An Estimation of Survival and Hazard Rate Functions of Exponential Rayleigh Distribution

Lamyaa Khalid Hussein
Department of Mathematics,
College of Science,
Mustansiriyah University
lamyaakhalid8242@gmail.com

Iden Hasan Hussein
Department of Mathematics,
College of Science for Women,
University of Baghdad,Iraq.
Idenalkanani58@gmail.com

Huda Abdulla Rasheed
Department of Mathematics,
College of Science,
Mustansiriyah University

iraqalnoor1@gmail.com

Abstract

In this paper, we used the maximum likelihood estimation method to find the estimation values for survival and hazard rate functions of the Exponential Rayleigh distribution based on a sample of the real data for lung cancer and stomach cancer obtained from the Iraqi Ministry of Health and Environment, Department of Medical City, Tumor Teaching Hospital, depending on patients' diagnosis records and number of days the patient remains in the hospital until his death.

Keyword: Exponential Rayleigh *ER* distribution; Maximum Likelihood Estimation; Chi Square Test; Real - Life Data Application.

1.Introduction

Statistical tests are essential for scientific research. Some of these tests are called parametric tests and others are called non-parametric tests. These tests help the researcher analyze the data accurately and thus are useful in summarizing the results an essentially and appropriately way and drawing general conclusions. In parametric tests, some assumptions are based the suitability of data for a normal distribution, while non-parametric tests do not require that your data follow the normal distribution.

Non-parametric tests include various tests. In this study we will use the chi-square test which is a non-parametric test and it represents the oldest known goodness of fit test introduced by Karl Pearson in (1900). This test is used to how likely the observed

distribution of data fits with the distribution that is expected (i.e., to test the goodness of fit), it is used to analyze categorical data.

The goodness of fit test is used to check whether a given sample of data follows from a proposed distribution. The formula for calculating a chi-square statistic is [4,6]:

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

Where

k represents the number of classes.

O_i represents the observed frequency in class i

E_i represents the expected frequency in class i

The characteristics of the chi-square test is that it can be easily calculated and applied to both continuous and discrete variables. It is not recommended for small sample size (less than 25). Also the asymptotic distribution of the test statistic is the chi-square distribution with (k-1) degrees of freedom. The null and alternative hypotheses are as follows:

H_0: The failure time data is distributed as Exponential Rayleigh p.d.f.

H_1: The failure time data is not distributed as Exponential Rayleigh p.d.f.

The null hypothesis is rejected when comparing the tabulated value at level of significance (α) and degree of freedom ($k - 1$) is less than the calculated value $\chi^2 > \chi^2_{\alpha,(k-1)}$.

The exponential distribution is the most commonly used distribution for lifetime data analysis. Its simplicity and mathematical feasibility made it the most widely used lifetime model in reliability (survival) theory. It is commonly used to model the time until something occurs in the process.

A continuous non-negative random variable Z is called to have an Exponential distribution with parameter α, if its probability density function is given by [3]:

$$f(z; \alpha)_E = \alpha\, e^{-\alpha z} \qquad\qquad ; z \geq 0;\ \alpha > 0 \qquad\qquad \dots(1)$$

And zero otherwise, where α is the scale parameter.
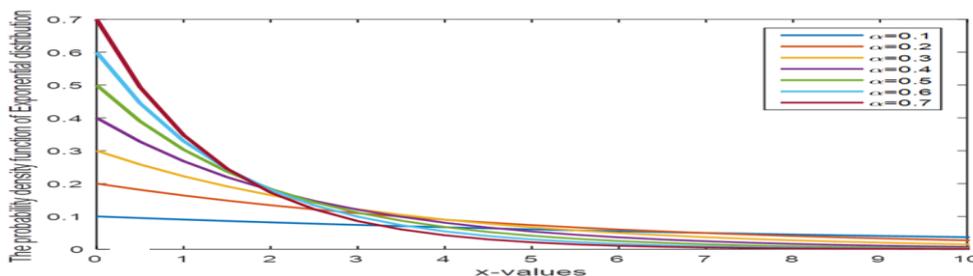

Figure 1: probability density function of Exponential distribution for different values of ($\alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The cumulative distribution function is:

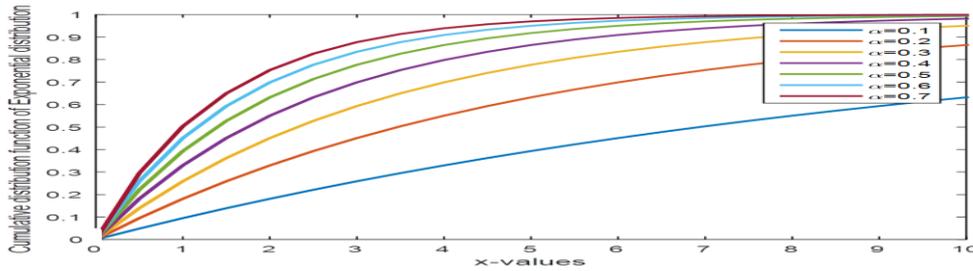$$F(z; \alpha)_E = 1 - e^{-\alpha z} \qquad ; z \geq 0; \ \alpha > 0 \qquad (2)$$

Figure 2: The cumulative distribution function of Exponential distribution for different values of ($\alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

Rayleigh distribution with one parameter is one of the most widely used distributions. It is an essential distribution in statistics and operation research. In (1880) Lord Rayleigh presented the Rayleigh distribution [5]. It shows the main role in connection with a problem in various branches modeling and analyzing lifetime data for instance project effort loading modeling, communication, survival and reliability theory, physical sciences, technology, diagnostic imaging, clinical subjects, and applied statistics [7].

A continuous non-negative random variable Y is called to have a Rayleigh distribution with parameter λ, if its probability density function is given by [2]:

$$f(y; \lambda)_R = \lambda y \, e^{-\frac{\lambda}{2} y^2} \qquad ; y \geq 0; \ \lambda > 0 \qquad (3)$$

And zero otherwise, where λ is the scale parameter.
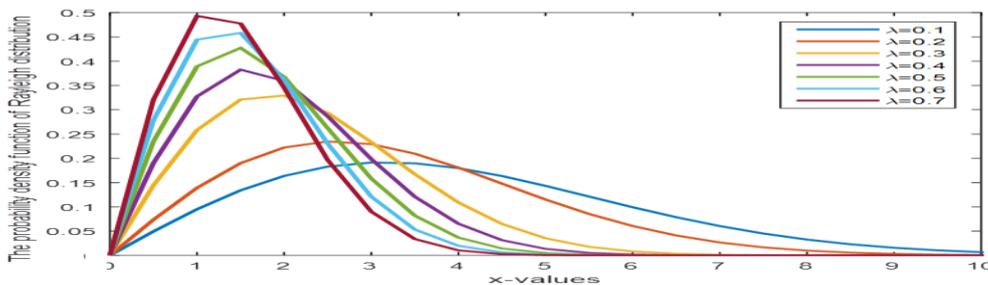

Figure 3: The probability density function of Rayleigh distribution for different values of ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The cumulative distribution function is:

$$F(y; \lambda)_R = 1 - e^{-\frac{\lambda}{2} y^2} \qquad ; y \geq 0; \ \lambda > 0 \qquad (4)$$
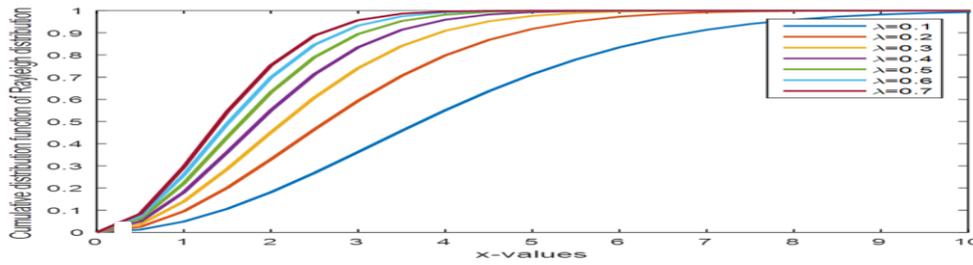
Figure 4: The cumulative distribution function of Rayleigh distribution for different values of ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The Exponential Rayleigh distribution is obtained based on mixed between cumulative distribution function of Exponential distribution in equation (2) and cumulative distribution function of Rayleigh distributions in equation (4) as follows [1]:

Let $X = max(Z, Y)$ where Z and Y are two independent random variables then:

$$F(x; \alpha, \lambda)_{ER} = pr(X \leq x)$$

$$F(x; \alpha, \lambda)_{ER} = 1 - pr(X > x)$$

$$F(x; \alpha, \lambda)_{ER} = 1 - pr(max(Z, Y) > x)$$

$$F(x; \alpha, \lambda)_{ER} = 1 - [pr(Z > x). pr(Y > x)]$$

$$F(x; \alpha, \lambda)_{ER} = 1 - [(\int_{x}^{\infty} f(z; \alpha) dz). (\int_{x}^{\infty} f(y; \lambda) dy)]$$

$$F(x; \alpha, \lambda)_{ER} = 1 - [(\int_{x}^{\infty} \alpha e^{-\alpha z} dz). (\int_{x}^{\infty} \lambda y e^{-\frac{\lambda}{2} y^2} dy)]$$

$$F(x; \alpha, \lambda)_{ER} = 1 - e^{-(\alpha x + \frac{\lambda}{2} x^2)} \qquad ; x \geq 0; \alpha, \lambda > 0 \qquad (5)$$

Figure 5: The cumulative distribution function of ER distribution for $\alpha = 0.5$ and different values of ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The probability density function is given by [1]:

$$f(x; \alpha, \lambda)_{ER} = (\alpha + \lambda x) e^{-(\alpha x + \frac{\lambda}{2} x^2)} \qquad ; x \geq 0; \alpha, \lambda > 0 \qquad (6)$$

And zero otherwise, where α and λ are scale parameters.

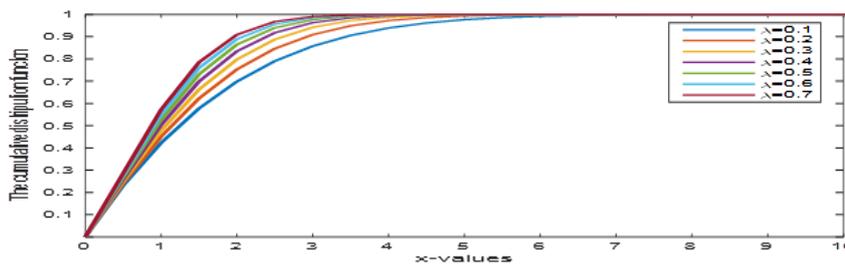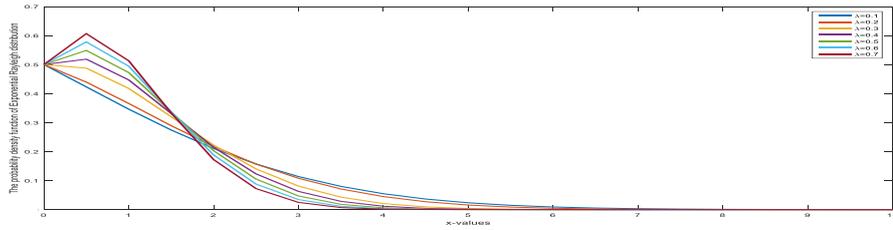**Figure 6:** probability density function of ER distribution for $\alpha = 0.5$ and different values of ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The Survival function can be expressed by [1]:

$$S(t; \alpha, \lambda)_{ER} = e^{-\left(\alpha t + \frac{\lambda}{2} t^2\right)} \qquad t \geq 0; \alpha, \lambda > 0 \tag{7}$$
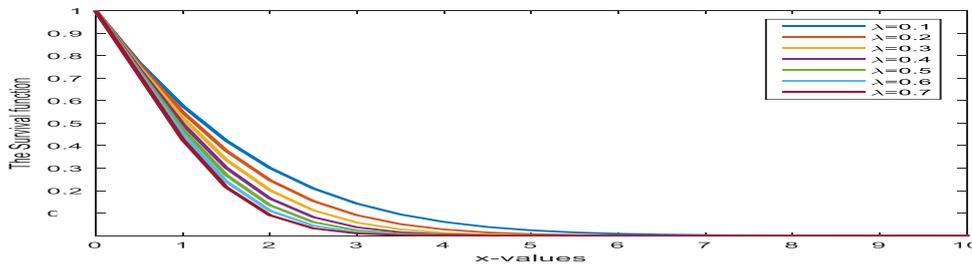

Figure 7: Survival function of ER distribution for $\alpha = 0.5$ and different values of ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The Hazard rate function is obtained by [1]:

$$h(t; \alpha, \lambda)_{ER} = \alpha + \lambda t \qquad t \geq 0; \alpha, \lambda > 0 \tag{8}$$
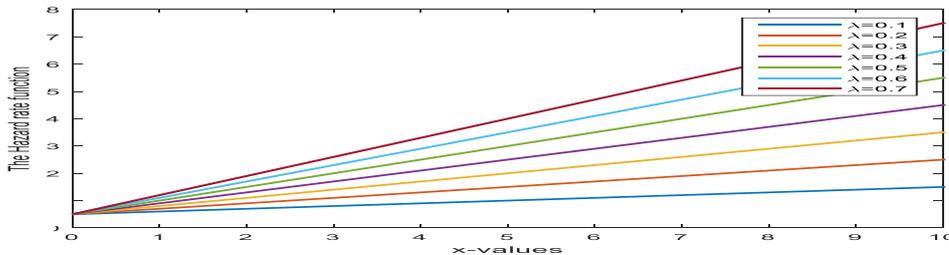
Figure 8: Hazard rate function of ER distribution for $\alpha = 0.5$ and different values of ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$) [Matlab R2015a].

The rth moment about the origin can be expressed by [1]:

$$E(X^r)_{ER} = \sum_{n=0}^{\infty} \frac{(-\alpha)^n}{n!} \frac{2^{\frac{r+n}{2}}}{\lambda^{\frac{r+n}{2}}} \left[\frac{\alpha}{\sqrt{2\lambda}} \Gamma\left(\frac{r+n+1}{2}\right) + \Gamma\left(\frac{r+n+2}{2}\right) \right] \quad ; r = 1,2,3, \tag{9}$$

The mean: The first moment which is named the mean is obtain by put $r = 1$ in equation (9) thus [1]:

$$E(X)_{ER} = \sum_{n=0}^{\infty} \frac{(-\alpha)^n}{n!} \frac{2^{\frac{1+n}{2}}}{\lambda^{\frac{1+n}{2}}} \left[\frac{\alpha}{\sqrt{2\lambda}} \Gamma\left(\frac{n+2}{2}\right) + \Gamma\left(\frac{n+3}{2}\right) \right] \tag{10}$$

The variance: The general formula $v(X)$ of ER distribution is given by [1]:

$$v(X)_{ER} = E(X^2)_{ER} - [E(X)_{ER}]^2$$

$$= \left[\sum_{n=0}^{\infty} \frac{(-\alpha)^n}{n!} \frac{2^{\frac{2+n}{2}}}{\lambda^{\frac{2+n}{2}}} \left[\frac{\alpha}{\sqrt{2\lambda}} \Gamma\left(\frac{n+3}{2}\right) + \Gamma\left(\frac{n+4}{2}\right) \right] \right] - \left[\sum_{n=0}^{\infty} \frac{(-\alpha)^n}{n!} \frac{2^{\frac{1+n}{2}}}{\lambda^{\frac{1+n}{2}}} \left[\frac{\alpha}{\sqrt{2\lambda}} \Gamma\left(\frac{n+2}{2}\right) + \right. \right.$$

$$\left. \left. \Gamma\left(\frac{n+3}{2}\right) \right] \right]^2 \tag{11}$$

The moment generating function of ER distribution can be obtained by [1]:

$$M_X(t)_{ER} = \sum_{n=0}^{\infty} \frac{(-(\alpha-t))^n}{n!} \frac{2^{\frac{n}{2}}}{\lambda^{\frac{n}{2}}} \left[\frac{\alpha}{\sqrt{2\lambda}} \Gamma\left(\frac{n+1}{2}\right) + \Gamma\left(\frac{n+2}{2}\right) \right] \tag{12}$$

2. Maximum Likelihood Estimation

Let $\underline{x} = (x_1, x_2, \dots, x_n)$ be a random sample of size (n) drawn from ER distribution with pdf given by equation (6). The complete data likelihood function $L(\alpha, \lambda | \underline{x})_{ER}^{MLE}$ for a given random sample can be expressed by,

$$L(\alpha, \lambda | \underline{x})_{ER}^{MLE} = \prod_{i=1}^{n} f(x_i; \alpha, \lambda)_{ER} = \prod_{i=1}^{n} \left[(\alpha + \lambda x_i) e^{-(\alpha x_i + \frac{\lambda}{2} x_i^2)} \right] \tag{13}$$

The natural log – likelihood function is:

$$\ell_{ER}^{MLE} = \ln L(\alpha, \lambda | \underline{x})_{ER}^{MLE} = \sum_{i=1}^{n} \ln(\alpha + \lambda x_i) - \sum_{i=1}^{n} (\alpha x_i + \frac{\lambda}{2} x_i^2) \tag{14}$$

We derive the natural log – likelihood function partially with respect to α and λ respectively and setting it equal to zero yields,

$$\frac{\partial \ell_{ER}^{MLE}}{\partial \alpha} = \sum_{i=1}^{n} \frac{1}{\alpha + \lambda x_i} - \sum_{i=1}^{n} x_i = 0 \tag{15}$$

$$\frac{\partial \ell_{ER}^{MLE}}{\partial \lambda} = \sum_{i=1}^{n} \frac{x_i}{\alpha + \lambda x_i} - \frac{1}{2} \sum_{i=1}^{n} x_i^2 = 0 \tag{16}$$

The maximum likelihood estimators denoted by $\hat{\alpha}_{MLE(ER)}$ and $\hat{\lambda}_{MLE(ER)}$ are the values of α and λ that maximizes $L(\alpha, \lambda | \underline{x})_{ER}^{MLE}$ can be obtained by the solution of equations (15), (16). Note that there are no closed solutions of these equations; therefore, Newton – Raphson method is iterative technique can be applied to find the solution.

In Newton – Raphson method, the solution of the likelihood equation at iteration $(h + 1)$ is extract through the following iterative process,

$$\begin{bmatrix} \hat{\alpha}_{MLE(ER)}^{(h+1)} \\ \hat{\lambda}_{MLE(ER)}^{(h+1)} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{MLE(ER)}^{(h)} \\ \hat{\lambda}_{MLE(ER)}^{(h)} \end{bmatrix} - J_{(h)}^{-1} \begin{bmatrix} \dfrac{\partial \ell_{ER}^{MLE}}{\partial \alpha} \\ \dfrac{\partial \ell_{ER}^{MLE}}{\partial \lambda} \end{bmatrix}_{\substack{\alpha = \hat{\alpha}_{MLE(ER)}^{(h)} \\ \lambda = \hat{\lambda}_{MLE(ER)}^{(h)}}} \qquad ; h = 0,1,2,\dots$$

Where,

$$J_{(h)} = \begin{bmatrix} \dfrac{\partial^2 \ell_{ER}^{MLE}}{\partial \alpha^2} & \dfrac{\partial^2 \ell_{ER}^{MLE}}{\partial \alpha \partial \lambda} \\ \dfrac{\partial^2 \ell_{ER}^{MLE}}{\partial \lambda \partial \alpha} & \dfrac{\partial^2 \ell_{ER}^{MLE}}{\partial \lambda^2} \end{bmatrix}_{\substack{\alpha = \hat{\alpha}_{MLE(ER)}^{(h)} \\ \lambda = \hat{\lambda}_{MLE(ER)}^{(h)}}}$$

Where the first partial derivatives as in equations (15), (16) and the second partial derivatives are obtained as follows,

$$\frac{\partial^2 \ell_{ER}^{MLE}}{\partial \alpha^2} = - \sum_{i=1}^{n} \frac{1}{(\alpha + \lambda x_i)^2} \qquad (17)$$

$$\frac{\partial^2 \ell_{ER}^{MLE}}{\partial \lambda^2} = - \sum_{i=1}^{n} \frac{x_i^2}{(\alpha + \lambda x_i)^2} \qquad (18)$$

$$\frac{\partial^2 \ell_{ER}^{MLE}}{\partial \alpha \partial \lambda} = \frac{\partial^2 \ell_{ER}^{MLE}}{\partial \lambda \partial \alpha} = - \sum_{i=1}^{n} \frac{x_i}{(\alpha + \lambda x_i)^2} \qquad (19)$$

Now, based on an invariant property of the MLE estimator, the survival function at mission time (t) of the ER distribution can be obtained by replacing α and λ in equation (7), by their MLE estimators as follows:

$$\hat{S}(t; \alpha, \lambda)_{ER}^{MLE} = e^{-(\hat{\alpha}_{MLE(ER)} t + \frac{\hat{\lambda}_{MLE(ER)}}{2} t^2)} \qquad (20)$$

3. Real-Life Data Applications

3.1 Practical Application (1)

In this section, real data for lung cancer disease is analyzed, because of the importance of this disease, we have collected data related to mortality from this disease from the Iraqi Ministry of Health and Environment, Department of Medical City, Tumor Teaching Hospital, from 1 / 1 / 2015 to 1 / 1 / 2021. The data related to this disease were not taken during 2021 due to the spread of the Covid-19 epidemic, as patients do not stay in the hospital for more than a day or two for fear of contracting this epidemic. The sample size consists (100) observations. It was noted that all patients died during different periods and this means that the data or sample used is a complete data set.

The following real data set represents the number of days of entering a patient with lung cancer to the hospital until death. The data set consist (100) observations:

[4 28 3 6 2 6 4 21 5 10 5 8 3 2 5 10 4 2 7 13 6 2 11 5 7 10 5 21 12 3 13
12 2 10 4 3   22 2 7 20 18 4 10 5 16 21 4 26 12 15 3 5 25 3 15 20 13 6 22
27 13 6 8 6 7 5 11 6 17 5 4 3 21 18 9 13 3 18 24 18 4 10 20 7 12 17 19 4 6
15 15 4 6 6 12 11 3 10 2 29].

Using the maximum likelihood estimation in practical application to calculate the estimated
value for the two parameters α and λ as follows;

Number of Observation	100
Initial Value of α	0.01
Initial Value of λ	0.01
Estimated Value of α	0.0412
Estimated Value of λ	0.0074

The following graphic shows the frequency histogram table for lung cancer patients, the
vertical axis represents the number of patients and the horizontal axis represents the number
of days from the patient's admission to the hospital until his death
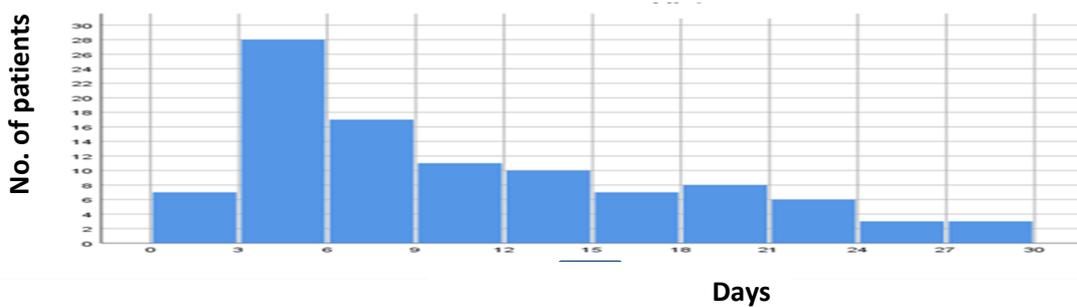
Figure 9: Frequency Histogram Table for Lung Cancer Patients

Now, we'll compute the expected frequency E_i as follows;

$$E_1 = F(3) \times 100$$

$$= \left[1 - e^{-[0.0412(3) + \frac{0.0074\,(9)}{2}]}\right] \times 100$$

$$E_1 = 14.6$$

$$E_2 = [F(6) - F(3)] \times 100$$

$$= \left[[1 - e^{-[0.0412(6) + \frac{0.0074\,(36)}{2}]}] - [1 - e^{-[0.0412(3) + \frac{0.0074\,(9)}{2}]}]\right] \times 100$$

$$E_2 = 17.1$$

$$E_3 = [F(9) - F(6)] \times 100$$

$$= \left[[1 - e^{-[0.0412(9) + \frac{0.0074\,(81)}{2}]}] - [1 - e^{-[0.0412(6) + \frac{0.0074\,(36)}{2}]}]\right] \times 100$$

$$E_3 = 17.2$$

$E_4 = [F(12) - F(9)] \times 100$

$$= \left[[1 - e^{-[0.0412(12) + \frac{0.0074 \, (144)}{2}]}] - [1 - e^{-[0.0412(9) + \frac{0.0074 \, (81)}{2}]}] \right] \times 100$$

$E_4 = 15.3$

$E_5 = [F(15) - F(12)] \times 100$

$$= \left[[1 - e^{-[0.0412(15) + \frac{0.0074 \, (225)}{2}]}] - [1 - e^{-[0.0412(12) + \frac{0.0074 \, (144)}{2}]}] \right] \times 100$$

$E_5 = 12.4$

$E_6 = [F(18) - F(15)] \times 100$

$$= \left[[1 - e^{-[0.0412(18) + \frac{0.0074 \, (324)}{2}]}] - [1 - e^{-[0.0412(15) + \frac{0.0074 \, (225)}{2}]}] \right] \times 100$$

$E_6 = 9.1$

$E_7 = [F(21) - F(18)] \times 100$

$$= \left[[1 - e^{-[0.0412(21) + \frac{0.0074 \, (441)}{2}]}] - [1 - e^{-[0.0412(18) + \frac{0.0074 \, (324)}{2}]}] \right] \times 100$$

$E_7 = 6.1$

$E_8 = [F(24) - F(21)] \times 100$

$$= \left[[1 - e^{-[0.0412(24) + \frac{0.0074 \, (576)}{2}]}] - [1 - e^{-[0.0412(21) + \frac{0.0074 \, (441)}{2}]}] \right] \times 100$$

$E_8 = 3.8$

$E_9 = [F(27) - F(24)] \times 100$

$$= \left[[1 - e^{-[0.0412(27) + \frac{0.0074 \, (729)}{2}]}] - [1 - e^{-[0.0412(24) + \frac{0.0074 \, (576)}{2}]}] \right] \times 100$$

$E_9 = 2.2$

$E_{10} = [F(30) - F(27)] \times 100$

$$= \left[[1 - e^{-[0.0412(30) + \frac{0.0074 \, (900)}{2}]}] - [1 - e^{-[0.0412(27) + \frac{0.0074 \, (729)}{2}]}] \right] \times 100$$

$E_{10} = 1.2$

Now, we'll compute

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

$$\chi^2 = \frac{(7-14.6)^2}{14.6} + \frac{(28-17.1)^2}{17.1} + \frac{(17-17.2)^2}{17.2} + \frac{(11-15.3)^2}{15.3} + \frac{(10-12.4)^2}{12.4}$$

$$+ \frac{(7-9.1)^2}{9.1} + \frac{(8-6.1)^2}{6.1} + \frac{(6-3.8)^2}{3.8} + \frac{(3-2.2)^2}{2.2} + \frac{(3-1.2)^2}{1.2}$$

$$\chi^2 = 17.915$$

It is discovered that the calculated value is (17.915); when comparing this value with tabulated value at the level of significance (0.01) and degrees of freedom (9) we find out that the calculated value is less than the tabulated value (21.67). That means accepting the null hypothesis H_0 and the data is distributed according to Exponential Rayleigh ER distribution.

Now, using estimate values for two parameters in Exponential Rayleigh distribution by MLE method to find numerical values for probability density function $\hat{f}(t)$, cumulative distribution function $\hat{F}(t)$, survival function $\hat{S}(t)$ and hazard rate function $\hat{h}(t)$ as follows;

Table 1 : Estimated values for functions $\hat{f}(t)$, $\widehat{F}(t)$, $\hat{S}(t)$ and $\hat{h}(t)$

t	f_i	$\hat{f}(t)$	$\widehat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t)$
2	7	0.0507	0.093	0.907	0.056
3	9	0.054	0.146	0.854	0.0634
4	10	0.056	0.201	0.799	0.0708
5	9	0.057	0.259	0.741	0.0782
6	10	0.0584	0.317	0.683	0.0856
7	5	0.0581	0.375	0.625	0.093
8	2	0.056	0.433	0.567	0.1004
9	1	0.055	0.489	0.511	0.1078
10	7	0.052	0.543	0.457	0.1152
11	3	0.049	0.594	0.406	0.1226
12	5	0.046	0.642	0.358	0.13
13	5	0.043	0.687	0.313	0.1374
15	4	0.035	0.766	0.234	0.1522
16	1	0.031	0.8	0.2	0.1596
17	2	0.028	0.83	0.17	0.167
18	4	0.024	0.857	0.143	0.1744
19	1	0.021	0.88	0.12	0.1818
20	3	0.018	0.901	0.099	0.1892
21	4	0.016	0.918	0.082	0.1966
22	2	0.013	0.933	0.067	0.204
24	1	0.0096	0.956	0.044	0.2188
25	1	0.0079	0.965	0.035	0.2262
26	1	0.0065	0.972	0.028	0.2336
27	1	0.0053	0.978	0.022	0.241
28	1	0.0042	0.983	0.017	0.2484
29	1	0.0033	0.987	0.013	0.2558

Here, we will discuss the following important notes on the previous results table (1):

1. The values of probability density function are increasing until $t = 6$. Then the probability density function are decreasing when the failure times ($7 \leq t \leq 29$), so $t = 6$ is the mode of this function. Noting that the differences between all the values of probability density function are very small and converged.

2. The cumulative distribution function values are increasing with the increase of failure times for the lung cancer patients in the hospital. This means there is a direct relationship between failure times and cumulative distribution function.

3. The values of the probability of survival for patients are great at small failure times and vice versa. There is a reverse relationship between failure times and survival function. This means the values of the survival function $\hat{S}(t)$ are decreasing with the increasing of the failure times. So the patient who remains $(t = 2)$ days has greatest probability of survival with (0.907), but the patient who remains $(t = 29)$ days in a hospital has the smallest probability of survival with (0.013).

4. The values of hazard function $\hat{h}(t)$ are increasing with the increasing of the failure times for the lung cancer patients in the hospital. This means there is a direct relationship between the failure times and hazard function. Noting that the probability of hazard for patients is small when small failure times and vice versa. The patient who remains $(t = 2)$ days in a hospital has a smallest probability of hazard for death with (0.056), while the patient who remains $(t = 29)$ days in a hospital has a greatest probability of hazard for death with (0.2558).

3.2 Practical Application (2)

Stomach cancer is an abnormal growth of cells that begins in the stomach. Because of the importance of this disease, we have collected data related to mortality from this disease from the Iraqi Ministry of Health and Environment, Department of Medical City, Tumor Teaching Hospital, from 1 / 1 / 2015 to 1 / 1 / 2021.The data related to this disease were not taken during the year 2021 due to the spread of the Covid-19 epidemic, as patients do not stay in the hospital for more than a day or two for fear of contracting this epidemic. The sample size consists (81) observations. It was noted that all patients died during different periods and this means that the data or sample used is a complete data set.

The following real data set represents the number of days from entering a patient with stomach cancer to the hospital until death. The data set consist (81) observations:

[6, 11, 27, 5, 2, 16, 8, 25, 7, 26, 14, 8, 19, 5, 24, 22, 16, 20, 17, 11, 3, 15, 2, 6, 15, 29, 3, 24, 15, 7, 3, 13, 19, 6, 4, 2, 11, 16, 27, 25, 15, 12, 7, 26, 9, 4, 2, 16, 22, 17, 6, 24, 23, 11, 15, 20, 6, 4, 22, 10, 8, 4, 13, 9, 7, 2, 5, 26, 14, 19, 25, 22, 4, 9, 3, 8, 13, 27, 25, 16, 8].

Using the maximum likelihood estimation in practical application to calculate the estimated value for the two parameters α and λ as follows;

Number of Observation	81
Initial Value of α	0.01
Initial Value of λ	0.01
Estimated Value of α	0.0182
Estimated Value of λ	0.0063

The following graphic shows the frequency histogram table for stomach cancer patients, the vertical axis represents the number of patients and the horizontal axis represents the number of days from the patient's admission to the hospital until his death.
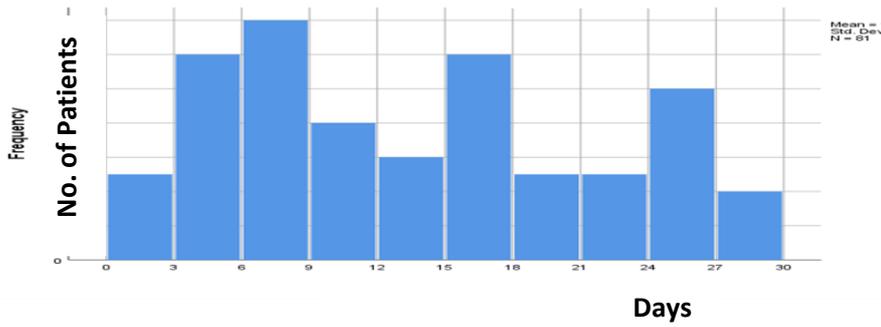
Figure 10: Frequency Histogram Table for Stomach Cancer Patients

Now, we'll compute the expected frequency E_i as follows;

$E_1 = F(3) \times 81$

$= \left[1 - e^{-[0.0182(3) + \frac{0.0063(9)}{2}]}\right] \times 81$

$E_1 = 6.48$

$E_2 = [F(6) - F(3)] \times 81$

$= \left[[1 - e^{-[0.0182(6) + \frac{0.0063(36)}{2}]}] - [1 - e^{-[0.0182(3) + \frac{0.0063(9)}{2}]}]\right] \times 81$

$E_2 = 9.72$

$E_3 = [F(9) - F(6)] \times 81$

$= \left[[1 - e^{-[0.0182(9) + \frac{0.0063(81)}{2}]}] - [1 - e^{-[0.0182(6) + \frac{0.0063(36)}{2}]}]\right] \times 81$

$E_3 = 11.583$

$E_4 = [F(12) - F(9)] \times 81$

$= \left[[1 - e^{-[0.0182(21) + \frac{0.0063(144)}{2}]}] - [1 - e^{-[0.0182(9) + \frac{0.0063(81)}{2}]}]\right] \times 81$

$E_4 = 11.907$

$E_5 = [F(15) - F(12)] \times 81$

$= \left[[1 - e^{-[0.0182(15) + \frac{0.0063(225)}{2}]}] - [1 - e^{-[0.0182(12) + \frac{0.0063(144)}{2}]}]\right] \times 81$

$E_5 = 11.016$

$E_6 = [F(18) - F(15)] \times 81$

$= \left[[1 - e^{-[0.0182(18) + \frac{0.0063(324)}{2}]}] - [1 - e^{-[0.0182(15) + \frac{0.0063(225)}{2}]}]\right] \times 81$

$E_6 = 9.315$

$E_7 = [F(21) - F(18)] \times 81$

$= \left[[1 - e^{-[0.0182(21) + \frac{0.0063(441)}{2}]}] - [1 - e^{-[0.0182(18) + \frac{0.0063(324)}{2}]}]\right] \times 81$

$E_7 = 7.209$

$E_8 = [F(24) - F(21)] \times 81$

$= \left[[1 - e^{-[0.0182(24) + \frac{0.0063(576)}{2}]}] - [1 - e^{-[0.0182(21) + \frac{0.0063(441)}{2}]}]\right] \times 81$

$E_8 = 5.265$

$E_9 = [F(27) - F(24)] \times 81$

$= \left[[1 - e^{-[0.0182(27) + \frac{0.0063(729)}{2}]}] - [1 - e^{-[0.0182(24) + \frac{0.0063(576)}{2}]}]\right] \times 81$

$E_9 = 3.564$

$E_{10} = [F(30) - F(27)] \times 81$

$$= \left[\left[1 - e^{-[0.0182(30) + \frac{0.0063(900)}{2}]} \right] - \left[1 - e^{-[0.0182(27) + \frac{0.0063(729)}{2}]} \right] \right] \times 81$$

$E_{10} = 2.187$

Now, we'll compute

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

$$\chi^2 = \frac{(5 - 6.48)^2}{6.48} + \frac{(12 - 9.72)^2}{9.72} + \frac{(14 - 11.583)^2}{11.583} + \frac{(8 - 11.907)^2}{11.907} + \frac{(6 - 11.016)^2}{11.016}$$

$$+ \frac{(12 - 9.315)^2}{9.315} + \frac{(5 - 7.209)^2}{7.209} + \frac{(5 - 5.265)^2}{5.265} + \frac{(10 - 3.564)^2}{3.564} + \frac{(4 - 2.187)^2}{2.187}$$

$\chi^2 = 19.526$

It is discovered that the calculated value is (19.526), when comparing this value with the tabulated value at the level of significance (0.01) and degrees of freedom (9), we find out that the calculated value is less than the tabulated value (21.67). That means accepting the null hypothesis H_0 and the data is distributed according to Exponential Rayleigh ER distribution .

Now, using estimate values for two parameters in Exponential Rayleigh distribution by MLE method to find numerical values for probability death density function $\hat{f}(t)$, cumulative distribution function $\hat{F}(t)$, survival function $\hat{S}(t)$ and hazard rate function $\hat{h}(t)$

Table 2 : Estimated values for functions $\hat{f}(t)$, $\widehat{F}(t)$, $\hat{S}(t)$ and $\hat{h}(t)$

t	f_i	$\hat{f}(t)$	$\widehat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t)$
2	5	0.0293	0.0479	0.9521	0.0308
3	4	0.03414	0.0797	0.9203	0.0371
4	5	0.0383	0.116	0.884	0.0434
5	3	0.0419	0.1562	0.8438	0.0497
6	5	0.0448	0.1996	0.8004	0.056
7	4	0.0469	0.2456	0.7544	0.0623
8	5	0.0484	0.2934	0.7066	0.0686
9	3	0.0492	0.3423	0.6577	0.0749
10	1	0.0493	0.3917	0.6083	0.0812
11	4	0.0489	0.4409	0.5591	0.0875
12	1	0.0478	0.4894	0.5106	0.0938
13	3	0.0463	0.5366	0.4634	0.1001
14	2	0.0444	0.5819	0.4180	0.1064
15	5	0.0422	0.6254	0.3746	0.1127
16	5	0.0396	0.6664	0.3336	0.119
17	2	0.0370	0.7047	0.2953	0.1253
19	3	0.0312	0.7731	0.2269	0.1379
20	2	0.0284	0.8029	0.1971	0.1442
22	4	0.0228	0.8542	0.1458	0.1568
23	1	0.0202	0.8757	0.1243	0.1631
24	3	0.0178	0.8948	0.1052	0.1694
25	4	0.0155	0.9115	0.0885	0.1757
26	3	0.0134	0.9259	0.0740	0.182
27	3	0.0115	0.9385	0.0615	0.1883
29	1	0.0083	0.9583	0.0417	0.2009

Here we will discussion the following important notes on the previous results table (2):

1. The values of probability density function are increasing until $t = 10$. Then the probability density function are decreasing when the failure times ($11 \leq t \leq 29$), so $t = 10$ is the mode of this function. Noting that the differences between all the values of probability density function are very small and converged.
2. The cumulative distribution function values are increasing with the increase of failure times for the stomach cancer patients in the hospital. This means there is a direct relationship between failure times and cumulative distribution function.
3. The values of the probability of survival for patients are great at small failure times and vice versa. There is a reverse relationship between failure times and survival function. This means the values of the survival function $\hat{S}(t)$ are decreasing with the increasing of the failure times. So the patient who remains ($t = 2$) days has greatest probability of survival with (0.9521), but the patient who remains ($t = 29$) days in a hospital has the smallest probability of survival with (0.0417).
4. The values of hazard function $\hat{h}(t)$ are increasing with the increasing of the failure times for the stomach cancer patients in the hospital. This means there is a direct relationship between the failure times and hazard function. Noting that the probability of hazard for patients is short when small failure times and vice versa, The patient who remains ($t = 2$) days in a hospital has a smallest probability of hazard for death with (0.0308), while

the patient who remains $(t = 29)$ days in a hospital has a greatest probability of hazard for death with (0.2009).

4. Conclusion

Based on real data of lung cancer and stomach cancer the differences between all the values of probability density function $\hat{f}(t)$ are very small and converged. The values of cumulative distribution function $\hat{F}(t)$ are increasing with the increasing of failure times for the patients in the hospital. The values of the probability of survival for patients was great at small failure times and vice versa. There is a reverse relationship between failure times and survival function $\hat{S}(t)$. The values of hazard function $\hat{h}(t)$ are increasing with the increasing of the failure times for the patients in the hospital, that means there is a direct relationship between the failure times and hazard function.

References

1. Hussein L. K., Hussein I. H. and Rasheed H. A.; A Class of Exponential Rayleigh Distribution and New Modified Weighted Exponential Rayleigh Distribution with Statistical Properties, *Journal of Physics conference series*. **2021**.

2. Merovci, F. and Elbatal, I.; Weibull Rayleigh distribution: *Theory and applications, Applied Mathematics and Information Sciences.* **2015**, *9, 4*, 2127 – 2137.

3. Oguntunde, P. E.; Balogun, O. S.; Okagbue, H. I.; Bishop, S. A.; The Weibull-Exponential Distribution: Its Properties and Applications, *Journal of Applied Sciences*. **2015**, *15, 11*, 1305 – 1311.

4. Rana, R. ; Singhal, R.; Chi-square Test and its Application in Hypothesis Testing, *Journal of the Practice of Cardiovascular Sciences*. **2015**, *1, Issue 1*, 69 – 71.

5. Rayleigh, J. W. S.; On the resultant of a large number of vibrations of the some pitch and of arbitrary phase, *Philosophical Magazine Series 5*. **1880**, *10, 60*, 73 – 78.

6. Turhan, N. S.; Karl Pearson's chi-square tests, *Educational Research and Reviews*. **2020**, *15, 9*, 575 – 580.

7. Tomy, L.; Gillariose, J.; A generalized Rayleigh distribution and its application, Biom. Biostat. Int. J. **2019**, *8,4,* 139 – 143.