Reliability Analysis in the Software R

Fernanda Alves Araújo, Paulo César de Resende Andrade

Institute of Science and Technology, UFVJM University, Brazil.

Received: 1 Sept 2020; Received in revised form: 22 Oct 2020; Accepted: 3 Nov 2020; Available online: 7 Nov 2020

***Abstract*— *The reliability study can be used to analyze the times of faults of the equipment and to determine which distribution is adjusted better to the data. There is software that executes this type of procedure, however, most of them are developed for application in the industrial sector, being generally paid and closed. This work aims to develop a code in R software capable of analyzing and determine which distribution is adjusted better to the data, using goodness of fit tests (numerics and graphics). Famous distributions such as Weibull and Lognormal were implemented, as well as complex distributions, such as the Generalized Gamma. For analysis, the code displays results from various tests such as Likelihood Ratio Test (LRT), Kolmogorov-Smirnov (KS) and Chi-square (χ2), in addition to generating graphics of density, accumulated density, reliability, risk and showing the roles of the probability of distributions. All the code was developed in R, as it is a free platform, so it facilitates the work of researchers and companies in the reliability sector. Three sets of equipment failure time data were analyzed, the results found has been coherent and some superior cases when compared with other works and software.*

***Keywords*— Data modeling, failure times, probability distributions, R code, survival analysis.**

I. INTRODUCTION

The modeling of failure times is important for the study of reliability. Therefore, to model mathematically the studied objects, it is necessary to use the probability distributions that relate a given value of the studied variable with the probability of occurrence (CRESPO, 2005).

According to Haviaras (2005) the distributions most used to model failure times are Exponential, Gamma, Lognormal and Weibull. The analysis consists of selecting the distribution that best fits the failure times (WUTKE; SELLITTO, 2008). After defining the distribution that characterizes the data, it is possible to obtain the reliability function, risk function, density function and the average time to failure (FOGLIATTO; RIBEIRO, 2009)

In general, this modeling is done using software that indicates which distributions best fit the failure times. The software R (R CORE TEAM, 2020) allows the application of numerical and analytical methods, being a differential about software that allows only analytical calculations. Besides, the R software is free, making it easier for users to obtain fault data modeling, usually limited to paid software and developed for restricted environments, such as corporate ones.

Moreover, some methods can be used to determine or to indicate which better model describes the failure time data. These methods can be divided into graphs and numeric. The paper of the probability, for example, is a graphical method in which it linearizes the accumulated density function (FALCETTA, 2000). Regarding numerical methods, there are tests such as AIC (Akaike Information Criterion), Loglik (Log-Likelihood) and LRT (Likelihood Ratio Test).

The objective of the work is to create a function (RELPF) using software R to analyze two groups of failure times and to determine which model is the most suitable. This function contains analytical and numerical distributions, as well as the use of graphical and numerical methods to conclude about the adequation or not of the specified distribution.

II. METHODOLOGY

To perform the RELPF function, it is necessary to supply failure times of a product or equipment. Through these failure times, the reliability analysis is performed.

The R is software and a programming language. It manipulates, analyzes and simulates the probability of the

https://dx.doi.org/10.22161/ijaers.711.7 ISSN: 2349-6495(P) | 2456-1908(O)

data to follow the most diverse models of existing distributions. For this, it is required to install and load packages, such as survival, flexsurv, zoo, sfsmisc, fitdistrplus, weibullness. All this process of installation and/or loading of packages were developed so that it is not necessary to execute manually.

The distributions implemented in RELPF for data modeling were: Generalized Gamma, Gamma, Weibull, Weibull with 3 parameters, Lognormal, Normal and Exponential. It is by means of these distributions that if which verifies the best one if it adjusts to the equipment failure time.

For each distribution, the Maximum Likelihood Method was used to estimate the parameters. The first distribution to be interpreted by the R is Generalized Gamma, later is possible to get the Loglik that serves of measure of adjustment to effect the LRV of the other distributions, with the use of the package flexsurv. Moreover, it is possible to get the AIC of the Generalized Gamma distribution.

In relation to the other distributions, beyond the AIC and Loglik, the LRV was carried through. The LRV serves as a statistical test of the fit quality between two models. In addition, the p-value was performed, it is characterized as a descriptive number for decision making in the tests. In addition, MTTF (mean time between failures), t10 and t50 (the period in which 10% and 50% of equipment fail, respectively) were also calculated. However, of the Weibull distribution of 3 parameters, only its parameters were extracted.

Ahead, the tack tests were carried out. The $\chi 2$ and KS tests had been gotten for all distributions, except for the distribution Gamma Generalizada and Weibull of 3 parameters. These tests can be performed in several ways, then, the results had a small difference from the results of famous software such as ProConf (FRITSCH; RIBEIRO, 1998). However, the studies of FERNANDES (2013) and FERREIRA (1996) were observed and considered to perform $\chi 2$ and KS tests.

Regarding the graphical presentation of the distributions, they are used to show the behavior of the density, reliability and risk of the equipment over time. Only the Generalized Gamma distribution has no graphical representation. Each graph was assembled using functions from the survival and flexsurv packages and generated with simple functions such as the plot.

After that, the paper of probability of the distributions Exponential, Lognormal, Weibull and Normal was configured using linearization functions such as abline, available in the basic libraries of R. For all these generated graphs, had been used functions that carry these graphs to a graphical window, in order to get an easy and organized visualization.

Finally, all numerical information is organized into five tables that are displayed in software R when executing the RELPF function. Table 1 presents the Loglik, AIC, LRV and p-value for all distributions, except in the particularities of the Generalized Gamma and Weibull of 3 parameters. Table 2 presents the parameters, which vary in quantity depending on each distribution. Table 3 shows the t10, t50 and MTTF for the distributions. Finally, Tables 4 and 5 show the results of the $\chi 2$ and KS tests respectively, indicating the rejection or not of the hypothesis.

To validate the RELPF function, the code was tested with three data sets. The first was the study of the failures of an electronic overseer of bottles in a line of production (SILVA; ANDRADE, 2018), with 43 data. The second contains the failure time of 14 optical fiber transmitters (SILVA et al, 2017). These two data sets have different amounts of data with the intention to verify the behavior of function RELPF. The third was a data set provided by Shanker (2016) with 15 data on failure times of an electronic component and has for objective the justification and comparison of data analyzed when comparing with software Proconf.

III. RESULTS AND DISCUSSIONS

To start the reliability analysis using the code, the user must insert the data organized in a column with the failure times of the equipment to be analyzed. From these times, a histogram related to the occurrence of these failures with the time is generated, as shown in Figure 1.

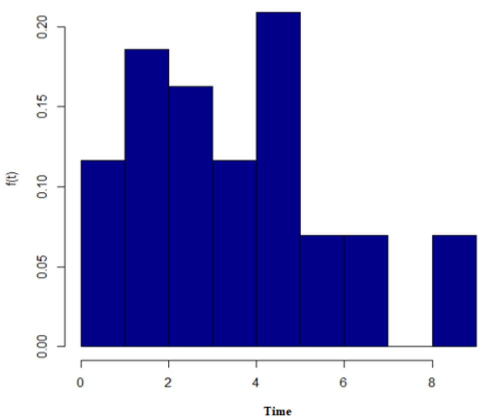


Fig. 1: Histogram generated by the software R of the failure time of electronic overseer of bottles.

Source: Authors.

Using the survreg, flexsurv and fitdistrplus functions from the library previously installed in the RELPF function, all parameters, shape and/or scale, depending on the distribution, are calculated. These calculations are carried through utilizing the mathematical operations in the software R, the results are stored to be presented to the end of the execution in the Table of the R. Tables 1 and 2 show these parameters for each distribution in relation to the first two data sets tested.

Table 1. Values of the shape and scale parameters generated by the RELPF function of the failure time of electronic overseer of bottles.

Distributions	Shape parameter 1	Scale parameter	Shape parameter 2
Lognormal	1.00647	0.80075	-
Exponential	0.28379	-	-
Weibull	1.64422	3.93472	-
Normal	3.52372	2.16192	-
Gamma	2.12776	1.65611	-
Generalized Gamma	1.52238	-0.65387	1.52561
Weibull 3 parameters	1.62808	3.90628	0.02136

Source: Authors.

Table 2. Values of the shape and scale parameters generated by the RELPF function of the failure time of fiber optic transmitters.

Distributions	Shape parameter 1	Scale parameter	Shape parameter 2
Lognormal	5.49773	0.77844	-
Exponential	0.00311	-	-
Weibull	1.47672	357.52894	-
Normal	321.92857	225.26	-
Gamma	1.95811	164.43926	-
Generalized Gamma	5.64848	-0.28054	0.38932
Weibull 3 parameters	1.05807	272.28662	55.2844

Source: Authors.

Table 3 presents the results of the parameters of the distributions using the fiber optic data by the ProConf software. Comparing Table 2 with Table 3, it is observed that the values are very close, except for the gamma that presents a significant difference.

Table 3. Values of the distribution parameters using the ProConf software for fiber optic failure data.

Distribution	Shape parameter	Scale parameter
Lognormal	5.4977	0.606
Exponential	0.0031	-
Weibull	1.4765	357.5035
Normal	321.9286	233.7633
Gamma	1.586	212.5604

Source: Authors.

Regarding the Gamma, this difference can be justified when comparing the results published by Shanker (2016) with ProConf and the RELPF function. See Table 4. Note that the results of Shanker (2016) and RELPF are very close, the existing differences are due to rounding. Shanker (2016) and RELPF use the maximum likelihood estimate, however, ProConf does not have complex mathematical tools for calculating efficiently for the Gamma distribution.

Table 4. Comparing the values of the parameters of the gamma distribution.

	Shape parameter	Scale parameter
Shanker (2016)	1.442	19.231
RELPF	1.44194	19.10058
ProConf	1.198	24.3493

Source: Authors.

Moreover, the function implemented in R has a precision in decimal places greater than the ProConf software, getting more accurate results. The precision of the decimal places in the R can increase or decrease since the user has control over this aspect. Another important factor is that the RELPF function is able to calculate the parameters for more complex distributions such as the Generalized Gamma and Weibull of 3 parameters.

Observing the values of the parameters for the data of the bottle inspector (Table1) with those generated by ProConf (Table 5) it is possible to see a similarity between the results with a greater difference in the Gamma as justified previously. Therefore, the RELPF function can generate good results for all distributions, as much for small how much for great amount of data.

Table 5. Values of the distribution parameters using the ProConf software for data from bottle inspectors.

Distribution	Shape parameter	Scale parameter
Lognormal	1.0065	0.6412
Exponential	0.2838	-
Weibull	1.6445	3.9349
Normal	3.5237	2.1875
Gamma	1.9945	1.7876

Source: Authors.

After obtaining the distribution parameters, it is possible to calculate the mean time between failures (MTTF), the t10 and t50, which vary according to each model. Table 6 shows these values taken directly from the RELPF function for the data set obtained from overseers of bottles.

Table 6. MTTF, t10 and t50 of data from electronic overseers of bottles gotten by the function implemented in R.

Distributions	MTTF	t10	t50
Lognormal	3.76997	2.96401	4.08304
Exponential	3.52372	0.37126	2.44246
Weibull	3.51947	1.00119	3.14851
Normal	3.52372	0.79971	3.52372
Gamma	3.52380	0.57513	1.39406

Source: Authors.

The results for MTTF, t10 and t50 do not need to be generated for the fiber optic data set or compared with ProConf, because they are mathematical operations that use the parameters. As seen previously, the parameters are correct; therefore, the results for MTTF, t10 and t50 are correct and the RELPF function can efficiently calculate them for any data set size.

After the attainment of the parameters of each distribution, is possible to generate the graphs of reliability, risk, density and accumulated density of the distributions. Figures 2 and 3 show the reliability and risk graphs for the Weibull function for the two studies used.

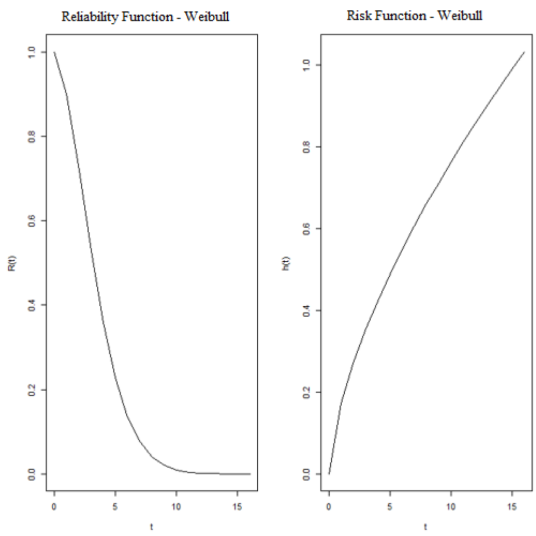


Fig. 2: Reliability and risk graphs for Weibull distribution of failure time for electronic R bottle inspectors

Source: Authors.

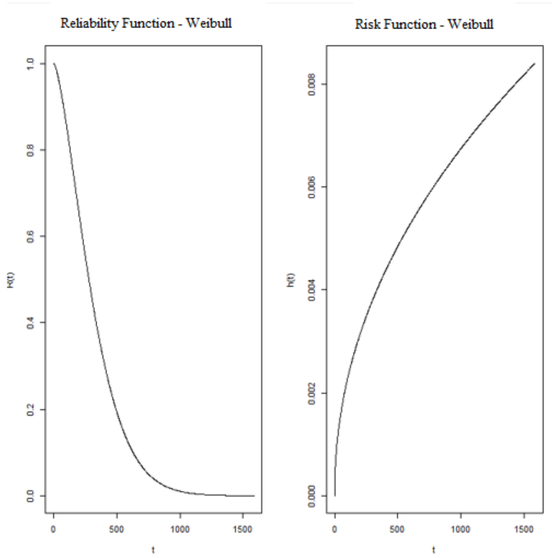


Fig. 3: Density, reliability and risk graphs for the Weibull distribution of the failure time of fiber optic transmitter.

Source: Authors.

Another advantage of the use of software R is the control of the axes of the graph. The RELPF function automatically adjusts these axes according to the data, but the user can still modify them and to modify the scale of the graphs, obtaining a better response compared to other software that do not allow such action.

Ahead, the implemented function performs the LRV, AIC and Loglik tests, in addition to presenting the p-value. Tables 7 and 8 present all this information for each one of

https://dx.doi.org/10.22161/ijaers.711.7 ISSN: 2349-6495(P) | 2456-1908(O)

the distributions of the two analyzed data sets. Moreover, this information is not gotten in some software such as ProConf.

Table 7. Table generated by the RELPF function with the results of LRV, AIC and Loglik for the failure time of the bottle inspectors

Distributions	Loglik	AIC	LRT
Lognormal	-93.32693	195.4134	0.13568
Exponential	-94.84062	191.6812	3.16306
Weibull	-93.39699	190.7940	0.27581
Normal	-95.70671	195.4134	4.89525
Gamma	-93.25909	190.6060	0.08782
Generalized Gamma	-93.25909	192.5182	-

Source: Authors.

Table 8. Table generated by the RELPF function with the results of LRV, AIC and Loglik for the failure time of fiber optic transmitters.

Distributions	Loglik	AIC	LRT	p-value
Lognormal	94.73752	192.3342	9.13289	0.00251
Exponential	97.15928	196.3186	13.97642	0.00092
Weibull	90.58656	185.1731	0.83097	0.36199
Normal	94.16712	192.3342	7.9921	0.0047
Gamma	90.17107	186.7666	2.42445	0.11945
Generalized Gamma	90.17107	186.3421	-	-

Source: Authors.

After the exhibition of all these specific information of the distributions, the code presents for the user the graphical method for verification of the adequacy or not of the distributions by means of the probability paper. In Figures 4 and 5 the papers of the probability of the distributions Weibull and Lognormal for the two analyzed studies are presented.

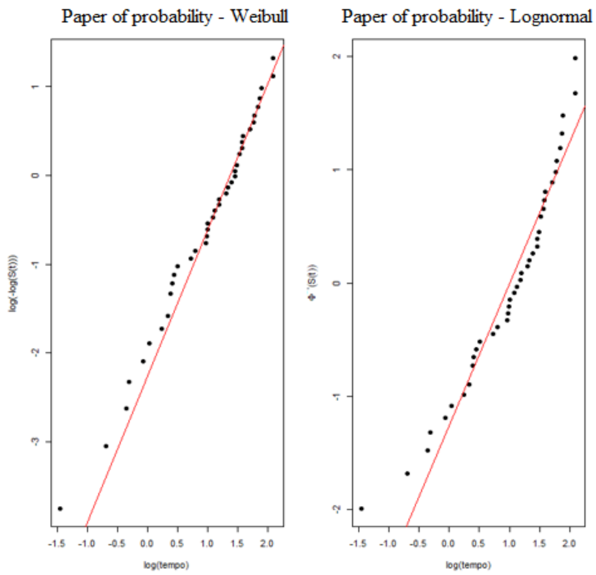


Fig. 4: Papers of the probability of the distributions Weibull and Lognormal for the failure time of the bottle inspectors.

Source: Authors.

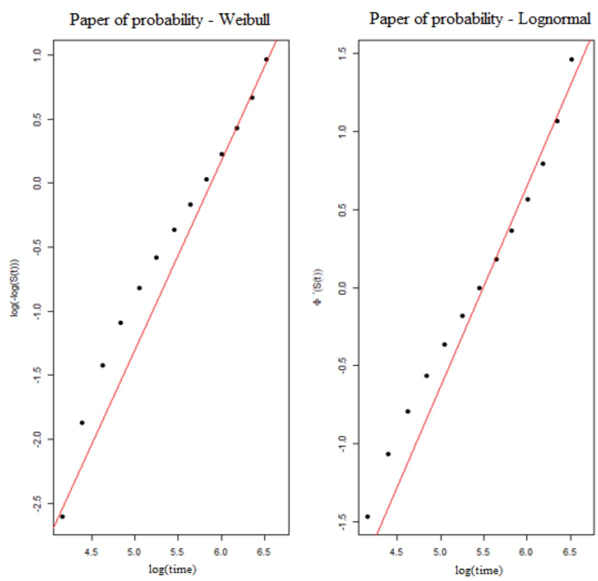
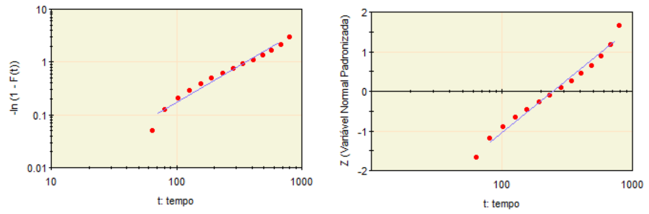


Fig. 5: Papers of the probability of Weibull and Lognormal distributions for the failure time of fiber optic transmitters.

Source: Authors.

The papers of probability of the Weibull and Lognormal distributions differ from ProConf, see Figure 6. The difference is found in the abscissa axis, the RELPF function uses the neperian logarithm of the time, while ProConf uses only time.

https://dx.doi.org/10.22161/ijaers.711.7 ISSN: 2349-6495(P) | 2456-1908(O)



a) Weibull b) Lognormal

Fig.6: Paper of probability of the Weibull and Lognormal distributions made by the Proconf software for fiber optic data.

Source: Authors.

The last tests for adequacy of the sample to the distributions are KS and χ2 tests, these are presented in two individual tables that show the value of the statistics of the test, beyond indicating if the hypothesis can or not be rejected. Tables 9 and 10 show the results of KS and X2 for data from electronic bottle supervisors. Figures 11 and 12 show for fiber optic data.

Table 9. Result of the χ2 test for the failure time of the electronic overseers of bottle.

Distribution	Statistics X2	p-value X2	Hypothesis
Lognormal	6.31313	0.17695	It cannot be rejected
Exponential	12.88477	0.01185	Rejected
Weibull	1.86555	0.76047	It cannot be rejected
Normal	4.22802	0.37603	It cannot be rejected
Gamma	-	-	-

Source: Authors.

Table10. Result of the KS test for the failure time of the electronic overseers of bottle.

Source: Authors.

Distribution	Statistics KS	D-critic	Hypothesis
Lognormal	0.13212	0.20282	It cannot be rejected
Exponential	0.17752	0.20282	It cannot be rejected
Weibull	0.08723	0.20282	It cannot be rejected
Normal	0.10787	0.20282	It cannot be rejected
Gamma	-	-	-

Table 11. Result of the χ2 test for the optical fiber failure time.

Distribution	Statistics X2	p-value X2	Hypothesis
Lognormal	0.14286	0.70543	It cannot be rejected
Exponential	0.14002	0.70826	It cannot be rejected
Weibull	0.39915	0.52753	It cannot be rejected
Normal	2.66742	1.10242	It cannot be rejected
Gamma	-	--	It cannot be rejected

Source: Authors.

Table 12. Result of the KS test for the optical fiber failure time.

Distribution	Statistics KS	D-critic	Hypothesis
Lognormal	0.10039	0.34878	It cannot be rejected
Exponential	0.18029	0.34878	It cannot be rejected
Weibull	0.10253	0.34878	It cannot be rejected
Normal	0.1535	0.34878	It cannot be rejected
Gamma	-	--	It cannot be rejected

Source: Authors.

The KS and χ2 tests have the same objective, however, the χ2 test is more sensitive when used with a larger number of data. This fact can be seen when comparing Table 9 and 10.

Regarding the results of the ProConf software, there are differences due to the way in which the tests are calculated, as there are steps in which the considerations vary from author to author. Tables 13 and 14 show the values of the KS and χ2 tests for optical fiber failure times. Note that the χ2 statistics and the p-value of the χ2 test differ from the results presented by RELPF because of the choices of the intervals for the calculation of the observed and waited for frequencies. Concerning KS test, note that the statistics are similar, however, ProConf performed the analysis using the p-value, while the RELPF function used the comparison of the statistics with the critical value.

Table 13. Values of the χ2 tests of the distributions using the ProConf software for fiber optic failure data.

Distribution	Statistic	p-value	Hypothesis
Lognormal	0,1004	0,3866	It cannot be rejected
Exponential	0,1803	0,1779	It cannot be rejected
Weibull	0,1006	0,386	It cannot be rejected
Normal	0,2078	0,1031	It cannot be rejected
Gamma	0,0878	0,4261	It cannot be rejected

Source: Authors.

Table 14. Values of the KS tests of the distributions using ProConf software for fiber optic failure data.

Distribution	Statistic	p-value	Hypothesis
Lognormal	0,22	0,6405	It cannot be rejected
Exponential	0,42	0,8093	It cannot be rejected
Weibull	0,16	0,6862	It cannot be rejected
Normal	2,13	0,1445	It cannot be rejected
Gamma	0,03	0,8559	It cannot be rejected

Source: Authors.

IV. CONCLUSION

In the reliability analysis, tools that help to determine the fit of the sample distributions are important for decision making, mainly when these choices are made in the industrial sector and have involved cost in the process. By means of the developed code, it was possible to present resulted trustworthy in the modeling of these distributions. The RELPF function showed similar results to ProConf (except for Gamma) and Shanker (2016). It obtained a better quality result for the Gamma distribution when compared to ProConf. Another advantage of the RELPF function was the implementation of more complex distributions, such as Generalized Gamma and Weibull of 3 parameters.

The implementation in the R software was done efficiently. The R allowed greater flexibility for the user to control the results and allowed a better analysis for the user, such as controlling the axes of the graphs and changing their scale. In addition, R is free and open source, extending the accessibility for users interested in performing equipment reliability analysis.

The next step is to optimize the RELPC function with the implementation of new distributions such as Extreme Value.

Beyond extracting more information of the Weibull distribution of 3 parameters. In order to make the RELPF function more complete and to obtain to analyze of efficient form other data that need different distributions to the implemented ones.

ACKNOWLEDGEMENTS

To the National Council for Scientific and Technological Development (CNPQ) for the scientific initiation scholarship.

REFERENCES

[1] CRESPO, A.A. Estatística Fácil. 17ª Edição. LTC Editora, 2005.
[2] HAVIARAS, G.J. Metodologia para análise de confiabilidade de pneus radiais em frota de caminhões de longa distância. Dissertação (Mestrado em Engenharia Automotiva), Escola Politécnica da Universidade de São Paulo, São Paulo, p.124 2005.
[3] WUTTKE, R.A.; SELLITTO, M.A. Cálculo da disponibilidade e da posição na curva da banheira de uma válvula de processo petroquímico. Revista Produção On Line, Santa Catarina, v. 8, n. 4, dez. 2008.
[4] FOGLIATO, F.S.; RIBEIRO, J.L.D. Confiabilidade e Manutenção Industrial. Elsevier, Rio de Janeiro, 2009.
[5] R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2020.
[6] FALCETTA, E.F. Análise da Confiabilidade de Produtos baseada em Dados de utilização da Garantia. 2000.45p. Dissertação (Mestrado profissionalizante em Engenharia),

https://dx.doi.org/10.22161/ijaers.711.7 *ISSN: 2349-6495(P) | 2456-1908(O)*

Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

[7] FRITSCH, C.; RIBEIRO, J. ProConf: Um software orientado para análises de confiabilidade. In. XVIII Encontro Nacional de Engenharia de Produção.Niterói, 1998.

[8] FERNANDES, F.S. Testes de Ajustes a Distribuição Estatísticas e Métodos para Estimação dos Parâmetros em Análise de Fiabilidade. Instituto Superior de Engenharia de Lisboa. 2013.

[9] FERREIRA, D.F. Estatística Básica. Ministério da Educação e do Desporto. Universidade Federal de Lavras. 1996.

[10] SILVA, E. M.; ANDRADE, P. C. Análise de confiabilidade de um inspetor eletrônico de garrafas. Revista da Universidade Vale do Rio Verde, v. 16, n. 2. 2018.

[11] SILVA, E. C.; FERREIRA; G. S.; SANTOS, M. M.; ANDRADE, P. C.; ROSA, R. M. Análise de dados de falha de um transmissor de fibra óptica. Revista THEMA, v. 8, n. 4. 2017.

[12] SHANKER, R. A Quase Sujatha Distribution. International Journal of Probability and Statistics, 5(4): 89-100. 2016.