University of Tehran Press

DESERT

Home page: https://jdesert.ut.ac.ir/

Online ISSN: 2345-475X

Comparison of normality test methods for some soil properties in the arid land of South Khorasan

Moslem Rostampour[1*], Farhad Azarmi-Atajan[2]

[1] Department of Rangeland and Watershed Management, Research Group of Drought and Climate Change, Faculty of Natural Resources and Environment, University of Birjand, Birjand, Iran. E-mail: rostampour@birjand.ac.ir

[2] Department of Soil Science, Research Group of Drought and Climate Change, Faculty of Agriculture, University of Birjand, Birjand, Iran.

Article Info.	ABSTRACT

Article type:
Research Article

Article history:
Received: 07 Sep. 2023
Received in revised from: 18 Nov. 2023
Accepted: 25 Nov. 2023
Published online: 27 Dec. 2023

Keywords:
Normal distribution,
Parametric tests,
Soil and plant relationships,
Soil properties.

Statistical assumptions are the basis of many univariate and multivariate statistical tests. Normality is the most basic assumption of multivariate analysis in plant ecology. If the normality assumption is violated, some specific statistical tests are not valid. Therefore, the present study compares the methods of normality assessment of some soil properties in the arid land of South Khorasan. It also examines the effect of increasing the number of soil samples from 25 to 50 on the normality results. Histogram, box plot, Q-Q plot, CV, skewness, and univariate and multivariate normality tests were used. The results showed that EC, K, Ca, Mg, Na, Cl, HCO3, and SAR data had a very high variation (CV 75–100%) and saturation moisture and pH had a low variation (CV <15%). Based on the results of most statistical tests and the skewness coefficient, saturation moisture, pH, N, P, CaCO3, sand, and silt were normal. EC, K, Ca, Mg, Na, Cl, HCO3 and SAR had the right skewed distribution. The results showed multivariate normality was violated, and the use of these data was not suitable for multivariate analysis. The results of the goodness-of-fit test showed that P, sand, and silt follow a normal distribution. Other soil properties do not follow any of the studied probability distributions ($p \geq 0.05$). Therefore, the use of nonparametric is recommended for the physical and chemical properties of the soil in the area. Although in general, the increase in the number of samples has a positive effect on the actual distribution of the community, but due to the high spatial variability of some soil properties such as salinity, the status of nutrients, particle size, etc., the CV and the range of variations in most of soil properties are wide.

Cite this article: Rostampour, M., Azarmi Atajan, F. (2023). Comparison of normality test methods for some soil properties in the arid land of South Khorasan. DESERT, 28 (2), DOI: 10.22059/jdesert.2023.95765

1. Introduction

Soil, as a part of the biosphere, plays an important role in food production and environmental stability. Soil is a dynamic, complex and heterogeneous ecological system (Hartemink and Krasilnikov, 2020; Nunes et al., 2020). Parent materials and factors of soil formation influence the environment, the physical, chemical and biological properties of the soil, as well as the functioning of the soil (Al-Kaisi et al., 2017). Any changes in soil location and time, and thus soil heterogeneity due to changes in the availability of soil resources such as water and nutrients, will affect soil-plant relationships and vegetation-biodiversity interactions (Jafari and Rostampour, 2019; et al., 2020; Li et al., 2021). According to Liu et al. (2021) soil heterogeneity refers to the uneven distribution pattern of soil resources, including nutrients, water, and organisms. It is widely recognized that there is spatial variability in soils both horizontally and vertically due to environmental changes especially soil formation factors and processes. This spatial variability inherent in soils becomes the dominant source of uncertainty in the assessment of soil properties based on in field studies (Han et al., 2020). Sampling error and collection of an unrepresentative sample can be reasonably high in random composite sampling procedures, On the other hand and the mean of soil physical and chemical characteristics is also greatly affected by outlier data.

Many of the statistical methods used for hypothesis testing, which are essential for explaining soil-plant relationships (regression and correlation analyses), are based on probabilities (Kent, 2012). Moreover, the high cost of soil property analysis makes the sample size insufficient, which limits the use of many statistical tests. Statistical tests used to analyze data required assumptions for the results to be valid, typical assumptions for statistical tests, including normality, homogeneity of variances and independence (Tsagris and Pandis, 2021).

Statistical assumptions underlie many univariate and multivariate tests. One of the most important conditions for multivariate data analysis is the establishment of assumptions of normality, linearity and the same data distribution. When one or more of these assumptions are ignored, the statistical results show bias or distortion (Meyers et al., 2012). Parametric tests require random data, normal distribution of data, appropriate sample size and homogeneity of variances (Kim and Park, 2019). However, empirical evidence, including real data from some scientific journal reviews, shows that these assumptions are not always fulfilled (Blanca et al., 2017).

The normal distribution of the data depends on the type of test used. In exploratory factor analysis, where the researcher has no idea about the structure of the data or the number of dimensions of his variables, the problem of normal distribution is a partial problem, as problems such as the presence of outliers, data standardization and transformation are the main problems. In confirmatory factor analysis, where the structure of factors and variables is predefined and specified, the problem of statistical inference is raised. In such a case, if the data is supposed to represent society, it is necessary to know the statistical distribution of the data and, especially, the assumption of normality of the data (Greenacre & Primicerio, 2013).

Of course, all confirmatory factor analysis methods do not require the assumption of normality, for example, among the model estimation methods, the maximum likelihood (ML) method is used when the observed variables are normal and linear. The Weighted Least Squares (WLS) method does not depend on the normality assumption and requires a very large sample size (more than 1000 cases) (Li, 2021).

The assumption of data normality is the most important assumption for statistical testing of general linear models (Knief and Forestmeier, 2021), which is widely used in the distribution of experimental designs in agricultural sciences and natural resources. If this is not the case, some specified statistical tests are invalid and cannot be used (Hair et al., 2010). There has been little research on the normalization of soil survey data. Some studies show that most soil properties follow a skewed distribution and are log-normally distributed (Awal et al., 2019). Mcgrath and Zhang, (2003) and Fukomasu et al. (2022) acknowledged that soil organic carbon is not normal. Frophifar et al., (2012) investigated the abundance of soil physical and chemical properties in Dash-e-Tabriz, northwestern Iran. Their results showed that clay, silt, CaCO3, bulk density, pH and CEC are normally distributed and EC, total nitrogen, sand, organic carbon, P (available) and SAR are log-normally distributed. Awal et al., (2019) stated in their study that characteristics such as porosity, bulk density, soil surface temperature and soil organic seem to follow neither a normal nor a log-normal distribution.

Although soil science and ecology experts who intend to use parametric tests based on normal distribution, need to check the normality of the data, some methods of classification and ordination plant communities do not require a normality test, and multivariate statistical analyses can also be performed on non-normal data (Legendre & Legendre, 2012). Among univariate and multivariate analyses, analyses such as ANOVA, MANOVA, simple regression and multiple regression, discriminant analysis and canonical correlation analysis (CCA) require normal data (Oppong and Agbedra, 2016). For continuous data such as soil physical and chemical properties, the normality test is very important because the normality of the data is the basis for choosing parametric and non-parametric statistical tests (Mishra et al., 2019). Therefore, this study compares the methods used to check the normality of some soil physical and chemical properties in dry and desert rangelands of South Khorasan, Iran. It also examines the effect of increasing the number of soil samples from 25 to 50 on the normality of the data.

2. Materials and Methods

This study was conducted in the rangelands of Zirkouh, South Khorasan (Figure 1), during a *Salsola richteri-Tamarix stricta* (halophyte) to *Ammodendron persicum-Salsola richteri* (psammophyte) salinity gradient. The study area with 100 hectares with is located at a latitude of 33° 37´ N and longitude of 60° 03´ E with warm semi-arid climate. the mean elevation about 960 m above the sea level. The average annual rainfall is 147.75 mm and the average annual temperature is 20.53 ° C (Rostampour, 2022). From the beginning to the end of the gradient, 25 samples and 50 soil samples were randomly taken from the depth of 0-30 cm from the under canopy of the studied species, and after sample preparation, some physical and chemical properties of the soil, including soil texture components, were determined by Bouyoucos hydrometer method, soil saturation by weight method (Jafari and Rostampour, 2019), pH and electrical conductivity (EC) with pH meter and EC meter (Jafari Haghighi, 2003), nitrogen, phosphorus, potassium, calcium, magnesium, sodium, chloride and HCO3 were analyzed by graphite furnace atomic absorption spectroscopy (ContrAA 700) (with a stock concentration of 1 mg/mL and compute calibration adjusted R2 =0.99), sodium absorption ratio (SAR) determined by formula (Jafari and Rostampour, 2019) and calcium carbonate using the volumetric calcimetric method (Jafari Haghighi, 2003) in the soil and environmental pollution laboratories of University of Birjand.

2.1. Statistical Analysis

After recording the soil data, descriptive and inferential methods (statistical tests) were used to measure the normal data. In the present study, Pearson's coefficient of skewness (second method) was calculated along with the standard error and the test statistics (by dividing the skewness values by their standard errors). If the test statistic is greater than 2, the data distribution is asymmetric and does not follow the normal distribution (Table 1). One of the quick methods to check the normality of the data is the coefficient of variation (CV), and if the standard deviation is less than half of the mean (i.e. %CV <50) data are considered normal (Mishra et al., 2019).

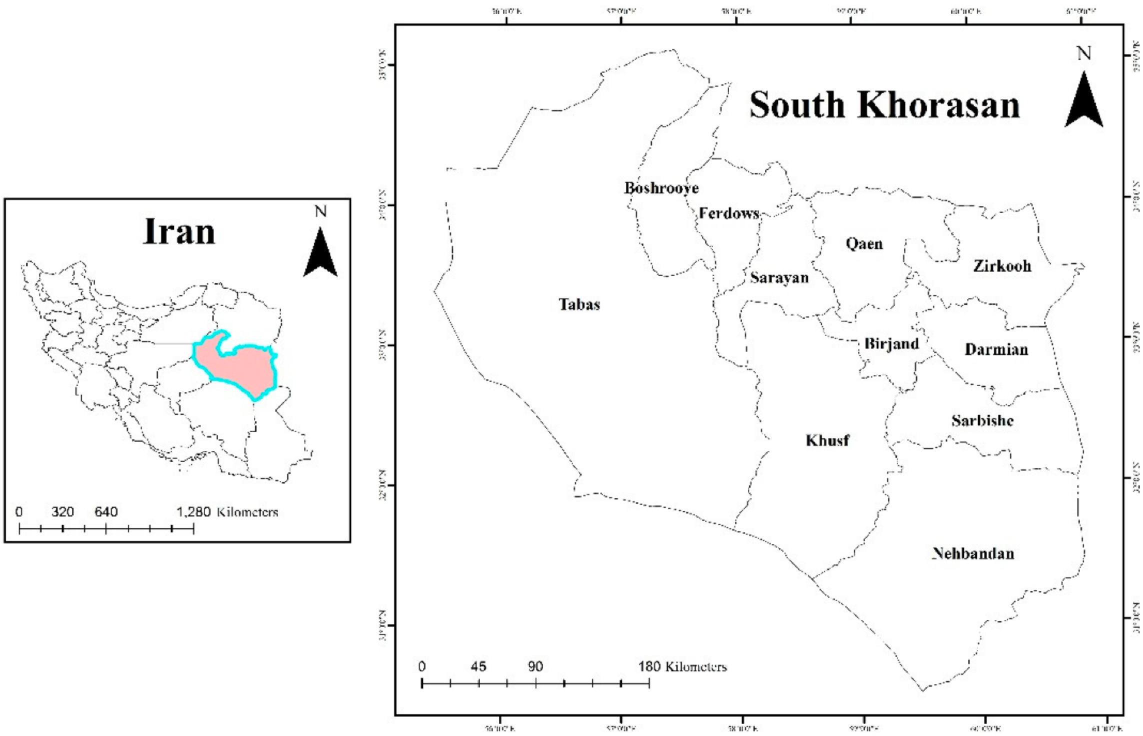
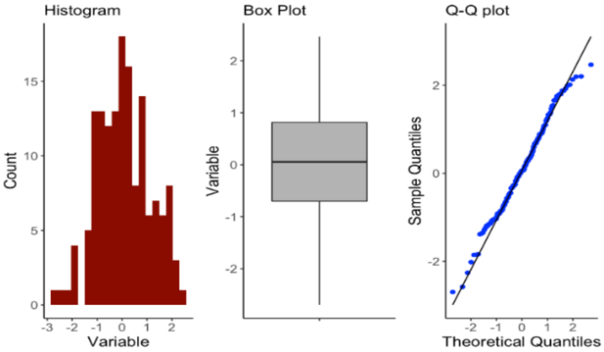


Fig. 1. The region geographical location, South Khorasan Province, Iran

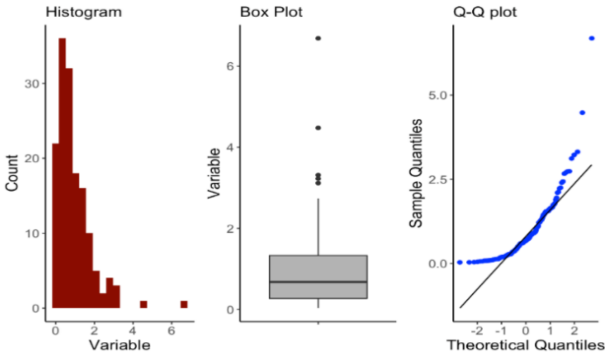
Table 1. Classification of coefficient of skewness

Interval	Skewness
-2.25<	Very extreme negative
-2.25, -1.76	Extreme negative
-1.75, -1.26	High negative
-1.25, -0.76	Moderate negative
-0.76, -0.26	Slight negative
-0.25, 0.25	Near normal
0.26, 0.75	Slight positive
0.76, 1.25	Moderate positive
1.26, 1.75	High positive
1.76, 2.25	Extreme positive
>2.25	Very extreme positive

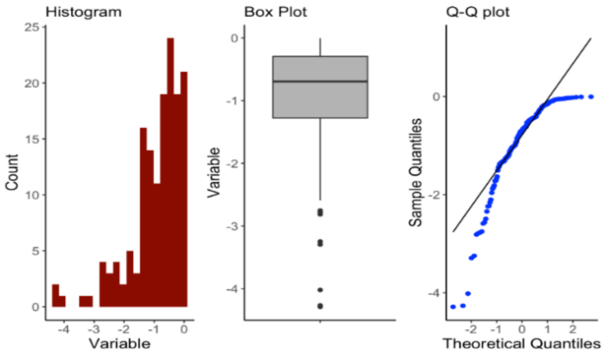
A histogram is a data visualization that shows the distribution of a variable. If the bars follow a roughly symmetrical bell or hill shape, the distribution is approximately normal (Cooksey, 2020). Another visual way to check the normal distribution is to use a Q-Q plot (quantile-quantile plot), which plots the correlation between a given sample and the normal distribution. A 45 degree reference line is also drawn. Q-Q plots are used to visually check the normality of the data. Since all points fall roughly along this reference line, we can assume normality (Weine et al., 2023). A box plot is a standard method for showing the distribution of data. If the box plot is perfectly symmetrical (the median is only in the middle of the data), then the data may be normally distributed (Mishra et al., 2019) (Figure 2).



A: Normal distribution



B: Right skewed distribution



C: Left skewed distribution

Fig. 2. Histograms, boxplot and Q-Q plot of a symmetric (A), right-skewed (B), and left-skewed (C) data set

2.2. Univariate normality tests

To test the normality of the data, the null hypothesis was tested with a 5% error level based on the assumption that the data were normally distributed. Therefore, if the larger test statistic is 0.05, there is no reason to reject the null hypothesis on the basis that the data are normal. In other words, the dissemination of information is normal. The null and alternative hypotheses for these tests are, respectively:

H_0: The data follow a normal distribution

H_1: The data do not follow a normal distribution.

Eight tests were used in this study: Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-Von Misses, Anderson-Darling, Pearson Chi-Square, Shapiro-France, Lilliefors (modified Kolmogorov-Smirnov) and Jarque-Bera. In addition to classical tests, robust tests such as Robust Jarque-Bera and Robust Sj Test were used in this study. The Robust Jarque–Bera (RJB) is the robust version of the Jarque–Bera (JB) test of normality, which is based on the ratio of the classical standard deviation S to the robust standard deviation J (Average Absolute Deviation from the Median, MAAD) of the sample data (Gel et al., 2007).

2.3. Multivariate normality tests

In the present study, the tests of Shapiro-Wilk, Mardia, Henze-Zirkler, Royston Dornik-Haansen,. E-Statistics were used. The Mardia test can also was calculated corrected version of skewness coefficient for small sample size ($n< 20$). For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample size less than 20 then p.value. Small should be used as significance value of skewness instead of p.value.skew (Korkmaz, 2022).

2.4. Fitting data into probability distributions

Four continuous statistical distributions were selected from the continuous probability distributions for the soil physical and chemical property data: normal, lognormal, exponential, and gamma, and a chi-square goodness-of-fit test was fitted to the data. This test contains two assumptions:

H_0: The data follow a specified distribution.

H_1: The data do not follow the specified distribution.

If the p-value of the chi-square test is less than 0.05, the null hypothesis will be rejected, and it can be concluded that the data does not follow a specified distribution. The best model was selected for soil properties that follow several statistical distributions according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). In this case, the model with the smaller AIC and BIC values for each selected fitting distribution is known as the best-suited model.

Based on the coefficient of variation (CV), confidence level (0.05) and acceptable error rate (d), the number of soil samples required were determined (Table 2).

Table 2. Sample sizes required for estimating the true mean μ using a prespecified relative error and the coefficient of variation (adapted from Carter and Gregorich, 2008)

Confidence level	Relative error, d	Coefficient of Variation (CV), %				
		10	20	40	50	100
0.90	0.10	2	12	45	70	271
	0.25			9	12	45
	0.50				2	13
0.95	0.10	4	17	63	97	385
	0.25			12	17	62
	0.50				4	16

All analysis were performed with nortest, norttest, mvnormtest, MVN, mvnormaltest, fitdistrplus and lawstat packages in the R environment (R core Team, 2021).

3. Results

The histogram shows the soil properties of the soil studied, showing that the diagrams of saturation (SP), nitrogen (N), phosphorus (P), sand and silt are symmetrical and possibly have a normal distribution. Other properties have a skew to the right or left (Figure 3). The box plot of the standard values of the studied soil properties data shows that soil saturation moisture, pH, nitrogen, phosphorus, CaCO3, sand and silt do not have outliers and are probably normal. However, the abundant outliers in electrical conductivity (EC), potassium, calcium, magnesium, chlorine, HCO3, sodium absorption ratio (SAR) and clay indicate a deviation from the normality of the data for these characteristics. (Figure 4). Also, in the quartile-quartile diagram, a number of soil samples deviate from the 45-degree line, which shows that they do not follow the normal distribution (Figure 4).

3.1. Data normality with a sample size of 25 soils

Some descriptive statistics are shown in Table 3. Soil physical and chemical properties with a coefficient of variation (CV) above 50% are not normal. Therefore, EC, Ca, Mg, Na, Cl and HCO3 probably do not follow a normal distribution. The skewness coefficients and standard error are calculated for all soil properties (Table 4). If the ratio of skewness to its standard error is less than -2 or greater than +2, its data is not normal ($p \leq 0.01$). As can be seen from Table 3, soil properties such as N, P, CaCO3, sand, silt and clay are normally distributed ($P \geq 0.05$).

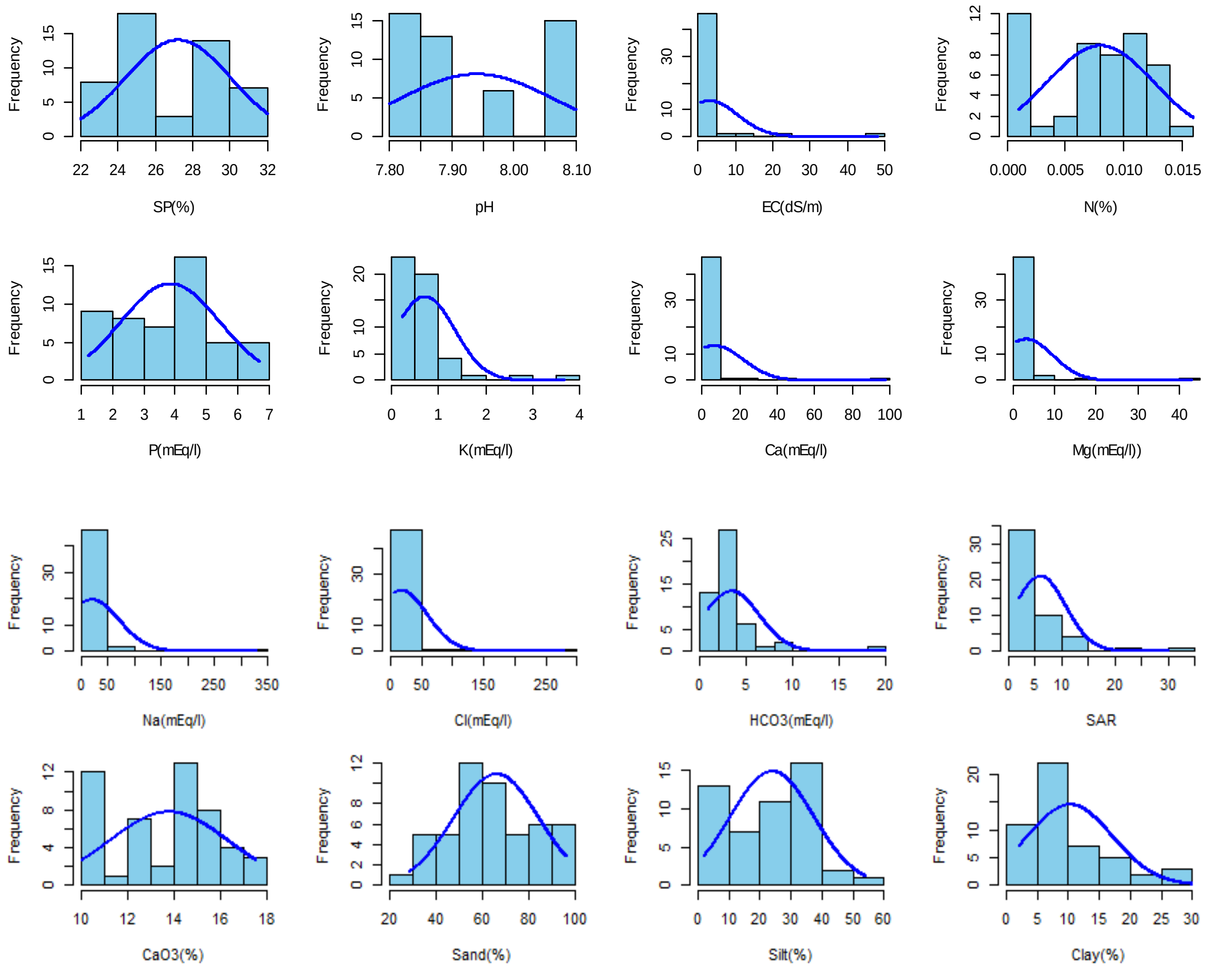


Fig. 3. Histogram with normal distribution curve for the soil properties data set in the study area

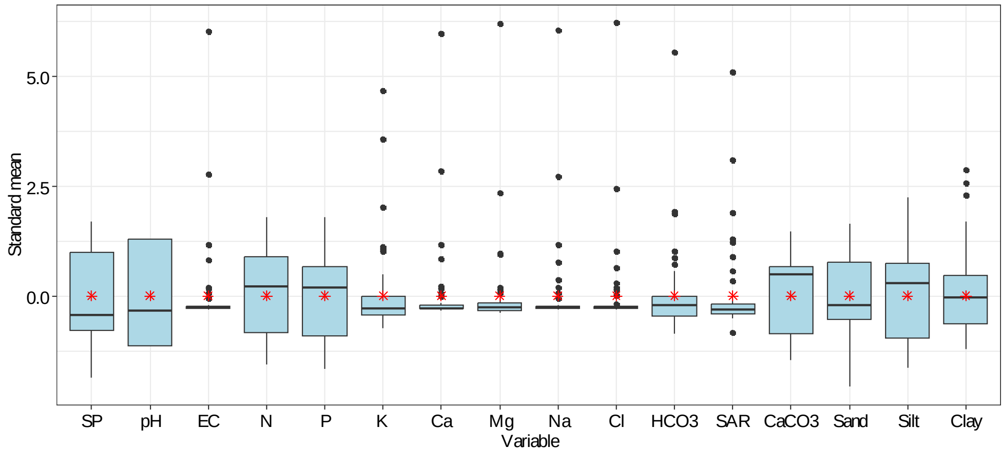


Fig. 4. Boxplot with the variables with the standardized data for the soil properties data set in the study area

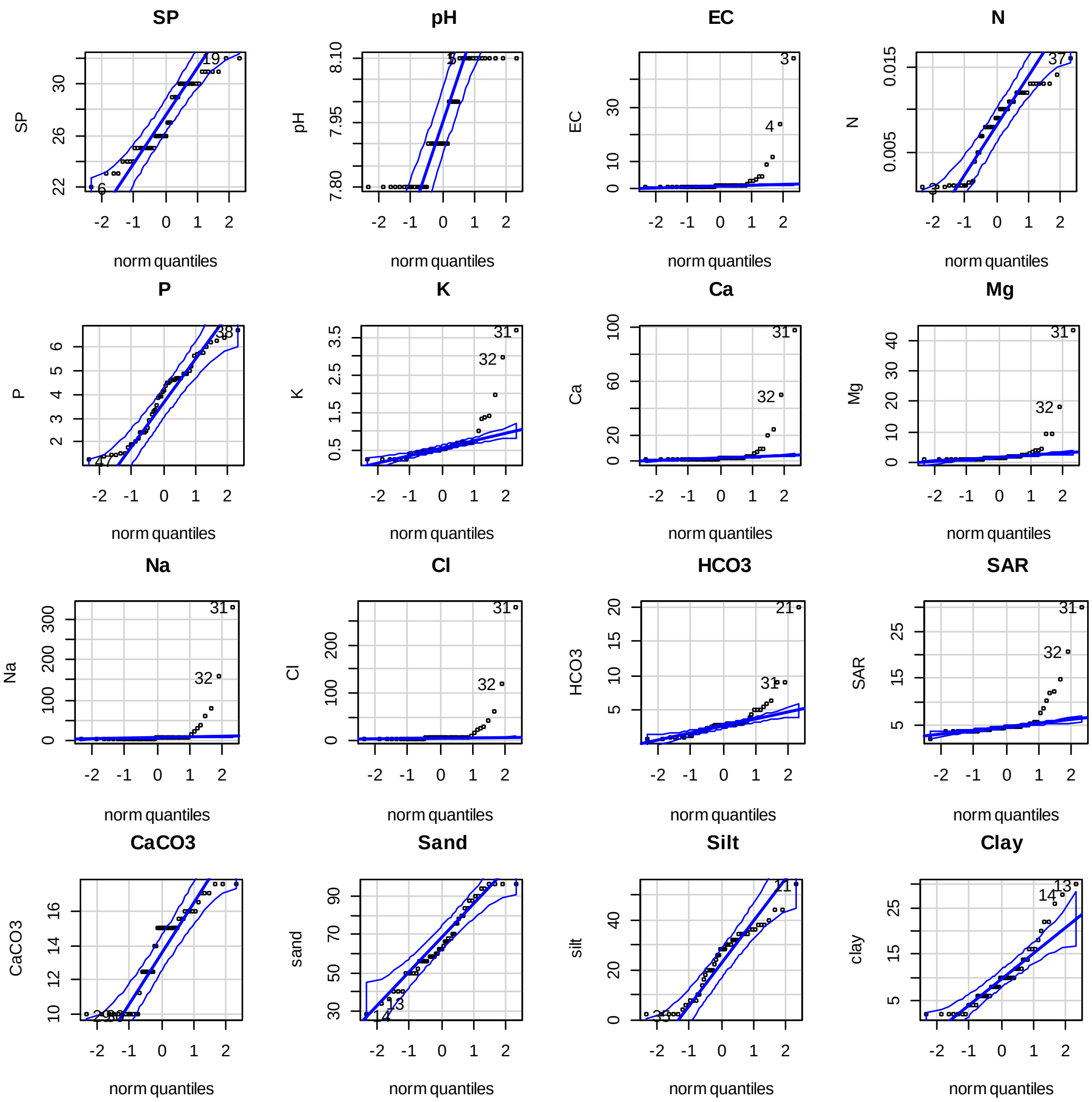


Fig. 5. Q-Q plot for the soil properties data set in the study area

Table 3. Descriptive statistics for the soil properties in the study area (N = 25)

Properties	Mean	Max	Min	SD	SE	CV	Normal
SP	27.5	35	23	3.86	0.77	14.1	Yes
pH	8.03	8.2	7.4	0.19	0.04	2.32	Yes
EC	3.01	8.43	0.93	2.43	0.48	**80.7**	**No**
N	0.01	0.023	0.004	0.005	0.00	46	Yes
P	2.48	4.7	1.4	0.84	0.17	33.8	Yes
K	0.90	1.9	0.4	0.44	0.09	49.3	Yes
Ca	4.14	13.2	2.05	2.56	0.51	**62**	**No**
Mg	2.93	8.54	1.45	1.67	0.33	**57**	**No**
Na	12.6	41.5	0.78	9.4	1.88	**74.8**	**No**
Cl	9.75	32	1.5	7.18	1.44	**73.7**	**No**
HCO3	5.28	20	1.8	3.85	0.77	**72.9**	**No**
SAR	5.84	12.6	2.4	2.16	0.43	37	Yes
CaCO3	11.1	20	5	3.91	0.78	35.2	Yes
Sand	61.8	80	52	6.73	1.35	10.9	Yes
Silt	29.3	36	18	4.89	0.98	16.7	Yes
Clay	9.28	16	2	3.82	0.76	41.2	Yes

Table 4. Results of skewness statistics and normality test of distribution of the soil properties (N=25)

Properties	Skewness	z	p	Type of skewness
SP	0.94	2.02	0.04^{*}	Near normal
pH	-1.97	4.26	0.00^{**}	Near normal
EC	1.11	2.40	0.02^{*}	Very extreme positive
N	0.19	0.40	**0.69**	Slight negative
P	0.82	1.76	**0.08**	Near normal
K	0.99	2.14	0.03^{*}	Very extreme positive
Ca	2.08	4.48	0.00^{**}	Very extreme positive
Mg	1.81	3.90	0.00^{**}	Very extreme positive
Na	1.85	3.98	0.00^{**}	Very extreme positive
Cl	1.66	3.58	0.00^{**}	Very extreme positive
HCO3	2.35	5.08	0.00^{**}	Very extreme positive
SAR	1.54	3.32	0.00^{**}	Very extreme positive
CaCO3	0.11	0.25	**0.80**	Slight negative
Sand	0.68	1.47	**0.14**	Near normal
Silt	-0.72	1.56	**0.12**	Near normal
Clay	0.16	0.35	**0.72**	Moderate positive

**, * Significant at 1% and 5% probability level, respectively

The statistics and significance level of the univariate normality tests are calculated in Table 5. As the results show, according to the Shapiro-Wilk, Cramer-von Mises, Anderson-Darling, Shapiro-Francia and Lilliefors tests, the properties such as N, P, sand, clay and sometimes silt follow the normal distribution. The result of the Kolmogorov-Smirnov test shows that most of the soil properties are normal except for saturation moisture, pH, EC and Na. The result of the

Pearson chi-square test shows that P, SAR, sand and silt are normal. The result of the Jarque-Bera test is almost identical to the result of the skewness test.

Table 5. Univariate normality test results of the soil properties (N = 25)

Properties	Shapiro-Wilk		Kolmogorov-Smirnov		Cramer-von Mises		Anderson-Darling	
	Statistics	p.value	Statistics	p.value	Statistics	p.value	Statistics	p.value
SP	0.83	0.00**	0.29	0.03*	0.30	0.00**	1.74	0.00**
pH	0.73	0.00**	0.31	0.00**	0.36	0.00**	2.15	0.00**
EC	0.77	0.00**	0.26	0.05*	0.46	0.00**	2.50	0.00**
N	0.94	**0.15**	0.14	**0.74**	0.06	**0.31**	0.49	**0.20**
P	0.93	**0.12**	0.13	**0.76**	0.07	**0.29**	0.46	**0.24**
K	0.87	0.00**	0.20	**0.25**	0.21	0.00**	1.17	0.00**
Ca	0.74	0.00**	0.24	**0.09**	0.39	0.00**	2.12	0.00**
Mg	0.78	0.00**	0.23	**0.15**	0.33	0.00**	1.87	0.00**
Na	0.75	0.00**	0.26	0.05*	0.46	0.00**	2.47	0.00**
Cl	0.81	0.00**	0.24	**0.10**	0.28	0.00**	1.64	0.00**
HCO3	0.75	0.00**	0.18	**0.35**	0.24	0.00**	1.59	0.00**
SAR	0.84	0.00**	0.23	**0.13**	0.27	0.00**	1.47	0.00**
CaCO3	0.91	0.02*	0.21	**0.20**	0.19	0.00**	1.06	0.00**
Sand	0.94	**0.17**	0.16	**0.48**	0.07	**0.22**	0.46	**0.23**
Silt	0.91	0.03*	0.16	**0.56**	0.11	**0.06**	0.76	0.04*
Clay	0.95	**0.26**	0.14	**0.67**	0.08	**0.19**	0.49	**0.20**

**,* Significant at 1% and 5% probability level, respectively

Table 5. Continued

Properties	Pearson chi-square		Shapiro-Francia		Lilliefors		Jarque-Bera	
	Statistics	p.value	Statistics	p.value	Statistics	p.value	Statistics	p.value
SP	29.72	0.00**	0.84	0.00**	0.29	0.00**	3.90	**0.14**
pH	32.28	0.00**	0.72	0.00**	0.31	0.00**	34.57	0.00**
EC	22.04	0.00**	0.78	0.00**	0.27	0.00**	5.27	**0.07**
N	16.92	0.00**	0.95	**0.22**	0.14	**0.26**	0.52	**0.77**
P	1.56	**0.91**	0.94	**0.12**	0.13	**0.28**	2.85	**0.24**
K	15.64	0.00**	0.88	0.00**	0.20	0.01**	4.12	**0.13**
Ca	15.64	0.00**	0.73	0.00**	0.24	0.00**	37.92	0.00**
Mg	16.92	0.00**	0.77	0.00**	0.23	0.00**	33.96	0.00**
Na	34.84	0.00**	0.74	0.00**	0.26	0.00**	22.36	0.00**
Cl	16.92	0.00**	0.81	0.00**	0.24	0.00**	17.69	0.00**
HCO3	12.44	0.03*	0.73	0.00**	0.19	0.03*	67.81	0.00**
SAR	10.52	**0.06**	0.83	0.00**	0.23	0.00**	16.55	0.00**
CaCO3	29.72	0.00**	0.91	0.03*	0.21	0.00**	0.27	**0.87**
Sand	5.40	**0.37**	0.94	**0.14**	0.17	**0.07**	2.00	**0.37**
Silt	9.88	**0.08**	0.91	0.04*	0.16	**0.10**	2.30	**0.32**
Clay	12.44	0.03*	0.96	**0.23**	0.15	**0.18**	0.51	**0.77**

**,* Significant at 1% and 5% probability level, respectively

The results of the Mardia and Henze-Zirkler multivariate normality tests show that, in general, the soil data matrix has a normal distribution ($p \geq 0.05$). Other tests violated the assumption of normality of the data ($p \leq 0.01$) (Table 6).

Table 6. Multivariate normality test results of the soil properties (N=25)

Test		Statistics	p.value	Normality
	Skewness	696.2	**0.99**	Yes
Mardia	Skewness, small sample corrected	786.94	**0.73**	Yes
	Kurtosis	-1.81	**0.07**	Yes
Henze-Zirkler		0.99	**0.08**	Yes
Royston		112.58	0.00**	No
Doornik-Hansen		410.54	0.00**	No
E-statistic		2.08	0.01**	No
Shapiro-Wilk		0.20	0.00**	No

**, * Significant at 1% and 5% probability level, respectively

3.2. Data normality with a sample size of 50 soils

The results show that based on the coefficient of variation (CV), only saturation moisture, pH, phosphorus, CaCO3 and sand follow a normal distribution (Table 7). Based on the skewness coefficient, properties such as saturation moisture, pH, nitrogen, phosphorus, CaCO3, sand and silt are normal ($p \geq 0.05$). Electrical conductivity (EC), potassium, calcium, magnesium, sodium, chlorine, HCO3 and the sodium absorption ratio (SAR) have an extreme skewness to the right (Table 8).

The results of the Shapiro-Wilk, Cramer-von Mises and Anderson-Darling tests show that only sand is normal ($p \geq 0.05$). The results of the Pearson chi-square, Shapiro-Francia and Lilliefors tests show that besides sand, phosphorus is also normal ($p \geq 0.05$). Here, the results of the Kolmogorov-Smirnov and Jarque-Bera tests are similar (Table 9).

Table 7. Descriptive statistics for the soil properties data set in the study area (N = 50)

Properties	Mean	Max	Min	SD	SE	CV	Normal
SP	27.2	32	22	2.82	0.34	10.4	Yes
pH	7.94	8.1	7.8	0.12	0.02	1.55	Yes
EC	3.03	48	0.75	7.48	1.06	**247**	**No**
N	0.008	0.016	0.001	0.004	0.0006	**56.2**	**No**
P	3.85	6.7	1.24	1.58	0.22	41.1	Yes
K	0.71	3.67	0.24	0.63	0.09	**89.2**	**No**
Ca	6.42	97.6	1.14	15.3	2.16	**238**	**No**
Mg	3.15	43	0.73	6.43	0.91	**204**	**No**
Na	20.2	329	4.79	51.2	7.23	**253**	**No**
Cl	17.1	280	4.39	42.3	5.99	**248**	**No**
HCO3	3.44	20	0.89	2.99	0.42	**86.8**	**No**
SAR	6.02	30.1	2.12	4.75	0.67	**78.8**	**No**
CaCO3	13.8	17.5	10	2.56	0.36	18.6	Yes
Sand	65.9	96	28	18.3	2.59	27.8	Yes
Silt	24	54	2	13.4	1.9	**55.9**	**No**
Clay	10.3	30	2	6.89	0.97	**67.1**	**No**

Table 8. Skewness statistics and results of the test of the normality of distribution of the soil properties (N = 50)

Properties	Skewness	z	p	Type of skewness
SP	0.09	0.27	**0.78**	Near normal
pH	0.20	0.59	**0.55**	Near normal
EC	4.95	14.71	0.00**	Very extreme positive
N	-0.41	1.23	**0.22**	Slight negative
P	-0.09	0.27	**0.78**	Near normal
K	3.20	9.51	0.00**	Very extreme positive
Ca	4.90	14.55	0.00**	Very extreme positive
Mg	5.20	15.44	0.00**	Very extreme positive
Na	4.97	14.78	0.00**	Very extreme positive
Cl	5.25	15.59	0.00**	Very extreme positive
HCO3	3.74	11.12	0.00**	Very extreme positive
SAR	3.50	10.40	0.00**	Very extreme positive
CaCO3	-0.36	1.08	**0.28**	Slight negative
Sand	0.09	0.27	**0.78**	Near normal
Silt	-0.21	0.61	**0.54**	Near normal
Clay	1.06	3.15	0.00**	Moderate positive

**, * Significant at 1% and 5% probability level, respectively

Table 9. Univariate normality test results of the soil properties data set in the study area (N=50)

Properties	Shapiro-Wilk		Kolmogorov-Smirnov		Cramer-von Mises		Anderson-Darling	
	Statistics	p.value	Statistics	p.value	Statistics	p.value	Statistics	p.value
SP	0.92	0.00**	0.19	0.06	0.34	0.00**	1.87	0.00**
pH	0.81	0.00**	0.21	0.03*	0.48	0.00**	3.44	0.00**
EC	0.32	0.00**	0.40	0.00**	2.75	0.00**	13.24	0.00**
N	0.89	0.00**	0.16	**0.14**	0.28	0.00**	2.00	0.00**
P	0.95	0.02*	0.12	**0.47**	0.14	0.00**	0.88	0.02*
K	0.59	0.00**	0.33	0.00**	1.32	0.00**	6.90	0.00**
Ca	0.33	0.00**	0.40	0.00**	2.69	0.00**	13.00	0.00**
Mg	0.34	0.00**	0.36	0.00**	2.41	0.00**	11.79	0.00**
Na	0.32	0.00**	0.42	0.00**	2.79	0.00**	13.39	0.00**
Cl	0.30	0.00**	0.41	0.00**	2.77	0.00**	13.36	0.00**
HCO3	0.61	0.00**	0.29	0.00**	0.93	0.00**	5.04	0.00**
SAR	0.52	0.00**	0.37	0.00**	1.80	0.00**	8.87	0.00**
CaCO3	0.87	0.00**	0.24	0.00**	0.39	0.00**	2.76	0.00**
Sand	0.96	0.07	0.10	**0.64**	0.11	0.09	**0.67**	0.07
Silt	0.94	0.02*	0.14	**0.30**	0.17	0.01**	1.08	0.01**
Clay	0.90	0.00**	0.18	**0.09**	0.22	0.00**	1.41	0.00**

**, * Significant at 1% and 5% probability level, respectively

Table 9. Continued

Properties	Pearson chi-square		Shapiro-Francia		Lilliefors		Jarque-Bera	
	Statistics	p.value	Statistics	p.value	Statistics	p.value	Statistics	p.value
SP	24.00	0.00**	0.92	0.00**	0.19	0.00**	3.77	**0.15**
pH	87.20	0.00**	0.83	0.00**	0.21	0.00**	5.22	0.07
EC	200.40	0.00**	0.30	0.00**	0.40	0.00**	1563	0.00**
N	28.20	0.00**	0.90	0.00**	0.16	0.00**	4.07	**0.13**
P	10.40	**0.17**	0.95	**0.06**	0.12	**0.07**	2.83	**0.24**
K	79.20	0.00**	0.57	0.00**	0.33	0.00**	326.30	0.00**
Ca	150.40	0.00**	0.31	0.00**	0.40	0.00**	1496	0.00**
Mg	131.20	0.00**	0.32	0.00**	0.36	0.00**	1910	0.00**
Na	199.60	0.00**	0.30	0.00**	0.42	0.00**	1597	0.00**
Cl	178.00	0.00**	0.28	0.00**	0.41	0.00**	1947	0.00**
HCO3	65.20	0.00**	0.59	0.00**	0.29	0.00**	743.50	0.00**
SAR	175.60	0.00**	0.50	0.00**	0.37	0.00**	464.90	0.00**
CaCO3	43.20	0.00**	0.88	0.00**	0.24	0.00**	4.39	**0.11**
Sand	8.00	**0.33**	0.97	**0.14**	0.10	**0.18**	1.56	**0.46**
Silt	13.20	**0.07**	0.95	0.03*	0.14	0.02*	2.00	**0.37**
Clay	25.20	0.00**	0.90	0.00**	0.18	0.00**	10.57	0.01**

**,* Significant at 1% and 5% probability level, respectively

The result of all multivariate normality tests (Table 10) shows that the matrix of the soil's physical and chemical properties is not normal (p≤0.01).

Table 10. Multivariate normality test results of the soil properties data set in the study area (N=50)

Test		Statistics	p.value	Normality
Mardia	Skewness	2333.70	0.00**	No
	Kurtosis	16.88	0.00**	No
Henze-Zirkler		1.13	0.00**	No
Royston		197.30	0.00**	No
Doornik-Hansen		1122.80	0.00**	No
E-statistic		3.49	0.00**	No
Shapiro-Wilk		0.12	0.00**	No

**Significant at 1% probability level

The result of the goodness-of-fit chi-square test shows that phosphorus follows normal and exponential distributions (p≥0.05). Also, sand follows normal, log-normal and exponential distributions (p≥0.05). Silt follows the normal distribution, and clay follows the exponential distribution (p≥0.05). Other soil properties do not follow any of the studied probability distributions (Table 11). To select the best distribution for soil phosphorus, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling statistics and AIC and BIC values were used. Because the normal distribution has the smallest AIC value, it can be concluded that phosphorus has the best fit with the normal distribution (Table 12).

Histograms of empirical and theoretical distributions of the normal and exponential distributions of phosphorus, along with the Q-Q plot, cumulative distribution (CDF) and probability-probability (P-P) plot, also confirm this result (Figure 6).

Table 11. Goodness-of-Fit test results of soil properties for normal, lognormal, gamma, and exponential with Chi-square test

Properties	p-value			
	Normal	Log Normal	Gamma	Exponential
SP	0.00**	0.00**	0.00**	0.00**
pH	0.00**	0.00**	0.00**	0.00**
EC	0.00**	0.00**	0.00**	0.00**
N	0.00**	0.00**	0.00**	0.00**
P	**0.17**	0.00**	0.00**	**0.08**
K	0.00**	0.00**	0.00**	0.00**
Ca	0.00**	0.00**	0.00**	0.00**
Mg	0.00**	0.00**	0.00**	0.00**
Na	0.00**	0.00**	0.00**	0.00**
Cl	0.00**	0.00**	0.00**	0.00**
HCO3	0.00**	0.00**	0.00**	0.00**
SAR	0.00**	0.00**	0.00**	0.00**
CaCO3	0.00**	0.00**	0.00**	0.00**
Sand	**0.33**	**0.15**	0.00**	**0.13**
Silt	**0.07**	0.00**	0.00**	0.00**
Clay	0.00**	0.02*	0.00**	**0.11**

**, * Significant at 1% and 5% probability level, respectively

Table 12. Comparison of goodness-of-fit statistics of four distributions of P by Akaike's Information Criterion and Bayesian Information Criterion

	Normal	log-Normal	Gamma	Exponential
Goodness-of-fit statistics				
Kolmogorov-Smirnov statistic	0.12	0.18	0.16	0.28
Cramer-von Mises statistic	0.15	0.43	0.25	1.44
Anderson-Darling statistic	0.91	3.19	1.38	7.61
Goodness-of-fit criteria				
Akaike's Information Criterion	**190.72**	201.76	192.66	236.77
Bayesian Information Criterion	**194.54**	205.59	196.48	238.68

Also, to choose the best distribution for sand, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling statistics and AIC and BIC values were used. Since the normal distribution has the smallest AIC value, it can be concluded that sand also has the best fit with the normal distribution (Table 13). Histograms of empirical and theoretical distributions of the normal, log-normal, and exponential distributions of sand, along with the Q-Q plot, cumulative distribution (CDF) and probability-probability (P-P) plot, also confirm this result (Figure 7). Normal and log normal distribution curves are similar.

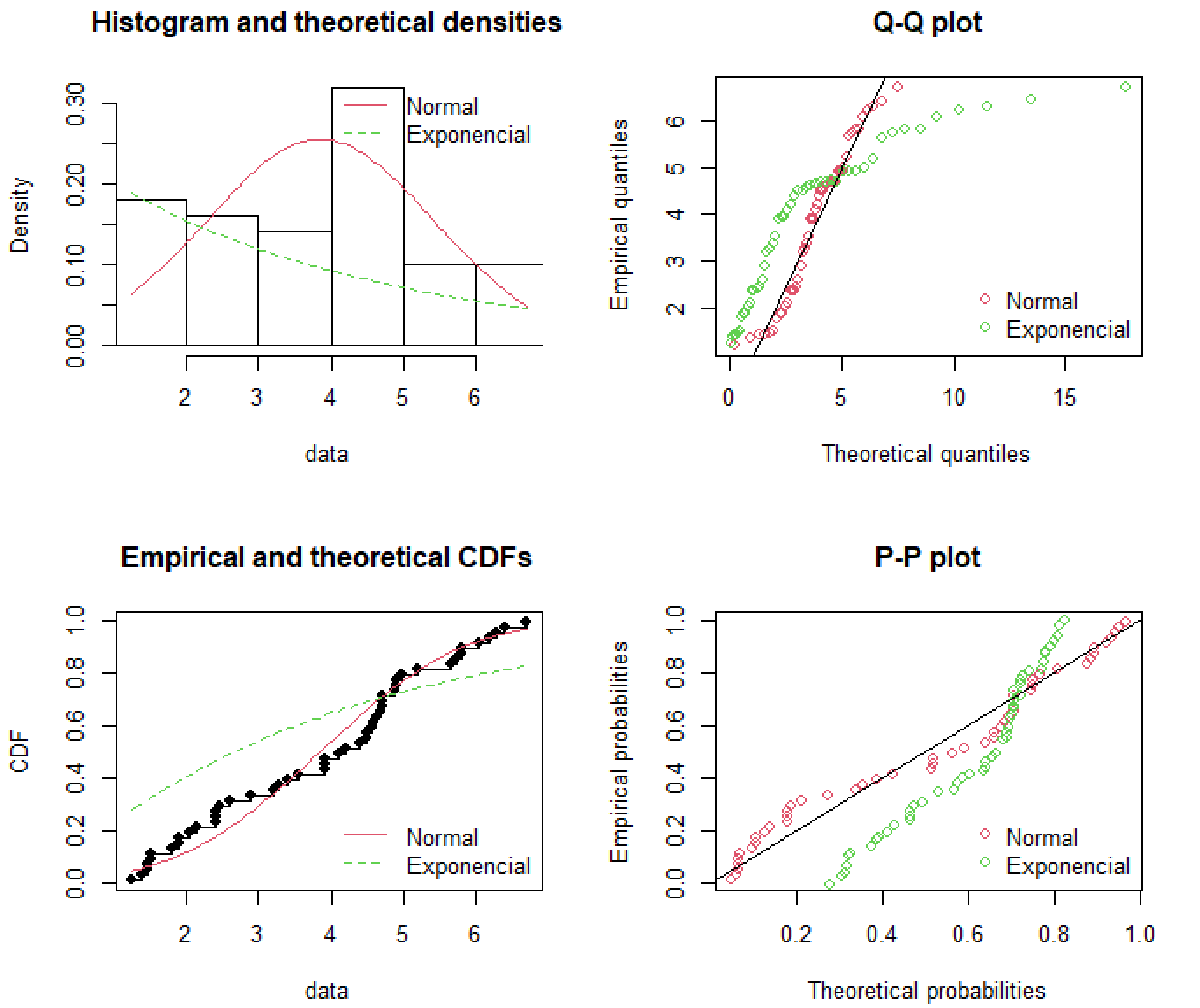


Fig. 6. Plot of empirical and theoretical distributions for P. The density plot or histogram plot in the top-left corner, quantile-quantile (Q-Q) plot or probability plot in the top-right corner, cumulative distribution function plot (CDF) in the bottom-left corner, and probability-probability (P-P) plot in the bottom-right corner.

Table 13. Comparison of goodness-of-fit statistics of four distributions of sand by Akaike's Information Criterion and Bayesian Information Criterion

	Normal	log-Normal	Gamma	Exponential
Goodness-of-fit statistics				
Kolmogorov-Smirnov statistic	0.11	0.09	0.08	0.41
Cramer-von Mises statistic	0.11	0.08	0.07	2.48
Anderson-Darling statistic	0.70	0.77	0.55	12.06
Goodness-of-fit criteria				
Akaike's Information Criterion	**435.74**	438.93	436.20	520.84
Bayesian Information Criterion	**439.56**	442.75	440.02	522.76

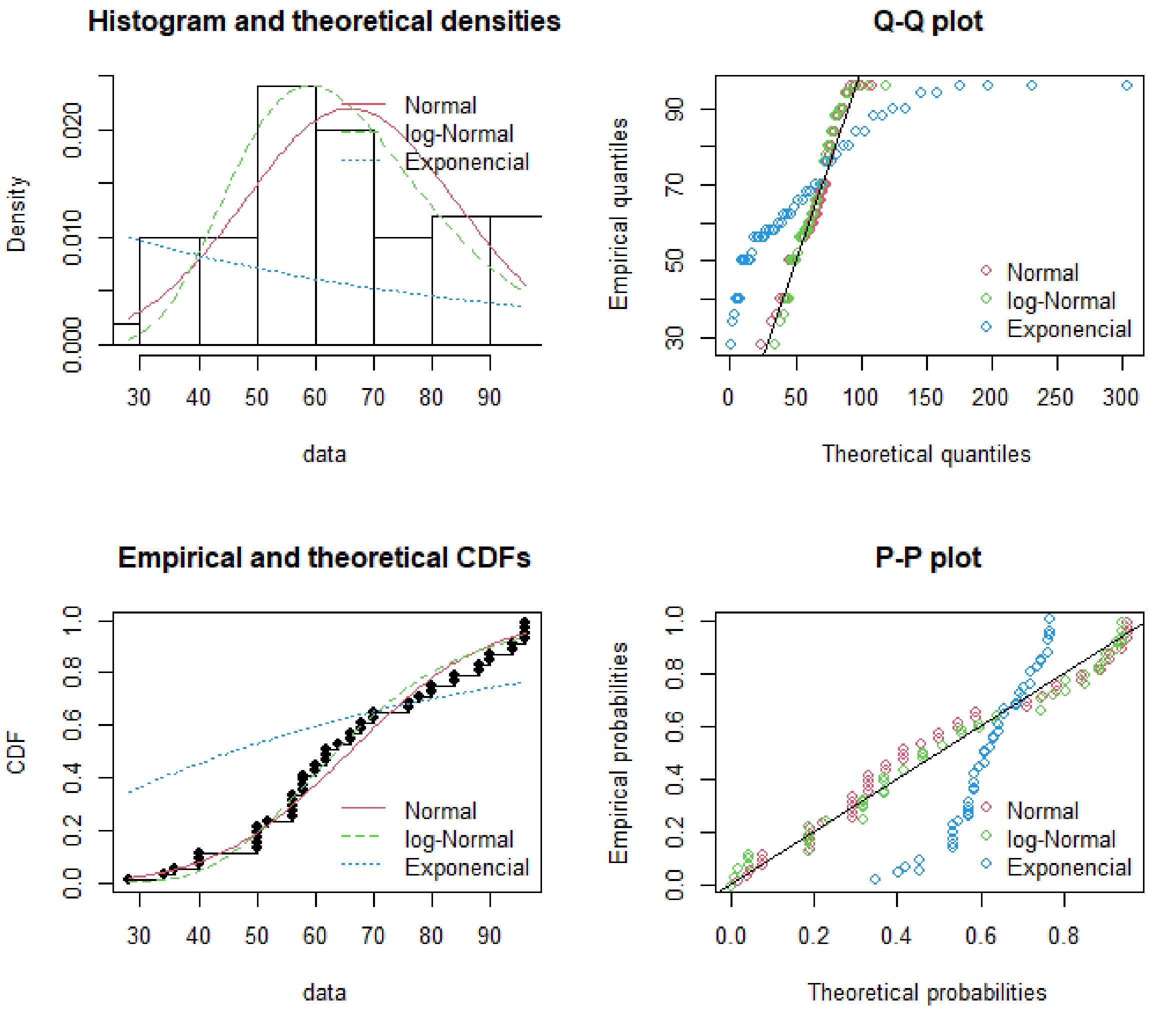


Fig. 7. Plot of empirical and theoretical distributions for sand. The density plot or histogram plot in the top-left corner, quantile-quantile (Q-Q) plot or probability plot in the top-right corner, cumulative distribution function plot (CDF) in the bottom-left corner, and probability-probability (P-P) plot in the bottom-right corner.

4. Discussion

The results of the present study showed that most soil properties do not follow a normal distribution. The results showed that in both halophyte and psammophyte habitats, sand had a normal distribution. These results are similar to those reported by Shukla and Sharma (2023) who observed that most soil properties were not normally distributed. The measured properties were transformed using a natural logarithm, and some of the properties became near normal. A series of recent studies has indicated that most of the soil properties were not normally distributed (Piotrowska-Długosz et al., 2019; Obi et al., 2020; Aggag and Alharbi, 2022; Šestak et al., 2022). Obi et al. (2020) have found that the profile characteristics that were non-normally distributed were those that are usually influenced by pedogenesis and since the landscape characteristics and genesis are different, they are from different populations, and their characteristics should not be normally distributed.

The results of the present study showed that most of the soil properties do not follow normal and lognormal distributions. Similar to the present study, Awal et al. (2019) reported that porosity, bulk density, soil surface temperature, and soil organic carbon seem to follow neither a normal nor a log-normal distribution. Although some researchers have pointed out that the data transformation of skewed soil properties using a natural logarithm improved the normality of most of the properties (Piotrowska-Długosz et al., 2019; Moharana et al., 2021; Löfman and Korkiala-Tanttu, 2021; Li et al., 2022; Aggag and Alharbi, 2022; Shukla and Sharma, 2023), but Bagheri Bodaghabadi (2018) discovered that data transformation had no effect on the normalization of the distribution of the predictions and the errors. Therefore, since some soil data do not follow the log normal distribution, it seems that log transformation does not help improve this research data.

In the present study, different methods were used to measure the normality of the data. Data from this study suggest that the Shapiro–Wilk test is a more appropriate method for small sample sizes (n <50). During his research, Khatun (2021) investigated the power of common univariate normality tests. The results showed that with the increase in the number of samples, the overall power increased, but the Shapiro-Wilk test, Shapiro-Francia test and Anderson-Darling test were the most powerful tests among the others. Cramer-Von-Mises test has a better performance than the Pearson Chi-Square. Lilliefors test has better power than the Jarque-Bera test. Jarque-Bera test has less power among other tests. Yap and Sim (2011) compared different types of normality tests and concluded that Shapiro-Wilk and D'Agostino tests have better power for symmetric short-tailed distributions and asymmetric distributions. Razali and Yap (Razali & Yap, 2011) compared the power of the Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests and concluded that the Shapiro-Wilk test is the most powerful normality test. The Shapiro-Wilk test is suitable for small sample sizes (N<50), although it can be performed in larger sample sizes, while the Kolmogorov-Smirnov test is used for N ≥50 (Mishra et al., 2019).

The limitation of the Kolmogorov-Smirnov test is its high sensitivity to outliers. The Shapiro-Wilk test provides better power than the Kolmogorov-Smirnov test, even after Lilliefors correction. Therefore, the Kolmogorov-Smirnov test should no longer be used due to its low power (Ghasemi & Zahediasl, 2012). The Jarque Bera test, which is known in statistics as the D'Agostino-Pearson or Bowman-Shenton test and is mostly used in economics (Kim, 2016), had a similar result to the results of the skewness and Kolmogorov-Smirnov tests. The Henze-Zirkler (HZ) tests are more recommended in the multivariate normality tests (Zhou & Shao, 2014).It is worth mentioning that in multivariate analyses of soil and plant relationships, multivariate normality is important, not univariate, so that it can be said that a univariate normal distribution does not guarantee the presence of a multivariate normal distribution (Wulandari et al., 2021).

Although the Kolmogorov-Smirnov test is generally used in rangeland and desert science research, this test is not suitable for research whose sample size is less than 100 and has outlier data (Schoder et al., 2006). Because normality tests have low statistical power, Liu et al., (2023) recommend the simultaneous use of graphical representations such as Q-Q plot and normality tests. In this research, the simultaneous use of numerical and graphic descriptive methods along with statistical tests was investigated, and the results were complementary to each other.

Ghasemi and Zahediasl (2012) pointed out that, it should be noted that normality tests are sensitive to sample size. Small samples often pass normality tests. Also, if the sample size is greater than 50, the Q-Q plot is more recommended than statistical methods because, in larger sample sizes, the Shapiro-Wilk test is very sensitive even to a small deviation from normality (Ghasemi & Zahediasl, 2012). For studies with a small sampling size, none of these methods are satisfactory. On the other hand, for studies with numerous samples, some normality tests may be very sensitive, so it is recommended to use numerical and graphical descriptive methods simultaneously with statistical tests (Yang and Berdine, 2021).

The results of the goodness-of-fit test showed that phosphorus, sand and silt have a better fit with a normal distribution. Other soil properties do not follow any of the normal, log-normal, gamma, or exponential probability distributions. The results of Amer et al. (2021) and Aggag and Alharbi (2022) in the arid regions of Egypt and Saudi Arabia also showed that none of the soil physical and chemical properties follow a normal distribution.

As stated in the introduction, the soil environment is a heterogeneous environment. The greater the internal changes between the soil data, the greater the standard deviation, and as a result, the coefficient of variation increases, therefore, in heterogeneous areas, the possibility of outliers are high. The result of this research also showed that the data of electrical conductivity, potassium, calcium, magnesium, sodium, chlorine, HCO3 and sodium absorption ratio (SAR) have a very high coefficient of variation, while nitrogen, phosphorus, silt and clay have a high coefficient of variation. CaCO3 and sand have medium coefficient of variation, and saturation moisture and pH have low coefficient of variation. This classification is somewhat consistent with that of Carter and Gregorich (2008) (Table 13).

Now the question is, based on the results of this research, is the number of 50 soil samples enough to perform parametric and multivariate statistical tests or not? If the objective is to determine bulk density (CV<15%) at a level of 0.05 and the relative error rate is 0.10, approximately 4 soil samples are required for one plant type. To measure soil electrical conductivity (CV 75%-100%) at the same level of 0.05 but with an error of 0.25, about 62 soil samples are required (See Table 2).

Table 13. Variability of soil properties

Coefficient of Variation (Carter & Gregorich, 2008)			
Low (CV<15%)	Moderate (CV 15%-35%)	High (CV 35%-75%)	Very high (CV 75%-100%)
Soil hue and value pH A horizon Thickness Silt content Porosity Bulk density	Sand content Clay content CEC % BS $CaCO_3$ equivalent Soil organic C	Solum thickness Exchangeable Ca, Mg and K Soil nitrate N Soil-available P Soil-available K	Nitrous oxide flux Electrical conductivity Saturated hydraulic conductivity Solute dispersion coefficient
In this study			
SP pH	CaCO3 Sand	N, P Silt Clay	EC K, Ca, Mg, Na, Cl HCO3 SAR

5. Conclusion

The present research examined some of the most widely used methods and tests for evaluating the normality of data in the environmental sciences. The results showed that although graphical methods and coefficient of variation (CV) are quick methods for detecting normality, they require experience and statistical knowledge. In 25 soil samples, the results of the coefficient of variation method and skewness test were almost similar to each other. The results showed that the Kolmogorov-Smirnov test was the only one that reported the most characteristics with a normal distribution. Some multivariate normality tests showed the normality of the soil properties matrix and can be used for multivariate analyses where multivariate normality is one of the assumptions.

In total, the results of this research showed that nitrogen, phosphorus, sand and clay have a normal distribution. However, in statistics sources, it is stated that in large samples (N> 30) based on the central limit theorem, there is no need to worry about the normality of the data. For this purpose, the number of soil samples increased from 25 to 50 to investigate this issue. The results showed that with the increase in the number of samples, the coefficient of variation and skewness increased several times, so that the number of outliers in the box plot of some soil properties increased from two to six. The results of the Kolmogorov-Smirnov test and skewness test were similar, and the number of normal soil characteristics decreased from 12 items to 6 items (N = 25). The results showed that with the increase in the number of soil samples, multivariate normality was violated, and using the same data for multivariate analysis is not suitable.

The last point is that the statistical outliers of the data do not necessarily mean an error in sampling, laboratory analysis, or data recording, but the random sampling method and the variability of soil properties in desert areas also lead to the observation of very large or very small data compared to the mean, and these data violate normality. In the present study, it was shown that with an increase in the number of samples, the probability of the presence of outlier data increases. Although the increase in the number of samples increased the cost of the research, it had no effect on improving the normality of the data. Therefore, to perform valid statistical tests, solutions such as removing outliers, using a trimmed mean, transforming the data, using robust parametric tests and finally using non-parametric tests are recommended.

References

Aggag A.M., and Alharbi A. 2022. Spatial Analysis of Soil Properties and Site-Specific Management Zone Delineation for the South Hail Region, Saudi Arabia. *Sustainability*. 14(23):16209.

Al-Kaisi M. M., Lal R., Olson K. R., and Lowery B. 2017. Fundamentals and Functions of Soil Environment. In: Al-Kaisi M. M. and Lowery B. (Ed.), Soil Health and Intensification of Agroecosystems (pp. 1-23). Elsevier Inc.

Amer, B.S., Moussa, K.F., Sheha, A.A. and Abdel-Fattah, M.K. 2021. Delineation of site-specific management zones using multivariate analysis and geographic information system technique. Plant Archives, 21: 1385–1390

Awal R, Safeeq M, Abbas F, Fares S, Deb SK, Ahmad A, Fares A. 2019. Soil Physical Properties Spatial Variability under Long-Term No-Tillage Corn. *Agronomy*. 9(11):750.

Awal R, Safeeq M., Abbas F., Fares S., Deb S.K., Ahmad A., and Fares A. 2019. Soil Physical Properties Spatial Variability under Long-Term No-Tillage Corn. *Agronomy*, 9(11):750.

Bagheri Bodaghabadi M. 2018. Is it necessarily a normally distributed data for kriging? A case study: soil salinity map of Ghahab area, central Iran. *Desert*, *23*(2), 284-293.

Blanca M. J., Alarcón R., Arnau J., Bono R., and Bendayan R. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29(4): 552–557.

Carter M.R., and Gregorich E.G., Eds. 2008. Soil Sampling and Methods of Analysis. 2nd Edition, CRC Press, Taylor & Francis Group, Boca Raton. 1224 p.

Cooksey R. W. 2020 Descriptive Statistics for Summarizing Data. *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, 61–139.

Froughifar H., Jafarzadah A.A., Torabi Gelsefidi H. and Aliasgharzadah, N. 2012. Effect of different landforms on spatial variability and frequency distribution of soil biological properties in Tabriz Plain. *Water and Soil Science (Agricultural Science)*, 21(4): 35-52. (In Persian)

Fukumasu J., Jarvis N., Koestel J., Katterer T., and Larsbo M. 2022. Relations between soil organic carbon content and the pore size distribution for an arable topsoil with large variations in soil properties. *European Journal of Soil Science*, 73(1), e13212.

Gel Y.R., Miao W., and Gastwirth J.L. 2007. Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, 51(5), 2734–2746.

Greenacre M.J., and Primicerio R. 2013. Multivariate analysis of ecological data. Fundación BBVA. 336 p.

Hair J.F., Black W.C., Babin B.J., and Anderson R.E. 2010. Multivariate Data Analysis. 7th Edition, Pearson, New York. 816 p.

Han J., Liu X., Wu Y., and Zhou X. 2020. Importance of Distribution Type on Bearing Capacity of an Embedded Foundation in Spatially Varying Soils. *Advances in Civil Engineering*, 1-15.

Hartemink A. E., Zhang Y., Bockheim J. G., Curi N., Silva S. H. G., Grauer-Gray J., Lowe D. J., and Krasilnikov P. 2020. Soil horizon variation: A review. In D. L. Sparks (Ed.), *Advances in Agronomy* (1 ed., Vol. 160, pp. 125-185). (Advances in Agronomy; Vol. 160, No. 1). Academic Press.

Hwang S. I., and Choi S. I. 2006. Use of a lognormal distribution model for estimating soil water retention curves from particle-size distribution data. *Journal of Hydrology*, 323(1-4): 325-334.

Jafari Haghighi M., 2003. Method of Soil Analysis sampling and Important Physical and Chemical Analysis with emphasis on theoretical and applied principles. Publication Nedaye zohi. 236 pp. (In Persian)

Jafari M., and Rostampour M. 2019. Soil - Plant Relationships: Ecology, Statistics and analysis (Vol. 1). University of Tehran Press, Tehran, 468 p. (In Persian)

Kent M. 2012. Vegetation Description and Data Analysis: A Practical Approach, 2nd Ed. Wiley-Blackwell, 448 p.

Khatun N. 2021. Applications of Normality Test in Statistical Analysis. *Open Journal of Statistics*, 11: 113-122.

Kim N. 2016. A robustified Jarque–Bera test for multivariate normality. *Economics Letters*, 140: 48-52.

Kim T. K., and Park J. H. 2019. More about the basic assumptions of t-test: normality and sample size. *Korean journal of anesthesiology*, 72(4): 331–335.

Knief U., and Forstmeier W. 2021. Violating the normality assumption may be the lesser of two evils. *Behaviour Research,* 53: 2576–2590.

Korkmaz, S. 2022. Package “MVN”. Multivariate Normality Tests. https://cran.r-project.org/web/packages/MVN/MVN.pdf

Legendre P., and L. Legendre 2012. Numerical ecology. Third edition. Elsevier, Amsterdam, The Netherlands. 853 p.

Li G., Wang M., Ma C., Tao R., Hou F., and Liu Y. 2021. Effects of Soil Heterogeneity and Species on Plant Interactions. *Frontiers in Ecology and Evolution*, 9:756344.

Li CH. 2021. Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables. *Behav Res 53*, 2191–2213

Li X., Chen, Y., Lv G., Wang J., Jiang L., Wang H. and Yang X. 2022. Predicting spatial variability of species diversity with the minimum data set of soil properties in an arid desert riparian forest. *Front. Plant Sci.* 13:1014643.

Liu Y., Li G., Wang M., Yan W., and Hou F. 2021. Effects of three-dimensional soil heterogeneity and species composition on plant biomass and biomass allocation of grass-mixtures. *AoB PLANTS*, 13(4): plab033.

Liu Z., Al Amer F.M., Xiao M. Xu C., Furuya-Kanamori L., Hong H., Siegel L., and Lin L. 2023. The normality assumption on between-study random effects was questionable in a considerable number of Cochrane meta-analyses. *BMC Medicine* 21, 112.

Löfman M.S., Korkiala-Tanttu L.K. 2021. Inherent Variability of Geotechnical Properties for Finnish Clay Soils. In: Matos, J.C., *et al.* 18th International Probabilistic Workshop. IPW 2021. Lecture Notes in Civil Engineering, vol 153. Springer, Cham.

McGrath D., Zhang C. 2003. Spatial distribution of soil organic carbon concentrations in grassland of Ireland. *Appl Geochem*, 18:1629–1639.

Meyers L. S., Gamst G. C., and Guarino A. J. 2012. Applied Multivariate Research: Design and Interpretation. SAGE Publications, Inc. 1104 p.

Mishra P., Pandey C. M., Singh U., Gupta A., Sahu C., and Keshri A. 2019. Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, 22(1): 67–72.

Moharana P.C., Jena R.K., Kumar N., Singh R.S., and Rao S.S. 2021. Assessment of soil organic and inorganic carbon stock at different soil depths after conversion of desert into arable land in the hot arid regions of India. *Carbon Manag*, 12, 153–165.

Nunes F. C., Lander A., Claudia C., Eduardo G., Thaís S., and Prasad M. N. V. 2020. Chapter 9 - Soil as a complex ecological system for meeting food and nutritional security. In: Prasad, M. N. V., Pietrzykowski, M. (Eds.) Climate Change and Soil Interactions. Elsevier. 229-269.

Obi J. C., Udoh İ. B., and Obi İ. C. 2020. Modelling soil properties from horizon depth functions and terrain attributes: An example with cation exchange capacity. Eurasian Journal of Soil Science, 9(1), 10-17.

Oppong F. B., and Agbedra S. Y. 2016. Assessing Univariate and Multivariate Normality, a Guide for Non-Statisticians. Mathematical Theory and Modeling, 6(2): 26-33.

Piotrowska-Długosz A., Breza-Boruta B. and Długosz J. 2019. Spatio-temporal heterogeneity of soil microbial properties in a conventionally managed arable field. *J Soils Sediments* **19**, 345–355.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Razali N., and Yap B. W. 2011. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2: 21-33.

Rostampour M. 2022. Rangeland ecosystems monitoring in different climatic regions of Iran, South Khorasan Province, Zirkouh site. Research Institute of Forests and Rangelands, Tehran, Iran. (In Persian).

Schoder V., Himmelmann A., and Wilhelm K. P. 2006. Preliminary testing for normality: some statistical aspects of a common concept. *Clinical and experimental dermatology*, 31(6): 757–761.

Šestak I., Pereira P., Telak L.J., Perčin A., Hrelja I., Bogunović I. 2022. Soil Chemical Properties and Fire Severity Assessment Using VNIR Proximal Spectroscopy in Fire-Affected Abandoned Orchard of Mediterranean Croatia. *Agronomy*. 12(1):129.

Shukla M.K., and Sharma P. 2023. Fuzzy K-Means and Principal Component Analysis for Classifying Soil Properties for Efficient Farm Management and Maintaining Soil Health. *Sustainability*. 15(17):13144.

Tsagris M., and Pandis N. 2021. Normality test: Is it really necessary? *American journal of orthodontics and dentofacial orthopedics*: official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics, 159(4): 548–549.

Weine E., McPeek M. S., and Abney M. 2023. Application of Equal Local Levels to Improve Q-Q Plot Testing Bands with R Package qqconf. *Journal of statistical software*, *106*(10), 10.

Wulandari D., Sutrisno S., and Nirwana, M. B. 2021. Mardia’s Skewness and Kurtosis for Assessing Normality Assumption in Multivariate Regression. *Enthusiastic*: *International Journal of Applied Statistics and Data Science*, 1(1): 1–6.

Yang S., and Berdine G. 2021. Normality tests. *The Southwest Respiratory and Critical Care Chronicles*, 9(37):87–90.

Yap B.W., and Sim C.H. 2011. Comparisons of Various Types of Normality Tests. *Journal of Statistical Computation and Simulation*, 81: 2141-2155.

Zhang L., Mi X., Harrison R. D., Yang B., Man X., and Ren H. 2020. Resource heterogeneity, not resource quantity, plays an important role in determining tree species diversity in two species-rich forests. *Frontiers in Ecology and Evolution*, 8:224.

Zhou M., and Shao Y. 2014. A Powerful Test for Multivariate Normality. *Journal of applied statistics*, 41(2): 351–363.