

Revisiting the critical values of the Lilliefors test: towards the correct agrometeorological use of the Kolmogorov-Smirnov framework

Gabriel Constantino Blain (*)

Instituto Agronômico (IAC), Avenida Barão de Itapura, 1481, 13020-902 Campinas (SP), Brasil.

(*) Corresponding author: gabriel@iac.sp.gov.br

Received: Feb. 17, 2014; Accepted: Mar. 18, 2014

Abstract

Several studies have applied the Kolmogorov-Smirnov test (KS) to verify if a particular parametric distribution can be used to assess the probability of occurrence of a given agrometeorological variable. However, when this test is applied to the same data sample from which the distribution parameters have been estimated, it leads to a high probability of failure to reject a false null hypothesis. Although the Lilliefors test had been proposed to remedy this drawback, several studies still use the KS test even when the requirement of independence between the data and the estimated parameters is not met. Aiming at stimulating the use of the Lilliefors test, we revisited the critical values of the Lilliefors test for both gamma (gam) and normal distributions, provided easy-to-use procedures capable of calculating the Lilliefors test and evaluated the performance of these two tests in correctly accepting a hypothesized distribution. The Lilliefors test was calculated by using critical values previously presented in the scientific literature (KSL_{crit}) and those obtained from the procedures proposed in this study ($\text{NKSL}_{\text{crit}}$). Through Monte Carlo simulations we demonstrated that the frequency of occurrence of Type I (II) errors associated with the KSL_{crit} may be unacceptably low (high). By using the $\text{NKSL}_{\text{crit}}$ we were able to meet the significance level in all Monte Carlo experiments. The $\text{NKSL}_{\text{crit}}$ also led to the lowest rate of Type II errors. Finally, we also provided polynomial equations that eliminate the need to perform statistical simulations to calculate the Lilliefors test for both gam and normal distributions.

Key words: goodness of fit, gamma distribution, normal distribution.

Revisão dos valores críticos do teste Lilliefors: em direção ao correto uso agrometeorológico do algoritmo de Kolmogorov-Smirnov

Resumo

Diversos estudos têm aplicado o teste de Kolmogorov-Smirnov (KS) para verificar se determinada distribuição paramétrica pode ser utilizada para estimar a probabilidade de ocorrência de variáveis agrometeorológicas. Contudo, a probabilidade de não rejeição de uma falsa hipótese de nulidade (H_0 ; erro tipo II) torna-se elevada quando o KS é aplicado à mesma amostra de dados utilizada para estimar os parâmetros da distribuição. Embora uma adaptação denominada Lilliefors tenha sido proposta para permitir o uso do algoritmo de Kolmogorov-Smirnov na condição anteriormente mencionada, o KS em sua forma original é ainda frequentemente utilizado. Objetivando estimular o correto uso do KS e de sua referida adaptação, este trabalho revisou os valores críticos do Lilliefors para as distribuições gama e normal, desenvolveu procedimentos computacionais capazes de calcular esses testes de aderência e comparou a habilidade do KS e do Lilliefors em aceitar corretamente uma H_0 . O teste de Lilliefors foi calculado utilizando-se tanto valores críticos previamente apresentados na literatura científica (KSL_{crit}) quanto novos valores propostos neste estudo ($\text{NKSL}_{\text{crit}}$). Por meio de simulações de Monte Carlo demonstrou-se que a frequência de ocorrência de erros tipo I (II) associada ao KSL_{crit} pode ser consideravelmente baixa (elevada). A frequência de erros tipo I associada ao $\text{NKSL}_{\text{crit}}$ permaneceu próxima ao nível de significância adotado em todos os experimentos. O $\text{NKSL}_{\text{crit}}$ também demonstra menor taxa de erros tipo II. Considerando-se as distribuições gama e normal, este trabalho também desenvolveu equações polinomiais que eliminam a necessidade de realizar simulações estatísticas para calcular o teste Lilliefors.

Palavras-chave: teste de aderência, distribuição gama, distribuição normal.

1. INTRODUCTION

Parametric distributions have been widely used to represent meteorological data. Although the use of these mathematical functions may be regarded as a theoretical idealization of the actual data, it is not free from empirical considerations (Wilks, 2011). In fact, before adopting a particular distribution to assess the probability of occurrence of a given data value one needs to check if the parametric function does provide a reasonable fit. The selection of an appropriate candidate distribution requires the use of actual data (Wilks, 2011) and it is frequently based on statistical methods called goodness-of-fit tests. The goodness-of-fit tests are usually computed to obtain evidences in favor of the null hypothesis (H_0) that states that the data under analysis were drawn from a particular distribution (Wilks, 2011).

The one-sample Kolmogorov-Smirnov test (KS) is a widely used goodness-of-fit test (Wilks, 2011). It compares the empirical (edf) and the cumulative distribution function (cdf). And, for continuous data, the KS test tends to be more powerful than other tests that compare the data histogram with the probability density function (e.g. the chi-square test, Wilks, 2011). However, the original KS is applicable if, and only if, the parameters of the theoretical distribution have not been estimated from the same bunch of data used to apply this goodness-of-fit test. If this requirement is not fulfilled, the probability of accepting a false H_0 becomes unacceptable high (Crutcher, 1975; Lilliefors, 1967; Steinskog et al., 2007; Vlček and Huth, 2009; Wilks, 2011). Yet, the original KS test has still been frequently used when the requirement of independence between the data and the estimated parameters is not met (Steinskog et al., 2007; Vlček and Huth, 2009). Naturally, this statement suggests that further efforts are required to avoid this misuse of the original KS.

In situations where the parameters of the distribution have been estimated from all available data or from the same bunch of data used to calculate the goodness of fit, a modification of the original KS test must be adopted. This modification is frequently called Kolmogorov-Smirnov/Lilliefors test or simply Lilliefors test (Lilliefors, 1967; 1969). In order to draw the attention of the scientific community to Lilliefors' findings, Crutcher (1975) tabulated critical values for the Lilliefors test (KSL_{crit}) that can be used for several univariate distributions, such as the 2-parameter gamma (gam) and normal distributions (it is worth emphasizing that the chi-square and the exponential distributions may be regarded as particular cases of the gam distribution). The KSL_{crit} values provided by Crutcher (1975) have been widely used to evaluate the fit of meteorological data to the gam

or to the normal distributions (Chen et al., 2013; Vlček and Huth, 2009; Wilks 2011). However, Vlček and Huth (2009) indicate that it may be difficult to use the KSL_{crit} values because they are tabulated only for integer values of the shape parameter of the gam distribution. Moreover, it is worth recalling that these critical values were obtained from a computer-based technique called Monte Carlo simulation. Thus, it becomes natural to suppose that the current capacity of the personal computers may provide better estimates for the critical values of the Lilliefors test ($NKSL_{crit}$).

Thus, the aims of this study are: to (i) review the critical values of the Lilliefors test for both gam and normal distributions, to (ii) provide easy-to-use procedures capable of calculating the Lilliefors test for these two distributions and to (iii) evaluate the performance of the KS and Lilliefors tests in correctly accepting or rejecting a candidate distribution to assess the probability of occurrence of a given variable. The Lilliefors test was calculated by using the critical values presented by Crutcher (1975) and those obtained from the easy-to-use procedures proposed in this study. We hope this study stimulates the use of the Lilliefors test in meteorological studies.

2. MATERIAL AND METHODS

Theoretical background

The null hypothesis (H_0) of a hypothesis test defines a particular logical frame (Wilks, 2011) that allows us to generate the sample null distribution for the test statistic. As in other goodness-of-fit tests, the H_0 of the KS test states that the data under analysis were drawn from the hypothesized distribution. Accordingly, if the calculated test statistic falls in the rejection region of the null distribution, the proposed H_0 is rejected. Otherwise, the calculated test statistic is regarded as consistent with the null hypothesis (the H_0 is not rejected). At this point, it is worth recalling that the rejection region of the null distribution is defined by the adopted significance level. Accordingly, the probability of a Type I error (rejecting a true H_0) must be equal to the adopted significance level. By way of analogy, a Type II error occurs if a false H_0 is not rejected. The alternative hypothesis (H_A) is another fundamental element of a hypothesis test and is frequently defined as: the null hypothesis is not true (this definition was adopted in this study). Further information regarding the fundamental elements of any hypothesis test can be found in several studies such as Wilks (2011) which is recommended reading.

The Kolmogorov-Smirnov and the Lilliefors tests

The KS test is based on the comparison between the theoretical $[F'(x)]$ and the empirical cumulative distributions $[F(x)]$ (equation 1).

$$D = \max |F'(x) - F(x)| \tag{1}$$

The statistic D describes the largest discrepancy between $[F'(x)]$ and $[F(x)]$. Naturally, sufficiently large D values lead to the rejection of H_0 . According to Stephens (1974) the critical values of the original KS test may be approximated by:

$$KS_{crit} = \frac{K_\alpha}{\sqrt{n + 0.12} + 0.11/\sqrt{n}} \tag{2}$$

Where n is the sample size. K_α is set to 1.358 (1.224) when the test is performed at the 5% (10%) significance level.

The KS test may be regarded as a distribution-free test because its null distribution does not depend upon the explicit form of the distribution under analysis. In other words, the critical values obtained from equation 2 are applicable to any distribution regardless the values of its parameters. As previously described, equation 2 can only be used if the parameters of the distribution have not been estimated from the same data sample used to calculate the KS_{crit} value.

Although the Lilliefors test is also based on equation 2, it cannot be regarded as a distribution-free test because for distributions, such as the gam distribution, the sample distribution of D depends on the sample size and on the values of the shape parameter. Crutcher (1975) tabulated KS_{crit} values for both gam and normal distributions (Table 1). The studies of Husak et al. (2007), Steingnov (2009), Blain (2011) and Wilks (2011) used these critical values to evaluate the suitability of the gam distribution in describing rainfall series.

In order to review the critical values tabulated by Crutcher (1975) we generated new critical values ($NKSL_{crit}$) for both

gam and normal distributions. The Monte Carlo simulations used to approximate the $NKSL_{crit}$ is described as follow:

The simulation procedure starts by generating a large number of samples from the hypothesized distribution. For the gam distribution, we generated $N_s=500000$ samples for each assumed shape and n value. For the normal distribution, we generated $N_s=500000$ data samples for each assumed n value. The parameters of the gam or normal distribution were then estimated from each synthetic data samples. $F(x)$ and $F'(x)$ were then estimated for each synthetic dataset generating 500000 D values for each shape and n values (gam) or for each n value (normal). At this point, it is worth emphasizing that the H_0 is true for each D value. Accordingly, each collection of D values is, in fact, the null distribution of the hypothesis test. Thus, the $NKSL_{crit}$ values were approximated as the $(1-\alpha)$ quantile of each null distribution (α is the adopted significance level). Further information regarding this procedure can be found in several studies such as Wilks (2011) and Shin et al. (2012). Note that this procedure can be used to derive $NKSL_{crit}$ for any shape, sample size and significance level (α) of interest. The r-codes used to run these procedures are exemplified as follow.

```
### Lilliefors test for the 2-parameter gamma distribution
Ns=500000 # number of synthetic samples
n=50 # sample size (given by the user)
shape= 3 # shape parameter (given by the user)
beta=30 # scale parameter (given by the user)
x=matrix(NA,n,1)
lilliefors=matrix(NA,Ns,1)
probpar=matrix(NA,n,1)
pos=matrix(1: n, n, 1)/n
for (i in 1:Ns){
  x[,1]=rgamma(n,shape,1/beta)
  A=log(mean(x))-((sum(log(x)))/n)
  alfali=(1/(4*A))*(1+sqrt(1+(4*A/3)))
  betali=mean(x)/alfali
  probpar[,1]=pgamma(sort(x), alfali, 1/betali, lower.tail = TRUE,
  log.p = FALSE)
  Dmax=max(abs(pos- probpar))
  lilliefors[i,1]=Dmax}
NKScrit5=quantile(lilliefors, probs=0.95) # 5% significance level
NKScrit10=quantile(lilliefors, probs=0.90) # 10% significance level
format(NKScrit5, digits=3)
format(NKScrit10, digits=3)

###Lilliefors test for the normal distribution
Ns=50000 # number of synthetic samples
n=50 # sample size (given by the user)
average=0 # (given by the user)
standard_deviation=2 # (given by the user)
lilliefors=matrix(NA,Ns,1)
probpar= matrix(NA,n,1)
x=matrix(NA,n,1)
lilliefors=matrix(NA,Ns,1)
probpar= matrix(NA,n,1)
pos=matrix(1: n, n, 1)/n
for (i in 1:Ns){
  x[,1]=rnorm(n,average,standard_deviation)
  xsd=data.frame(x)
  averages=mean(x)
```

Table 1. Critical values for the Lilliefors test. Extracted from Crutcher (1975). α is the significance level, Shape is the shape parameter of the 2-parameter gamma distribution and n is the sample size

Shape	$\alpha = 10\%$		$\alpha = 5\%$	
	Sample Size (n) n = 30	Sample Size (n) n > 30	Sample Size (n) n = 30	Sample Size (n) n => 30
2-parameter gamma distribution				
1	0.169	0.95/ \sqrt{n}	0.184	1.05/ \sqrt{n}
2	0.161	0.91/ \sqrt{n}	0.175	0.97/ \sqrt{n}
3	0.151	0.86/ \sqrt{n}	0.165	0.94/ \sqrt{n}
4	0.148	0.83/ \sqrt{n}	0.163	0.91/ \sqrt{n}
8	0.146	0.81/ \sqrt{n}	0.161	0.89/ \sqrt{n}
Normal distribution				
	0.144	0.805/ \sqrt{n}	0.161	0.886/ \sqrt{n}

```

standard_deviations=sapply(xsd,sd)
propar[,1]=pnorm(sort(x), averages, standard_deviations, lower.
tail = TRUE, log.p = FALSE)
Dmaxs=max(abs(pos- propar))
lilliefors[i,1]=Dmaxs}
NKScrit5=quantile(lilliefors, probs=0.95) # 5% significance level
NKScrit10=quantile(lilliefors, probs=0.90) # 10% significance
level
format(NKScrit5, digits=3)
format(NKScrit10, digits=3)
#####

```

On the performance of the tests: type I and II errors

To evaluate the probability of a type I error associated with KS_{crit} , KSL_{crit} and $NKSL_{crit}$ we generated 50000 series from gam (normal) distributions with shape parameter equal to 1, 2 and 3 (with 0 mean and standard deviation equal to 1) and sample sizes equal to 30, 40, 50, 60, 70, 80 and 90. The scale parameter was set to 40. We fitted the gam (normal) distribution to each simulated series. The agreement, for each simulated series, between $F(x)$ and $F^*(x)$ were assessed by using the KS_{crit} , KSL_{crit} and $NKSL_{crit}$. As one may note the H_0 is true for all simulated series. Thus, the frequency of occurrence of type I errors is simply the ratio between the number of cases in which H_0 was erroneously rejected and the number of simulated series (50000).

As previously described, a Type II error occurs when a false H_0 is not rejected. To evaluate the frequency of occurrence of Type II errors obtained by using each of the three critical values we generated 50000 series from gam distributions with shape parameter equal to 1 and 2 and sample size equal to 30, 40, 50, 60, 70, 80 and 90. After that, we fitted a normal distribution to each series and assessed this fit by using the KS_{crit} , KSL_{crit} and $NKSL_{crit}$. In this procedure, the H_0 is false. Thus, the frequency of occurrence of type II errors is simply the ratio between the number of cases in which H_0 was not rejected and the number of simulated series (50000).

As a case of study, the fit of monthly rainfall data, obtained from the weather station of Ribeirão Preto (State of São Paulo, Brazil; 1970-2010), to the gam distribution were evaluated by using the KS_{crit} , KSL_{crit} and $NKSL_{crit}$ tests. This weather station belongs to the Agronomic Institute of Campinas (Instituto Agronômico, IAC/APTA-SAA). These monthly series present no missing values and their consistency have already been verified in previous studies (Blain, 2011). The maximum likelihood estimates of the parameters of the gam distribution were calculated as described in Husak et al. (2007). The goodness-of-fit tests were performed at the 5% significance level (the r-code is described in appendix I). The quantil-quantil plot was used to evaluate the different conclusions indicated by the three tests. The QQ plots may be regarded as a qualitative

method of goodness-of-fit (Wilks, 2011). It is capable of indicating where the parametric representation of the data is inadequate (Wilks, 2011).

Regression models

Theoretically the statistical simulation procedure described in section “The Kolmogorov-Smirnov and the Lilliefors tests” can be used to derive $NKSL_{crit}$ values for any distribution at any significance level (Wilks, 2011). However, we are aware that many researchers around the world may not be familiar with statistical simulations techniques. This fact may be one of the reasons why the KS test has been erroneously used in several scientific papers. Accordingly, by following authors such as Shin et al. (2012) we provided the $NKSL_{crit}$ as a function of shape parameter and sample size (gam) or as a function of sample size (normal) for two frequently used significance levels (5% and 10%). We applied the response function form described in Tolikas and Heravi (2008) to approximate the $NKSL_{crit}$ values for the gam distribution (equation 3). We also verified that the rational model described by equation 4 is suited to approximate the $NKSL_{crit}$ values for the normal distribution.

$$NKSL_{crit} = a + (b/n) + (c/n^2) + d \cdot \text{shape} + (e \cdot \text{shape}^2) \quad (3)$$

$$NKSL_{crit} = (p1 \cdot n^2 + p2 \cdot n + p3) / (n + q1) \quad (4)$$

Where: n is the sample size; shape is the shape parameter of the 2-parameter gamma distribution; a, b, d, e, p1, p2, p3 and q1 are the parameters of the regression models.

The coefficient of determination (r^2), the mean absolute error (MAE) and the square root of the mean square error (RMSE) were used to assess the agreement between the $NKSL_{crit}$ generated from the Monte Carlo simulations (section 2.b) and the $NKSL_{crit}$ calculated from the regression equations. We also performed an additional step to validate the above-mentioned regression models. $NKSL_{crit}$ values were generated from another set of Monte Carlo simulations in which the shape parameter was set to vary from 0.8 to 6.3 (by steps of 0.5) and the sample size was set to vary from 35 to 170 (by steps of 10). The r^2 , MAE and RMSE were used to evaluate the agreement between the $NKSL_{crit}$ values obtained from the Monte Carlo simulations and from the regression equations. It is worth mentioning that this additional step was based on $NKSL_{crit}$ values that were not used to estimate the parameters of the regression models.

3. RESULTS AND DISCUSSION

Tables 2 and 3 presents $NKSL_{crit}$ values obtained by using the r-codes described in section “The Kolmogorov-Smirnov and the Lilliefors tests” As can be noted, the shape parameter

Table 2. Critical values for the Lilliefors test obtained from the procedure described in section 2.b (5% significance level)

Shape	Sample size									
	30	40	50	60	70	80	90	100	110	120
2-parameter gamma distribution										
Critical Values - 5% significant level										
0.5	0.165	0.144	0.131	0.120	0.112	0.106	0.100	0.095	0.091	0.088
1.0	0.155	0.136	0.123	0.113	0.105	0.099	0.093	0.089	0.085	0.082
1.5	0.153	0.134	0.121	0.111	0.103	0.097	0.092	0.087	0.084	0.080
2.0	0.152	0.133	0.120	0.110	0.103	0.097	0.091	0.087	0.083	0.080
2.5	0.151	0.132	0.120	0.110	0.102	0.096	0.091	0.086	0.082	0.079
3.0	0.151	0.132	0.119	0.110	0.102	0.096	0.091	0.086	0.083	0.079
3.5	0.151	0.132	0.119	0.109	0.102	0.096	0.090	0.086	0.082	0.079
4.0	0.150	0.132	0.119	0.109	0.102	0.096	0.090	0.086	0.082	0.079
4.5	0.150	0.132	0.119	0.109	0.102	0.095	0.090	0.086	0.082	0.079
5.0	0.150	0.132	0.119	0.109	0.102	0.095	0.090	0.086	0.082	0.079
5.5	0.150	0.131	0.118	0.109	0.102	0.095	0.090	0.086	0.082	0.079
6.0	0.150	0.131	0.118	0.109	0.101	0.095	0.090	0.086	0.082	0.079
6.5	0.150	0.131	0.118	0.109	0.101	0.095	0.090	0.086	0.082	0.079
7.0	0.150	0.131	0.118	0.109	0.101	0.095	0.090	0.086	0.082	0.079
7.5	0.150	0.131	0.118	0.109	0.101	0.095	0.090	0.086	0.082	0.078
8.0	0.150	0.131	0.118	0.109	0.101	0.095	0.090	0.086	0.082	0.079
Normal distribution										
	0.148	0.130	0.118	0.108	0.101	0.095	0.089	0.085	0.081	0.078

Table 3. Critical values for the Lilliefors test obtained from the procedure described in section 2.b (10% significance level)

Shape	Sample size									
	30	40	50	60	70	80	90	100	110	120
2-parameter gamma distribution										
Critical Values - 10% significant level										
0.5	0.149	0.131	0.119	0.110	0.102	0.096	0.091	0.087	0.083	0.080
1.0	0.141	0.124	0.112	0.103	0.096	0.090	0.085	0.081	0.077	0.074
1.5	0.139	0.122	0.110	0.101	0.094	0.089	0.084	0.080	0.076	0.073
2.0	0.138	0.121	0.109	0.101	0.093	0.088	0.083	0.079	0.076	0.073
2.5	0.137	0.121	0.109	0.100	0.093	0.087	0.083	0.079	0.075	0.072
3.0	0.137	0.120	0.109	0.100	0.093	0.087	0.083	0.079	0.075	0.072
3.5	0.137	0.120	0.108	0.100	0.093	0.087	0.083	0.079	0.075	0.072
4.0	0.137	0.120	0.108	0.100	0.093	0.087	0.082	0.078	0.075	0.072
4.5	0.137	0.120	0.108	0.099	0.093	0.087	0.082	0.078	0.075	0.072
5.0	0.137	0.120	0.108	0.099	0.093	0.087	0.082	0.078	0.075	0.072
5.5	0.137	0.120	0.108	0.099	0.093	0.087	0.082	0.078	0.075	0.072
6.0	0.137	0.120	0.108	0.099	0.092	0.087	0.082	0.078	0.075	0.072
6.5	0.137	0.120	0.108	0.099	0.092	0.087	0.082	0.078	0.075	0.072
7.0	0.136	0.120	0.108	0.099	0.092	0.087	0.082	0.078	0.075	0.072
7.5	0.136	0.119	0.108	0.099	0.092	0.087	0.082	0.078	0.075	0.072
8.0	0.136	0.120	0.108	0.099	0.092	0.087	0.082	0.078	0.075	0.072
Normal distribution										
	0.135	0.119	0.107	0.099	0.092	0.086	0.082	0.078	0.074	0.071

of the gam distribution was set to vary from 0.50 to 8.0 by steps of 0.50. The scale parameter was set to 30 without loss of generality. By assuming these different values of the shape parameter the probability density function of the gam distribution varies from an exponential form to a bell-shaped form that tends to approach the normal distribution. These

new critical values were also generated for several sample sizes (n) that vary from 30 (a climatological normal period) to 120 by steps of 10.

Surprisingly, the $NKSL_{crit}$ values (Tables 2 and 3) are considerably lower than the KSL_{crit} presented by Crutcher (1975; Table 1). Accordingly, one may indicate that the

probability of occurrence of Type I and II errors associated with the use of the KS_{crit} and $NKSL_{crit}$ will also differ from each other. This inference is further evaluated from the results presented in tables 3 and 5.

As expected (Crutcher, 1975; Lilliefors, 1967; Steinskog et al., 2007; Vlček and Huth, 2009; Wilks, 2011), the frequency of occurrence of Type I errors associated with the use of the KS_{crit} is much lower than the adopted significant levels (Tables 4). The rejection rates obtained by using the KS_{crit} were always lower than 1% (at both 5 and 10% significance level). The frequency of Type I errors obtained by using the KSL_{crit} values are higher than those associated with the KS_{crit} . However, they are also considerably

lower than the adopted significant level. Therefore, the results presented in table 4 indicate that the Lilliefors test calculated by using the KSL_{crit} values may be regarded as a conservative test with respect to the Type I errors. This statement holds for both gam and normal distributions. The rejection rates obtained from the $NKSL_{crit}$ were capable of meeting the adopted significant level. The frequency of Type I errors associated with these latter critical values varied from 4.83% to 5.34% (5% significance level) and from 9.45% to 10.53% (10% significance level). It is also worth emphasizing that the frequencies of occurrence of Type I errors, obtained from the $NKSL_{crit}$ were little affected by

Table 4. Frequency of Type I errors (%) associated with the use of the KS_{crit} (equation 2), KSL_{crit} (Crutcher, 1975) and $NKSL_{crit}$ (section The Kolmogorov-Smirnov and the Lilliefors tests)

Sample Size	30	40	50	60	70	80	90
5% significance level shape= 1; scale = 40							
KS_{crit}	0.04	0.01	0.03	0.01	0.01	0.03	0.03
KSL_{crit}	0.80	0.75	0.71	0.68	0.82	0.92	0.83
$NKSL_{crit}$	4.86	5.07	5.00	5.11	5.05	5.18	4.81
shape= 2; scale = 40							
KS_{crit}	0.00	0.00	0.00	0.02	0.02	0.02	0.01
KSL_{crit}	1.31	1.27	1.38	1.52	1.64	1.67	1.68
$NKSL_{crit}$	4.98	5.10	4.81	5.12	5.34	5.12	4.95
shape= 3; scale = 40							
KS_{crit}	0.01	0.01	0.02	0.01	0.01	0.02	0.02
KSL_{crit}	2.36	1.60	1.88	2.03	2.10	2.21	2.29
$NKSL_{crit}$	4.97	4.83	5.18	4.90	5.02	5.14	4.80
10% significance level shape= 1; scale = 40							
KS	0.11	0.12	0.14	0.1	0.08	0.19	0.18
KSL_{crit}	2.2	1.78	2.13	2.15	2.54	2.53	2.75
$NKSL_{crit}$	10.53	9.45	9.51	9.75	10.36	10.35	10.36
shape= 2; scale = 40							
KS_{crit}	0.04	0.05	0.1	0.05	0.11	0.05	0.05
KSL_{crit}	3.16	2.33	2.62	2.74	3.02	3.16	3.14
$NKSL_{crit}$	10.24	10.07	9.99	10.44	10.52	9.95	10.02
shape= 3; scale = 40							
KS_{crit}	0.05	0.09	0.05	0.1	0.09	0.09	0.08
KSL_{crit}	4.69	3.83	4.08	4.38	4.79	4.53	4.18
$NKSL$	9.89	9.98	9.74	9.99	10.2	10.06	9.81
Normal distribution 5% significance level							
KS_{crit}	0.00	0.00	0.01	0.00	0.01	0.01	0.01
KSL_{crit}	2.61	2.72	2.80	2.85	3.26	3.30	3.43
$NKSL_{crit}$	4.97	5.16	4.81	4.89	4.85	4.82	5.23
10% significance level							
KS_{crit}	0.05	0.04	0.05	0.07	0.05	0.06	0.06
KSL_{crit}	6.49	5.79	6.38	6.65	6.94	7.23	7.45
$NKSL_{crit}$	10.22	9.53	9.94	9.55	9.71	10.13	9.56

Table 5. Frequency of Type II errors (%) associated with the use of the KS_{crit} (equation 2), KSL_{crit} (Crutcher, 1975) and $NKSL_{crit}$ (section “The Kolmogorov-Smirnov and the Lilliefors tests”)

	30	40	50	60	70	80	90
5% significance level							
shape= 1 scale =40							
KS	88.93	77.99	65.12	51.01	37.76	26.00	16.79
KSLcrut	25.17	12.52	5.66	2.18	0.75	0.24	0.06
NKSL	16.26	7.14	1.46	0.52	0.15	0.05	0.01
shape= 2 scale =40							
KS	97.74	95.25	91.83	87.80	82.61	77.01	70.74
KSLcrut	55.73	42.63	31.30	22.97	16.06	11.01	7.39
NKSL	43.69	31.21	13.99	10.31	7.36	4.57	3.26
10% significance level							
shape= 1 scale =40							
KS	75.41	59.24	43.55	29.87	18.56	11.04	5.79
KSLcrut	12.90	6.00	2.26	0.72	0.20	0.08	0.02
NKSL	8.44	3.32	1.16	0.40	0.09	0.03	0.00
shape= 2 scale =40							
KS	92.87	87.42	80.91	73.64	65.76	57.98	50.22
KSLcrut	38.56	29.01	19.13	12.78	7.99	5.06	3.02
NKSL	29.73	20.89	13.15	9.14	5.61	3.43	1.28

Table 6. Monthly rainfall series of Ribeirão Preto, State of São Paulo, Brazil (1970-2010). Shape and Scale are the parameters of the gamma distribution, # zeroes is the number of zeroes in each monthly series and KSL_{crit} and $NKSL_{crit}$ are the critical values for the Lilliefors test respectively obtained from Crutcher (1975) and from section “The Kolmogorov-Smirnov and the Lilliefors tests” of this study

Month	Shape	Scale	Dmax	# zeroes	KSL_{crit}	$NKSL_{crit}$
January	6.70	41.35	0.061	0	0.139	0.130
February	3.31	67.97	0.087	0	0.147	0.131
March	4.49	35.98	0.073	0	0.142	0.130
April	1.30	62.32	0.084	0	0.164	0.133
May	1.25	49.90	0.153	0	0.164	0.133
June	0.75	49.74	0.115	8	0.183	0.152
July	0.75	39.35	0.080	6	0.177	0.148
August	0.74	49.95	0.083	13	0.198	0.163
September	1.12	56.72	0.095	0	0.164	0.133
October	3.51	34.15	0.123	0	0.142	0.131
November	5.89	30.43	0.055	0	0.142	0.130
December	7.93	35.21	0.108	0	0.139	0.130

the different shape parameters and/or sample sizes adopted in this study.

Regarding the probability of a Type II error, the results presented in table 5 indicate that the KS_{crit} must not be used when the parameters of the distribution have been fitted from the same data set used to apply the KS test. In other words, when erroneously applied, the KS_{crit} frequently leads to the conclusion that a given bunch of data was drawn from a parent distribution that may be considerably different from the process that has generated the data. For instance, when n was set to 30 the rate of type II errors were, approximately, 89% ($\alpha=0.05$) and 75% ($\alpha=0.10$). Different from what is observed for the Type I errors, the frequency of Type II errors, obtained from both KSL_{crit} and $NKSL_{crit}$, is a decreasing function of the sample size. In other

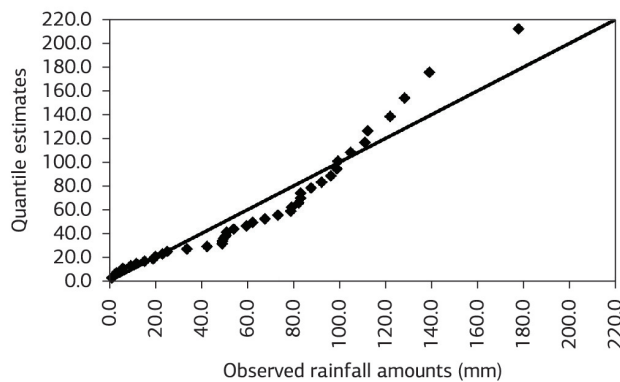


Figure 1. Quantile-quantile plot. Observed monthly rainfall data of the location of Ribeirão Preto, State of São Paulo, Brazil (1970-2010) and quantile estimates obtained from a 2-parameter gamma distribution.

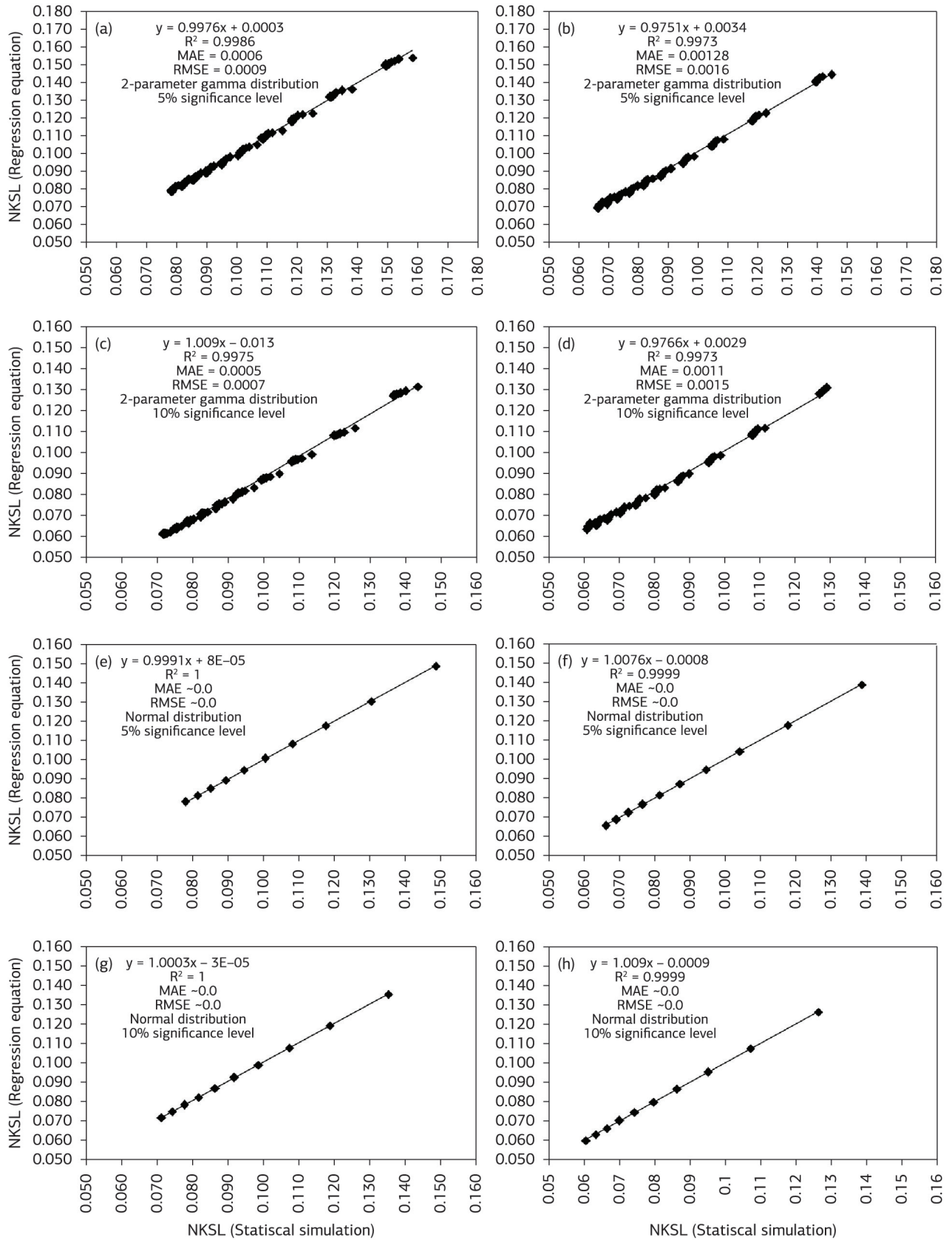


Figure 2. Simple linear regression analysis between the $NKSL_{crit}$ generated from the Monte Carlo simulations (section 2.b) and the $NKSL_{crit}$ calculated from the regression equations. MAE is the mean absolute error and RMSE is the square root of the mean. Figures on the left are based on $NKSL_{crit}$ values that were not used to estimate the parameters of the regression models (equations 5 and 6).

words, the smaller the sample size, the lower the power of the tests. The results presented in table 5 also indicate that when compared to the KSL_{crit} the $NKSL_{crit}$ increases the power of the Lilliefors test. As can be noted, the frequency of Type II errors obtained from the $NKSL_{crit}$ values are lower than those obtained from the KSL_{crit} values for all simulations performed in this study. Thus, by considering the results presented in tables 4 and 5, we concluded that the $NKSL_{crit}$ improves the capability of the lilliefors test in correctly reject a false H_0 .

As previously described, the KSL_{crit} are tabulated only for discrete values of the shape parameter. Authors such Vlček and Huth (2009) and Wilks (2011) deal with this difficulty by adopting the KSL_{crit} associated with the tabulated shape parameter closest to the estimated one. The same procedure was used in this study because the same difficult were observed when we used the KSL_{crit} to evaluate the fit of the monthly rainfall series of Ribeirão Preto to the gam distribution (table 6).

The $NKSL_{crit}$ and KSL_{crit} lead to different conclusions only for the rainfall amounts observed during the month of May (Table 6). The KSL_{crit} values indicates that the gam distribution can be used to assess probability of the monthly rainfall amounts of weather station of Ribeirão Preto for any month (including the month of May). However, according to the $NKSL_{crit}$ the gam distribution cannot be used to assess the probability of occurrence of the rainfall amounts observed during the months of May. At this point it becomes worth emphasizing that if the gam distribution is an appropriate model for representing the above-mentioned monthly rainfall amounts, the points of figure 1 (Q-Q plot) should lie close to the unit diagonal (Coles, 2001). However, the visual inspection of figure 1 reveals substantial departures from linearity. Thus, we concluded that the results depicted in figure 1 are more consistent with the conclusion reached by using the $NKSL_{crit}$. This last inference is also consistent with the results presented in tables 3, in the sense that the $NKSL_{crit}$ improves the capability of the lilliefors test in correctly reject a false H_0 .

Regression equations

Although we claim that the r-code presented in this study may be regarded as an easy-to-use procedure, one may correctly argue that they require some level of knowledge of the r-software. Thus, by considering that the main goal of this study is to stimulate the use of the lilliefors test, we have also provided polynomial equations (equations 5 and 6) capable of calculating the $NKSL_{crit}$ for the range of shape parameters and sample sizes presented in tables 1 and 2. In this view, figure 2 depicts the linear regression between the $NKSL_{crit}$ values obtained from the Monte Carlo simulations and the $NKSL_{crit}$ values obtained from equations 5-6. The r^2 , MAE and RMSE indicate that the regression equations can

be used to approximate the $NKSL_{crit}$ values for the gam and normal distributions.

$$NKSL_{crit(5\%)} = 0.04889 + (4.601/n) + (-42.24/n^2) - 0.001945shape + 0.0001695shape^2 \quad (5.1)$$

$$NKSL_{crit(10\%)} = 0.04435 + (4.201/n) + (-39.07/n^2) - 0.001639shape + 0.0001416shape^2 \quad (5.2)$$

$$NKSL_{crit(5\%)} = (((-8.530 \cdot 10^{-5})(n^2)) + 0.05645n + 4.906) / (n + 13.90) \quad (6.1)$$

$$NKSL_{crit(10\%)} = (((-8.225 \cdot 10^{-5})(n^2)) + 0.05255n + 4.422) / (n + 13.83) \quad (6.2)$$

4. SUMMARY

The Kolmogorov-Smirnov test is a frequently used goodness-of-fit test. However, in its original form this test requires that the parameters of the theoretical distribution have not been estimated from the same bunch of data used to apply this goodness-of-fit test. As described in several studies, if the original KS test is applied under this above-mentioned situation the probability of accepting a false H_0 becomes unacceptably high. The results obtained in this study are consistent with these previous studies. This study demonstrated that the use of the original KS_{crit} leads to a frequency of occurrence of Type I errors much lower than the adopted significant level and to a high probability of failure to reject a false H_0 (Type II error). The so-called Lilliefors test has been proposed to remedy this drawback. This latter test is capable of achieving a better balance between type I and II errors. In this sense, by using the critical values described in Crutcher (1975), for both 2-parameter gamma and normal distributions, one is capable of bringing the probability of a Type I error more close to the adopted significance level. However, this study also demonstrated that the frequency of occurrence of Type I errors associated with the above-mentioned values is still considerably lower than the adopted significance level. In addition, especially for small sample sizes, the probability of a Type II error associated with the critical values described in Crutcher (1975) may be unacceptably high. For instance, from the results presented in table 5, one may verify that when the sample size was set to 30 the frequency of occurrence of Type II errors, obtained by using these critical values (at the 5% significance level) were greater than 25%. From the above-mentioned results, one may also note that the probability of a Type II error decreases as the sample size increases.

We revisited the critical values presented by Crutcher (1975). Based on sets of 500000 simulated series with different sample sizes we derived new critical values for the Lilliefors test that can be used to assess the fit of the

2-parameter gamma and normal distributions. By using these new critical values (or the r-code devised to calculate these critical values) we were able to meet the adopted significance level in all simulations carried out in this study. In addition, these new critical values also led to the lowest frequency of occurrence of Type II errors observed in this study. Finally, by assuming that many researchers around the world may not be familiar with statistical simulations techniques (or with the R-software), we have also provided polynomial equations that eliminate the need to use the r-codes to calculate the $NKSL_{crit}$ values presented in this study. We hope this study stimulates the correct use of the Lilliefors test.

ACKNOWLEDGEMENTS

The author greatly appreciates Dr. Patrícia Cia's constructive suggestions.

REFERENCES

- BLAIN, G.C. Monthly values of the standardized precipitation index in the State of São Paulo, Brazil: trends and spectral features under the normality assumption. *Bragantia*, v.71, p.122-131, 2011. <http://dx.doi.org/10.1590/S0006-87052012005000004>
- CHEN, B.; YANG, J.; PU, J. Statistical Characteristics of Raindrop Size Distribution in the Meiyu Season Observed in Eastern China. *Journal of the Meteorological Society of Japan Ser. II*, v.91, p.215-227, 2013. <http://dx.doi.org/10.2151/jmsj.2013-208>
- COLES, S. An introduction to statistical modeling of extreme value. London: Springer, 2001. <http://dx.doi.org/10.1007/978-1-4471-3675-0>
- CRUTCHER, H.L. A note on the possible misuse of the Kolmogorov-Smirnov test. *Journal of Applied Meteorology*, v.14, p.1600-1603, 1975. [http://dx.doi.org/10.1175/1520-0450\(1975\)014<1600:ANOTPM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1975)014<1600:ANOTPM>2.0.CO;2)
- HUSAK, G.J.; MICHAELSEN, J.; FUNK, C. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *International Journal of Climatology*, v.27, p.935-944, 2007. <http://dx.doi.org/10.1002/joc.1441>
- LILLIEFORS, H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, v.62, p.399-402, 1967. <http://dx.doi.org/10.1080/01621459.1967.10482916>
- LILLIEFORS, H.W. On the Kolmogorov-Smirnov test on the exponential distribution with mean unknown. *Journal of the American Statistical Association*, v.64, p.387-389, 1969. <http://dx.doi.org/10.1080/01621459.1969.10500983>
- SHIN, H.; JUNG, Y.; JEONG, C.; HEO, J. Assessment of modified Anderson-Darling test statistics for the generalized extreme value and generalized logistic distributions. *Stochastic Environmental Research and Risk Assessment*, v.26, p.105-114, 2012. <http://dx.doi.org/10.1007/s00477-011-0463-y>
- TOLIKAS, K.; HERAVI, S. The Anderson-Darling goodness-of-fit test statistic for the three-parameter lognormal distribution. *Communications in Statistics - Theory and Methods*, v.37, p.3135-3143, 2008. <http://dx.doi.org/10.1080/03610920802101571>
- STEINSKOG, D.J.; TJØSTHEIM, D.B.; KVAMSTØ, N.G. A cautionary note on the use of the Kolmogorov-Smirnov test for normality. *Monthly Weather Review*, v.135, p.1151-1157, 2007. <http://dx.doi.org/10.1175/MWR3326.1>
- STEPHENS, M.A. EDF Statistics for goodness of fit. *Journal of the American Statistical Association*, v.69, p.730-737, 1974. <http://dx.doi.org/10.1080/01621459.1974.10480196>
- VLČEK, O.; HUTH, R. Is daily precipitation Gamma-distributed? Adverse effects of an incorrect application of the Kolmogorov-Smirnov test. *Atmospheric Research*, v.93, p.759-766, 2009. <http://dx.doi.org/10.1016/j.atmosres.2009.03.005>
- WILKS, D.S. Statistical methods in the atmospheric sciences. 2nd ed. San Diego: Academic Press, 2011. 629p.

APPENDIX I

```
#####
# datamatrix is a matrix in which each column corresponds to each month
datamatrix= as.matrix(read.table("datamatrix.txt", head=T))
shape=matrix(NA,12,1)
scale=matrix(NA,12,1)
Dmax=matrix(NA,12,1)
NKSLcrit5=matrix(NA,12,1)
NKSLcrit10=matrix(NA,12,1)
pvalue=matrix(NA,12,1)
for (month in 1:12){
data=datamatrix[,month]
data1=data> 0
datap=data[data1] # the 2-parameter gamma is undefined for x ≤ 0
n=length(data)
np=length(datap)
nz=n-np
probzero=(n-nz)/n
Ns=50000
probacum=matrix(NA,np,1)
lilliefors=matrix(NA,Ns,1)
probpar= matrix(NA,np,1)
A=log(mean(datap))-((sum(log(datap)))/np)
shape[month,1]=(1/(4*A))*(1+sqrt(1+(4*A/3)))
scale[month,1]=mean(datap)/shape[month,1]
pos=matrix(1:np, np, 1)/np
probacum[,1]= (pgamma(sort(datap), shape[month,1], 1/scale[month,1], lower.tail = TRUE, log.p = FALSE))
Dmax[month,1]=max(abs(pos- probacum))
#####Lilliefors
x=matrix(NA,np,1)
lilliefors=matrix(NA,Ns,1)
probpar=matrix(NA,np,1)
poss=matrix(1: np, np, 1)/np
for (i in 1:Ns){
x[,1]=rgamma(np,shape[month,1],1/scale[month,1])
As=log(mean(x))-((sum(log(x)))/np)
alfals=(1/(4*As))*(1+sqrt(1+(4*As/3)))
betals=mean(x)/alfals
probpar[,1]=pgamma(sort(x), alfals, 1/betals, lower.tail = TRUE, log.p = FALSE)
Dmaxs=max(abs(poss- probpar))
lilliefors[i,1]=Dmaxs}
NKSLcrit5[month,1]=quantile(lilliefors, probs=0.95)
NKSLcrit10[month,1]=quantile(lilliefors, probs=0.90)
m=lilliefors>Dmax[month,1]
pvalue[month,1]=(length(lilliefors[m]))/Ns}
Goodness=c("shape", shape, "scale", scale, "Dmax", Dmax, "NKSLcrit5%", NKSLcrit5, "NKSLcrit10%", NKSLcrit10,
"p-value", pvalue)
write.csv(Goodness, "GoodnessGamma.csv")
#####
```