Descriptive statistics analysis of the variable in the data of toothbrushing simulator system modelling

Salihatun Md Salleh¹,*, Mohamad Norani Mansor¹, Hadirah Hassan², Ainul Husna Mohd Yusoff³, Badrul Aisham Md Zin¹, Musli Nizam Yahya¹, S. T. Wijianto³

¹Mechanical Failure and Reliability (MRPOVE), Department of Engineering Mechanics, Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat 86400 Johor, Malaysia.
²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Seremban 3, 70300, Negeri Sembilan, Malaysia.
³Faculty of Mechanical Engineering, Universitas Muhammadiyah Surakarta, Surakarta 57162, Jawa Tengah, Indonesia.

E-mail: saliha@uthm.edu.my

Abstract. The controller of a system can be designed by constructing the model from the known the input and output. Modelling the dynamic of the system can help to visualize the real behavior of the system to develop the suitable controller parameter. The measured data from the real system need to be analyzed statistically to support that the current system needs a controller to improve the system performance. This study analyzed the normality and linearity of the variables in the research data of Toothbrushing Simulator System Modelling which are Speed (RPM) as the output (Y) and Voltage (V) as the input (X). There were five different data sets with 1000 observations respectively. All the data had been analyzed by using IBM SPSS Statistics 23 in which it will be explained by graphical method of histogram, scatter plot and descriptive measures of coefficient determination between variables. The results for this study turn out that all the data sets were not normally distributed and not linear. Hence, the result from this statistical analysis has proven that the controller development is very crucial for the toothbrushing simulator system to improve the system performance.

1. Introduction

Statistically, it is conventional to assume that the observations or samples collected in a research are normally distributed. Normality is often a maintained assumption in estimation and finite-sample inference [1]. The assumption of normality is useful as significant test of the models and to examine the confidence intervals around parameters [2]. The normality distribution can be examined by using graphical method and numerical method [3]. Compared to numerical method, graphical methods are easy to interpret and intuitive as they visualize the distribution of random variables. Whereas, numerical methods provide objective ways of examining normality as its present summary statistics included skewness, kurtosis and statistical test of normality [4]. Linearity is the ability within a given range to give out the results that are directly proportional to the concentration of the analyte in the test sample [5]. Linearity also can be defined when the input (X) and corresponding output (Y) are tied by the rules of superposition and homogeneity, then the device
or system behavior is said to be linear, otherwise nonlinear [6]. The relationship between two variables should always be made on the basis of a scatter plot that will determine the linearity [7].

The purpose of this study is to analyses the normality and linearity of the variables in the research data of Toothbrushing Simulator System Modelling which are Speed (RPM) as the output(Y) and Voltage (V) as the input(X). This study is primarily concerned on the normality and linearity test of the real time input and output signal of the system to support that the current system needs a controller to improve the system performance.

2. Methodology

2.1 Sample

The purpose of this statistical analysis is to prove that the controller development is very crucial for the toothbrushing simulator. Hence, this study used the research data of Toothbrushing Simulator test rig [8],[9]. There were 1000 samples in each of five different data sets of Pulse Width Modulation (PWM) respectively which were PWM 175, 195, 215, 235 and 255. Whereby each data set consist of 1000 samples of real time input and output data. All input and output signals involved are represented in Fig. 1 and 2.

![Figure 1. Input signal of five data set of PWM 255, 235, 215, 195 and 175.](image-url)
2.2 Normality and Linearity Test

It is interesting to study the distribution of the data referring to the arithmetic mean to characterizing the dynamics of the average. Naive intuition suggests that most data will be randomly distributed about some preferred set of values when the system is in a steady state, as there are many ways in which data can be randomly distributed. For normality, plotting a histogram of the variable of interest is one of the graphical methods for checking the normality of data in which the observed value is plotted against their frequency [10]. The ideal appearance type which is a bell-shape curve will give an indication that the data is normally distributed [11].

In addition, normality test using skewness and kurtosis is another method in accessing the normality of the distribution as it may be more reliable for small and large samples. A measure of asymmetry of the distribution of variables is called as skewness while kurtosis is a measure of the peakedness of a distribution [12], [13]. The SPSS Software will give the value of skewness and kurtosis along with the value of standard error. The formula used in [14] to measure the standard error, SE of skewness and kurtosis are given by equation (2) below.

$$SE \text{ of Skewness} = \sqrt{\frac{3}{n}}$$  \hspace{1cm} $$SE \text{ of Kurtosis} = \sqrt{\frac{24}{n}}$$

where $n$ is the sample size.

Meanwhile, a scatter plot will help in examining the linearity for a data set by plotting the graph of input at x axis versus output at y axis [7]. A scatter plot is a simple method in identifying the relationship between two variables of input(X) and output(Y) [15]. The linear equation can be represented as in equation (2).

$$y = mx + c$$

where $m$ is the slope and $c$ are constant. In order to get the linear relationship, as stated in [6], $c$ value must be zero. All the analysis had been analysed by using a software of IBM SPSS Statistics 23 [16].
3. Result and Discussion

3.1 Normality

A common pattern which is the bell-shaped curve with the zero skew value will indicate the data is normally distributed. The descriptive statistic of skewness value for both variables in all PWM had been presented in Table 1. Standard error was written as 0.077 as this value is consistent at all reading. The skewness for both variables of all the dataset can be seen through the histogram plotted using SPSS in the Table 2. The skewness shows that the distribution is not in normal in most conditions. Overall, the PWM of speed and voltage variables were positive skew value as it was skewed to the right.

However, PWM 255 shown the voltage had a negative skew, as it has left skewed. Table 2 illustrates the distribution of variables in all PWM. Most of the sample mean values were clustered on the left side for speed variable. While, there was a spike plotted in each histogram of voltage as it represents the most common values except for PWM 255 that shown the skewness was heavily to the left. It shows that the data was not normally distributed.

| Data Set | Skewness Statistic | Speed | Voltage |
|----------|--------------------|-------|---------|
| 175      | 2.313              | 1.410 |
|          | 0.077              | 0.077 |
| 195      | 3.547              | 1.814 |
|          | 0.077              | 0.077 |
| 215      | 3.879              | 2.170 |
|          | 0.077              | 0.077 |
| 235      | 4.292              | 3.395 |
|          | 0.077              | 0.077 |
| 255      | 4.110              | -9.214|
|          | 0.077              | 0.077 |

Table 2. Histogram with Normal Curve Plotted of Speed (RPM) and Voltage (V).
(Cont.) Table 2: Histogram of Speed (RPM) and Voltage (V).

| PWM | Speed (RPM) | Voltage (V) |
|-----|-------------|-------------|
| 255 | ![Histogram of Speed (RPM) and Voltage (V)](image) |

3.2 Linearity
Based on Fig. 3 to 7, they visualize that the plotting are not randomly scattered. Thus, it can be concluded that there was no linear relationship between speed and voltage in all dataset of PWM since there were a random pattern in the graph plotted. In addition, these also can be proven by checking the statistical measure, which is coefficient of determination, R² that will measure the strength of the linear relationship between the dependent and independent variables.

![Figure 3. Linearity of Speed (RPM) and Voltage (V) in PWM175.](image)

![Figure 4. Linearity of Speed (RPM) and Voltage (V) in PWM195.](image)

![Figure 5. Linearity of Speed (RPM) and Voltage (V) in PWM215.](image)

![Figure 6. Linearity of Speed (RPM) and Voltage (V) in PWM235.](image)
4. Conclusion

According to the results, all the data sets can be concluded that the distribution of both variables, input and output are not normally distributed. As for linearity, the figures of scatter plot gave evidence that the association between input, voltage and output, speed variables is not linear. It is also proven that the system is a nonlinear. Adding the sample of observations is recommended for further study to imply the normal distribution and non linear system. In conclusion, the proposed testing can be used to determine the system condition to proof that the controller development is very crucial for the toothbrushing simulator system to improvise the system performance.

Acknowledgement

The authors graciously acknowledge the financial support provided by Universiti Tun Hussein Onn Malaysia, UTHM.

References

[1] Bai, J., & Ng, S. (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data. Journal of Business & Economic Statistics, 49-60.
[2] Field, P. A. (2016). Discovering Statistic, 1-20.
[3] Lund, D. A., & Lund, M. (2015). Laerd Statistics. Retrieved 25 October 2018, from Laerd Statistics: https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php
[4] Park, H. M. (2008). Univariate Analysis and Normality Test Using SAS, Stata, and SPSS. University Information Technology Services Center for Statistical and Mathematical Computing Indiana University, 1-38.
[5] LGC Maine Standards. (2017). Retrieved 24 October 2018, from LGC: https://www.mainestandards.com/support/faq-calibration-verification.php.
[6] Electrical System. (September, 2010). Retrieved 24 October 2018, from Electrical System: http://www.skm-eleksys.com/2010/09/linear-and-non-linear-systems-in.html
[7] Schneider, A., Hommel, G., & Blettner, M. (2010). Linear Regression Analysis. Series on Evaluation of Scientific Publications, 776-781.
[8] Ainul H.M.Y, Salleh S. M., Halib N., Taib H., Fathi, M.S. (2018). Analysis of Toothbrush Rig Parameter Estimation Using Different Model Orders in Real-Coded Genetic Algorithm (RCGA). International Journal of Engineering & Technology (IJET), 7, 443-447.

[9] Yusoff A.H.Y., Salleh S.M., Mahzan S., Ahmad S., Taib A., Rahman H.A., Ismail A., Choiron M.A. (2019). Evaluation of Varied Model Order in GA-optimised Parameter Estimation of Toothbrush Rig System. International Journal of Integrated Engineering.

[10] Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology & Metabolism, 487-489.

[11] Das, K. R., & Imon, A. R. (2016). A Brief Review of Test for Normality. American Journal of Theoretical and Applied Statistics, 5-12.

[12] Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restorative Dentistry & Endodontics, 52-54.

[13] Rose, S., Spinks, N., & Canhoto, A. I. (2015). Management Research: Applying the Principle. Test for The Assumption That A Variable Is Normally Distributed, 1-4.

[14] Shukla, Satishprakash, S. (2015). Normality Test. Conference in New Trend in Educational Research, Department of Education, Gujerat University, Ahmedabad, Gujerat, India.

[15] Dastmard, B. (2013). A statistical analysis of the connection between test results and field claims for ECUs in vehicles. Stockholm, Sweden: Royal Institute of Technology School of Engineering Sciences.