

Unit Monsef distribution with regression model

M.M.E. Abd El-Monsef

Faculty of Science, Tanta University

N.M. Sohsah

Faculty of Science, Tanta University

W.A. Hassanein

Faculty of Science, Tanta University

Abstract

The aim of this paper is introduced a new distribution with one parameter called unit Monsef distribution (UMD) with it is regression model since it is more fitting than well-known distributions in educational attainment data , derived from a transformation on Monsef distribution. its probability function has been generated also distribution function, reliability function and hazard function, some of statistical measures are investigated such like moments of origin point, central moments, Lorenz and Bonferroni curves also stress strengths parameter are investigated. parameter estimated is obtained by using more than method like maximum likelihood method, method of moment and least square method. the simulation scheme for unit Monsef parameters. Finally, real data is used for studying the flexibility of UMD also applied on UMD regression model compared with others well Known distributions.

Keywords: Unit distributions, Monsef distribution, Stress strength model, Random variable transformation, Estimation methods, re-parametrization, regression model.

1. Introduction

There are many phenomena have the uncertainty of a bounded, which lie on the interval $(0, 1)$ such as the scores of ability tests and different rates. The transformation of the random variable generates new distribution with domain $(0,1)$, which be more flexible and more fitting to data which lie in this interval. Unit distributions are the best to describe this phenomenon one of the most important distributions that use in this data is the Beta distribution (or the Pearson type IV distribution) because of its flexibility there is many unit distributions which introduced as alternative to beta distribution and Kumaraswamy distribution like the unit-Gamma distribution [1], the unit-Logistic distribution [3], a gamma regression model [8] obtained depend on unit

gamma distribution. unit Burr XII distribution and regression model [14] of the most important regression models are beta regression model [5] and Kumaraswamy regression model [13]. to in this paper a new unit distribution will be obtained depend on Monsef distribution [13], which has more flexibility, this paper has organized as follow, the first section introduced the method of transformation , pdf and cdf of the UMD, the survival and hazard rate and the behavior of the pdf and hazard function, the moments and ,Lorenz and Bonferroni curve are obtained in section 2, stress strength parameter were obtained in section 3, more than one estimation method were obtained to estimate the parameter of the new distribution in section 4, section 5 and 6 introduced the simulation study and application.

2. the Unit Monsef distribution

Monsef distribution was introduced by Monsef [13] with cdf:

$$F(x) = 1 - e^{-x^\alpha}, \quad (1)$$

and the pdf of *Monsef* distribution is:

$$f(x) = \alpha x^{\alpha-1} e^{-x^\alpha}, \quad (2)$$

By using the transformation in equation (1) hence the new distribution was proposed with support on the unit interval. The cdf and the pdf of the UMD are:

$$(3)$$

$$(4)$$

The survival and hazard functions can be obtained as following:

The behavior of the pdf and the hazard rate of the new distribution at different values of x will be showing in the below figure.

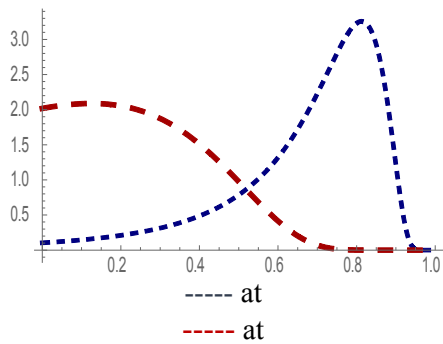


Fig. (1). The behavior of the unit Monsef pdf

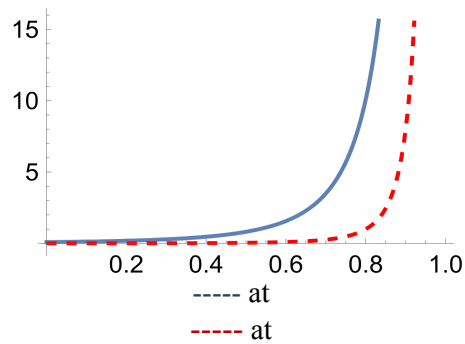


Fig. (2). The behavior of the unit Monsef hazard function

3. Statistical measures

Some of the important statistical properties will be obtained in this section like moments, skewness, and Kurtosis coefficients, Lorenz and Bonferroni curves also the stress strength parameter were obtained.

3.1 Moments about the origin (raw moments).

The r^{th} moment about the origin of the unit Monsef distribution is defined by

The first four moments about the origin are written as:

,

3.2 Central moments.

The r^{th} moment about the mean of the unit Monsef distribution is defined by

Where

the variance

The skewness and kurtosis measures can be obtained by:

coefficient The of variation (CV) for the new distribution can be written as :

3.3 Lorenz and Bonferroni curves

The Lorenz and Bonferroni curves have importance in studying economics and in reliability applications.

The Lorenz curve for the unit Monsef distribution is defined as

The Bonferroni curve for the unit Monsef distribution is defined as

3.4 Stress strength reliability

Studying the survival and failure rate under conditions is an important to the system the system will perform its intended function adequately. Let X and Y are two independent unit Monsef r.vs with parameters and then the stress strength measures can be obtained as:

4. The Estimation Methods

In this section some methods of parameter estimation will be introduced for the new distribution unit Monsef.

4.1 Maximum likelihood Estimation

Suppose there is a sample of iid observations. Coming from a distribution with probability density function identified in equation (4). The log-likelihood function is given by

The partial derivatives of the log-likelihood for the parameter can be written:

The MLEs can be obtained by solving the derivatives of the to zero. For the UMD the solution of the nonlinear equations has no closed form, also some numerical methods are needed for the solution.

4.2 the method of moment Estimator

By letting be a r.s with size n from the UMD, the moment estimates (MOM), for the parameter is given solving equation

By solving the above equation

The solution of above equation has not closed form, but by using, we can compute the estimated values of .

4.3 Method of Least Square Estimation

The LSE and WLSE were proposed by Swain, Venkatraman & Wilson [4] to estimate the parameters of Beta distributions.

The least square estimators can be obtained by minimizing

the least square estimator of can be introduced as:

The partial derivatives of the Least square method for the parameters can be obtained as follow:

4.4 Method of Cramér- Von- Mises

Another method of estimation parameter is Cramér-von-Mises estimators of can be obtained by:

5. Simulation study

A simulation Scheme for the UMD will be made by generating 5000 samples for the parameters at different sample sizes. The simulation nodes were chosen at different values at .

5.1 Simulation study for the new distribution using MLE.

Table (1): The MSE and the Bias for the parameter .

| N | | | | | | | | |
|-----|---------|--------|--------|--------|--------|--------|--------|--------|
| | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias |
| 10 | 0.00041 | 0.0010 | 0.0251 | 0.0319 | 0.1213 | 0.0775 | 0.9700 | 0.2518 |
| 20 | 0.00017 | 0.0038 | 0.0107 | 0.0084 | 0.0495 | 0.0210 | 0.3713 | 0.0826 |
| 30 | 0.00011 | 0.0009 | 0.0069 | 0.0079 | 0.0318 | 0.0099 | 0.2337 | 0.0761 |
| 40 | 0.00009 | 0.0006 | 0.0059 | 0.0053 | 0.0273 | 0.0269 | 0.2023 | 0.0750 |
| 50 | 0.00006 | 0.0004 | 0.0041 | 0.0034 | 0.0187 | 0.0192 | 0.1363 | 0.0609 |
| 60 | 0.00006 | 0.0002 | 0.0034 | 0.0032 | 0.0157 | 0.0127 | 0.1143 | 0.0569 |
| 70 | 0.00005 | 0.0002 | 0.0033 | 0.0029 | 0.0151 | 0.0077 | 0.1099 | 0.0399 |
| 80 | 0.00004 | 0.0003 | 0.0025 | 0.0022 | 0.0116 | 0.0068 | 0.0845 | 0.0255 |
| 90 | 0.00003 | 0.0004 | 0.0021 | 0.0021 | 0.0098 | 0.0059 | 0.0702 | 0.0241 |
| 100 | 0.00003 | 0.0003 | 0.0019 | 0.0011 | 0.0089 | 0.0058 | 0.0643 | 0.0214 |

5.2 study for the new distribution using LSE method.

Table (2): The MSE and the Bias for the parameter

| N | | | | | | | | |
|----|----------|---------|---------|---------|---------|---------|-------------|---------|
| | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias |
| 10 | 0.0042 | 0.0013 | 0.02599 | 0.01116 | 0.12817 | 0.02981 | 1.0691 | 0.12038 |
| 20 | 0.00021 | 0.00081 | 0.01247 | 0.00650 | 0.05923 | 0.01677 | 0.4640 4 | 0.06643 |
| 30 | 0.00014 | 0.00075 | 0.00827 | 0.00595 | 0.03902 | 0.01419 | 0.3004 1 | 0.05410 |
| 40 | 0.00011 | 0.00065 | 0.00664 | 0.00514 | 0.03119 | 0.01202 | 0.2397 6 | 0.04709 |
| 50 | 0.000085 | 0.00051 | 0.00511 | 0.00395 | 0.02384 | 0.00894 | 0.1809 4 | 0.03659 |
| 60 | 0.000069 | 0.00034 | 0.00416 | 0.00263 | 0.01935 | 0.00575 | 0.1449 5 | 0.02633 |
| 70 | 0.000059 | 0.00028 | 0.00354 | 0.00211 | 0.01647 | 0.00445 | 0.1230 9 | 0.02239 |
| 80 | 0.000053 | 0.00027 | 0.00317 | 0.00306 | 0.01466 | 0.00636 | 0.1092 9 | 0.02134 |
| 90 | 0.000046 | 0.00024 | 0.00274 | 0.00199 | 0.01267 | 0.00387 | 0.0945 4 | 0.02037 |

| | | | | | | | | |
|-----|----------|---------|---------|---------|---------|---------|-------------|---------|
| 100 | 0.000042 | 0.00022 | 0.00248 | 0.00164 | 0.01146 | 0.00299 | 0.0856 3 | 0.01779 |
|-----|----------|---------|---------|---------|---------|---------|-------------|---------|

5.3 Simulation study for the new distribution using Cramér- Von- Mises method.

Table (3): The MSE and the Bias for the parameter

| N | | | | | | | | |
|-----|---------|--------|--------|--------|--------|--------|--------|--------|
| | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias |
| 10 | 0.00042 | 0.0018 | 0.0261 | 0.0155 | 0.1296 | 0.0406 | 1.091 | 0.1570 |
| 20 | 0.00021 | 0.0011 | 0.0125 | 0.0087 | 0.0596 | 0.0224 | 0.4699 | 0.0854 |
| 30 | 0.00013 | 0.0009 | 0.0083 | 0.0075 | 0.0392 | 0.0180 | 0.3035 | 0.0671 |
| 40 | 0.00011 | 0.0008 | 0.0066 | 0.0063 | 0.0313 | 0.0149 | 0.2414 | 0.0568 |
| 50 | 0.00009 | 0.0006 | 0.0051 | 0.0048 | 0.0239 | 0.0112 | 0.1821 | 0.0444 |
| 60 | 0.00007 | 0.0004 | 0.0042 | 0.0034 | 0.0194 | 0.0076 | 0.1458 | 0.0329 |
| 70 | 0.00006 | 0.0003 | 0.0035 | 0.0027 | 0.0165 | 0.0061 | 0.1237 | 0.0281 |
| 80 | 0.00005 | 0.0004 | 0.0032 | 0.0036 | 0.0147 | 0.0078 | 0.1098 | 0.0323 |
| 90 | 0.00004 | 0.0003 | 0.0027 | 0.0025 | 0.0127 | 0.0052 | 0.0949 | 0.0248 |
| 100 | 0.00004 | 0.0002 | 0.0024 | 0.0021 | 0.0114 | 0.0041 | 0.0859 | 0.0218 |

The result of simulation shows that the MLE give the best results of biases and MSEs. LSE is the second good estimator, followed by the Cramér- Von- Mises method. we can argue that the MLE and least squares estimators are among the best performing estimators for UMD.

6. Regression model and application

In this section, a regression structure for the unit-Monsef distribution will be obtained, also a real data is applied to shows the flexibility of UMD. The MLE is used to estimate the parameters, The statistics criteria (WT, AIC, BIC, AICC, HQIC, AD and CVM), K-S and p-value (K-S) were calculated for the fitted distributions.

In regression analysis it is very common to model the mean of the response. Since the unit-Monsef distribution has closed form expression for mean it can be used in this context. It is noteworthy that the re-parametrized pdf of unit Monsef can be written as follow:

(5)

Let be n independent random variables, where X_i . The regression model is defined assuming that the mean of X_i satisfies the following functional relation

where β is a p -dimensional vector of regression coefficients and X_i denotes the observations on known covariates.

The proposed regression model is

The Education attainment

The used data set comes from the BLI of OECD countries. The data source is <https://stats.oecd.org/index.aspx?DataSetCode=BLI>. The educational attainment values of the OECD countries

Table (4): The MLEs for the parameters, -log and K-S for different distributions.

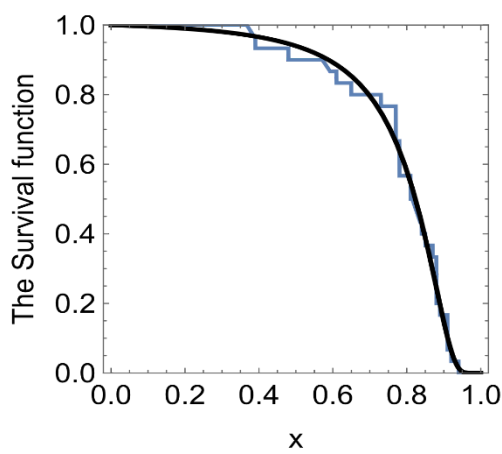
| Distribution | MLEs | -log | K-S | P-value |
|--------------|------|-------------|------------|---------|
| Unit Monsef | | - 24.298 | 0.101 3 | 0.9181 |
| Unit Lindley | | - 23.623 | 0.155 9 | 0.4595 |
| Kumaraswamy | | - 21.096 | 0.753 2 | 0.000 |
| Beta | | - 20.519 | 0.202 7 | 0.1698 |
| Unit Weibull | | - 19.142 | 0.210 1 | 0.1416 |

Table (5): The statistics criteria for different distributions.

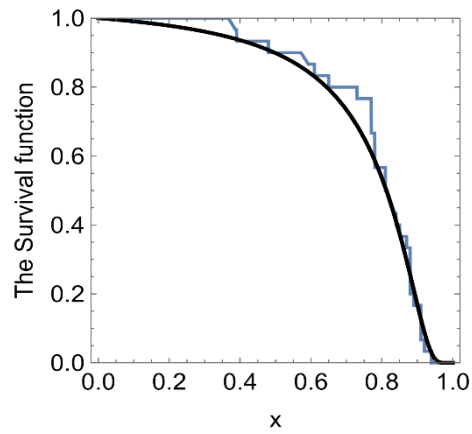
| Distribution | AIC | BIC | AICC | HQIC | AD | CVM | WT |
|--------------|-----|-----|------|------|----|-----|----|
|--------------|-----|-----|------|------|----|-----|----|

| | | | | | | | |
|--------------|---------|---------|---------|---------|---------|---------|---------|
| Unit Monsef | -46.596 | -45.195 | -45.596 | -46.158 | 0.2537 | 0.03849 | 0.0384 |
| Unit Lindley | -45.247 | -43.846 | -45.104 | -44.799 | 0.42934 | 0.0724 | 0.05647 |
| Beta | -37.519 | -34.237 | -36.595 | -36.143 | 1.0531 | 0.1755 | 0.1383 |
| Kumaraswamy | -38.192 | -35.389 | -37.747 | 37.295 | 67.523 | 6.895 | 1.443 |
| Unit Weibull | -34.283 | -33.839 | -33.834 | -33.387 | 1.2077 | 0.1975 | 0.1800 |

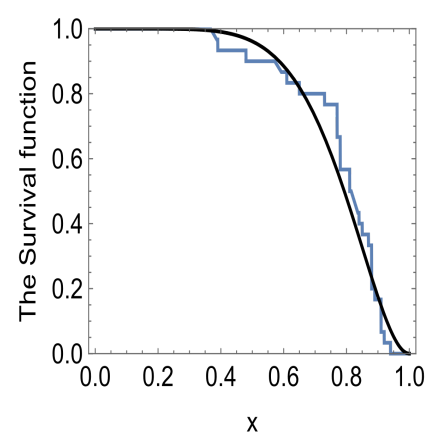
Figure (3): The survival curves for the educational attainment data.



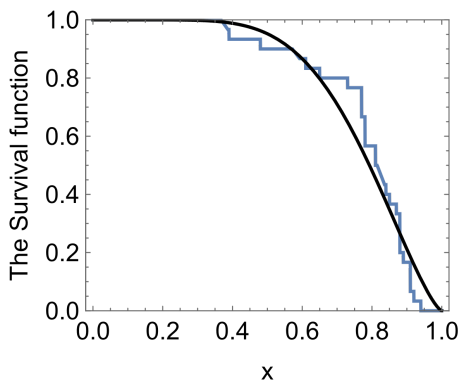
Unit Monsef Distribution



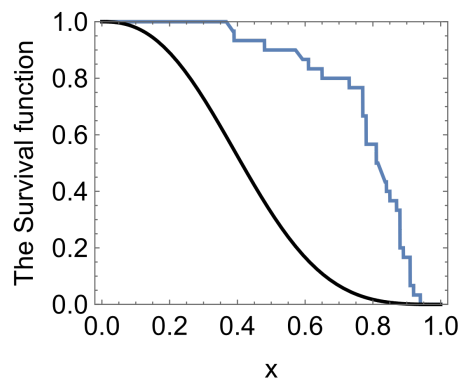
Unit Lindley Distribution



Beta Distribution



Unit Weibull Distribution



Kumaraswamy Distribution

The Education attainment

The used data set comes from the BLI of OECD countries. The educational attainment values of the OECD countries (y) is considered as response (dependent) variable. The goal is to explore the effects of following covariates on the conditional mean of the response variable: homicide rate (HR), dwellings without basic facilities (DWBF), and labor market insecurity (LMI). The logit link function which ensures that the estimated mean lies between 0 and 1, is used for all fitted regression models. The fitted regression model is

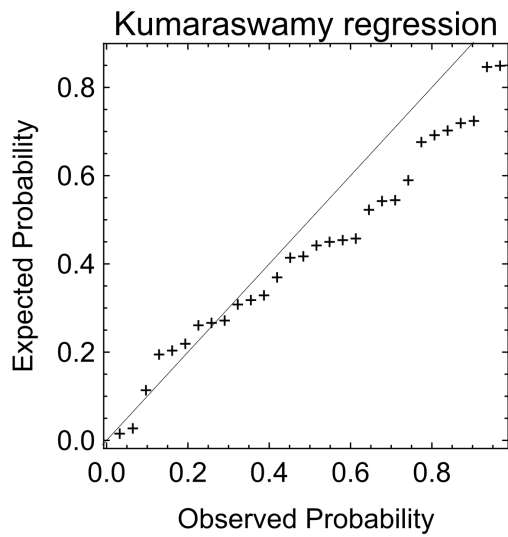
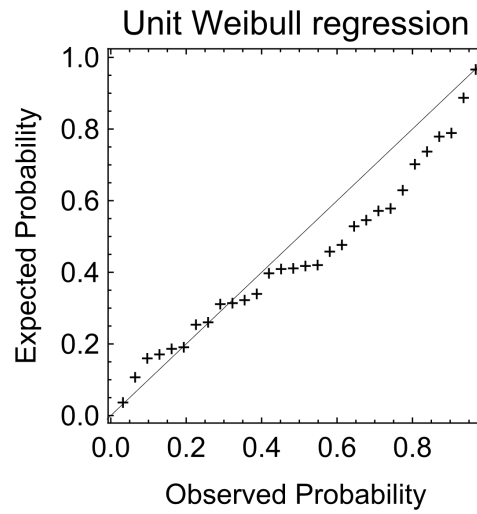
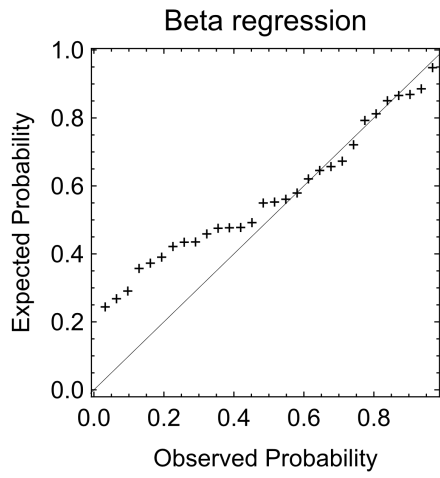
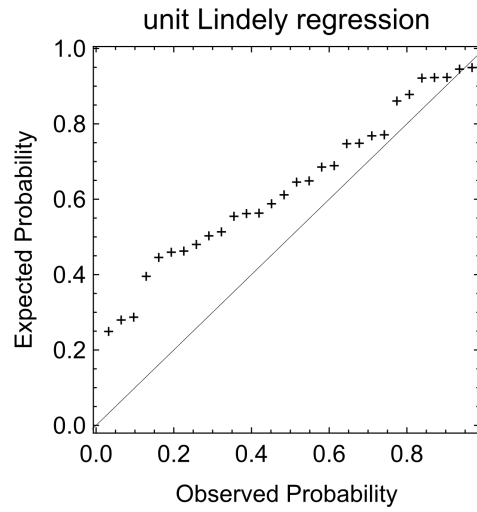
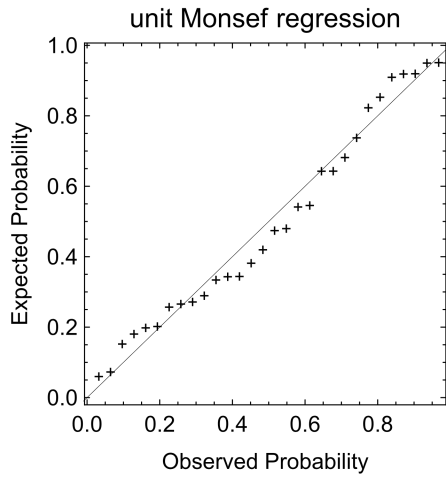
Table (6): The estimated value of the covariates for the Unit Monsef regression model comparing regression models.

| MLE | Unit Monsef | Unit Lindley | Unit Weibull | Kumaraswamy | Beta |
|------------|--------------------|---------------------|---------------------|--------------------|-------------|
| | ----- | ----- | 1.8894 | 3.2183 | 13.1724 |
| | 1.668 | 1.502 | 1.9856 | 1.8935 | 1.7749 |
| | 0.0213 | -0.0241 | 0.05202 | 0.00756 | -0.0216 |
| | -0.0320 | -0.0288 | -0.09362 | -0.06828 | -0.0618 |
| | -0.0446 | -0.0411 | -0.0667 | 0.04769 | -0.0451 |

Table (7): The estimated criteria for the Unit Monsef regression model comparing regression models.

| Criteria | Unit Monsef | Unit Lindley | Unit Weibull | Beta | Kumaraswamy |
|-----------------|--------------------|---------------------|---------------------|-------------|--------------------|
| AIC | -47.1035 | -44.045 | -45.352 | -42.2007 | -39.4771 |
| BIC | -41.4987 | -38.440 | -38.346 | -35.1947 | -32.4711 |
| -log | -27.5518 | -26.023 | -27.676 | -26.1004 | -24.7385 |

Figure (4): Q-Q Plots for the educational attainment data.



The results of the above shows that the UMD is more flexible and more fitting comparing with other distributions, since has the smallest statistic criteria and the smallest log likelihood, hence the above values of measures will lead to the pdf of the UMD has its own shape and may be difficult to replace by any other known distribution.

7. Conclusion

In this paper, the pdf and cdf of UMD were derived in section 2 also the behavior of the pdf and hazard rate are drawn at different values of the parameter. The most important statistical properties such like moments, stress strength parameters and other are derived in section 3. The parameter estimates is obtained using maximum likely hood, method of moment, least square method , weighted least square and Cramér- Von- Mises method in section 4. In section 5 simulation study is introduced using MLE, LSE and Cramér- Von- Mises method, Finally the regression model for the proposed distribution is obtained and a real data is applied in UMD and compared with other well known distribution and the results show that the UMD is more flexible and well fit then the other distributions in section 6.

Reference

- [1] Kumarasamy, P.; (1980). *A generalized probability density function for double random processes*. Journal of hydrology.46 (1-2):79—88.
- [2] Tadikamalla, P.R.; (1981). *On a family of distributions obtained by the transformation of the gamma distribution*. Journal of Statistical Computation and Simulation 13 (3-4), 209–214.
- [3] Tadikamalla, P. R.; Johnson, N. L.;(1982). *Systems of frequency curves generated by transformations of logistic variables*. Biometrika. 69: 461–465.
- [4] Swain. James. J; Venkatraman, Sekhar; Wilson, James. R. (1988). *Least-squares estimation of distribution functions in johnson's translation system*. Journal of Statistical Computation. 29:4, 271-297.
- [5] Ferrari, S; Cribari-Neto,F. (2004). *Beta regression for modeling rates and proportions*. Journal of applied statistics. 31(7):799—815.
- [6] Mitnik,P. A; Baek, S.(2013). *The Kumaraswamy distribution: median-dispersion re-parametrizations for regression modeling and simulation-based estimation*. Statistical Papers. 54(1):177—192.
- [7] Alzaatreh, A; Lee, C; Famoye, F (2013). *A new method for generating families of continuous distributions*. METRON. 71, 63-79.
- [8] Mousa, A. M., El-Sheikh, A. A., Abdel-Fttah, M. A. (2016). *A gamma regression for bounded continuous variables*. Advanced and Applications in Statistics. 49 (4):305—326.
- [9] Bayes, C.L; Bazán, J. L; De Castro, M. (2017). *A quantile parametric mixed regression model for bounded response variables*. Statistics and its interface. 10(3):483—493.
- [10] Dey, S; Raheem, E; Mukherjee, S (2017). *Statistical Properties and Different Methods of Estimation of Transmuted Rayleigh Distribution*. Revista Colombiana de Estadística. 40(1):165-203.
- [11] Mazucheli, J; Menezesa, A. F.; Fernandes, Chakraborty, S. (2018). *On the one parameter Unit-Lindley distribution and its associated regression model for proportion data*. Applied Statistics, 46:4, 700-714.

- [12] Mazucheli, J; Menezesa, A. F.; Fernandes, L. B; de Oliveira, R. P. ; M. E. Ghitany, M. E.(2019). *The Unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates*. APPLIED STATISTICS.
- [13] Abd El-Monsef, M.M.E; (2021). *Erlang mixture distribution with application on COVID-19 cases in Egypt*. *Intrentional Journal of Biomathematics*.
- [14] Korkmaz, M.Ç., Chesneau, C.; (2021). *On the unit burr-xii distribution with the quantile regression modeling and applications*. *Computational and Applied Mathematics*. 40(1):1—26.