Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/130128

Formulating an Efficient Statistical Test Using the Goodness of Fit Approach with Applications to Real-life Data

S. A. Qaid [1,2,], S. E. Abo Youssef [2] and M. M. M. Mansour[3]*

[1]Department of Mathematics, Faculty of Education, Abyan University, Abyan, Yemen
[2]Department of Mathematics, Faculty of Science, Al-Azhar University, Nasr City, Cairo, Egypt
[3]Department of Basic Science, Faculty of Engineering, The British University in Egypt, El Sherook City, Cairo, Egypt

Received: 28 Mar. 2023, Revised: 11 Jul. 2023, Accepted: 24 Jul. 2023.
Published online: 1 Jan. 2024.

Abstract: Statistical tests are very important for researchers to make decisions. In particular, when the tests are non-parametric, they are of greater importance because they can be applied to a wide range of data sets regardless of knowing the distribution of these data. Researchers are therefore racing to obtain efficient tests for making good decisions based on the results of these tests. In this study, NBU (2)L was used based on the goodness of fit approach to present an efficient statistical test. The efficiency of the proposed test was computed, and the results were compared to those of other tests. Critical values were computed and tabulated, and the power of this test was estimated. Finally, this test was applicable to both real, complete data and censored data.

Keywords: Aging, Statistical tests, Incomplete data, Exponential distribution, NBU (2)L.

1 Introduction

Recently, Reliability has become more crucial across every sector. especially in the field of industrial Engineering as it is defined as the ability of an operating system performs its mission successfully over a period of time. The single product now contains a large group of components that work as one system, which increases the possibility of its failure in the event of the failure of one of its components. Therefore, a product that does not fail to be used within the specified time period is considered reliable. Thus, how to measure, analyses, and evaluate the product's ability is the basis of the reliability theory. To read more about reliability, see Barlow and Proschan [1]. A product, system, component, or element's reliability may be assessed using statistical tests, and through these tests, we can search for faults and work to ensure that they do not occur during the specified time period for this product (see Abu- Youssef et al [2] and Hassan and Said [3]).

Exponential distribution has numerous uses outside industrial engineering, it is used to determine how far apart DNA mutations are from one another, see Duan et al [4]. Moreover, the exponential distribution is appreciable for figuring out how long it will take the radioactive particle to decay, see Poston [5]. Also, aids in determining the height of various molecules in a gas under steady conditions of pressure and temperature in a constant gravitational field, see Beckers et al [6]. Data resulting from regular rainfall and river discharge volumes can be modeled to the Exponential distribution, see Tomy et al [7].

The exponential test plays a vital role in reliability theory, and it was applied to different life distributions to represent phenomena that are not limited by the statute of limitations because it ensures that the longevity of the phenomenon is not affected by its previous duration. The exponential distribution has been tested against many other classes of life distribution; for example, Gadallah et al [8], EL-Sagheer et al [9] and Mansour [10] submitted a paper to examine the use of the idea of testing exponentiality in medical research.

For the exponential test against many life distributions and their applications in various fields of science such as medical, industrial, economic, life sciences, etc. (see Bakr et al [11], El-Morshedy et al [12]). Based on the goodness of fit methodology, Abu-Youssef and Silvana [13] developed a non-parametric test and applied it to real data. For testing against new better than used in second order (NBU (2)), see Kayid et al [14]. Mansour [15] developed a new method of exponential testing to evaluate the effectiveness of all different treatment modalities in all medical fields. These are some of the basic definitions that smoothed the way for deriving our class:

Definition 1 If and only if, a continuous random variable Y with a cumulative distribution function (CDF), F, meets the following criteria, it is said to be NBU:

$$\bar{F}(t+y) \leq \bar{F}(t)\bar{F}(y) \quad , \forall\, t, y > 0 \tag{1}$$

Definition 2 In the case of a continuous random variable Y with a CDF, F is said to be NBU (2) if:

$$V(t+y) - V(t) \leq \bar{F}(t)V(y) \quad , \forall\ t, y > 0 \tag{2}$$

*Corresponding author e-mail: salahabdoh767@gmail.com

where $V(y)=\int_0^y \bar{F}(u)du,\ \bar{F}(u)=1-F(u).$

The duration life of a new item is stochastically longer than that of a consumed one at age, t > 0, according to the definition provided by Kayid et al. [14]. A new class of life distribution, known as new better than used in second order and the Laplace formula (NBU (2)L), is shown in the definition below:

Definition 3 Y is said to be NBU (2)L if:

$$s^2\int_0^\infty e^{-sy}\bar{V}(t+y)dy \geq s\ \bar{V}(t)-\bar{F}(t)\left(1-\phi(s)\right), \forall\, y,t>0, s\geq 0. \tag{3}$$

Where, $\bar{V}(t)=\int_t^\infty \bar{F}(u)du$,$\phi(s)=E[e^{-sY}]=\int_0^\infty e^{-sy}dF(y)$.

In the current study, using goodness of fit, we tested the exponential against (NBU (2)L). Asymptotic features of our test are explored in Section 2 based on the U-statistic. Monte Carlo simulations of the critical points for null distributions with sample sizes of n=10,11,…,50,63 and estimations of the test's power are reported as well. The information in Section 3 has been right-censored. Finally, we explore a few examples in Section 4 to show how the suggested test may be used in survival analysis.

2 Testing exponentiality against NBU (2)L

In order to compare exponentiality to NBU (2)L in this part, a test statistic built on the goodness of fit approach is developed. The subsequent lemma is necessary:

Lemma 1. A random variable Y with CDF, F, and it is a member of the NBU (2)L class, then:

$$\phi(1)\left(1-\phi(s)\right)\leq s\ \phi(s)\left(1-\phi(1)\right),\quad s\geq 0\quad ,s\neq 1. \tag{4}$$

Where, $\phi(s)=Ee^{-sY}=\int_0^\infty e^{-sy}dF(y)$, $\phi(1)=Ee^{-Y}=\int_0^\infty e^{-y}dF(y)$.

Proof:

Since F is NBU (2)L then,

$$s^2\int_0^\infty e^{-sy}\bar{V}(t+y)dy \geq s\ \bar{V}(t)-\bar{F}(t)\left(1-\phi(s)\right), \forall\, y,t>0, s\geq 0.$$

where, $\bar{v}(y)=\int_y^\infty \bar{F}(u)\,du$

At this point, we multiply both sides by e^{-t} and implement the integral over $(0,\infty)$ with respect to t, we get:

$$s^2\int_0^\infty\int_0^\infty \bar{V}(t+y)e^{-t}\ e^{-sy}dy\ dt\ \geq\ s\int_0^\infty e^{-t}\bar{V}(t)dt\ -\left(1-\phi(s)\right)\int_0^\infty e^{-t}\bar{F}\ (t)dt \tag{5}$$

We put

$$I_1=\int_0^\infty e^{-t}\bar{F}(t)dt=E\int_0^\infty e^{-t}I(Y>t)dt$$

$$=E\int_0^y e^{-t}dt=(1-Ee^{-y}),$$

it's easy to show

$$I_1=(1-\phi(1)). \tag{6}$$

We put,

$$I_2=\int_0^\infty e^{-t}\bar{V}(t)dt=\int_0^\infty\int_t^\infty e^{-t}\bar{F}(u)du\ dt$$

$$=\int_0^\infty\int_0^t e^{-u}\bar{F}(t)dudt=\int_0^\infty(-(e^{-t}-1)\bar{F}(t)dt$$

$\Rightarrow \quad I_2 = -\left(1-\phi(1)\right)+\mu$ (7)

We put

$$I_3 = \int_0^\infty \int_0^\infty \bar{V}(t+y)e^{-t}\ e^{-sy}dy\ dt$$

I_3 can so be expressed as follows:

$$I_3 = \int_0^\infty \int_V^\infty \bar{V}(u)e^{-v}e^{-s(u-v)}dudv$$

$$= \int_0^\infty \int_0^v \bar{V}(v)e^{-u}\ e^{-sv+su}du\ dv$$

$$= \frac{-1}{1-s}\int_0^\infty e^{-v}\quad \bar{V}(v)dv\ + \frac{1}{1-s}\int_0^\infty e^{-sv}\ \bar{V}(v)dv$$

therefore

$$I_3 = \frac{1}{(1-s)}\left(1-\phi(1)\right) - \frac{1}{(1-s)}\mu - \frac{1}{s^2(1-s)}\left(1-\phi(s)\right) + \frac{\mu}{s(1-s)} \quad (8)$$

Substituting (6), (7) and (8) into (5), we get

$\phi(1)(1-\phi(s)) \le s\ \phi(s)(1-\phi(1))\ ,\quad s\ge 0\quad ,s\ne 1.$

This completes the proof.

2.1 Test procedures based on complete data

Let, $Y_1, Y_2, \ldots, Y_n$ a sample drawn at random from the distribution F. the issue is to verify:

H_0 : F is exponential against H_1: F is NBU (2)L and not exponential.

The criterion of deviation from H_0 in the return of H_1 can be considered to define the proposed test as follows:

$\zeta(s) = s\ \phi(s)(1-\phi(1)) - \ \phi(1)(1-\phi(s))$

Therefore, under H_0 it is $\zeta(s) = 0$, under H_1 it is $\zeta(s) > 0$.

Define the test statistic $\hat{\zeta}_n(s)$ as follows:

$$\hat{\zeta}_n(s) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[se^{-sy_i}(1-e^{-y_j}) - e^{-y_i}(1-e^{-sy_j})\right].$$

To ensure the invariant of the test, let

$$\boldsymbol{\Delta}_n(s) = \frac{\zeta_n(s)}{\mu^2},$$

which estimated by

$$\widehat{\boldsymbol{\Delta}}_n(s) = \frac{\hat{\zeta}_n(s)}{\bar{y}^2},$$

Where, $\bar{y} = \sum_{i=1}^{n}\frac{y_i}{n}$ is the sample mean.

Then,

$$\widehat{\Delta}_n(s) = \frac{1}{n^2\bar{y}^2}\sum_i \sum_J \phi(y_i, y_j), \quad (9)$$

where

$$\phi(y_i, y_j) = se^{-sy_i}(1-e^{-y_j}) - e^{-y_i}(1-e^{-sy_j}) \quad (10)$$

The test's asymptotic characteristics are enumerated in the subsequent theorem.

Theorem As $n \to \infty, \sqrt{n}\left(\hat{\Delta}_n(s) - \Delta(s)\right)$ is asymptotically normal with mean 0 and variance $\sigma^2(s)$ where $\sigma^2(s)$ is stated in (11). Under H_0 , the variance yields to (12).

Proof

One can note that using the conventional U-statistic theory (see Lee [16]),

$\sigma^2 = V\{E[\phi(y_1, y_2)|y_1] + E[\phi(y_1, y_2)|y_2]\}$.

Recall the definition of $\phi(y_i, y_j)$ in Equation (10), thus it is easy to show that:

$$E[\phi(y_1, y_2)|y_1] = se^{-sy} - se^{-sy}\int_0^\infty e^{-y}dF(y) - e^{-y} + e^{-y}\int_0^\infty e^{-sy}dF(y)$$

Similarly,

$$[\phi(y_1, y_2)|y_2] = s\int_0^\infty e^{-sy}dF(y) - se^{-y}\int_0^\infty e^{-sy}dF(y) - \int_0^\infty e^{-y}dF(y) + e^{-sy}\int_0^\infty e^{-y}dF(y).$$

Hence,

$$\sigma^2(s) = var\left\{s\ e^{-sy} - e^{-y} + (e^{-sy} - se^{-y} - 1)\int_0^\infty e^{-y}dF(y) + \ (e^{-y} - se^{-y} + s)\int_0^\infty e^{-sy}dF(y)\right\}\ . \quad (11)$$

Under H_0

$$\sigma_0^2(s) = \frac{s^2(s-1)^2(3s+2)}{12(s+1)^2(2s+1)(s+2)} \quad (12)$$

2.2 The Pitman asymptotic efficiency

To assess the effectiveness of this procedure, we compare the asymptotic Pitman efficiency (PAE) for this test with a number of other tests in Table 1 to see how effective this approach is based on the following three alternatives:

i) The Weibull distribution: $\bar{F}_1(y) = e^{-y^\theta}, y \geq 0, \theta \geq 1.$

ii) The linear failure rate distribution (LFR): $\bar{F}_2(y) = e^{-y-\frac{\theta}{2}y^2}, y \geq 0, \theta \geq 0.$

iii) The Makeham distribution: $\bar{F}_3(y) = e^{-y-\theta(y+e^{-y}-1)}, y \geq 0, \theta \geq 0.$

Note that for case (i), when θ=1 it reduces to the exponential distribution while cases (ii) and (iii) are shortened to the exponential case when θ = 0. The PAE is defined by:

$$PAE\ \ (\Delta_n(s)) = \frac{1}{\acute{o}_\circ(s)}\left|\frac{d}{d\theta}\zeta(s)\right|_{\theta\to\theta_\circ}.$$

$At\ \ s = 0.05\ \Rightarrow\ \sigma_\circ = 0.0127515$

This leads to:

$PAE\ [\Delta_n(0.05), Weibull] = 1.20314,\qquad PAE\ [\Delta_n(0.05), LFR] = 0.844686$

and $PAE\ [\Delta_n(0.05), Makeham] = 0.288429.$

Table 1: The proposed test's PAE in exchange for other tests' PAEs

Test	Wiebull	LFR	Makeham
Mahmoud and Abdul Alim [17]	0.405	0.433	0.289
Kaid [14]	2.3238	0.5809	0.2582
Walid et al [18]	1.046	0.932	0.233
Our test $\Delta_n(0.05)$	1.20314	0.844686	0.288429

Our test is clearly the most effective in most cases. It is noted that the proposed test $\Delta_n(0.05)$ has a high efficiency that was calculated with different values of s to obtain the best efficiency for all alternatives. Furthermore, notably at s = 0.05 its effectiveness was superior to other tests for all alternatives.

2.3 Monte Carlo null distribution critical points

Using 10000 samples drawn from the standard exponential distribution via the Mathematica 13 programme, this subsection simulates the critical points of the Monte Carlo null distribution for the test statistic $\Delta_n(0.05)$. Table 2 lists the Critical points as separators among the rejection and acceptance regions for $\Delta_n(0.05)$ at n = 10,11,…,50,63.

Table 2: Critical points as separators among the rejection and acceptance regions for $\Delta_n(0.05)$

n	**90%**	**95%**	**99%**	**n**	**90%**	**95%**	**99%**
10	0.00612915	0.00760218	0.00989814	31	0.00333719	0.00406059	0.00543005
11	0.00588476	0.00714011	0.00953379	32	0.00318551	0.00391056	0.00530333
12	0.00555664	0.00678451	0.00918894	33	0.0031573	0.00388968	0.00523328
13	0.00519294	0.00636804	0.00859580	34	0.00307764	0.00382248	0.0051996
14	0.00509508	0.00623100	0.00836150	35	0.00299854	0.003777753	0.00513871
15	0.00 494740	0.00599825	0 .00797861	36	0.00297054	0.00366989	0 .00499191
16	0 .00482736	0 .00585247	0 .00782165	37	0 .00295291	0 .00366769	0 .00501863
17	0.00444979	0.00553989	0.00740572	38	0.00285803	0.00352807	0.00498007
18	0.0044503	0.00548825	0.00719673	39	0.00285756	0.00353565	0.00475777
19	0.00432873	0.00520294	0.00716689	40	0.00282486	0.0034995	0.00476739
20	0.00419273	0.00510081	0.00684665	41	0.00271959	0.00343164	0.0046395
21	0.00403961	0.00494897	0.00667254	42	0.00275841	0.00346057	0.00468684
22	0.00390940	0.00481833	0.00649891	43	0.00274193	0.00339537	0.00465622
23	0.00382211	0.00471217	0.00624157	44	0.00267388	0.00333365	0.00450734
24	0.00393391	0.004586677	0.00614525	45	0.0026659	0.00329629	0.00450942
25	0 .00360499	0.00450507	0.00609004	46	0 .00265002	0.00328499	0.00444521
26	0 .00354129	0.00438897	0.00602348	47	0 .00254544	0.00316547	0.00424521
27	0.00353747	0.00434625	0.00578534	48	0.00256074	0.00319813	0.00436355
28	0.00346002	0.00427667	0.00565991	49	0.00257031	0.00315432	0.00425847
29	0.00433247	0.00422013	0.00580109	50	0.00250879	0.00311966	0.00426324
30	0.00329124	0.00410801	0.00544473	63	0.00223175	0.00278172	0.00371889

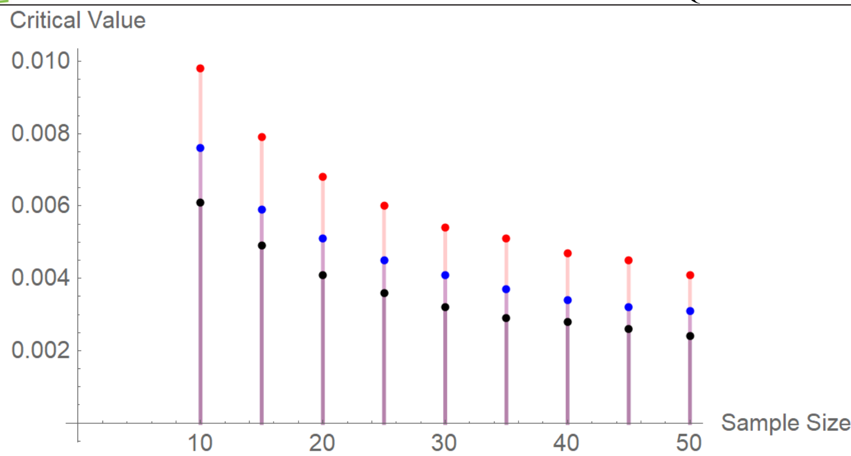

Fig. 1: Relation between critical values, sample size and confidence level.

Table 2 and Figure 1 make it evident that while the critical values are decreasing as sample numbers increase, they are increasing as confidence levels grow.

2.4 Power of the test

In this subsection, the power of our test $\Delta_n(0.05)$ will be calculated at a significance level $\alpha = 0.05$ based on 10000 samples generated from linear failure rate (LFR), Weibull, and gamma distributions. The power estimates with parameters $\theta = 2, 3$ and 4 for n = 10 , 20 , and 30 are shown in Table 3.

Table 3: The Power Estimates of the test $\Delta_n(0.05)$

n	θ	LFR	Weibull	Gamma
10	2	0.7418	0.6486	0.8803
	3	0.8715	0.9894	0.9814
	4	0.9159	0.9999	0.9973
20	2	0.9899	0.9834	0.9499
	3	0.9962	1.0000	0.9966
	4	0.9979	1.0000	1.0000
30	2	0.9998	0.9992	0.9753
	3	0.9999	1.0000	0.9994
	4	1.0000	1.0000	0.9998

It is evident that the sample size increases, the power estimates increase for each value of the parameter θ.

3 Test procedures using censored data

In this section, using data that has been randomly right censored, a test statistic is suggested to compare H_0 with H_1 . In a clinical trial or life test model where patients may be lost before the study is finished (censored), such censored data are typically the only information available. This experimental situation can be formulated as follows: suppose n objects to be tested, and let $X_1, X_2, \dots, X_n$ connote their actual lifetimes. Let $X_1, X_2, \dots, X_n$ be independently and identically distributed (i.i.d) according to the continuous lifetime distribution F. Let $Y_1, Y_2, \dots, Y_n$ be (i.i.d) according to a continuous life distribution G. Also, we assume that X's and Y's are independent. The pairs (ψ_j, δ_j), j = 1, 2, …, n, are observed in the randomly right-censored model, where $\psi_j = \min(\psi_j, \delta_j)$,

and , $$\delta_j = \begin{cases} 1, & \text{if } \psi_j = X_j \ (j-\text{thobservation is uncensored}) \\ 0, & \text{if } \psi_j = Y_j \ (j-\text{thobservation is censored}) \end{cases}$$

Let $\psi_0 = 0 < \psi_1 < \psi_2 < \dots < \psi_n$ connote the orderd ψ's and $\delta_{(j)}$ is δ_j corresponding to $\psi_{(j)}$.

Using the censored data (ψ_j, δ_j) , j = 1, 2, …, n. The following is how Kaplan and Meier [19] suggested the product limit estimator:

$\bar{F}(X) =_{[j:z_{(j)}\leq x]} \{(n-j) \ / \ (n-j+1)\}^{\delta_{(j)}} {}_{,X\in 0,Z_{(n)}]}$

Now, for testing $H_0 \ : \zeta_C(s) = 0$, against $H_1: \zeta_C(s) > 0$ based on the censored scenario, we recommend the following test statistic:

$$\hat{\zeta}_n(s) = s \ \phi_\theta(s)(1-\phi_\theta(1)) - \phi_\theta(1)(1-\phi_\theta(s)).$$

for computational purpose, $\hat{\zeta}_n(s)$ may be rewritten as

$$\hat{\zeta}_n(s) = s \ \eta(1-\tau) - \tau(1-\eta), \ where$$

$$\eta = e^{-s\psi_{(j)}}[\sum_{p=1}^{j-2} C_p^{\delta(p)} - \sum_{p=1}^{j-1} C_p^{\delta(p)}] \quad , \tau = \sum_{j=1}^{n} e^{-\psi_{(j)}}[\sum_{p=1}^{j-2} C_p^{\delta(p)} - \sum_{p=1}^{j-1} C_p^{\delta(p)}] \quad \text{and}$$

$$dF_n(\psi_j) = \bar{F}_n(\psi_{j-1}) - \bar{F}_n(\psi_j), \qquad C_k = [n-k][n-k+1]^{-1}$$

To ensure the invariant of the test, let

$$\hat{\Delta}_c(s) = \frac{\hat{\zeta}_c(s)}{\bar{\psi}^2}, \text{where } \bar{\psi} = \sum_{i=1}^{n} \frac{\psi_{(i)}}{n}.$$

3.1 Critical values of the Monte Carlo null distribution in the censored scenario

Using Mathematica 13 program, the Monte Carlo null distribution critical values of Δ_n at s = 0.05 for sample sizes n = 5, 10, …, 50, 60, 70, 80, and 81with 10000 replications are simulated from the standard exponential distribution. Table 4 reviews the upper percentile points of the statistic $\hat{\Delta}_n(0.05)$.

Table 4: The Upper critical values of $\Delta_c(0.05)$

n	90%	95%	99%
5	1.51731×10^8	2.29620×10^8	5.02146×10^8
10	1.22997×10^8	1.61204×10^8	2.90207×10^8
15	1.09461×10^8	1.35370×10^8	2.02127×10^8
20	1.02951×10^8	1.25058×10^8	1.78230×10^8
25	9.78957×10^7	1.16134×10^8	1.57300×10^8
30	9.52259×10^7	1.10444×10^8	1.46502×10^8
35	9.29083×10^7	1.04983×10^8	1.34740×10^8
40	8.97560×10^7	1.01687×10^8	1.29630×10^8
45	8.87678×10^7	1.00519×10^8	1.28349×10^8
50	8.66336×10^7	9.68911×10^7	1.19124×10^8
51	8.60768×10^7	9.55256×10^7	1.15565×10^8
60	8.43037×10^7	9.30341×10^7	1.13702×10^8
70	8.32422×10^7	9.18085×10^7	1.07699×10^8
80	8.12362×10^7	8.86478×10^7	1.03852×10^8
81	8.12972×10^7	8.87211×10^7	1.04341×10^8

According to Table 4, the key spots grow as confidence levels rise and shrink as sample sizes grow.

4 Real-Data Applications

In this section, our test is run with a 95% confidence level on a few real data sets.

4.1 Case of Complete Data

In this subsection, three examples are provided with the assumption that s = 0.05.

Example 1 Numbers in cells below indicate the cumulative information on how long it takes a patient to feel better after taking an analgesic (in minutes). The data was provided by Gross and Clark [20] and it contains the following twenty observations:

1.10	1.40	1.70	1.30	1.90
1.60	2.20	1.70	2.70	4.10
1.50	1.20	1.40	3.00	1.70
1.60	2.00	1.80	1.80	2.30

Since $\Delta_n(0.05)$ = 0.00596368 which is more than Table 2's critical value. The conclusion is that the NBU (2)L property applies to this data collection.

Example 2 Tensile strengths of 1.5 cm glass fibers were tested at the National Physical Laboratory in England, and those results are shown in the data. This data set has been discussed by Adepoju et al [21].

0.55	0.93	1.25	1.36	1.49	1.52	1.58	1.61	1.64
1.68	1.73	1.81	1.04	1.27	1.39	1.49	1.53	1.59
1.61	1.66	1.68	1.76	1.82	2.01	0.77	1.11	1.28
1.42	1.50	1.54	1.60	1.62	1.66	1.69	1.76	1.84
2.24	0.81	1.13	1.29	1.48	1.50	1.55	1.61	1.62
1.66	1.70	1.77	1.84	0.84	1.24	1.30	1.48	1.51
1.55	1.61	1.63	1.67	1.70	1.78	1.89	0.74	2.00

Since $\Delta_n(0.05)$ = 0.00814731 and this number surpasses Table 2's percentile. The conclusion is thus that the provided data set has the NBU (2)L property.

Example 3 Take into account the data from Abouammoh et al [22]. This information pertains to a group of 40 Saudi Arabian Ministry of Health hospitals patients who had (leukemia), a kind of blood cancer.

2.211	2.162	2.370	0.315	0.496	0.616	1.145	1.208
1.263	1.414	2.025	2.532	2.693	2.805	2.910	2.912
3.912	3.263	3.348	3.348	3.427	3.499	3.34	3.767
3.751	3.858	3.986	4.049	4.244	4.323	4.381	4.392
4.397	4.647	4.753	4.929	4.973	5.074	4.381	2.036

Since $\Delta_n(0.05)$ = 0.00232696 and this value is less than the critical value in Table 2. Then we accept the null hypothesis (exponential) and we reject the alternative hypothesis (NBU (2)L).

Case of Incomplete Data

In this subsection two examples are provided with the assumption that s = 0.05.

Example 1 Consider the data from Susarla and Vanryzin [23]. These numbers indicate 81 melanoma patient survival periods. Of these, 35 represent censored (incomplete) data and 46 represent entire life histories (complete data). These values were noticed:

13	14	19	19	20	21	23	23	25	26	26	27
27	31	32	34	34	37	38	38	40	46	50	53
54	57	58	59	60	65	65	66	70	85	90	98
102	103	110	118	124	130	136	138	141	234	16 +	21 +
44 +	50 +	55 +	67 +	73 +	76 +	80 +	81 +	86 +	93 +	100 +	108 +
114 +	120 +	124 +	125 +	129 +	130 +	132 +	134 +	140 +	147 +	148 +	151 +
152 +	152 +	158 +	181 +	190 +	193 +	194 +	213 +	215 +			

(+ indicates censored observations)

We get $\Delta_c(0.05) = 4.92514 \times 10^{-7}$ which is less than the critical value of Table 4. Then we reject the alternative hypothesis of NBU (2)L property and accept the null hypothesis of exponentiality property.

Example 2 The data shows the estimated death times (in weeks) for tongue cancer patients with aneuploidy DNA profiles. Previous users of the data include sickle-Santanello et al. [24]., Klein and Moeschaberger [25]. the 51 observations make up the data:

1	3	3	4	10	13	16	16	24
28	30	30	32	41	51	61+	65	67
72	73	74+	77	79+	80+	81+	87+	87+
89+	91	93+	96	97+	100	101+	104	104+
109+	120+	150+	131+	157	167	13	231+	240+
24	27	70	88+	108+	400			

(+ indicates censored observations)

We get $\Delta_c(0.05) = 7.81316 \times 10^{-7}$ which is below Table 4's critical threshold. The alternative hypothesis (NBU (2)L) is therefore rejected, and the exponentiality property null hypothesis is accepted.

Conclusion

An efficient non-parametric statistical test is introduced based on a developed class of life distributions, NBU (2)L, for testing the exponential property to a wide range of real data sets in different applied fields. The features of this test have been studied whether the PAE criterion or the power of the proposed test. Furthermore, a comparison study is introduced between the proposed test and other previous tests to emphasize the impact of the proposed test in the field of testing exponentiality. The test is applied to the cases of complete and censored data. As the introduced test is efficient, it is recommended to be applied in practical studies.

Conflict of interest

The authors declare that there is no conflict regarding the publication of this paper.

Reference

[1] R. E. Barlow and F. Proschan. Statistical theory of reliability and life testing: probability models to begin with, Florida State Univ Tallahassee, Silver Spring, MD, (1981).

[2] S. A. Abu-Youssef, N. S. A. Ali & A.A. El-Toony. Nonparametric Test for a class of Lifetime Distribution UBAC (2) Based on the Laplace Transform. Journal Egypt Math Soc 30: 9, 1-11 (2022).

[3] E. M. A. Hassan & M. M. Said .A New Class of Life Distributions Based on Moment Inequalities. Asian Journal of Probability and Statistics, 13(4), 47--57. (2021).

[4] Wen-Qi Duan, Zahid Khan, Muhammad Gulistan, & Adnan Khurshid. "Neutrosophic Exponential Distribution: Modeling and Applications for Complex Data Analysis", Complexity, vol. 2021, Article ID 5970613, 1-8, (2021). https://doi.org/10.1155/2021/5970613.

[5] J. W. Poston. Encyclopedia of Physical Science and Technology (Third Edition), & Academic Press. Pages 603-650, (2003). https://doi.org/10.1016/B0-12-227410-5/00185-X.

[6] J Beckers, W. W. Stoffels, T. Ockenga, M. Wolter & H. Kersten. Microparticles in a RF plasma under hyper gravity conditions. IEEE International Conference on Plasma Science - Abstracts, San Diego, CA, USA, pp. 1-1, (2009). https://doi: 10.1109/PLASMA.2009.5227229.

[7] L. Tomy , M. Jose & G. Veena . A review on recent generalizations of exponential distribution. Biom Biostat Int J. ;9(4):152-156. (2020). httb:// DOI: 10.15406/bbij.2020.09.00313

[8] A. M. Gadallah, B. I. Mohammed, Abdulhakim A. Al-Babtain, Saima Khan Khosa, Mutua Kilai, M. Yusuf, M. E. Bakr, "Modeling Various Survival Distributions using a Nonparametric Hypothesis Testing Based on Laplace Transform Approach with Some Real Applications", Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 5075716, 1-8, (2022). https://doi.org/10.1155/2022/5075716

[9] R. M. EL-Sagheer, M. A. W. Mahmoud & W. B. H. Etman. Characterizations and Testing Hypotheses for NBRUL-t_0 Class of Life Distributions. J Stat Theory Pract 16:31, 1-20, (2022).

[10] M. M. M. Mansour. Non-parametric statistical test for testing exponentiality with applications in medical research. Statistical Methods in Medical Research, 29(2):413-420, (2020).

[11] M. E. Bakr, A. M. Gadallah, A. Abdulhakim & Al-Babtain. Solving Fundamental Life Testing Issues: An Application in Medical Science. Research Square 23 February, PREPRINT (Version 1), 1-11, (2022). https://doi.org/10.21203/rs.3.rs-1376290/v1.

[12] El-Morshedy, Mahmoud, Afrah Al-Bossly, M. Rashad. EL-Sagheer, Bader Almohaimeed, B. H. Waleed. Etman, & S. E. Mohamed. "A Moment Inequality for the NBRULC Class: Statistical Properties with Applications to Model Asymmetric Data" Symmetry, vol. 14, no. 11: 2353, 1-13. (2022). https://doi.org/10.3390/sym14112353.

[13] S. E. Abu-Youssef & T. Silvana. Gerges. Based On the goodness of fit approach, a new test statistic for testing NBUCmgf class of life distributions. Pakistan Journal of statistics, vol. 38(1):129-144. (2022).

[14] M. Kayid , L. S. Diab & A. Alzughaibi .Testing NBU (2) class of life distribution based on goodness of fit approach. Journal of King Saud University – Science, vol. 22 (4):241-245. (2010).

[15] M. M. M. Mansour. Assessing Treatment Methods via Testing Exponential Property for Clinical Data. Journal of Statistics Applications & Probability, vol. 11, No. 1:109-113. (2021).

[16] A. J. Lee. U-Statistics Theory and Practice, Marcel Dekker, Inc, State New York, NY. (1990).

[17] M. A. W. Mahmoud & A. N. Abdul Alim. A goodness of fit approach for Testing NBUFR (NWUFR) and NBAFR (NWAFR) properties. International Journal on Reliability Applications, vol. 9:125--140. (2008).

[18] B. H. Walid. Etman, M. Rashad EL-Sagheer, Shaban E. Abu-Youssef & Amr Sadek . On Some Characterizations to NBRULC Class with Hypotheses Testing Application. Journal Aplied Mathematics & Information sciences, vol. 16, No. 2, 139-148. (2022).

[19] E. L. Kaplan & P. Meier. Nonparametric estimation from incomplete observation. Journal of the American Statistical Association, vol. 53, 457-481. (1958).

[20] A. J. Gross, & V. A . Clark. "Survival Distributions: Reliability Applications in the Biometrical sciences", John Wiley, New york, USA. (1975).

[21] K. A. Adepoju, A. U. Chukwu & M. wang. The Beta Power Exponential Distribution. Journal of Statistical Science and Application, vol. 2,37-46. .(2014).

[22] A. M. Abouammoh, S. A. Abdulghani & I. S. Qamber. On partial orderings and testing of new better than used classes. Reliability engineering & system Safety, vol.43,37-41. (1994)

[23] V. Susarla & J. Vanryzin. Empirical Bayes Estimations of A Survival Function Right Censored Observation. The Annals of statistics, vol. 6 No 4, 710-755. (1978).

[24] B. J. Sickle-Santanello, W. B. Farrar, J. K. DeCenzo, S. Keyhani-Rofagha, J. Klein, D. Pearl, H. Laufman & R. V. O. Toole. Technical and statistical improvements for flow cytometric DNA analysis of paraffin-embedded tissue. Cytometry, vol. 9(6), 594–599. (1988). https://doi.org/10.1002/cyto.990090613

[25] J. P. Klein & M. L. Moeschberger. Survival Analysis: Techniques for Censored and Truncated Data. 2nd Edition, Springer, New York, USA. (2003).

.