https://africanjournalofbiomedicalresearch.com/index.php/AJBR
Afr. J. Biomed. Res. Vol. 28(2s) (January 2025); 451-457
Research Article

Statistical Evaluation of Glucose Level Distributions among Women: Implications for Thyroid Function Analysis

T. Manjusha[1, 2], P. Pranay[3*], B. Navatha[4], V.V. Haragopal[5]

[1]Research Scholar, Department of Mathematics and Statistics, Chaitanya (Deemed to be) University Kishanpura, Hanamkonda Telangana, India
[2]Assistant Professor, Department of Statistics, Little Flower Degree College, Uppal, Hyderabad Telangana, India, tmanjusha99@gmail.com
[3*] Associate Professor, Department of Mathematics and Statistics, Chaitanya (Deemed to be) University Kishanpura, Hanamkonda Telangana, India, pettempranay@gmail.com
[4]Assistant Professor, Department of Mathematics, Presidency University, Bengaluru, avatha.b@presidencyuniversity.in
[5]Professor, Department of Statistics, BITS Pilani, Hyderabad, haragopal_vajjha@yahoo.com.

***Corresponding Author:** P.Pranay
*Associate Professor, Department of Mathematics and Statistics Chaitanya (Deemed to be) University Kishanpura, Hanamkonda, Telangana, India

ABSTRACT
Background: The current study takes up issues of glucose levels in the background of thyroid health among women aged 15-49, with 188 and 365 samples from the NFHS-5 dataset.
Objectives: To study the relationship between glucose levels and thyroid health among women. Whether or not the glucose levels depart from a symmetric distribution. Abnormalities in thyroid function and related metabolic processes shall be investigated.
Methods: Data were sourced from the NFHS-5 survey. The parameters included the levels of glucose and indicators of thyroid health. A proper study of these data was conducted by fitting 25 distributions, which would indicate the most appropriate model for glucose levels. There are three primary statistical tests that can be employed to evaluate the quality of each distribution's fit: the Kolmogorov-Smirnov one-sample test, the Anderson-Darling test, and the chi-square test.
Results: The four-parameter generalised gamma distribution was ranked as the best fitting model for the glucose dataset, with the beta distribution ranked second. Using all 365 samples, the beta distribution was ranked highest; with a reduced sample size of 188, the best ranking distribution was the gamma distribution.
Novelty: Its novelty comes from a way of understanding the relationship between glucose levels and thyroid disorders in women aged 15-49 years. According to that study, glucose levels were found not symmetrically standardised, hence the potential abnormalities of thyroid function and its associated metabolic process. This has not been extensively documented in the literature and thus presents a novel contribution to both the fields of endocrinology and biostatistics.

Keywords: Glucose levels, Kolmogorov -Smirnov test; Anderson-Darling test; Chi-Square test; distribution fitting: Probability distribution.

***Authors for correspondence: E-mail Id:** pettempranay@gmail.com*

Received - 02/10/2024 *Acceptance - 11/10/2024*

DOI: https://doi.org/10.53555/AJBR.v28i2S.2543

1. INTRODUCTION

Thyroid disorders in women are very common and affect gross health and well-being. Being one of the most important endocrine glands and regulating metabolism by producing hormones, the thyroid affects many physiologic processes in the body. Approximately forty million people in India are already suffering from such disorders, which are assuredly evolving into a significant public health issue. The prevalence of these disorders was remarkable in India, and, as found among young women, there was modest TSH elevation [1]. Thyroid disorders are part of the most prevalent conditions mankind is aware of, close only to diabetes. The normal functioning of the thyroid glands is vital for maintaining normal levels of thyroid hormone, hence normal reproductive behaviour and physiology. Hypothyroidism leads to several complications in the form of depression, memory loss, cognitive impairments, and a plethora of problems with the nervous and muscular systems [2]. The relationship of glucose levels and thyroid health in women is very essential, especially because the prevalence of thyroid disorders and metabolic syndromes, incidentally, is on the rise globally. Recent studies have defined a complex interrelation between thyroid function and glucose metabolism, indicating that disturbances in thyroid hormones could profoundly affect glucose homeostasis. For instance, the increased prevalence of thyroid nodules accompanies plasma glucose levels that make up the metabolic syndrome, a reflection of the association between thyroid health and glucose metabolism [3]. Moreover, reduced free hormone levels of the thyroid accompany high glucose levels in the blood and contribute to resistance of insulin, showing the metabolic aspects of thyroid disease [4]. Most of all, not much has been written on precisely how glucose levels distribute in relation to thyroid health from the available literature. Studies that have taken place previously rely on conventional statistical models that do not suitably reflect the difficulties of glucose metabolism for this population. This fills the gap because it uses a more subtle approach and fits 25 different distributions to the glucose data sourced from the NFHS-5 survey. This methodological advancement is important since it enabled obtaining more precise comprehension of glucose level distribution, which was best expressed by four-parameter generalized gamma distribution and a beta distribution rather than previously applied models [5], [6].

This paper is novel for not only identifying the best-fitting distribution for glucose levels but also reflecting on its implications on our understanding of thyroid health. However, since the glucose levels are not symmetrically distributed, this might tell us of the abnormalities related to thyroid function and similar metabolic pathways. This works in aggregation with the findings by Pillay, who observed that thyroid hormones indeed regulate glucose homeostasis, and thyroid dysfunction is one of the commonly encountered pathologies among T2DM patients [7]. In addition, thyroid disease has for a long period of time been associated with insulin resistance; however, studies have shown that impaired glucose metabolism will often be compounded by thyroid dysfunction to further develop diabetes risk factors [8].

The relationship of thyroid health with different metabolic disorders has been studied through multiple different studies in recent times. In that respect, Zaho et al. explored gender differences in associations that exist between anxiety symptomatology and thyroid hormones in demonstrating how a healthy thyroid matters for the broader implications of metal and metabolic well-being [9]. Furthermore, Mehran et al. highlighted the idea that thyroid dysfunction combined with metabolic syndrome in diabetic patients; therefore, comprehensive research in this area is required [10].

The research gap identified for this particular study regarding glucose levels and thyroid health among women is mainly based on the fact that the literature currently available does not fully provide a scope that explains the distribution patterns of glucose levels with respect to thyroid function. Several studies have reported on relationships concerning thyroid disorders and metabolic parameters, yet they do not give an account of how glucose levels deviate from symmetric standards and the implications of such deviations toward thyroid health.

2. METHODOLOGY

The results of the National Family Health Survey, conducted in 2019 and 2021, are utilized in the present analysis. Being mainly based on the NFHS-5 survey, the information source is relatively holistic to provide an adequate dataset for analysis that gives insights into different health related factors. In analysing production data, users often seek to characterise its behaviour by fitting it to a probability distribution. Semi-parametric, non-parametric, and parametric distributions are the three main categories that are usually taken into account throughout this procedure.

2.1 Fitting of the Probability Distribution

The fitting of the distributions involved assessing twenty five distributions: Beta, Chi-Squared, Chi-Squared (2P), Exponential, Exponential (2P), Gamma, Gamma (3P), Gen. Extreme Value, Gen.Gamma, Gen.Gamma (4P), Gen.Pareto, Gumbel Max, Gumbel Min, log-Gamma, Log-Logistic, Log-Logistic 3P), Logistic, Lognormal, Lognormal (3P), Normal, Pareto, Pareto 2, Student's t, Weibull, Weibull(3P). A few of the research methods are given below.

2.1.1 Log-Logistic Distribution

The three parameter log logistic probability density function is provided by

$f(x)= \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha-2}\right)$; $\Upsilon \leq x < +\infty$

α- indicates shape (α>0)

β- Indicates scale (β>0)

ϒ- Indicates location

2.1.2 Beta Distribution

The Beta distribution's probability density function is provided by

$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$

Where

x is the variable
α and β are shape parameters, and
B(α,β) is the beta function defined as B(α,β)= $\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}$ dt

2.1.3 Generalized Gamma distribution

k-indicates shape (k>0)
α-indicates shape (α>0)
β-indicates scale (β>0)
ϒ-indicates location
The Four-Parameter Gamma distribution in generalized form, the density function of probability is provided by
f(x)= $\frac{k(x-\Upsilon)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)}$exp(-((x- ϒ)/ β)^k); ϒ≤x≤+∞

2.2 Goodness of fit tests

As recommended by Darling D.A. [11] And Anderson T. W. et al. [12], the importance of goodness-of-fit tests in supporting the decision- making process to identify the best appropriate distribution that accurately reflects the data is elaborated upon in the section that follows.

2.2.1 Kolmogorov-Smirnov test

Definition:
The Kolmogorov-Smirnov statistic (D) is predicted by the highest vertical discrepancy between the cumulative distribution function, which is empirical and theoretical.
D=Max.1≤i≤n [F(X$_i$)-$\frac{i-1}{n}, \frac{i}{n} - F(x_i)$]

2.2.2 Anderson darling test

The Anderson-Darling approach is a genetic test to see whether any experienced cumulative distribution function fits a predicted cumulative distribution function.
Definition: The definition of the Anderson Darling statistics (A^2) is
A^2 = -n $\frac{1}{n}\sum_{i=1}^{n}(2i-1)$[ln F(x$_i$)+ln(1-F(x$_{n-i+1}$))]

2.2.3 Chi-square tests

It is frequently employed in research to evaluate theories on the correlations between various demographic variables.
Definition: The Chi Squared statistic is defined as χ^2 =$\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$
Where O$_i$ is observed frequency and E$_i$ is expected frequency

3. Result & Discussion

3.1 Descriptive Statistics

The descriptive statistics of both sample sizes at 188 and 365 reveal negligible skewness of the glucose level. This shows that the distributions of glucose levels are platykurtic, along with minor magnitudes of excess kurtosis and skewness statistics. Furthermore, the distribution nearly approach normality but shows a slight trend to the side of right skewness. In general, the finding was that the glucose levels have relatively balanced distributions in the analyzed samples.

3.2 Probability Density Function Plots for glucose levels of sizes n=188 & 365

For sample size n=188 For sample size n=365

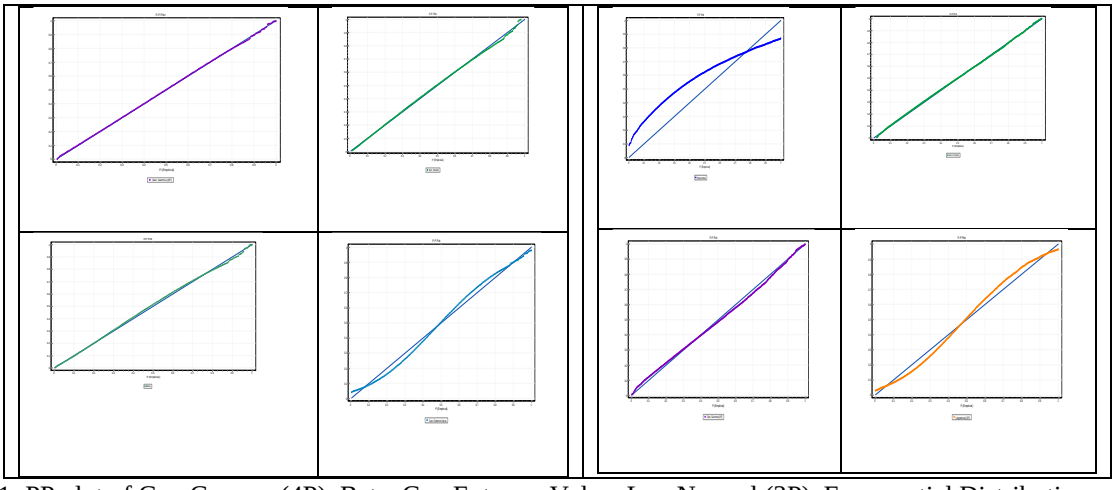

Figure 1: PP plot of Gen Gamma (4P), Beta, Gen Extreme Value, Log Normal (3P), Exponential Distribution, and Pareto Distribution

For sample size n=188 For sample size n=365

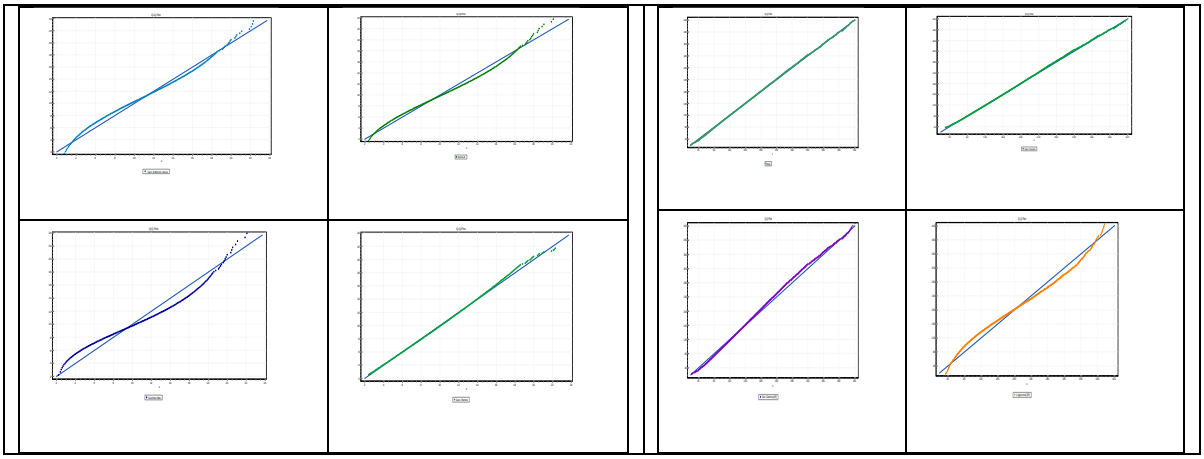

Figure 2: QQ plot of Gen Gamma (4P), Beta, Gen Extreme Value, Log Normal (3P), Exponential Distribution, and Pareto Distribution

From both QQ and PP plots for both sample sizes of n=188 and n=365, it can be seen that if the points are close to the diagonal, a relatively straight line indicates that the data are likely to follow the assumed distributions. The best-fit distributions: Gen gamma 4P, Beta and Gen Pareto are follow accordingly.

3.3 Goodness-of-fit tests

Table 1. Summary goodness –of- fit for various distributions for n=188

S.No	**Distribution**	**Kolmogorov Smirnov**		**Anderson Darling**		**Chi-Squared**	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Beta	0.02627	2	0.16818	2	1.1707	2
2	Chi-Squared	0.13958	20	9.858	19	42.84	20
3	Chi-Squared (2P)	0.06833	11	1.681	9	9.4753	9
4	Exponential	0.2	23	19.922	23	70.867	21
5	Exponential (2P)	0.15393	21	10.762	20	30.546	19
6	Gamma	0.07936	14	3.363	13	16.539	15
7	Gamma (3P)	0.06096	10	1.5611	8	9.7577	10
8	Gen. Extreme Value	0.04374	4	0.97853	3	5.3496	3
9	Gen. Gamma	0.07783	12	2.2974	11	11.572	11
10	Gen. Gamma (4P)	0.0143	**1**	0.06199	**1**	0.25716	**1**
11	Gen. Pareto	0.02796	3	15.675	21	N/A	
12	Gumbel Max	0.08957	15	4.2753	16	22.97	17
13	Gumbel Min	0.10349	17	5.64	17	21.543	16
14	Log-Gamma	0.12604	19	6.9978	18	24.137	18
15	Log-Logistic	0.1049	18	3.9322	15	15.431	13
16	Log-Logistic (3P)	0.05571	7	1.7265	10	9.02	8
17	Logistic	0.07915	13	2.7712	12	16.482	14
18	Lognormal	0.09609	16	3.7241	14	14.325	12
19	Lognormal (3P)	0.05375	6	1.4386	6	8.7136	7
20	Normal	0.05663	8	1.3774	5	7.2886	5
21	Pareto	0.29143	24	35.332	24	87.098	23

22	Pareto 2	0.18445	22	18.067	22	74.049	22
23	Student's t	0.92014	25	696.17	25	9407.0	24
24	Weibull	0.05101	5	0.99343	4	5.4144	4
25	Weibull (3P)	0.05847	9	1.4732	7	7.7605	6

Table 2. Summary goodness –of- fit for various distributions for n=365

S.No	**Distribution**	**Kolmogorov Smirnov**		**Anderson Darling**		**Chi-Squared**	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Beta	0.01383	**1**	0.16007	**1**	0.41498	**1**
2	Chi-Squared	0.38379	24	1588.8	24	868.54	22
3	Chi-Squared (2P)	0.05382	8	3.0297	7	16.695	8
4	Exponential	0.17803	22	31.08	22	94.056	20
5	Exponential (2P)	0.15424	20	23.832	20	61.548	18
6	Gamma	0.07517	12	7.3054	13	35.834	14
7	Gamma (3P)	0.05582	10	2.929	6	15.562	5
8	Gen. Extreme Va lue	0.04176	4	2.0609	3	9.2368	3
9	Gen. Gamma	0.07587	13	4.6431	11	18.183	10
10	Gen. Gamma (4P)	0.02994	3	4.6279	10	N/A	
11	Gen. Pareto	0.0155	2	23.404	19	N/A	
12	Gumbel Max	0.08231	15	7.7854	14	42.75	16
13	Gumbel Min	0.10586	18	13.005	18	56.461	17
14	Log-Gamma	0.11185	19	10.78	17	36.613	15
15	Log-Logistic	0.10283	17	8.1006	16	28.015	12
16	Log-Logistic (3P)	0.05289	7	3.3659	9	16.077	7
17	Logistic	0.07848	14	5.8194	12	34.112	13
18	Lognormal	0.09565	16	7.9716	15	26.457	11
19	Lognormal (3P)	0.05177	5	2.8628	5	15.879	6
20	Normal	0.05582	11	3.0474	8	17.184	9
21	Pareto	0.33149	23	81.582	23	232.29	21
22	Pareto 2	0.15938	21	27.47	21	92.523	19
23	Student's t	0.99875	25	3406.8	25	2.1907E+6	23
24	Weibull	0.05497	9	1.9843	2	6.2664	2
25	Weibull (3P)	0.05215	6	2.6286	4	13.929	4

From the comparative analysis for goodness of fit of several probability distributions, such as generalized Gamma (4P) and Beta, it can be seen that the selection of the distribution has a very significant bearing on how data is interpreted at different sample sizes. In both the cases of N= 188 and N= 365, generalized gamma (4P) emerged out to be the best fit model at both sample sizes, and therefore, it is best suited for any data fitting irrespective of the variations of sample size. Generalized gamma distribution has been found robust as it emerged

in almost all the analyses as the best fit model without an influence of the sample size on the model of glucose levels [11]. It was remarkably improved in the larger sample size case N=365, where it performed well in all three goodness of fit tests applied; namely, Kolmogorov-Smirnov, Anderson-Darling, and Chi-Square tests. This therefore, implies that the bets might be especially adequate for larger datasets, hence giving new clues to researchers as to optimal distribution selection for glucose level analysis. This indicates that enhancement of the Kolmogorov-Smirnov statistic also holds valid in its usage if more samples are considered for research, like from 0.02627 for N= 188 to 0.01383 for N= 365, further supporting its applicability in studies involving larger samples [12],[13]. Moreover, the application of more than one goodness of fit test in this study goes well with the modern approach of statistics, which advises the performance of a comprehensive methodology assessment.

4. CONCLUSION

The study highlights the novelty of glucose levels that are to be analyzed in the case of 188 women between ages 15-49 who are suffering from thyroid disorders. Due to significant skewness in the data, two distributions, namely the Weibull and generalized gamma (4P), were applied to the data. Once their parameters are estimated, goodness of fit tests have been carried out using the Kolmogorov-Smirnov one sample test, Anderson-Darling test, and Chi-Square test. These tests were used to compare and rank the distributions in terms of how well each fitted the data. The outcome was that the generalized gamma (4P) distribution best fits the 188 cases among the women analysed.

Interestingly, this article also illustrated the effect of doubling the sample size, which resulted in the beta distribution being the best fit based on the same goodness-of-fit criteria. This seems to indicate that the distribution of glucose levels in women with thyroid disorders might be sensitive to sample size and that there doesn't exist a consistent trend between thyroid-related glucose irregularities and the ages of 15-49. No studies, according to our knowledge of the existing literature, have ever been found that apply this type of distributional analysis to the situation of thyroid-related glucose levels in women, and thus it is a new and important feature of the methodology and results of the study.

One limitation of this study is that as the sample size increases, the identification of the disease within the group may not follow a unique distribution. This variability suggests that the patterns in glucose levels and their relation to thyroid disorders might become more diverse with larger datasets. For further studies, while this study targeted an age group of 15-49, later studies would be more effective for searching out other age groups to see whether similar distributional patterns may happen or if a different distributions may better reflect the behaviour of glucose levels within these populations. This expanding approach will help to determine whether the changes in distribution are actually driven by age or by increasing sample size in order to furnish deeper insight into the relationship between thyroid disorder and glucose levels across demographic groups.

5. REFERENCES

1. Singh R, Gupta S, Tyagi M, Gupta M, Yadav SK. Anemia and its effect on thyroid profile in pregnant women in tertiary care centre. International Journal of Reproduction, Contraception, Obstetrics and Gynecology. 2023 Jul 1;12 (7):2021-6.
2. Yadav M, Kose VD, Vijay NR, Dwidmuthe KS. Frequency of Thyroid Disorder in Pre-and Post-Menopausal Women and Its Association with Menopausal Symptoms–A Scoping Review. Journal of Datta Meghe Institute of Medical Sciences University. 2023 Apr 1;18(2):327-30.
3. Liu J, Wang C, Tang X, Fu S, Jing G, Ma L, Sun W, Li Y, Wu D, Niu Y, Niu Q. Correlation analysis of metabolic syndrome and its components with thyroid nodules. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy. 2019 Aug 30:1617-23.
4. Gu L, Yang J, Gong Y, Ma Y, Yan S, Huang Y, Wang Y, Peng Y. Lower free thyroid hormone levels are associated with high blood glucose and insulin resistance; these normalize with metabolic improvement of type 2 diabetes. Journal of Diabetes. 2021 Apr;13(4):318-29. [6**]** Kumar V, Bala S. Best fit probability distribution analysis of precipitation and potential evapotranspiration of India's highly dense population state - Bihar. *MAUSAM*. 2022;73(1):139-150.
5. Fadhil M, Razaq S, Al-Kareem A, Al-Kazaz A. Evaluation the correlation between IL-17 level and autoimmune antibodies in hypo and hyper thyroidisms Iraqi patients. Iraqi Journal of Science. 2019 Sep 29;60(9):1967-76.
6. Kaur I, Kaur D, Kaur M, Sibia RP, Massaon HK. Correlation of T3, T4, TSH with fasting plasma glucose and glycosylated haemoglobin in patients of type 2 diabetes mellitus. Int J Clin Biochem Res. 2020;7(3):338-42.
7. Pillay M, Mosili P, Akinnuga A, Sibiya N, Ngubane P, Khathi A. Association between Altered Thyroid Function and Prediabetes in Diet-Induced Prediabetic Male Sprague Dawley Rats. Diabetology. 2023 Sep 12;4(3):406-17.
8. Elmageed Mohammed RM, Hafez Ahmed MH. Thyroid disorders and diabetes mellitus: prevalence and assosciation. Journal of Advances in Medicine and Medical Research. 2021 Dec 15;33(23):220-8.
9. Zhao Y, Liu JC, Yu F, Yang LY, Kang CY, Yan LJ, Liu ST, Zhao N, Wang XH, Zhang XY. Gender differences in the association between anxiety symptoms and thyroid hormones in young patients with first-episode and drug naïve major depressive disorder. Frontiers in Psychiatry. 2023 Aug 29;14:1218551.
10. Mehran L, Amouzegar A, Rahimabad PK, Tohidi M, Tahmasebinejad Z, Azizi F. Thyroid function and metabolic syndrome: a population-based thyroid study. Hormone and Metabolic research. 2017 Mar;49(03):192-200.

11. Liu C. Reweighted and circularised Anderson-Darling tests of goodness-of-fit. Journal of Nonparametric Statistics. 2023 Oct 2;35(4):869-904.
12. Silva JR, Taveira MK, Serrano RO, Mesquita AA, Moreira JG. Probability of rainfall for the city of Cruzeiro do Sul, Acre, Brazil. Revista Ambiente & Água. 2021 Feb 10;16(1):e2593.
13. Aslam M. Introducing Kolmogorov–Smirnov tests under uncertainty: an application to radioactive data. ACS omega. 2019 Dec 31;5(1):914-7.