ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 11, No. 2, April 2023, pp.275-283

Comparison of Weibull and Fréchet distributions estimators to determine the best areas of rainfall in Iraq

Inam Abdul Rahman Noaman[1], Huda Amer Abdul Ameer[2], Wahhab Salim Mohammed[3]
[1]College of Administration and Economics, University of Diyala, Iraq
[2]College of Science, University of Diyala , Iraq
[3]College of Administration and Economics, University of Diyala, Iraq

ABSTRACT

In this research, an appropriate distribution of the amount of rain will be found in the Iraqi governorates for the period (2006-2014) and the researcher used two important distributions, namely, the Weibull distribution and the Fréchet distribution. Where the specific distribution was determined based on the minimum criteria (the criteria of goodness of fit) and the tests used are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Rainfall in the Iraqi governorates for the stations (Mosul, Kirkuk, Tikrit, Khanaqin, Rutba, Baghdad, Karbala) is a Weibull distribution using the greatest possible estimation method, while the stations in other provinces (Najaf, Diwaniyah, Maysan, Basra) the Fréchet distribution was the distribution It is better to represent the data of these stations using the method of estimating the greatest possible as well. We also note the superiority of the method of maximum likelihood of least squares.

Keywords: Weibull distribution, Fréchet distribution, Maximum likelihood, Least square, Goodness of fit.

Corresponding Author:

Inam Abdul Rahman Noaman
College of Administration and Economics
University of Diyala, Iraq

E-mail: Inaamsta@uodiyala.edu.iq

1. Introduction

Water is one of the vital natural resources and it is the first condition for survival because it plays a fundamental and important role in sustaining human life, whether in agriculture or in industry and in our daily life. The shortage of water supply will cause a significant negative impact on the country, but its excessive supply can also contribute to natural disasters such as landslides or floods, and some diseases are also linked to unhealthy water supplies and sewage management, so it is important to manage water materials to rise optimally because they have great effects on a country, if water is the main source of income In the water resource system as well as the main component of the life cycle, it is responsible for depositing most of the fresh water on the earth and provides suitable conditions for many types of ecosystems, as well as water for hydropower stations, irrigation, and crops [1].

On the theoretical, the distributions of rain will be clarified using some mathematical distributions to test the best appropriate distribution of rainfall in Iraq between the periods 2004 to 2016. These distributions are the Weibull distribution and Frechet distribution. To obtain the estimations of the parameters of the probability density functions of the distributions, the maximum likelihood method, least squares have been used, and these methods are considered among the classic methods of estimation, and the values of numerical estimates differ from one estimator to another, and the research aims to find the appropriate probability density function for the distribution of rainfall data. In some Iraqi governorates, the practical aspect also included analyzing the data of rainfall data in several Iraqi governorates and determining which area is better than finding the appropriate distribution for it [2-7].

Studies related to the amounts of depression and rainfall in Iraq are considered few in general, and in the statistical aspect in particular, and the most important thing is to determine the distributions and the appropriate statistical model for estimating and predicting rainfall.

This paper aims to determine the appropriate distribution of rainfall rates in Iraq based on some probability density functions.

2. Material and methods

The study of the phenomenon of rain in the world is of interest to all countries of the world, whether rich or poor, for the distinction of the phenomenon as it increases the economic resources of countries, and attention will be paid to the probability distribution of the Weibull and the distribution of Fréchet (inverse Weibull) to study the random behavior of the rates of rainfall recorded in the stations located in All governorates of Iraq, as well as estimating landmarks using maximum likelihood method and least squares method.

2.1. Weibull distribution

The distribution was discovered in 1939 by the physicist Weibull, and it was named after him. It is considered one of the important α distributions and has many uses in the vital fields, the environment, the meteorological environment, as well as the economy, and it has a probability density function p.d.f as follows [1]

$$f(x,\alpha,\beta) = \frac{\beta}{\alpha^{\beta}}(x)^{\beta-1}e^{-\left(\frac{x}{\alpha}\right)^{\beta}} \qquad X,\alpha,\beta > 0 \qquad (1)$$

Where β represents the shape parameter, α represents the scale parameter, and the cumulative distribution function (C.D.F) is as follows:

$$F(x,\alpha,\beta) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}} \qquad (2)$$

2.2. Fréchet distribution

It is considered a Fréchet distribution (called the inverse Weibull distribution) and it is considered one of the important distributions in the data. It is considered one of the continuous distributions. It is used to analyze lifetime data and is concerned with distributing anomalous values. It has many uses in the health and environmental fields. It has a probability density function and a distribution function as follows [3]:

$$f(x,\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{\alpha}{x}\right)^{\beta+1} e^{-\left(\frac{\alpha}{x}\right)^{\beta}} \qquad X > 0,\alpha,\beta > 0 \qquad (3)$$

Where β represents the shape parameter, α represents the scale parameter, and the cumulative distribution function (C.D.F) is as follows:

$$F(x,\alpha,\beta) = e^{-\left(\frac{\alpha}{x}\right)^{\beta}} \qquad X > 0,\alpha,\beta > 0 \qquad (4)$$

2.3. Statistical Inference

In this paper, two methods for estimating the parameters of distributions will be addressed, namely, the Maximum Likelihood method least squares method, which will be presented in this section as follows:

2.3.1. Maximum likelihood method (MLM)

The scientist "C.F. Gauss" was the first to formulate the method of the maximum likelihood function, while the researcher "R.A.Fisher" applied it for the first time through many studies. The extracted capabilities according to this method are characterized by being efficient estimators and having the property of the least possible variance (Minimum). variance Unbiased estimators), as well as a very important property, which is the invariant property, which is more accurate than other estimation methods, especially when the sample size (n) increases. If $x_1,x_2,\ldots x_n$ are the vocabulary of a random sample of size n drawn from a population that has a known probability density function f(x,α,β), then (the maximum potential function of the sample data is defined as the joint distribution of that data) (L), and the method aims To make likelihood function for random variables the greatest possible, which can be written as follows:

2.3.1.1. Weibull distribution estimators

The withdrawn samples are characterized by independence and therefore:

$$\boldsymbol{L}\,\boldsymbol{f}(\mathrm{x}_1, \mathrm{x}_2, \ldots .\mathrm{x}_n; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{\boldsymbol{\beta}}{\boldsymbol{\alpha}}\left(\frac{\boldsymbol{\alpha}}{\boldsymbol{x}}\right)^{\boldsymbol{\beta}+\mathbf{1}} \boldsymbol{e}^{-\left(\frac{\alpha}{x}\right)^{\beta}} \tag{5}$$

To find the estimators of the maximum likelihood function concerning (β,α) by taking the natural logarithm of the above equation and differentiating the function for the two parameters and by setting it equal to zero, we get the following two estimations [4]:

$$\frac{\partial Log\ L}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} log\ xi - \frac{1}{\alpha}\sum_{i=1}^{n} x_i^{\beta}\ Logx_i = 0 \tag{6}$$

$$\frac{\partial Log\ L}{\partial \alpha} = -\frac{n}{\alpha} + \frac{1}{\alpha^2}\sum_{i=1}^{n} x_i^{\beta} = 0 \tag{7}$$

After simplifying equation (7), we get the estimator of the scale parameter$\hat{\alpha}$:

$$\hat{\alpha}_{ML} = \frac{\sum_{i=1}^{n} t_i^{\beta_{ML}}}{n} \tag{8}$$

Substituting equation (8), we get the following equation:

$$\frac{\sum_{i=1}^{n}(x_i^{\beta} Logx_i)}{\sum_{i=1}^{n} x_i^{\beta}} - \frac{1}{\beta} - \frac{1}{n}\sum_{i=1}^{n} log\ xi = 0 \tag{9}$$

Equation (9) is solved by one of the numerical methods used to solve nonlinear mathematical equations, such as Newton-Raphson's method for obtaining the shape parameter $\hat{\beta}$.

2.3.1.2. Frechet distribution estimators

The maximum likelihood function of this distribution is as follows [3]:

$$L\,f(\mathrm{x}_1, \mathrm{x}_2, \ldots .\mathrm{x}_n; \beta, \alpha) = \prod_{I=1}^{N} \frac{\beta}{\alpha}\left(\frac{\alpha}{x}\right)^{\beta+1} e^{-\left(\frac{\alpha}{x}\right)^{\beta}} \tag{10}$$

$$L\,f(\mathrm{x}_1, \mathrm{x}_2, \ldots .\mathrm{x}_n; \beta, \alpha) = \left(\frac{\beta}{\alpha}\right)^{n} \prod_{i=1}^{n} \left(\frac{\alpha}{x}\right)^{\beta+1} e^{-\sum_{i=1}^{n}\left(\frac{\alpha}{x}\right)^{\beta}} \tag{11}$$

$$\frac{\partial Log\ L}{\partial \alpha} = \frac{n\beta}{\hat{\alpha}} - \beta\hat{\alpha}^{\beta}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)^{\beta} = 0 \tag{12}$$

After simplifying Equation (12), we get the estimator of the scale parameter$\hat{\alpha}$:

$$\hat{\alpha}_{ML} = \left(\frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)^{\beta}}\right)^{\frac{1}{\beta}} \tag{13}$$

To estimate the value of the shape parameter β by substituting Equation (13):

$$\frac{\partial Log\ L}{\partial \beta} = \frac{n}{\hat{\beta}} - n\ ln\hat{\alpha} - \sum_{i=1}^{n} ln\ x_i + \hat{\alpha}^{\beta}\sum x_i^{-\beta}(-1)Ln\ x_i - \alpha^{\beta}Ln\hat{\alpha}\sum x_i^{-\beta} = 0 \tag{14}$$

Solve Equation (14) by numerical methods, such as Newton-Raphson's method, because it is difficult to solve using ordinary methods.

2.3.2. Least square method (LSM)

The least squares method is considered one of the important and widely used methods in estimating and it can be used to find solutions to mathematical and engineering problems. The method depends on the existence of a relationship between two or more variables, in which the distribution parameters are estimated, which makes the sum of the error squares at its least end. It is considered one of the most important methods that have been widely used. In estimating parameters because they have the characteristics of a good estimator, including unbiasedness and consistency [5].

2.3.2.1. Weibull distribution estimators

This method has been proposed by researchers (Venkatraman, Wilson, and Swain) to estimate the parameters of the Weibull distribution, and this method depends on the cumulative distribution function (CDF) and is not calculated using default values of the parameters, but is estimated using nonparametric methods and according to the following formula [6]:

$$\hat{F}(t_i) = \frac{i}{n+1}$$

$$LSM = \sum_{i=1}^{n}\left[\frac{i}{n+1} - F(x)\right]^2$$

The least squares estimates of (β, α) are obtained by minimizing the amount of least squares (L.S):

$$LSM = \sum_{i=1}^{n}\left[\frac{i}{n+1} - F(x)\right]^2$$

$$LSM = \sum_{i=1}^{n}\left[\frac{i}{n+1} - (1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}\right]^2 \qquad (15)$$

To estimate the shape parameter by performing partial differential concerning β :

$$\frac{\partial LSM}{\partial \beta} = 2\sum_{i=1}^{n}\left[\frac{i}{n+1} - (1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}})\right] e^{-\left(\frac{x}{\alpha}\right)^{\beta}} Ln\, e^{-\frac{x}{\alpha}} \qquad (16)$$

The shape parameter is estimated and assuming that the scale parameter is known by minimizing the sum of squares of equation (15) concerning α and as follows:

$$\frac{\partial LSM}{\partial \alpha} = 2\sum_{i=1}^{n}\left[\frac{i}{n+1} - (1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}})\right] \beta e^{-\left(\frac{x}{\alpha}\right)^{\beta-1}} \left(-\alpha\, e^{-\frac{x}{\alpha^2}}\right) \qquad (17)$$

It is also an implicit function of the two observations and parameters (β and α) and it is necessary to use an iterative method as mentioned above to get the two estimators (β and α) in this way from the last two equations.

2.3.2.2. Fréchet distribution estimators

The least squares method involves minimizing the sum of the squares of the error as follows [7]:

$$M(a,b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2$$

By partial differentiation of (a, b) and equalizing the result to zero, two equations are obtained, two equations are obtained, and by simplification, the values of the estimators for least squares $(\hat{a}, \hat{b})$are obtained. For the two parameters, one of the advantages of this method is that it is unbiased and based on a cumulative distribution function and my agencies [8]:

$$F(x,\alpha,\beta)=e^{-\left(\frac{\alpha}{x}\right)^{\beta}} \qquad X>0,\alpha,\beta\ >0 \tag{18}$$

By taking the natural logarithm of both sides, we get:

$$Ln\,F(x,\alpha,\beta)=-\left(\frac{\alpha}{x}\right)^{\beta} \qquad X>0,\alpha,\beta\ >0$$

$$Ln(Ln\,F(x,\alpha,\beta))=-\beta\,Ln\,x+\beta\,Ln\,\alpha \tag{19}$$

By matching with the simple linear regression model, the Y_i values were obtained, and to find the values of the estimators by the method of least squares, we apply the previous equation and get:

$$T=\sum_{i=1}^{n}(Y_i+\beta\,Ln\,x_i\ -\beta\,Ln\,\alpha)^2 \tag{20}$$

By performing partial differentiation concerning parameters, setting it equal to zero, and then solving the two equations, we get:

$$\hat{\alpha}_{ols}=e^{\left(\frac{\bar{Y}}{\hat{\beta}}+\frac{\sum_{i=1}^{n}ln\ x_i}{n}\right)} \tag{21}$$

Substituting in the value of ($\hat{a}$) we get the function $g(\beta)$

$$g(\beta)=\sum_{i=1}^{n}Y_i\,Ln\ \ x_i+\beta\sum_{i=1}^{n}(Ln\ \ x_i)^2+\bar{Y}\sum_{i=1}^{n}Ln\ \ x_i\ -\frac{\beta}{n}\sum_{i=1}^{n}(ln\ \ x_i)^2 \tag{22}$$

As for the equation g(β), it is difficult to solve it by ordinary methods, because of the high non-linear degree in it, but by using one of the numerical methods to solve it, such as Newton-Raphson's method.

2.4. Goodness of fit criterions

Finding the appropriate and appropriate distribution of data is one of the basics in the process of analyzing data and finding results, and these criteria are used to determine whether they follow a certain probability distribution, as well as obtaining accurate information in the field of rainfall and the amount of depression in Iraq and these criteria are [2]:

$$AIC=-2LogL+2K \tag{23}$$

$$BIC=-2LogL+K\ log(n) \tag{24}$$

Where logL is a natural logarithm of the likelihood function and K represents the number of parameters, n is the sample size.

3. Results and discussion

Rainfall data were obtained from the Iraqi Meteorological and Seismic Monitoring Authority, as these data were recorded in the authority's stations distributed in the Iraqi governorates, where the data for the stations (Mosul, Kirkuk, Tikrit, Khanaqin, Rutba, Baghdad, Karbala, Najaf, Diwaniyah) were taken. Maysan, Basra) and that the inspection mechanism in these stations is carried out with one reading per month, and the unit of data measurement is (mm).To find the appropriateness of the distributions contained in the theoretical side of the rainfall data in the studied governorates, the good-matching tests mentioned on the theoretical side were used. Tables 1 and 2 show the results of estimations for the parameters of the Weibull distribution and the

Fréchet distribution using two estimation methods, namely, the Maximum likelihood method (MLM) and the least squares method (LSM).

Table 1. Weibull Distribution Parameters Estimators

City	*Methods*	*Weibull Distribution*	
		$\hat{\beta}$	$\hat{\alpha}$
Mosul	*MLM*	0.8120978	6.4821796
	LSM	1.00977	21.63656
Kirkuk	*MLM*	0.8293579	35.0147123
	LSM	0.7907966	40.3592225
Tikrit	*MLM*	1.368504	6.992495
	LSM	0.9770808	7.1584271
Khanaqin	*MLM*	1.314923	7.846441
	LSM	0.8776261	8.2707997
AL Rutba	*MLM*	1.129345	6.480127
	LSM	0.7760085	7.2207973
Baghdad	*MLM*	1.163829	7.458939
	LSM	0.8166443	8.0442702
Karbala	*MLM*	1.298732	7.642182
	LSM	0.9224426	7.7198501
Najaf	*MLM*	1.592454	2.507202
	LSM	1.431353	2.387897
Diwaniyah	*MLM*	1.566056	3.401532
	LSM	0.9770073	3.5791892
Maysan	*MLM*	1.533157	3.417710
	LSM	0.9770975	3.5792505
Basra	*MLM*	1.430606	3.301644
	LSM	0.9770605	3.5792278

Table 2. Fréchet Distribution Parameters Estimators

City	**Methods**	**Fréchet Distribution**	
		$\hat{\beta}$	$\hat{\alpha}$
Mosul	MLM	0.8120978	6.4821796
	LSM	1.009980	8.708746
Kirkuk	MLM	0.5957507	7.1217949
	LSM	0.7910488	12.6267262
Tikrit	MLM	1.136404	2.896736
	LSM	0.9771104	2.7938222
Khanaqin	MLM	1.057811	3.012745
	LSM	0.8776269	2.9017733
Al Rutba	MLM	1.122696	2.403545
	LSM	0.776007	2.208614
Baghdad	MLM	1.035279	2.661930
	LSM	0.8167523	2.6104700

City	Methods	Fréchet Distribution	
		$\hat{\beta}$	$\hat{\alpha}$
Karbala	MLM	1.045796	2.907406
	LSM	0.9226814	2.8509042
Najaf	MLM	2.102914	1.341390
	LSM	1.431426	1.256343
Diwaniyah	MLM	1.590155	1.642574
	LSM	0.9771315	1.3972572
Maysan	MLM	1.59996	1.65627
	LSM	0.9771315	1.3972572
Basra	MLM	1.624228	1.569639
	LSM	0.9771315	1.3972572

The good fit (GOF) tests that were mentioned in the theoretical aspect were used to find the best match for the probability distributions on the rainfall data for each of the Iraqi governorates, where the results were as follows:

Table 3. The results of Goodness of fit with rainfall data in the studied governorates

City	*Distributions*	*Methods*	*Goodness of fit criteria*	
			AIC	**BIC**
Mosul	Weibull	*MLM*	**565.5808**	**570.1062**
		LSM	567.7367	572.2621
	Fréchet	MLM	598.3941	602.9195
		LSM	622.1110	626.6364
Kirkuk	Weibull	*MLM*	**659.5740**	**664.0994**
		LSM	660.9513	665.4766
	Fréchet	MLM	691.6448	696.1701
		LSM	741.8439	746.3692
Tikrit	Weibull	*MLM*	**399.6647**	**404.1901**
		LSM	411.6156	416.1410
	Fréchet	*MLM*	418.1370	422.6624
		LSM	421.4186	425.9440
Khanaqin	Weibull	*MLM*	**420.1319**	**424.6573**
		LSM	436.1324	440.6578
	Fréchet	*MLM*	440.6081	445.1334
		LSM	445.2527	449.7781
AL Rutba	Weibull	*MLM*	**403.1952**	**407.7206**
		LSM	419.2542	423.7796
	Fréchet	*MLM*	403.8356	408.3609
		LSM	419.0860	423.6113
Baghdad	Weibull	*MLM*	**421.4782**	**426.0036**
		LSM	434.8867	439.4121
	Fréchet	*MLM*	432.5044	437.0298
		LSM	439.0931	443.6185
Karbala	Weibull	*MLM*	**417.2843**	**421.8096**
		LSM	428.7102	433.2355
	Fréchet	*MLM*	437.7915	442.3169

City	*Distributions*	*Methods*	*Goodness of fit criteria*	
			AIC	**BIC**
		LSM	439.9443	444.4697
Najaf	Weibull	*MLM*	238.1673	242.6927
		LSM	239.5572	244.0825
	Fréchet	*MLM*	**212.5377**	**217.0631**
		LSM	**227.6436**	**232.1689**
Diwaniyah	Weibull	*MLM*	285.4845	290.0099
		LSM	307.5057	312.0311
	Fréchet	*MLM*	**284.1503**	**288.6757**
		LSM	310.1597	314.6851
Maysan	Weibull	*MLM*	287.4215	291.9468
		LSM	308.2083	312.7336
	Fréchet	*MLM*	**283.7410**	**288.2664**
		LSM	310.6397	315.1651
Basra	Weibull	*MLM*	288.0308	292.5562
		LSM	304.8255	309.3509
	Fréchet	*MLM*	**276.4996**	**281.0249**
		LSM	302.3420	306.8674

Table 3 shows the values of the tests of goodness of fit results for each station in the Iraqi governorates. We note that the lowest value for all these tests appeared with the Weibull distribution by using the maximum likelihood method for the stations (Mosul, Kirkuk, Tikrit, Khanaqin, Rutba, Baghdad, Karbala), and this means that The Weibull distribution is the distribution corresponding to the data on rainfall in these stations, while the lowest values of the tests that appeared with Fréchet distribution and by using maximum likelihood method for the stations (Najaf, Diwaniyah, Maysan, Basra), and this means that the distribution of Fréchet is the distribution corresponding to the data of rainfall in these stations.

4. Conclusions

Through the presented results, the study concluded that the best probability distribution for representing rainfall data in the Iraqi governorates for the stations (Mosul, Kirkuk, Tikrit, Khanaqin, Rutba, Baghdad, Karbala) is the Weibull distribution using the maximum likelihood method, while the stations are in the governorates. The other (Najaf, Diwaniyah, Maysan, Basra) was the best distribution of Fréchet to represent the data of these stations using the method of estimating maximum likelihood as well. We also note the superiority of the maximum likelihood method over the least squares.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

[1] S.M. Salh and S.A. Ahmed, Statistical Distribution of Rainfall in Kurdistan-Iraq Region, *Al-Mustansiriyah Journal of Science*, vol.30, no.4, pp.18-28, 2019.

[2] A. A.-H.-M. Al-Wakeel, "Estimating the parameters of the best distribution of rainfall rates in Iraq", *Journal of Economic and Administrative Sciences*, vol.23, no. 15, 2017.

[3] I. A. Ibragimov and R. Z. Has’minskii, Eds., *Statistical Estimation*. New York, NY: Springer, 2014.

[4] D. H.-H. Badr, Comparing some methods of estimating the reliability of the two-parameter weibull distribution using simulations, *Journal of Al-Rafidain University College of Science*, vol.415, pp.50-65, 2019.

[5] A. S. Hassan, "The invers Weibull Generator of distribution: Properties and Applications", Journal *of Data Science*, vol.16, no.4, pp.723-741, 2018.

[6] I. A. Noman, "Designing Admission Sample Plans for the General Company for Electronic Industries using General Exponential Distribution", PhD thesis, College of Administration and Economics, Department of Statistics, University of Baghdad, 2012.

[7] N. A. Hussein, "The effect of atmospheric depressions in determining the amount of rain falling on Iraq during the period (2006-2014)", PhD thesis, College of Education, University of Diyala, 2017.

[8] B. F. Al-Saadi, "Some Parametric and Non-parametric Methods for Estimating the Reliability Function with Practical Application", Master Thesis, University of Baghdad, College of Administration and Economics, Department of Statistics, 2010.