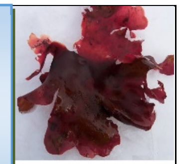
HIBISCUS PUBLISHER

BULLETIN OF ENVIRONMENTAL SCIENCE & SUSTAINABLE MANAGEMENT

Website: http://journal.hibiscuspublisher.com/index.php/BESSM/index



BESSM VOL 4 NO 2 2020

Iridaea cordata

Test for Outlier and Normality of the Residuals for the Pseudo-1st Order Kinetic Modelling of Glyphosate Adsorption onto Palm Oil Fronds

Aisami Abubakar[1]*

[1]Department of Biochemistry, Gombe State University, Gombe, Nigeria.

*Corresponding author:

Aisami Abubakar
Department of Biochemistry,
Gombe State University,
Gombe,
Nigeria.
Email: abubakar.aisami05@gmail.com

HISTORY

Received: 15th Oct 2020
Received in revised form: 16th Dec 2020
Accepted: 17th Dec 2020

KEYWORDS

nonlinear regression
normality test
outlier
Grubb's test
biosorption

ABSTRACT

Generally, in nonlinear regression, the residuals of the curve must be distributed normally, and residuals must be tested for the presence of outliers. This is normally done using normality tests such as the Kolmogorov-Smirnov, Wilks-Shapiro and D'Agostino-Pearson and the Grubb's test, the latter test for the presence of an outlier, which is the focus of this study. Normality tests for residues used in general nonlinear regression revealed that the usage of the pseudo-1st order model in the fitting of glyphosate biosorption to palm oil fronds was adequate due to lack of an outlier. The Grubb's test was applied to the residual results. The critical value of *Z* from statistical table for Grubbs' test for a single outlier using mean and SD was 2.289 (n=10). The Grubbs (Alpha = 0.05) $\boldsymbol{g}$ value was 1.844. As the test statistic is lesser than the critical assigned value, no outlier was deemed present.

INTRODUCTION

In several herbicide formulations used successfully to suppress weeds, glyphosate is a selective, broad-spectrum active ingredient. In the last decade, glyphosate dominates the world's herbicides usage and its uses has increased nearly 15-fold and almost 8.6 billion kilos have been applied worldwide [1]. Substantial and continuous glyphosate-based herbicide application culminated in soils and stream runoff build-up. Glyphosate can contaminate the water supply through field runoff, flooding, leak, floating or spray drift, and this is closely associated with its polar and high solubility characteristics in water (11.6 g L^{-1}, 25 ◦C) [2–4].

Glyphosate pollution incidence has occurred in numerous incidents including river, surface waters, groundwater and marine water, and this is recorded globally [5–8]. In Malaysia, residues of glyphosate and its aminomethylphosphonic acid (AMPA) metabolite have been found in the surface waters at 1.0 to 2.0 mg/L and soil and sediments from 5.0 to 6.0 mg/k near oil palm plantations at Tasik Chini, Pahang Malaysia. [9].
Glyphosate pollution, represented by a growing array of recent toxicological tests [10–14], is a significant toxicity issue for both the atmosphere and human health. Its removal via bioremediation using degrading microorganisms [15–19] and biosorption [20–25] are reported in the literature. Biosorption is characterized as a process independent of physicochemical metabolism that results in the removal of biological materials from the solution. The process was historically used to extract metals and associated materials [26–34], but the application is now being extended to remove other organic target substances such as dyes, steroids, pharmaceuticals, medicines and pesticides [35–37]. A variety of experiments have been performed on the adsorption of glyphosate using different materials such as biopolymer membrane, resin, magnetic nanocomposite and readily accessible activated carbon and biochar [38–41].

Previously, the absorption kinetics data of biosorption isotherm on the biosorption of glyphosate on palm oil fronds activated carbon were analyzed using three models—pseudo-1st, pseudo-2nd and Elovich, and fitted using non-linear regression. The best was pseudo-1st order [24]. Nevertheless, in nonlinear regression the residuals of the curve must be distributed normally, and residuals must be tested for the presence of outliers [at 95 or 99% of confidence). This is normally done using normality tests such as the Kolmogorov-Smirnov, Wilks-Shapiro and D'Agostino-Pearson and the Grubb's test, the latter test for the presence of an outlier, which is the focus of this study.

METHODOLOGY

Residual data were acquired from a previously published work [24].

Residuals

Residuals are very important in assessing the health of a curve from a particular used model. Mathematically, residual for the i^{th} observation in a given data set can be defined as follows (**Eqn. 1**);

$$e_i = y_i - f\left(x_i; \hat{\beta}\right) \qquad \textbf{(Eqn. 1)}$$

where y_i denotes the i^{th} response from a given data set while x_i is the vector of explanatory variables to each set at the i^{th} observation corresponding values in the data set.

Grubbs' Statistic

In an average value, a single data point with deformation can lead to gross error in the fitting of a nonlinear curve. Therefore, searching for an outlier is an integral aspect of curve fitting. The Grubbs test is used to evaluate the outlier in the univariate environment and the data is normally distributed [42]. The test can be applied to the maximal or minimal observed data from a Student's t distribution (**Eqn. 2**) and to test for both data simultaneously (**Eqn. 3**).

$$G_{\min} = \frac{\bar{X} - \min(X)}{s}$$

$$G_{\max} = \frac{\max(X) - \bar{X}}{s} \qquad \textbf{(Eqn. 2)}$$

$$p_G = 2n.p_t\left(G\frac{\sqrt{n(n-2)}}{n-1}, n-2, 1\right)$$

$$G_{all} = \frac{\max(\bar{X} - \min(X), \max(X) - \bar{X}}{s}$$

$$p_G = n.p_t\left(G\frac{\sqrt{n(n-2)}}{n-1}, n-2, 2\right) \qquad \textbf{(Eqn. 3)}$$

Normality test

Residuals from the pseudo-1st order model were subjected to three normality tests- Kolmogorov-Smirnov [43,44], Wilks-Shapiro [45] and the D'Agostino-Pearson omnibus K2 test [46]. Using graphical and numerical methods are two ways to search for normality. The simplest and easiest way to assess the normality of data is via graphical methods such as the normal quantile–quantile (Q-Q) plots, histograms or box plots [47]. The normality tests were carried out using the GraphPad Prism® software (Version 6.0, GraphPad Software, Inc., USA).

RESULTS AND DISCUSSION

Statistics often used in nonlinear regressions rely on the use of residual data, which is the difference between the expected and the actual values. Statistical analyses should be done to evaluate the adequacy of residues in randomness, do not include outliers, obey normality, and do not demonstrate autocorrelation. Usually, the greater the discrepancy between the expected and the observable values, the less well off the model. [48]. The Grubbs' test deals with one aspect at a time. Outliers are eliminated and test replicated before test passes without revealing any outliers. As a general rule, sample sizes of 6 or less results in biased data sets. Many variations of the same model alter the probability of identification.

Table 1. Residual data from the pseudo-1st order model.

	Residuals
	0.00052
	0.00120
	-0.00037
	-0.00337
	-0.00216
	-0.00104
	-0.00112
	-0.00154
	-0.00098
	0.00063
Mean	-0.000823
Std. Deviation	0.0013809

The Grubb's test was applied to the residual results (**Table 1**). Grubbs test statistic defines the highest absolute variance from the survey mean in the sample standard deviation units. The critical value of *Z* from statistical table for Grubbs' test for a single outlier using mean and SD was 2.289 (n=10). The Grubbs (Alpha = 0.05) ***g*** value was 1.844. Individual Z value indicates that the residual with a value of -0.00337 (row 4) was far from the rest but is deemed not a significant outlier (p > 0.05) (**Table 2**). As the test statistic is lesser than the critical assigned value, no outlier was deemed present (**Table 1**).

Table 2. Calculated Z value for residual data.

Row	Value	Z	Significant Outlier?
1	0.00052	0.97257	
2	0.0012	1.46501	
3	-0.00037	0.328052	
4	-0.00337	1.844479	Furthest from the rest, but not a significant outlier (P > 0.05).
5	-0.00216	0.968225	
6	-0.00104	0.157146	
7	-0.00112	0.215081	
8	-0.00154	0.519235	
9	-0.00098	0.113696	
10	0.00063	1.052229	

The fitness of a mathematical model is generally calculated exactly via the use of residual measures. Residuals are the difference between the sum expected and observed using a specific mathematical model. The basic notion is that the greater the gap between the expected and the observable values, the weaker the model. Residual plot (observed-predicted) was tested and the study revealed that the data was randomly distributed for all experiments (**Fig. 1**). Evidently, a possible outlier is an unusual data point that the researcher marks as impossible in terms of a variety of unique criteria. More precisely, the outlier in the study may be an exceptional attribute that is definitely too unusual. By way of illustration, most are known to be outliers only if they are statistically too high for the distribution to the limit in the sample model [49].

An easy way to mark prospective outliers in tests is to add boxplot, although a bit more complex methodology is often used, like the Chauvenet criteria in engineering and the 3-sigma criterion, along with the Z-score in chemometrics. Given the fact that these methods are very plain and quick, there is a far more effective way of utilizing the statistical test for outlier detection. Relevant assessments differ from the Dixon Q-test or the Grubbs ESD-test with one outlier. The key restriction of the Grubbs test is generally that the thinking quantity of the outliers, *k*, must be stated specifically. If *k* is not clarified appropriately, the results of the experiments can be skewed. In the event that the outliers are multiple, or the exact quantity of outliers is not known, the Rosner Generalized Severe Studentized Deviate or ESD-test is prescribed [50]. This is since the existence of more than one outlier will distort the Grubbs test results and, as this occurs, the Ferguson sample skew test is more resilient against the disguising impact than the Grubb test [51].

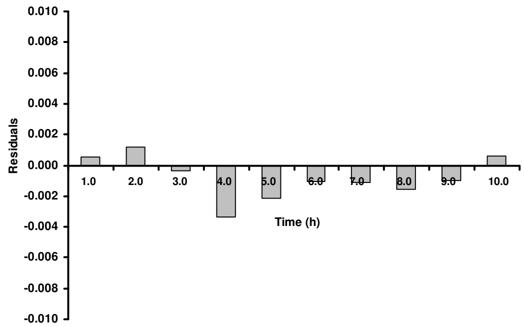


Fig. **1**. Residual plot for the pseudo-1st order model model.

The number of bins and samples examined dictated the form of the distribution. In the Wilks-Shapiro test, the W2 statistic is calculated on the basis of the predicted values of the order statistics between both the identically distributed random variables as well as their independent covariance plus the regular normal distribution. If the importance of the test statistics-W2 is big, the agreement is refused. [45]. The Kolmogorov-Smirnov statistic is a non-parametric numerical test which calculates the cumulative residual frequency. It measures the relation between both the model and the values observed. It may also be used as a comparison of two series of observations. The p value is determined for the discrepancy of two combined distributions and the size of the sample. [52–54]. In general, the Central Limit Theorem (CLT), that states that on any continuous variables (even for discrete variables such as Binomial or Poison distributions), when n tends to infinite (in practice n>30) the frequency of distribution of probabilities tends to fit Gaussian distribution [55,56].

The distribution skewness and kurtosis were measured as a measure for quantifying the disparity between the distributions of the samples to the usual distribution in the D'Agostino-Pearson normality test method. The p-value of the sum of these contradictions or discrepancies is then determined. D'Agostino developed several normality tests but the most often utilized is the omnibus K2 test [46].

Graphical diagnostic of residuals normality

The normal probability Q-Q plot of residuals for the pseudo-1st order model was almost in a straight line and appears to show no underlying pattern (**Fig. 2**). The resulting histogram overlaid with the resulting normal distribution curve (**Fig. 3**) indicates the residuals were truly random and the model used was appropriately fitted.

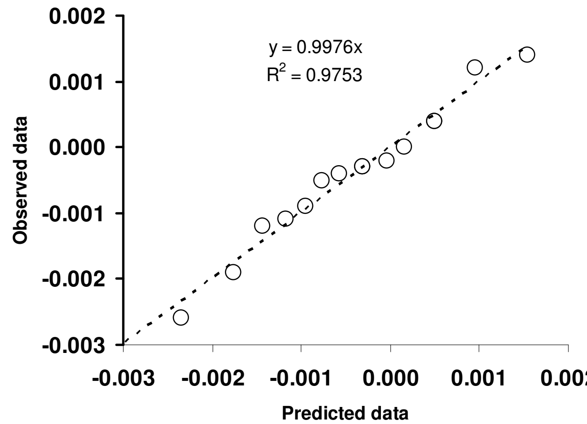


Fig 2. Normal Q-Q plot for the observed sample against theoretical quantiles.

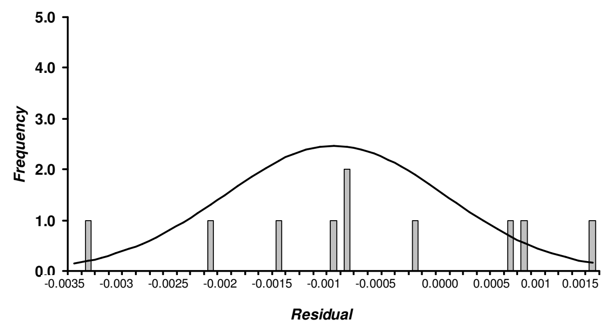


Fig. **3**. Histogram of residual for the pseudo-1st order model overlaid with a normal distribution (mean -0.000823 and standard deviation 0.001380878).

Table 3. Numerical normality test for the residual from the pseudo-1st order model after removal of an outlier.

Normality test	Analysis
KS normality test	
KS distance	0.1453
P value	> 0.10
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
D'Agostino & Pearson omnibus normality test	
K2	0.1963
P value	0.9065
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
Shapiro-Wilk normality test	
W	0.9680
P value	0.8715
Passed normality test (alpha=0.05)?	Yes
P value summary	ns

CONCLUSION

In conclusion, the normality checks for the residues used in this work showed that the usage of the pseudo-1st order model in the fitting of glyphosate biosorption to palm oil fronds was satisfactory due to the lack of an outlier. It is well established that several articles have not further expanded on the application of the model employed in the mathematical diagnosis of residues. This may result in a data violation in Gaussian or regular distribution. This assertion is an essential condition for many of the parametric predictive estimation techniques used in nonlinear regression. Methods such as the RMSE, the Pearson correlation coefficient either standard or modified, the F-test and the t-test are based on the residuals conforming to the normal distribution. These assumptions could avoid errors of the Type I and II errors. In addition, in the case that diagnostic tests indicate that contaminants have broken any of the assumptions that multiple nonparametric therapies should be used or that they should be changed to a new form, the condition may be remedied in operation.

REFERENCES

1. Benbrook CM. Trends in glyphosate herbicide use in the United States and globally. Environ Sci Eur. 2016 Dec;28(1).
2. Hu YS, Zhao YQ, Sorohan B. Removal of glyphosate from aqueous environment by adsorption using water industrial residual. Desalination. 2011 Apr;271(1–3):150–6.
3. Ng C. Agriculture and Water Pollution Risks. UTAR Agriculture Science Journal. 2017;3:34–44.
4. Bonansea R, Filippi I, Wunderlin D, Marino D, Amé M. The Fate of Glyphosate and AMPA in a Freshwater Endorheic Basin: An Ecotoxicological Risk Assessment. Toxics. 2017;6(1):3.
5. Mercurio P, Flores F, Mueller JF, Carter S, Negri AP. Glyphosate persistence in seawater. Mar Pollut Bull. 2014;85(2):385–90.
6. Grandcoin A, Piel S, Baurès E. AminoMethylPhosphonic acid (AMPA) in natural waters: Its sources, behavior and environmental fate. Water Res. 2017;117:187–97.
7. Castro Berman M, Marino DJG, Quiroga MV, Zagarese H. Occurrence and levels of glyphosate and AMPA in shallow lakes from the Pampean and Patagonian regions of Argentina. Chemosphere. 2018;200:513–22.
8. Lupi L, Bedmar F, Puricelli M, Marino D, Aparicio VC, Wunderlin D, et al. Glyphosate runoff and its occurrence in rainwater and subsurface soil in the nearby area of agricultural fields in Argentina. Chemosphere. 2019;225:906–14.
9. Mardiana-Jansar K, Ismail BS. Residue determination and levels of glyphosate in surface waters, sediments and soils associated with oil palm plantation in Tasik Chini, Pahang, Malaysia. In Selangor, Malaysia; 2014.p. 795–802. Available from: http://aip.scitation.org/doi/abs/10.1063/1.4895304
10. Albañil Sánchez JA, da Costa Klosterhoff M, Romano LA, De Martinez Gaspar Martins C. Histological evaluation of vital organs of the livebearer *Jenynsia multidentata* (Jenyns, 1842) exposed to glyphosate: A comparative analysis of Roundup® formulations. Chemosphere. 2019;217:914–24.
11. de Brito Rodrigues L, Gonçalves Costa G, Lundgren Thá E, da Silva LR, de Oliveira R, Morais Leme D, et al. Impact of the glyphosate-based commercial herbicide, its components and its metabolite AMPA on non-target aquatic organisms. Mutat Res - Genet Toxicol Environ Mutagen. 2019;842:94–101.
12. Li Y, Ding W, Li X. Acute exposure of glyphosate-based herbicide induced damages on common carp organs via heat shock proteins-related immune response and oxidative stress. Toxin Rev. 2019;
13. Maskey E, Crotty H, Wooten T, Khan IA. Disruption of oocyte maturation by selected environmental chemicals in zebrafish. Toxicol In Vitro. 2019;54:123–9.
14. Smith CM, Vera MKM, Bhandari RK. Developmental and epigenetic effects of Roundup and glyphosate exposure on Japanese medaka (*Oryzias latipes*). Aquat Toxicol. 2019;210:215–26.
15. Sabullah MK, Rahman MF, Ahmad SA, Sulaiman MR, Shukor MS, Shamaan NA, et al. Isolation and characterization of a molybdenum-reducing and glyphosate-degrading *Klebsiella oxytoca* strain Saw-5 in soils from Sarawak. Agrivita. 2016;38(1):1–13.
16. Manogaran M, Shukor MY, Yasid NA, Johari WLW, Ahmad SA. Isolation and characterisation of glyphosate-degrading bacteria isolated from local soils in Malaysia. Rendiconti Lincei. 2017;28(3):471–9.
17. Manogaran M, Yasid NA, Ahmad SA. Mathematical modelling of glyphosate degradation rate by *Bacillus subtilis*. J Biochem Microbiol Biotechnol. 2017 Jul 31;5(1):21–5.
18. Manogaran M, Ahmad SA, Yasid NA, Yakasai HM, Shukor MY. Characterisation of the simultaneous molybdenum reduction and glyphosate degradation by *Burkholderia vietnamiensis* AQ5-12 and *Burkholderia* sp. AQ5-13. 3 Biotech. 2018 Feb 7;8(2):117.
19. Xu B, Sun Q-J, Lan JC-W, Chen W-M, Hsueh C-C, Chen B-Y. Exploring the glyphosate-degrading characteristics of a newly isolated, highly adapted indigenous bacterial strain, Providencia rettgeri GDB 1. J Biosci Bioeng. 2019;128(1):80–7.
20. Cheah U-B, Kirkwood RC, Lum K-Y. Adsorption, desorption and mobility of four commonly used pesticides in Malaysian agricultural soils. Pestic Sci. 1997;50(1):53–63.
21. Salman J, Abid F, Muhammed AA. Batch study for pesticide glyphosate adsorption onto palm oil fronds activated carbon. Asian J Chem. 2012 Dec 1;24:5646–8.
22. Zhu X, Li B, Yang J, Li Y, Zhao W, Shi J, et al. Effective adsorption and enhanced removal of organophosphorus pesticides from aqueous solution by Zr-Based MOFs of UiO-67. ACS Appl Mater Interfaces. 2015;7(1):223–31.
23. Herath I, Kumarathilaka P, Al-Wabel MI, Abduljabbar A, Ahmad M, Usman ARA, et al. Mechanistic modeling of glyphosate interaction with rice husk derived engineered biochar. Microporous Mesoporous Mater. 2016 May 1;225:280–8.
24. Marbawi H, Khudri MAMRS, Othman AR, Halmi MIE, Gansau JA, Yasid NA. Kinetic analysis of the adsorption of glyphosate onto palm oil fronds activated carbon. Bioremediation Sci Technol Res. 2019;7(1):29–33.
25. Marbawi H, Khudri MAMRS, Othman AR, Halmi MIE, Gansau JA, Yasid NA, et al. Isothermal modelling of the adsorption of glyphosate onto palm oil fronds activated carbon. J Environ Microbiol Toxicol. 2019;7(1):21–6.
26. Torma AE. Use of biotechnology in mining and metallurgy. Biotechnol Adv. 1988;6(1):1–8.
27. Lokesha S, Somashekar RK. Biosorption potency of heavy metals by some fungi. Curr Sci. 1989;58(10):571–3.
28. Holan ZR, Volesky B, Prasetyo I. Biosorption of cadmium by biomass of marine algae. Biotechnol Bioeng. 1993;41(8):819–25.
29. Weppen P, Hornburg A. Calorimetric studies on interactions of divalent cations and microorganisms or microbial envelopes. Thermochim Acta. 1995;269–270(C):393–404.
30. Suyama K, Fukazawa Y, Suzumura H. Biosorption of precious metal ions by chicken feather. Appl Biochem Biotechnol - Part Enzyme Eng Biotechnol. 1996;57–58:67–74.
31. Yetis Ü, Özcengiz G, Dilek FB, Ergen N, Erbay A, Dölek A. Heavy metal biosorption by white-rot fungi. Water Sci Technol. 1998;38(4-5–5 pt 4):323–30.
32. Lyalikova-Medvedeva NN, Khijniak TV. Biosorption of long-lived radionuclides. Process Metall. 1999;9(C):327–34.
33. Lin Z-Y, Fu J-K, Wu J-M, Liu Y-Y, Cheng H. Preliminary Study on the Mechanism of Non-enzymatic Bioreduction of Precious Metal Ions. Acta Phys - Chim Sin. 2001;17(5):479–80.
34. Mallick N. Biotechnological potential of Chlorella vulgaris for accumulation of Cu and Ni from single and binary metal solutions. World J Microbiol Biotechnol. 2003;19(7):695–701.
35. Gadd GM. Biosorption: critical review of scientific rationale, environmental importance and significance for pollution treatment. J Chem Technol Biotechnol. 2009;84(1):13–28.
36. Ahmad T, Rafatullah M, Ghazali A, Sulaiman O, Hashim R, Ahmad A. Removal of Pesticides from Water and Wastewater by Different Adsorbents: A Review. J Environ Sci Health Part C. 2010;28(4):231–71.
37. Fomina M, Gadd GM. Biosorption: current perspectives on concept, definition and application. Bioresour Technol. 2014;160:3–14.

38. Carneiro RTA, Taketa TB, Gomes Neto RJ, Oliveira JL, Campos EVR, de Moraes MA, et al. Removal of glyphosate herbicide from water using biopolymer membranes. J Environ Manage. 2015;151:353–60.
39. Chen F, Zhou C, Li G, Peng F. Thermodynamics and kinetics of glyphosate adsorption on resin D301. Arab J Chem. 2016;9:S1665–9.
40. Fiorilli S, Rivoira L, Calì G, Appendini M, Bruzzoniti MC, Coïsson M, et al. Iron oxide inside SBA-15 modified with amino groups as reusable adsorbent for highly efficient removal of glyphosate from water. Appl Surf Sci. 2017;411:457–65.
41. Dissanayake Herath GA, Poh LS, Ng WJ. Statistical optimization of glyphosate adsorption by biochar and activated carbon with response surface methodology. Chemosphere. 2019;227:533–40.
42. Grubbs F. Procedures for detecting outlying observations in samples. Technometrics. 1969;11(1):1–21.
43. Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. G Dell' Ist Ital Degli Attuari. 1933;4:83–91.
44. Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat. 1948;19:279–81.
45. Royston P. Wilks-Shapiro algorithm. Appl Stat. 1995;44(4):R94.
46. D'Agostino RB. Tests for Normal Distribution. In: D'Agostino RB, Stephens MA, editors. Goodness-Of-Fit Techniques. Marcel Dekker; 1986.
47. Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J Off Publ Fed Am Soc Exp Biol. 1987;1(5):365–74.
48. López S, Prieto M, Dijkstra J, Dhanoa MS, France J. Statistical evaluation of mathematical models for microbial growth. Int J Food Microbiol. 2004;96(3):289–300.
49. Barnett V, Lewis T. Outliers in Statistical Data. 3rd ed. Chichester ; New York: Wiley; 1994. 604 p.
50. Rosner B. Fundamentals of biostatistics. 7th ed. Boston: Brooks/Cole; 2011.
51. Bendre SM, Kale BK. Masking effect on tests for outliers in normal samples. Biometrika. 1987 Dec 1;74(4):891–6.
52. Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. G Dell' Ist Ital Degli Attuari. 1933;4:83–91.
53. Kolmogorov A. Confidence limits for an unknown distribution function. Ann Math Stat. 1941;12(4):461–3.
54. Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat. 1948;19:279–81.
55. Fischer H. A History of the Central Limit Theorem: From Classical to Modern Probability Theory. New York: Springer-Verlag; 2011. (Sources and Studies in the History of Mathematics and Physical Sciences). Available from: https://www.springer.com/gp/book/9780387878560
56. Pandoo P. Which normality test is more appropriate on residuals with sample size 1000?. ResearchGate. 2014 [cited 2021 Jan 4]. Available from: https://www.researchgate.net/post/Which-normality-test-is-more-appropriate-on-residuals-with-sample-size-1000. Accessed date 4th January 2021.