RECENT J. (2018), 56:173-177
https://doi.org/10.31926/RECENT.2018.56.173

Analysis of the Probability of Acceptance of the Normality Hypothesis by the Goodness-of-Fit Tests

Ramona CLINCIU
Transilvania University of Brasov, Romania, r.clinciu@unitbv.ro

Abstract
The paper presents and emphasizes the behaviour of different goodness-of-fit tests, towards the acceptance of the normality hypothesis. The goodness of fit tests considered are general goodness-of-fit tests - Kolmogorov-Smirnov, Cramer-von-Mises, Anderson-Darling, and normality goodness-of-fit tests - Lilliefors, Shapiro-Wilk, D'Agostino, Massey, Filliben, Z, Cox. These goodness-of-fit tests are conducted on normally distributed data in order to test the normality of the data. The need for testing the normality of the data appears especially in metrology, for the analysis of the metrological reliability. General metrology uses especially the normal distribution, despite the fact that positive and asymmetrical distributions (e.g. Weibull distribution) are frequently met in the analysis of the metrological reliability. In these cases, it is necessary to perform goodness-of-fit tests in order to ascertain that the normal distribution fits the data. The quality of the results depends on the goodness-of-fit test which was chosen to determine if the normal distribution fits the data best.

Keywords
metrological reliability, goodness-of-fit tests, normal distribution

1. Introduction

Goodness-of-fit tests are essential for the quantitative evaluation of a system reliability and maintainability. The most important issue for data analysis is to find the best, or the most appropriate distribution which describes the experimental collected data and goodness-of-fit tests are used to test whether the selected distribution fits the data.

Positive and asymmetrical distributions (e.g. Weibull distribution) are met frequently in the analysis of the metrological reliability. Despite this fact, metrology uses especially the normal distribution and, in these cases, it is necessary to perform goodness-of-fit tests in order to ascertain that the normal distribution fits the data [1]. The quality of the results obtained depends on the goodness-of-fit test which was chosen to test if the normal distribution fits the data best.

The present paper aims to present and emphasize the results obtained by conducting different goodness-of-fit tests on normally distributed data, in order to test the probability of acceptance of the normality hypothesis by the goodness-of-fit tests which were considered: general goodness-of-fit tests - Kolmogorov-Smirnov, Cramer-von-Mises, Anderson-Darling, and normality goodness-of-fit tests - Lilliefors, Shapiro-Wilk, D'Agostino, Massey, Filliben, Z, Cox. These tests are conducted on normally distributed data grouped in samples of different sizes, in order to test the normality of the data. Their behaviour towards the considered samples, generated according to the normal distribution function is to be studied. The goodness-of-fit tests considered in this research were applied in accordance with the speciality literature [1-9]. Although in [7] Cox goodness-of-fit test is considered to be adequate for n > 100, it was conducted on small sizes of samples too, in order to analyse its behaviour towards the small sizes of samples. The results are presented in this paper. Goodness of fit tests are applied in researches [10, 11, 12] in order to test the normality of data.

The general goodness-of-fit tests were used in their modified form, according to the situation in which the parameters of the normal distribution are not known and they need to be estimated by [2, 3]:

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2} \tag{2}$$

where x_i, i = 1, ... n, are the values in the sample, n is the size of the sample, m is the mean, and σ is the standard deviation.

2. Case Study: Programmed Generation of Normally Distributed Data

In this case study, the considered goodness-of-fit tests were performed on normally distributed data, data being obtained by programmed generation according to the formulas below [4, 5, 6]:

$$x_i = m \pm z_i \cdot \sigma \ , \quad i = 1, \ldots, n \tag{3}$$

$$z_i = F^{-1}(\alpha_i) \tag{4}$$

$$\alpha_i = \frac{i - 3/8}{n + 1/4} \ , \quad i = 1, \ldots, n \tag{5}$$

$$F(z) = \frac{1}{2 \cdot \pi} \int_0^z e^{-\frac{t^2}{2}} dt \tag{6}$$

The tests were conducted on samples with different sizes $n \in$ {5, 10, 20, 30, 40, 50, 60, 80, 100}, each sample being generated according to the equation (3).

Table 1 and Table 2 present the results of the tests in the case the considered samples are normally distributed, being obtained by programmed generation of normally distributed data, the parameters of the normal distribution being m = 0 and σ = 1.

Table 1. Tests statistics and their critical values

Sample size (n)	5	10	20	30	40
General goodness-of-fit tests / decision criteria					
Kolmogorov - Smirnov ($d \le d_{n,\alpha}$)	0.2872 < 0.895	0.1989 < 0.895	0.1407 < 0.895	0.1238 < 0.895	0.1075 < 0.895
Cramer-von-Mises ($W^2 < W^2_{n,\alpha}$)	0.0186 < 0.126	0.009 < 0.126	0.0045 < 0.126	0.0031 < 0.126	0.0024 < 0.126
Anderson-Darling ($A^2 < A^2_{n,\alpha}$)	0.105 < 0.787	0.0899 < 0.787	0.0521 < 0.787	0.037 < 0.787	0.029 < 0.787
Normality goodness-of-fit tests / decision criteria					
Lilliefors ($L \le L_{n,\alpha}$)	0.1102 < 0.337	0.0581 < 0.258	0.0302 < 0.190	0.0167 < 0.161	0.0167 < 0.1401
Shapiro-Wilk (3<n<50) ($W \ge W^2_{n,\alpha}$)	0.9965> 0.986	0.9966 > 0.978	0.9972 > 0.983	0.9961 > 0.985	0.9949 > 0.987
D'Agostino (n>50) $Y \in (Y_{n,\alpha/2}, Y_{n,1-\alpha/2})$	-	-	-	-	-
Massey (10 <n<30) ($d \le d_{n,\alpha}$)	-	0.0689 < 0.130	0.0364 < 0.117	0.026 < 0.102	-
Filliben ($r_c \ge r_{n,\alpha}$)	1 > 0.995	1 > 0.99	0.9999 > 0.992	1 > 0.994	1 > 0.995
Z ($Z \le Z_{n,\alpha}$)	0.0088 < 1.4047	0 < 0.8744	0 < 0.6135	0 < 0.5045	0 < 0.4392
Cox (b_1 / b_2) $b_1 \in$[-0.05, 0.05] $b_2 \in$[2.95, 3.05]	0.0074 / 1.8547	0 / 2.1855	0 / 2.4314	0 / 2.5391	0 / 2.6089

By analysing the results indicated in Tables 1 and Table 2, the conclusion is that all the considered goodness-of-fit tests accept the normal hypothesis of the programmed distribution, except for the Cox goodness-of-fit test for parameter b_2.

Table 2. Tests statistics and their critical values

Sample size (n)	50	60	80	100
General goodness-of-fit tests / decision criteria				
Kolmogorov - Smirnov ($d \leq d_{n,\alpha}$)	0.0968 < 0.895	0.0952 < 0.895	0.0862 < 0.895	0.0808 < 0.895
Cramer-von-Mises ($W^2 < W^2_{n,\alpha}$)	0.002 < 0.126	0.0017 < 0.126	0.0013 < 0.126	0.0011 < 0.126
Anderson-Darling ($A^2 < A^2_{n,\alpha}$)	0.0243 < 0.787	0.021 < 0.787	0.0166 < 0.787	0.0138 < 0.787
Normality goodness-of-fit tests / decision criteria				
Lilliefors ($L \leq L_{n,\alpha}$)	0.0135 < 0.1253	0.0121 < 0.1144	0.0095 < 0.0991	0.008 < 0.0886
Shapiro-Wilk (3<*n*<50) ($W \geq W^2_{n,\alpha}$)	0.9935 > 0.988	-	-	-
D'Agostino (*n*>50) $Y \in (Y_{n,\alpha/2}, Y_{n,1-\alpha/2})$	0.4322 ∈ (-2.2, 0.923)	0.4297 ∈ (-2.179, 0.986)	0.4036 ∈ (-2.118, 1.076)	0.3874 ∈ (-2.075, 1.137)
Massey (10 <*n*<30) ($d \leq d_{n,\alpha}$)	-	-	-	-
Filliben ($r_c \geq r_{n,\alpha}$)	1 > 0.996	1 > 0.996	1 > 0.997	1 > 0.998
Z ($Z \leq Z_{n,\alpha}$)	0 < 0.3943	0 < 0.3609	0 < 0.3137	0 < 0.2812
Cox (b_1 / b_2) $b_1 \in [-0.05, 0.05]$ $b_2 \in [2.95, 3.05]$	0 / 2.6496	0 / 2.6815	0 / 2.7349	0 / 2.7688

3. Case Study: Random Generation of Normally Distributed Data

Two methods were used for random generation of normally distributed data. In the first method used for random generation of normally distributed data, *n* uniform random variables Q_i, $i = 1, ..., n$, are generated and then, variable *t* is calculated according to [8]:

$$t = \sqrt{\ln \frac{1}{Q^2}} \quad (7)$$

$$Q = \begin{cases} Q & 0 < Q \leq 0.5 \\ 1 - Q & 0.5 < Q < 1 \end{cases} \quad (8)$$

Random variables having a normal distribution can be generated by [8]:

$$x = t - \frac{c_0 + c_1 \cdot t + c_2 \cdot t^2}{1 + d_1 \cdot t + d_2 \cdot t^2 + d_3 \cdot t^3} + \varepsilon(Q), \quad (9)$$

where the error $\varepsilon(Q)$ is given by (10) and the constants are given by (11) [8]:

$$|\varepsilon(Q)| < 4.5 \cdot 10^{-4} \quad (10)$$

$$\begin{array}{ll} c_0 = 2.515517 & d_1 = 1.432788 \\ c_1 = 0.802853 & d_2 = 0.189269 \\ c_2 = 0.010328 & d_3 = 0.001308 \end{array} \quad (11)$$

The second method used for random generation of normally distributed data is the one derived from the central limit theorem. According to this method, random variables having a normal distribution can be generated by [3]:

$$x = \sum_{i=1}^{12} U_i - 6, \qquad (12)$$

where U_i, $i = 1, ..., n$, are uniform random variables.

The tests have been conducted on samples with different sizes $n \in \{5, 10, 20, 30, 40, 50, 60, 80, 100\}$ in each case being considered 2000 samples of each size n. The samples were generated according to the equation (9) or (12), respectively, and the considered goodness-of-fit tests were performed on the samples.

Tables 3 and 4 present the results obtained in case the samples were generated according to both equations (9) and (12) - the probability of acceptance, in percentage, of the normality hypothesis by the goodness-of-fit tests which were considered.

The adopted significance level is $\alpha_{ad} = 0.05$.

Table 3. Probability of acceptance of the normality hypothesis

Sample size (n)	5	10	20	30	40
Samples generated according to equation (9) / Samples generated according to equation (12)					
General goodness-of-fit tests					
Kolmogorov - Smirnov	95.5/94	95/95	96/91	95.5/95	96/96
Cramer-von-Mises	95/97	96.5/96	96/93	98/95	94/96
Anderson-Darling	99/98	95.3/92	93.8/94	95/95	95.4/96
Normality goodness-of-fit tests					
Lilliefors	97/98	97/96	97/96	97/97	97/97
Shapiro-Wilk ($3<n<50$)	5/0	5.1/5	4.5/5	5.9/7	5.4/6
D'Agostino ($n>50$)	-	-	-	-	-
Massey ($10<n<30$)	-	28/25	54.6/51	60.6/55	-
Filliben	2.7/0	4.1/1.4	6.4/6	5.4/5	6/6
Z	92/96	91/95	90/89	91/85	90/87
Cox ($n>100$)	0/0.1	0.1/0.1	0.2/0.2	0.5/0.5	0.6/0.6

Table 4. Probability of acceptance of the normality hypothesis

Sample size (n)	50	60	80	100
Samples generated according to equation (9)/ Samples generated according to equation (12)				
General goodness-of-fit tests				
Kolmogorov - Smirnov	95.5/93	95/96	95.5/97	96/96
Cramer-von-Mises	95/92	95/93	96/95	95/93
Anderson-Darling	98/92	95/95	95/94	95/93
Normality goodness-of-fit tests				
Lilliefors	97/97	97/97	97/98	97/98
Shapiro-Wilk ($3<n<50$)	4.8/4	-	-	-
D'Agostino ($n>50$)	90/93	90/92	90/90	92/91
Massey ($10<n<30$)	-	-	-	-
Filliben	3.4/3	7/7	6/7	3/4
Z	90/97	90/93	93/92	92/92
Cox ($n>100$)	0.85 /0.6	0.9/0.7	1.35/1.4	1.75/1.5

4. Conclusions

The analysis of the results obtained in this research indicates the fact that the probability of acceptance of the normality hypothesis in case of applying the normality goodness-of-fit tests is smaller than the probability of acceptance of the normality hypothesis in case of the general goodness-of-fit tests.

Another conclusion drawn is that, among the normality goodness-of-fit tests, the smallest values for the probability of acceptance of the normality hypothesis are obtained in case the Cox or Filliben goodness-of-fit tests are conducted. The greatest values for the probability of acceptance of the normality hypothesis are obtained in case of application of Z, D'Agostino and Lilliefors goodness-of-fit tests, these values being comparable to those obtained in case general goodness-of-fit tests are conducted.

The results obtained in the research described in the present paper indicate the fact that the probability of acceptance of the normality hypothesis by the goodness-of-fit tests increases with the size of the sample in most cases.

References

1. Vodă V.Gh. (1979): *Metode de testare a normalității datelor de măsurare* (*Methods for testing the normality of the measuring data*). Metrologia aplicată, nr. XXVI (1), p. 11-14 (in Romanian)
2. Craiu V., Enache R., Basca O. (1988): *Teste de concordanţă cu programe în FORTRAN* (*Goodness of fit tests with programmes in FORTRAN*). Editura Ştiinţifică şi Enciclopedică, Bucureşti, Romania (in Romanian)
3. Kececioglu D. (1991): *Reliability Engineering Handbook.* Prentice Hall, 1st ed., ISBN 978-0137722945
4. Iliescu D.V., Vodă V.Gh. (1977): *Statistică şi toleranţe* (*Statistics and tolerances*). Editura Tehnică, Bucureşti, Romania (in Romanian)
5. Martinescu I., Popescu I. (1995): *Fiabilitate* (*Reliability*). Gryphon, ISBN 973-604-007-0, Brasov, Romania (in Romanian)
6. Wiener U., Isaic-Maniu A., Vodă V.Gh. (1983): *Aplicaţii ale reţelelor probabiliste în tehnică* (*Applications of the probabilistic networks in technique*). Editura Tehnică, Bucureşti, Romania (in Romanian)
7. Motoiu R. (1994): *Ingineria calității* (*Quality Engineering*). Editura Chiminform Data, Bucureşti, Romania (in Romanian)
8. Hewlett-Packard HP-25 (1975): *Application Programs*. Hewlett-Packard Company
9. Bentley J. (1993): *An Introduction to Reliability and Quality Engineering*. John-Wiley& Sons Inc., ISBN 978-0201331325, New York
10. Clinciu R. (2013): *Determining the Calibration Life of Measuring Devices by Means of Parameter Deviation Function*. Applied Mechanics and Materials, ISSN 1660-9336, No. 371, p. 695-699, DOI: 10.4028/www.scientific.net/AMM.371.695
11. Clinciu M.R. (2011*): Statistical Analysis of Data Samples Collected in an Experimental Installation*. Bulletin of the Transilvania University of Brasov, Series I, ISSN 2065-2119, Vol. 4(53), no. 1, p. 25-30
12. Clinciu R., Clinciu M.R. (2018): *Study on the Metrological Reliability of Water Meters*. IOP Conference Series Materials Science and Engineering, vol. 399(1), p. 012009, DOI: 10.1088/1757-899X/399/1/012009, https://www.researchgate.net/publication/328050309_Study_on_the_metrological_reliability_of_the_water_meters

Received: 12 December 2018; Accepted: 17 December 2018