

ANALYSIS OF PROBABILITY DISTRIBUTION FOR CORONA VIRUS MORTALITY IN IRAQ BEFORE AND AFTER THE DISCOVERY OF THE EPIDEMIOLOGICAL VACCINES

Abbas Gulmurad Beg Murad

Sulamani University, Collage of Administration and Economics, Department of Statistics and Informatics, Email: abbas.beg@univsul.edu.iq

Sozan Saber Haider Ali

Sulamani University, Collage of Administration and Economics, Department of Statistics and Informatics, Email: sozan.haider@univsul.edu.iq

ABSTRACT

Modern times created modern problems and many of those problems have data, In this research, we try to analyze the behavior of the Mortality data distribution resulting from the Corona epidemic probabilistically before and after the discovery of vaccines globally and their distribution in Iraq of different types of vaccines using some statistical tools, which can have a significant impact on increasing immunity and reducing the mortality rate of healthy people and people with infection, and thus changing the shape of the probability distribution of mortality On a daily basis, and reduce the proportion of deaths resulting from this epidemic by knowing the probability distribution with the best fit of the probability represents the data that has wide uses in survival analysis to describe the incidence of deaths before and after the discovery of vaccines with high fitness value using the method of estimating the maximum likelihood of Appropriate parameter estimating by using Anderson-Darling distance to reduce the distance, insides use some other statistical techniques to analyze the differences between mortality during the two years 2020 and 2021 (before and after vaccines discovery) depend upon programming language r and rely on a set of packages, the most important of which is (fitdistrplus package) and Other statistical functions in R language insides easy fit program for Iraqi mortality time series analysis from Corona virus.

Keywords: Corona Virus Mortality, Johnson SB distribution, Beta distribution, Dagum (4p) distribution, Anderson Darling test, fitdistrplus package.

1. Introduction

To address random phenomena and data that arise in many applied problems in the life sciences, probability distributions for medical science can be applied to make informed predictions and decisions under uncertain conditions. Therefore, the statistical treatment of these data is an important aspect of their analysis and interpretation. in this paper we apply various probabilistic models to characterize these data. Therefore, a better selection of the most appropriate probability distribution can help extrapolate the observations to the most important values.

The quantification and management of deaths is a major challenge for health care workers to know the rise, cause or distribution of deaths. It is particularly difficult to deal with severe

mortality events that come from a rapidly spreading epidemic that can have extreme economic consequences., but there is at least a viable modeling tool for how it can be used to measure the distribution of mortality. (30)

Predictive decisions are often data driven. These prediction rules are simple if the data contains consistent, clear information, but this information is often represented by differences. Statistics provides tools for describing the variety of data and making informed decisions based on the data.

Quantitative data can be described using the main properties of location or shape, and scale measurements. The shape of the data distribution can be described as symmetric, skewed, flat, or bell-shaped, and can be determined by its statistical center (such as mean or median) and statistical distribution (such as normal). Variances or range ...etc.). These statistics can be used to compare different distributions numerically or visually using graphs. Knowing the center and distribution is not enough to describe the distribution. The statistics you compare, the scenarios you use, and what the results of the comparison mean depend on the investigating question and the specific action taking.

Randomness can be said to have two main uses in deriving statistical results. The first is to draw valid conclusions about the entire community by gathering data from a random sample of the population. This can only be assessed under randomized conditions. It is important to consider the study design, how the data were collected, the analyzes used, and their summary A probabilistic model is a description of possible events, In a probabilistic model, sample points represent outcomes and clusters represent events. (15)

To understand the global impact of the COVID-19 pandemic, the number of mortality from this pandemic for 2020 and 2021 are determined. Vaccine availability is limited until 2021, so it is necessary to study the effect of the vaccine on mortality in this analysis as well as the probability distribution Mortality behavior in Iraq before and after vaccine availability. (14)

probability of a particular event occurring can be determined by evaluating the occurrence of the same event in previous observations under some conditions as we have seen this is an interpretation of what will happen and is based on the relative nature of the events observed, in the previous analysis (1).

For more explanation Insurance companies need mortality estimates in older ages for annuities and mortgages, for example. Most workers use Johnson SB distribution, Beta distribution, Gompertz, Makeham, ..etc, and other parametric models to fit mortality accounts.

So, some models cannot optimally explain deaths from COVID-19, the local demand for a fully comprehensive sample of data in Iraq is often also required to make inferences across the entire data range using the best data representation model.(2)

2. COVID-19 pandemic

The Covid-19 infections speed rise in Wuhan City in Hubei Province in China caused to thousands of deaths (24). patients at high risk. We identified clinical predictors of mild and severe patient outcomes (26).

Health crises need for effective data collection and dissemination. The data is updated by region according to age or cause of death. This information can be quickly collected and

processed by health information systems. Where more resources should be used to develop better action plans and tools at the national as well as regional levels. By intervening with effective preventive measures to control the spread of the virus, the epidemic can be confronted throughout the country and this includes Iraq. Another reason it is difficult to ascertain the long-term effects of the epidemic is that changes vary from country to country, as do health services and vaccines available to all (27).

3. An Introduction to probability fundamentals

The probability of a particular event can be determined by evaluating the occurrence of the same event under conditions as similar as possible to those observed [this is a frequentist definition of probability, the probability of an observed event Based on relative frequency observed in the previous conditions (3)]. In other words, probability describes the likelihood of an event occurring given a set of circumstances. This is a form of reasoning, a way of predicting what will happen based on what has happened before in the same (but not exactly the same) situation. Probabilities range from 0 (the predicted event has never been observed and should not occur) to 1 (or 100%, the event is almost certain).(4)

$$\sum P(X = x) = 1 \quad \dots (1)$$

Kolmogorov described it as one of the three axioms of probability.(5)

Probabilities can be described using a functions or graph where each event is associated with its probability of occurrence. This describe of these probabilities is called a probability distribution. (6)

It is better to think of the possibility in another way. We use probability theory to talk explicitly and quantitatively about the degree of certainty or uncertainty we have about a question. In other words, if we develop a theory about how perfectly rational a person will be sure of a given outcome, we end up with something very similar to probability theory. (7)

We've just finished our introduction to the most important ideas in probabilistic models. Never lose sight of this: we will build a model for a set of coronavirus deaths and assign the distribution of key events to that space. We will almost certainly map this distribution by some simpler distribution assigned to a simpler space.(8)

3.1. Probability & Probability Distributions (8,9)

In the early stages of development of probability theory, there were various definitions and methods of computing probability, including classical probability, geometric probability, and frequency. In 1933, Kolmogorov established a system of axioms of probability theory based on measurement theory, and laid the foundation for modern probability theory. The axioms of probability theory: Let Ω be the set of points ω and F the set of subsets A of Ω . The letter F is called σ -algebra for if it meets the following conditions:

- (i) $\Omega \in F$;
- (ii) if $A \in F$, then its complement set $A^c \in F$;
- (iii) if $A_n \in F$ for $n = 1, 2, \dots$, then $\coprod_{n=1}^{\infty} A_n \in F$.

Let $P(A)(A \in F)$ be a real function valued in σ -algebra Fields, The Probability satisfies:

- (1) $0 \leq P(A) \leq 1$ for every $A \in F$;

(2) $P(\Omega) = 1$;

(3) $P(\prod_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ holds for $A_n \in F$, $n = 1, 2, \dots$, where $A_i \cap A_j = \emptyset$ for $i \neq j$, and \emptyset is the empty set.

Then P is the measure of probability, or probability, of F . Also, the set of F is called an event, and (Ω, F, P) is called the probability space.

A random variable X can represent values over periods and is called a continuous random variable if the probability of X falling within a sub-interval is constant. For a continuous r.v. X , if there is a non-negative integral function $f(x)$, then $P\{a \leq X \leq b\} = \int_a^b f(x)dx$, holds for $-\infty < a < b < \infty$, and $\int_{-\infty}^{\infty} f(x)dx = 1$, $f(x)$ is called the density function of r.v. (x). then if $F(x) = P\{X \leq x\}$ for $-\infty < x < \infty$, then $F(x)$ is called the distribution function of X .

When X is a discrete r.v. its distribution function is $F(x) = \sum_{i: x_i \leq x} p_i$ like when X is a continuous random variable, distribution of it is:

$$F(x) = \int_{-\infty}^x f(t)dt. \quad \dots (2)$$

Much work has been done to discover and improve probabilistic mortality models that represent population mortality.

3.2. Connections to Functions and Modeling (12)

The functions used to describe data, If the behavior of the data show a linear relationship, this relationship can be represented by a regression line, and the strength and direction of the correlation coefficient.

It is better to define modeling alongside other criteria rather than as an isolated issue. Mathematical modeling is a standard of mathematical practice, and specific modeling standards are marked with an asterisk at each high school

A statistically significant result is not due to chance and can be checked under random conditions. The conditions under which the data were collected are important in drawing conclusions from the data, it is important to consider the study design, data collection methods, applied analyses, data summaries, and conclusions. and so on.

Stochastic processes can be described mathematically using stochastic models. A probabilistic model is a list or description of sample space, each assigned a probability. Whether flipping a coin, rolling a dice, or drawing a card, it would be reasonable to assume that the outcomes would be equal. In probabilistic models, sample points represent outcomes that together form an event. The interpretation of these probabilities depends on the interpretation of the independent and conditional probabilities.

3.3. Probability distributions in survival analysis (19)

The age of the individual is described by a non-negative variable T , with support for a subset of true positive numbers $[0, +\infty)$. Suppose $F(x)$ is the cumulative T function of the variable, which is provided by the following and describes the behavior of the random variable (T):

$$F(x) = P(T \leq x), \quad x \geq 0$$

The probability that an individual will survive x years is called the survival function of T , denoted by $S(x)$ and given by the relation:

$$S(x) = P(T \leq x) = 1 - F(x) \quad \dots (3)$$

With conditions:

$$S(x) = 1 \ \& \ \lim_{x \rightarrow \infty} S(x) = 0.$$

... (4)

Mortality model selection One of the mainly objectives of this paper is the description of the mechanisms that describes mortality (The mortality models), as previously mentioned and explained as follows,

3.4. Johnson-SB Distribution function (16)(17)

Johnson proposed the Johnson system In 1949 called the curve system consisting of Johnson-SB, Johnson-SL and Johnson-SU. The symbol SL stands for "common logarithmic system", SB stands for "finite system" and SU stands for "infinite system". The Johnson system can be approximated by several continuous distributions in one of three functional forms, making it very flexible to accommodate different curves. Many commonly used continuous distributions, such as normal, logarithmic, gamma, beta, and exponential, are special cases of the Johnson system; Therefore, curve fitting to the Johnson system has more advantages than any other single distribution.

Johnson-SB is based on the three divisions of the Johnson system. The skew-elastic properties presented in this paper are useful for comparing distribution skew with corona mortality. It corresponds to the distribution of a continuous random variable x where a transformation is used to obtain a normally distributed transformation such as:

$$Z = V + \delta \ln\left(\frac{X-\mathcal{E}}{\lambda + \mathcal{E}-x}\right) \quad \dots (5)$$

where x is a given continuous r.v. $\in (\mathcal{E}, \mathcal{E} + \lambda)$, $\mathcal{E} = \min(x)$, $\lambda = \max(x) - \min(x)$, and V and δ are shape parameters, $\delta > 0$, $V \in (-\infty, +\infty)$. Z is a standard normal r.v., and its p.d.f. as follows:

$$P(z) = \frac{1}{\sqrt{2P}} e^{-z^2/2} \quad \dots (6)$$

Where:

$$Z = V + \delta f(y) \quad \dots (7)$$

And:

$$y = \left(\frac{X-\mathcal{E}}{\lambda + \mathcal{E}-x}\right)$$

Then:

$$f(y) = \ln\left(\frac{X-\mathcal{E}}{\lambda + \mathcal{E}-x}\right) \quad \dots (8)$$

Since (z) is the inverse function of y . According to the continuous random variable transform as follow:

$$p(y) = \delta \hat{f}(y) p(z) = \frac{\delta}{\sqrt{2P}} \hat{f}(y) e^{-1/2[V+\delta f(y)]^2} \quad \dots (9)$$

Below is a typical Johnson-SB probability density function curve with different parameters. the is the horizontal index from 0 to 255, and the vertical index is the value corresponding to the probability density function. As Fig. 3(a) shows curves with different δ and $V=0$, and this mean | Fig. 3(b)~3(e) shows curves δ and $V=0$ separately, respectively, each curve. have better characteristics especially on both sides As we can see, c controls the position of the function and the distribution of the function is a normal distribution. Negatively skewed distribution and positively skewed distribution when $V=0$, $V > 0$ and $V < 0$ respectively. (18)

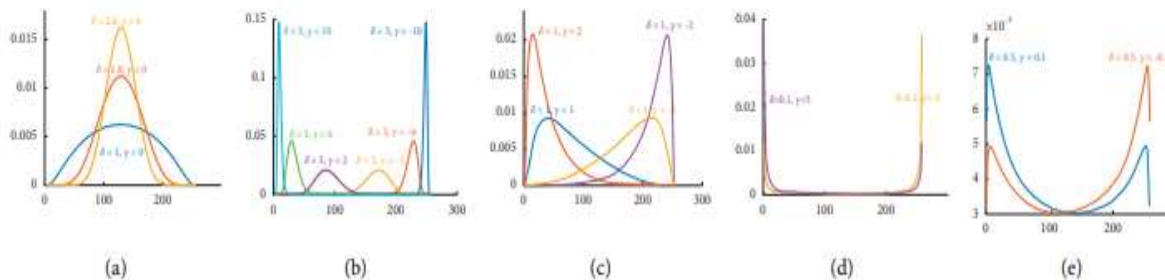


Fig.(1): Pdf of (Johnson-SB Model) curves with different parameters (18)

3.5. Beta Distribution function (13)

The beta distribution is a set of continuous probability distributions defined in interval $[0, 1]$ with two positive shape parameters represented by α and β . It controls the various features and distribution structure.

A common use of this distribution is to model uncertainty about the probability of success in a random experiment. In project management the uses a three-point method called the "beta distribution" that describes the uncertainty in project time estimates. There are powerful statistical tools combined with special statistics to calculate confidence levels for expected times of completion.

3.5.1. Properties of Beta Distribution (9, 10,11)

The Pdf of the r.v. (x) is:

$$f(x; a, b) = f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{if } 0 < x < 1 \\ 0 & \text{Otherwise} \end{cases}, \quad \dots (10)$$

When $a > 0, b > 0$, then we say that X follows the Beta distribution with parameters a and b, and show that $(x) \sim B(a, b)$.

$$\text{Where: } B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad \dots (11)$$

$$\text{and } E[X] = a/a + b \text{ and } \text{Var}(X) = \frac{ab}{(a+b)^2 (a+b+1)}.$$

$$BE(x; a, b) = \begin{cases} 1, & x > 1, \\ I_x(a, b), & 0 < x \leq 1, \\ 0, & x \leq 0, \end{cases} \quad \dots (12)$$

Since $I_x(a,b)$ is the ratio of the incomplete beta function. Like the gamma distribution, the name of beta distribution because of the shape of the function it is similar to the beta function. In

particular, if $a = b = 1$, then the standard uniform distribution is $U(0, 1)$, there are Some characteristics of a beta distribution as follow:

1. If $X \sim B(a, b)$, then $(1-X) \sim B(b, a)$.
2. The Pdf of beta distribution has the following: the density function is monotonically decreasing when $a < 1, b \geq 1$, if $a \geq 1, b < 1$, the density is monotonic increasing, the density function curve is U-shaped When $a < 1, b < 1$, the density function curve may be peaked If $a > 1, b > 1$, the density function curve is symmetric about $x = 1/2$ When $a = b$.
3. The k -th moment of X is $E(X^k) = \frac{B(a+k, b)}{B(a, b)}$ If $X \sim B(a, b)$.
4. The expectation and variance of X are $E(X) = \frac{a}{(a + b)}$ and $\text{Var}(X) = \frac{ab}{((a + b + 1)(a + b)^2)}$, respectively If $X \sim B(a, b)$.
5. The skewness of X is $s = \frac{2(b-a)(a+b+1)^{1/2}}{(a+b+2)(ab)^2}$ and the kurtosis of X is $\kappa = \frac{((3(a+b)(a+b+1)(a+1)(2b-a)) / ab(a+b+2)(a+b+3)) + (a-a-b) / a+b - 3}{(a+b+2)(ab)^2}$ If $X \sim B(a, b)$.
6. The moment-generating function and the characteristic function of X are $M(t) = \frac{\Gamma(a+b)}{\Gamma(a)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(a+b+k)} \frac{t^k}{\Gamma(k+1)}$ and $\psi(t) = \frac{\Gamma(a+b)}{\Gamma(a)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(a+b+k)} \frac{it^k}{\Gamma(k+1)}$, respectively If $X \sim B(a, b)$.

It is also possible to estimate a and b using maximum likelihood, but the method requires an iterative procedure and hence is more computationally intensive. (13)

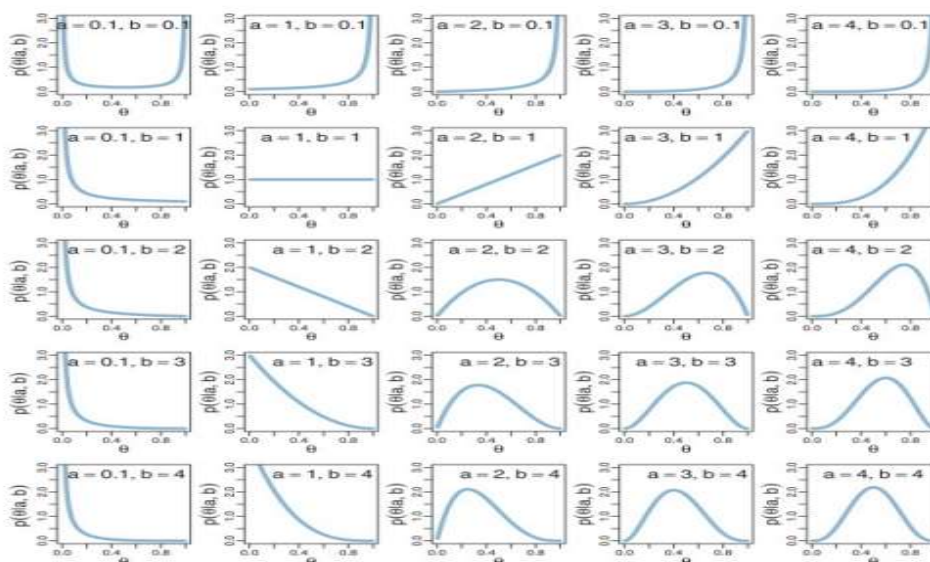


Fig.(2): Examples of The shape parameter (a) arises from left to right across the columns, while the shape parameter (b) arises from top to bottom across the rows. (13)

3.6. Dagum Distribution function (29)

Camillo Dagum was a pioneer in the search for mathematical distributions corresponding to empirical distributions of income and wealth. In 1970, satisfied with the traditional distribution used to collect data, Dagum (1977) promoted a model based on empirical observation, which

is a joint distribution function. CDF). Yield income elasticity $\eta(F, x)$ is a decreasing and limiting function of F . from the differential equation.

$$\eta(F, x) = \frac{d \log F(x)}{d \log x} = ap \left\{ 1 - [F(x)]^{\frac{1}{p}} \right\}, x \geq 0, \dots$$

(13)

s. t.: $p > 0$ and $ap > 0$,

$$F(x) = \left[1 + \left(\frac{x}{b} \right)^{-a} \right]^{-p}, x > 0. \dots$$

(14)

Developed in several articles on the growth processes of income distribution systems presented by Dagum, this approach starts with the specific characteristics of the empirical distribution of income and wealth and leads to the generation process which translated as follows:

$$\frac{d \log \{F(x) - \delta\}}{d \log x} = \vartheta(x)\varphi(F) \leq k, 0 \leq x_0 < x < \infty,$$

Since $k > 0$, $\vartheta(x) > 0$, $\varphi(x) > 0$, $\delta < 1$, and $d\{\vartheta(x)\varphi(F)\}/dx < 0$. the income elasticity of the CDF constraints is a positive, it is bound and decrease function of F .

The CDF of Dagum type II distribution as follow:

$$F(x) = \delta + (1 - \delta) \left[1 + \left(\frac{x}{b} \right)^{-a} \right]^{-p}, x \geq 0, \dots$$

(15)

$a, b, p > 0$ and $\delta \in (0, 1)$.

3.6.1. Basic properties of Dagum Distribution function

The b parameter of the Dagum distribution is the threshold and the other two parameters (a and p) are the shape parameters. However:

$$f(y) = \frac{ap e^{ap(y-\log b)}}{[1 + e^{a(y-\log b)}]^{p+1}}, -\infty < y < \infty \dots$$

(16)

a and $\log b$ are scale and location parameters, respectively, but p is a shape (or skewness) parameter.

this mode is at x :

$$\text{Mode}(x) = b \left(\frac{ap-1}{a+1} \right)^{1/a}. \dots$$

(17)

Note that a and ap determine the decreasing rate (increasing) from zero to $x \rightarrow 0$ ($x \rightarrow \infty$), It should also be noted,

Specifically,

$$E(X) = \frac{b\Gamma(p+1/a)\Gamma(1-1/a)}{\Gamma(p)} \dots$$

(18)

and

$$\text{Var}(X) = \frac{b^2\{\Gamma(p)\Gamma(p+2/a)\Gamma(1-2/a) - \Gamma^2(p+1/a)\Gamma^2(1-1/a)\}}{\Gamma^2(p)}. \dots$$

(19)

3.6.2. Dagum Distribution (4p) function

It's A continuous non-negative random variable, X , is said to have a Dagum (4P) distribution if its probability density function (pdf) and cumulative distribution function (cdf) are respectively given by: (30)

Dagum distribution (4P) is one of the most common statistical techniques in fields such as actuarial science, biology, engineering, medicine, and reliability ... etc.

In the 1970s, Camilo Dagum proposed the Dagum distribution (4P) in his research on income distribution, which examines the distribution of people's income. It is a continuous probability distribution defined for positive real numbers (31)(32).

Dagum distribution (4P) of a continuous nonnegative random variable X as its probability density function (pdf) and cumulative distribution function (cdf) as follow: (30)

$$f(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha k-1}}{\beta \left[1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right]^{k+1}} \quad \dots (20)$$

And

$$F(x) = \left[1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right]^{-k} \quad \dots (21)$$

since $k (> 0)$: shape parameter; $\alpha (> 0)$: shape parameter; $\beta (> 0)$: scale parameter; location parameter; and domain: $\gamma \leq x < \infty$, Dagum (4P) classification is also known as Inverse Burr (4P) distribution. When $\gamma = 0$, Dagum (4P) reduces to Dagum (3P) distribution The possible graphs of pdf and cdf of Dagum (4P) distribution are given below for some selected values as

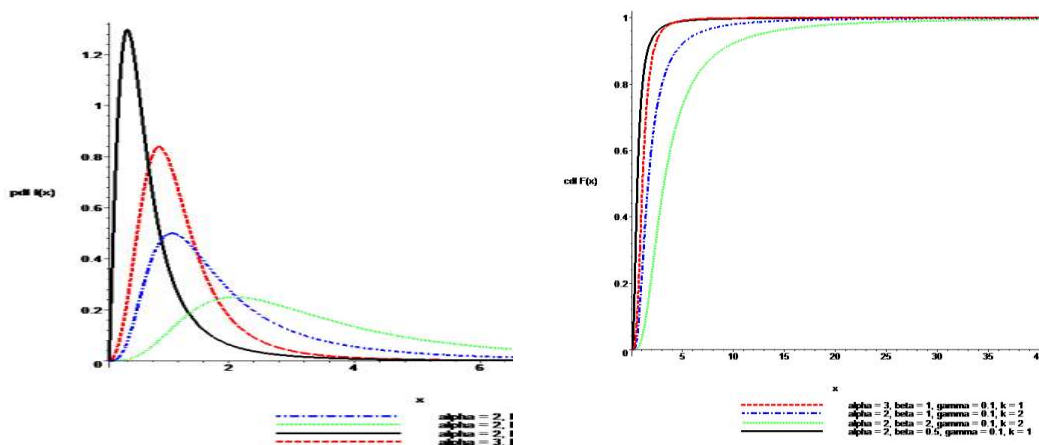


Fig.(3):Plots pdf and cdf for the Dagum (4P).(30)

The parameters effect is easily seen in these graphs, it is clear from this plot that the Dagum (4P) distribution is positively skewed with a long and heavy right tail for selected values of the parameter. (30)

4. Anderson Darling test (Anderson and Darling) (21)(22)

The quality of statistical model results depends on the fit of the predicted probability distribution to the data. Therefore, many efforts have been made to develop different sets of probability distributions and their corresponding statistical methods [33].

The Anderson-Darling test is a statistical test that judges whether a sample of data is drawn from a given probability distribution, but then the test and the set of significant values are independent of the distribution and must be brought to his attention, however this test is often used in the context of a family of test distributions, in which case the family parameters must be estimated and the test statistic or its significant value adjusted with that in mind, the Anderson Darling statistic as follows:

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F(x) - F_n(x)]^2}{F(x)[1 - F(x)]} dF(x) \quad \dots (22)$$

Below are observed order statistics of Anderson Darling, its becomes:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\log(zi) + \log(1 - z(n + 1 - i))] \quad \dots (23)$$

Anderson Darling statistic gives more weight to the tail of the distribution compared to the Cramer von Mises statistic, making it useful for detecting outliers.

Everywhere, Assume X_1, X_2, \dots, X_n is a r. s. from the distribution specified by $H_0: x_1, x_2, \dots, x_n$ represent their observed values; $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ represent the order statistics of X_1, X_2, \dots, X_n ; $x_{1:n} < x_{2:n} < \dots < x_{n:n}$ represent the observed order statistics; $z(i) = F(x_{i:n})$ for a hypothesized CDF, F ; F_n represent the empirical CDF (ECDF) of the r. s.; $\overline{F}_n = 1 - F_n$ denotes the empirical survival function (23)

Critical values of the Anderson-Darling test depend on the distribution under consideration.

5. Corona Virus pandemic

Coronaviruses are widely distributed in many countries in a variety of animals, including bats, cats, birds and camels (40). Coronavirus is also one of the pathogens of human respiratory tract infection. Four other human coronaviruses (HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1) induce mild upper respiratory tract disease, similar to the usual Cool (38), (39).

As of May 22, 2022, nearly 1 billion people in low-income countries will not be immune. Only 57 countries have vaccinated 70% of their populations - most of them high-income countries. Support all countries to achieve 70% vaccination coverage as much as, including 100% of people over 60 years; 100% of healthcare workers; 100% of people with underlying medical conditions Improved vaccine availability, but vaccination rates have not kept up. In certain countries, we experience gaps in operational or financial capacity in certain circumstances, we are seeing reluctance to get vaccinated due to misinformation.

Officials in Wuhan, China, first reported the first human case of COVID-19, the disease caused by the novel coronavirus that causes COVID-19, in December 2019. Wuhan city is the source of the outbreak or played a role in it. Amplified at the beginning of the epidemic. SARS-CoV-2 was discovered in early January, and its genetic sequence was made public on January 11-12. Complete genetic sequencing of the early human disease SARS-CoV-2, as well as sequencing of several other viruses isolated from human cases in China and around the world, suggest that SARS-CoV-2 has an ecological origin in bats. The virus is of natural animal origin, not from a manipulated or engineered virus. Several researchers managed to study the genetic characteristics of SARS-CoV-2 (36)

After that, the disease quickly spread from Wuhan to other regions. By the first week of January 2020, the Chinese Centers for Disease Control and Prevention (CDC) identified the novel coronavirus, named 2019 novel coronavirus (2019-nCoV), and based on throat swab samples taken from these patients. While some of the initial transmission involved exposure to seafood and animal markets, an increasing number of patients reported no exposure to animal markets, suggesting human-to-human transmission. 2019-nCoV has broken out and spread rapidly to other parts of China and some other countries. Due to the rapid global spread, the World Health Organization (WHO) declared the epidemic a public health emergency on January 30. disease . Coronavirus The virus is considered a sister type of severe acute respiratory syndrome (37).

5.1. Coronavirus disease (COVID-19) Vaccines

Several COVID-19 vaccines have been approved for use by the World Health Organization, the first mass vaccination program began in early December 2020, and the number of vaccine doses administered is updated daily on the COVID-19 dashboard.

The WHO Emergency Use List process determines whether a product can be recommended based on all available data on safety, efficacy and suitability in low- and middle-income countries. Vaccines are evaluated against clinical trial data, manufacturing processes, and quality control to ensure they meet accepted standards of quality, safety, and efficacy. The evaluation weighs the danger posed by the emergency and the benefits of using the product against the potential risks. Countries are free to issue emergency use authorizations for any medical product. Local emergency use authorizations are issued at the discretion of countries and do not require World Health Organization approval. As of January 12, 2022, the following vaccines have received an end-user license certificate:

- Pfizer/BioNTech Comirnaty Vaccine 31 December 2020
- SII/COVISHIELD and AstraZeneca/AZD1222 vaccines February 16, 2021
- Janssen/Ad26.COV 2.S vaccine developed by Johnson & Johnson March 12, 2021
- Moderna COVID-19 vaccine (mRNA 1273) April 30, 2021
- Sinopharm COVID-19 vaccine May 7, 2021
- Sinovac-CoronaVac vaccine June 1, 2021
- Bharat Biotech BBV152 COVAXIN Vaccine Nov 3, 2021
- Covovax vaccine (NVX-CoV2373) December 17, 2021
- Novavax vaccine (NVX-CoV2373) December 20, 2021

More than two years since the first SARS-CoV-2 infection was reported, the COVID-19 pandemic remains an acute global emergency. WHO's 2022 Strategic Preparedness, Preparedness and Response Plan outlines a number of key strategic adjustments whose rapid and sustained implementation at national, regional and global levels will enable the world to end the acute phase of the pandemic. (35)

The global spread of the virus has created a unique opportunity for joint learning around the world. Citizens, communities and governments everywhere faced the same urgent task of containing the spread of COVID-19. This has led to a variety of policy responses, including drastic attempts at government action, from total lockdowns in Italy and strict quarantines in

China to deliberate inaction in Sweden and Brazil. These policy responses have been the subject of detailed comparative research. The ways in which governments, societies and citizens came together to make these decisions and act were quite varied. (34)

6. Appropriate distribution and Statistical analysis. (20)

Fitting a distribution in statistics is a very common task, including choosing a probability distribution to model a random variable and finding parameter estimates for that distribution. This requires judgment and skill, and is often an iterative process of selecting distributions, estimating parameters, and assessing appropriate quality. Mass (Venables and Ripley, 2010) In the R package (R Development Core Team, 2013), the `fitDSR` function is used to calculate the maximum likelihood; Other R functions can be used to complete other steps in the fitting process, and we suggest the R package `fitdisrplus` (Deligette-Muller et al., 2014), which implements several methods to fit distributions of univariate parameters. The `fitDSR` function estimates the parameters of the distribution by maximizing the probability function using the best fit. Parameters with different roles as important parameters such as maximum fit estimation (also known as minimum distance estimation) and perturbation factors.

Before fitting one or more distributions to a data set, it is necessary to select good candidates from the predefined distributions. This choice can be guided by knowledge of the stochastic process governing the parameters of the model or, in the absence of knowledge of the underlying process, by observing its empirical distribution. To solve this problem, we use a program (Eastern Compatibility) to solve this problem and use functions to plot and characterize empirical distributions. Before discussing data and models, we will analyze graphical models. A graphical model is a method that uses graphs to display and estimate families of probability distributions.

The study focused on four groups: health professionals, scientists, physicians, statisticians, and the general public. We want to encourage public health and medical professionals to use more statistical methods, which are not always easy to implement. Healthcare professionals and statisticians need to work more closely together: not only on the data, but also on the research design. Collaborative work enables the inference (statistical inference) of results that are promising to the public, the ultimate goal of research, and assists clinicians and public health scientists in the planning, implementation, and analysis of statistical methods by providing illustrative examples. Understanding what others are doing and contributing to a more effective collaboration, this research could eventually become part of a biostatistics course in public health or an applied course in medicine. We can use the capabilities of Covid-19 to help doctors and nurses prepare or order hospital beds, ventilators, other essential hospital equipment and more. (28)

6.1. Description of the Data and Data sources

The data were collected for Mortality from the Corona epidemic in Iraq and on a daily basis for the extended period (1/6/2020 till 30/11/2021) and an attempt was made in this research to

know the probability distribution of the behavior of Mortality resulting from the Corona epidemic in Iraq over the previous period

As well as the probability distribution of deaths before the discovery of vaccines and the probability distribution after the discovery of vaccines, based on a set of packages and functions in the programming language R mainly, especially the (fitdistrplus) package, as well as the (Easyfit) software in a secondary way. The results were as follows, knowing that the data were collected through interactive database (WorldOmeter.info)

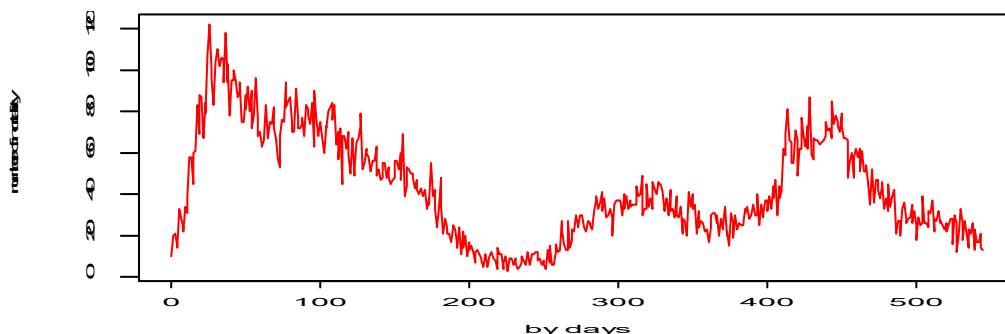


Fig.(4): Shows Iraqi corona virus mortalities from ()

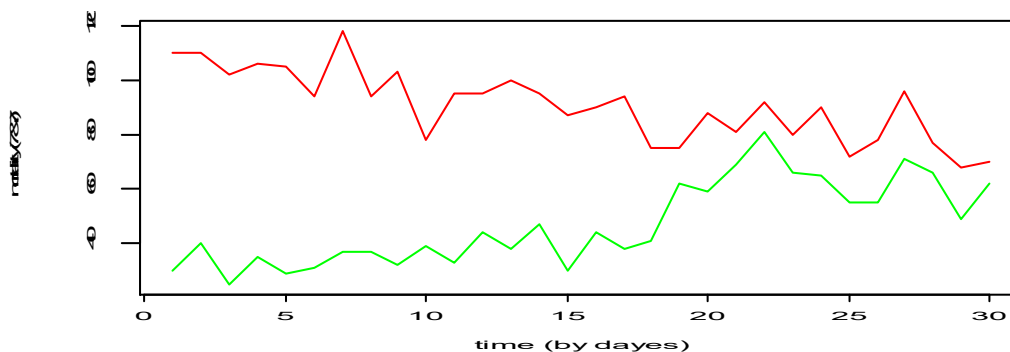


Fig.(5): show the mortality in month of July for (2020 & 2021)

We observe some outliers in both variables, especially in the upper region. This was also observed in the time series plot. The aim of this material is just to show some basic functions from the library(fitdistr) and library(fitdistrplus) which are used for fitting probability distributions to a vector of numerical data.

The function descdist provides a skewness-kurtosis graph will help you choose the best filter(s) to fit a given dataset. If we want to use it for discrete distributions we may use argument discrete=TRUE.

Table(1): shows Anderson-Darling results for Johnson SB Distribution

Johnson SB Distribution
Anderson-Darling

Sample Size	545					Days
Statistic	1.1398					
Rank	1					
α	0.2	0.1	0.05	0.02	0.01	
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074	
Reject?	No	No	No	No	No	
Johnson SB	$\hat{\alpha}=0.76551$	$\hat{\beta}=0.94596$	$\hat{\gamma}=125.9$	$\hat{\delta}=0.08084$		

H0: The data follow a specified distribution.

Ha: The data does not follow the specified distribution.

Under the null hypothesis, test statistics Johnson SB. H0 is rejected if the Lagrange multiplier LMP is greater than the Johnson SB critical point at a given critical level. α .(527)

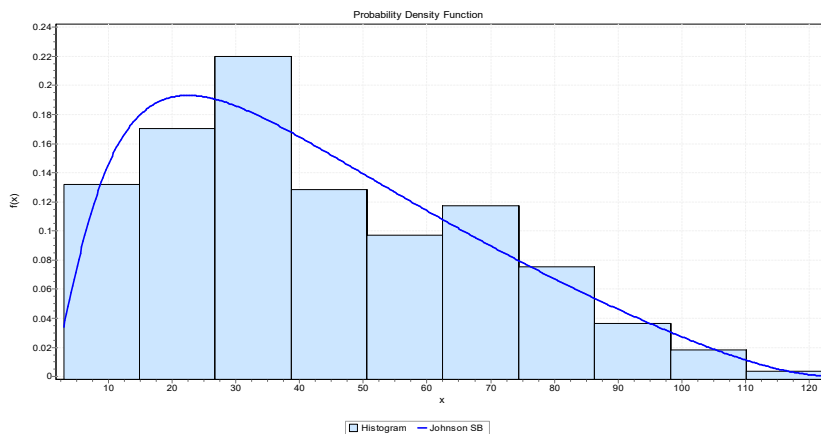


Fig.(6): Shows Johnson SB Distribution of Iraqi mortality before and after vaccine discovery

We use this graphical test to compare the fitted distributions with the real observations. We expect the distributions to follow the patterns of the real observations in the histogram.

Table(2): shows Anderson-Darling results for Dagum (4P) Distribution

Dagum (4P) Distribution
Anderson-Darling

Sample Size	301					Days
Statistic	1.2827					
Rank	1					
α	0.2	0.1	0.05	0.02	0.01	
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074	
Reject?	No	No	No	No	No	
Dagum (4P)	k=0.03103	$\alpha=26.425$	$\beta=99.135$	$\gamma=2.9998$		

To observe more clearly the fitting between the two most accurate fitting we can construct graphics above only for these two distributions, We are searchign the minimum value of the test statistic and the minimum value of the factor score. The lowest value match among the comparisons is considered the most accurate. In this case the "best fitting" of the data.

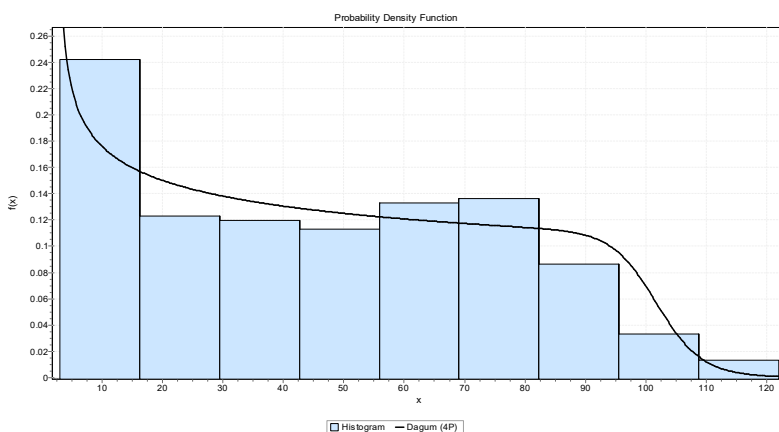


Fig.(7): Shows Dagum (4P) Distribution of Iraqi mortality before vaccine discovery

Table(3): shows Anderson-Darling results for Beta Distribution

Beta Distribution						
Anderson-Darling						
Sample Size	153					Days
Statistic	1.1402					
Rank	1					
α	0.2	0.1	0.05	0.02	0.01	
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074	
Reject?	No	No	No	No	No	

Beta	$\alpha_1=1.2583$	$\alpha_2=1.8034$	$a=11.839$	$b=88.733$
------	-------------------	-------------------	------------	------------

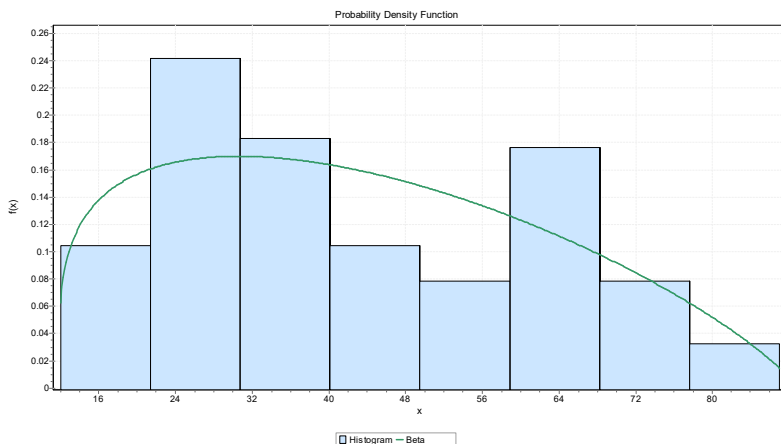


Fig.(8): Shows Beta Distribution of Iraqi mortality after vaccine discovery

7. Strengths and limitations of this study

Our study revealed trends and short-term characteristics in the analysis of deaths resulting from the Corona epidemic in Iraq, The accuracy of the data is reliable because the sample includes the entire population in all regions of Iraq based on the global monitoring network, Because of the regional divisions in Iraq and the lack of specific data for each region, we did not analyze (e.g. differences between rural and urban areas or between governorates, or on the basis of gender, age, or whether or not the patient suffers from chronic diseases) either what was the secondary cause of death, or whether corona pandemic caused the deceased to have a stroke, whether there is a significant relationship between coronary artery disease and stroke, unfortunately, it did not exist.

8. Conclusion

This paper uses probability distributions to help estimate mortality from the COVID-19 pandemic. After analyzing and comparing the results of probability distributions, a common way of describing the incidence of COVID-19 deaths, this paper provides evidence that the mortality behavior of COVID-19 is a statistical variable, The goal is to use a probability distribution to determine the nature of the corona-caused mortality variable whose behavior we are trying to describe, determining the correct category would allow for appropriate application of the distributions (e.g. Johnson, beta or dagum distributions), so the behavior of mortality data in Iraq was different before and after vaccine discovery. For the total mortality data, the Johnson distribution was the best for representing the data as we indicated in the application side tables. As for the behavior of the data before the discovery of the vaccine, it was beta variance. It is the best compared to other probability distributions. As for the mortality behavior

after the discovery of the vaccine, the Dugum distribution was the best, relying on (Anderson Darling test) to compare the distribution and choose the best, which may help improve predictions of virus epidemics in Iraq in the future.

It can be compared internationally, helping to understand how different national strategies and policies affect the spread and severity of epidemics, as well as to understand the number of staff and the need to take care of hospitals. with increasing COVID patient admissions or decreasing devices, however, we do not believe this limitation needs to be explained in a biological or demographic sense due to genetics or some other natural mechanism.

9. Recommendations

There are many recommendations that we have noticed that can be worked on in future studies, including:

- Analysis of age-specific mortality rates of deaths classified as COVID-19-related deaths.
- Government make suggestions to evaluate the current policy and formulate more regulations related to epidemic and infectious diseases in order to improve health care for the Iraqi community.
- Robust and timely data collection and dissemination systems, with disaggregated and disaggregated data by sex and cause of death.
- High quality information be collected and processed quickly and that we have more resources to develop COVID-19 action plans and better tools to track the disease. Iraqi health systems are even less computerized and staff lack contingency plans to keep accurate information in health records to create databases for the future or make a suggestion to sensitize health workers.
- Implementing different statistical procedures to estimate the health system load (hospital, ward, intensive care, etc.) and operational needs (number of beds, doctors, nurses, ventilators, etc.) in order to manage with success any possible health problems. System overload, during a possible next wave of Covid-19. Which may have to do with a very large number of patients requiring critical medical care.

References

1. Daniel WW. , 1995, "a foundation for analysis in the health sciences", John Wiley & Sons.
2. Samuel Gbari et al, 2017, "Extreme Value Analysis of Mortality at the Oldest Ages: A Case Study Based on Individual Ages at Death", Society of Actuaries.
3. Daniel WW. eds. Biostatistics: a foundation for analysis in the health sciences. New York: John Wiley & Sons, 1995.
4. Andrea Viti et al, 2014, "A practical overview on probability distributions", Journal of Thoracic Disease.
5. Kolmogorov AN. , 1950, "Foundations of Theory of Probability", Chelsea Publishing.
6. J Thorac Dis, 2015, "A practical overview on probability distributions", Journal of thoracic disease.

7. John Goldsmith, 2001, " Probability for linguists ".
8. Chow, YS, Teicher, H. , 1988, Probability Theory: Independence, Interchangeability, Martingales. New York: Springer.
9. Fang, K, Xu, J. , 1987, Statistical Distributions. Beijing: Science Press.
10. Krishnamoorthy, K. , 2006, Handbook of Statistical Distributions with Applications. Boca Raton: Chapman and Hall/CRC.
11. Patel, JK, Kapadia, CH, and Owen, DB. , 1976, Handbook of Statistical Distributions. New York: Marcel Dekker.
12. High School Statistics and Probability, 2010, "Common Core State Standards for Mathematics", Oregon department of Education.
13. S. Sinharay, 2010, Continuous Probability Distributions ,in International Encyclopedia of Education (Third Edition).
14. Gabriel Demombynes Damien de Walque Paul Gubbins Beatriz Piedad Urdinola Jeremy Veillard, 2021, COVID-19 Age-Mortality Curves for 2020 Are Flatter in Developing Countries Using Both Official Death Counts and Excess Deaths, Human Development Global Practice & Development Economics Development Research Group JEL Classification: I18, O57, J10.
15. Zdenek Karpisek, 2007, Applied Statistics, Department of Statistics and Optimization-Insitute of Mathematics-Brno University of Technology.
16. N. L. Johnson, "Systems of frequency curves generated by methods of translation," *Biometrika*, vol. 36, no. 1-2, pp. 149–176, 1949.
17. A. Alzaatreh, C. Lee, and F. Famoye, "A new method for generating families of continuous distributions," *Metron*, vol. 71, no. 1, pp. 63–79, 2013.
18. Yueqin Dun and Yu Kong Efficient, 2022, Johnson-SB Mixture Model for Segmentation of CT Liver Image, *Journal of Healthcare Engineering* Volume 2022, Article ID 5654424, 21 pages <https://doi.org/10.1155/2022/5654424>.
19. Andreopoulos Panagiotis , Bersimis. G. Fragkiskos , Tragaki Alexandra and Rovolis Antonis, 2016, MORTALITY MODELLING USING PROBABILITY DISTRIBUTIONS. APPLICATION IN GREEK MORTALITY DATA, 4 rd SMTDA Conference Proceedings, 1-4 June 2016, Valletta, Malta C. H. Skiadas (Ed).
20. Marie Laure Delignette-Muller , 2020, fitdistrplus: An R Package for Fitting Distributions, Universit´e de Lyon.
21. Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, 193-212.
22. Anderson, T. W. and Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association*, 49, 765-769.
23. Chu, J., Dickin, O., & Nadarajah, S. (2019). A review of goodness of fit tests for Pareto distributions. *Journal of Computational and Applied Mathematics*, 361, 13-41. <https://doi.org/10.1016/j.cam>

24. Mahase E (2020) Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* 368:m641. <https://doi.org/10.1136/bmj.m641>
25. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, T Yu, Zhang X, Zhang L (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. [https://doi.org/10.1016/s0140-6736\(20\)30211-7](https://doi.org/10.1016/s0140-6736(20)30211-7)
26. Qiurong Ruan, 2020, Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China, *Intensive Care Med* (2020) 46:846–848 <https://doi.org/10.1007/s00134-020-05991-x>.
27. Helena Cruz Castanheira, 2020, COVID-19 mortality Evidence and scenarios, DEMOGRAPHIC OBSERVATORY Latin America and the Caribbean/United Nations Population Fund (UNFPA) Regional Programme on Population and Development.
28. Jorge Luis Romeu,, 2020, Covid-19 ICU Staff and Equipment Requirements using the Negative Binomial, https://www.researchgate.net/publication/341282217_A_Proposal_for_Fighting_Covid-19_and_its_Economic_Fallout.
29. Christian Kleiber, 2008, A Guide to the Dagum Distributions, <https://www.researchgate.net/publication/46450919>, DOI: 10.1007/978-0-387-72796-7_6.
30. M. Shakil, B. M. G. Kibria, and M. Ahsanullah, 2021, Some Inferences on Dagum (4P) Distribution: Statistical Properties, Characterizations and Applications, *WSN* 154 (2021) 1-33 EISSN 2392-2192 : <https://www.researchgate.net/publication/349363355>
31. Dagum, C., A model of income distribution and the conditions of existence of moments of finite order. *Bulletin of the International Statistical Institute*, 46 (1975) 199-205 [3].
32. Dagum, C., A new model of personal income distribution: Specification and estimation. *Economie Appliquée*, 30 (1977) 413-437. https://doi.org/10.1007/978-0-387-72796-7_1.
33. H. M. Yousof, A. Z. Afify, A. E. H. N. Ebraheim, G. G. Hamedani, and N. S. Butt, “On six-parameter frechet distribution: properties and applications,” *Pakistan Journal of Statistics and Operation Research*, vol. 12, no. 2, pp. 281–299, 2016.
34. Afsahi, Afsoun, Emily Beausoleil, Rikki Dean, Selen A. Ercan, Jean-Paul Gagnon. 2020. Democracy in the Time of COVID-19 (A Special Issue of Democratic Theory).
35. WHO, 2022, Coronavirus disease (COVID-19): Vaccines, <https://www.who>.
36. WHO, 2020, Coronavirus disease 2019 (COVID-19) Situation Report – 94, <https://www.who>.
37. Gorbalenya, Alexander E. , 2020, Severe acute respiratory syndrome-related coronavirus – The species and its viruses, a statement of the Coronavirus Study Group, , bioRxiv: 2020.02.07.937862. (Accessed on 13 Feb 2020).

38. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. , 2016, Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol*; 24(6):490-502. 8.
39. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al., 2020, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet.*; 395(10223):507-513.
40. Sheikh Salahuddin Ahmed, 2020, The Coronavirus Disease 2019 (COVID-19), *Journal of Advances in Medicine and Medical Research*, Article no.JAMMR.55245 ISSN: 2456-8899.