Hakemli - Süreli - Bilimsel Dergi
İNTERNET UYGULAMALARI ve YÖNETİMİ
Journal of
INTERNET APPLICATIONS and MANAGEMENT

2023 / 14(2)

What should be the measure of conformity to normal distribution (normality) test in Likert type digital and face-to-face survey data?

Likert tipi dijital ve yüz yüze anket verilerinde normal dağılıma uygunluk (normallik) testi ölçüsü ne olmalıdır?

Selçuk Burak HASILOGLU[1], selcukburak@hasiloglu.com

Melda HASILOGLU-CIFTCILER[2], hasilogm@tcd.ie

Received: 19.08.2023; **Accepted**: 23.10.2023 **DOI**: 10.34231/iuyd.1346463

In statistics, it is assumed that the data suitable for parametric tests exhibit a normal distribution. In other words, before running parametric tests, it is essential to check whether the data set follows a normal distribution. Parametric tests can only be applied to data that adhere to a normal distribution. The Shapiro-Wilk and Kolmogorov-Smirnov (Lilliefors modification) tests are among the most commonly utilised methods for testing normality. However, these tests were originally developed for rational numbers. The use of these tests for Likert-type digital or face-to-face survey data has always been a topic of discussion. Even if the measurement tools consist of at least three items, the data sets are dominated by repeated values. Again, there are different opinions on the acceptable range for kurtosis and skewness values. With this study, it was determined that the Kolmogorov-Smirnov (Lilliefors modification) test is not suitable for testing the normality of Likert-type digital or face-to-face survey data. In addition, the coefficients of kurtosis and skewness within the range of (-1, +1) range are acceptable for normality. For this reason, it is recommended not to use goodness-of-fit tests such as Shapiro-Wilk or Kolmogorov-Smirnov for testing the normality of interval scale data collected through questionnaires.

Keywords: *Normal distribution, Normality test, Kurtosis Skewness, Shapiro-Wilk test, Kolmogorov-Smirnov test, Lilliefors modification (Lilliefors Significance Correction), Goodness of fit, Parametric tests, Likert type scale, Digital questionnaire, Face-to-face questionnaire*

İstatistikte, parametrik testlere uygun verilerin normal dağılım sergilediği kabul edilir. Başka bir ifade ile parametrik testler yapabilmek için öncelikli olarak veri setinin normal dağılıma uygun olup olmadığına bakılır. Normal dağılıma uygun olan verilerle parametrik testler yapılabilir. Shapiro-Wilk ve Kolmogorov-Smirnov (Lilliefors modifikasyonu), normallik testleri arasında en çok kullanılanlardır. Ancak bu testler rasyonel sayılar temel alınarak geliştirilmiştir. Likert tipi dijital veya yüz yüze anket verileri için bu testlerin kullanımı her zaman bir tartışma konusu olmuştur. Ölçme araçları en az üç maddeden meydana geliyor olsa dahi veri setlerinde tekrarlı değerler hakimdir. Yine basıklık çarpıklık değerleri için kabul edilebilir uygun aralık üzerine de farklı görüşler bulunmaktadır. Bu çalışma ile Kolmogorov-Smirnov (Lilliefors modifikasyonu) testinin Likert tipi dijital veya yüz yüze anket verilerinin normallik testi için uygun olmadığı tespit edilmiştir. Ayrıca (-1, +1) aralığındaki basıklık ve çarpıklık katsayıları, normallik için kabul edilebilirdir. Bu nedenle anket yolu ile toplanan aralıklı ölçek verileri normallik testi için Shapiro-Wilk veya Kolmogorov-Smirnov gibi uyum iyiliklerinin kullanılmaması önerilmektedir.

Anahtar Kelimeler: *Normal dağılım, Normallik testi, Basıklık ve Çarpıklık, Shapiro-Wilk testi, Kolmogorov-Smirnov testi, Lilliefors modifikasyonu, Uyum iyiliği, Parametrik testler, Likert tipi ölçek, Dijital anket, Yüz yüze anket*

[1] Professor, Management Information Systems, Pamukkale University
[2] PhD Canditade; 1) Industrial Engineering, Eskisehir Technical University
2) Trinity College Dublin, The University of Dublin

1. INTRODUCTION

Visualise the trees and their sizes in a centuries-old, untouched pine forest. They are not all the same height or size, of course. While the majority are almost the same height, there are a few very short or very tall trees. With this observation, those that are many will be normal, while those that are few will be abnormal. When we rank the trees from the largest to the smallest, the tree in the middle will be considered as the most normal tree height that is most common in the forest. When we measure the height of all trees and take the average, the tree closest to the height we obtain will correspond to the height of the tree we define as the most normal. In this way, the heights of the trees in a forest have a normal distribution. We can see a distribution close to the normal distribution in many natural event data. This distribution can be seen not only in natural measurement data such as weight, length, temperature, etc., but also in data such as attitudes, beliefs and behaviours of people. This approach forms the cornerstones of statistics based on probability and hypothesis, including variance and averages.

Parametric tests such as T-test, ANOVA (ANalysis Of Variance), MANOVA (Multivariate ANalysis Of Variance), ANCOVA (ANalysis of COVAriance) and Pearson Correlation Analysis are used to analyse normally distributed data. Researchers use normality tests to measure the conformity of the data to the normal distribution. W/S test, Jarque-Bera test, Moment test, Shapiro-Wilk test, D'Agostino test, Cramer-von Mises test, Anderson-Darling test, or Kolmogorov-Smirnov test are known as normality tests. However, these tests have developed based on rational numbers called ratio data. The use of these tests for Likert-type digital or face-to-face survey data has always been a matter of debate.

Likert scale is an interval scale type introduced to the literature by Rensis Likert in 1932. This study, which was a part of Likert's doctoral thesis, provides a technique for the measurement of attitudes. In his study, Likert, a US social psychologist, aimed to measure attitudes such as internationalism and imperialism by using items and equally spaced 5 response alternatives (5-point Likert scale: "strongly approve, approve, undecided, disapprove and strongly disapprove"). Surveys based on this principle (Likert-type scale) are still used today. This type of survey, which started with face-to-face surveys, is gradually losing its popularity to digital (online surveys). Face-to-face surveys are used on the road, on the street, in the bazaar, in the market, and even during shopping. Researchers first recognised digital surveys (online surveys) as a digitalisation of the data collection process. However, according to Haşıloğlu (2022: 114-116), digital surveys are not an alternative to traditional surveys and this distinction will be more clearly recognised in the future. The most important advantages are as follows: Hasiloglu (2022:116)

- It is ideal for studies where different questions need to be brought automatically according to the answer given.
- It can be easily and quickly stored and evaluated with the database created.
- It reduces research costs in terms of reaching respondents, collecting, storing and evaluating data.
- Respondents' response time and even item-by-item response times can be learnt more easily than traditional media applications.
- Showing what the next question is can be avoided.

- Inattentive or reluctant respondents can be dynamically detected by using control questions. They can be asked to return when necessary.
- Experimental questionnaires can be faster and easier.
- The applicability of graphic scale is possible.
- Surveys for digital products can be very easy to realise.

The Kolmogorov-Smirnov (Lilliefors modification) test, which is used in the normality tests of the data obtained with Likert-type scales, does not give appropriate results for researchers. Another test used for the normality test is the kurtosis-skewness measure. This measure is more suitable for testing data collected with Likert-type scales. However, there are different opinions about the acceptable range for kurtosis-skewness values.

2. NORMAL DISTRIBUTION AND GOODNESS OF FIT TESTS

Normal distribution function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

is in the form. Here

μ: arithmetic mean of the data set and

σ: standard deviation of the data set

it is expressed as. The normal distribution probability function corresponds to the area between the curve and the horizontal axis. Naturally, an integral is needed to calculate this area. In addition, the area formed by the standard normal distribution function for all values that x can take is equal to 1. This means that,

$$F_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}} \cdot dx = 1$$

If F(x) is 1, it means that the probability is 100% realised. When two values such as a and b are used in the integral instead of -∞, +∞, the probability value in this range is calculated. However, it will not be enough to use pen and paper to calculate the integral of this function. Although there are many package programs that can solve such problems, Microsoft Excel is the most practical among them. For example, let us assume that the distribution of trouser waist sizes of the customers of a store is normally distributed. If the mean waist size is 78 cm and the standard deviation is 19 cm, the probability that the waist size of a customer entering the store door is less than 85 cm is 64,37%. If 556 new trousers are to be produced for this store, 19,43% of these trousers, i.e. a total of 108 trousers, should have waist sizes in the range of (66 cm ; 76 cm]. (Hasiloglu, 2022: 214-227)

To analyse the conformity of the data to normal distribution, methods such as W/S test, Jarque-Bera test, Moment test, Shapiro-Wilk test, Zhang Test, D'Agostino test, Cramer-von Mises test, Anderson-Darling test or Kolmogorov-Smirnov test are used. Among these, especially Kolmogorov-Smirnov (Lilliefors modification: Lilliefors Significance Correction) and Shapiro-Wilk tests are easily accessible in IBM SPSS (*SPSS*→ *Analysis*→ *Descriptive Statistics*→ *Explore…*→ *Plans…*→ *Normality plots with tests*), so they are the most frequently used measures among the users of this package programme in recent years. However, the suitability of these two tests for the normality analysis of Likert-type scale data is a matter of debate. The kurtosis and skewness values of the data can give more meaningful results in examining the conformity to the normal distribution (Hair et al., 2019: 48; George & Mallery, 2019: 114-115).

Shapiro-Wilk Test

In 1965, it was developed by Shapiro and Wilk. The authors presented the method in detail in their paper (Shapiro & Wilk, 1965). The general term of the test statistic

$$W = \frac{(\sum_i^n a_i \cdot y_i)^2}{\sum_i^n (y_i - \bar{y})^2}$$

is of the form. Here y_i's represent the test data sorted from smallest to largest. The a_i's in the general term are drawn from the table values given in the paper. These values are listed separately in the table according to the sample size (n) and conform to the normal distribution curve. Therefore, the W value reveals the relationship between the test data y_i and the data a_i, which are in accordance with the normal distribution. As W approaches 1, the strength increases between the test data and the ai data representing the normal distribution vector. The paper also provides a table describing the power of W with respect to sample size. To prove that the test data is normally distributed, the Shapiro-Wilk p statistic must be greater than 0,05.

There is evidence that the Shapiro-Wilk test is more significant/powerful than the Kolmogorov-Smirnov test (Mendes & Pala, 2003; Keskin, 2006; Razali & Wah, 2011). However, this test was developed for small sample ($n \leq 50$) data (Shapiro & Wilk, 1965).

Kolmogorov-Smirnov (Lilliefors modification) Test

This test, which was first developed by A.N. Kolmogorov in 1933 and then by N.V. Smirnov in 1939, is referred to as Kolmogorov-Smirnov since they are close methods. The general term of this test statistic, which is based on a single sample, is the largest of the absolute differences of the cumulative relative frequency of the theoretical (assumed) distribution and the sample frequency value,

$$D = max|F_n(X) - F(X)|$$

is expressed in the form. After 28 years, with the modification made by Lilliefors (1967), this method became widely used. Using the Monte Carlo method, Lilliefors (1967) presented a table to be used with the Kolmogorov-Smirnov statistic when the mean and variance are not specified and need to be estimated from the sample. In this table of critical values, the value of D depends on the sample size. For sample size (n) above 30;

$$D = \frac{886}{\sqrt{n}} \cdots p \cong 0{,}05$$

$$D = \frac{1031}{\sqrt{n}} \cdots p \cong 0{,}01$$

is; where is the Kolmogorov-Smirnov (Lilliefors modification) test statistic,

$$D = max|F^*(X) - S_n(X)|$$

is in the form. The Kolmogorov-Smirnov p statistic must be greater than 0,05 to prove that the test data is normally distributed.

Since digital or face-to-face surveys collect Likert-type interval scale data, the Kolmogorov-Smirnov statistic for normality testing is a matter of debate. This study proved that the Kolmogorov-Smirnov (Lilliefors modification) test is not sufficient for normal distribution analysis.

Kurtosis Skewness Coefficients

The kurtosis coefficient is a measure of the sharpness or flatness of the curve formed by the data set. The skewness coefficient is an indicator of whether the curve is skewed to the right or left. Formulas,

$$Kurtosis = \frac{\sum_i^n (x_i - \bar{x})^4}{n \cdot s^4} - 3$$
$$Skewness = \frac{\sum_i^n (x_i - \bar{x})^3}{n \cdot s^3}$$

is in the form. As both coefficient values approach zero, the degree of normality will increase. When the skewness coefficient value is zero, the curve is perfectly symmetric; that is, the mean, mode and median values are equal. There is no constant -3 in the original kurtosis formula (Newbold, 1988: 420); however, as in the case of skewness, this constant was added to the formula so that zero indicates complete normality. Hair et al. (2019:48) suggested that if the skewness coefficient is outside the range [-1, +1], the data set indicates a skewed distribution. George & Mallery (2019: 114-115), on the other hand, have a similar opinion, but stated that depending on certain applications, the range [-2, +2] may be acceptable. This study proved that the test data does not exhibit a normal distribution when the skewness coefficient is outside (-1, +1).

In addition to these, there are also studies that use [-3, +3] values as the appropriate range by citing references (Groeneveld & Meeden, 1984; Moors, 1986; Hopkins & Weeks, 1990; De Carlo, 1997). Unfortunately, there is no such evidence in any of the mentioned references. In summary, there is no study with evidence that values in the [-3, +3] range of kurtosis skewness coefficients can be sufficient for a normal distribution.

NORMALITY FOR INTERVAL (1-5 LİKERT TYPE) SCALE DATA

The subject of measurement-based data type dates back to the 1940s. Naturally, there have been minor changes in the types of scales and accordingly in their characteristics since then until today. Since this data type consists of equal intervals, it is also considered as a measurement tool. Interval scales are a type of measurement which are frequently used in marketing research. The rating starts from 1 and continues with 1 increment. For this reason, the most important feature of the interval scale is that it is graded, and the intervals are equally ordered in this grading (Hasiloglu; 2022: 147-148). The ideal number of answer alternatives varies according to the research topic, model, and the characteristics of the sample, and 1-5 interval (Likert type) scale type is preferred by researchers (Kalburan et al., 2019). Since interval scale data are collected with Likert-type scales, even if the measurement tool consists of at least three items, repetitive values dominate in the data sets. The use of normality tests based on rational numbers, such as the Shapiro-Wilk test or the Kolmogorov-Smirnov test, to analyse the conformity of a data set with such characteristics and repeated values to a normal distribution has been a matter of debate. There are also different opinions on the acceptable range for kurtosis-skewness values.

Inadequacy of the Kolmogorov-Smirnov Test

Firstly, the adequacy of the Kolmogorov-Smirnov (Lilliefors modification: Lilliefors Significance Correction) test for the analysis of conformity to normal distribution in the data collected with an interval scale such as Likert was examined. For this purpose, three data sets were produced as survey data and suitable for normal distribution. The first data set (Z_i) consists of single-item 1-5 interval scale (integers in the form of 1,2,3,4,5) data. The second data set (R_i) consists of values representing the average of the 1-5 interval scale with at least three items. Both are dominated by repeated values. In addition, the third data set ($R_i \pm \varepsilon$), which was developed for control purposes, was generated to be ±0.09 away from each value in the second data set (ε<0.09; random value) to minimize the repetitive value. Table 1 shows the frequencies of the values in these three data sets produced in accordance with the normal distribution.

Table 1. Frequencies of the values of the data sets

Data set-1		**Data set-2**		**Data set-3**	
Z_i	**f**	**R_i**	**f**	**R_i ±ε (ε<0,09)**	**f**
1	24	1	5	1 + ε	5
2	120	1,2	8	1,2 ± ε	8
3	192	1,4	11	1,4 ± ε	11
4	120	1,6	15	1,6 ± ε	15
5	24	1,8	19	1,8 ± ε	19
Total	**480**	2	24	2 ± ε	24
		2,2	29	2,2 ± ε	29
		2,4	33	2,4 ± ε	33
		2,6	37	2,6 ± ε	37
		2,8	39	2,8 ± ε	39
		3	40	3 ± ε	40
		3,2	39	3,2 ± ε	39
		3,4	37	3,4 ± ε	37
		3,6	33	3,6 ± ε	33
		3,8	29	3,8 ± ε	29
		4	24	4 ± ε	24
		4,2	19	4,2 ± ε	19
		4,4	15	4,4 ± ε	15
		4,6	11	4,6 ± ε	11
		4,8	8	4,8 ± ε	8
		5	5	5 - ε	5
		Total	**480**	**Total**	**480**

According to the Kolmogorov-Smirnov and Shapiro-Wilk tests, the single-item 1-5 interval scale sample data set (Data set-1: Z_i) and the data set representing the 1-5 interval scale mean with at least three items (Data set-2: R_i) do not exhibit normal distribution (Table 2). However, as can be seen in Figure 1, the data sets exhibit a normal distribution. Therefore, Kolmogorov-Smirnov (Lilliefors modification) is not an appropriate method for normality tests of single or at least three-item data sets with repetitive values collected with Likert-type scale.

Table 2. Kolmogorov-Smirnov and Shapiro-Wilk test results of the data sets

	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**			**Skewness**	**Kurtosis**
	D	df	p	W	df	p		
Data Set-1: Zi	0,200	480	<0,001	0,903	480	<0,001	0	-0,399
Data Set-2: Ri	0,049	480	0,007	0,988	480	0,001	0	-0,576
Data Set-3: Ri_Epsln	0,025	480	0,200	0,992	480	0,008	0,0005	-0,601

[a] Lilliefors Significance Correction

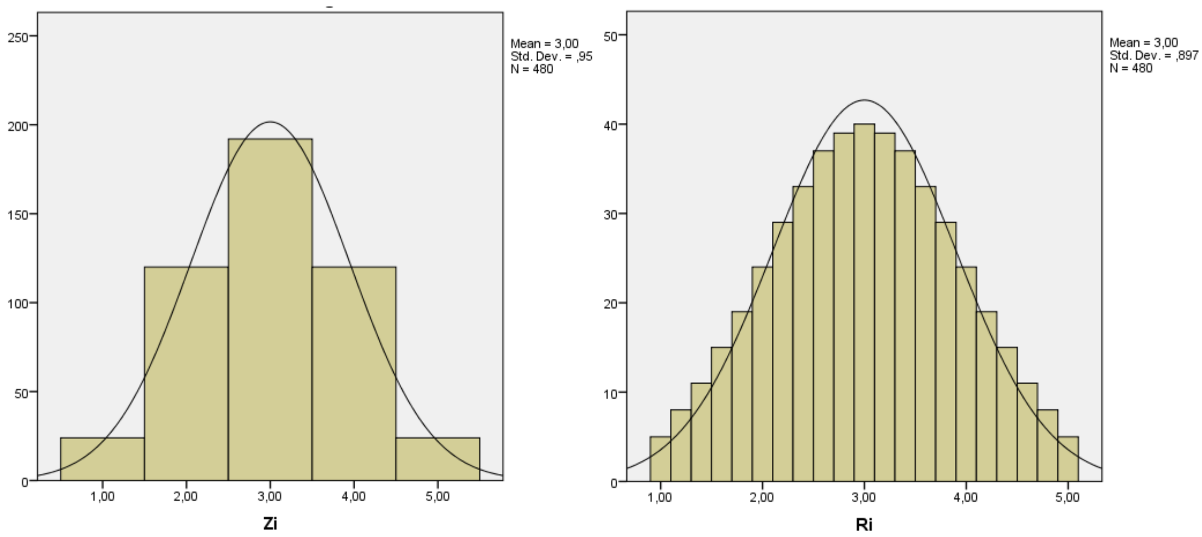

Figure 1. Histogram graphs of Zi and Ri data sets

On the other hand, since there are no repeated values in the Data Set-3: R_i_Epsilon data set, although these data are normally distributed according to the Kolmogorov-Smirnov test, they are not normally distributed according to the Shapiro-Wilk test. In order to avoid this contradiction arising from the sample size, the sample size is reduced to 32. This data set, selected by drawing data with 15 systematic intervals, is normally distributed according to both Kolmogorov-Smirnov test (p=0,200) and Shapiro-Wilk test (p=0,717).

Acceptable Skewness Coefficient Value Range

In the second stage of the study, the adequacy of the skewness coefficient value of the data compatible with the Likert-type (1-5 interval) scale outside the range (-1, +1) was examined for its conformity to the normal distribution.

Table 3 shows the frequencies and skewness coefficient of the 1-5 normalization values of the data compatible with the standard normal distribution. Since the data are fully symmetrically normally distributed, it is an expected result that the skewness coefficient is 0.

Table 3. Yi ∈ [1, 5] normalized values and skewness coefficient

Xi∈ [-3, 0]	**Yi ∈[1, 3]**	**f**		**Xi∈ (0, 3]**	**Yi ∈(3, 5]**	**f**
-3	1	1		0,25	3,17	38
-2,75	1,17	2		0,5	3,33	35
-2,5	1,33	3		0,75	3,50	30
-2,25	1,50	4		1	3,67	24
-2	1,67	5		1,25	3,83	18
-1,75	1,83	8		1,5	4	13
-1,5	2	13		1,75	4,17	8
-1,25	2,17	18		2	4,33	5
-1	2,33	24		2,25	4,50	4
-0,75	2,50	30		2,5	4,67	3
-0,5	2,67	35		2,75	4,83	2
-0,25	2,83	38		3	5	1
0	3	40		**Total**		**402**
				Skewness		**0**
				Kurtosis		**0,07**

When only Y_i ∈ [1, 3] data of Table 3 is used, only the left side of the vertically bisected curve will remain. We cannot say that this increasing curve exhibits a normal distribution (Figure 2). However, the skewness coefficient for the Y_i ∈ [1, 3] range is -1.014. The skewness coefficient for the range [1, 3,17] is -0,922 and the coefficient approaches zero as the range increases. Therefore, it can be said that for this sample data set, the skewness coefficient value outside the range (-1, +1) is not sufficient for the analysis of conformity to normal distribution.

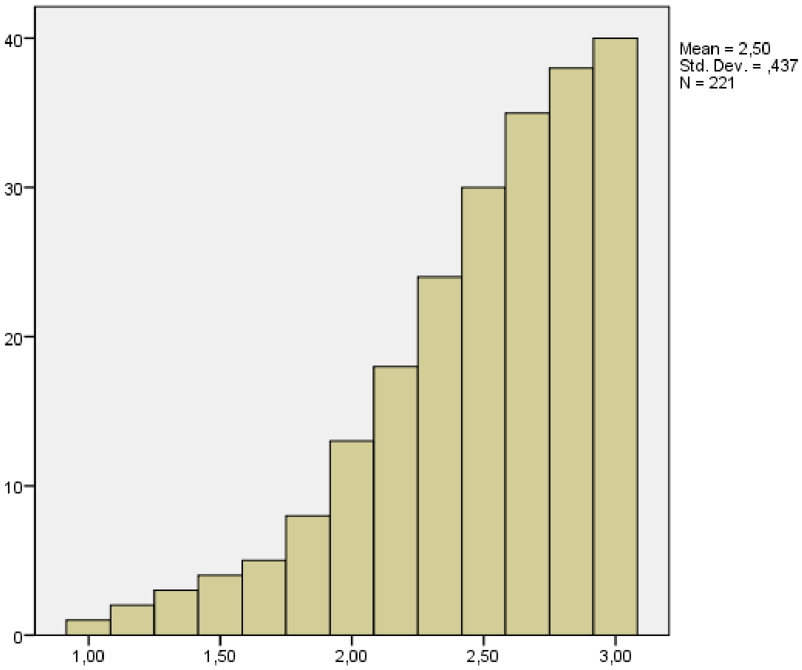

Figure 2. Yi ∈[1, 3] data plot not conforming to normal distribution (skewness= -1,014)

CONCLUSION

Questionnaires are the most intensively used data collection tool by social scientists since they contain scales. The Likert-type scale, which forms the basis of the scales, dates back to the 1930s. It is a measurement tool for factors such as attitudes, intentions, tendencies and beliefs of many social and educational scientists. This type of scale consists of equal intervals and 5 response alternatives are frequently used by researchers. Although the mean of at least three items is used for measurement, the collected answers are weighted with repeated data. Depending on all these, the use of goodness of fit for rational numbers in the normality test of the data collected through questionnaires has always been a matter of debate. Among these, Shapiro-Wilk and Kolmogorov-Smirnov tests have been the most preferred ones. In this study, it was concluded that both methods were not suitable for the survey data. Therefore, researchers should not resort to Shapiro-Wilk and Kolmogorov-Smirnov (Lilliefors modification) tests to prove the suitability of digital or face-to-face survey data based on Likert-type scale for normal distribution.

As an alternative to Shapiro-Wilk or Kolmogorov-Smirnov tests, some researchers suggest looking at the kurtosis-skewness coefficients. However, there are different opinions about the acceptable range of skewness coefficient values. This study proved that it is not possible to say that the data set belonging to the skewness coefficient outside the range (-1, +1) exhibits a normal distribution. In addition, when the equal numbers of 1, 2, 3, 4 and 5 values are used to represent the 5-point Likert-type scale, the kurtosis coefficient reaches to the -1.3 limit. When the equal numbers of rational numbers in the range of 1-5 are used, it can reach -1 limit. Researchers are advised not to rely on the information that the range [-3, +3] is acceptable. Because there is never such a proven study in the literature. References (Groeneveld and Meeden, 1984; Moors, 1986; Hopkins and Weeks, 1990; De Carlo, 1997) are not realistic.

REFERENCES

De Carlo, L.T. (1997). "On the meaning and use of kurtosis". *Psychological Methods*, *2*, 292-307

George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Fifteenth edition, Routledge.

Groeneveld, R. A., & Meeden, G. (1984). "Measuring Skewness and Kurtosis". *Journal of the Royal Statistical Society. Series D (The Statistician)*, *33*(4), 391–399.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2019). *Multivariate data analysis, Eighth Edition. Cengage Learning EMEA*.

Haşıloğlu, S.B. (2022). *Pazarlama Araştırması ve Analitiği*, Ankara: Nobel Yayınevi

Hopkins, K.D. & Weeks, D.L. (1990). "Tests for normality and measures of skewness and kurtosis: Their place in research reporting". *Educational and Psychological Measurement,* 50, 717-729.

Kalburan, C., Hasiloglu, S.B., & Bardakci, A. (2019). "Does a difference in the number of response categories change the results for ACSI in the mobile phone sector?". *International Journal of Mobile Communications*, 17(6), 746-759.

Keskin, S. (2006). "Comparison of several univariate normality tests regarding type I error rate and power of the test in simulation based small samples". *Journal of Applied Science Research*, *2*(5), 296-300.

Lilliefors, H.W. (19767). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown". *Journal of the American Statistical Association,* 62(318), 399-402.

Mendes, M., & Pala, A. (2003). "Type I error rate and power of three normality tests". *Pakistan Journal of Information and Technology*, *2*(2), 135-139.

Moors, J. J. A. (1986). "The meaning of kurtosis: darlington reexamined". *The American Statistician,* 40, 283-284.

Newbold, P. (1988). *Statistics for Business and Economics*. Prentice Hall Inc., USA.

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.

Shapiro, S. S., & Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*, *52*(3/4), 591-611.

Likert tipi dijital ve yüz yüze anket verilerinde normal dağılıma uygunluk (normallik) testi ölçüsü ne olmalıdır?

(GENİŞLETİLMİŞ TÜRKÇE ÖZET METİN)

1. GİRİŞ

T-testi, ANOVA (ANalysis Of VAriance: Varyans Analizi), MANOVA (Multivariate ANalysis Of VAriance: Çok değişkenli Varyans Analizi), ANCOVA (ANalysis of COVAriance: Kovaryans Analizi) ve Pearson Korelasyon Analizi gibi testlerin geneli için adlandırılan parametrik testler, normal dağılıma uygun verilerin analizinde kullanılır.

Araştırmacılar, verinin normal dağılıma uygunluğunu ölçmek için normallik testine başvururlar. W/S testi, Jarque-Bera testi, Moment testi, Shapiro-Wilk testi, D'Agostino testi, Cramer-von Mises testi, Anderson-Darling testi veya Kolmogorov-Smirnov testi bilinen normallik testlerindendir. Ancak bu testler genel olarak oran verisi olarak adlandırılan rasyonel sayılar temel alınarak geliştirilmiştir. Likert tipi dijital veya yüz yüze anket verileri için bu testlerin kullanımı her zaman bir tartışma konusu olmuştur.

Likert ölçeği, Rensis Likert tarafından 1932 yılında literatüre kazandırılan aralıklı ölçek türüdür. Likert'in doktora tezinin bir parçası olan bu çalışma, tutumların ölçümü için bir teknik sunmaktadır. ABD'li sosyal psikolog Likert, çalışmasında enternasyonalizm ve emperyalizm gibi tutumları, maddeler ve eşit aralıklı 5 cevap alternatifli seçenekler ("kesinlikle onaylıyorum, onaylıyorum, kararsızım, onaylamıyorum ve kesinlikle onaylamıyorum" şeklinde 5'li Likert) kullanarak ölçmeyi amaçlamıştır. Bugün hala bu prensibe dayalı (Likert tipi ölçek) anketler kullanılmaktadır. Yüz yüze ile başlayan bu anket türü popüleritesini adım adım dijitale (çevrimiçi/online anket) bırakmaktadır. Yüz yüze anketler, yolda, sokakta, çarşı pazarda ve hatta alışveriş sırasında uygulan bir türdür. Araştırmacılar, ilk olarak, dijital anketleri (online anketler) veri toplama sürecinin dijital platforma taşınmış hali şeklinde tanımıştır. Ancak Haşıloğlu (2022:114-116)'na göre dijital ortam anketleri, geleneksel ortam anketlerinin alternatifi değildir ve gelecekte bu ayrım çok daha net bir şekilde fark edilecektir. En önemli avantajları şunlardır: Haşıloğlu (2022:116)

- Verilen cevaba göre farklı soruların otomatik getirilmesine ihtiyaç duyulan çalışmalar için idealdir.
- Kolaylıkla ve hızlıca depolanabilir ve oluşturulan veri tabanı ile değerlendirilebilir.
- Cevaplayıcılara ulaşma, verileri toplama, depolanma ve değerlendirilmesi açısından araştırma maliyetlerini düşürmektedir.
- Cevaplayıcıların anketi cevaplama süresi ve hatta madde madde cevaplama süreleri, geleneksel ortam uygulamalarına göre daha kolay öğrenilebilir.
- Bir sonraki sorunun ne olduğunun gösterilmesinin önüne geçilebilir.
- Kontrol soruları kullanılarak dikkatsiz veya isteksiz cevaplayıcılar dinamik olarak tespit edilebilir. Gerektiğinde geri dönmesi istenebilir.

- Deneysel anketler daha hızlı ve kolay olabilir.
- Grafik ölçeğin uygulanabilirliği mümkündür.
- Dijital ürünlere yönelik anketlerin gerçekleşmesi çok da kolay olabilir.

Likert tipi ölçeklerle elde edilen verilerin normallik testlerinde kullanılan, özellikle Kolmogorov-Smirnov (Lilliefors modifikasyonu) testi, araştırmacılar için uygun sonuçlar vermemektedir. Normallik testi için baş vurulan bir diğer test ise basıklık çarpıklık ölçüleridir. Bu ölçü, Likert tipi ölçeklerle toplanan verilerin testi için daha uygundur. Ancak basıklık çarpıklık değerleri için kabul edilebilir uygun aralık konusunda farklı görüşler bulunmaktadır.

2. NORMAL DAĞILIM VE UYUM İYİLİĞİ TESTLERİ

Normal dağılım fonksiyonu

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

şeklindedir. Burada

μ: veri setinin aritmetik ortalaması ve

σ: veri setinin standart sapması

olarak ifade edilir. Normal dağılım olasılık fonksiyonu eğri ile yatay eksen arasındaki kalan alana karşılık gelmektedir.

Verilerin normal dağılıma uygunluğunun analizi için W/S testi, Jarque-Bera testi, Moment testi, Shapiro-Wilk testi, Zhang Testi, D'Agostino testi, Cramer-von Mises testi, Anderson-Darling testi veya Kolmogorov-Smirnov testi gibi yöntemlere başvurulmaktadır. Bunlar arasında özellikle Kolmogorov-Smirnov (Lilliefors modifikasyonu) ve Shapiro-Wilk testleri IBM SPSS'te kolaylıkla erişilebilir (*SPSS*→ *Analyze*→ *Descriptive Statistics*→ *Explore…*→ *Plots…*→ *Normality plots with tests*) olduğundan, bu paket programın kullanıcıları arasında son yıllarda en sık kullanılan ölçülerdir. Ancak Likert tipi ölçek verilerinin normallik analizi için bu iki testin uygunluğu bir tartışma konusudur. Verinin basıklık çarpıklık değerleri normal dağılıma uygunluğu incelemede daha anlamlı sonuçlar verebilmektedir (Hair vd., 2019: 48; George & Mallery, 2019: 114-115).

Shapiro-Wilk Testi

1965 yılında, Shapiro ve Wilk tarafından geliştirilmiştir. Yazarlar makalelerinde (Shapiro & Wilk, 1965) yöntemi detaylı olarak vermişlerdir. Test istatistiğinin genel terimi

$$W = \frac{(\sum_i^n a_i \cdot y_i)^2}{\sum_i^n (y_i - \bar{y})^2}$$

şeklindedir. Burada y_i'ler test verisinin küçükten büyüğe sıralanmış halini temsil eder. Genel terimdeki a_i'ler ise makalede de verilmiş olan tablo değerlerinden çekilir. Bu değerler örneklem büyüklüğüne (n) göre tabloda ayrı ayrı sıralanmıştır ve normal dağılım eğrisine uygun içeriktedir. Dolayısıyla W değeri, y_i şeklindeki test verisinin, normal dağılıma uygun özelliğe sahip a_i verisi ile ilişkisini ortaya koymaktadır. W, 1'e yaklaştıkça test verisi ile normal dağılım vektörünü temsil eden a_i verisi arasında güç artar. Söz konusu makalede aynı zamanda, örnek büyüklüğüne göre W gücünü tanımlayan tablo verilmiştir. Test verisinin normal dağıldığını kanıtlamak için Shapiro-Wilk p istatistiğinin 0,05'ten büyük olması gerekmektedir.

Shapiro-Wilk testinin Kolmogorov-Smirnov testinden daha anlamlı/güçlü olduğuna dair kanıtlar bulunmaktadır (Mendes & Pala, 2003; Keskin, 2006; Razali & Wah, 2011). Ancak bu test küçük örneklem (n≤50) verisi için geliştirilmiştir (Shapiro & Wilk, 1965).

Kolmogorov-Smirnov (Lilliefors modifikasyonu)

İlk olarak 1933 yılında A.N.Kolmogorov, sonrasında da 1939 yılında N.V. Smirnov tarafından geliştirilen bu test, birbirine yakın yöntemler olduğundan Kolmogorov-Smirnov adıyla ifade edilmektedir. Tek örnekleme dayalı olan bu test istatistiğinin genel terimi teorik (varsayılan) dağılımın kümülatif bağıl frekansı ile örneklem frekansı değerinin mutlak farklarının en büyüğü olup,

$$D = max|F_n(X) - F(X)|$$

şeklinde ifade edilir. Aradan 28 yıl sonra, Lilliefors (1967)'un yaptığı modifikasyon ile bu yöntem yoğun kullanılır hale gelmiştir. Lilliefors (1967), Monte Carlo yöntemini kullanarak, ortalama ve varyans belirtilmediğinde ve örneklemden tahmin edilmesi gerektiğinde, Kolmogorov-Smirnov istatistiğiyle birlikte kullanılmak üzere bir tablo sunmuştur. Bu kritik değerler tablosunda D değeri örneklem büyüklüğüne değişmektedir. 30'un üzerindeki örneklem büyüklüğü (n) için;

$$D = \frac{886}{\sqrt{n}} \cdots p \cong 0{,}05$$

$$D = \frac{1031}{\sqrt{n}} \cdots p \cong 0{,}01$$

olup; burada Kolmogorov-Smirnov (Lilliefors modifikasyonu) test istatistiği,

$$D = max|F^*(X) - S_n(X)|$$

şeklindedir. Test verisinin normal dağıldığını kanıtlamak için Kolmogorov-Smirnov p istatistiğinin 0,05'ten büyük olması gerekmektedir.

Dijital veya yüz yüze anketlerde özellikle Likert tipi aralıklı ölçek verisi toplandığından normallik testi için Kolmogorov-Smirnov istatistiği bir tartışma konusudur. Bu çalışma ile Kolmogorov-Smirnov (Lilliefors modifikasyonu) testinin normal dağılıma uygunluk analizi için yeterli olmadığı kanıtlanmıştır.

Basıklık Çarpıklık Katsayıları

Basıklık katsayısı veri setinin meydana getirdiği eğrinin sivriliğini veya basıklığını gösteren bir ölçüdür. Çarpıklık katsayısı ise eğrinin sağa veya sola doğru çarpıklığına ait bir göstergedir. Formülleri,

$$Basıklık = \frac{\sum_i^n (x_i - \bar{x})^4}{n \cdot s^4} - 3$$

$$Carpıklık = \frac{\sum_i^n (x_i - \bar{x})^3}{n \cdot s^3}$$

şeklindedir.

Her iki katsayı değer sıfıra yaklaştıkça normallik dereceleri artacaktır. Çarpıklık katsayı değeri sıfır olması durumda eğri tam simetriktir; yani ortalama, mod ve medyan değerleri birbirine eşittir. Basıklık formülünün orijinalinde -3 sabiti bulunmamaktadır (Newbold, 1988: 420); ancak çarpıklıkta olduğu gibi sıfırın tam normalliği işaret etmesi amacıyla formüle bu sabit eklenmiştir. Hair vd. (2019:48) çarpıklık katsayısının [-1, +1] aralığı dışında olması durumda veri setinin çarpık bir dağılımı işaret ettiğini öne sürmüşlerdir. George & Mallery (2019: 114-115) ise benzer görüşte olmakla birlikte, belirli uygulamalara bağlı olarak, [-2, +2] aralığının kabul edilir olabileceğini ifade etmişlerdir. Bu çalışma ile test verisinin çarpıklık katsayısının [-1, +1] dışında olması durumda normal dağılım sergilemediği kanıtlanmıştır.

Bunların yanında [-3, +3] değerlerinin uygun aralık olduğunu, Groeneveld & Meeden (1984), Moors (1986), Hopkins & Weeks (1990) ve De Carlo (1997) kaynaklarını göstererek; bu aralıkta buldukları basıklık çarpıklık katsayılarına ait verilerin normal dağıldığını kabul eden çalışmalar da bulunmaktadır. Ancak ne yazık ki söz konusu kaynakların hiçbirinde böyle bir kanıt bulunmamaktadır. Özetle, normal dağılım için basıklık çarpıklık katsayılarının [-3, +3] aralığındaki değerlerin yeterli olabileceğine dair kanıtı olan bir çalışma bulunmamaktadır.

ARALIKLI (1-5 LİKERT TİPİ) ÖLÇEK VERİLERİ İÇİN NORMALLİK

Aralıklı ölçekler pazarlama araştırmalarında sıklıkla kullanılan bir ölçüm türüdür. Derecelendirme 1'den başlar ve 1'er artışla devam eder. Bu nedenle aralıklı ölçeğin en önemli özelliği derecelendirmeli olmasıdır ve bu derecelendirmede aralıklar eşit sıralıdır (Haşıloğlu;2022: 147-148). İdeal cevap alternatifi sayısı araştırma konusu, modeli ve örneklemin özelliğine göre değişmekte olup, araştırmacılar tarafından yoğun olarak 1-5 aralıklı (Likert Tipi) ölçek tipi tercih edilmektedir (Kalburan vd., 2019). Likert tipi ölçekler ile aralıklı ölçek verisi toplandığından, ölçme aracı en az üç maddeden meydana geliyor olsa dahi veri setlerinde tekrarlı değerler hakimdir. Bu özellikte ve tekrarlı değerlerin bulunduğu bir veri setinin normal dağılıma uygunluğunun analizi için Shapiro-Wilk testi veya Kolmogorov-Smirnov testi gibi, rasyonel sayıların temel alındığı normallik testlerinin kullanılması bir tartışma konusu olmuştur. Yine basıklık çarpıklık değerleri için kabul edilebilir uyun aralık üzerine de farklı görüşler bulunmaktadır.

Kolmogorov-Smirnov Testinin Yetersizliği

Çalışmada ilk olarak, Likert gibi aralıklı ölçek ile toplanan verilerde Kolmogorov-Smirnov (Lilliefors modifikasyonu) testinin normal dağılıma uygunluk analizi için yeterliliği incelenmiştir. Bu amaçla anket verisi niteliğinde ve normal dağılama uygun üç veri seti üretilmiştir. Birinci veri seti (Z_i) tek maddeli 1-5 aralıklı ölçek (1,2,3,4,5 şeklindeki tam sayılar) verisinden meydana gelmektedir. İkinci veri seti (R_i) ise en az üç maddeli 1-5 aralıklı ölçek ortalamasını temsil eden değerlerden meydana gelmektedir. Her ikisinde de tekrarlı değerler hakimdir. Ayrıca kontrol amaçlı geliştirilen üçüncü veri seti ($R_i \pm \varepsilon$), ikinci veri setindeki her bir değere ±0,09 uzaklıkta olacak şekilde (ε<0,09; rastgele değer) üretilerek tekrarlı değer en aza indirilirmiştir. Tablo 1'de (Table 1) normal dağılıma uygun olarak üretilen bu üç veri setlerindeki değerlerin frekansları bulunmaktadır.

Kolmogorov-Smirnov ve Shapiro-Wilk testlerine göre tek maddeli 1-5 aralıklı ölçek örneği veri seti (Data set-1: Z_i) ve en az üç maddeli 1-5 aralıklı ölçek ortalamasını temsil eden veri seti (Data set-2: R_i) normal dağılım sergilememektedir (Table 2). Ancak Resim 1'den (Figure 1) görüleceği üzere veri setleri normal dağılım sergilemektedir. Dolayısıyla Likert tipi ölçek ile toplanan tekrarlı değerlerin hakim olduğu tek veya en az üç maddeli veri setlerinin normallik testlerinde Kolmogorov-Smirnov (Lilliefors modifikasyonu) uygun bir yöntem değildir.

Diğer yandan Data Set-3: R_i_Epsilon veri setinde tekrarlı değerler olmadığından Kolmogorov-Smirnov testine göre bu veriler normal dağıldığı halde Shapiro-Wilk testine göre normal dağılmamaktadır. Örneklem büyüklüğünden kaynaklanan bu çelişkinin önüne geçilmesi için örneklem büyüklüğü 32'ye düşürülmüştür. 15 sistematik aralıkla veri çekilerek seçilen bu veri seti hem Kolmogorov-Smirnov testi (p=0,200) hem de Shapiro-Wilk testine (p=0,717) göre normal dağılım sergilemektedir.

Kabul Edilebilir Çarpıklık Katsayısı Değer Aralığı

Çalışmanın ikinci aşamasında Likert tipi (1-5 aralıklı) ölçek ile uyumlu verilerin (-1, +1) aralığı dışındaki çarpıklık katsayı değeri normal dağılıma uygunluğu için yeterliliği incelenmiştir.

Tablo 3'te (Table 3) standart normal dağılıma uygun verilerin 1-5 normalizasyon değerlerine ait frekanslar ve çarpıklık katsayısı yer almaktadır. Veriler tam simetrik normal dağıldığından çarpıklık katsayısının 0 olması beklenen bir sonuçtur. Tablo 3'ün sadece $Y_i \in [1, 3]$ verileri kullanıldığında dikey olarak ikiye bölünmüş eğrinin sadece sol yanı kalacaktır. Artan nitelikte olan bu eğrinin normal dağılım sergilediğini söyleyemeyiz (Figure 2). Bununla birlikte $Y_i \in [1, 3]$ aralığı için çarpıklık katsayısı -1,014'tür. [1, 3,17] aralığı için çarpıklık katsayısı -0,922 olup, aralık genişledikçe katsayı sıfıra yaklaşmaktadır. Dolaysıyla bu örnek verisi seti için (-1, +1) aralığı dışındaki çarpıklık katsayı değeri normal dağılıma uygunluk analizi için yeterli olmadığı söylenebilir.

SONUÇ

Anketler, içerisinde ölçekler barındırdığından, sosyal bilimcilerin en yoğun kullandığı veri toplama aracıdır. Ölçeklerin temelini oluşturan Likert tipi ölçeğin geçmişi 1930'lu yıllara kadar uzanmaktadır. Birçok sosyal ve eğitim bilimcinin tutum, niyet, eğilim ve inanç gibi faktörlere yönelik ölçüm aracıdır. Bu ölçek türü eşit aralıklardan meydana gelmekte olup, araştırmacılar tarafından sıklıkla 5 cevap alternatifi kullanılmaktadır. Ölçüm için en az üç maddenin aritmetik ortalaması kullanılıyor olsa da toplanan cevaplar tekrarlı veri ağırlıklıdır. Tüm bunlara bağlı olarak anketler aracılığı ile toplanan verilerin normallik testinde rasyonel sayılar için uygun olan uyum iyiliklerinin kullanılması bugüne kadar hep bir tartışma konusu olmuştur. Bunlar arasında Shapiro-Wilk ve Kolmogorov-Smirnov testi en çok tercih edilen olmuştur. Bu çalışma ile her iki yöntemin de anket verisi için uygun olmadığı sonucuna ulaşılmıştır. Bu nedenle araştırmacıların Likert tipi ölçeğe dayalı dijital veya yüz yüze anket verilerinin nöral dağılıma uygunluğunu kanıtlamak için Shapiro-Wilk ve Kolmogorov-Smirnov (Lilliefors modifikasyonu) testlerine başvurmamaları gerekmektedir.

Shapiro-Wilk veya Kolmogorov-Smirnov testlerine alternatif olarak bazı araştırmacılar basıklık çarpıklık katsayılarına bakılmasını önermektedirler. Ancak burada da kabul edilebilir çarpıklık katsayısı değer aralığı hakkında farklı görüşler bulunmaktadır. Bu çalışma ile (-1, +1) aralığı dışında kalan çarpıklık katsayısına ait veri setinin normal dağılım sergilediğini söylememizin mümkün olamayacağı kanıtlanmıştır. Bununla birlikte, 5'li Likert tipi ölçeğini temsilen, 1, 2, 3, 4 ve 5 değerlerinden eşit miktarda kullanıldığında basıklık katsayısı -1,3 sınırına kadar; 1-5 arlığında eşit miktarda rasyonel sayı kullanıldığında ise -1 sınırına kadar ulaşabilmektedir. Araştırmacıların özellikle [-3, +3] aralığının kabul edilebilir olduğu bilgisine itibar etmemeleri önerilir. Çünkü literatürde kanıtlanmış böyle bir çalışma asla bulunmamaktadır. Kaynak (Groeneveld ve Meeden, 1984; Moors, 1986; Hopkins ve Weeks, 1990; De Carlo, 1997) olarak belirtilen referanslar gerçekçi değildir.