}$ e b_2 are respectively the second, third and fourth central moments. The JB statistics has asymptotic distribution $\chi^2_{(2)}$ under null hypothesis.

G. Shapiro-Wilk Test

The Shapiro-Wilk test statistic is expressed by:

$$W = \frac{\left\{ \sum_{i=1}^{n} a_i x_{(i)} \right\}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (12)$$

and a_i is computed as the solution of:

$$(a_1, a_2, \cdots, a_n) = \frac{m' V^{-1}}{(m' V^{-1} V^{-1} m)^{1/2}} \quad (13)$$

where:
$x_{(i)}$ is the i-th order statistic;
$m = (m_1, m_2, \cdots, m_n)$ is the expected values vector of the order statistics;
V is the variance-covariance matrix of the random sample order statistics.

H. Shapiro-Francia Test

Shapiro and Francia introduced a change in the W statistics, so W' depends only of the expected value (m) of the order statistic. The Shapiro-Francia test statistic is given by:

$$W' = \frac{\left\{ \sum_{i=1}^{n} b_i x_{(i)} \right\}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (14)$$

where: $b' = (b_1, b_2, \cdots, b_n) = \frac{m'}{(m' m)^{1/2}}$. (15)

III. MATERIAL AND METHOD

The aforementioned test efficiency assessment was carried out through the Monte Carlo simulation and used the *nortest* and *fBasics* libraries available in the R system. Were generated ten thousand (10,000) samples, with the sizes equal to 10, 20, 30, 40, 50, 70, 90, 100, 150, 200 and 500, according to different distributions, which are: Standard Normal distribution $N(0,1)$, used to assess type I error percentage, that is, the null hypothesis (H_0) is true, but is rejected; Gamma distribution with parameters 5 and $1/5$; Laplace or Double Exponential distribution with parameter 0 and 1; Chi-squared distribution with parameter 5; Exponential distribution with parameter 3 and Normal Contaminated distribution, where 50% of the values were obtained using Standard Normal distribution and the other 50% using the two-parameter Exponential distribution (3 and 1). The respective distribution figures are as follows: Fig. 1 shows the frequency histogram of random variable X with Gamma distribution $(5;1/5)$ and sample size $n = 50$. It is possible to observe that this symmetry is slightly positive. Fig. 2 shows the frequency histogram of random variable X with Double Exponential distribution $(0\ ;\ 1)$, with sample size $n = 50$. This symmetric distribution can be considered as the distribution of the difference between two independently and identically distributed (i.i.d) random variables with exponential distribution. When the graphic is doubled around the y axis it is proportional to Exponential distribution density.

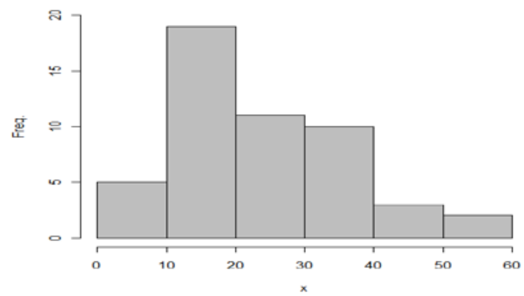

Fig. 1 - Frequency histogram of random variable X with Gamma distribution

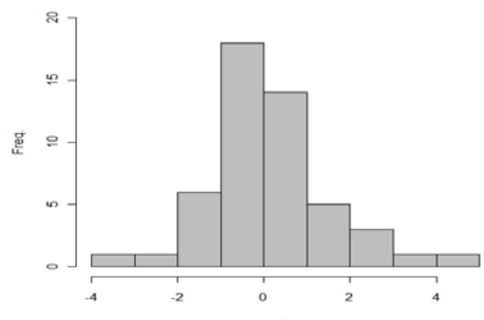

Fig. 2 - Frequency histogram of random variable X with Double Exponential distribution

Fig. 3 shows frequency histogram of the random variable X with distribution $\chi^2(5)$ and $n = 50$, indicating positive asymmetry.

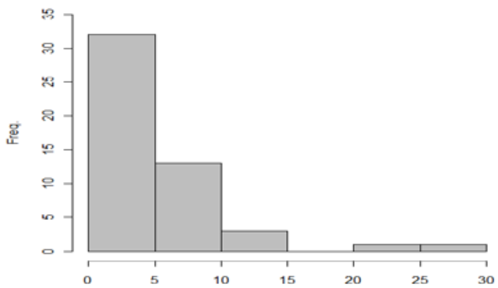

Fig. 3 – Frequency histogram of the random variable X with Chi-squared distribution

The frequency histogram of the random variable X with distribution $\text{Exp}(3)$ and sample size $n = 50$ presented in Fig. 4 also indicates positive asymmetry.

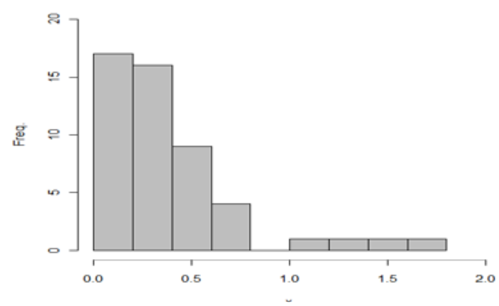

Fig. 4 – Frequency histogram of the random variable X with Exponential distribution

Finally, the histogram of a Contaminated Normal distribution with sample size equal to 50, where 50% of values were generated through Standard Normal distribution and the other 50% through a two-parameter (3 and 1) Exponential distribution, with sample size n=50. Fig. 5 shows that distribution has negative asymmetry.

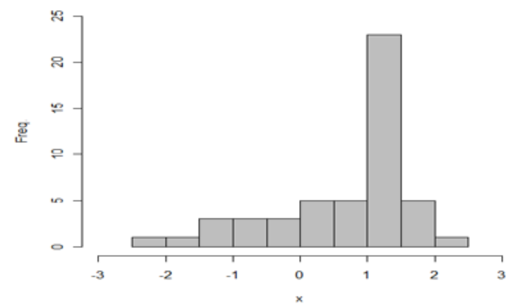

Fig. 5 – Frequency histogram of the random variable X with Contaminated Normal distribution

Significance levels of 1% and 2% were used to test the hypothesis H_0: the data follow normal distribution, against the H_1: the data do not follow normal distribution. This choice is due the fact that when performing a hypothesis test we want to commit minor mistake as possible.

Concerning normal distribution data and test power of other distributions (Gamma, Chi-squared, Exponential and Contaminated Normal distribution), type I error rates were obtained from ratios between number of H_0 rejections (p-value < 0.01 or p-value < 0.05) and total samples generated to each test.

IV. RESULTS AND DISCUSSION

Tables 1 and 2 show type I error rates of different tests applied to normal distribution data for pre-defined sample sizes.

TABLE 1 – TYPE I ERROR RATES ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR NORMAL DISTRIBUTION DATA, USING SIGNIFICANCE LEVEL OF 1%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.0113	0.0103	0.0114	0.0102	0.0113	0.0105	0.0095	0.0097	0.0100	0.0092	0.0092
Shapiro-Wilk	0.0114	0.0096	0.0092	0.0086	0.0072	0.0082	0.0098	0.0095	0.0088	0.0085	0.0097
Chi-squared	0.0141	0.0121	0.0111	0.0116	0.0098	0.0097	0.0100	0.0109	0.0109	0.0113	0.0104
Anderson-Darling	0.0099	0.0094	0.0091	0.0090	0.0087	0.0090	0.0095	0.0095	0.0091	0.0091	0.0108
Cramer-von Mises	0.0093	0.0097	0.0097	0.0100	0.0087	0.0083	0.0096	0.0089	0.0097	0.0096	0.0099
Shapiro-Francia	0.0110	0.0096	0.0099	0.0091	0.0091	0.0088	0.0101	0.0091	0.0092	0.0097	0.0096
D'Agostino-Pearson	-	0.0169	0.0180	0.0162	0.0163	0.0140	0.0138	0.0143	0.0142	0.0145	0.0120
Jarque-Bera	-	0.0098	0.0131	0.0143	0.0153	0.0145	0.0153	0.0155	0.0165	0.0159	0.0138

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

When distribution is normal and used significance level of 1%, almost all the test are efficient since type I error rates are close to 1%, except for Chi-squared tests to n = 10 and Jarque-Bera and D'Agostino-Pearson tests that show higher error taxes if compared to the others. The Jarque-Bera is an asymptotical test, that's to say a large sample test. Thus, we will only consider the results of samples equal or over size 20. Concerning D'Agostino-Pearson, we also took into consideration the equal or over 20-sized samples, since this is a limitation of the R system.

TABLE 2 – TYPE I ERROR RATES ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR NORMAL DISTRIBUTION

DATA, USING SIGNIFICANCE LEVEL OF 5%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.0556	0.0479	0.0486	0.0519	0.0481	0.0488	0.0490	0.0494	0.0485	0.0479	0.0483
Shapiro-Wilk	0.0514	0.0489	0.0484	0.0461	0.0491	0.0471	0.0495	0.0494	0.0473	0.0488	0.0491
Chi-squared	0.0649	0.0492	0.0531	0.0571	0.0525	0.0546	0.0544	0.0502	0.0529	0.0516	0.0534
Anderson-Darling	0.0522	0.0491	0.0485	0.0486	0.0493	0.0479	0.0490	0.0485	0.0491	0.0486	0.0490
Cramer-von Mises	0.0507	0.0498	0.0499	0.0487	0.0500	0.0504	0.0517	0.0504	0.0503	0.0486	0.0477
Shapiro-Francia	0.0555	0.0518	0.0500	0.0487	0.0488	0.0482	0.0500	0.0523	0.0483	0.0484	0.0494
D'Agostino-Pearson	-	0.0545	0.0559	0.0567	0.0545	0.0521	0.0517	0.0531	0.0536	0.0499	0.0470
Jarque-Bera	-	0.0215	0.0305	0.0329	0.0344	0.0370	0.0389	0.0407	0.0393	0.0411	0.0411

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

The tests were also efficient when we used the significance level of 5%, except for the Chi-squared test that showed type I error rate of 0.0649 for $n = 10$. On the other hand, the Jarque-Bera test showed the lowest type I error rates for all the sample sizes.

According to [8], concerning normal data the four criteria (Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises and Shapiro-Wilk) are equivalent, except for Kolmogorov-Smirnov. The significance level used was of 1%.

The authors in [6] say that the Jarque-Bera test was superior for Normal distribution and Standard Normal, because it showed lower type I error rate. Shapiro-Wilk and Kolmogorov-Smirnov with Lilliefors correction tests can also be used for practical purposes.

According to [4], among the five tests studied (Anderson-Darling, Lilliefors, Jarque-Bera, D'Agostino-Pearson and Shapiro-Francia) the best is the Anderson-Darling test. The Jarque-Bera test is the most popular and widely used in economy area, but results suggest that Anderson-Darling is superior. The author carried out the tests taking into consideration the significance levels of 1%, 5% and 10%.

The test power assessment for Gamma distribution data (table 3) with significance level of 1% showed that the Shapiro-Wilk test power is superior to 90.0% for sample size $n = 150$.

Using the significance level of 5% (table 4), the Shapiro-Wilk test showed a power equal to 89.92%, for sample size $n = 100$.

TABLE 3 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR GAMMA DISTRIBUTION WITH SIGNIFICANCE LEVEL OF 1%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.0274	0.0486	0.0708	0.1037	0.1344	0.2110	0.2959	0.3329	0.5405	0.7040	0.9952
Shapiro-Wilk	0.0415	0.0995	0.1798	0.2639	0.3533	0.5298	0.6920	0.7563	0.9332	0.9854	1.0000
Chi-squared	0.0227	0.0345	0.0494	0.0581	0.0719	0.1151	0.1523	0.1811	0.3301	0.4932	0.9792
Anderson-Darling	0.0353	0.0764	0.1308	0.1923	0.2569	0.3979	0.5402	0.6013	0.8279	0.9378	1.0000
Cramer-von Mises	0.0313	0.0644	0.1102	0.1606	0.2069	0.3317	0.4525	0.5071	0.7458	0.8819	1.0000
Shapiro-Francia	0.0407	0.0981	0.1686	0.2437	0.3222	0.4845	0.6386	0.7088	0.9088	0.9774	1.0000
D'Agostino-Pearson	-	0.1256	0.1831	0.2467	0.3060	0.4263	0.5500	0.6042	0.8296	0.9439	1.0000
Jarque-Bera	-	0.0868	0.1507	0.2215	0.2858	0.4163	0.5476	0.6051	0.8365	0.9473	1.0000

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

TABLE 4 - POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR GAMMA DISTRIBUTION WITH SIGNIFICANCE LEVEL OF 5%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.0950	0.1476	0.2064	0.2648	0.3206	0.4356	0.5408	0.5959	0.7813	0.8910	0.9995
Shapiro-Wilk	0.1184	0.2349	0.3570	0.4776	0.5804	0.7494	0.8599	0.8982	0.9820	0.9976	1.0000
Chi-squared	0.1058	0.1069	0.1383	0.1890	0.2039	0.2746	0.3568	0.3905	0.5730	0.7287	0.9965
Anderson-Darling	0.1101	0.2001	0.2932	0.3858	0.4725	0.6245	0.7541	0.8007	0.9408	0.9866	1.0000
Cramer-von Mises	0.1045	0.1794	0.2587	0.3358	0.4159	0.5590	0.6816	0.7301	0.8920	0.9638	1.0000
Shapiro-Francia	0.1251	0.2313	0.3403	0.4494	0.5498	0.7153	0.8353	0.8742	0.9731	0.9957	1.0000
D'Agostino-Pearson	-	0.2154	0.3031	0.3864	0.4715	0.6177	0.7455	0.8036	0.9509	0.9906	1.0000
Jarque-Bera	-	0.1387	0.2346	0.3241	0.4147	0.5817	0.7206	0.7840	0.9453	0.9900	1.0000

NOTE: Monte Carlo Simulation NMC = 10,000.

According to [6], for $t(30)$ distribution, all the tests showed low power. We have to keep in mind that the Student t distribution developed by William Sealy Gosset is a symmetrical and bell-shaped distribution similar to the Standard Normal curve but with wider tails. The only parameter that defines and characterizes it is the degree of freedom (ν). The larger the parameter, the closer to Normal the distribution is. For degree of freedom equal to 30, the distribution is closer to Standard Normal distribution.

The following table (table 5) shows the power of different tests and sample sizes for Double Exponential distribution data, which is a symmetrical but not mesokurtic distribution.

It was observed that Shapiro-Wilk, Anderson-Darling, Cramer-von Mises, Shapiro-Francia and Jarque-Bera tests show test power greater than 92.00% for the sample size when using the significance level of 1 %.that Shapiro-Wilk, Anderson-Darling, Cramér-von Mises, Shapiro-Francia and Jarque-Bera tests show test power higher than 92.00% for sample size $n = 200$ when using significance level of 1%. Thus, the closer to normal the distribution is the data analyzed, the large the sample size must be, so that the tests for normality checking will be efficient.

TABLE 5 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR DOUBLE EXPONENTIAL DISTRIBUTION DATA WITH SIGNIFICANCE LEVEL OF 1%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.0536	0.0886	0.1288	0.1671	0.2161	0.3182	0.4119	0.4577	0.6661	0.8170	0.9985
Shapiro-Wilk	0.0707	0.1310	0.2032	0.2661	0.3392	0.4717	0.5790	0.6317	0.8278	0.9271	1.0000
Chi squared	0.0352	0.0590	0.0750	0.0918	0.1122	0.1775	0.2315	0.2689	0.4217	0.5707	0.9762
Anderson-Darling	0.0649	0.1292	0.1973	0.2682	0.3400	0.4775	0.6035	0.6568	0.8523	0.9451	1.0000
Cramer-von Mises	0.0615	0.1247	0.1879	0.2570	0.3298	0.4653	0.5946	0.6485	0.8489	0.9423	0.9998
Shapiro-Francia	0.0756	0.1592	0.2427	0.3186	0.3941	0.5292	0.6299	0.6806	0.8595	0.9408	0.9999
D'Agostino-Pearson	-	0.1894	0.2544	0.3026	0.3460	0.4409	0.5173	0.5571	0.7206	0.8364	0.9982
Jarque-Bera	-	0.1493	0.2522	0.3330	0.4085	0.5333	0.6314	0.6760	0.8385	0.9231	0.9996

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

Using significance level of 5% (table 6), the Anderson-Darling, Cramer-von Mises and Shapiro-Francia tests show power of over 98.00%, for sample sizes equal to 200, thus proving that for the tests are efficient, If distributions are closer to Normal distribution, sample sizes needs to be considerably large.

TABLE 6 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR DOUBLE EXPONENTIAL DISTRIBUTION DATA WITH SIGNIFICANCE LEVEL OF 5%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.1435	0.2168	0.2904	0.3606	0.4315	0.5526	0.6632	0.7079	0.8652	0.9452	0.9999
Shapiro-Wilk	0.1567	0.2607	0.3638	0.4433	0.5194	0.6490	0.7529	0.7919	0.9217	0.9742	1.0000
Chi-squared	0.1379	0.1520	0.1962	0.2560	0.2679	0.3582	0.4419	0.4757	0.6475	0.7817	0.9955
Anderson-Darling	0.1625	0.2778	0.3736	0.4615	0.5372	0.6828	0.7835	0.8265	0.9414	0.9843	1.0000
Cramer-von Mises	0.1575	0.2724	0.3651	0.4578	0.5318	0.6786	0.7797	0.8242	0.9412	0.9844	0.9999
Shapiro-Francia	0.1893	0.3264	0.4330	0.5199	0.5976	0.7200	0.8099	0.8428	0.9444	0.9823	1.0000
D'Agostino-Pearson	-	0.3051	0.3873	0.4517	0.5157	0.6155	0.6967	0.7302	0.8655	0.9394	0.9996
Jarque-Bera	-	0.2213	0.3380	0.4321	0.5184	0.6451	0.7388	0.7738	0.9070	0.9613	1.0000

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

Concerning χ^2 distribution of data (table 7 and table 8), the Shapiro-Wilk test showed test power of 89.56% and 97.31% for significance levels of 1% and 5%, respectively for sample size $n = 70$.

TABLE 7 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR CHI-SQUARED DISTRIBUTION WITH SIGNIFICANCE LEVEL OF 1%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.0488	0.1062	0.1753	0.2573	0.3349	0.4999	0.6432	0.7034	0.9040	0.9751	1.0000
Shapiro-Wilk	0.0765	0.2226	0.4038	0.5764	0.7159	0.8956	0.9699	0.9858	1.0000	1.0000	1.0000
Chi-squared	0.0390	0.0872	0.1369	0.1979	0.2519	0.4296	0.5613	0.6474	0.8873	0.9782	1.0000
Anderson-Darling	0.0671	0.1859	0.3263	0.4702	0.5927	0.7974	0.9106	0.9433	1.0000	1.0000	1.0000
Cramer-von Mises	0.0603	0.1599	0.2802	0.4035	0.5128	0.7147	0.8446	0.8874	0.9870	0.9989	0.9949
Shapiro-Francia	0.0732	0.2106	0.3671	0.5253	0.6605	0.8564	0.9511	0.9740	0.9994	1.0000	1.0000
D'Agostino-Pearson	-	0.2193	0.3255	0.4322	0.5423	0.7115	0.8423	0.8883	0.9904	0.9997	1.0000
Jarque-Bera	-	0.1598	0.2798	0.3992	0.5212	0.7105	0.8503	0.8954	0.9910	0.9996	1.0000

NOTE: Monte Carlo Simulation $\text{NMC} = 10{,}000$.

TABLE 8 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR CHI-SQUARED DISTRIBUTION WITH SIGNIFICANCE LEVEL OF 5%

TESTES	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.1409	0.2595	0.3802	0.4889	0.5879	0.7460	0.8541	0.8893	0.9795	0.9975	1.0000
Shapiro-Wilk	0.1929	0.4274	0.6406	0.7924	0.8862	0.9731	0.9954	0.9978	0.9996	1.0000	1.0000
Chi-squared	0.1605	0.2126	0.3003	0.4214	0.4937	0.6620	0.8038	0.8350	0.9666	0.9958	1.0000
Anderson-Darling	0.1800	0.3734	0.5495	0.6907	0.8007	0.9214	0.9754	0.9863	1.0000	1.0000	1.0000
Cramer-von Mises	0.1648	0.3376	0.4935	0.6253	0.7298	0.8734	0.9467	0.9653	0.9978	0.9999	0.9968
Shapiro-Francia	0.2002	0.4066	0.6001	0.7534	0.8538	0.9582	0.9917	0.9956	1.0000	1.0000	1.0000
D'Agostino-Pearson	-	0.3421	0.4810	0.6083	0.7157	0.8746	0.9564	0.9776	0.9994	1.0000	1.0000
Jarque-Bera	-	0.2389	0.3935	0.5457	0.6737	0.8540	0.9490	0.9747	0.9994	1.0000	1.0000

NOTE: Monte Carlo Simulation $\text{NMC} = 10{,}000$.

Concerning Exponential distribution data, presented in table 9 and table 10, the Shapiro-Wilk test showed power of 97.91% for sample size $n = 40$ with significance level of 1%. When using significance level of 5%, the same test showed test power of 96.44% for sample size $n = 30$.

According to [5], who assessed the Shapiro-Wilk and D'Agostino-Pearson tests, the first was more powerful than the second for different distributions and sample sizes, thus inferring that normality test performance is strongly affected by the shape of distribution data and sample size. The significance level used was of 5%.

TABLE 9 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR EXPONENTIAL DISTRIBUTION WITH SIGNIFICANCE LEVEL OF 1%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.1289	0.3305	0.5396	0.7259	0.8566	0.9706	0.9959	0.9987	1.0000	1.0000	1.0000
Shapiro-Wilk	0.2323	0.6248	0.8730	0.9791	0.9949	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Chi-square	0.1274	0.4212	0.6566	0.7884	0.8691	0.9791	0.9898	0.9975	1.0000	1.0000	1.0000
Anderson-Darling	0.2044	0.5678	0.8136	0.9369	0.9834	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000
Cramer-von Mises	0.1883	0.5117	0.7510	0.8980	0.9621	0.9964	0.9997	1.0000	0.9990	0.9803	0.0013
Shapiro-Francia	0.2108	0.5762	0.8257	0.9492	0.9878	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
D'Agostino-Pearson	-	0.4430	0.6344	0.7790	0.8774	0.9750	0.9968	0.9989	1.0000	1.0000	1.0000
Jarque-Bera	-	0.3541	0.5870	0.7565	0.8735	0.9785	0.9970	0.9992	1.0000	1.0000	1.0000

NOTE: Monte Carlo Simulation $\text{NMC} = 10{,}000$.

TABLE 10 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR EXPONENTIAL DISTRIBUTION WITH SIGNIFICANCE LEVEL OF 5%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.2929	0.5719	0.7795	0.9013	0.9615	0.9963	0.9994	0.9999	1.0000	1.0000	1.0000
Shapiro-Wilk	0.4320	0.8314	0.9644	0.9954	0.9993	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Chi-squared	0.3879	0.6556	0.8493	0.9543	0.9840	0.9994	0.9999	1.0000	1.0000	1.0000	1.0000
Anderson-Darling	0.4024	0.7746	0.9314	0.9859	0.9965	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Cramer-von Mises	0.3738	0.7282	0.8991	0.9676	0.9918	0.9998	1.0000	1.0000	0.9991	0.9844	0.0018
Shapiro-Francia	0.4212	0.7937	0.9474	0.9914	0.9986	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
D'Agostino-Pearson	-	0.5954	0.7842	0.9026	0.9675	0.9989	0.9999	1.0000	1.0000	1.0000	1.0000
Jarque-Bera	-	0.4787	0.7278	0.8803	0.9579	0.9979	0.9999	1.0000	1.0000	1.0000	1.0000

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

Finally, if the data analyzed come from negative asymmetrical distributions, such as the Contaminated Normal used in this work (table 11), with significance levels of 1%, the Kolmogorov-Smirnov with Lilliefors correction test showed power of 96.26% for sample size $n = 40$. Using significance level of 5% (table 12), the test power is of 95.50% for sample size $n = 30$.

When analyzing the Cramér-von Mises test, it was observed that test power for samples sizes higher then 200 are reduced for distributions with strong and moderate asymmetry.

TABLE 11 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR CONTAMINATED NORMAL DISTRIBUTION DATA WITH SIGNIFICANCE LEVEL OF 1%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.1306	0.5406	0.8466	0.9626	0.9928	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
Shapiro-Wilk	0.1001	0.3015	0.5440	0.7321	0.8530	0.9698	0.9946	0.9985	1.0000	1.0000	1.0000
Chi-squared	0.1430	0.3311	0.6122	0.7654	0.8993	0.9815	0.9989	0.9997	1.0000	1.0000	1.0000
Anderson-Darling	0.1128	0.4224	0.7235	0.8946	0.9607	0.9958	0.9999	1.0000	1.0000	1.0000	1.0000
Cramer-von Mises	0.1172	0.4571	0.7741	0.9242	0.9765	0.9980	1.0000	1.0000	1.0000	1.0000	0.0033
Shapiro-Francia	0.0737	0.2444	0.4766	0.6753	0.8178	0.9595	0.9934	0.9976	1.0000	1.0000	1.0000
D'Agostino-Pearson	-	0.0975	0.1406	0.1907	0.2573	0.4210	0.6130	0.7021	0.9282	0.9863	1.0000
Jarque-Bera	-	0.0534	0.1065	0.1630	0.2381	0.4324	0.6389	0.7279	0.9378	0.9883	1.0000

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

TABLE 12 – POWER ACCORDING TO DIFFERENT TESTS AND SAMPLE SIZES FOR CONTAMINATED NORMAL DISTRIBUTION DATA WITH SIGNIFICANCE LEVEL OF 5%

TESTS	SAMPLE SIZES										
	10	20	30	40	50	70	90	100	150	200	500
Lilliefors (K-S)	0.3543	0.7969	0.9550	0.9939	0.9985	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Shapiro-Wilk	0.3068	0.5999	0.8051	0.9161	0.9666	0.9958	0.9995	0.9996	1.0000	1.0000	1.0000
Chi-squared	0.3512	0.5477	0.8154	0.9221	0.9724	0.9968	1.0000	1.0000	1.0000	1.0000	1.0000
Anderson-Darling	0.3376	0.7146	0.9055	0.9747	0.9949	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000
Cramer-von Mises	0.3447	0.7570	0.9312	0.9855	0.9971	0.9998	1.0000	1.0000	1.0000	1.0000	0.0046
Shapiro-Francia	0.2819	0.5736	0.7870	0.9081	0.9629	0.9950	0.9996	0.9996	1.0000	1.0000	1.0000
D'Agostino-Pearson	-	0.2244	0.3142	0.4333	0.5585	0.7771	0.8903	0.9217	0.9873	0.9981	1.0000
Jarque-Bera	-	0.1144	0.2165	0.3427	0.4872	0.7468	0.8795	0.9161	0.9865	0.9983	1.0000

NOTE: Monte Carlo Simulation $NMC = 10{,}000$.

Aiming at confirming the aforementioned discussed results, we carried out normality tests in different sample sizes obtained through the Monte Carlo simulation for some selected distributions.

The following tables show p-values of different normality tests applied to samples generated through the Monte Carlo simulation. It was observed that Normal distribution samples in all the tests discussed herein indicate data normality whatever their sizes (n). It is important to emphasize that asymmetry coefficients are very close to zero, thus indicating symmetric distribution (table 13).

TABLE 13 – P-VALUES FOR NORMALITY TESTS APPLIED TO NORMAL DISTRIBUTION SAMPLES ACCORDING TO DIFFERENT SAMPLE SIZES AND ASYMMETRY COEFFICIENTS

TESTS	SAMPLE SIZES AND ASYMMETRY COEFFICIENTS				
	$n = 10$ $As = 0.5790$	$n = 20$ $As = 0.1983$	$n = 30$ $As = -0.0089$	$n = 50$ $As = 0.1578$	$n = 100$ $As = -0.2068$
Lilliefors (K-S)	0.9580	0.7863	0.5183	0.5444	0.4754
Shapiro-Wilk	0.8492	0.7560	0.5887	0.9087	0.6439
Chi-squared	0.3618	0.6626	0.3920	0.7327	0.5378
Anderson-Darling	0.8580	0.7241	0.5200	0.8032	0.5690
Cramer-von Mises	0.8960	0.6777	0.4551	0.6958	0.6007
Shapiro-Francia	0.8008	0.8515	0.7178	0.9042	0.5921
D'Agostino-Pearson	-	0.5817	0.5539	0.8204	0.5634
Jarque-Bera	-	0.6576	0.6437	0.7566	0.6793

NOTE: Samples from distribution $N(380, 19^2)$.

For Exponential distribution with small samples (n=10) and with asymmetry coefficient lower than 1, all the tests indicate data normality. However, if the sample size is equal to 20 and asymmetry coefficient higher than 2, all tests indicates non-normality (table 14).

TABLE 14 – P-VALUES FOR NORMALITY TESTS APPLIED TO EXPONENTIAL DISTRIBUTION SAMPLES ACCORDING TO DIFFERENT SAMPLE SIZES AND ASYMMETRY COEFFICIENTS

TESTS	SAMPLE SIZES AND ASYMMETRY COEFFICIENTS				
	$n = 10$ $As = 0.7352$	$n = 20$ $As = 2.8894$	$n = 30$ $As = 0.9259$	$n = 50$ $As = 0.9742$	$n = 100$ $As = 1.2995$
Lilliefors (K-S)	0.8867	1.981E-04	1.530E-03	1.144E-04	4.191E-09
Shapiro-Wilk	0.7575	1.641E-05	4.499E-03	1.951E-04	4.245E-08
Chi-squared	0.5724	8.687E-03	9.029E-03	1.124E-04	1.085E-09
Anderson-Darling	0.6760	2.636E-05	2.115E-03	9.450E-05	4.622E-10
Cramer-von Mises	0.6862	1.022E-04	2.036E-03	3.151E-05	1.395E-07
Shapiro-Francia	0.6056	3.261E-05	8.738E-03	6.733E-04	4.399E-07
D'Agostino-Pearson	-	8.544E-08	9.267E-02	2.371E-02	4.784E-06
Jarque-Bera	-	7.772E-16	1.449E-01	2.426E-02	8.805E-08

NOTE: samples $Exp(1/380)$.

The Gamma distribution with small sample size (n=10) and asymmetry coefficient higher than 1, all the tests indicate data normality with significance level of 1%.

The D'Agostino-Pearson and Jarque-Bera tests indicate data normality even for sample with sizes equal to 20 and 30 and asymmetry coefficients close to 1 (table 15).

TABLE 15 – P-VALUES FOR NORMALITY TESTS APPLIED TO GAMMA DISTRIBUTION SAMPLES ACCORDING TO DIFFERENT SAMPLE SIZES AND ASYMMETRY COEFFICIENTS

TESTS	SAMPLE SIZES AND ASYMMETRY COEFFICIENTS				
	$n = 10$ $As = 1.6353$	$n = 20$ $As = 1.2944$	$n = 30$ $As = 1.0422$	$n = 50$ $As = 1.5508$	$n = 100$ $As = 1.6033$
Lilliefors (K-S)	0.1227	2.699E-03	4.761E-04	6.071E-05	5.305E-06
Shapiro-Wilk	0.0138	1.528E-03	6.131E-04	5.048E-06	1.608E-08
Chi-squared	0.0154	3.449E-03	1.180E-03	1.847E-03	5.534E-06
Anderson-Darling	0.0209	5.100E-04	1.228E-04	8.783E-07	1.705E-09
Cramer-von Mises	0.0255	4.960E-04	1.608E-04	6.327E-06	2.374E-07
Shapiro-Francia	0.0156	2.699E-03	1.525E-03	2.058E-05	1.352E-07
D'Agostino-Pearson	-	3.750E-02	6.435E-02	4.334E-05	3.471E-09
Jarque-Bera	-	8.995E-02	8.383E-02	2.380E-06	<2.200E-16

NOTE: Samples from distribution $Gamma(1, 9)$.

Taking into consideration samples with chi-squared distribution with small size (n=10) with asymmetry coefficient a little higher than 1, all the tests also indicate data normality for significance level of 1% and 5%, except for chi-squared. Even when we increase the sample size to 30 with low asymmetry coefficient, all the tests indicate data normality for significance level of 5%, except for Lilliefors (K-S) which the level is 4% (table 16).

TABLE 16 – P-VALUES FOR NORMALITY TESTS APPLIED TO CHI-SQUARED DISTRIBUTION SAMPLES ACCORDING TO DIFFERENT SAMPLE SIZES AND ASYMMETRY COEFFICIENTS

TESTS	SAMPLE SIZES AND ASYMMETRY COEFFICIENTS				
	$n=10$ $As=1.3152$	$n=20$ $As=1.9089$	$n=30$ $As=0.6948$	$n=50$ $As=1.0105$	$n=100$ $As=1.4568$
Lilliefors (K-S)	0.2416	1.180E-02	0.0435	3.195E-02	1.258E-03
Shapiro-Wilk	0.0792	1.287E-03	0.0663	2.456E-03	5.298E-06
Chi-squared	0.0268	6.905E-02	0.1637	2.986E-01	5.738E-03
Anderson-Darling	0.1000	2.687E-03	0.1245	5.109E-03	2.868E-05
Cramer-von Mises	0.1150	4.505E-03	0.1621	9.200E-03	2.146E-04
Shapiro-Francia	0.0717	1.387E-03	0.1030	4.720E-03	1.812E-05
D'Agostino-Pearson	-	1.445E-04	0.2604	1.070E-02	2.113E-06
Jarque-Bera	-	9.296E-05	0.3119	1.391E-02	5.972E-09

NOTE: Samples from distribution $\chi^2(4)$.

V. CONCLUSION

The present study concluded that normality test performance is affected by the significance level used in the hypothesis test, the shape of data being analyzed and sample size.

Concerning normal distribution data, the different tests analyzed herein showed similar performances. When using significance level of 1%, the D'Agostino-Pearson, Jarque-Bera and Chi-squared tests with size samples equal or smaller than 20 are the less efficient. But, when using $\alpha = 5\%$, the Jarque-Bera test performance is the best.

Shapiro-Wilk, Anderson-Darling, Cramer-von Mises, Shapiro-Francia and Jarque-Bera tests can be used to assess distributions similar to normal distribution, that's to say almost symmetrical, for sample sizes equal to or higher than 200 and significance level of 1%. However, when using significance level of 5%, we recommend the Anderson-Darling, Cramér-von Mises and Shapiro-Francia for sample sizes close to 200.

Concerning slightly positive asymmetric distributions, such as $Gamma(5, 1/5)$, Shapiro-Wilk and Shapiro-Francia tests are efficient for sample sizes close to 150 when using significance level of 1%. The same tests are efficient for sample sizes higher than 100, when using significance level of 5%.

Concerning moderate positive asymmetric distributions such as $\chi^2(5)$, Shapiro-Wilk, Shapiro-Francia and Anderson-Darling tests are efficient for samples sizes close to 90 with significance level of 1%. When using significance level of 5%, the same tests are efficient for sample sizes close to 70.

When data show strong positive asymmetry such as $Exp(3)$, Shapiro-Wilk, Anderson-Darling and Shapiro-Francia tests are efficient for sample sizes close to 40 with significance level of 1%. When using significance level of 5%, the same tests are efficient for sample sizes close to 30.

Concerning negative asymmetry data, such as the Contaminated Normal distribution described in the present study, Kolmogorov-Smirnov with Lilliefors correction and Cramér-von Mises tests are efficient for sample sizes close to 40 with significance level of 1%. When using significance level of 5%, the same tests are efficient for sample sizes close to 30. It is important to consider that Cramér-von Mises test should be used for sample sizes (n) smaller than or equal to 200, since for higher values of n, the test power can decrease.

The present study permits us to infer that to apply a normality test when distribution is not normal the more asymmetric it is, as $Exp(1/380)$ shown in table 14, the smaller the sample size can be. On the other hand the closer to Normal the distribution is, almost symmetric, the larger the sample size should be. This fact shows the importance of carrying out a previous exploratory analyzes before starting any statistical inference process. We recommend the use of *Normal Q-Q plot* or *Normal P-P plot* graphic, when sample sizes are small ($n < 30$) and distributions are symmetric, for it helps data normality checking analysis.

REFERENCES:

[1] D'AGOSTINO, Ralph B., et al. A suggestion for using powerful and informative tests of normality. The American Statistician, v. 44, n. 4, pp. 316-321, nov. 1990.

[2] D'AGOSTINO, Ralph B. An omnibus test of normality for moderate and large size samples. Biometrika, v. 58, n. 2, pp. 341-348, ago. 1971.

[3] GUJARATI, Damodar N. Econometria básica. São Paulo: Pearson Makron Books. 3. ed. 2000.

[4] ISLAM, Tanweer ul. Normality Testing – A new direction. http://mpra.ub.uni-muenchen.de/16452/ . Accessed on 10/02/2010.

[5] KESKIN, Siddik. Comparison of Several Univariate Normality Tests Regarding Type I Error Rate and Power of the Test in Simulation based Small Samples. Journal of Applied Science Research, v. 5, n. 2, pp. 296-300, 2006.

[6] ÖZTUNA, Derya et al. Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. Turkish Journal of Medical Sciences, v. 3, n. 36, 171-176,2006.

[7] SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). Biometrika, v. 52, n. 3/4, pp. 591-611, dez. 1965.

[8] TORMAN, V. B. L. ; BIRCK, Alan Rodrigues ; RIBOLDI, João. Comparação dos testes de aderência à normalidade Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises e Shapiro-Wilk por simulação. In: 11 Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO), 2005. Anais da 50ª RBRAS e do 11 SEAGRO, 2005.