

Using weibull distribution in the forecasting by applying on real data of the number of traffic accidents in sulaimani during the period (2010-2013)

Samira Muhammad Salh

University of Sulaimani - College of Administration and Economic / Statistic Department.
Sulaimani / Iraq
Samira196902@yahoo.com

Abstract:

In this study we tried the application of one continuous probability distributions ,especially Weibull probability distribution , which is used in the study of the reliability and quality control in the forecasting .Real data of the number of traffic accidents for the city of Sulaimani city during (2010-2013) has been applied:

- 1) Convert(transference) Weibull distribution to linear regression to estimate the parameters.
- 2) How to use the generalized Weibull distribution in finding probability of the occurrence of traffic accidents.

Keywords: Weibull distribution, reliability, regression, ordinary least square and linear estimation techniques.

1. Introduction:

The problem of estimating the unknown parameters of the probability distributions is important issues that have received and interest to researchers and others that interested in mathematical statistics, by giving the development of estimation methods and contrast, which recall that for accuracy estimation we can find the best estimator for these parameters. In this paper we focused on the Essence of the distribution of Weibull , In order to estimate the parameter of this distribution .In 1990 (AL-Baidhani) reached to an alternative formula for the distribution of Weibull named "generalized Weibull distribution " , to avoid the problems that facing researchers in estimating the parameters of the distribution , especially when his parameters exceeds the value of (3.6) , so suggest a new formula in several ways , it was the most important method of least squares .

Research Problem:

In this study we focus on the basic research into the possibility of using Weibull probability distribution to forecast and determine the modalities and methods used to estimate unknowns parameters .The aim of the research on the theoretical knowledge and applied to the estimation methods for the parameters of the Weibull distribution of Weibull and his methods and focus on the possibility to use it in forecasting. That applied on real data of the number of traffic accidents in Sulaimani during (2010-2013).

2. Methodology and research hypotheses:

in this research, we relied on the descriptive analytical approach in the characterization of the distribution of Weibull, and how to estimate its features. The data of traffic accidents to the city of Sulaimani derived from the Directorate of Traffic in Sulaimani during the extended period of time between (2010-2013) .In this research, we focused on a basic premise of the appropriateness of the data distribution and generalized Weibull

distribution with two parameters. Also how similar the real values through probabilistic distribution and Weibull.

3. The theoretical side:

3.1 Essence Weibull Distribution:

Here we will discuss the theoretical framework for the distribution and and situations through the reliability and hazard function and general linear methods to estimate parameters of this distribution. Suppose that (X) have continuous random variable with Weibull three parameters distribution which are $(\theta, \lambda, \kappa)$, (k) represented as shape parameter, (λ) scale parameter, and (θ) Location parameter, and as long as: $(\lambda > 0), (k > 0), (\theta)$ real value, and it gives a density continued as follows:

$$f(x) = \begin{cases} 0, & \text{for } X \leq 0 \\ \frac{\kappa}{\lambda} \left(\frac{X - \theta}{\lambda} \right)^{\kappa-1} e^{-\left(\frac{X - \theta}{\lambda}\right)^\kappa}, & \text{for } X > 0 \end{cases} \quad (1)$$

Let $\left(\frac{\kappa}{\lambda} = \alpha\right), (\beta = \kappa)$ and $\left(\left(\frac{X - \theta}{\lambda}\right) = X\right)$ can we write the equation 1 as follow :

$$f(x) = \begin{cases} 0, & \text{for } X < 0 \\ \alpha (X)^\beta e^{-(X)^\beta} & \text{for } X > 0 \end{cases} \quad (2)$$

The cumulative probability function is given

$$F(x) = \begin{cases} 0, & \text{for } X < 0 \\ 1 - e^{-(X)^\beta} & \text{for } X > 0 \end{cases} \quad (3)$$

And the adoption of the Weibull density distribution function we get generalized Weibull distribution and when $(\theta = 0)$, we get the Weibull distribution function with two parameters and, density function, are:

$$f(x) = \begin{cases} 0, & \text{for } X \leq 0 \\ \frac{\kappa}{\lambda} \left(\frac{X}{\lambda} \right)^{\kappa-1} e^{-\left(\frac{X}{\lambda}\right)^\kappa}, & \text{for } X > 0 \end{cases} \quad (4)$$

Because the Weibull distribution in relation to practical applications, it is used today in several fields, including:

1. The analysis of survival.
2. Industrial reliability analysis.
3. Theoretical maximum value.
4. The weather, and in the description of the distribution of wind speed.
5. Communication Systems Engineering.
6. The field of general insurance.

Weibull distribution. This is one of the extreme distributions function , and it is given as unique for the smallest value for a large number of random variables independent (X), It is also a continuous distributions function called the distribution test, such as the distribution of

(χ^2) and distribution (t) and distribution (F), due to the importance in mathematical statistic methods. Those depending on the probability distribution of the extreme values, we can express the generalized distribution and (Generalized Weibull Distribution), the function of the distribution of a random variable (X). As follows:

$$G(X, \theta, \alpha, \kappa) = p(X \leq x) = \exp \left[- \left\{ 1 - \frac{\kappa}{\theta} \right\} (x - \alpha)^{\frac{1}{\kappa}} \right], \kappa \neq 0 \quad (5)$$

$$G(X, \theta, \alpha) = \exp \left[\left\{ - \exp \left\{ - \frac{x - \alpha}{\theta} \right\} \right\} \right], \kappa = 0 \quad (6)$$

Where ;

(κ) shape parameter , (θ) Scale parameter and (β) Location parameter.

As the relationship between the Weibull distribution and generalized extreme value distribution, are: $(T = -X)$: note that (T) represents the random variable distribution and Weibull generalized, and we can write:

$$p(T \leq t) = p(-X \leq x) = 1 - G(x; \theta, \alpha, \kappa)$$

Where : (t) the value of a random variable (T),

and be specific from the $\left(\beta - \frac{\theta}{\kappa} \right)$, and in order to get the Weibull distribution function with two and three parameters, we replace (x) with (-t) and $(-\alpha)$ with value $((\beta))$ in the two equations (5) and (6), it becomes:

$$F(t, \theta, \beta, \kappa) = 1 - \exp \left[\left\{ - \left(1 + \frac{\kappa}{\theta} \right) (t - \beta) \right\}^{\frac{1}{\kappa}} \right], \kappa \neq 0 \quad (7)$$

$$F(t, \theta, \beta) = 1 - \exp \left\{ - \exp \left(\frac{t - \beta}{\theta} \right) \right\}, \kappa = 0 \quad (8)$$

To get probability density function for the generalized Weibull distribution with three and a two parameters, we take a derivative of both two relationships (5) and (6), we get ;

$$f(t; \theta, \beta, \kappa) = \frac{1}{\theta} \left(1 + \frac{\kappa}{\theta} (t - \beta) \right)^{\frac{1}{\kappa} - 1} \exp \left[- \left(1 + \frac{\kappa}{\theta} (t - \beta) \right)^{\frac{1}{\kappa}} \right], \kappa \neq 0 \quad (9)$$

$$f(t; \theta, \beta) = \frac{1}{\theta} \exp \left(\frac{t - \beta}{\theta} \right) \exp \left[- \exp \left(\frac{t - \beta}{\theta} \right) \right], \kappa = 0 \quad (10)$$

Consequently we can write a Reliability Function as follows

$$R(t) = \exp\left[-\exp\left(\frac{t-\beta}{\theta}\right)\right] \quad (11)$$

And the (Hazard rate) for this distribution, as follows:

$$h(t) = \frac{1}{\theta} \exp\left[\frac{t-\beta}{\theta}\right]$$

3.2 Linear Estimation Techniques:

Let (T) be the random variable with Weibull distribution function with two parameters, and according to the relationship (8) we write:

$$F(t; \theta, \beta) = 1 - \exp\left[-\exp\left(\frac{t-\beta}{\theta}\right)\right] \quad ; \kappa = 0$$

Let ($t_1, t_2, t_3, \dots, t_n$) the data distributed to the Weibull distribution with two parameters, then we can estimate this parameters, as follows:

$$\hat{\theta} = \sum_{i=1}^n c_i t_i \quad , \quad \hat{\beta} = \sum_{i=1}^n a_i t_i$$

As: i is (1,2,3, ..., n)

(a_i, c_i) is the linear weights (Linear Weighting Factors).

Accordingly, we write mathematical relationship to the inverse function of the Weibull distribution with two parameters, as follows:

$$t_p = \beta + \theta x_p \quad (12)$$

$$x_p = \ln(-\ln(1-p))$$

And then we can find the value of each (θ, β), considers represent a slop and the lump in the equation of the regression line (15), through several methods enable us to estimate two parameters of Weibull distribution, but we will limit ourselves here only to use a White (White Method), based on the theory of regression analysis, and so transforming the distribution function of the relationship (8) to form a simple linear regression model as

$$F(t, \theta, \beta) = 1 - \exp\left\{-\exp\left(\frac{t-\beta}{\theta}\right)\right\} \quad (13)$$

or
$$1 - F(t, \theta, \beta) = \exp\left\{-\exp\left(\frac{t-\beta}{\theta}\right)\right\}$$

And we get to take twice the logarithm of the function above, on:

$$\ln\{-\ln[1 - F(t; \beta, \theta)]\} = \frac{t - \beta}{\theta} \quad (14)$$

Assuming that ($t_n > \dots > t_3 > t_2 > t_1$), is order sample arranged for a random sample size (n), then it write previous relationship after taking the logarithm twice, as follows

$$\ln\{-\ln[1 - F(t_{(i)}; \beta, \theta)]\} = \frac{t_{(i)} - \beta}{\theta}$$

And including

$$t(F_i) = \beta + \theta \ln \left[-\ln(1 - \hat{F}_i) \right]$$

And if we put x_i and instead of $\ln \left[-\ln(1 - \hat{F}_i) \right]$ and y_i instead of $t(F_i)$, then become linear regression equation is as follows:

$$y_i = \beta + \theta x_i \quad (15)$$

Because that is the point estimate for, which is created through one of the following nonparametric methods?

$$1) \hat{F}_i = \frac{1}{n+1}$$

$$2) \hat{F}_i = \frac{i - \left(\frac{3}{8}\right)}{n + 0.25}$$

$$3) \hat{F}_i = \frac{i - 0.5}{n}$$

But the goal of estimating the two parameters of generalized Weibull distribution according to (White method), using the method of least squares minimizing estimated the following:

$$G(\beta, \theta) = \sum_{i=1}^n \left[t(F(i)) - \beta - \theta \left\{ \ln \left[-\ln(1 - \hat{F}_i) \right] \right\} \right]^2 \quad (16)$$

We will get the two parameters of the regression equation above, by applying the following relationship:

$$\hat{\theta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (17)$$

$$\hat{\beta} = \bar{y} - \hat{\theta} \bar{x}$$

Where

$$1. \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \dots \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

2. $y_i = t(F_i)$, represented average values are Views phenomenon studied and ranked in ascending order.

$$3. x_i = \ln \left[-\ln(1 - \hat{F}_i) \right]$$

$$4. \hat{F}_i = \frac{i - 0.05}{n}$$

We can also obtain it by estimating the first order for a smooth Taylor series ranked in ascending order, as follows:

$$Z_i = \ln(-\ln(1 - p_i)), t(p_i) = \beta + \theta Z_i$$

$$\frac{\partial t}{\partial \beta} = 1, \quad \frac{\partial t}{\partial \theta} = Z_i$$

$$f(a+h) = f(a) + h_1 f'(a) + h_2 f''(a)$$

as

$$f(a) = \beta_0 + \theta_0 Z_i$$

$$E(t_i) = \beta_0 + \theta Z_i + (\beta - \beta_0).1 + (\theta - \theta_0).Z_i$$

$$= \beta_0 + \theta Z_i + \beta - \beta_0 + \theta Z_i - \theta_0 Z_i$$

$$= \beta + \theta Z_i$$

And we can use method (OLS) based on the style of matrices, estimating parameters of Weibull distribution, where the extract estimated values for the following relationship:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\theta} \end{pmatrix} = (H H)^{-1} H t(i) z \tag{18}$$

As

$$H = \begin{bmatrix} 1 & Z_1 \\ 1 & Z_2 \\ \cdot & \\ \cdot & \\ 1 & Z_n \end{bmatrix}$$

IJOART

In order to determine the appropriate data of any phenomenon studied the Weibull distribution (GSP), it is possible to subject these data to test goodness of fit, and based on the table values (Anderson, A²n) for the distribution of Weibull with two parameters and the distribution of the value of the extremist, made by (D, agostion, R, B., Stephens, M, A., 1986, P.86-90), it can reach a table values (A²n) for the Weibull distribution with two parameters, as follows: Let we have a sample of the size ranked (t₁, t₂, t₃, ..., t_n) and estimated values (β̂, θ̂) for the Weibull distribution with two parameters, as follows:

$$F(t; \hat{\theta}, \hat{\beta}) = 1 - \exp \left\{ - \exp \left(\frac{t - \hat{\beta}}{\hat{\theta}} \right) \right\} \quad ; \kappa = 0 \tag{19}$$

Suppose that

$$Z_i = - \exp \left[\frac{t_i - \hat{\beta}}{\hat{\theta}} \right]$$

Substituting this (suppose) hypothesis in equation (19), we get:

$$F(t; \hat{\theta}, \hat{\beta}) = 1 - \exp [Z_i]$$

Accordingly, to applying equation (20) we get a values of (A2n) .

$$A^2n = -n - 2 \left\{ \sum_{i=1}^n P_i \ln(F_i) + (1 - P_i) \ln(1 - F_i) \right\} \quad (20)$$

Where $P_i = \frac{i - 0.5}{n}$, $i = 1, 2, 3, \dots, n$

4. Applications part:

After our review of probability distributions ,spatially the Weibull distribution, we can say that the use of probability distributions is one of the important statistical method and essential in forecasting the phenomena of life. In order to reach statistical models that designed to diagnose these phenomena, estimating and analyzing the different of the interaction among them, by using mathematical and statistical methods to determine trends , explained by its finding of indicators and accurate estimates have ,that would be the basis for the development of possible solutions and how to deal with them in the forecasting . In our research we focused on study trends in the rate of traffic accidents in the city of Sulaimani by duration (2010-2013) based on the Weibull distribution and estimate parameters for this distributions by using the least squares method to find the probability of occurrence of the expected rate of traffic accidents.

In this study, we used data on monthly rates for a number of traffic accidents to the city of Sulaimani during (2010-2013) from the records of the directorate traffic.

Table (1) Ascending order of the number of traffic accidents to the city of Sulaimani during (2010-2013)

| Number of accidents | Month | Number of accidents | Month |
|---------------------|-------|---------------------|-------|
| 44 | 1 | 89 | 25 |
| 52 | 2 | 94 | 26 |
| 53 | 3 | 96 | 27 |
| 59 | 4 | 96 | 28 |
| 59 | 5 | 98 | 29 |
| 60 | 6 | 99 | 30 |
| 63 | 7 | 99 | 31 |
| 63 | 8 | 101 | 32 |
| 64 | 9 | 103 | 33 |
| 65 | 10 | 103 | 34 |
| 67 | 11 | 104 | 35 |
| 68 | 12 | 104 | 36 |
| 72 | 13 | 108 | 37 |
| 73 | 14 | 108 | 38 |
| 76 | 15 | 125 | 39 |
| 77 | 16 | 127 | 40 |
| 77 | 17 | 128 | 41 |
| 79 | 18 | 129 | 42 |
| 79 | 19 | 129 | 43 |
| 82 | 20 | 133 | 44 |
| 83 | 21 | 139 | 45 |
| 86 | 22 | 143 | 46 |

| | | | |
|----|----|-----|----|
| 86 | 23 | 154 | 47 |
| 87 | 24 | 159 | 48 |

According to the above review in theatrical part for "Weibull generalized distribution", after ascending order of data of traffic accidents for the city of Sulaimani (as shown in Table (1), and after estimating the parameters for Weibull with two parameters(β, θ) for those data with applying equation (17), we have obtained and estimate its features according to the least squares method (OLS), using software "easy fit" to test, and estimate the values of the parameters, as shown in the following tables (2) and (3) respectively :

Table (2) value of A^2n

| Anderson-Darling | | | | | |
|--------------------------|----------------|--------|--------|--------|-----------------|
| Sample Size | 48 | | | | |
| Statistic | 0.7654 | | | | |
| Rank | 34 | | | | |
| α | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 |
| Critical Value parameter | 1.3749 | 1.9286 | 2.5018 | 3.2892 | estimate 3.9074 |
| Reject? | No | No | No | No | 100.67 No |
| | $\hat{\theta}$ | | | | 3.8444 |

Table (3) Shows the estimated values of the two parameters of Weibull distribution

Through our use of the (Anderson Darling) test, with the equation(19) by using the "easy fit" software .we achieved the value of (A^2n) to ensure the appropriateness of data traffic accidents, according to the estimated values in Table (3). for the distribution Weibull with two parameters, for the value of the accident number (0.7654) given in table (2), while comparing these values at level of significance (0.05) and equal to (2.50), we accept the data for the number of accidents, because the calculated value is

less than the value, so can we say about these data it is belong to Weibull distribution with two parameters. From equation (15) we can find the estimate number for number of accidents as shown in table (4), and Figure (1) Shows similarity between real values of number accidents and estimate values.

Table (4) Show the estimate number for number of accidents

| number of accidents | estimate number | | number of accidents | estimate number |
|---------------------|-----------------|--|---------------------|-----------------|
| 44 | 83 | | 87 | 99 |
| 52 | 87 | | 89 | 99 |
| 53 | 89 | | 94 | 100 |
| 59 | 91 | | 96 | 100 |
| 59 | 92 | | 96 | 100 |
| 60 | 93 | | 98 | 100 |
| 63 | 93 | | 99 | 100 |
| 63 | 94 | | 99 | 101 |
| 64 | 94 | | 101 | 101 |
| 65 | 95 | | 103 | 101 |
| 67 | 95 | | 103 | 101 |
| 68 | 96 | | 104 | 102 |
| 72 | 96 | | 104 | 102 |
| 73 | 96 | | 108 | 103 |
| 76 | 96 | | 108 | 103 |
| 77 | 97 | | 125 | 103 |
| 77 | 97 | | 127 | 103 |
| 79 | 98 | | 128 | 104 |
| 79 | 98 | | 129 | 104 |
| 82 | 98 | | 129 | 104 |
| 83 | 98 | | | |
| 86 | 99 | | | |
| 86 | 99 | | | |

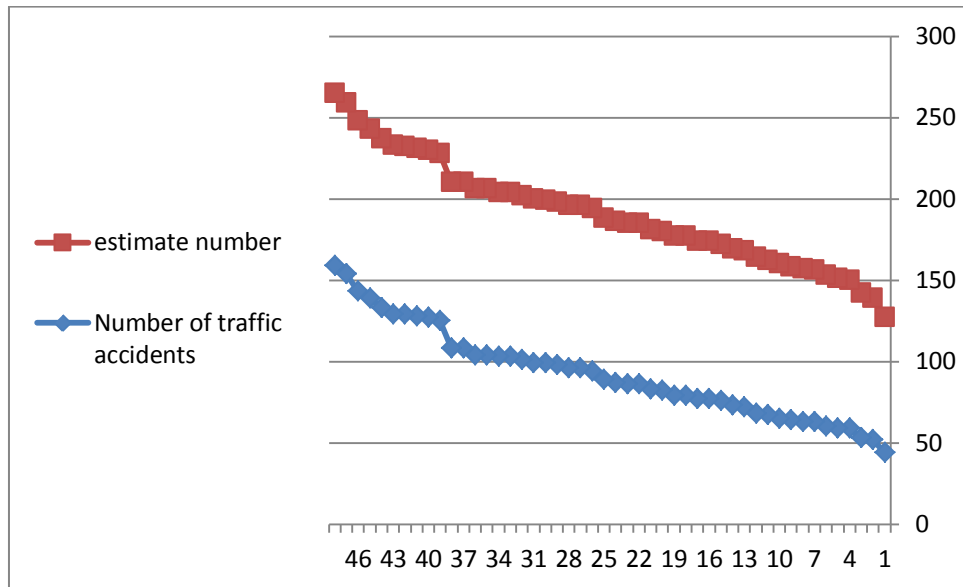


Figure (1) Show similarity between real and estimate values

Also we can find expected probability of occurrence of the number of accidents for (N) years, by using equation (19) and estimate value of (β, θ) .

$$G = 1 - [1 - F(t)]^N$$

For example, if the rate of accidents for the December is equal (44), F(t) will be :

$$F(t) = 0.01$$

Similarity for two years and four year equal (0. 29), (0.06) respectively ,and so on .

4. Conclusion:

Through this study, using Weibull distribution to analysis the number of accidents we conclude these points:

- 1) Can be using mathematical form of Weibull distribution to forecasting.
- 2) Can be using regression technique to estimate Weibull parameters.
- 3) From table (4) for average number of accidents are increasing.

At the end we recommend for future work can be using these theorems of probability distribution in realized process.

Reference :

1. AI-Baidhdhani,F.A.(1990),Reliability Theory in operational Research ,University of St. Andrews,U.K.
- 2.D,agostion,R.B.,and Stephens ,M.A.,(1986),Goodness of Fit Techniques ,Marcel Dekker, inc, New York.
3. Grosh,D.L,(1989)" A Primer of Reliability Theory" John Wiley & Sons, inc, Canada.
- 4.Hung,W.L,(2004),Estimation of Weibull parameters using a Fuzzy ,Least squares method, International Journal of Uncenation,vol.12,No.,pp:701-711.
- 5.Kahadawala,C.,(2006),Estimation of Weibull Distribution :the add Weibull family ,Statistical modeling ,6;pp:265-277.
6. Zacks,S.,(1992),Introduction to Reliability Analysis Probability Models analysis Statistical Methods, New York.
- 7.<http://www.erpt.org/014Q/nelsa-06.htm>.Retrived.
- 8.Chin-Diew Lai,(2014)"Generalized Weibull Distributions "Springer ,Heidelberg, New York, Dordrecht ,London.
9. Horst Rinne,(2009)"The Weibull Distribution A Handbook", Justus-Liebig-University Giessen, Germany, Chapman & Hall/CRC, Taylor & Francis Group.

IJOART