New Approaches in Frequency Table Construction for Continuous Symmetrical Data

[1]M. B. Mohammed∗, **[2]M. J. Subhi and [3]A. A. J. Wafi**

[1]Department of Mathematics and Statistics, Federal University of Kashere
0182 Gombe State, Nigeria.

[2]Department of Science and Biotechnology, Universiti Selangor Batang Berjuntai
45600, Selangor, Malaysia.

[3]Department of Mathematics, Universiti Putra Malaysia
43400, Serdang, Selangor, Malaysia.

∗Corresponding author: muhdb1@gmail.com

Abstract In computing statistical measures from a frequency table, the midpoint represents the observations in each class. The assumption that the observations are distributed uniformly and concentrated around the midpoint is not always valid. In this article, two new methods of constructing a frequency table, using the mean and median of the classes, are compared with the existing method, which uses the midpoint. The simulation study using symmetrical positive data shows that the method which used the classes' mean performed well compared with the other methods. Freedman and Diaconis (FD) rule is the recommended rule when choosing the number of classes.

Keywords Frequency Table, Midpoint, Continuous Data, Statistic Value.

Mathematics Subject Classification 46N60, 92B99.

1   Introduction

Descriptive statistics plays a significant role of organizing, displaying, and describing data using frequency tables, graphs, and other summary measures. Exploratory data analysis (EDA), a key component of descriptive statistics, employs many methods (mostly graphical) to reveal the underlying structure of the data. Another advantage of the EDA includes detection of outliers and irregularities if present in a data set [13].

Data in original form do not display any meaningful representation unless if organized in a systematic form. The raw data are divided into suitable size classes, showing variates with their corresponding frequencies. When a dataset is systematically organized in this form, the process is referred to as frequency distribution table [14]. This table shows how many data values are in each class. The classes are constructed through which each observation can fall

into one class. A frequency table displays data along with the midpoint, cumulative frequency, and the class boundary [2].

Though condensing the data set when it is large enough is sometimes inevitable, the process can lead to considerable errors compared to the original data. Even if the number of classes is chosen carefully to obtain an adequate scatter in each class, the measures of center, variability, and type of distribution may not be reliable. Nonetheless, frequency tables are necessary for organizing a large dataset and making generalizations [4].

Furthermore, one of the two essential parameters, the number of classes and the class width, is used to construct the frequency table. While the former describes the number of partitions of the dataset, the latter is the distance between lower and upper-class limits [24]. These two parameters are dependent on one another, knowledge of anyone serves. Determining the appropriate number of classes to be used in constructing the frequency table and histogram remains a long-existing problem in statistics. Though several classification rules were reported in the literature; however, none of them is perfect [15]. Another problem of interest in statistical studies is how to minimize the errors generated due to the use of the midpoint in computing statistical measures. These errors are due to the assumption that the arithmetic mean of observations in each class of the frequency distribution table coincides with the midpoint [3,4].

This study is carried out to minimize the amount of error when using the midpoint by proposing two new methods of constructing the frequency table. The methods substitute the midpoint with the mean and median of variates in each class of the frequency table. The efficacy of the methods is ascertained using the root-mean-square error (RMSE). Following this, we recommend the appropriate rule among Sturges, Scott and Freedman and Diaconis rules for choosing the number of classes.

2   Components of a Frequency Table

2.1   Class Interval

The first step in constructing a frequency table for continuous data is to determine the appropriate number of classes. Thus, the table is created such that all data points fall into exactly one of the possible classes. The following guidelines should be observed while dividing continuous data into classes:

1. The classes should be mutually exclusive; that is, non-overlapping. Two different classes should not contain the same value(s).

2. The classes should be exhaustive; That is, they should cover the whole data set.

3. The number of classes should neither be too small nor too large. In other words, there should be relatively fewer classes if the range is small and many classes if the range is large.

4. The classes should be of equal width.

2.2 Number of Classes

In statistics, the rule of thumb mostly suggested in the literature is to have the number of classes between 5 and 25 [10]. Because more than enough classes can widely spread the data through which little advantage can be derived when compared with the raw data. Equally, the use of fewer classes tends to condense the data much, which can lead to the loss of information and render the shape of the distribution undetermined [10]. De Beer and Swanepoel (1997) also observed that, when plotting a histogram fewer than enough bins produces an over smoothed histogram which leads to little information concerning the true density. Meanwhile, a relatively large number of bins can lead to under smoothed histogram, which can display the features of other distributions rather than the estimated density. Increasing or decreasing the number of classes arbitrarily does not guarantee better frequency table. What is important is to choose suitable number of classes or class width. The classes should be large enough to capture the shape features of the distribution and small enough to retain good details in the presence of random sampling error.

2.3 Class Width

The class width is the distance between lower and upper class interval of a class. The choice of class width is directly related to the number of classes. This choice is preferable if all the classes are of equal width; nevertheless, unequal class width is compelled to be used in rare cases [8]. Mathematically, the class width can be defined as

$$w = \frac{R}{k},$$

where k is the number of classes, and R is the range of the data set.

2.4 Class Limits and Class Boundaries

The pairs of numbers written in the column of class intervals are called the class limits. Meanwhile, the class boundaries are values halfway between the upper limit of one class and the lower limit of the next class.

2.5 Midpoint

The midpoint is the average of the upper and lower class limits. It is also the center of bars on a histogram. The assumptions that allow the use of the include the following:

1. The observations within the class interval occur at the midpoint.

2. The frequencies are uniformly distributed within the classes.

3. The observations are concentrated around the midpoint or that the arithmetic mean and the midpoint of the classes are equal [3].

These assumptions are not always valid, especially when the dataset is asymmetric.

2.6   Frequency (f)

In a frequency table,the frequencies usually written as $f_1, f_2, \cdots, f_k$ are the number of oc-curences in the k class intervals.

2.7   Cummulative Frequency (cf)

The cummulative frequency at a particular class is the total frequency up to the upper class boundary of that class. if $f_1, f_2, \cdots, f_k$ are the respective frequencies of k classes in a frequency table, the cummulative frequencies of the classes are $f_1, f_1 + f_2, \cdots, f_1 + f_2 + \cdots + f_k$.

3   Construction of a Frequency Table

The following steps should be followed when constructing a frequency distribution table.

1. Determine the range of the data. The range is the difference between the largest and smallest observations in the dataset (that is, Range = largest observation - smallest observation).

2. Decide the number of classes. Apart from the rule of thumb, several rules were reported in the literature. However, in this article, we consider the three popular rules; Sturges' (1926), Scott's (1979), and Freedman and Diaconis (1981) rules to determine the suitable number of classes.

3. Determine the class width. The class width is the range divided by the appropriate number of classes.

4. Decide the starting point. Obtaining the lower limit of the first class is the starting point, and it should be decided such that the smallest observation in the raw data is included.

5. Determine the classes. When the lower limit of the first class has been decided, the upper limit of the class can be obtained by adding the class width to the lower limit. The remaining lower and upper class limits can be determined by adding the class width until the last class containing the highest observation in the data set is obtained [14, 16].

3.1   Existing Rules for Choosing the Number of Classes (Bins)

The Frequency table is as old as Graunt's work in the 17th century. However, the lack of systematic guidelines for choosing the number of classes has been the existing problem in statistical studies [22]. The problem remained unattended until a research work by Sturges in 1926. Theoreticians such as Doane (1976) and Scott (1979) have developed procedures for obtaining an optimal number of classes; nevertheless, these procedures generally make strong assumptions about the shape of the distribution. Depending on the data's distribution and the aim of the analysis, a different number of classes may be optimal. Therefore, different experiments are often needed to determine a suitable number of classes. The work of Sturges (1926) has immensely contributed to obtaining a suitable number of classes. The Sturges rule was based on using a frequency histogram with k bins where the Binomial coefficient

$\binom{k-1}{i}$, $i = 1, 2, \cdots, k-1$ is its i-th bin count. As k increases, frequency histogram approaches a Gaussian density. The total number of observations is

$$n = \sum_{i=0}^{k-1} \binom{k-1}{i} = (1+1)^{k-1} = 2^{k-1}.$$

Taking the logarithm of both sides, the number of bins formula is

$$k = 1 + \log_2 n,$$

where n is the sample size.

Following this, Doane (1976) argued that the sample data are rarely normal and that Sturges' (1926) rule does not often provide enough classes to accommodate skewed data. Doane (1976) proposed a modification to Sturges' rule by incorporating the skewness statistic.

$$\sqrt{\eta_1} = \frac{\sum (x - \bar{x})^3}{[\sum (x - \bar{x})^2]^{3/2}}.$$

It is a popular statistic used for measuring asymmetry. The standard deviation of $\sqrt{b_1}$ depends only on sample size and becomes small as sample size increases:

$$\sigma \sqrt{\eta_1} = \sqrt{\frac{6 \times (n-2)}{(n+1) \times (n+3)}}.$$

The rule for additional classes is

$$k_e = \log_2 \left(1 + \frac{\sqrt{\eta_1}}{\sigma \sqrt{\eta_1}} \right).$$

Therefore the optimal number of bins formula is [7]

$$k = 1 + \log_2 (n) + \log_2 (\gamma),$$

where

$$\gamma = \sqrt{\frac{(n+1)(n+3) \sum_{i=1}^{n} (x - \bar{x})^2}{\sqrt{6(n-2) \sum_{i=1}^{n} (x - \bar{x})^2}}}.$$

Though both Sturges (1926) and Doane (1976) were among the first researchers to come up with scientific rules for choosing a suitable number of classes, the validity of their works has been questioned by Hyndman (1995). The argument is that the two rules were derived based on a Binomial approximation to the normal distribution approach, which is inappropriate. Hyndman (1995) suggested two alternative rules, the Scott (1979) and Freedman and Diaconis (1981) rules. Scott's (1979) rule was derived using the L_2 risk function parameter selection

criterion. The method is also known as the mean integrated square error (MISE) as a measure of error between the estimated density $\hat{f}(x)$ and the unknown density $f(x)$.

Let $X_1, X_2, \cdots, X_n$ be a real-valued random sample having density function f. The histogram estimator of unknown density function $f(x)$, based on dividing the real-line into bins B_j of width w, is given by

$$\hat{f}(x; w) = \frac{\text{No. of } X_i^{'s} \text{ in } B_j}{nw}, \qquad x \in B_j.$$

Before the bin width, w, is selected scientifically, there is a need to aim at having optimum result. For the reason of simplicity, the optimal bin width is taken to be w_{MISE}, the bin width that minimizes

$$MISE = E \int (\hat{f}(x; w) - f(x))^2 dx.$$

If f is sufficiently regular, then as $w = w_n \to 0$ and $nw \to \infty$ then [9, 21]

$$w_{MISE} \sim \left(\frac{6}{-\psi_2 n} \right)^{1/3} \text{ as } n \to \infty.$$

Using this with the normal density as the unknown density of reference, Scott arrived at a bin width

$$w = \frac{3.49s}{n^{1/3}}.$$

The number of classes (bins) is thus

$$k = \frac{R \times \sqrt[3]{n}}{3.49 \times s},$$

where, R is the range, n is the number of observations, and s is the estimated standard deviation.

The Freedman and Diaconis (1981) rule is a special case of Scott's rule using interquartile range(IQ) as the variability measure, which is less sensitive to outliers compared with the standard deviation. The bin width rule is

$$w = \frac{2 \times (IQ)}{n^{1/3}},$$

where n is the number of observations and IQ is the interquartile range of the data set. This basically amount to choosing the number of bins

$$k = \frac{R \times \sqrt[3]{n}}{2 \times IQ}.$$

Choosing the number of classes is guided by the existing rules; however, care is needed since an appropriate number of classes can reveal the significant features of the data and often provide insights into the information therein [20]. Though computation using some formulas is straightforward, almost all of them yield decimal points for the value of k, it should be rounded up to the next whole number.

Table 1: Typical frequency table for grouped data

	CI		CB		Freq	Cum f
c	l_c	u_c	l_b	u_b	(f)	(cf)
1	$^-h_1$	$^+h_1$	$^-h_1 - \frac{\delta}{2}$	$^+h_1 + \frac{\delta}{2}$	f_1	f_1
2	$^-h_2$	$^+h_2$	$^-h_2 - \frac{\delta}{2}$	$^+h_2 + \frac{\delta}{2}$	f_2	$f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	^-h_k	^+h_k	$^-h_k - \frac{\delta}{2}$	$^+h_k + \frac{\delta}{2}$	f_k	$\sum_{i=1}^{k} f_i$

Table 2: The three methods

Mean (M_e)	Median (M_d)	Midpoint (x^*)
M_{e1}	M_{d1}	$\frac{^-h_1 + ^+h_1}{2}$
M_{e2}	M_{d2}	$\frac{^-h_2 + ^+h_2}{2}$
$\vdots$	$\vdots$	$\vdots$
M_{ek}	M_{dk}	$\frac{^-h_k + ^+h_k}{2}$

Remark: c is the class, k is the number of classes, δ is the unit measurement of the dataset. CI is the class interval, l_c and u_c are the lower and upper class intervals. CB is the class boundary, l_b and u_b are lower and upper class boundaries, ^-h_i and ^+h_i are the lower and upper limits of class i. f_i is the frequency of class i and cf is the cumulative frequency. $i = 1, 2, \cdots, k$. The frequency distribution table is complete if one column of Table 2 is added to Table 1.

In addition, M_{ei} is the mean of observations in class i, M_{di} is the median of class i, which is equal to x_s if the number of observations in the class is odd and $\frac{1}{2}(x_s + x_{s+1})$ if the number of observations in the class is even. The variate X_s is the ordered middle observation within the class and $\frac{1}{2}(^-h_i + ^+h_i)$ is the midpoint of observations in class i, where $i = 1, 2, \cdots, k$. k is the number of classes [17–19].

3.2 Grouping Error

Grouping error is the error introduced when raw data are grouped into a frequency table. Grouping assumes that the scores are uniformly distributed around the midpoint of each class,

and this assumption may not always be valid. Several studies were carried out on grouping error since 1873, but William Fleetwood Sheppard proposed the most popular method for symmetrically distributed data in 1898 [1]. This article addresses this problem using two different approaches: mean and median. These statistics are used in place of the midpoint in each class of the frequency table [18].

4   Root-mean-squared error(RMSE)

The root-mean-square error (RMSE) is an important measure of the performance of an estimator. This estimator can be a model or a method used to estimate a statistic value of a sample or value of a population parameter. In other words, the RMSE measures the accuracy of methods in estimating the actual values of a sample or population.

RMSE is always positive. However, sometimes a rare value zero can be obtained; which indicates a perfect model or method. Generally, a method with the smallest RMSE is the best. Mathematically, the RMSE is given by

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{\vartheta}_i - \vartheta_i)^2},$$

where m is number of samples, $\hat{\vartheta}_i$ is the estimated statistic value of sample i, and ϑ_i is the actual statistic value of sample i, $i = 1, 2, \cdots, m$ [11].

In this study, the est_i are the estimated means from the frequency tables using the three methods: Mean, Median and Midpoint. The formulas used in the computation of the est_i are given in Table 3. Simulation of 100 samples from the normal distribution were used to construct 100 frequency tables using the three different methods. The statistical measure, arithmetic mean, was estimated from the constructed frequency tables ($\hat{\vartheta}$). Meanwhile, the raw arithmetic mean formula was used to compute the actual means (ϑ) of the 100 samples.

Table 3 presents the formulas for computing the statistical measure using the three methods. M_{ei}, M_{di} and x_i^* are respectively the mean, median and Midpoint of observations in class i, $i = 1, 2, \cdots, k$ and k is the number of classes. The sample standard deviation is obtained by taking the square root of the sample variance.

5   Results and Discussion

5.1   Results

The primary interest of this research was two-fold: (a) to propose new methods of constructing frequency distribution table, and (b) to identify a more accurate rule for choosing the number of classes. Two methods of constructing a frequency table, using the mean and median of the class intervals, are proposed and compared with the existing one, which uses the midpoint. Meanwhile, the number of classes was chosen using Sturges, Scott's, and Freedman and Diaconis's (FD) rules. Simulation using 100 samples of size 500 each from the normal distribution with parameters $\mu = 10$, $\sigma = 2$ is performed. In choosing the number of classes for the construction of frequency tables, the Sturges rule gave the smallest number of classes $k = 10$, followed by

Table 3: Formulas for computing the measures of location and variability using the three methods

	Mean$(\bar{x})$	**Variance**(s^2)
Mean(M_e)	$\dfrac{\sum\limits_{i=1}^{k} f_i M_{ei}}{\sum\limits_{i=1}^{k} f_i}$	$\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{k} f_i m_{ei}{}^2 - \dfrac{(\sum\limits_{i=1}^{k} f_i m_{ei})^2}{\sum\limits_{i=1}^{k} f_i}\right)$
Median(M_d)	$\dfrac{\sum\limits_{i=1}^{k} f_i M_{di}}{\sum\limits_{i=1}^{k} f_i}$	$\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{k} f_i m_{di}{}^2 - \dfrac{(\sum\limits_{i=1}^{k} f_i m_{di})^2}{\sum\limits_{i=1}^{k} f_i}\right)$
Midpoint(x^*)	$\dfrac{\sum\limits_{i=1}^{k} f_i x_i^*}{\sum\limits_{i=1}^{k} f_i}$	$\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{k} f_i x_i^{*2} - \dfrac{(\sum\limits_{i=1}^{k} f_i x_i^*)^2}{\sum\limits_{i=1}^{k} f_i}\right)$

Scott's $k = 13$, and the largest is Freedman and Diaconis $k = 18$. Following these, the minimum and maximum values of the mean (see Table 4) and standard deviation (see Table 5) are computed from the constructed frequency distribution tables using the three methods.

Since the results obtained by the simulations are large enough, the tables were given to summarize the results by reporting the minimum and maximum values. To observe the performance of the proposed methods of constructing the frequency tables as well as the rules for choosing the number of classes, we further determined the Root-Mean-Square-Error (RMSE). While the RMSE for the means are displayed in Table 6, Table 7 displayed the RMSE for the standard deviations. The RMSE of the methods can be observed in the rows of Table 6. It shows that Method 2 has the lowest score of RMSE for the mean since the method record the lowest scores in both Sturges and Scott (see the second row of Table 6).

Moreover, the columns of Table 6 displayed the RMSE of the rules of choosing the number of classes. FD rule records the minimum scores in all the methods (see the last column of Table 6). When comparing the RMSE for the standard deviations, the existing method reported the lowest RMSE scores (Table 7). Meanwhile, FD still maintained the minimum RMSE values for the standard deviation (see the last column of Table 7).

5.2 Discussion

This study proposed new methods for constructing the frequency table and identified the most accurate rule of choosing the number of classes. The methods and the classification rules were compared using the RMSE. Furthermore, the mode has not been considered as the appropriate statistical tool to describe the continuous data. It is because the mode is recognized as the observation(s) which occur most frequently. When it involves continuous data, the mode may not be the right tool to represent the centrality of the dataset. The possibility of having two continuous observations with the same decimal points value is rare. Researchers normally

Table 4: Maximum and Minimum means computed from the frequency distribution tables constructed using the three methods

		Sturges	Scott	FD
Method 1	Min	9.79342	9.81864	9.82902
	Max	10.19585	10.19754	10.17481
Method 2	Min	9.81859	9.81714	9.81695
	Max	10.15981	10.16138	10.15997
Midpoint	Min	9.79990	9.78011	9.79908
	Max	10.18961	10.17916	10.15526

Table 5: Maximum and Minimum standard deviations computed from the frequency distribution tables constructed using the three methods

		Sturges	Scott	FD
Method 1	Min	1.79126	1.81429	1.84068
	Max	2.10709	2.12840	2.13141
Method 2	Min	1.81982	1.83603	1.84491
	Max	2.13862	2.15634	2.16035
Midpoint	Min	1.87380	1.86312	1.84159
	Max	2.19125	2.19716	2.20517

Table 6: The root-mean-square error (RMSE) of the actual means and the estimated means from the frequency distribution tables constructed using the three methods

	Sturges	Scott	FD
Method 1	0.0820109	0.0810426	**0.0781552**
Method 2	0.0790262	0.0793549	**0.0787304**
Midpoint	0.0801568	0.0800521	**0.0788342**

Table 7: The root-mean-square error (RMSE) of the actual standard deviations and the estimated standard deviations from the frequency distribution tables constructed using the three methods

	Sturges	Scott	FD
Method 1	0.0975016	0.0800714	**0.0690634**
Method 2	0.0771617	0.0698055	**0.0666188**
Midpoint	0.0712161	0.0677481	**0.0663331**

truncate the values into two or three decimal places to suit purpose. If the observations are not approximated, they cannot be the same. For example, when measuring some distance, the first and second observations are 4.112356 cm and 4.113465 cm, respectively. These values can be truncated into two decimal places, and both can be written as 4.11 cm. These two values were not the same repeated values; they were truncated to ease presentation. As for the median, for normally distributed data, theoretically, it should be the same as the mean. Hence the mean is the most appropriate statistical tool to describe the central location of a symmetrical continuous data.

Moreover, the suitable method that can be use to construct the frequency table according to the result obtained is using Method 2, which substitutes the midpoint by the mean of the classes. The proposed Method 2 can be used as the best alternative for constructing the frequency table for continuous data, which will reduce the error when estimating statistical measures, such as mean and variance, from the frequency table. An improvement can be made by using the FD rule because when the value of n increases, the Sturges rule loses its effectiveness in describing the data by choosing the least number of classes among the three rules. For a larger sample size, n, when the dataset is symmetrical, the alternative will be the Scott and FD rules. It coincides with a comment made by Hyndman in 1995 that "Sturges' rule has probably survived as long as it was because, for moderate n (less than 200), it gives similar results to the alternative such as Scott, and so produces reasonable histograms. However, it does not work for large n."

In reducing the grouping error, future research should propose using different statistics, such as the midrange, trimmed mean, and a random selection, in place of the midpoint. Another interesting topic is to consider a distribution different from normal such as a skewed distribution. The method of minimizing the grouping error by Sheppard correction can also be improved [6].

Acknowledgements

References

[1] Anders, H. On the History of the Correction for Grouping, *Scandinavian Journal of Statistics*. 2001. 28(3): 417–428.

[2] Brase, C. H., and Brase, C. P. *Understanding basic statistics* Boston: CA Houghton Mifflin 2001.

[3] Canning, J. B. Formation of Frequency Distributions, *Journal of the American Statistical Association.* 1926. 21(154): 133–188.

[4] Davies, G.R. The Analysis of Frequency Distributions *Journal of the American Statistical Association.* 1929. 24(168): 349–366.

[5] De Beer, C. F. and Swanepoel, J. W. H. Simple and effective number-of-bins circumference selectors for a histogram, *Statistics and Computing.* 1997. 9: 27–35.

[6] Di Nardo, E. A new approach to Sheppard's corrections. *Mathematical methods of statistics.* 2010. 19(2): 151–162.

[7] Doane, D. P. Aesthetic Frequency Classifications, *The American Statistician.* 1976. 30(4): 181–183

[8] Dogan, N. and Dogan, Determination of the Number of Bins/Classes Used in Histograms and Frequency Tables: A Short Bibliography. *TurkStat, Journal of Statistical Research.* 2010. 7: 77–86.

[9] Freedman, D. and Diaconis, P. On the histogram as a density estimator: L2 theory. *Probability theory and related fields.* 1981. 57(4): 453–476.

[10] Gardiner G. and Gardiner V. *Analysis of Frequency Distributions* Concepts and Techniques in Modern Geography, No. 19. 1979. ISBN 902246984. 2010.

[11] Hyndman, R. J. and Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting.* 2006. 22(4): 679–688.

[12] Hyndman, R.J. *The Problem With Sturges Rule for Constructing Histograms* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.218.7126, 19/09/2018. 1995.

[13] Jeffers, J. N. R. The Importance of Exploratory Data Analysis Before the Use of Sophisticated Procedure *Biometrics.* 1994. 50(3): 881–883.

[14] Kenney, J. F. *Mathematics of Statistics.* Boston: Technical Composition Co. 1939.

[15] Lucien, B. and Yves, R. How many Bins Should beput in a regular histogram, *ESAIM: Probability and Statistics.* 2006. 10: 24–45.

[16] Manikandan, S. Frequency distribution. *Journal of pharmacology & pharmacotherapeutics.* 2011. 2(1): 54–56.

[17] Mohammed, M. B., Adam, M. B., Ali, N., and Zulkafli, H. S. Improved Frequency Tables Measures of Skewness and Kurtosis with Application to Weather Data. *Communications in Statistics-Theory and Methods.* 2020. 51(3): 581–598.

[18] Mohammed, M. B., Adam, M. B., Zulkafli, H. S., and Ali, N. Improved Frequency Table with Application to Environmental Data. *Mathematics and Statistics*. 2020. 8(2): 201–210.

[19] Mohammed, M. B., Zulkafli, H. S., Ali, N., Olaniran, O. R. & Ahmed,H. (2022) New two-way discrete frequency table with application to English Premier League data. *Research in Mathematics*. 2022. 9:1, 1–11.

[20] Sanders, D. H. and Smidt, R. K. *Statistics: A First Course*. Boston: McGraw-Hill 2000.

[21] Scott, D. W. On Optimal and Data-Based Histograms. *Biometrika*. 1979. 66(3): 605–610.

[22] Scott, D. W. *Multivariate Density Estimation: Theory, Practice and Visualisation*. New York: John Wiley & Sons, Inc. 1992.

[23] Sturges, H.A. The Choice of a Class Interval. *Journal of the American Statistical Association*. 1926. 21(153): 65–66.

[24] Wand, M.P. Data-Based Choice of Histogram Bin Width. *The American Statistician*. 1997. 51(1): 59–64.