AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

Selecting the best probability distribution of infection with MERS-COV in Wasit Governorate

SARAH ADEL MADHIOOM (1)

Middle Technical University

Suwaira Technical Institute

SHAYMAA QASIM MOHSIN (2)

BASRA UNIVERSITY

COLLEGE OF ADMINISTRATION AND ECONOMICS

Sarah adel@mtu.edu.iq

shaymaa.qassim79@gmail.com

HAYDER RAAID TALIB (3)

UNIVERSITY OF SUMER / COLLEGE OF ADMINISTRATION & ECONOMICS

haider.raid@uos.edu.iq

Abstract

This research aims to study the probability distribution of cases infected with Coronavirus in Wasit Governorate for the period from 2020 to 2021 using three types of probability distributions: the logistic distribution, the two-parameter Weibull distribution, and the least-valued Campbell distribution. Maximum, where the parameters of these distributions were estimated using the maximum likelihood method. Applying several criteria to determine the optimal distribution of cases infected with Coronavirus in the governorate, as these criteria were consistent Akaike (CAIC), Bayesian Akaike (BIC), and Akaike (AIC), and the probability distribution with the lowest value for these criteria is considered the best to represent a good model for studying this data.

Based on the applied part, the researchers concluded that the most appropriate probability distribution for coronavirus infection data in Wasit Governorate is the two-parameter Weibull distribution. The research is to provide a better understanding of the pattern of spread of the virus in the governorate, and to enable health officials to take more effective measures to limit the spread of infection in the future.

Keywords: Camble distribution, Weibull distribution, Logistic distribution, Maximum likelihood estimators.

Introduction

COVID-19, caused by a new (or novel) coronavirus, is a serious epidemic disease that spread rapidly around the world starting in December 2019. The spread of this disease has killed hundreds of thousands of people and infected many others, prompting Health officials in all countries should implement precautionary health measures to face this health challenge. Determining the pattern of virus spread and the temporal distribution of infections is crucial to understanding the epidemiological situation of the disease and taking the necessary preventive measures for that. This research aims to determine the most appropriate probabilistic distribution of cases of coronavirus infection in Wasit Governorate during the period from 2020 to 2021. This will be done using three different types of probability distributions (the logistic distribution, the two-parameter Weibull distribution, and the Camble distribution with the least limit).

Depending on the results and analyses, the optimal statistical model will be determined that accurately and reliably represents the temporal spread of the virus in the province. These results will be of great importance to the medical community and health officials to make proactive health decisions and develop effective plans to confront future risks of the virus in the province. The goal of the research will be achieved through Collecting data on coronavirus infections in the governorate during the specified period and using the method of greatest possibility to determine the optimal distribution of data and analyze it extensively. The results of the research will allow a better understanding of the spread of the disease and provide important information for making scientific and accurate decisions to address future health challenges associated with the coronavirus in the governorate.

1.1. The research problem

The lack of studies and statistical information available on cases of coronavirus infection in Wasit Governorate during the period 2020-2021 hinders the ability to determine the appropriate probabilistic distribution of the virus in the governorate, and this constitutes a major challenge for researchers wishing to better understand the spread of the disease and determine the optimal statistical model to accurately describe the phenomenon.

1.2. The search objective

The research aims to determine the most appropriate probability distribution of coronavirus cases in Wasit Governorate during the period 2020-2021 by using three types of probability distributions (logistical, two-parameter Weibull distribution, and Camble distribution with the least limit) and to determine the optimal statistical model that accurately describes the prevalence and temporal distribution. The research will contribute to providing a better understanding of the spread of the disease and providing health officials with valuable information to make effective health decisions to face future health challenges related to the virus in the governorate.

1.3. The previous studies

Ali (2011) conducted a study on the comparison of approximate estimation methods for the two parameters of the logistic distribution. He has compared them with approximate estimates derived mainly from the method of (White) in estimation, considering the logistic distribution as one of the exponential probability distributions, which are both the ordinary least squares method and the letter regression method (9).

Nuba et al (2015), presented a study on Gamble distribution applications using the R program, and concluded that when setting position and measurement parameters, the maximum likelihood estimation method (MLE) is better than the probability-weighted moment method (PWM) in parameter estimation, and their research was aimed at defining Gamble distribution methods for parameter estimation using the MLE method and the PWM method [11].

Alwakeel (2016), found the Mixed Weibull Distribution and concluded through the data used in the research that this distribution appeared, the research showed the existence of six parameters for the Weibull Distribution. In other cases, five or more parameters were found depending on the number of distributions involved and the number of parameters in each distribution [5].

Al-Quraishi (2018), presented a study aimed at estimating the reliability function of the two-parameter Whipple distribution, and four methods were chosen to estimate the reliability function, namely the method of MLE, the method of moments, the mixed method and the method of least squares, and the Monte Carlo method was applied in order to generate data using the program (Matlab) [3].

Raed (2018), presented a study on estimating the reliability of a three-parameter composite distribution using the method of MLE and the method of least squares using the standard of comparison (mean absolute errors). The study was applied to data from the General Company for Textile Industries in Wasit governorate. It was concluded that the best method of parameter estimation is the MLE method [8].

2. Theoretical framework

This chapter aims to identify some of the basic principles and definitions related to continuous distributions, including (logistical distribution, Whipple distribution, Campbell distribution) and also aims to estimate the parameters of distributions using the greatest possibility method, and the criteria for good fit tests will be clarified.

2.1. Logistical distribution

It means that it is one of the distributions of the exponential family, and it is often used in applications on the growth curve with a close and strong relationship to the subject of regression with a dependent variable in the two-response, and one of the important distributions that was taken from this transformation is the logarithmic distribution, which can be applied in the field of dependency functions equally Whether this distribution is used highly and one of the types of regression is called logistic regression [12] [9] :-

$$f(x) = \frac{e^{-\frac{(x-\alpha)}{\beta}}}{\beta\left(1 + e^{-\left(\frac{x-\alpha}{\beta}\right)^2}\right)} \quad ,0 \leq \alpha \leq \infty, \beta > 0 \;, -\infty < x < \infty, \qquad . \; . \; . \quad (2-1)$$

The cumulative distribution function is as follows:

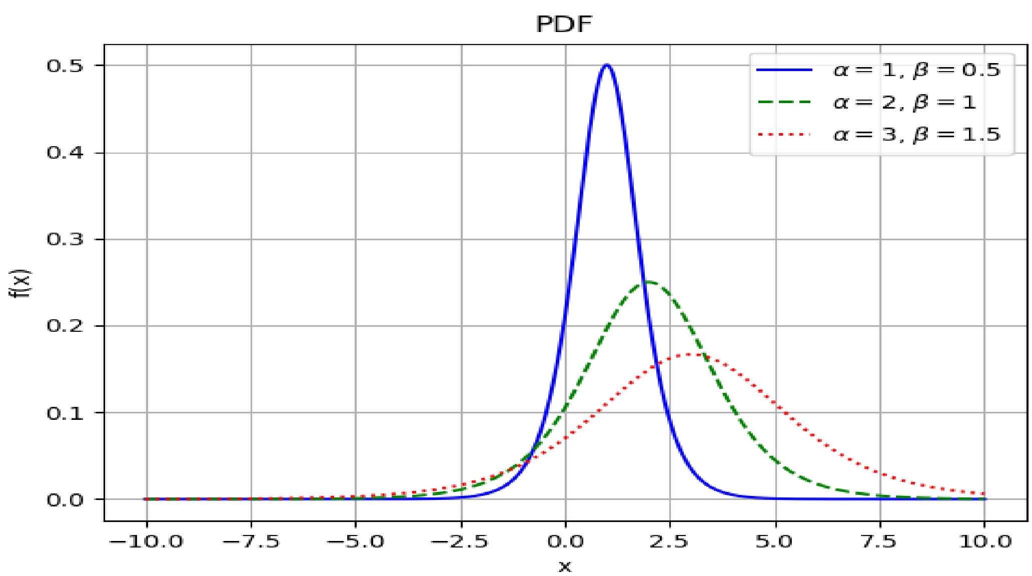
$$\mathbf{f(x \leq x) = \frac{1}{1+e^{\frac{-(x-\alpha)^2}{\beta}}}} \qquad . \ . \ . \ \ \mathbf{(2-2)}$$

As for the arithmetic mean and variance, the following formulas are taken [9]:

- ✓ Arithmetic mean $\boldsymbol{E(x) = \alpha}$

- ✓ Variance $\boldsymbol{Var(x) = \frac{\pi^2\beta^2}{3}}$

The following figure shows the different types of logistic distribution according to the location and measurement parameter:



As for the quantile function, it is as follows:

$$f^{-1} = \alpha + \beta \ln\left(\frac{p}{1-p}\right), p \in (0,1) \qquad . \ . \ . \ \ (2-3)$$

2.2. Weibull Distribution

The Whipple distribution is one of the continuous probability distributions and was named after the name of the mathematician (Whipple) in 1951 and was used for experimental presentation to see the change in the expansion of iron, which is one of the common models in reliability and in the distribution of survival times [1][6].

The two-parameter Weibull distribution function is [10]:

$$f(x) = \frac{\theta x^{\theta-1}}{c^{\theta}} e^{-\left[\left(\frac{x}{c}\right)^{\theta}\right]} \quad ,0 \leq x \leq \infty \qquad . \ . \ . \ (2-4)$$

whereas:

θ: scaling parameter.

C: shape parameter.

- ✓ The Probability Density Function (PDF):

$$f_i(x_i, c, \theta) = \begin{bmatrix} \frac{c}{\theta_i^c} x_i{}^{c-1} \exp[-(x_i/\theta_i)^c] & \\ o & 0.w' \end{bmatrix} \quad (x > 0, \theta > 0) \qquad . \ . \ . \ (2-5)$$

- ✓ The Cumulative Distribution Function (CDF):

$$f(x_i, c, \theta) = 1 - \exp\left[-(c\theta)^{\theta}\right], \quad c > 0, \theta > 0, \qquad . \ . \ . \ (2-6)$$

As for the arithmetic mean and variance of the distribution, it is as follows:

- ✓ Arithmetic mean

$$\boldsymbol{Mean = \theta r\left[1 + \frac{1}{c}\right]}$$

- ✓ variance

$$\boldsymbol{variance = \theta^2\left[r\left(\frac{2}{c} + 1\right) - r\left(\frac{1}{c} + 1\right)\right]^2}$$

2.3. Gumble Distribution (Gumble Min Distribution)

This distribution is considered to be one of the continuous distributions ("distributive value distribution"). It is named by a German mathematician named Emil Julius Gamble. When a random variable is distributed according to Gamble's distribution, the probability density function takes the following form [11]:

$$f(x) = \frac{1}{\sigma} \exp\left[\frac{x-\mu}{\sigma} \exp - \exp\frac{x-\mu}{\sigma}\right]; \quad x, \mu \in R, \sigma > 0 \qquad . \ . \ . \ (2-7)$$

whereas:

M: position parameter.

σ: scaling parameter.

✓ The Cumulative Distribution Function (CDF):

$$f(x)\exp\left[-\exp - \frac{x-\mu}{\sigma}\right] \qquad . \ . \ . \ (2-8)$$

As for the arithmetic mean and variance of the distribution, it is as follows:

✓ Arithmetic mean

$$\boldsymbol{E(x) = \mu + \sigma y}$$

Where y is a primary constant whose value is approximately 0.5572.

✓ variance

$$\boldsymbol{variance = \frac{1}{\sigma} \pi^2 \sigma^2}$$

3. Maximum likelihood estimation of distribution parameters

3.1. Estimation of logistic distribution parameters

To find the Maximum likelihood estimation (MLE) for the parameters α and β of the logistic distribution, based on the given probability density function, we find the possibility function first where the probability function is obtained as follows[12]:

$$f(x: \alpha, \beta) = \frac{e^{\left(\frac{x-\alpha}{\beta}\right)}}{\beta\left(1 + e^{-\left(\frac{x-\alpha}{\beta}\right)}\right)^2} \quad , -\infty < x < \infty \qquad . \; . \; . \quad (2-9)$$

The maximum likelihood estimators of the logistic distribution for each α, β is the solution of a system of rates that we obtain by partial derivation of (log likelihood) as follows:

$$\frac{dlnf}{d\varepsilon} = \frac{n}{\alpha} - \frac{\varepsilon}{\alpha}\sum_{i=1}^{n}\left[\frac{e^{-\left(\frac{x_i-\varepsilon}{\alpha}\right)}}{1 + e^{-\frac{(x_i-1)}{\alpha}}}\right] = 0 \;\; . \; . \; . \; (2-10)$$

$$= \sum_{i=1}^{n}\left(\frac{x_i - \varepsilon}{\alpha^2}\right) - \frac{n}{\alpha} - \frac{2}{\alpha^2}\sum_{i=1}^{n}\left[\frac{(x_i - \varepsilon)e^{\frac{-(x_i-\varepsilon)}{\alpha}}}{\left[1 + e^{\frac{-(x_i-\varepsilon)}{\alpha}}\right]}\right] = 0 \qquad . \; . \; . \; (2-11)$$

simplifying this system yields the following equations, which are estimates that can only be found numerically.

$$\frac{n}{2} = \sum_{i=1}^{n}\left[\frac{1}{1 + e^{\frac{-(x_i-\varepsilon)}{\alpha}}}\right] \qquad . \; . \; . \; (2-12)$$

$$\sum_{i=1}^{n}\frac{(x_i - \varepsilon)}{\alpha}\left[\frac{1 - e^{\frac{-(x_i-\varepsilon)}{\alpha}}}{1 + e^{\frac{-(x_i-\varepsilon)}{\alpha}}}\right] = n \qquad . \; . \; . \; (2-13)$$

We note that equations (12-2) and (13-2) are non-linear equations that cannot be solved by the usual methods. Therefore, we will use a numerical method to obtain the solution, and this method is the Newton Raphson method.

3.2. The Bi-Parametric Weibull Distribution

The probability density function for this distribution is [1][4]:

$$f_i(x_i, c, \theta) = \begin{cases} \frac{c}{\theta_i^c} {x_i}^{c-1} \exp[-(x_i/\theta_i)^c] & (x > 0, \theta > 0) \\ 0 & \text{ow} \end{cases} \qquad . \; . \; . \; (2-14)$$

It is clear that when (c = 1), the probability density function here transforms into a function of the exponential distribution, and that the maximum likelihood function for the bi-parameter Weibull distribution is:

$$L = c^{2n}\left[\pi_{i=1}^{2}\pi_{j=1}^{2}x_{ij}^{c-1}\right]\exp\left(-\sum_{i=1}^{2}\sum_{j=1}^{n}x_{ij}^{c}\Big/\pi_{i=1}^{2}\theta_{i}^{c}\right. \quad . \quad . \quad . \quad (2-15)$$

Then, taking the natural logarithm of this function and transforming it into a linear form:

$$\ln L = 2n\ln c + (c-1)\sum_{i=1}^{2}\sum_{j=1}^{n}\ln x_{ij} - \sum_{i=1}^{2}\sum_{j=1}^{n}\frac{x_{ij}^{c}}{\theta_{i}^{c}} - nc\sum_{i=1}^{2}\ln\theta_{i} \quad . \quad . \quad . \quad (2-16)$$

Then, to find the values of the estimates for each of the shape and size parameters θ, c where (i=1,2) and which render the potential maximum likelihood function, we need to calculate the maximum bounds of the function in the following way [7]. We take the partial derivative of the function with respect to the parameters (θ, c), setting the partial derivative equal to zero.

$$\frac{\partial \ln l}{\theta_i} = \frac{-n}{\hat{\theta}_i^{c}} + \sum_{i=1}^{2}\frac{\sum_{i=1}^{n}x_{ij}^{c}}{\theta_i^{2c}} = 0$$

$$\frac{n}{\theta_i^{c}} = \frac{\sum_{j=1}^{n}x_{ij}^{c}}{\theta_i^{2c}}$$

$$\hat{\theta}_i^{c} = \frac{\sum_{j=1}^{n}x_{ij}^{c}}{n}$$

$$\hat{\theta}_i = \left[\frac{\sum_{j=1}^{n}x_{ij}^{c}}{n}\right]^{1/\hat{c}}$$

$$\hat{\theta}_1 = \left[\sum_{j=1}^{n}\frac{x_{1j}^{c}}{n}\right]^{1/\hat{c}}$$

$$\hat{\theta}_2 = \left[\sum_{j=1}^{n}\frac{x_{2j}^{c}}{n}\right]^{1/\hat{c}}$$

Furthermore, the partial derivative of the function is relative to the measurement parameter [2]:

$$\frac{\partial \ln l}{c} = \frac{2n}{\hat{c}} + \sum_{i=1}^{2}\sum_{j=1}^{n} \ln x_{ij} - \sum_{i=1}^{2}\sum_{j=1}^{n} \frac{x_{ij}^{c} \ln\left(x_{ij}/\theta\right)}{\hat{\theta}\hat{c}} \quad . \ . \ . \ (2-17)$$

Due to its high degree of non-linearity, this equation cannot be solved by the usual methods. It can therefore be solved using one of the numerical methods for solving non-linear equations, such as the (Newton-Raphson) method:

$$\hat{c}_j = \hat{c}_{j=1} - \frac{g\left(\hat{c}_j - 1\right)}{\acute{g}\left(\hat{c}_j - 1\right)} \text{ if } \acute{g}\left(\hat{c}_j - 1\right) \neq 0$$

$$g(\hat{c}) = \frac{\sum_{j=1}^{n} x_j^{c} \ln x_j}{\sum_{j=1}^{n} x_j^{c}} - \frac{1}{\hat{c}} - \frac{\sum_{j=1}^{n} \ln x_j}{n} \quad . \ . \ . \ (2-18)$$

$$\acute{g}(\hat{c}) = \frac{\partial g(\hat{c})}{\hat{c}} = \frac{\sum_{j=1}^{n} x_j^{\hat{c}} \sum_{j=1}^{n} x_j^{\hat{c}} \left(\ln x_j\right)^2 - \left(\sum_{j=1}^{n} x_j^{\hat{c}} \ln x_j\right)^2}{\left(\sum_{j=1}^{n} x_j^{\hat{c}}\right)^2} + \frac{1}{\hat{c}^2} \ . \ . \ . \ (2-19)$$

$$\hat{c} = \left[\frac{1}{2}\left\{\sum_{i=1}^{2}\sum_{j=1}^{n} x_{ij}^{\hat{c}} \ln x_{ij}\right\}\left\{\sum_{j=1}^{n} x_{ij}^{\hat{c}}\right\} - \frac{1}{2^n}\sum_{i=1}^{2}\sum_{j=1}^{n} \ln x_{ij}\right] \ . \ . \ . \ (2-20)$$

The above equation is solved by an iterative procedure to estimate the parameter of the form c, i.e. c ̂.

3.3. Estimating the parameters of the Gumble Min Distribution

The probability density function is defined by the following relation [11] [1]:

$$f(x) = \frac{1}{\sigma} \exp\left(\frac{x-\mu}{\sigma}\right) \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right], \quad x, \mu \in R, \ \ \sigma > 0 \ . \ . \ . \ (2-21)$$

We find that the bulk density function of the Campbell value distribution is the following:

$$f(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]; \quad x, \mu \in R \ \ ; \sigma > 0, \ . \ . \ . \ (2-22)$$

Its probability density function is:

$$f(x) = \frac{1}{\sigma}\exp\left[-{}^{1}/_{\delta}\,(x-\mu)\right]\exp\left[-\exp-\left(\frac{x-\mu}{\delta}\right)\right]\ ;\ \ x,\mu \in R \quad ,\delta > 0\ .\ .\ .\ (2-23)$$

Where x is a variable (x= x_1,x_2,x_3,.........x_n), for each (x∈n) a complete sample taken from a population with a Gumble distribution.

The maximum likelihood function for these complete data is given by the following relationship:

$$L = n\ln\frac{1}{\sigma^2}\exp\left[-\frac{1}{\sigma}\sum_{i=1}^{n}(x_i-\mu)\right]\ \exp\left[-\sum_{i=1}^{n}\exp-\left(\frac{x_i-\mu}{\sigma}\right)\ .\ .\ .\ (2-24)\right]$$

If we take the natural logarithm, we obtain:

$$\ln l = n\ln\sigma - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left(\frac{x_i-\mu}{\sigma}\right)\ .\ .\ .\ (2-25)$$

And by taking the derivative with respect to the position and measurement parameters, as follows:

$$\frac{\delta\ln l}{\delta u} = \frac{n}{\sigma} - \frac{1}{\sigma}\sum\exp-\left(\frac{x_i-\mu}{\sigma}\right)\ .\ .\ .\ (2-26)$$

$$\frac{\delta\ln l}{\delta b} = \frac{-n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) - \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\exp-\left(\frac{x_i-\mu}{\sigma}\right)\ .\ .\ .\ (2-27)$$

$$\frac{n}{\hat\sigma} - \frac{1}{\hat\sigma}\sum_{i=1}^{n}\exp-\left(\frac{x_i-\hat\mu}{\hat\sigma}\right) = 0$$

Equalization of the two parameters to zero gives:

$$-\frac{n}{\hat\sigma} + \frac{1}{\hat\sigma^2}\sum_{i=1}^{n}(x_i-\hat\mu) - \frac{1}{\hat\sigma^2}\sum_{i=}^{n}(x_i-\mu)\exp-\left(\frac{x_i-\hat\mu}{\hat\sigma}\right) = 0,\ .\ .\ .\ (2-28)$$

Simplifying, we obtain:

$$n - \exp\left(\frac{\mu}{\hat{\sigma}}\right)\sum_{i=1}^{n}\exp\left(\frac{x_i}{\hat{\sigma}}\right) = 0$$

$$n\hat{\sigma} + \sum_{i=1}^{n}(x_i - \mu) - \sum_{i=1}^{n}(x_i - \hat{\mu})\exp\left(\frac{x_i - \mu}{\hat{\sigma}}\right) = 0, \quad . \ . \ . \ (2-29)$$

Simplifying the above equation and taking the natural logarithm of the two sides, we obtain:

$$\hat{\mu} = -\hat{\sigma}\left[\ln\sum_{i=1}^{n}\left(\frac{\exp-\left(\frac{x_i}{\hat{\hat{\sigma}}}\right)}{n}\right)\right] \qquad . \ . \ . \ (2-30)$$

We also find that:

$$-n\hat{\sigma} + \sum_{i=1}^{n}x_i - n\hat{\mu} - \sum_{i=1}^{n}x_i\exp-\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) + \hat{\mu}\sum_{i=1}^{n}\exp\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) = 0, \quad . \ . \ . \ (2-31)$$

From the above equation and after substitution and abbreviations, we obtain the following relationship:

$$\hat{\sigma} = \frac{\sum_{i=1}^{n}x_i}{n} - \frac{\sum_{i=1}^{n}x_i\exp-\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)}{\exp\left(\frac{\hat{\mu}}{\hat{\hat{\sigma}}}\right)\sum_{i=1}^{n}\exp-\left(\frac{x_i}{\hat{\sigma}}\right)} \quad . \ . \ . \ (2-32)$$

$$\hat{\sigma} = \bar{x} - \frac{\sum_{i=1}^{n}x_i\exp-\left(\frac{x_i}{\hat{\hat{\sigma}}}\right)}{\exp\left(\frac{\hat{\mu}}{\hat{\hat{\sigma}}}\right)\sum_{i=1}^{n}\exp-\left(\frac{x_i}{\hat{\hat{\sigma}}}\right)} \qquad . \ . \ . \ (2-33)$$

Since the above relationship is in the form (g(x)= x), both estimators can be found using the fixed-point iteration method, noting that both estimators are unique for each sample:

$$n = \exp\left(\frac{\mu}{\hat{\sigma}}\right)\sum_{i=1}^{n}\left(-\frac{x_i}{\hat{\sigma}}\right) = 0 \qquad . \ . \ . \ (2-34)$$

4. Goodness of Fit (GOF)

The question of the appropriate distribution of data is one of the important topics, as it is confronted with numerous criteria that can be used to determine whether the data used follow a particular probability distribution. In order to obtain accurate information on the distribution of coronavirus infection data in Wasit, we used three types of goodness-of-fit criteria, which were applied to coronavirus infection data [5].

4.1. Akaike information criterion (AIC)

The general form of this test is the following:

$$\mathrm{AIC} = -2\log \mathrm{l} + 2\mathrm{k} \quad . \quad . \quad . \quad (2-35)$$

log L: logarithm of the model

K: number of parameters in the distribution function

4.2. Conditional Akaike Information Criterion (CAIC)

$$\mathrm{CAIC} = -2\log \mathrm{l} + \frac{2\mathrm{nk}}{\mathrm{n}-\mathrm{k}-1}, \quad . \quad . \quad . \quad (2-37)$$

log L: logarithm of the model

K: number of parameters in the distribution function

n: sample size.

5. Practical framework

This section gives concrete form to what has been discussed in the theoretical aspect and in the context of the research objective. We compared the three distributions (The Bi-Parametric Weibull Distribution, the Gumble distribution and the logistic distribution) and applied them to data relating to cases of coronavirus infection in Wasit governorate, extracted from the Iraqi Ministry of Health's health bulletins for the year (2020-2021), to determine whether the distributions presented in the theoretical aspect are suitable. For the Corona virus infection data, goodness-of-fit tests that are mentioned in the theoretical aspect according to the program (Matlab) were used.

5.1. Parameter estimation results

We present here the results obtained by estimating the characteristics for each of the distributions estimated using the maximum likelihood method (MLM), in order to compare them with each other.

Table (1) the results of parameter estimation for each distribution

City	Distribution	Parameters	The value
Wasit	logistic	α	128.34
		β	58.48
	Weibull2	C	165.06
		θ	1.37
	Gumbel Min	μ	221.00
		σ	176.56

Table (1) shows the results of parameter estimation for the logistic distribution, the Gumble Min distribution and the two-parameter Weibull distribution, using the maximum likelihood method (MLM) for parameter estimation.

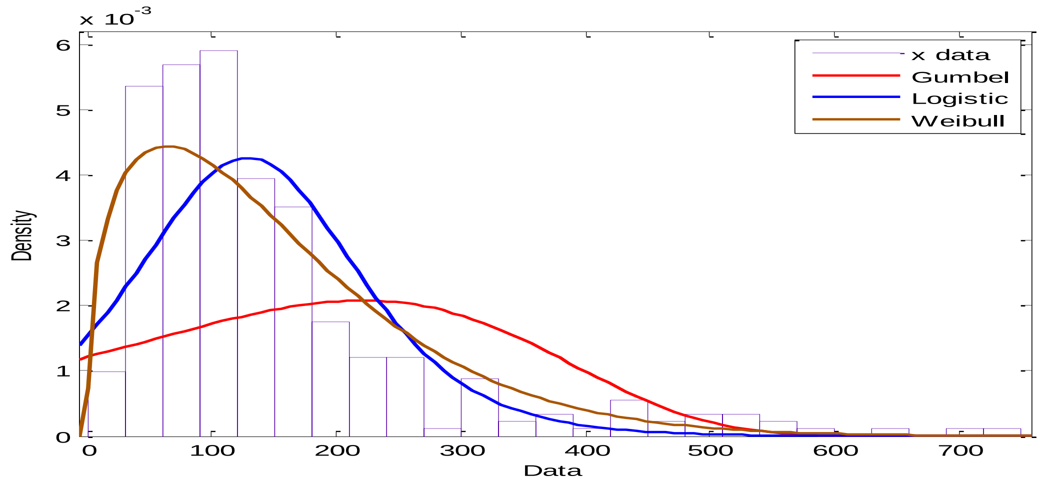


Figure (1): The probability density function of distributions.

From the above figure, we see clearly that the best distribution to represent incidence data is the two-parameter Weibull distribution, as we can see from the above figure that it is the closest to the normal distribution, so it best represents the data.

5.2. Goodness-of-fit results

The goodness-of-fit (GOF) tests mentioned in the theoretical section were used to find the most appropriate probability distributions to represent the incidence data. In table (2), the results of the criteria have been included for each distribution.

Table (2) Results of goodness-of-fit criteria for coronavirus infection data

Distribution	Goodness Fit		
	AIC	BIC	CAIC
logistic	3727.30	3729.02	3724.30
Weibull(2para)	3608.24	3609.96	3605.24
Gumbel Min	4003.71	4005.42	4007.71

Table (2), which represents the results of the goodness-of-fit tests for the coronavirus infection data, shows that the lowest value of the goodness-of-fit tests is obtained with the two-parameter

Weibull distribution (Weibull(2para)). Thus, the two-parameter Weibull distribution is the best distribution to represent coronavirus infection data in Wasit governorate.

6. Conclusions and recommendations

6.1. Conclusions

Our results allow us to conclude the following:

- ✓ Three types of probability distributions were studied from continuous distributions, namely the logistic distribution, the two-parameter Weibull distribution and the **Gumble Min Distribution**, and the necessary mathematical derivations that were discussed in the estimation methods were performed in order to obtain the parameters of the distribution.
- ✓ The results of goodness-of-fit (GOF) tests for coronavirus infection data in Wasit governorate showed that the best probability distribution for coronavirus infection data in Wasit governorate is the two-parameter Weibull distribution (Weibull,2para).

6.2. Recommendations

Through what has been presented in the Theoretical and Applied Framework, we recommend the following:

- ✓ Using continuous distributions to represent the data of infection with the Corona virus and compare it with the results obtained in this research.
- ✓ Using the two-parameter distribution to study other random phenomena.
- ✓ Using other estimation methods to find distribution estimators, and we also recommend research into numerical algorithms to find systemic solutions to the non-linear equations resulting from the maximum likelihood method.

References:

1- Ali, O. A. M. (2011). to review methods that used to estimate Logistic distribution parameters. An exact estimators method which is the Moment method, compared with other approximate estimators obtained essentially. *Journal of Economics and Administrative Sciences*, *17*(62), 273–283.

2- Al-Nasser, A. al-M. H., & Al-Bayati, H. N. A. (2005). Comparison Berween the Methods of Estimation Weibull Distribution Using Simulation. *Iraqi Journal of Statistical Sciences*, *7*. Retrieved from http://search.mandumah.com/Record/866538

3- Al-Quraishi, O. A.-A. (2019). *Estimating the Reliability function for monitored data subject to a Weibull distribution.* (MA thesis). University of Karbala - College of Administration and Economics.

4- Al-Sarraf, N. M., Alrawi, A. G., & Kamal, G. I. (2016). *Statistical Estimation*. Al-Jazeera Printing and Publishing Office.

5- Al-Tawash, M. S. (1990). *A statistical study on Webull and extreme value distribution with applied applications* (MA thesis). College of Administration and Economics - University of Baghdad.

6- Al-Wakil, A. A.-H. (2016). *Finding the Mixed Weibull Distribution*. College of Administration and Economics - University of Baghdad.

7- Al-Wakil, A. A.-H., & Al-Moussawi, N. S. (2017). Estimate the distribution parameters for the best rates of rainfall in Iraq. *Journal of Economics and Administrative Sciences*, *23*(100), 473–489.

8- Erkut Tekeli, Güzin Yüksel. (2022) Estimating the parameters of twofold Weibull mixture model in right-censored reliability data by using genetic algorithm. Communications in Statistics - Simulation and Computation 51:11, pages 6621-6634

9- Hussein, A. N. (2015). *Finding the Efficient Algorithm in Estimating the Parameters of the Mixed Whipple Distribution (An Application to Wind Speed in Iraq* (Ph.D). College of Business and Economics - University of Baghdad.

10- Kamar, S. A. H., & Abood, H. N. (2016). The use of a simulation for comparing among the estimations of parameters the function of whipple distribution according to greatest possible and base methods. *ECONOMICS AND ADMINSTRATION STUDIES JOURNAL (EASJ) (Formerly AL-DANANEER JOURNAL)*, *1*(9), 531–540.

11- Nouba, E. A., & Jabbar, Y. M. (2016). *umble distribution applications using R*. Sebha University - Faculty of Science - Department of Statistics. Retrieved from http://dspace.sebhau.edu.ly//handle/1/1238

12- Publications of the World Health Organization. (2020). Retrieved January 1, 2020, from https://www.who.int/publications.

13- Singh, S., & Singh, U. (2021). Estimation of parameters of the logistic exponential distribution under progressive type-I hybrid censored sample. Journal of Statistical Theory and Applications, 20(3), 337-354

14- Talib, H. R. (2017). sing some methods to estimate the parameters and reliability of the complex probability model (exponential-Webble) with practical application. *IRAQI JOURNAL FOR ADMINISTRATIVE SCIENCES*, *13*(52), 246–268. Retrieved from https://www.iasj.net/iasj/article/154133