*Journal of Statistics & Management Systems*
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 8, pp. 1691–1700
DOI : 10.47974/JSMS-1268

# Using the Bayesian method to estimate the exponential survival regression model for stomach cancer patients in Iraq

Ahmed Salam Mezher *
Wadhah S. Ibrahim †
*Department of Statistics*
*College of Administration and Economics*
*Al-Mustansiriya University*
*Baghdad 10046*
*Iraq*

**Abstract**

In this paper, we will rely on a probability distribution, which is the exponential distribution, to build a parametric survival regression model, which is the exponential survival regression model, relying on the Cox regression model to be used in forming this model. The data used in this research were obtained through the Cancer Council in Iraq, which is affiliated with the Ministry of Health. The data was represented by a figure of patients with stomach cancer for a period of two years (2021-2022), who were registered from the onset of symptoms to hospitalization and then death or recovery., and they numbered 200 sick, the data represents the length of stay of patients in the hospital in months. This aims to study the effect of some explanatory variables on the length of survival of stomach cancer patients. The model was used to analyze the data and estimate the parameters of this model using the Bayesian method, and using goodness-of-fit tests provided by the statistical program. The ready-made Easy Fit 5.6 test is the Kolmokrov-Smirnov test, the Anderson-Darlink test, and the Chi-Squared test. It turns out that the exponential survival regression model is the most appropriate model for the data of this research, use a program (R 4.3.1), one of the most important conclusions reached was the great convergence between the exponential survival regression model for the estimated values Use a Bayesian method and the exponential survival regression model for the original values. This indicates the effectiveness of the Bayesian method in estimating parameters and its ability to approach the real data. Accurately based on the available data, and when the estimated values of the model are close to the true values of the model, this means that the Bayesian method that

* *E-mail:* **ahmed.salam@uomustansiriyah.edu.iq (Corresponding Author)**
† *E-mail:* **dr_wadhah_stat@uomustansiriyah.edu.iq**

was used for estimation is highly accurate and that the model that was used for estimation has estimated it well and reflects the facts accurately.

***Subject Classification:*** *62C10, 62F15.*

***Keywords:*** *Cox regression model, Exponential distribution, Bayesian method, Goodness of fit test.*

## 1. Introduction

Survival regression models are statistical tools used to study the persistence of a particular event, such as death, recovery, or a medical event, by analyzing the period of Time over which it lasts. It is one of the important branches of statistics, and it has two names. The first is Survival, which means the study of the Time for living organisms that extends from the beginning of infection with a disease or taking treatment until the occurrence of the event that represents (death or recovery). The other name is Reliability, which means the study of Time for machines and machines. It is defined as a set of statistical procedures for analyzing data in which the dependent variable is Time, and Time may be years, months, weeks, or days from the beginning of follow-up of an individual until the occurrence of the event, and the meaning of the event is death, illness, etc., as survival analysis is concerned with A group of words that have a known event, usually death, which occurs after a period of Time, knowing that the event occurs only once for each word. [4].

The study of survival regression models has become of great importance in recent years, especially in the fields of scientific research and statistical analysis. This interest is attributed to the main role that these models play in understanding and analyzing the Time required until a specific event occurs, whether that is in the field of medicine, biological sciences, or other fields, the importance of these studies is clearly demonstrated in analyzing the survival of organisms after a specific period of Time (t), These models play a vital role in identifying risk factors in various fields such as medicine and bio industry. Given this increasing interest, studies of survival regression models have become taught courses, at various levels of education, this science has become independent as it plays an essential role in analyzing events, estimation, forecasting, and improvement. These studies are considered a powerful tool for researchers and analysts in understanding how different variables affect survival over Time, which contributes to the development and improvement of many Scientific and industrial fields [8].

## 2. Cox Regression Model

It is one of the proportional hazards models [2], [4].

$$\frac{h_i(t)}{h_0(t)} = \exp(\eta_i),\ \eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} \tag{1}$$

Where p represents the number of explanatory variables, $x_{ji}$ represents the value of the ith observation from the jth-sequenced explanatory variable, $\beta_j$ is the regression parameter for the explanatory variable, and $h_0(t)$ represents the hazard function of the probability distribution. It is noted that the Cox model does not contain $\beta_0$, and this is due to This is because $h_0(t)$ replaces it.

## 3. Exponential distribution

One of the advantages of this distribution is that its hazard rate is a fixed amount, and this indicates that the causes of failure are not due to obsolescence, but rather due to random accidents. The parametric exponential regression model is one of the easiest parametric survival models, because it assumes that the survival time follows an exponential distribution [1]. The probability density function (p.d.f) for this model is: [2].

$$f_i(t) = \lambda \exp(-\lambda t_i) \qquad t_i > 0,\ \ \lambda > 0 \tag{2}$$

Its survival function $S_i(t)$ is: [6].

$$S_i(t) = e^{-\lambda t_i},\ \ h_i(t) = \lambda,\ \ H_i(t) = \lambda t_i,\ F_i(t) = 1 - e^{-\lambda t_i} \tag{3}$$

$h_i(t)$: The Hazard Function, $H_i(t)$: The Hazard Function Cumulative, $F_i(t)$: The Cumulative Density Function [5].

The survival function can be written based on the proportional hazard model if it is assumed that the survival time follows an exponential distribution in the following form: [9] [10].

$$S_i(t) = (S_0(t))^{\exp(\eta_i)} \rightarrow S_i(t) = (\exp(-\lambda t_i))^{\exp(\eta_i)} \tag{4}$$

Generating data for an exponential survival regression model Using the inverse transformation method as follows: [3].

$$F_i(t_i) = U = 1 - U = (\exp(-\lambda t_i))^{\exp(\eta_i)} \tag{5}$$

Since: U and (1-U) follow the uniform distribution (0, 1) then:

$$U = (\exp(-\lambda t_i))^{\exp(\eta_i)} \tag{6}$$

Taking the natural logarithm of both sides and doing some mathematics:

$$t_i = \frac{-\ln(U)}{\lambda \exp(\eta_i)} \tag{7}$$

## 4. Estimating the parameters of the exponential survival regression model using the Bayesian method

We assume a probability distribution (The Prior distribution) for the parameter $(\lambda_i)$ that has the following formulas: [11].

$$\lambda_i \sim N(\mu_\lambda, \sigma_\lambda^2) \qquad i = 1, \ldots, n \tag{8}$$

Since the parameter $(\sigma_\lambda^2)$ is a variance of the parameter $(\lambda_i)$, according to Jeffery's rule, that it has an inverse chi-square distribution, that is: [7].

$$P\left(\frac{\sigma_\lambda^2}{y\lambda}\right) \propto (\sigma_\lambda^2)^{-\left(\frac{y_\lambda}{2}+1\right)} \exp\left(\frac{-y_\lambda}{2\sigma_\lambda^2}\right) \tag{9}$$

The joint posterior distribution of the parameter $(\lambda_i)$ is determined as follows: [11].

$$\int_0^\infty (\sigma_\lambda^2)^{-(n+y_\lambda+1)/2} \exp\left[\frac{-[y\lambda + \Sigma_i(\lambda_i - \bar{\lambda})^2]}{2\sigma_\lambda^2}\right] d\sigma_\lambda^2 \propto [y_\lambda + \Sigma_i(\lambda_i - \bar{\lambda})^2]^{-(n+y_\lambda-1)/2} \tag{10}$$

Using the equation, the maximum likelihood function for the exponential survival regression model [12]:

$$P(\lambda_i / t) = \lambda_i^{\,n} \exp\left(\sum_{i=1}^{n} \beta' X\right) \prod_{i=1}^{n} (\exp(-\lambda_i t_i))^{\exp(\beta' X)} * [y_\lambda + \sum_i (\lambda_i - \bar{\lambda})^2]^{-(n+y_\lambda-1)/2} \tag{11}$$

Integration of the joint posterior distribution above is inappropriate despite making it a normal distribution. By taking the natural logarithm and then finding the derivative and setting it equal to zero, the formula will be as follows: [11].

$$f(\lambda_i) = \frac{n}{\widehat{\lambda}_i} - \sum_{i=1}^{n} \exp(\beta' X)(t_i) + (\widehat{\lambda}_i - \widehat{\bar{\lambda}}) / \vartheta_{\lambda_i} \tag{12}$$

$$w(\beta) = \sum_{i=1}^{n} X - \lambda_i \sum_{i=1}^{n} \exp(\hat{\beta}' X) X(t_i) \tag{13}$$

The first derivative of functions (12) and (13) is as follows:

$$f'(\lambda_i) = -\frac{n}{\widehat{\lambda}_i^{\,2}} + \left[ \vartheta_{\lambda_i} \left( 1 - \frac{1}{n} \right) - 2(\widehat{\lambda}_i - \widehat{\widehat{\lambda}})^2 / (n + y_\lambda - 1) \right] / (\vartheta_{\lambda_i})^2 \tag{14}$$

$$w'(\beta) = -\lambda_i \sum_{i=1}^{n} \exp(\hat{\beta}' X) X^2(t_i) \tag{15}$$

We use the Newton-Raphson method to find the estimators as follows:

$$\hat{\lambda}_i^{t+1} = \hat{\lambda}_i^t - \frac{f(\hat{\lambda}_i^t)}{f'(\hat{\lambda}_i^t)}, \ \hat{\beta}_i^{t+1} = \hat{\beta}_i^t - \frac{z(\hat{\beta}_i^t)}{z'(\hat{\beta}_i^t)} \tag{16}$$

## 5. Practical side

Data on stomach cancer were obtained from the Cancer Council of Iraq, which is affiliated with the Iraqi Ministry of Health. The data includes information on 200 patients with stomach cancer during the period from 2021 to 2022. These patients were recorded from the onset of symptoms to hospitalization, and then death or recovery. The analysis focuses on the length of stay of patients in the hospital, which was measured in months from the date of their admission to the date of death. Or recovery. The aim of this research is to study the effect of some explanatory variables on this period, these variables can include several aspects, such as treatment received, biological factors, and other health conditions that may affect the course of the disease. Directing research toward understanding how these variables affect the length of hospital stay can provide important insights into improving the care of stomach cancer patients. Statistical analysis of this data will be useful to identify influencing factors and identify contexts in which health care and treatment for these patients can be improved.

### 5.1 *Description of variables*

The research variables were determined by the factors that have a fundamental role in determining the severity of the disease, which are as follows:

t: represents the length of stay in the hospital in months.

X1: represents age in months.

X2: represents sex and includes: a) Male. (1), b) Female (2).

X3: Occupation includes: a) Housewife (1), b) Office Gang (4), c) Retired. (7), d) Operator (Manual). (2), e) Earner (Business). (5), f) Others. (8), g) Unknown (9)

X4: represents the area in which Adders Code resides: a) Baghdad (Baghdad=1),b) Northern Provinces: (Erbil = 2, Sulaymaniyah = 3, Kirkuk = 4, Nineveh = 5, Dohuk = 6).

c) Southern Provinces: (Babylon=7, Wasit=8, Karbala=9, Najaf=10, Basra=11, Dhi Qar=12, Maysan=13, Muthanna=14, Diwaniya=15), d) Middle Provinces: (Diyala=16, Anbar=17, Salah al-Din=18), e) Unknown = 19 (by which we mean the governorates that were not mentioned).

X5: represents the grade of the tumor (Grad), divided into: a) Grade 1 tumor (1), b) Grad II tumor. (2), c) Grade III tumor. (3), d) Grade IV tumor (4), e) More Than Grad IV. (9)

X6: Chemotherapy includes: a) Useful success. (1), b) Unhelpful. (2), c) He did not undergo chemotherapy (Unknown (9)).

X7: Radiotherapy includes: a) Useful success. (1), b) Unhelpful. (2), c) He did not undergo radiotherapy (Unknown (9)).

X8: represents the behavior of the tumor, including: a) Malignant. (3), b) Hamid (Benign). (0), c) In its location (1), d) Uncertain (Uncourt Bergman (2)).

X9: represents the extent of the tumor (Extent) and includes: a) In the original location (In situ). (1), b) Regional Direct Extent. (2), c) Local extension of the lymph nodes (Regional (3)), d) Distant Metastasis (7), e) Not Applicable. (8), f) Unknown. (9).

### 5.2 *Goodness-of-fit tests*

In order to ensure that the data of the dependent variable follow the exponential survival regression model, goodness-of-fit tests provided by the ready-made statistical program Easy Fit 5.6, namely the Kolmokrov-Smirnov (KS) test, the Anderson-Darlink (AD) test, and the Chi-square test ($\chi$^2) were used to test all From the following:

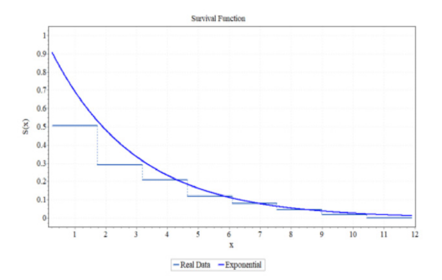
$$H_{0(Exp)}:\ Y \sim \text{Exponential survival regression model}$$

**Table 1**

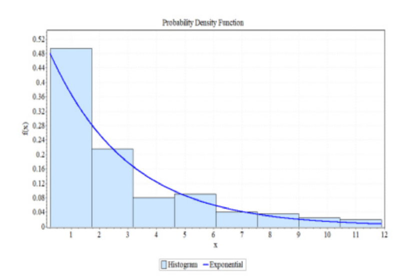
**Goodness-of-fit tests**

| | KS | AD | $\chi^2$ |
|---|---|---|---|
| **Statistic** | 0.0926 | 2.2361 | 10.37 |
| **Critical Value** | 0.09603 | 2.5018 | 14.067 |
| **Reject $H_0$ at 0.05?** | **No** | **No** | **No** |

It is clear from Table 1 that the calculated value for the three tests is less than the tabulated value at the 0.05 level of significance, and thus the null hypothesis is not rejected in the sense that the data for the t variable follow the exponential survival regression model, and the following figure shows the matching of the probability density function and the survival function for this model:




Survival function for exponential survival regression model




P.d.f. for the exponential survival regression model

**Figure 1**

**Data matching to an exponential survival regression model**

5.3 *Model estimation*

The Bayesian method was used to estimate the survival function according to the exponential survival regression model, and the regression parameters were estimated as shown in table 2 as follow:

**Table 2**

**Estimates of the parameters**

| **Parameters** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\beta_8$ | $\beta_9$ | $\lambda$ |
|---|---|---|---|---|---|---|---|---|---|---|
| **Exponential** | -0.00202 | -0.4598 | -0.051 | 0.0475 | -0.028 | -0.08 | 0.0807 | -0.075 | 0.1814 | 1.134 |

Based on these estimates, the survival function was estimated according to the exponential survival regression model, as follows:

$$S(t_i) = (\exp(-1.13414 t_i))^{\exp(\hat{\eta}_i)} \tag{17}$$

Since:

$$\begin{aligned}\hat{\eta}_i &= \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{\beta}_5 X_{5i} + \hat{\beta}_6 X_{6i} + \hat{\beta}_7 X_{7i} + \hat{\beta}_8 X_{8i} + \hat{\beta}_9 X_{9i} \\ &= -0.00202 X_{1i} - 0.4598 X_{2i} - 0.051 X_{3i} + 0.0475 X_{4i} - 0.028 X_{5i} - 0.08 X_{6i} \\ &+ 0.0807 X_{7i} - 0.075 X_{8i} + 0.1814 X_{9i}\end{aligned}$$

Based on this equation, the survival function was estimated and its results were as in Table 3 and Figure 1:

**Table 3**

**Estimated values of the exponential survival regression model**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.98909 | 0.80368 | 0.71907 | 0.64647 | 0.51981 | 0.44784 | 0.29826 | 0.09984 |
| 0.97754 | 0.80327 | 0.71880 | 0.64428 | 0.51611 | 0.44677 | 0.29197 | 0.09704 |
| 0.93364 | 0.79882 | 0.71182 | 0.64188 | 0.51428 | 0.44271 | 0.28948 | 0.09538 |
| 0.92862 | 0.79819 | 0.70876 | 0.63657 | 0.51125 | 0.44042 | 0.28790 | 0.07396 |
| 0.92144 | 0.79362 | 0.70806 | 0.63345 | 0.50608 | 0.42824 | 0.26278 | 0.06273 |
| 0.91981 | 0.79087 | 0.69795 | 0.62715 | 0.50318 | 0.42762 | 0.25777 | 0.04548 |
| 0.90858 | 0.78822 | 0.69567 | 0.61671 | 0.50217 | 0.42456 | 0.25445 | 0.04120 |
| 0.88840 | 0.78728 | 0.69254 | 0.60938 | 0.50179 | 0.40461 | 0.24782 | 0.03729 |
| 0.87965 | 0.77906 | 0.69231 | 0.60875 | 0.50141 | 0.39841 | 0.24406 | 0.03626 |
| 0.87906 | 0.76616 | 0.69225 | 0.60726 | 0.49884 | 0.38018 | 0.23627 | 0.03316 |
| 0.87550 | 0.76041 | 0.68958 | 0.60710 | 0.49794 | 0.37255 | 0.21678 | 0.02639 |
| 0.87441 | 0.75601 | 0.68718 | 0.60332 | 0.49615 | 0.36137 | 0.21493 | 0.02591 |
| 0.87167 | 0.75288 | 0.68592 | 0.59593 | 0.49352 | 0.35964 | 0.17915 | 0.02578 |
| 0.87161 | 0.75199 | 0.68527 | 0.59205 | 0.49162 | 0.35839 | 0.16873 | 0.02206 |
| 0.86528 | 0.74985 | 0.68255 | 0.58408 | 0.49062 | 0.35763 | 0.16828 | 0.01884 |
| 0.86446 | 0.74403 | 0.68080 | 0.56910 | 0.48647 | 0.35029 | 0.16423 | 0.01553 |
| 0.86180 | 0.74057 | 0.67934 | 0.56742 | 0.47906 | 0.34657 | 0.15107 | 0.01431 |
| 0.85309 | 0.73919 | 0.67757 | 0.56658 | 0.47592 | 0.34649 | 0.14326 | 0.00923 |
| 0.84637 | 0.73302 | 0.67607 | 0.55630 | 0.47330 | 0.34505 | 0.13782 | 0.00781 |
| 0.84002 | 0.73215 | 0.67075 | 0.55287 | 0.47287 | 0.34418 | 0.13727 | 0.00721 |
| 0.82905 | 0.72808 | 0.66898 | 0.54147 | 0.47093 | 0.34025 | 0.13567 | 0.00605 |
| 0.82493 | 0.72753 | 0.66863 | 0.53883 | 0.47045 | 0.33657 | 0.12226 | 0.00506 |
| 0.82418 | 0.72380 | 0.65784 | 0.53413 | 0.46704 | 0.31768 | 0.10964 | 0.00467 |
| 0.82283 | 0.72087 | 0.65330 | 0.52919 | 0.45718 | 0.31070 | 0.10759 | 0.00063 |
| 0.80519 | 0.72070 | 0.64766 | 0.52044 | 0.44944 | 0.30771 | 0.10653 | 0.00020 |

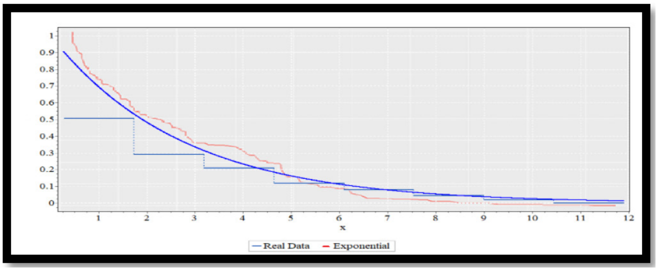



**Figure 2**

**Estimated exponential survival regression model**

## 6. Discussion and Conclusion

1. It is clear from Figure 2 It is clear that the results of the exponential survival regression model using the Bayesian method are very similar to those using the original values., and this indicates the effectiveness of the Bayesian method in estimating parameters and its ability to approach the real data accurately based on the available data, and when the estimated values of the model are close to the values. This means that the Bayesian method used for estimation is very accurate and that the model used for estimation has estimated it well and reflects the facts accurately.
2. From Table 3, The survival period for people with stomach cancer has been shown to decrease over Time. This means that the longer the disease lasts, the greater the likelihood of the individual with this disease dying.

## References

[1] A. S. Mezher and W. S. Ibrahim, "Using the Bayesian method to estimate and compare the regression of the exponential and gamma survival (Simulation)," *J. Coll. Basic Educ.*, vol. 30, no. 124, pp. 1-10 (2024).

[2] J. A. Anber, "The use of Cox model in the survival time analysis of lung cancer in Iraq," *Int. J. Agric. Stat. Sci.*, vol. 17 (2021).

[3] R. Bender, T. Augustin, and M. Blettner, "Generating survival times to simulate Cox proportional hazards models," Stat. Med., vol. 24, no. 11, pp. 1713-1723 (2005).

[4] J. Fox and S. Weisberg, "Cox proportional-hazards regression for survival data," An R and S-PLUS Companion to Applied Regression (2002).

[5] W. S. Ibrahim and A. K. Aliwi, "Using the method of maximum likelihood estimation to estimate the survival function of the new expanded transformer Weibull distribution," *J. Stat. Sci.*, vol. 19, pp. 23-33 (2023).

[6] W. S. Ibrahim and B. J. Khaleel, "Comparison of some classical methods for estimating the survival function of the two-parameter Lindley distribution," *J. Stat. Sci.*, vol. 17, pp. 70-81 (2023).

[7] W. S. Ibrahim and D. I. Mhadi, "A comparison of some methods for estimating Rasch model parameters," *J. Coll. Basic Educ.*, vol. 22, no. 93, pp. 245-260 (2016).

[8] A. A. Karim, "Analysis of the survival function when the risk factor is proportional to time (an applied study)," M.S. thesis, Coll. Admin. Econ., Univ. Karbala (2018).

[9] M. R. Karim and M. A. Islam, Reliability and Survival Analysis, Springer Singapore (2019).

[10] A. H. Muse, O. Ngesa, S. Mwalili, H. M. Alshanbari, and A. A. H. El-Bagoury, "A flexible Bayesian parametric proportional hazard model: Simulation and applications to right-censored healthcare data," *J. Healthcare Eng.* (2022).

[11] H. Swami Nathan and J. A. Gifford, "Bayesian estimation in the Rasch model," *J. Educ. Stat.*, vol. 7, no. 3, pp. 175-191 (1982).

[12] M. M. Abbood, H. H. Ebrahim, A. Al-Fayadh, and A. J. Obaid, "Measure defined on Γ–algebra and some of their generalizations," *J. Interdiscip. Math.*, pp. 1-7 (2024), doi: 10.47974/JIM-1502.

*Received February, 2024*