DIFFERENCE OF POWER TEST AND TYPE II ERROR (β) ON MARDIA MVN TEST, HENZE ZIKLER'S MVN TEST, AND ROYSTON'S MVN TEST USING MULTIVARIATE DATA ANALYSIS

Wahyul Anis[1], Kuntoro[2], Soenarnatarlina Melaniani[3]
[1]Faculty of Medicine, Universitas Airlangga, 60132 Surabaya, East Java, Indonesia
[2,3]Faculty of Public Health, Universitas Airlangga, 60115 Surabaya, East Java, Indonesia
Corresponding Author: Wahyul Anis
E-mail: wahyul.anis@fk.unair.ac.id

ABSTRACT

The Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test are the most widely used tests to analyze multivariate normal (MVN) data, but there have not been many studies explaining the advantages and disadvantages of these tests. The research objective was to analyze the difference in test strength and type II (β) error in the Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test. The research data were analyzed using three MVN tests, namely the Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test. The results of the analysis in the form of test strength and type II error (β) would be compared at alpha (α) 1%, 5%, 10%, 15%, 20%, and 25%. The comparison results explained that the Mardia test had the greatest test strength and the smallest type II (β) error. The study concluded that the Mardia MVN test was a multivariate normal test better than Henze Zikler's MVN test and Royston's MVN test.

Keywords: multivariate normal test, Mardia MVN test, Henze Zikler's MVN test, Royston's MVN test

ABSTRAK

Uji Mardia MVN, uji Henze Zikler's MVN dan uji Royston's MVN adalah uji yang paling banyak dilakukan untuk menganalisis kenormalan data multivariat/multivariate normal (MVN). Namun, belum banyak penelitian yang menjelaskan kelebihan dan kekurangan dari uji tersebut. Penelitian ini bertujuan untuk menganalisis perbedaan kekuatan uji dan kesalahan tipe II (β) pada uji Mardia MVN, uji Henze Zikler's MVN dan uji Royston's MVN. Data penelitian dianalisis menggunakan tiga uji MVN yaitu uji Mardia MVN, uji Henze Zikler's MVN dan uji Royston's MVN. Hasil analisis berupa kekuatan uji dan kesalahan tipe II (β) dibandingkan dengan alpha (α) 1%, 5%, 10%, 15%, 20% dan 25%. Hasil perbandingan menunjukkan bahwa uji Mardia memiliki kekuatan uji terbesar dan kesalahan tipe II (β) terkecil. Uji Mardia MVN merupakan uji normal multivariat yang lebih baik dibandingkan uji Henze Zikler's MVN dan uji Royston's MVN.

Kata kunci: *uji normal multivariat, uji Mardia MVN, uji Henze Zikler's MVN, uji Royston's MVN*

INTRODUCTION

The statistical test is a data analysis method used to support quantitative research. Multivariate data must be tested for normal multivariate to allow further statistical tests to be determined according to the results of data normality. Multivariate normal (MVN) is one of the most important assumptions for multivariate parametric tests. The assumption of normality explains that the sample group is distributed and has the assumption of variance from the same sample and is under the population (Nahm, 2016). If the data are not normal, then it cannot be tested under multivariate parametrics because it does not meet the assumptions. If the assumption is ignored, the results of the statistical analysis will be invalid or biased, especially in small or fewer sample sizes (Vogt, 2015).

The normal distribution is the probability distribution of the most important continuous random variable because it is most often encountered both theoretically and practically in statistical analysis. The normal distribution is also known as the Gaussian distribution. The normal distribution equation is initially investigated through residuals or errors with repeated measurements of the same simulation.

The sign possessed by normally distributed data is that it has the mean parameter

(µ) and standard deviation (σ) which will determine the location and shape of the distribution. The image on the curve has one peak, the mean is in the middle position of the distribution, the total area under the curve image is one, and the two tails of the curve will be elongated and will not intersect with the horizontal axis to depict a bell-like shape. The standard deviation will determine the width of the curve in the data distribution.

The multivariate normality test has four techniques that underlie the formation of the multivariate data normality test: the goodness of fit technique, procedures based on skewness and kurtosis, univariate test, graphical approach, and correlation.

Currently, there are many kinds of MVN to test the normality of multivariate data, namely graphical approaches (Chi-square, QQ, and plot approaches) and numerical approaches such as the Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test.

Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test are the most widely used tests to analyze normality in multivariate data. Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test are data normality tests that have similarities. Namely, the test results are not in the form of images, but are in the form of probability values.

Each MVN test has advantages and disadvantages. The advantage of the Mardia MVN test is that it can be used for sample sizes of less than 20, but has the disadvantage that it requires correction when the sample size is less than 20 to control type I error (α). Henze-Zikles's MVN test has the advantage that it has good test strength, especially for large sample sizes (n>100), but has the disadvantage that it requires logs from the normal distribution when calculating.

Royston's MVN test has the advantage that it is effective for detecting deviations from multivariate normality and can be used for small and large sample sizes (3≤ n ≤2000). It has the disadvantage that there is a provision if the kurtosis in the data is greater than three, then the Shapiro Franchia test needs to be used for leptokurtic distribution while the Shapiro Wilk test is used for platykurtic distribution.

Test power is defined as the probability or probability that the test results will reject the false null hypothesis. The strength of the test is the complement of type I error (α), which is an error for rejecting the null hypothesis (H0), even though the null hypothesis is true (Zhou and Shao, 2015). Type II error (β) is an error because it rejects the alternative hypothesis (Ha), even though the alternative hypothesis is true. These errors will greatly affect the actual interpretation of the data. The important thing in performing statistical hypothesis analysis is to have the smallest possible error values of type I (α) and II (β). Health research is recommended to minimize type I (α) and II (β) error values.

Based on the similarities, advantages, and disadvantages of MVN, the Mardia MVN test, Henze Zikler's MVN test,, and Royston's MVN test need to be investigated further to determine the best multivariate normality test based on the strength of the test and type II error (β). This study aimed to analyze the difference in test power and type II error (β) in the Mardia MVN test, Henze Zikler's MVN, and Royston's MVN tests at 1%, 5%, 10%, 15%, 20%, and 25%.

METHOD

This study used secondary data on the achievement of maternal health services from the annual report of the East Java Province Health Office. The achievement data included the percentage of pregnant women's first visit (K1), the achievement of complete visits to pregnant women (K4), the achievement of high-risk early detection by health workers (DN), the achievement of delivery assistance by health workers (PN), the achievement of delivery assistance in facilities health (PF), the achievement of handling maternal complications (PK), and the achievement of postpartum maternal visits (KF) in East Java.

The data obtained were descriptions of the achievements of 38 regencies or cities in East Java and the sample size used was 38. Multivariate data were analyzed using three tests for normality of multivariate data, namely the Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test. An advanced statistical application was used with the principle of unit accountability that is the data file that will be used in line with how the data was created, who compiled it, and the computational results that could be

verified (Gelfond *et al.*, 2018). Each test had their p-value, test power, and type II error (β) at 1%, 5%, 10%, 15%, 20%, and 25%, then the results of the test analyses were compared with each other.

RESULTS

The data were analyzed using the univariate normality test before the data were analyzed using the multivariate normality test. This is because the data that had been declared normally distributed in the univariate normality test would have a greater chance of having a normal distribution in the multivariate normality test.

Univariate normality test results are presented through numerical and graphical descriptive statistics. Numerical statistical descriptions are presented through mean, mode, standard deviation, variance, range, lowest value, and highest value. Univariate normality analysis is descriptive by looking at the average value, median value, and mode, in addition to looking for the coefficient of variance. Data were also presented in graphical form using histograms. The data were then analyzed using three MVN tests to compare their p-value, test power, and type II error (β). The results of the univariate and multivariate normality tests were compared with each other to see if there were differences in the results of the normality test.

Table. 1 is the result of a descriptive analysis of the data. The analysis results obtained include the average, mean, mode, standard deviation, variance, range, lowest value, and highest value. The average, mean, and mode are often referred to as measures of central tendency. If the average, the mean, and the mode have the same value, it will form a symmetrical frequency distribution curve and can be interpreted as a normal distribution.

Table 1 shows that each variable does not have the same average, mean, and mode. It can be interpreted that each variable is not normally distributed.

Other parameters that can be used as a reference in determining the normal distribution are the average value (mean) and standard deviation. The average value is used as the center of the distribution. The standard deviation is used to determine the possibility of the data having a curve area that exceeds the center of the curve or a curve wider than a normally distributed curve. The values in the results of a descriptive analysis that can be used to determine normality are the coefficient of variance, the ratio of skewness, and the ratio of kurtosis. Table 1 obtained the results of the coefficient of variance using the formula:

$$\frac{\text{Standard Deviation}}{\text{Average value}} \times 100\%$$

Data are normally distributed if a variance coefficient is less than 30%. The results of the calculation of the coefficient of variance in Table 1 explain that each variable tested had a coefficient of variance of less than 30%. It can be concluded that based on the analysis of the coefficient of variance, all variables were normally distributed.

The next univariate normality test was a graphical approach, namely a histogram. The histogram is popular and often used because it has eased in reading the results or interpretation. However, several important things must be considered in interpreting the results of the analysis, namely seeing the slant and sharpness of the image. The results of the analysis using histograms have a high possibility of subjectivity. Some sources recommend to also look at the value of the skewness ratio and the kurtosis ratio.

Table 1. Description of maternal health service statistics in East Java Province

Criteria	K1	K4	DN	PK	PN	PF	KF
Average value	98.07	90.45	24.35	19.38	95.05	94.03	91.00
Middle value	98.46	91.23	22.07	19.41	94.97	93.80	91.00
Mode	100.00	91.83	21.48	-	-	98.64	93.30
Standard deviation	2.80	4.30	6.49	2.89	3.85	3.84	2.39
Variance	8.07	18.49	42.22	8.33	14.83	14.79	5.71
Range	13.88	17.15	30.02	13.19	14.50	16.45	10.29
Lowest value	90.47	81.26	15.41	13.55	86.15	84.01	85.86
The highest score	104.35	98.41	45.43	26.74	100.65	100.46	96.15
Coefficient of variance	2.85	4.75	26.65	14.81	4.05	4.08	2.62

The histogram for each variable to explain univariate normality can be described as follows:
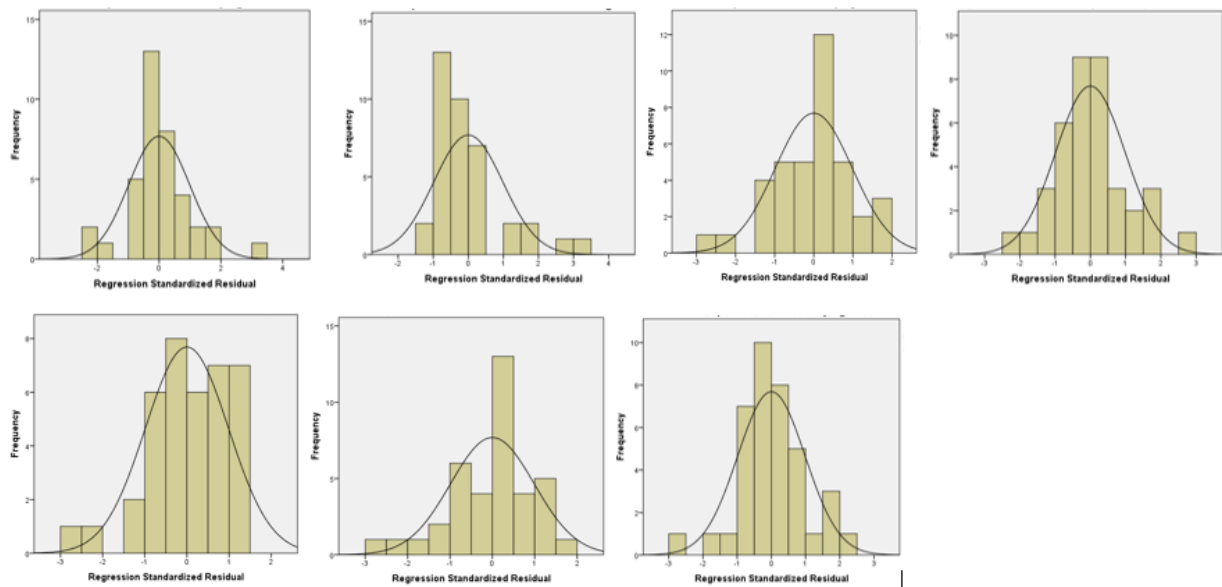

Figure 1. The histogram on each variable (univariate)

Figure. 1 shows that all variables have skewness while judging from the kurtosis or sharpness. It can be explained that all variables are more pointed than the standard normal distribution. It can be concluded that all variables are not normally distributed. Data that are normally distributed will have a slope (skewness) of 0 (zero) and a kurtosis of 3.

The results of the univariate normality test using the average, median, and mode approach give the same results as the histogram. All variables were not normally distributed, but based on the coefficient of variance, it was found that the distribution of all variables was normal. This can happen because each statistical test has advantages and disadvantages; therefore, that the selection of the statistical test used is important.

Univariate data analysis was followed by multivariate normality analysis using Mardia MVN, Henze Zikler's MVN, and Royston's MVN tests. Based on the results of the univariate normality test, the results of the multivariate normality test can be predicted to have abnormal distributions.

The results of the Mardia MVN test, Henze Zikler's MVN, and Royston's MVN test were presented in the form of p values, test strength, and type II error (β). Test power and type II error (β) were compared at alpha (α) 1%, 5%, 10%, 20%, and 25%. The Mardia MVN test had a p-value of skewness and kurtosis while Henze Zikler's MVN test and Royston's MVN

test had p-values only. The test power value was obtained by first finding the effect size value on the sample size. The type II error (β) was obtained through the calculation results of the formula: 1-value of the test strength.

The results of the Mardia MVN test analysis on the "R" application were the p-value of skewness 2.2×10^{-8}, while the Mardia estimated value was 28.80. The p-value of kurtosis was 46.2×10^{-5}, while the estimated value of the Mardia multivariate kurtosis was 78.65. The results of the comparison of the p-value with Mardia's estimated value concluded that the multivariate data were not normally distributed because the p-value was lower than the estimated value.

Henze Zikler's MVN test analysis results on the R application showed a Hanze value of 1.33, while the p-value was 0 (zero). It can be interpreted that the multivariate data did not have a normal distribution in alpha (α) equal to 5% (0.05).

The results of Royston's MVN test analysis on the R application were a Royston value of 23.85 and a p-value of 0.0006. It can be interpreted that the multivariate data did not have a normal distribution at 5% alpha (0.05). The results of the analysis of the strength of the test were obtained by first finding the value of the effect size on the sample size of 38. The effect size of the Mardia MVN test was 0.20. The effect size in the Henze Zikler MVN test was 0.0. The effect size in Royston's MVN was 0.04. Effect size is a value that describes

Table 2. Value comparison p, test strength, and type II error (β) on the Mardia MVN test, Henze Zikler's MVN test, and Royston's MVN test on 1%, 5%, 10%, 20%, and 25%

MVN Test	p-value	Type I Error (α)	Test Strength	Type II Error (β)
MVN Mardia Test	2.2×10^{-8} (Skewness) 46.2×10^{-5} (Kurtosis)	1%	0.09	0.91
		5%	0.25	0.75
		10%	0.36	0.64
		15%	0.44	0.56
		20%	0.50	0.50
		25%	0.56	0.44
Henze Zikler's MVN Uji test	0.00	1%	0.01	0.99
		5%	0.05	0.95
		10%	0.10	0.90
		15%	0.15	0.85
		20%	0.20	0.80
		25%	0.25	0.75
Royston's MVN test	6×10^{-4}	1%	0.01	0.99
		5%	0.06	0.94
		10%	0.10	0.90
		15%	0.17	0.83
		20%	0.22	0.88
		25%	0.27	0.73

the size of the influence of a variable on other variables. The effect size calculated before the analysis or test has the aim of determining the size of the sample to be used to obtain meaningful test results. The effect size value is opposite to the sample size, meaning that the smaller the effect size value, the larger sample size required for meaningful test results. The results of the comparison of effect sizes in the three tests show that the largest effect size is in the Mardia MVN test. It can be explained that the Henze Zikler MVN test and Royston's MVN test require a larger sample size in order to get an effect akin to the Mardia MVN test.

Table 2 shows that at alpha (α) 1%, 5%, 10%, 15%, 20%, and 25%, the power of the Mardia MVN test was greater than Henze Zikler's MVN test and Royston's MVN test. Type II error (β) at alpha 1%, 5%, 10%, 15%, 20%, and 25% obtained the smallest value in the Mardia MVN test.

DISCUSSION

Normality of data can be assessed visually (graphic) and through normality tests (numeric). Visual normality tests can use PP and QQ plots, but some studies state that the results are very subjective and it is better to use normality tests through p-value (Mishra *et al.*, 2019). The graphic plot that is the easiest and

simplest method for determining univariate normality is a histogram. The normal distribution is observed through the shape of a bell in terms of slant and taper (Rani, 2016).

It can be concluded that all variables were not normally distributed from the results of the analysis of the univariate normality test on each variable using descriptive numerical analysis using the average, median, and mode in Table 1 and descriptive graphic using the histogram in Figure 1,

The first step to see the distribution of multivariate normal data is to check the normality of each variable using the univariate normality method on all variables (Cain, Zhang, and Yuan, 2017). Data that has been declared normally distributed in the first stage has a high probability of being normally distributed in the multivariate normality test (Ellis *et al.*, 2017).

The results of the multivariate normality test for the p-value of the Mardia MVN test, Royston's MVN, and Henze Zikler's MVN test at 5% alpha obtained the same conclusion that the multivariate data were not normally distributed. Some cases where the data are not normally distributed can be caused by outliers and a lack of sample size.

The normal distribution is the probability distribution of the most important continuous random variable because it is most

often encountered both theoretically and applied in statistical analysis. The normal distribution is also known as the Gaussian distribution. The normal distribution equation is initially investigated through residuals or errors with repeated measurements of the same simulation.

The normal distribution describes the situation quantitatively in both the natural and social sciences. Various kinds of statistical tests can be determined by following the normal distribution approach. The normal distribution is the test that is most often performed in statistics because some statistical test methods require requirements for normality of the data to allow further analysis to be carried out, for example, in parametric statistics (Kestin, 2018).

The central limit theorem explains that in random sampling from a population distribution (regardless of the type of distribution), the distribution of the sample obtained is close to a normal distribution. The larger the sample, the closer the distribution to the normal distribution, or the closer to the population distribution.

The size of the sample distribution that is close to the normal distribution is influenced by the distribution of the population. If the population is normally distributed, the distribution of each sample size is not always normal, especially if the sample size is small. If the population has a distribution that is very far from normal (abnormal), then a larger sample size is needed to get a sample distribution that is close to normal.

The difficulty of determining the shape of the distribution of the population can be overcome by making one important sign, namely to apply the 'central limit theorem', it must have a large sample. The 'central limit theorem' also explains that, in general, a sample size of more than 30 can be explained as having sufficient numbers to fulfill the 'central limit theorem'. It can be concluded that a larger sample size or a minimum of 30 samples is needed to get a distribution that is close to normal. The principle of the 'central limit theorem' is that if the research wants to get more information from the sample taken, the size of the sample that is determined should be as large as possible because the larger the sample, the more the results obtained will describe the population being studied.

The Mardia MVN test is a normal multivariate test based on the intensity of the skewness ($\hat{\gamma}1$, p) and kurtosis/curly ($\hat{\gamma}2$, p). Mardia's test can be used for sample sizes of less than 20 samples. Small sample size has strength and type I error that can be violated. Mardia introduced the term correction into the statistical skewness test which is used when the sample size is less than twenty (n < 20) to control for type I error (α) (Mardia, 1970).

The equation of the multivariate normal Mardia statistical test is as follows:

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=0}^{n} \sum_{j=1}^{n} [(x_i - \overline{x})S^{-1}(x_j - \overline{x})]^3$$

and

$$\hat{\gamma}_{2,p} = \frac{1}{n^2} \sum_{i=0}^{n} m_{ii}^2$$

$i = 0.1,2\ldots,n$
$j = 1,2,3,\ldots,n$

Description:

p is the number of the variable

Statistical test for skewness ($\hat{\gamma}1$,p)is about a chi-square distribution with p(p+1)(p+2)/6 degrees of freedom. Statistical test for kurtosis, ($\hat{\gamma}2$,p) is approximately a normal distribution with a mean of p(p + 2) and a variance of 8p (p + 2)/n.

The results of the Mardia MVN test can also be in the form of p values of skewness and kurtosis which will then be compared with the estimated values of Mardia skewness and kurtosis. If the value of p skewness and kurtosis is smaller than the estimated value of Mardia skewness and kurtosis, then H0 is rejected. The provisions for H0 are multivariate data with normal distribution, and H1 is multivariate data that are not normally distributed.

In his article entitled *Some Techniques for Multivariate Normality Assessment* based on Shapiro Wilk W, Royston explained that the advantage of the Royston test is that it can be used on sample sizes of more than 3 and less than 100 (3<n<100). Previous research suggests choosing Royston's MVN test as a normality test on multivariate data with a sample size of less than 50 (n<50) (Korkmaz, Goksuluk and Zararsiz, 2014).

Royston's MVN test uses Shapiro Wilk/Shapiro Francia statistics for the multivariate normality test. If the kurtosis in the data is greater than 3, then the Shapiro Franchia

test is used for the leptokurtic distribution, while the Shapiro Wilk test is used for the platykurtic distribution.

Royston's MVN test equation is as follows:

$$H = \frac{e \sum_{j=1}^{p} \psi_j}{p} \sim x_e^2$$

Information:

e is the equivalent degree of freedom

H is Royston's statistical test.

Royston's MVN test results can also be in the form of a p-value and then analyzed by comparing it with the alpha value (α). If the obtained p-value is smaller than the predetermined alpha (α) value, then H0 is rejected. The provisions for H0 are multivariate data with normal distribution and H1 is multivariate data that are not normally distributed.

Hanze Zikler's MVN test is a data normality test based on the distance between two distribution functions. If the data are multivariate normal distribution, the statistical test is the log-normal distribution. The mean, variance, and parameters are calculated first, followed by log-normal and the p-value (Henze and Zirkler, 1990). The calculation of the Henze Zikler MVN statistical test takes into account the number of variables, the distance of Mahalanobis, and the sample size.

The result of Henze Zikler's MVN test is the p-value analyzed by comparing it with the alpha value (α). If the p-value is smaller than the predetermined alpha (α) value, then H0 is rejected. The condition for H0 is multivariate data with normal distribution and H1 is multivariate data that are not normally distributed.

The Henze Zikler MVN statistical test equation is as follows:

$$HZ = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\frac{\beta^2}{2}D_{ij} - 2(1-\beta^2)} \sum_{i=1}^{n} e^{\frac{-\beta^2}{2(1+\beta^2)}D_i}$$
$$+ n(1 + 2\beta^2)^{\frac{-p}{2}}$$
$$\beta = \frac{1}{\sqrt{2}} \left(\frac{n(2p+1)}{4} \right)^{\frac{1}{p+4}}$$
$$D_{ij} = (x_i - x_j)' S^{-1} (x_i - x_j)$$
$$D_i = (x_i - \bar{x}) S^{-1} (x_i - \bar{x}) = m_{ij}$$

Information:

HZ is Henze Zikler statistical test

p is the number of variables

Dij is the Mahalanobis distance between i and j (observation)

n is the sample size

Table 2 shows the results of the comparison of p-value with alpha (α) 5% from each test. In Henze Zikler's MVN test, it is more difficult to conclude that the multivariate data are normally distributed compared to the Mardia MVN test and Royston's MVN test. This is because Henze Zikler's test MVN has a lower p-value than the Mardia MVN test and Royston's MVN test. Henze Zirkler's MVN test is one of the MVN tests that best maintains the nominal level of significance (Hanusz *et al.*, 2018). In Royston's MVN test, it is easier to conclude that the multivariate data is normally distributed compared to the Mardia MVN test and Henze Zikler's MVN test. This is because the p-value in Royston's MVN test is greater than the other tests.

The results of the comparison of test strengths show that the Mardia MVN test has the greatest test power value compared to Henze Zikler's MVN and Royston's MVN. The Mardia MVN test can be used for a sample size of fewer than twenty samples. The small sample size has strength and type I error that can be violated. This means that the sensitivity of the Mardia MVN test to type 1 error is still not good (Korkmaz, Goksuluk, and Zararsiz, 2014).

The results of research by Mecklin and Mundrom recommend using the Mardia MVN test because the test is based on skewness and kurtosis, meaning that it can be used to determine deviations from normal multivariate. Other studies have also shown that the Mardia MVN test is sensitive to specific symmetric violations (Mecklin and Mundfrom, 2005).

Many previous studies have concluded that the multivariate normality test, namely Henze Zirkler's MVN and Royston's MVN tests, are good multivariate normality tests. This is because they have type I error control (α), they are strong, and they can be used to find the cause of deviations from the normal multivariate. The Mardia test MVN can calculate the skewness coefficient correction for a small sample size of fewer than 20 samples (n < 20) (Farrell, Salibian-Barrera, and Naczk, 2007).

Table 2 also shows the multivariate normality test based on the type II error value (β) at different alphas. The Mardia test had a value of type II error smaller than Henze Zikler's MVN and Royston's MVN. It can be concluded that the Mardia MVN test has a

smaller chance of type II error (β) than Henze Zikler's MVN and Royston's MVN test.

The results of the comparative analysis of power and type II errors in the three MVN tests also explain that the larger the type I error (α), the smaller the type II error (β). The results of several studies prove that type I (α) and II (β) errors are opposite to each other. This analysis aimed to minimize errors of type I and II, but its implementation is difficult.

The final result of the study found that the Mardia MVN test was better at controlling type II errors (β). This differs from the results of research conducted by Mecklin and Mundrom, which analyzed several tests of normality of multivariate data through a comprehensive simulation study on type I (α) and II (β) errors. They found that there was no best normality test for all types of data, but also highly recommended Henze Zikler's MVN and Royston's MVN test for normality testing of multivariate data because they have good control on type I (α) and II (β) errors.

The strength of the statistical test depends on the level of significance set by the researcher, the sample size and effect size, or the extent to which the group differs based on the treatment. The power of statistical tests is critical for healthcare providers to decide how many patients to enroll in clinical studies. The strength of the test is closely related to the sample size, when the sample size is large, there is no problem with the strength of the test. However, if the sample size used is small, the power of the test will be low, and researchers must be aware of the possibility of type II errors (β). The larger the sample size in a study, the more likely it is that a researcher will reject the null hypothesis (Shreffler and Huecker, 2020; Kim, 2015). Research in the health sector, in addition to considering the p-value (type I error), must also pay attention to the strength of statistical tests when interpreting study results (Ueki and Sakaguchi, 2018). The strength of the test can be considered by researchers when choosing the most appropriate test for the data they have (Lee and Lee, 2018).

It is recommended that multivariate normality test assessments use several methods/tests because each test has advantages and disadvantages. The most widely used multivariate normality test is Royston's MVN test and the Mardia MVN test. Many alternative methods can be used when the assumption of normality cannot be maintained to allow using non-parametric tests or changing the data to approach or become normal (Oppong and Agbedra, 2016).

This study had a sample size of 38, which is included in the category of small sample size. The results of the study indicate that the Mardia test was the most appropriate test used for this data, but different sample sizes may produce different analytical conclusions. Future research can develop research using different sample sizes.

The limitation of this study is that the researcher did not use variations in the sample size. Sample size can affect the strength of the test. Large sample sizes will result in a greater sensitivity test than small sample sizes. The sample size is required for sensitivity evaluation and specificity detection or diagnostic testing based on various predefined parameters (Bujang and Adnan, 2016). Sample size should not depend on the availability of the number of respondents. In small sample sizes, it is highly recommended to carry out multicentric studies to provide good test power in identifying clinical differences (Borkowf, Johnson, and Albert, 2018).

CONCLUSIONS AND SUGGESTIONS

Conclusion

The Mardia MVN test is a better multivariate normality test than Henze Zikler's MVN test and Royston's MVN test on data that has a sample size of 38 because it has the largest test power value and the smallest type II error (β).

Suggestion

The researcher recommends using a varied sample size in the next study to produce a more detailed test difference.

REFERENCES

Borkowf, C.B, Johnson, L.L and Albert, PS. 2018. *Power and Sample Size Calculations*. [online] *In Principles and Practice of Clinical Research*. Elsevier.

Bujang, M.A and Adnan, T.H. 2016. Requirements for Minimum Sample Size for sensitivity and Specificity Analysis. *Journal of Clinical and Diagnostic Research*, 10(10), pp.1–6.

Cain, M.K., Zhang, Z. and Yuan, K.H. 2017. Univariate and Multivariate Skewness and Kurtosis for Measuring Nonnormality: Prevalence, Influence and Estimation. *Behavior Research Methods*, 49, pp.1716–1735.

Ellis, A.R, Burchett, W.W, Harrar., S.W, and Bathke, A.C. 2017. Nonparametric Inference for Multivariate Data: The R Package NPMV. *Journal of Statistical Software*, 12(4), pp.156–170.

Farrell, P.J, Salibian-Barrera, M. and Naczk, K. 2007. On Tests for Multivariate Normality and Associated Simulation Studies. *Journal of Statistical Computation and Simulation*, 77(12), pp.1065–1080.

Gelfond, J., Goros, M., Hernandez, B. and Bokov, A. 2018. A System for an Accountable Data Analysis Process in R. *R Journal*, 10(1), pp.6–21.

Hanusz, Z., Enomoto, R., Seo, T. and Koizumi, K., 2018. A Monte Carlo Comparison of Jarque–Bera Type Tests and Henze–Zirkler Test of Multivariate Normality. *Communications in Statistics: Simulation and Computation*, 01(01), pp.1–14.

Henze, N. and Zirkler, B., 1990. A Class of Invariant Consistent Tests for Multivariate Normality. *Communications in Statistics - Theory and Methods*, 19(10), pp.3595–3617.

Kestin, I. 2018. Statistics in Medicine. *Anaesthesia and Intensive Care Medicine*, 19(3), pp.136–143.

Kim, H.-Y. 2015. Statistical notes for clinical researchers: Type I and Type II Errors in Statistical Decision. *Restorative Dentisty and Endodontics*, 40(3), pp.249–252.

Korkmaz, S., Goksuluk, D. and Zararsiz, G., 2014. MVN: An R Package for Assessing Multivariate Normality. *R Journal*, 14(5), pp.1–14.

Lee, S. and Lee, D.K, 2018. What Is the Proper Way to Apply the Multiple Comparison Test?. *Korean Journal of Anesthesiology*, 71(5), pp.353–360.

Mardia, KV, 1970. Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, 57(3), pp.519–530.

Mecklin, CJ and Mundfrom, DJ, 2005. A Monte Carlo Comparison of the Type I and Type II Error Rates of Tests of Multivariate Normality. *Journal of Statistical Computation and Simulation*, 7(2), pp.93–107.

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), pp.67–72.

Nahm, FS, 2016. Nonparametric Statistical Tests for the Continuous Data: The Basic Concept and the Practical Use. *Korean Journal of Anesthesiology*, 69(1), pp.8–14.

Oppong, F., B and Agbedra, S.,Y, 2016. Assessing Univariate and Multivariate Normality, A Guide For Non-Statisticians. *Mathematical Theory and Modeling*, 9(2), pp.26–33

Rani D., K., 2016. A Brief Review of Tests for Normality. *American Journal of Theoretical and Applied Statistics*, 5(1), pp.5–12.

Shreffler, J. and Huecker, MR, 2020. *Type I and Type II Errors and Statistical Power*. [online] StatPearls Publishing LLC. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK557530/>.

Ueki, C. and Sakaguchi, G., 2018. Importance of Awareness of Type II Error. *The Annals of Thoracic Surgery*, 105(1), pp.333–342.

Vogt, W., 2015. Multivariate Normality. In: W.P. Vogt, ed. *Dictionary of Statistics and Methodology*, 3rd ed. SAGE Publications, Inc.pp.59–70.

Zhou, M. and Shao, Y., 2015. A Powerful Test for Multivariate Normality. *Journal of Applied Statistics*, 41(2), pp.351–163.