International Journal of Assessment Tools in Education
2020, Vol. 7, No. 2, 255–265
https://doi.org/10.21449/ijate.656077

Published at https://dergipark.org.tr/en/pub/ijate *Research Article*

Parametric or Non-parametric: Skewness to Test Normality for Mean Comparison

Fatih Orcan [1,*]

[1]Department of Educational Measurement and Evaluation, Trabzon University, Turkey

ARTICLE HISTORY

Received: Dec 06, 2019
Revised: Apr 22, 2020
Accepted: May 24, 2020

KEYWORDS

Normality test,
Skewness,
Mean comparison,
Non-parametric,

Abstract: Checking the normality assumption is necessary to decide whether a parametric or non-parametric test needs to be used. Different ways are suggested in literature to use for checking normality. Skewness and kurtosis values are one of them. However, there is no consensus which values indicated a normal distribution. Therefore, the effects of different criteria in terms of skewness values were simulated in this study. Specifically, the results of t-test and U-test are compared under different skewness values. The results showed that t-test and U-test give different results when the data showed skewness. Based on the results, using skewness values alone to decide about normality of a dataset may not be enough. Therefore, the use of non-parametric tests might be inevitable.

1. INTRODUCTION

Mean comparison tests, such as t-test, Analysis of Variance (ANOVA) or Mann-Whitney U test, are frequently used statistical techniques in educational sciences. The techniques used differ according to the properties of the data sets such as normality or equal variance. For example, if the data is not normally distributed Mann-Whitney U test is used instead of independent sample t-test. In a broader sense, they are categorized as parametric and non-parametric statistics respectively. Parametric statistics are based on a particular distribution such as a normal distribution. However, non-parametric tests do not assume such distributions. Therefore, they are also known as distribution free techniques (Boslaung & Watters, 2008; Rachon, Gondan, & Kieser, 2012).

Parametric mean comparison tests such as t-test and ANOVA have assumptions such as equal variance and normality. Equal variance assumption indicates that the variances of the groups which are subject to test are the same. The null hypothesis for this assumption indicated that all the groups' variances are equal to each other. In other words, not rejecting the null hypothesis shows equality of the variances. The normality assumption, on the other hand, indicates that the data were drawn from a normally distributed population. A normal distribution has some properties. For example, it is symmetric with respect to the mean of the distribution where the mean, median and mode are equal. Also, normal distribution has a horizontal asymptote (Boslaung & Watters, 2008). That is, the curve approaches but never touches the x-axis. With

CONTACT: Fatih Orcan ✉ fatihorcan@trabzon.edu.tr Trabzon University, Fatih Collage of Education, Room: C-110, Trabzon, Turkey

ISSN-e: 2148-7456 /

normality assumption, it is expected that the distribution of the sample is also normal (Boslaung & Watters, 2008; Demir, Saatçioğlu & İmrol, 2016; Orçan, 2020). In case for comparison of two samples, for example, normality assumption indicates that each independent sample should be distributed normally. Departure from the normality for any of the independent sample indicates that the parametric tests should not be used (Rietveld & van Hout, 2015) since the type I error rate is affected (Blanca, Alarcon, Arnua, et al., 2017; Cain, Zhang, & Yuan, 2017). That is, parametric tests are robust in terms of type I error rate (Demir et al., 2016) and as the distribution of the groups apart from each other type I error rate raises (Blanca et al., 2017)

For independent samples, test of normality should be run separately for each sample. Checking the normality of the dependent variable for entire sample, without considering the grouping variable (the independent variable), is not the correct way. For example, if a researcher wants to compare exam scores between male and female students, the normality of exam scores for male and female students should be tested separately. If one of the groups is normally and the other is non-normally distributed the normality assumption is violated. Only if both groups' tests indicate normal distribution then parametric tests (i.e., independent sample t-test) should be considered. On the other hand, for one sample t-test or paired samples t-test (testing difference between pairs), normalities of the dependent variables are tested for entire sample at once.

Normality could be tested with variety of ways, some of which are Kolmogorov-Smirnov (KS) test and Shapiro-Wilk (SW) test. These are two of the most common ways to check normality (Park, 2008; Razali & Wah, 2011). Both tests assume that the data is normal, H_0. Therefore, it was expected to not to reject the null (Miot, 2016). KS test is recommended to use when the sample size is large while SW is used with small sample sizes (Büyüköztürk et al., 2014; Demir et al., 2016; Razali & Wah, 2011). Park (2008) pointed that SW test is not reliable when sample size is larger than 2000 while KS is usefull when the sample size is larger than 2000. However, it was also pointed that SW test can be powerful with large sample sizes (Rachon et al., 2012). Besides, it was stated that KS test is not useful and less accurate in practice (Field, 2009; Ghasemi & Zahediasl, 2012; Schucany & Tong NG, 2006).

In addition, KS and SW tests, other ways are also available for checking the normality of a given data set. Among them, few graphical methods are also available: Histogram, boxplot or probability-probability (P-P) plots (Demir 2016; Miot, 2016; Park, 2008; Rietveld & van Hout, 2015). For example, shape of the histogram for a given data set is checked to see if it looks normal or not. Even though it is frequently used, the decisions made based only on it would be subjective. Nevertheless, using histogram with other methods to check the shape of the distribution can be informative. Therefore, it will be useful to use graphical methods with other methods.

Another way to check the normality of data is based on checking skewness and kurtosis values. Although the use of skewness and kurtosis values are common in practice, there is no consensus about the values which indicate normality. Some suggest skewness and kurtosis up to absolute value of 1 may indicate normality (Büyüköztürk, Çokluk, & Köklü, 2014; Demir et al., 2016; Huck, 2012; Ramos et al., 2018), while some others suggest much larger values of skewness and kurtosis for normality (Iyer, Sharp, & Brush, 2017; Kim, 2013; Perry, Dempster & McKay, 2017; Şirin, Aydın, & Bilir, 2018; West et al., 1996). Lei and Lomax (2005) categorized non-normality into 3 groups: "The absolute values of skewness and kurtosis less than 1.0 as slight nonnormality, the values between 1.0 and about 2.3 as moderate nonnormality, and the values beyond 2.3 as severe nonnormality" (p. 2). Similarly, Bulmer (1979) pointed skewness, in absolute values, between 0 and .5 shows fairly symmetrical, between .5 and 1 shows moderately skewed and larger than 1 shows highly skewed distribution.

Standard error of skewness and kurtosis were also used for checking normality. That is, z-scores for skewness and kurtosis were used as a rule. If z-scores of skewness and kurtosis are smaller than 1.96 (for %5 of type I error rate) the data was considered as normal (Field, 2009; Kim, 2013). Besides, for larger sample sizes it was suggested to increase the z-score from 1.96 up to 3.29 (Kim, 2013).

Sample size is also an important issue regarding normality. With small sample size normality of a data cannot be quarantined. In an example, it was shown that sample of 50 taken from normal distribution looked nonnormal (Altman, 1991, as cited in Rachon et al., 2012). Blanca et al. (2013) examined 693 data sets with sample sizes, ranging between 10 and 30, in terms of skewness and kurtosis. They found that only 5.5% of the distributions were close to normal distribution (skewness and kurtosis between negative and positive .25). It was suggested that even with small sample size the normality should be controlled prior to analysis.

Since parametric tests are more powerful (Demir et al. 2016) researchers may try to find a way to show that their data is normal. Sometimes only SW or KS test are used while sometimes values such as skewness and kurtosis are used. In fact, based on Demir et al. (2016) study, 24.8% of the studies which test normality used skewness and kurtosis values while 24.1% of them used KS or SW tests. Even though the difference between the percentages is small, more researchers used skewness and kurtosis to check normality. There might be different reasons why researchers use skewness and kurtosis values to check normality. One of which might be related to get broader flexibility on the reference values of skewness and kurtosis. As indicated, different reference points on skewness and kurtosis were available in the literature. Therefore, it seems that it is easier for the researchers to show normality by using skewness and kurtosis values.

Based on the criteria chosen to check normality it is decided to use parametric or nonparametric tests. If the criterion is changed, the test to be chosen might also change. For example, if one use “skewness smaller than 1” instead of “z-score of skewness” criteria t-test instead of U-test might need to be used. In fact, normality test results might change with respect to the test which is used to utilized (Razali & Wah, 2011). Therefore, the aim of this study is to see how much difference might occur on decisions made on the used of t-test and U-tests under different skewness criteria. It was not aimed to point whether parametric or non-parametric tests are more or less useful then the other one. For this purpose, a simulation study was conducted with different design factors.

2. METHOD

2.1. Study Design Factors

Three different design factors were used to simulate independent sample testing proses. The first design factor was sample size. In order to simulate data from small to large sample four different values were considered (60, 100, 300 and 1000). It was indicated that sample size of 30 is small, while around 400 is large (Abbott, 2011, as cited in Demir et al., 2016). Later, percentages of the independent groups (25%, 50% or 75%) within the sample were changed and only one of the independent groups’ normality was altered as the second design factor. For the third design factor, non-normality was added to the selected group. For non-normality, five conditions were utilized. The conditions were choosen to represent normal to non-normal distributions. The non-normality values were summarized at Table 1. For example, under Sk=0, the skewness values were constrained to be between .00 and .10 while kurtosis values were between .00 and .20. For SK=2*SE group, maximum values of skewness and kurtosis were constrained to be smaller than 1.96 time of their standard errors. These values were considered to represent normal (Sk=0), non-normal (Sk=1) and severe non-normal (Sk=1.75) distributions.

Data generation procedure was different for one sample and independent sample tests. First, the procedure for independent sample test was described. Namely, data were generated to simulated one factor structure which was estimated by five items. The values of the factor loadings were adapted from Demirdağ and Kalafat (2015) and set to .70, .78, .87, .77 and .53. The loadings represent small (.53) to large (.87) values.

2.2. Data Generation Procedure

To simulate independent sample testing, first, normally distributed factor scores with mean of 0 and standard deviation of 1 was generated in R. Then, Fleishman's power transformation method (Fleishman, 1978) was used to get non-normal factor scores. This is one of the recognized method to simulate non-normality (Bendayan, Arnau, Blanca & Bono, 2014). Only one of the two independent groups was non-normal.

Table 1. *Skewness and Kurtosis Values Used for Data Generation*

Condition	Skewness		Kurtosis	
	Min	Max	Min	Max
Sk=0	.00	.10	.00	.20
Sk=2*SE	1.70*SES	1.96*SES	1.50*SEK	1.96*SEK
Sk=1	.90	1.00	.80	1.00
Sk=1.5	1.40	1.50	1.50	2.50
Sk=1.75	1.60	1.75	5.00	-

Sk: Skewness; SES: Standard Error of Skewness, SEK: Standard Error of Kurtosis

For example, for 25% of the sample (group 1) was non-normal and 75% of the data (group 2) was normal. That is, for the specified percent of total sample was non-normally distributed and the rest of sample was normal. To ensure this structure, first a normal distributed data set was generated for a given sample size. After getting a normally distributed data set another data set with non-normal distribution was generated. Later these two data sets were merged to get one data set in which the grouping variable was also available. Before saving the merged data set equal variance assumption was tested in R. If the assumption was satisfied the merged data sets were saved for independent sample tests. In total of 500 data sets were generated for each condition. Therefore, totally 30,000 (500*4*3*5) data sets were generated for independent sample tests.

For the dependent sample (one sample) test, the same factor structure was used. Fleishman's power transformation method was used to get non-normal factor scores. The simulated scores were considered as if they were score differences between pre-test and post-test results. For the dependent sample tests, only sample size and level of non-normality was used as design factors. The replication number was 500. Namely, 500 data sets were simulated for each of the given conditions. In total, 10,000 (500*4*5) data sets were generated for the dependent sample tests.

2.3. Data Analysis

The simulated data sets were also tested in R. To run the t-test and Mann-Whitney U test (U-test) *t.test* and *wilcox.test* functions were used. Type I error rates for both test was set to .05. In other words, significancies of the U-test and t-test were tested at the .05 alpaha level. For independent sample t-test equal variance was assumed since it was controlled within data generation process. Simulated data sets were analyzed under both t-test and U-test. For empirical studies only the *p*-values of the tests were used to decide about the null hypothesis. Therefore, only the *p*-values for the t-test and U-test were checked under this study too. Consequently, the numbers of t-test and U-test which showed the same result based on the *p*-values (significant or not significant) were counted. In other words, *p*-values larger than .05 and

smaller than .05 for both t-test and U-test were counted. These results showed how much of conclusion made on the null hypothesis were the same between t-test and U-test.

3. RESULT / FINDINGS

3.1. One Sample Test Results

Based on the simulation conditions given above, one sample test results were given below. Based on the results, skewness (i.e., non-normality) of the data has effect on t-test and U-test. Figure 1 shows the discrepancy between one sample t-test and Mann-Whitney U test. As the skewness of the data was increased the dissimilarity between the tests was increased. For example, when skewness was 1, under sample size of 100, t-test and U-test were given different results for 10% of the time. However, under the same condition when the skewness was increased to 1.5 the difference was increased to 30%.

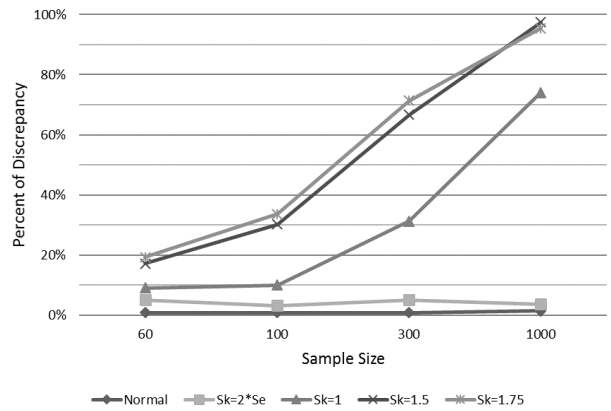

Figure 1. *Discrepancy between t-test and U-test for One-Sample tests*

The discrepancies were also dependent to sample sizes. As the sample size was increased the differences between t-test and U-test also increased for skewed data sets. For example, under the skewness of 1, when the sample size was increased from 100 to 300, the difference between the tests was increased from 10% to 31%.

When the data sets were normal the discrepancies between the tests were just about 1%. That is, when the data were normal, regardless of sample size, t-test and U-test gave the same results for 99% of the times. Figure 1 also shows the results for skewness equal to two times of its standard error (2*SES). Under this condition, the t-test and U-test were given the same results for 95% of the time on average. Table 2 gives the results of one sample tests in detail. For example, when sample size was 60 and skewness was 1.75 the discrepancy between t-test and U-test was 19%. As it is seen from the Table 2, for skewed data 2*SES rule gave the least discrepancies where the values were between 3 and 5 percents.

Table 2. *Discrepancy Values (%) between t-test and U-test for One Sample Tests*

Sample Size	Normal	Skewed Data			
		2*SES	1	1.5	1.75
60	1	5	9	17	19
100	1	3	10	30	34
300	1	5	31	67	71
1000	1	4	74	97	95

3.2. Independent Sample Test Results

Two independent groups were compared under this simulation study. Based on the results, sample size had an effect on the *p*-values for skewed data only as it was the case for one-sample test results. As the sample size was increased discrepancy between the *p*-values of tests also increased for skewed data. For example, under 25% of non-normal and skewness was 1, as the sample size was increased from 100 to 1000 the dissimilarity on the *p*-values increased from 4% to 20%. Left panel of Figure 2 shows the result for 25% of non-normal data while right panel shows the result for 50% (balanced) of non-normal data. Based on the results, under normally distributed data the *p*-values did not change much and the discrepancy was 2% at maximum. Thus, when the data were normal, regardless of sample size, t-test and U-test gave the same results for more than 98% of the times. Figure 2 also shows the results for skewness equal to two times of standard error of skewness (2*SES). Under this condition, the t-test and U-test gave the same results for more than 97% of the times in terms of *p*-values. Sample size did not affect the results under this condition. For example, as shown at left side of Figure 2, discrepancies for the *p*-values of the tests were about 3% for both sample sizes of 100 and 1000.

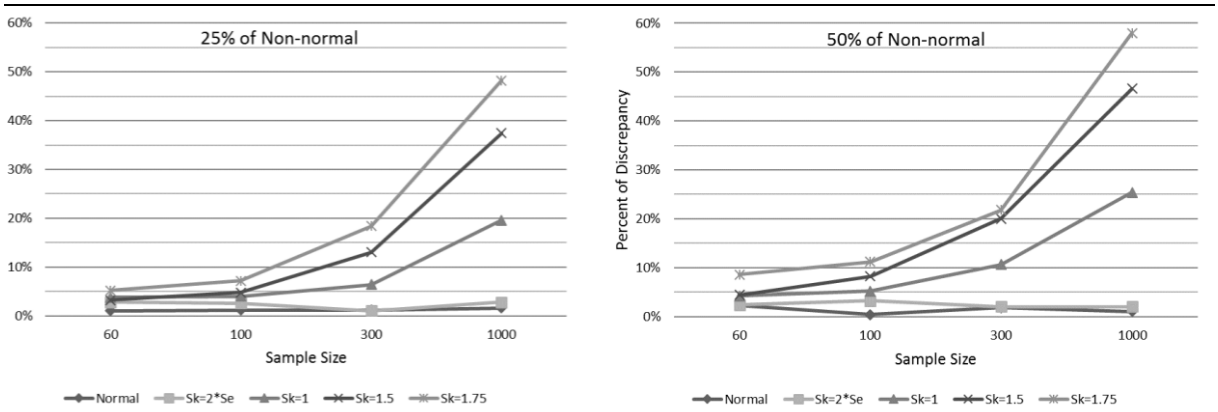

Figure 2. *Discrepancy between t-test and U-test for 25% and 50% (balanced) of non-normal data*

On the other hand, skewness also had effect on the *p*-values. As skewness was increased the difference between the *p*-values also increased. For example, on the left panel of Figure 2, as skewness was increased from 1 to 1.75 the difference between the *p*-values increased from 6% to 18%, under sample size of 300. Also, as sample size was increased the range of *p*-values also increased for skewed data. For example, the range was about 3% for sample size of 100 but 12% for 300 and 28% for 1000.

Percent of skewed data has also affected the results of t and U tests. Figure 3 shows the percent effects for sample sizes of 60 and 1000. When the sample size was small (60) the results of 25%, 50% and 75% non-normal data did not change much. Under these conditions, the discrepancies between the *p*-values were between 3% and 9%. However, as the sample size was increased, the effect of the percentages became more prominent. Interestingly, discrepancies between 25% and 75% of non-normality were similar. However, 50% of non-normality showed different and larger discrepancy as sample size was increased. On the other hand, when skewness was equal to two times its standard error (2*SES), percent of skewed data did not affect the results and discrepancies were between %1 and 3%. The results for independent tests were given at Table 3 in detail.

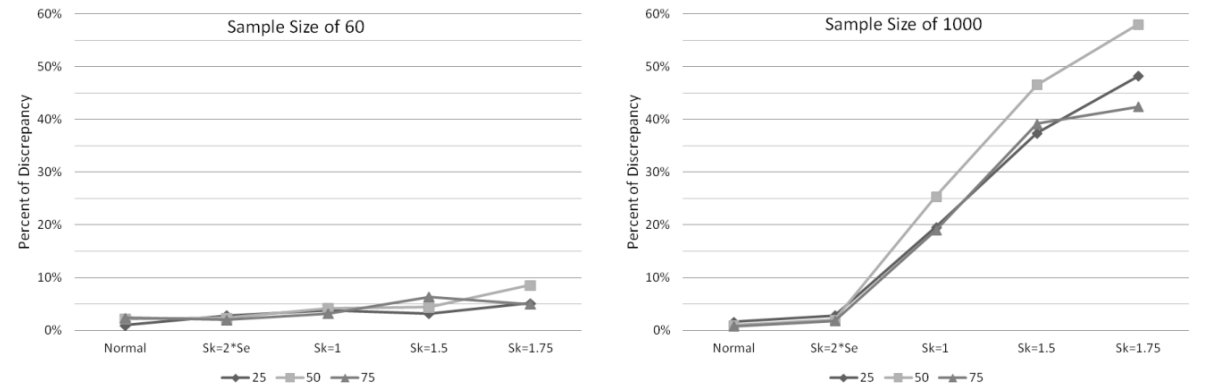

Figure 3. *Discrepancy between t-test and U-test for sample sizes of 60 and 1000*

Based on the results it was obvious that under skewed data sets t-test and U-test gave different results in terms of the *p*-values. The differences get clear as sample size and skewness of data were increased. However, under the 1.96 standard error rule, neither the sample size nor the percent of skewness were effective. Therefore, the results of this condition were investigated in detail.

Table 3. *Discrepancy Values (%) between t-test and U-test for Independent Sample Tests*

Sample Size	% of Skewness		Skewed Data			
		Normal	2*SES	1	1.5	1.75
60	25	1	3	4	3	5
	50	2	2	4	4	9
	75	2	2	3	6	5
100	25	1	3	4	5	7
	50	0	3	5	8	11
	75	1	2	2	7	13
300	25	1	1	6	13	18
	50	2	2	11	20	22
	75	1	3	8	15	15
1000	25	2	3	20	37	48
	50	1	2	25	47	58
	75	1	2	19	39	42

Table 4 shows average values of discrepancies between t-test and U-test with respect to SW tests. When the sample size was 60 about 92.8% of data was normal based on SW tests. Under this condition, when t-test was supposed to be used, 97.5% (90.5/92.8) of the U-test and when U-test was supposed to be used, 98.6% (7.1/7.2) of the t-test gave the same results. Even though SW tests results were different percent of similarities were alike across the sample sizes. For example, under sample size of 1000, when t-test was supposed to be used 97.5% (83.5/85.6) of the U-tests gave the same result in terms of p-values.

4. DISCUSSION and CONCLUSION

Checking the normality assumption is one of the critical steps for mean competition studies. Based on the results either parametric or non-parametric tests were considered to test mean differences. Literature suggests different approaches to check the assumption. Some of which are Kolmogorov-Smirnov test, Shapiro-Wilk test, checking skewness and kurtosis values or

basically looking the histogram of the dependent variables. Based on the test chosen the results of normality test might be different (Razali & Wah, 2011).

Table 4. *Discrepancy Average Discrepancy Values (%) for 2*SE Rule*

SW test Results		Sample Size			
		60	100	300	1000
Normal	Same	90.5	92.3	91.7	83.5
	Different	2.3	2.5	2.0	2.1
Non-normal	Same	7.1	5.0	6.3	14.3
	Different	.1	.1	.0	.1
Total of the Same		97.6	97.3	98.0	97.8

The use of skewness and kurtosis values to check normality is common in practice. Some suggest that the values can be up to as large as 2 in absolute values. On the other hand, standard errors of skewness and kurtosis were also used for normality tests. It was suggested that skewness and kurtosis values smaller than 1.96 times of their standard errors indicates normality (Kim, 2013; Field, 2009). However, there is no agreement on the values which indicate normality of a dataset. Therefore, this current study simulated different conditions to check the effect of skewness and kurtosis values on the decision made for mean comparison tests (a.k.a., t-test and U-test).

Based on the one-sample test results (see Table 2) when the data were normal or Sk < 1.96*SES, t-tests and U-tests were showed similar results with respect to *p*-values. Therefore, under these conditions, t-test can be used without any concerns. The results for normally distributed data were as expected. Nevertheless, under Sk < 1.96*SES condition, *p*-values of t-tests and U-tests were worth to point again. When skewness is smaller than its 1.96 standard error, t-tests and U-tests indicated the same results. Therefore, if Sk < 1.96*SES, t-tests can be used to test mean differences. However, when skewness is around 1 or larger, the t-tests and U-tests pointed different conclusions. Therefore, test of normality has to be considered carefully. There needs to be other evidences to show normality of data. If no evidence is found for normality and skewness is around or larger than 1, given the limitation of this study, U-tests should be used to test mean differences.

Similar results were obtained for two-sample tests as well. That is, when the data were normal or Sk < 1.96*SES, t-tests and U-tests were showed similar results with respect to *p*-values. Therefore, if Sk < 1.96*SES, t-tests can be used to test mean differences. However, if no other evidence found and skewness is around or larger than 1, U-tests should be used to test mean differences. This suggestion especially important for larger sample sizes. As the sample size was increased the effect of skewness become clear and the discrepancies between t-test and U-test increased.

On the other hand, a more detailed results for the 1.96*SE rule were given at Table 4. Based on the table, when SW test indicated that the data was normal, on average 97.6% of the t-test and U-test were the same in terms of *p*-values. Similarly, when SW test indicated that the data was not normal, on average 99.0% of the tests were the same in terms of *p*-values. Therefore, in order to use t-test for mean comparison the 1.96*SE rule can be used. Regardless of SW test results, if skewness and kurtosis of a given dataset are smaller than their 1.96 standard errors (about 2 standard errors), t-test can be preferred over U-test. However, based on the results of the simulation, when skewness and kurtosis of a given dataset are larger than 1 another proof to show normality (e.g., Shapiro-Wilk) is needed. Therefore, if no other proof is granted non-parametric U-test should be used for mean comparison. In other words, "skewness around and larger than 1" rules should not be used to decide between t-test and U-test.

For example, let's say that, a researcher wanted to test if there is difference on math achievement scores between male and female students. For this purpose, about 300 student's scores were collected in a data set. The researcher tested normality of the scores for each gender groups by Shapiro Wilk test. Let's say that, the test indicated that the data were non-normal. After the test, the researcher checked the skewness and kurtosis values. The values were about 1.5. Since the values were smaller than 2, the researcher decided to use the parametric test (e.g., t-test). In this case, there is 16% of chance (average of 13%, 20%, 15%) that the results of the t-test were different than U-test. Therefore, using only the skewness and kurtosis values to decide about the normality of a data set is too risky. That means that if only skewness and kurtosis values are used for normality it is possible that researchers may decide to use a wrong method to test their hypotheses. For example, they may decide to use t-test when U-test is supposed to be used. Regarding that, as far as this study showed, as skewness and sample size increased t-test and U-tests gave different conclusions in term of rejecting H_0. Therefore, it can be concluded that skewness and kurtosis values alone should not be used.

The literature also says that violation of normally assumption may not have serious effects on the results (Glass, Peckham, & Sanders, 1972, Blanca, Alarcon, Arnua, et al., 2017). However, uses of non-parametric tests are still very common in practice. Therefore, test of normally is still checked before mean comparison tests. The current study showed that the results changes based on the test chosen. The results of this study are limited with comparison of two means and predefined simulation conditions. Therefore, the results are limited to the conditions used within the study. For example, Ghasemi and Zahediasl (2012) and Kim (2013) and suggested the use of 2.58*SE or 3.29*SE rules under large sample size. Another study which simulated these conditions may also be useful. Under this study only the normality assumption was examined. Besides this, a simulation study where data are normal but equal variance assumption is violated can be informative as well.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

ORCID

Fatih Orcan http://orcid.org/0000-0003-1727-0456

5. REFERENCES

Abbott, M.L. (2011). *Understanding educational statistics using Microsoft Excel and SPSS*. United States: Wiley & Sons, Inc.

Altman, D.G. (1991). *Practical statistics for medical research*. London: Chapman and Hall

Bendayan, R., Arnau, J., Blanca, M.J. & Bono, R. (2014). Comparison of the procedures of Fleishman and Ramberg et al. for generating non-normal data in simulation studies. *Anales de Psicología, 30(1)*, 364-371. https://dx.doi.org/10.6018/analesps.30.1.135911

Bulmer, M. G. (1979). *Principles of statistics*. Mineola, New York: Dover Publications Inc.

Büyüköztürk, Ş., Çokluk, Ö. & Köklü, N. (2014). *Sosyal bilimler için istatistik* (15th Edition). Ankara: Pegem Akademik.

Blanca, M.J., Arnau, J., Lopez-Montiel, D., Bono, R. & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology, 9(2),* 78–84. https://dx.doi.org/10.1027/1614-2241/a000057

Blanca, M.J., Alarcon, R., Arnua, J., Bono, R. & Bendayan, R. (2017) Non-normal data: Is ANOVA still a valid option? *Psicothema, 29(4)*, 552-557. https://dx.doi.org/10.7334/psicothema2016.383

Boslaugh, S. & Watters, P.A. (2008). *Statistics in a nutshell*. Sebastopol, CA: O'REILLY.

Cain, M.K., Zhang, Z. & Yuan, K. (2017) Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behav Res, 49*, 1716–1735. https://dx.doi.org/0.3758/s13428-016-0814-1

Demir, E., Saatcioğlu, Ö. & İmrol, F. (2016). Uluslararası dergilerde yayımlanan eğitim araştırmalarının normallik varsayımları açısından incelenmesi, *Current Research in Education, 2*(3), 130-148. Retrieved from https://dergipark.org.tr/tr/pub/crd/issue/28292/300531

Demirdağ, S., & Kalafat, S. (2015). Meaning in life questionnaire (MLQ): The study of adaptation to Turkish, validity, and reliability. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *16(2),* 83-95. https://dx.doi.org/10.17679/iuefd.16250801

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd Edition). London: SAGE Publications Ltd

Fleishman, A.I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43,* 521-532. https://dx.doi.org/10.1007/BF02293811

Ghasemi, A. & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *Int J Endocrinology & Metabolism, 10*(2), 486-489. https://dx.doi.org/10.5812/ijem.3505

Glass, G., Peckham, P. & Sanders, J. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Measurement, 42,* 237-288.

Huck, S.W. (2012). *Reading statistics and research* (6th Edition). Boston, MA: Pearson

Iyer, D.N., Sharp, B.M. & Brush, T.H. (2017). Knowledge creation and innovation performance: An exploration of competing perspectives on organizational systems. *Universal Journal of Management, 5(6),* 261-270. https://dx.doi.org/10.13189/ujm.2017.050601

Kim, H. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Open lecture on statistics (NA),* 52-54. https://dx.doi.org/10.5395/rde.2013.38.1.52

Lei, M. & Lomax, R.G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling, 12(1)*, 1-27. https://dx.doi.org/10.1207/s15328007sem1201_1

Miot, H.A. (2016). Assessing normality of data in clinical and experimental trials. *Jornal Vascular Brasileiro 16(2)* 88-91. https://dx.doi.org/10.1590/1677-5449.041117

Orçan, F. (2020). Sosyal bilimlerde istatistik SPSS ve Excel uygulamaları (1st Edition). Ankara: Anı Yayıncılık.

Park, H.M. (2008). *Univariate analysis and normality test using sas, stata, and spss*. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University

Perry, J.L., Dempster, M. & McKay, M.T. (2017) Academic self-efficacy partially mediates the relationship between scottish index of multiple deprivation and composite attainment score. *Frontiers in Psychology, (8)*, NA. https://dx.doi.org/10.3389/fpsyg.2017.01899

Razali N.M. & Wah, Y.B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, *Journal of Statistical Modeling and Analytics, 2*(1), 21-33. Retrieved from: https://www.researchgate.net/publication/267205556

Rachon, J., Gordan, M. & Kieser, M. (2012). To test or not to test: Preliminary assessment of normality when comparing two independent samples, *BMC Medical Research Methodology, (12),*81. https://dx.doi.org/10.1186/1471-2288-12-81

Ramos, C., Costa, P.A., Rudnicki, T., et al. (2018). The effectiveness of a group intervention to facilitate posttraumatic growth among women with breast cancer. *Psycho-Oncology, (27)*, 258–264. https://dx.doi.org/10.1002/pon.4501

Rietveld, T. & van Hout, R. (2015). The t test and beyond: Recommendations for testing the central tendencies of two independent samples in research on speech, language and hearing pathology. *Journal of Communication Disorders, (58)*, 158-168. https://dx.doi.org/10.1016/j.jcomdis.2015.08.002

Schucany, W.R. & Tony N.G., H.K. (2006). Preliminary goodness-of-fit tests for normality do not validata the one-sample student t. *Communications in Statistics – Theory and Methods, 35*, 2275-2286. https://dx.doi.org/10.1080/03610920600853308

Şirin, Y.E., Aydın, Ö. & Bilir, F.P. (2018). Transformational-transactional leadership and organizational cynicism perception: physical education and sport teachers sample. *Universal Journal of Educational Research, 6(9*), 2008-2018. https://dx.doi.org/10.13189/ujer.2018.060920

West, S.G., Finch, J.F. & Curran, P.J. (1995). Structural equation models with nonnormal variables: problems and remedies. In RH Hoyle (Ed.). Structural equation modeling: Concepts, issues and applications. Newbery Park, CA: SAGE.