Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Estimation of Time of Survival Rate by Using Clayton Function for the Exponential Distribution with Practical Application

Muna Kadhim Jewi
Central Technical University / Center for Continuing Education
Baghdad/ Iraq
Joumuna38@gmail.com

Received:10/11/2021 **Accepted: 9/1/2022** **Published: June / 2022**

Abstract

Each phenomenon contains several variables. Studying these variables, we find mathematical formula to get the joint distribution and the copula that are a useful and good tool to find the amount of correlation, where the survival function was used to measure the relationship of age with the level of cretonne in the remaining blood of the person. The Spss program was also used to extract the influencing variables from a group of variables using factor analysis and then using the Clayton copula function that is used to find the shared binary distributions using multivariate distributions, where the bivariate distribution was calculated, and then the survival function value was calculated for a sample size (50) drawn from Yarmouk Hospital for renal patients, where nephritis directly affects human life because of its risks related to an increase in the possibility of permanent damage to the kidneys or infection that may cause failure. Renal life-threatening and the survival function represents the time when a particular system stops working After applying the model, it was found from the results of the research that there is an effect of an increase in the level of creatinine in the blood with age on the survival function ratio, as the higher the creatinine percentage leads to a decrease in survival, which means that the period of time lived by a person with an increase in creatinine in the blood decreases the higher it is in patients.

Keywords: **survival function, creatinine, factor analysis, Clayton bond function, unary distributions, co-binary distribution .**

1. Introduction

Connections are a good option to find the relationship between variables, where there is no phenomenon that does not contain variables that affect them and these variables are not independent of each other and Clayton function is one of the functions that are used to find the common binomial distributions of monolithic distributions, whose domain is $[0,1]^n$. Therefore, this function is a useful tool for constructing binomial distributions for two or several variables over the past years, where there has been an increasing interest in building distributions. For variables and the study of the reliability structure in 2017, Ivan Kojadinovic presented a model of continuous distribution of binary variables using the simulation principle and with iteration of 1000 sets of statistical software packages R, and the best value was selected from four values of the Kendall correlation and was applied to a set of data An insurance company for two variables are payout and loss, and measure the reliability of the study variables, and it was proven that the links are a good way to measure reliability with many statistics[I. Kojadinovic, 2017, 24-41][1]. In 2017, the researchers Zongwen Ann & Daoming Sun presented a set of links for a fixed-level failure-measuring system to Sandia National Laboratories of the China National Natural Science Foundation for Reliability and Safety Engineering and proved that the Campbell bond is the best[Z. An and D. Sun, 2017, 129 -138] [2].
Presented by Yao Yan, Xu jayu, Jin Di & Guo Gaopeng in 2019, a multi-purpose stochastic planning method to access the photovoltaic cells of the distribution network based on the theory of links and the cosine method, where a multi-level programming model was proposed to improve the location of photovoltaic access in the active distribution network. [K. Sun *et al.*,2019] [3]

Presented by Chen Sanqiang Chen Hongyu, Wu Xianguo Qin Wenwei Tang Yangyi in 2019 proposed a copula-cloud model to assess the levels of water leakage risk in tunnels. A model was established that combines the linkage function and the clouds model. The research sample is divided into blocks of five stations (Xunlimen Station, Zhongnan Road Station and Jianghan Road Station on the Wuhan Metro Line 2, Wangjiawan Station and Hong Kong Road Station on Line 3.
[& T.Y. Chen Sanqiang, Chen Hongyu, Wu Xianguo, Qin Wenwei,2019][4]

Annalisa Diclemente and Claudio Romano presented a study in 2021 on the importance of using links in financial applications to choose the best link that fits financial data in different types that was applied to a historical data set of Italian stock returns and prove that the couple is a useful tool to implement it to build simulation algorithms .
[Di Clemente, A,& Romano,c.2021][14]

Also, Mojca Sraj, Nejc Bezak, and Mitja Brilly presented a study in 2015 on the frequency of floods for the Litija station on the Sava River, using a set of links for binary variables to find the best association. The data is from the Slovenian Research Agency, the research[Mojca Sraj, Nejc Bezak,& Mitja Brilly,2015] [15]

To find the relationship between the variables in a mathematical formula to get the joint distribution and the links is a useful, good and better tool than the traditional methods for finding the amount of correlation of the function and finding the survival time.

The nonparametric Clayton function was used to find the co-distribution by estimating the value of ($\widehat{\theta}$)within the limits of Association, which represent same untre liability of function calculated for Clayton function through shared values and find time stay drawn from sample.

2. Marginal Distributions:

Known distributions of marginal distributions Regular and domain $[0,1]^n$ and it is a multivariate function that has a set of conditions for binary distributions: [R.B. Nelsen, 2007][5]
1- for each *v, u* in I be

C(u,0)=0=C(0,v)
C(u,1) = u and C(1,v)= v

2- each (u_1, u_2, v_1, v_2) **in (I) so that** $u_1 \leq u_2$ **and** $v_1 \leq v_2$
that $C(u_2, v_2) - (u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$

3. Theory (Sklar's) [R.B . Nelsen,][5]

The links were defined through Sklar's theorem, link which is considered the theoretical basis for the function. We suppose that *F* represents a distribution function.
So,$F_1(y_1)$ F_P.....(P_1) represents its boundary functions, then the C bond is present and in the following formula :

$$F(Y_1, \ldots\ldots Y_p) = C(F_1(y_1), \ldots\ldots F_p(y_p))$$

If F_1,..... F_P all continuous functions, then C is function single and unique, where C is determined only in field range (Ran F_1 x…..x Ran F_P...). On the contrary, if C is the connection method for p of the distribution functions, then the function F is defined as p from the dimensions of the continuous distribution functions F_1,..... F_P boundary, for u in the domain $[0,1]^p$ is:

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots\ldots F_p^{-1}(u_p))$$

4. The Clayton Bond

It is a useful non-parametric bond with domain $[0,1]^n$; it's a cumulative distribution function in the following form :
[A. Charpentier, 2003] [6]

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{\frac{-1}{\theta}} \quad 0 < \theta < \infty \qquad \ldots\ldots(1)$$

Formula for the probability density function (p.d .f) is as follows:

$$c(u,v) = (vu)^{-(\theta+1)}(1+\theta)\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}-2} \quad \text{.....(2)}$$

After a value is estimated $\hat{\theta}$ from the field of the function, the survival function is estimated from equation (6) equal to the value of the Clayton bond and after simplifying the equation by substituting for a value with its equivalent in the chosen exponential distribution to estimate the survival function, since R(t):

$$R(u,v) = 1 - \left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}} \quad \ldots\ldots (3)$$

We substitute for each u and with v equation (4)

$$R(y_1,y_2) = 1 - [(1-e^{-\sigma 1 y 1})^{-\theta} + (1-e^{-\sigma 2 y 2})^{-\theta} - 1]^{\frac{-1}{\theta}} \quad \text{....(4)}$$

5. Exponential Distribution

It is a continuous probability distribution whose name is derived from the exponential function and is used in estimating the time periods between the occurrence of events and to find the reliability of data life in public life tests and issues related to the measurement of time, to measure the duration of a telephone call, the waiting time for customer service to the mail and service counters, the period of unloading a freighter, the period of repairing a machine and research on loss of insurance for the financial field and in the exact sciences uses an exponential distribution where the median expresses the moment at which half of the original population remains in it and an exponential distribution is used to represent the life span of any phenomenon if its mean is constant 1/ σ . [Gupta, R.D ,&Kundu,D.2007][12]

Representing the probability distribution function (p.d.f) as follows:

$$f(y,\sigma) = \begin{cases} \sigma e^{-\sigma y} & ,y \geq 0 \\ 0 & ,y < 0 \end{cases}$$

where σ is the measure parameter and the distribution function:

$$F(y,\sigma) = 1 - e^{-\sigma y} \text{ for } y \geq 0 \quad \ldots\ldots(5)$$

The survival function is as follows: [Huang , M.L. , 2008][13]

Assuming T symbolizes the time life system, the survival of this system at moment time, t which denote by the $R(t)$ symbol, know following relationship:

$$R(t) = P(T > t)$$

called stay at moment t, $R(t)$ function stay and be a possibility that time can hang him system t over.

Survival theory is concerned with some problems that deal with calculating the actual bid probability of some systems at a specific time or at an optional time, or at a part of the time during which some systems operate efficiently and accurately.

The survival function can be expressed in terms of the cumulative distribution function of the random variable T as follows:

$$R(t) = 1 - P(T \leq t)$$

$$R(t_1, t_2) = 1 - F(t_1, t_2).....(6)$$

Where:

F(t_1, t_2) represents the cumulative distribution function.

6. Factor Analysis :

It is an important statistical method for analyzing multivariate data to simplify it and identify the main elements of that phenomenon and explore the common characteristics of multi-linked data to find the link between them, to identify the main factors and elements of the correlation of variables and to facilitate the process of their interpretation. [S.A. Mulaik, 2009][7]

Factor analysis leads to the extraction of factors to explain the correlation of a variable from a group of variables, and factor analysis is of two types. The first is exploratory (EFA) to identify the phenomenon and explore its main dimensions and the basic variables affecting it to the least number of variables and discover the basic variables affecting it. And the second confirmatory (CFA) is used to test the hypothesis that there is a connection between the variables and the relationship between them and to build a model according to the theory from which they started research in a statistical way to determine the accuracy of matching the data used and to extract the factor is the main component analysis (the main factors) (PCA) [D. F. Polit and C. T. Beck,2008][8] where it is the first stage. It aims to reduce the variables from a set of data as the PC provides the required variables, [D. Child, 2006][9]

7. Applied Aspect

Chronic Renal Failure disturbance in the kidneys that leads to the accumulation of waste products and excess fluid in the body. The kidneys are slow in the first stage, where no symptoms of the disease are distinguished, and the condition increases and worsens, and symptoms begin with the worsening of the condition. Therefore, the disease is not discovered until late stages.

[P. Chauveau, C. Combe, D. Fouque, and M. Aparicio, 2013] [10]

Symptoms of failure vary from person to person. Some people in the early stage of the disease may not feel symptoms or notice symptoms when they occur, when kidney failure at work and the accumulation of waste products in the blood and the body, which leads to an increase in the proportion of nitrogen in the blood, then few symptoms appear .

Food has clear effects on the prevention and development of kidney failure, especially for people with diabetes. Food that contains a high percentage of animal or vegetable proteins has a negative impact on kidney function on some people.

[M.S Rughooputh, R. Zeng, and Y. Yao, 2015] [11]

Kidney failure is diagnosed from clinical examinations with laboratory blood tests such as elevated urea and creatinine in the blood in addition to the degree of calcium, phosphorous, potassium, and sodium. These indicators reflect the performance level of the college.

8. Results and Discussion

The Spss program was used to calculate the factor analysis to extract the influencing factors from the set of factors in order to reduce the factors to the lowest number for the purpose of finding the values of the survival function for the binary co-distribution of the exponential distribution using the association function as in equation (2) we follow the following steps:

i. Estimation A value θ through the equation is the domain of the association function and it was chosen $\theta = 4$

ii. To give estimated values for the unilateral distribution parameter [$\sigma_1 = 0.3, \sigma_2$ =0.1] taken from previous studies.

iii. The value account survival function R(^{x,y}) through equation (3) for purpose calculating value R using MATLAB program, so our symbol ratio Alalkenininin blood symbol (x) and patient age symbol (y) as in following Table 1:

Table (1)

R(t)	Creatinine (X)	Age (Y)	Sequence
0.4909	10.8	45	1
0.7152	3.1	48	2
0.6549	4.1	53	3
0.5784	8.99	15	4
0.4860	11.1	63	5
0.58751	5.89	65	6
0.50299	9.89	70	7
0.4734	12.1	83	8
0.7178	3.02	64	9
0.5181	8.98	51	10
0.5799	6.15	66	11
0.5156	9.1	72	12
0.66707	3.91	25	13
0.5550	7.12	75	14
0.56725	6.36	50	15
0.51715	9.05	47	16

0.5553	7.11	61	17
0.5205	8.82	76	18
0.5113	13.39	22	19
0.79931	2.11	64	20
0.63737	4.5	37	21
0.8110	2.01	55	22
0.5382	7.9	2352	23
0.3214	14.29	2450	24
0.6792	3.63	2563	25
0.6047	5.37	2637	26
0.7043	3.89	2718	27
0.6928	3.4	2843	28
0.6658	3.88	2966	29
0.5682	6.68	32	30
0.3359	9.28	50	31
0.7470	3.1	3255	32
0.7084	11.6	35	33
0.7130	1.9	3465	34
0.7159	2.3	75	35
0.6973	3.6	3665	36
0.6705	9.89	3770	37
0.6837	12.1	3827	38
0.6979	5.1	3964	39
0.3590	8.98	4051	40
0.7090	5.3	4148	41
0.7003	9.1	25	42
0.7396	8.6	4322	43

0.7091	5.2	4448	44
0.3905	14.9	45	45
0.6957	5.8	4655	46
0.6138	7.9	4775	47
0.7076	11.5	80	48
0.8704	3.89	22	49
0.7891	3.4	64	50

Table (2) Rotated component matrix

	component		
	1	2	3
creatine	.759		
potassium	.721		
Sodium	-.677-		-.314-
urea	.497		
Age	.303	.848	
phosphate	.310	-.727-	
blood pressure			.856
hemoglobin		.302	.556

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.610
Bartlett's Test of Sphericity	Approx. Chi-Square	49.771
	df	28
	Sig.	.007

9. Conclusions

1- Through the results we reached in the practical side, it was shown that the survival function ratio generally decreases due to the high level of creatinine in patients, and through Table No. 1, we note that the survival rate was 0.8110 for patient No. 22 compared to a creatinine ratio of 2.1, while the survival rate for patient No. 24 was 0.4514 compared to a creatinine ratio of 14.29 in the blood, which leads to death in most cases.

2- The ratio of the value of the magnitude $\hat{\theta}$ has a significant impact when calculating the value of the survival function because it affects the proportion of the harmony that occurs for the joint distribution extracted from the link function.

10. Recommendations

i. **To find common distributions, we recommend using the Clayton function.**
ii. **The use of statistics to develop mathematical foundations that enable the measurement of survival time by using the Clayton association on other pathological cases of health institutions.**
ii. **The researcher recommends using the Clayton function for the purpose of measuring the survival time of the dependency function in industrial fields.**

References

[1] I. Kojadinovic, "Some copula inference procedures adapted to the presence of ties," Comput. Stat. Data Anal., vol. 112, pp. 24–41, 2017.
[2] Z. An and D. Sun, "Reliability modeling for systems subject to multiple dependent competing failure processes with shock loads above a certain level," Reliab. Eng. Syst. Saf., vol. 157, pp. 129–138, 2017.
[3] K. Sun et al., "Multi-objective bi-level photovoltaic access planning based on copula theory and included angle cosine method," in IOP Conference Series: Earth and Environmental Science, 2019, vol. 227, no. 3, p. 32016.
[4] & T. Y. Chen Sanqiang, Chen Hongyu, Wu Xianguo, Qin Wenwei, "Water leakage risk assessment of subway operating tunnels based on Copula-cloud model. Collection," 2019.
[5] R. B. Nelsen, An introduction to copulas. Springer Science & Business Media, 2007.
[6] A. Charpentier, "Tail distribution and dependence measures," in Proceedings of the 34th ASTIN Conference, 2003, pp. 1–25.
[7] S. A. Mulaik, Foundations of factor analysis. CRC press, 2009.
[8] D. F. Polit and C. T. Beck, Nursing research: Generating and assessing evidence for nursing practice. Lippincott Williams & Wilkins, 2008.
[9] D. Child, The essentials of factor analysis. A&C Black, 2006.
[10] P. Chauveau, C. Combe, D. Fouque, and M. Aparicio, "Vegetarianism: advantages and drawbacks in patients with chronic kidney diseases," J. Ren. Nutr., vol. 23, no. 6, pp. 399–405, 2013.
[11] M. S. Rughooputh, R. Zeng, and Y. Yao, "Protein diet restriction slows chronic kidney disease progression in non-diabetic and in type 1 diabetic patients, but not in type 2 diabetic patients: a meta-analysis of randomized controlled trials using glomerular filtration rate as a surrogate," PLoS One, vol. 10, no. 12, p. e0145505, 2015.

[12] Gupta,R.D,&Kundu,D.'' Generalized exponential distribution: Existing results and some recent developments '' Journal of statistical planning and inference, 137(11),3537-3547,2007.

[13] Huang , M.L. ''Aweighted estimation mwthod for survival function''App math Sci,2 (16), 62-753,2008 .

[14] Di clemente , A., & Romano,'' Calibrating and Simulation copula Function in Financial Application'' .font. Appl . math. stat. 7: 642210. Doi : 10.3389/ fams,2021.

[15] Sraj, M.,Bezak, N., &Brilly, M.'' Bivariate flood frequency analysis using the coupula function' 'a case study of the Litija station on the Sava River. Hydrological Processes, 29 (2) ,225-238, 2015.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.040	25.499	25.499	2.040	25.499	25.499	2.037	25.463	25.463
2	1.552	19.404	44.903	1.552	19.404	44.903	1.390	17.372	42.834
3	1.176	14.696	59.599	1.176	14.696	59.599	1.341	16.765	59.599
4	.992	12.403	72.002						
5	.739	9.241	81.243						
6	.552	6.902	88.145						
7	.527	6.583	94.728						
8	.422	5.272	100.000						

Extraction Method: Principal Component Analysis.

Total Variance Explained

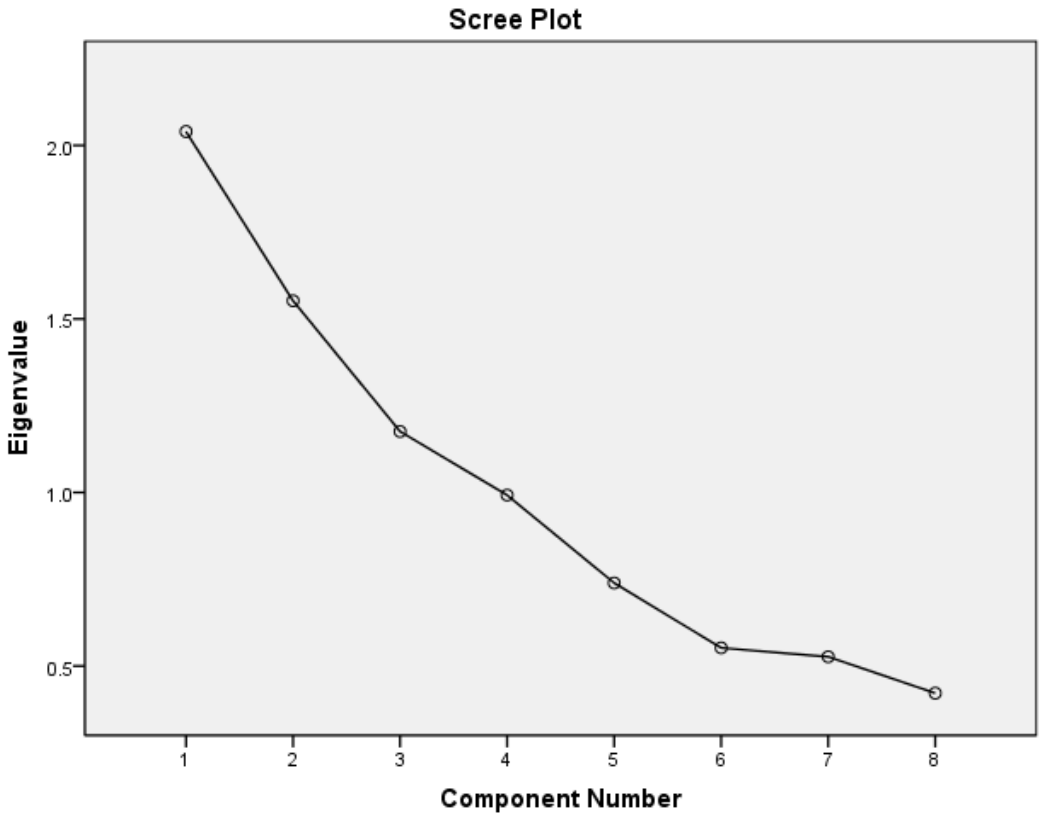
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.996	24.955	24.955	1.996	24.955	24.955	1.993	24.914	24.914
2	1.557	19.458	44.413	1.557	19.458	44.413	1.396	17.448	42.362
3	1.174	14.673	59.087	1.174	14.673	59.087	1.338	16.724	59.087
4	.985	12.316	71.402						
5	.779	9.737	81.139						
6	.571	7.134	88.273						
7	.521	6.515	94.788						
8	.417	5.212	100.000						

Extraction Method: Principal Component Analysis.

communalities

	initial	Extraction
creatine	1.000	.701
potassium	1.000	.540
Sodium	1.000	.564
urea	1.000	.333
Age	1.000	.831
phosphate	1.000	.594
blood pressure	1.000	.791
hemoglobin	1.000	.415

Extraction Method: Principal Component Analysis



تقدير زمن البقاء باستخدام دالة كلايتون للتوزيع الاسي مع تطبيق عملي

م.م. منى كاظم جوي
الجامعة التقنية الوسطى/ مركز التعليم المستمر
بغداد/ العراق
Joumuna38@gmail.com

Received:10/11/2021 Accepted: 9/1/2022 Published: June / 2022

مستخلص البحث

كل ظاهرة تحتوي على عدة متغيرات ولدراسة هذه المتغيرات علينا ايجاد صيغة رياضية للحصول على التوزيع المشترك والروابط اداة مفيدة وجيدة لايجاد مقدار الترابط حيث تم استخدام دالة البقاء لقياس علاقة العمر مع مستوى الكرياتينين بالدم المتبقي للشخص كما تم استخدام برنامج Spss لاستخراج المتغيرات المؤثرة من مجموعة من المتغيرات باستخدام التحليل العاملي وبعدها استخدام الدالة الرابطة كلايتون التي تستخدم لايجاد التوزيعات الثنائية المشتركة باستخدام التوزيعات الاحادية حيث تم احتساب التوزيع الثنائي المشترك ومن ثم حساب قيمة دالة البقاء لحجم عينة (50) مسحوبة من مستشفى اليرموك لمرضى الكلوي حيث يؤثر التهاب الكلوي بشكل مباشر على حياة الانسان لما له من مخاطر تتعلق بزيادة احتمالية ، ضرر دائم للكلى او قد تسبب عدوى بفشل كلوي تهدد الحياة وتمثل دالة البقاء زمن توقف نظام معين عن العمل، وبعد تطبيق النموذج تبين من نتائج البحث ان هنالك تاثير لارتفاع مستوى الكرياتينين بالدم مع العمر على نسبة دالة البقاء حيث كلما زادت نسبة الكرياتينين تؤدي الى انخفاض البقاء مما يعني ان الفترة الزمنية التي يعيشها الشخص المصاب بزيادة بالكرياتينين في الدم تقل كلما ارتفع لدى المرضى .

المصطلحات الرئيسة للبحث: دالة البقاء , الكرياتينين ,التحليل العاملي, دالة الرابطة كلايتون, التوزيعات الأحادية , التوزيع الثنائي المشترك .