Journal of Engineering and Applied Sciences 14 (Special Issue 2): 5483-5489, 2019
ISSN: 1816-949X

Estimation of Survival Function for Kumaraswam Distribution of BMI

Shrook A.S. AL-Sabbah and Enas Abdu Alhafudh Mohammed
Department of Statistics, Administration and Economics College, University of Kerbala, Kerbala, Iraq
shrookalsabah1968@gmail.com, Enas17stat@gmail.com

Abstract: This study is concerned with the study of the degree of relationship between BMI, age and sex and the estimation of the survival function of Kumaraswam distribution with two parameters by four methods of estimation: MLE, moment, order maximum and Linear moments to determine the effect of obesity on probability of survival by calculating the probabilistic survival values of each BMI value and by experimental study using simulation and by observations generated from the Kumaraswam distribution and for sample sizes 40 and 100, the four methods were linear moments was more efficiency based on statistical standards Akaike AIC, Standard Error (SE) and CV.

Key words: Kumarauamy distribution, BMI, estimate MLE, order MLE, moment, linear moment

INTRODUCTION

This research is an addition to the previous studies and to the researchers in the study of modern statistical distributions through the study of the distribution of Kumaraswam and its characteristics and its relationship with other distributions as well as describing the distribution of time taken until failure which contributes to the good expectations of probable failure times, this research deals real observations represent Body Mass Index (BMI) has a Kumaraswam distribution to determine the effect of obesity on probability of survival by calculating the probabilistic survival values of each (BMI). The study also examined four methods of estimation: the Maximum Likelihood (MLE) moments and order maximum and linear moments (Anonymous, 2011).

Object: Contribute to the detection of Kumaraswam distribution characteristics. Measurement of BMI by age and gender and degree of independence of relationship between them. Know the distribution of data (appropriate distribution) and identify the effect of BMI obesity on the probability of survival compared to estimates of the survival function of the distribution of Kumaraswam between the methods of estimation of the Maximum Likelihood (MLE), moments and order maximum and linear moments.

The theoretical side: The distribution of Kumarauamy distribution double bounded (Kum) is a continuous probability distribution defined on the interval (0.1) and is very similar to beta distribution. The distribution function was defined in 1980 and written as following Cordeiro *et al.* (2012), Elbatal (2013):

$$f_z(Z) = \frac{1}{b-c}pq\left(\frac{z-c}{b-c}\right)^{p-1}\left[1-\left(\frac{z-c}{b-c}\right)^{p}\right]^{a-1} \quad c<z<b$$

Where:
P = Continuous shape parameter (p>0)
q = Continuous shape parameter (q>0)
c, b = Continuous boundary parameter (c<b) domain c<x<b

Symbolizes kum (P, q, c, b). To extract the standard distribution formula when c = 0, b = 1 the function becomes as follow:

$$f_x(x) = pq(x)^{p-1}\left[1-(x)^{p}\right]^{q-1} \quad 0<x<1 \qquad (1)$$

Symbolizes kum (P, q, 0, 1).

Properties

Probability Density Function (PDF): The probability density function for kumaraswamy distribution with the two parameters (p, q) given as follows:

$$f_x(x) = pq(x)^{p-1}\left[1-(x)^{p}\right]^{q-1} \quad 0<x<1 \quad p,q>0$$

Corresponding Author: Shrook A.S. AL-Sabbah, Department of Statistics, Administration and Economics College, University of Kerbala, Kerbala, Iraq, shrookalsabah1968@gmail.com

Cumulative Distribution Function CDF: The cumulative distribution function is defined as a randomized distribution through the following equation:

$$F(x) = 1-(1-x^p)^q \quad 0<x<1$$

The survival function is as follows:

$$S(x) = 1-\left[1-(1-x^p)\right]^q$$

If we have a random variable (X) with a probability distribution known as Kumaraswamy distribution, the moment around zero is expressed as follow:

$$\dot{M}_r(x) = qB\left(1+\frac{r}{p},q\right)$$

where, B (α, β) is a beta function which is:

$$B(\alpha, \beta) = \int_0^1 X^{\alpha-1}(1-x)^{\beta-1}ds = \frac{\tau(\alpha)r(\beta)}{\tau(\alpha+\beta)}$$

where, Γ(v) is a gama function which is:

$$\Gamma(v) = \int_0^\infty t^{v-1}e^{-t}$$

The medium of kum distribution, it can be written as following (Kumaraswamy, 1980):

$$md(x) = w = \left(1-0.5^{\frac{1}{q}}\right)^{\frac{1}{p}} \quad median = f^+(0.5)$$

The rth' non-central moment:

$$E(x)^r = \dot{m}_r = \int_0^\infty x^r f(x)dx$$

When r = 1, then,

$$E(x) = \mu = \dot{m}_1(x) = qB\left(1+\frac{1}{p},q\right)$$

When r = 2, then,

$$E(x)^2 = qB\left(1+\frac{2}{p},q\right)$$

Therefore, the variance for the kum distribution is Lemonte (2011), Jones (2009) and Mitnik (2013):

$$V(x) = \mu_2 = \mu_2(x) - \mu^2$$

$$V(x) = qB\left(1+\frac{2}{p},q\right) - \left[qB\left(1+\frac{1}{p},q\right)\right]^2$$

The mode of the Kum distribution is:

$$Mo = \left(\frac{p-1}{pq-1}\right)^{\frac{1}{p}}$$

Find coefficient of Skewness:

$$C.S = \frac{E(x-\mu)^2}{\sigma^2}$$

$$C.S = \frac{m3(x)}{\sigma^2} - 6$$

Find coefficient of Kurtosis:

$$C.K = \frac{E(x-\mu)^4}{\sigma^4}$$

As m4 (x) represents the fourth moment:

$$C.K = \frac{m4(x)}{\sigma^4} - 3$$

MATERIALS AND METHODS

Estimation of parameters (p, q) (Salman, 2017; Safi and Ahmed, 2013)

Maximum likelihood method: This method is based on finding an estimator that makes the logarithm function possible at its extreme. If we have a random sample (x_1, x_2, x_3, ...,x_n) of the distribution defined by Eq. 1 the maximum likelihood function will be as follows:

$$L(x_1,x_2,...,x_n,p,q) = \prod_{i=1}^{n} f(x_i,p,q)$$

$$L = P^n q^n \prod_{i=1}^{n} x_i^{p-1} \prod_{i=1}^{n}\left[1-x_i^p\right]^{q-1} \quad (2)$$

Taking the logarithm of Eq. 2 as following:

$$Log\,L = n\log p + n\log q +$$
$$(p-1)\sum_{i=1}^{n}\log x_i + (q-1)\sum_{i=1}^{n}\log(1-x_i^p), ...,(\ )$$

By taking the first partial derivative of the Eq. 3-5 for (p, q) and equal to zero we get:

$$\frac{\partial \log L}{\partial p} = \frac{n}{p}\sum_{i=1}^{n}\log xi + (q-1)\sum_{i=1}^{n}\frac{1}{1-x_i^{\,p}}\left[-x_i^{\,p}\log x_i\right] = 0 \quad (3)$$

$$\hat{p}mle = \frac{n}{\sum_{i=1}^{n}\left(1-x_i^{\,p}\right)} \quad (4)$$

$$\frac{\partial \log 1}{\partial q} = \frac{n}{q} + (q-1)\sum_{i=1}^{n}\frac{x_i^{\,q}(\log x_i)}{1-x_i^{\,q}}\sum_{i=1}^{n}\log x_i = 0$$

$$\hat{q} = \frac{n}{k} \quad (5)$$

Where:

$$k = (q-1)\sum_{i=1}^{n}\frac{x_i^{\,q}(\log x_i)}{1-x_i^{\,q}}\sum_{i=1}^{n}\log x_i$$

Therefore, the estimation of the survival function in maximum likelihood method becomes:

$$S(x) = \left(1-x^{\hat{p}mle}\right)^{\hat{q}mle} \quad (6)$$

Method of order maximum likelihood: The estimation of parameter say (p, q) between (x_m, x_{m+1}) observation then the likelihood for ordered observation is :

$$L(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{m} f(x_i)\prod_{i=m+1}^{n} f(x_i) \quad (7)$$

$$L = p^m q^m \prod_{i=1}^{m} xi^{p-1}\prod_{i=m+1}^{n}\left(1-x_i^{\,q}\right)^{p-1} \quad (8)$$

$$\text{linL} = m\ln p + m\ln p + (p-1)\sum_{(i=1)}^{m\ln x} i(p-1) \sum_{(i=m+1)}^{n\ln}\left(1-x_i^{\,q}\right) \quad (9)$$

$$\frac{\partial \ln L}{\partial p} = \frac{m}{p} + \sum_{i=m+1}^{n}\left(1-x_i^{\,q}\right)$$

$$\frac{m}{\hat{p}} + \sum_{i=m+1}^{n}\left(1-x_i^{\,q}\right) = 0 \quad (10)$$

$$\hat{p}_{ORM} = -\frac{m}{\sum_{i=m+1}^{n}\ln\left(1+x_i^{\,q}\right)} \quad (11)$$

$$\frac{\partial \ln L}{\partial q} = \frac{m}{q} + \sum_{i=m+1}^{n}\ln x_i + (p+1)\sum_{i=m+1}^{n}\frac{-x_i^{\,q}\log x_i}{\left(1-x_i^{\,q}\right)}$$

$$\frac{m}{\hat{q}} + \sum_{i=m+1}^{n}\ln x_i + (p+1)\sum_{i=m+1}^{n}\frac{-x_i^{\,q}\log tx_i}{\left(1-xi^{q}\right)} \quad (12)$$

$$\hat{q} = \frac{m}{\sum_{i=m+1}^{n}\ln x_i + (p+1)\sum_{i=m+1}^{n}\frac{-x_i^{\,q}\log x_i}{\left(1-xi^{q}\right)}} \quad (13)$$

Therefore, the estimation of the function of survival in the way of the method of order maximum likelihood becomes:

$$S(x) = \left(1-x^{\hat{p}oml}\right)^{\hat{q}oml} \quad (14)$$

Moments method: Obtaining the moments estimators is achieved by making the sample of moments equal to the population moments as follows:

$$m_r = \frac{\sum_{i=1}^{n} x_i^{\,r}}{n} \quad (15)$$

$$\mu_r = E\left(x^r\right) \quad (16)$$

When applying the following definition:

$$\mu_r = m_r$$

We will get two equations as follows:

$$E\left(x^r\right) = \int_0^{\infty} x^r f(x)dx \quad (17)$$

$$\int_0^1 x^r pq(x)^{p-1}\left[1-(x)^p\right]^{q-1}dx \quad (18)$$

$$= pq\int_0^1 x^{r+p-1}\left[1-(x)^p\right]^{q-1}dx \quad (19)$$

Let

$$1-(x)^p = z \qquad x = (1-z)^{\frac{1}{p}};\ \frac{dx}{dz} = \frac{-dz}{p(1-z)^{1-\frac{1}{p}}}$$

After several algebraic operations we get:

$$E\left(x^r\right) = pq\frac{\Gamma\left(\frac{r}{p}+1\right)\Gamma(q)}{\Gamma\left(\frac{r}{p}+1+q\right)} \quad (20)$$

When r = 1, we get

Table 1: BMI classification

BMI	The weight	Risk of diseases of the age
<18.5	Weight less than normal	Few
18.5-24.9	Weight within normal range	Within the normal range the risk of infection is reduced
25-29.9	Weight gain more than normal	The risk of disease increases
39.9-30	Obesity	The risk of disease is large
more than 40	Excessive obesity	Very high risk of disease

$$E(x)^1 = pq\frac{\Gamma\left(\frac{1}{p}+1\right)\Gamma(q)}{\Gamma\left(\frac{1}{p}+1+q\right)} = \bar{x} \quad (21)$$

When r = 2, we get

$$E(x)^2 = pq\frac{\Gamma\left(\frac{2}{p}+1\right)\Gamma(q)}{\Gamma\left(\frac{2}{p}+1+q\right)} = \frac{\sum_{i=1}^{n} x^2}{n} \quad (22)$$

We have $pq\frac{\Gamma\left(\frac{1}{p}+1\right)\Gamma(q)}{\Gamma\left(\frac{1}{p}+1+q\right)} = \bar{x}$ From eqution:

$$\bar{x}\Gamma\left(\frac{1}{p}+1+q\right) = pq\Gamma\left(\frac{1}{p}+1\right)\Gamma(q) \quad (23)$$

$$\bar{x}\Gamma\left(\frac{1}{p}+1+q\right) = p\Gamma\left(\frac{1}{p}+1\right)\Gamma(q+1) \quad (24)$$

$$\hat{p}MOM = \frac{\bar{x}\Gamma\left(\frac{1}{p}+1+q\right)}{p\Gamma\left(\frac{1}{p}+1\right)\Gamma(q+1)} \quad (25)$$

To find the moments, we solve the Eq. (1-25) in numerical ways according to Newton-Raphson method. Therefore, the estimation of the function of survival in the manner of moments method becomes:

Applied side: Obesity increases the risk of many diseases, especially, heart disease, type 2 diabetes, sleep apnea, certain types of cancer and osteoporosis. Obesity usually results from a combination of excess calories with little physical activity and genetic effects. So, although, few cases occur primarily because of genes, endocrine disorders, medications and mental illness. To measure the amount of obesity we need to know BMI.

Body Mass Index (BMI): Body Mass Index (BMI) is the internationally recognized standard for distinguishing overweight or obese from ideal weight. It reflects the relationship between a person's weight and height and is recognized by the US national institute of health and the world health organization as the best standard for measuring obesity.

Method of calculating BMI: Equation: Body mass index = Weight (kg)/(length)^2 (meters) the weight (kg) is divided by the square length (meter) and the number produced is divided according to, Table 1 is it normal weight or less than normal or increased weight or obesity.

RESULTS AND DISCUSSION

The study sample: The study was conducted on a simple random sample with size (204) from 2011-2014 for adults and children in the primary health care center in Al-Abbas district in Al-Ghadeer District/Karbala Health Directorate. The BMI was calculated by dividing weight on square meters in length. The Statistical Application (SPSS) and (Easy Fit) were used to determine the distribution type and the matlab program.

For the purpose of application, the age was divided into the following age classes: <10 years with a No. of 1 and 10-19 years with a No. of 2 and from 20-29 years with a code No. 30-39. A year with a No. of 4 and 40-49 years with a symbol No. 5 and >50 symbols with a No. 6.

As a symbol of BMI values in symbols 1 to <18.5, 2 to (24.9-18.5), 3 to (29.9-25) and 4 to (39.9-30) and 5 to <40. Through the implementation of the program (SPSS) BMI was measured by age and degree of correlation between them and as Table 2.

Table 2 shows that the highest percentage was for the first age class (<10) and the value of BMI is less than the normal rate (few) (23.5%), followed by the second age class (10-19) and BMI is less than the normal rate (9.8%) it is equal to the percentage of the second age class that has a BMI within the normal range, followed by the fourth age class (30-39) and has a high BMI which is obese with 7.4%.

For the purpose of knowing the variables, the age of the person and the MBI value are independent or not. The value of the square (χ^2) was equal to (168.175) and the p-value (0.00) this indicates a significant relationship between the two variables.

Body Mass Index (BMI) was measured by sex and the degree of correlation between them and as in Table 3. For the purpose of knowing the variables, the age of the

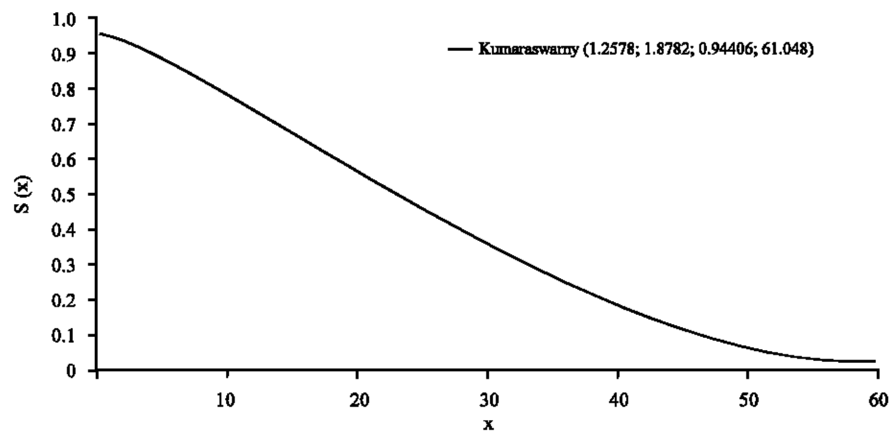

Fig. 1: Probability survival values (Survival function)

Table 2: Age ratios and corresponding BMI values

	Age						
BMI values	1.00	2.00	3.00	4.00	5.00	6.00	Total (%)
1.00 Count	48	20	3	1	0	0	72
Total (%)	23.5	9.8	1.5	0.5	0.0	0.0	35.3
2.00 Count	5	20	9	3	0	1	38
Total (%)	2.5	9.8	4.4	1.5	0.0	0.5	18.6
3.00 Count	0	10	12	9	5	4	40
Total (%)	0.0	4.9	5.9	4.4	2.5	2.0	19.6
4.00 Count	0	3	12	15	9	9	48
Total (%)	0.0	1.5	5.9	7.4	4.4	4.4	23.5
5.00 Count	0	0	1	3	2	0	6
Total (%)	0.0	0.0	0.5	1.5	1.0	0.0	2.9
Total Count	53	53	37	31	16	14	204
Total (%)	26.0	26.0	18.1	15.2	7.8	6.9	100.0

Table 3: The sex and BMI values corresponding to them

	Sex		
BMI values	Male	Male	Total (%)
1.00 Count	29	43	72
of Total (%)	14.2	21.1	35.3
2.00 Count	5	33	38
Total (%)	2.5	16.2	18.6
3.00 Count	5	35	40
Total (%)	2.5	17.2	19.6
4.00 Count	8	40	48
Total (%)	3.9	19.6	23.5
5.00 Count	0	6	6
Total (%)	0.0	2.9	2.9
Total Count	47	157	204
Total (%)	23.0	77.0	100.0

person and the MBI value are independent or not. We conduct the Chi-square test (χ^2) for independency where the p-value was (0.00) this indicates a significant relationship between the two variables.

To determine the distribution of the data, (Easy Fit) was used and it was found that the distribution of Kumaraswam was primarily by parameters ($\alpha1$ = 1.2578, $\alpha2$ = 1.8782, a = 0.94406, b = 61.043) according to the Kolmogorov-Smirnov test. To determine the effect of obesity on the probability of survival, the probability survival values for each BMI value in withdrawn sample were calculated as in Table 4. From Table 4, we observe that the probabilistic values of the survival function are reduced by increasing the body mass of Kumaraswamy distribution and to illustrate, the survival function was graphically shown as in Fig. 1.

Experimental side (simulation): By using simulation we generated observations from the Kumaraswam distribution for sample sizes 40 and 100 and the parameters estimated by researcher Salman (2017) to calculate the survival function estimates for each of the four methods and to choose a practical estimation method based on statistical criteria (Akaike AIC, standard error SE and CV coefficient in Table 5.

By using simulation we generated with sample sizes 40 and 100 by the researcher Muna Shaker Salman for examine the four methods based on statistical standards Akaike AIC, standard error SE and CV) as in Table 5. By using the statistical criteria SE, CV, AIC in the comparison it was found that the L-Moment method was the best for both samples (40,100) because it possessed the lowest values for each of the previous criteria.

Table 4: Probability survival values

BMI	Survival	BMI	Survival
10.170000	0.829580	20.500000	0.591930
10.679000	0.818230	20.700000	0.587320
11.223000	0.806000	21.000000	0.580430
11.342000	0.803330	21.500000	0.568960
11.747000	0.794170	21.800000	0.562100
12.000000	0.788420	22.000000	0.557530
12.002000	0.788380	22.321000	0.550210
12.004000	0.788330	23.000000	0.534790
12.228000	0.783230	23.800000	0.516720
12.311000	0.780890	23.804000	0.516630
12.356000	0.780320	23.880000	0.514930
12.500000	0.777030	24.000000	0.512230
12.977000	0.766110	25.000000	0.489880
13.000000	0.765590	25.200000	0.485440
13.080000	0.763750	25.300000	0.483220
13.219000	0.760560	26.000000	0.467760
13.417000	0.756010	26.130000	0.464910
13.879000	0.745370	26.250000	0.462280
14.000000	0.742580	26.400000	0.458990
14.096000	0.740370	26.500000	0.456800
14.120000	0.739810	27.000000	0.445910
14.123000	0.739740	27.200000	0.441570
14.343000	0.734660	27.500000	0.435080
14.363000	0.734200	28.000000	0.424330
14.366000	0.734130	29.000000	0.403050
14.574000	0.729320	29.600000	0.390440
14.609000	0.728510	29.660000	0.389180
14.619000	0.728280	29.700000	0.388350
14.792000	0.724280	30.000000	0.382100
14.860000	0.722700	30.400000	0.373810
15.000000	0.719460	30.500000	0.371750
15.158000	0.715800	30.850000	0.364560
15.240000	0.713900	31.000000	0.361490
15.420000	0.709720	32.000000	0.341240
15.600000	0.705540	32.390000	0.333450
15.606000	0.705400	32.400000	0.333250
15.608000	0.705360	32.800000	0.325320
15.700000	0.703220	33.000000	0.321380
16.000000	0.696260	33.200000	0.317460
16.020000	0.695790	33.800000	0.305790
16.159000	0.692560	34.000000	0.301930
16.349000	0.688150	35.000000	0.282890
16.500000	0.684640	35.170000	0.279700
16.640000	0.681390	35.400000	0.275400
16.833000	0.676900	36.000000	0.264310
17.000000	0.673020	37.000000	0.246180
18.000000	0.649780	37.700000	0.233770
18.219000	0.644700	38.000000	0.228530
18.400000	0.640500	39.000000	0.211380
18.860000	0.629830	40.000000	0.194750
19.000000	0.626590	40.500000	0.186630
19.578000	0.613210	41.000000	0.178650
20.000000	0.603460	46.000000	0.106830
		50.220000	0.058710

Table 5: Estimation of the parameters (p, q) and the survival function of the distribution (Kumaraswamy) at the sample sizes (40,100)

Sample size (n)/Method	$\hat{p}$	$\hat{q}$	SE($\hat{s}$)	CV($\hat{s}$)	AIC($\hat{s}$)	Best
40						
L-Moment	1.022	0.342	0.02690324	0.21474361	-26.7666	L-Moment
Moment	1.064	1.019	0.04491147	0.4947141	-17.8645	
MLE	1.113	1.112	0.04623380	0.51838991	-17.3604	
Order-MLE	2.298	1.486	0.04934431	0.44911795	-16.2293	
100						
L-Moment	1.026	0.341	0.02109537	0.286638	-63.5813	L-Moment
Moment	0.995	0.999	0.03040101	0.623019	-47.7112	
MLE	1.010	1.027	0.03056715	0.628284	-47.4745	
Order-MLE	2.015	1.409	0.03429697	0.585833	-42.4744	

CONCLUSION

In the study of the frequency table of ratios, it was observed that the highest percentage of children in the first age class (<10) and the second age class (10-19) years had a value of BMI less than the normal rate (few) they are malnourished. The fourth age group (30-39) has a high BMI and is obese and susceptible to disease. A strong relationship between sex and age with BMI. The probabilistic values of the survival function decrease with the increase of the BMI of the distribution Kumaraswamy, which means the increased probability of diseases leading to death. The simulation results showed that the L-Moment method was the best method because it achieved the lowest value for the statistical criteria (Akaike AIC, standard error SE and CV difference coefficient).

REFERENCES

Anonymous, 2011. Healthy snacks for children: 10 tips. Web Medicine, Kirkuk, Iraq. https://www.webteb.com/

Cordeiro, G.M., R.R. Pescim and E.M.M. Ortega, 2012. The Kumaraswamy generalized half-normal distribution for skewed positive data. J. Data Sci., 10: 195-224.

Elbatal, I., 2013. Kumaraswamy linear exponential distribution. Pioneer J. Theor. Appl. Stat., 5: 59-73.

Jones, M.C., 2009. Kumaraswamys distribution: A ß-type distribution with some tractability advantages. Stat. Methodol., 6: 70-81.

Kumaraswamy, P., 1980. A generalized probability density function for double-bounded random processes. J. Hydrol., 46: 79-88.

Lemonte, A.J., 2011. Improved point estimation for the Kumaraswamy distribution. J. Stat. Comput. Simul., 81: 1971-1982.

Mitnik, P.A., 2013. New properties of the Kumaraswamy distribution. Commun. Stat. Theor. Methods, 42: 741-755.

Safi, S.K. and R.H. Ahmed, 2013. Statistical estimation based on generalized order statistics from Kumaraswamy distribution. Proceedings of the 15th International Conference on Applied Stochastic Models and Data Analysis (ASMDA), June 25-28, 2013, ASMDA International Society, Mataro, Spain, pp: 1-8.

Salman, M.S., 2017. Comparing different estimators of two parameters Kumaraswamy distribution. J. Babylon Univ. Pure Appl. Sci., 25: 395-402.