Interdisciplinary Research Review

Volume 19, Number 3, Pages 13–19

An Alternative Statistical Approach for the DOE with the Attribute Response

Nichanach Katemukda[1,*]

[1] Department of Industrial Engineering, Faculty of Engineering, Rajamangala University of Technology Rattanakosin, Nakhon Pathom, 73170, Thailand

Abstract

The design of experiment (DOE) with the result as an attribute is generally used. On the other hand, the current DOE method is quite complicated in that the user is supposed to transform the attribute result into the quantitative result, then turn it back again for a conclusion on the final parameter setting. The purpose of the research is to find out the efficient method that has the same result as the DOE. Finally, the research discovered that the logistic regression can be applied instead and get the final result of the parameters setting the same as the original DOE method. According to the users, they do not need to transform the attribute result into the quantitative result; this is the main idea: the users save a lot of time on calculations and can conclude the parameter setting by only interpreting the result from a factorial plot of logistic regression.

Keywords: Design of experiment, Attribute response, Logistic regression, Factorial plot

Article history: Received 10 January 2023, Revised 19 April 2024, Accepted 22 April 2024

The current industrial problem yields attribute data or qualitative data, such as pass/fail, good/bad, and so on, and is also related to the optimization of a parameter or process that has a qualitative result [1]. The design of the experiment needed to be set up to find out the correct parameter setting. Generally, the method that has been used is to transform the qualitative result into a quantitative result [2]. With this method, it takes more time and requires conversion back and forth before interpreting the result. The study aimed to apply other statistical methods that still maintain qualitative results and can be used to determine the factors that need to be set as well as the factors that are influenced by the interaction of factors. The statistical tool proposed is called "logistic regression". The Minitab data blog provided the data for the comparative study of the two methods. This is the original data that was supplied and used in the experiment design, along with the attribute result. The benefit of the research was finding a new method that can significantly reduce the step with the same answer as parameter setting with the DOE method. The comparison of a step reduction between the two methods is shown in Table 1.

Table 1. A comparison step between the existing method and the proposal method

The existing method	The proposal method
Key data in the Minitab sheet.	Key data in the Minitab sheet.
Transform the attribute data to probability.	Perform Logistic regression tool in Minitab software.
Calculate square root of probability and Arcsine later.	Interpret data and conclude for parameter setting.
Perform DOE tool in Minitab software.	
Interpret data and conclude for parameter setting.	

*Corresponding author; e-mail: nichanach.kat@rmutr.ac.th

2. Objective

To compare the decision-making by applying logistic regression instead of the DOE with the attribute response.

3. Literature review

3.1 Design of experiment (DOE)

The Design of Experiments (DOE) is a statistical technique that involves experiment planning and conducting, and findings analyzing and interpreting. It is a subfield of applied statistics used to conduct scientific investigations of a system, process, or product in which input variables (Xs) are changed to see how they affect the measured response variable (Y). In the past, DOE was a very effective strategy for enhancing the quality and dependability of products [3]. Today, numerous sectors employ DOE to help with decision-making for process improvement, new product development, and production procedures. It is employed in administration, marketing, hospitals, pharmaceutical industry, and engineering [4]. Regarding experiments with several elements, factorial designs are generally used when it is important to examine the combined impact of the factors on a response. The most significant of these particular instances has two levels for each of the k factors. These levels can be qualitative, such as the "high" and "low" levels of a factor, or they may also be quantitative, such as two values of temperature or time. A 2^k factorial design is a full replication of such a design, which is called $2 \times 2 \times 2 \times ... \times 2 = 2^k$ observations. An ANOVA (ANalysis Of VAriance) is used to verify the significant influence of each input factor and the interactions on the response factor. The setting of the factor is then considered on the factorial plot for each factor and its interaction. The ARCSINE method was used to transform the response data for the qualitative or attribute response. Users need to do a replicate response data and provide as a proportion result, then apply square root on the inverse sine (ARCSINE) to maintain the normality of the data [5]. Then they follows the ANOVA and draws a conclusion about the input factor setting from the factorial plot.

3.2 Logistic regression

Logistic regression is a technique for estimating the likelihood of an attribute outcome given a dependent variable. Binary logistic regression, often known as two outcomes such as pass or fail, yes or no, and so forth [6] is generally used, including in the industrial sector [7], the service sector [8], etc. The fundamental assumption of logistic regression is that it is identical to linear regression [9]. Residual is the difference between the expected and actual values of the response variable (error). Typically, it is assumed that the residual will have a normal distribution with a mean of 0 and constant variance. The residuals do not correlate with one another. The dependent variable and the residual do not correlate as well as the residual and the response variable. As a result of the outcome being determined by the likelihood that the event will occur, the response variable result will range from zero to one. A probability of one means the fascinating event has a 100% chance of happening, whereas a probability of zero means it is unlikely to happen. Assign probability of interested event is represented by *P(Y)*, and can be written as Equation (1):

$$P(Y) = \frac{e^{b_0+b_1X_i}}{1+ e^{b_0+b_1X_i}} \quad (1)$$

Then the non-interested event is represented by *Q(Y)*, and equation can be written as Equation (2):

$$Q(Y) = 1 - P(Y) \quad (2)$$

b_0 and b_1 are the estimated coefficients that are calculated from the data. X_i is the independent variable and e is a Natural Logarithm.

Logistic regression can be transformed to a logit equation to link the dependent variable to the independent variable. The equation form of logit link can be written as Equation (3).

$$Logit(P) = Log\left[\frac{P}{Q}\right] \quad (3)$$

An important number in the logistic regression is the odds ratio. By comparing P and Q, the odds ratio describes the number of times an event occurs. The odds ratio can be written as Equation (4).

$$Odds\ ratio = \frac{P_y}{Q_y} \tag{4}$$

One of the most important criteria to test for the fit of the logistic regression model is the likelihood value. The statistical software will calculate the likelihood value and transform it into a Chi-square distribution test with the degree of freedom of the independent variable. The main hypothesis of the Chi-square test is that "all coefficients of the logistic regression model are zero." While the null hypothesis was rejected, the conclusion is that "logistic regression has a model." The statistical software also provided the P-value for making a decision on accept or reject the null hypothesis based on the type I error percentage set up. Type I error is typically 5%.

3.3 A challenge of attribute response

The industrial sector and the service sector are all facing the result as an attribute. For example, the product cracks or does not crack, which is defined as an attribute result. However, almost everyone avoids using attributes as the result for fine-tuning any process setting. In this case, they define the length of the crack as represented by the number [10], even though sometimes they may fully break up. On the other hand, the end result as attributed may represent the real situation of what we need to monitor. Such as the component crack in the electronics component, as long as the crack is defined, there is an impact on the quality because there are so small [11]. Analyzing the data on the attributes is more complicated than using quantitative data [12]; thus, any attribute data always are converted to quantitative data for easier decision-making [13]. To maintain the end result as attributes and propose a method that the end result is the same rather than to convert fort and back the data, this research will explore any method that can be applied.

4. Methodology

The research is a comparison study between the DOE with attribute response and the logistic regression. The result of the parameter setting of the input factor is supposed to be the same, and the time saving is a key part of the study for the calculation and result interpretation.

The hypothesis of the research: Logistic regression can be used to set up the input factor with the same result as DOE's attribute response.

Based on the research hypothesis, the data and results from the Minitab were used for a comparison study [14] [15]. The data is 3 factors with 2 levels and 10 replications. The design is $2^3 = 8$ Experiments with 10 replications mean $8 \times 10 = 80$ Runs.

Table 2. Raw data and results of experiment for a comparison study

Input Factor			Attribute Factor (OK/NOK)									
Factor 1	**Factor 2**	**Factor 3**	**Output factor with 10 Replicates**									
-1	-1	-1	NOK	OK	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK
1	-1	-1	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK
-1	1	-1	OK	NOK	OK	OK	OK	OK	OK	OK	OK	OK
1	1	-1	NOK	OK	OK	OK	OK	NOK	NOK	NOK	NOK	NOK
-1	-1	1	NOK	OK	OK	NOK	NOK	OK	OK	OK	OK	OK
1	-1	1	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK	NOK
-1	1	1	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
1	1	1	OK	OK	OK	OK	NOK	NOK	OK	OK	NOK	NOK

Table 3. Data transforming from Attribute to quantitative

Input Factor			Proportion (p)	Square root (p)	Arcsin (Sqrt (p))
Factor 1	Factor 2	Factor 3		Sqrt (p)	
-1	-1	-1	0.10	0.32	0.32
1	-1	-1	0.00	0.00	0.00
-1	1	-1	0.90	0.95	1.25
1	1	-1	0.40	0.63	0.68
-1	-1	1	0.70	0.84	0.99
1	-1	1	0.00	0.00	0.00
-1	1	1	1.00	1.00	1.57
1	1	1	0.60	0.77	0.89

The data transformation is an additional step for the DOE. In contrast, the logistic regression is not required. For logistic regression, the output "OK" is represented by 1 and the output "NOK" is represented by 0. The input factor put Low (L) and High (H) to represent -1 and 1. Then the logistic regression was performed and the factorial plot was provided in the result.

5. Result and discussion

5.1 DOE with attribute response and factorial plot result

Table 4. Summary result of DOE

Term	Coef	SE Coef	P-Value
Constant	0.71	0.06	0.000
Factor 1	-0.32	0.06	0.006
Factor 2	0.38	0.06	0.003
Factor 3	0.15	0.06	0.071

R-Sq = 94.84% and R-Sq adjust = 90.96%. There is no interaction between the factors.

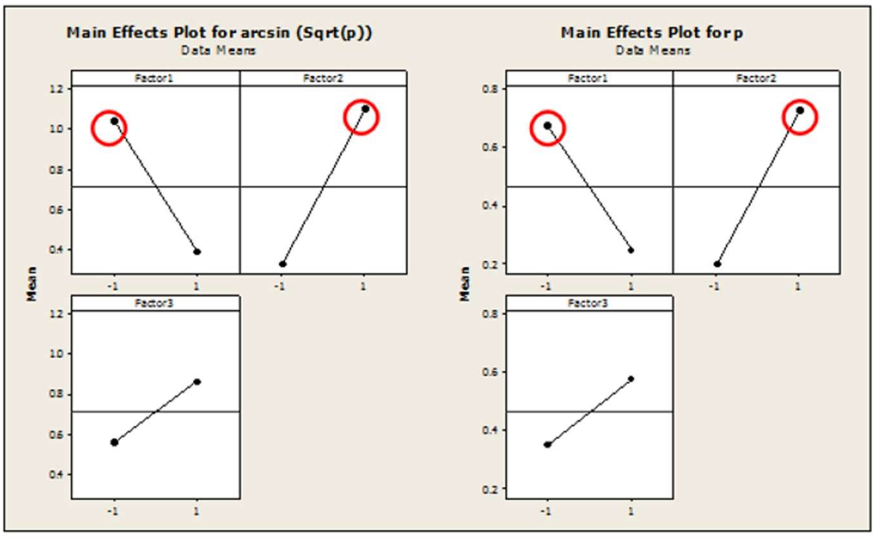

Figure 1. Factorial plot of DOE

Note. Optimizing Attribute Responses using Design of Experiments (DOE), Part 1. From https://blog.minitab.com/en/statistics-in-the-field/optimizing-attribute-responses-using-design-of-experiments-doe-part-1, by Jayakumar, M. Copyright 2022 by Minitab, LLC.

The conclusion based on the result analysis is to set up the maximum attribute response (a high "OK" rate). The analysis has shown that parameters for factors are set at the following levels:

Factor 1 setting at -1 (Low level).
Factor 2 setting at 1 (High level).
Factor 3 setting at 1 (High level, Factor 3 is less significant and might not have a high impact if changed to -1).

5.2 Logistic regression and factorial plot result (A proposal method)

Table 5. Summary result of DOE

Term	Coef	SE Coef	P-Value
Constant	0.83	0.59	0.000
Factor 1	3.82	1.11	0.001
Factor 2	-4.32	1.12	0.000
Factor 3	-1.83	0.69	0.008

R-Sq = 49.60% and R-Sq adjust = 46.88%. There is no interaction between the factors.

According to the P-value of each factor, it is lower than 0.05, which means all factors influence the response at a 95% confidence interval.

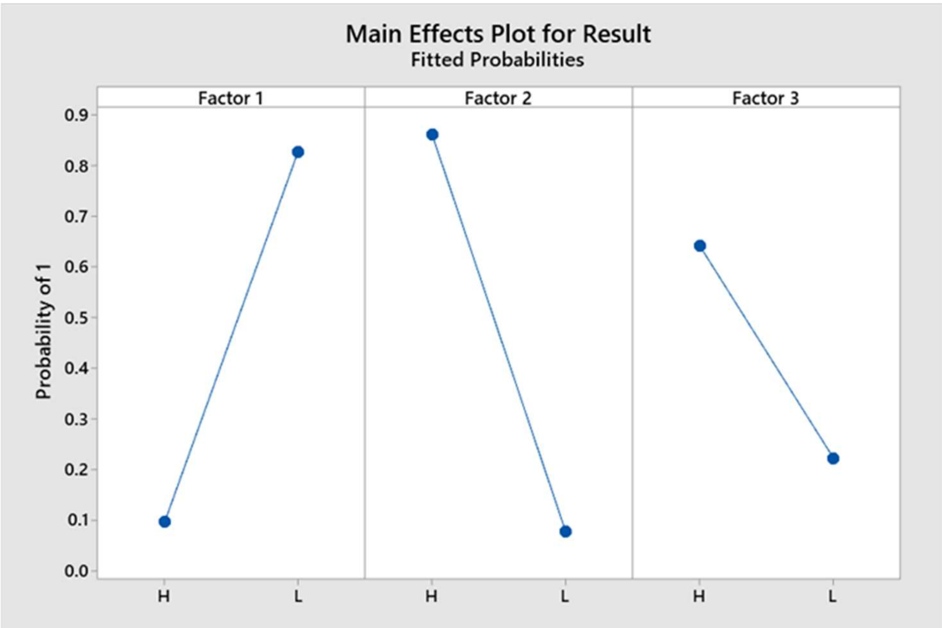

Figure 2. Factorial plot of Logistic regression

Note. Optimizing Attribute Responses using Design of Experiments (DOE), Part 2. From https://blog.minitab.com/en/statistics-in-the-field/optimizing-attribute-responses-using-design-of-experiments-doe-part-2, by Jayakumar, M. Copyright 2022 by Minitab, LLC.

The conclusion based on the result analysis is to set up the maximum attribute response (a high "OK" rate). The analysis has shown that parameters for factors are set at the following levels:

Factor 1 setting at -1 (Low level).
Factor 2 setting at 1 (High level).
Factor 3 setting at 1 (High level).

The difference between DOE with attribute response and logistic regression is factor 3. At a 95% confidence level, the DOE indicates that factor 3 has no effect. In contrast, logistic regression suggests factor 3 has a significant impact at a 95% confidence interval. In terms of the different values of the R-Sq between the two methods, the R-Sq indicated how well dependent variables can explain the independent variable (or output of the system). The point of view for the practitioner is that they need to know how to set parameters specific to the “low” or “high” settings of each factor. Anyway, the conclusion of the setting for the system can be reached with the same result. This is suitable for the practitioner in the factory due to the calculation of the time they are saving. Regarding the logistic regression, users key in the result and run the statistical test, then the conclusion can be made. On the other hand, DOE needed to transform the data before running the statistical test. This step is taking more time than the logistic regression method. Finally, after setting the model as suggested, the result must be validated with at least 30 samples in the final step.

6. Conclusion

Referring to the research hypothesis, an alternate statistical method, called logistic regression, has the same result as the system setting up of the factor with the DOE's attribute response. This proposal is applicable to the 2^k factorial and produces a binary response. Therefore, the new method is beneficial in terms of time savings and step reduction. Moreover, this is a suitable method for the practitioner who needs to set up any system, machine, and so forth.

Acknowledgment

The researcher is grateful to the Industrial Engineering Department, Faculty of Engineering, Rajamangala University of Technology Rattanakosin for supporting the publication of this study.

References

[1] El-Haik, B., and Roy, D. M. (2005). Service design for six sigma: a roadmap for excellence. New Jersey: John Wiley & Sons.

[2] Anderson, V. L. and McLean, R. A. (1974). Design of experiments: a realistic approach (vol 5). CRC Press.

[3] Durakovic, B. (2017). Design of experiments application, concepts, examples: State of the art. Periodicals of Engineering and Natural Sciences, 5(3), 421-439. DOI: http://dx.doi.org/10.21533/pen.v5i3.145

[4] Paulo, F. and Santos, L. (2017). Design of experiments for microencapsulation applications: A review. Materials Science and Engineering: C, 77(August), 1327-1340.

[5] Mathews, P. G. (2004). Design of Experiments with MINITAB. Quality Press.

[6] Harrell F. E. (2015). Binary logistic regression (2nd ed.), In: Harrell FE (Ed.). Regression modeling strategies (pp. 219-274). Cham: Springer.

[7] Mahajan, R., Garg, S., Sharma, P. (2014). An illustration of logistic regression technique: A case of processed food sector. International Journal of Business Excellence, 7(5), 659-676.

[8] Su, Q., Shao, P. and Ye, Q. (2012). The analysis on the determinants of mobile VIP customer churn: A logistic regression approach. International Journal of Services Technology and Management, 18(1-2), 61 - 74.

[9] Montgomery D. C., Peck E. A. and Vining G. G. (2006). Introduction to Linear Regression Analysis (4th ed.). New Jersey: John Wiley & Sons.

[10] Aabid, A., Hrairi, M., and Ali, J. S. M. (2020). Optimization of composite patch repair for center-cracked rectangular plate using design of experiments method. Materials Today: Proceedings, 27, 1713-1719.

[11] Katemukda, N. (2021). Applying the logistic regression for the quality of microelectronics product. Journal of Applied Statistics and Information Technology, 6(2), 25-35.

[12] Pena, J. C., Nápoles, G., and Salgueiro, Y. (2022). Normalization method for quantitative and qualitative attributes in multiple attribute decision-making problems. Expert Systems with Applications, 198, 116821.

[13] Rao, R., and Lakshmi, J. (2021). R-method: A simple ranking method for multi-attribute decision-making in the industrial environment. Journal of Project Management, 6(4), 223-230.

[14] Optimizing Attribute Responses using Design of Experiments (DOE), Part 1, Available online in August 20, 2012, Available from: https://blog.minitab.com/en/statistics-in-the-field/optimizing-attribute-responses-using-design-of-experiments-doe-part-1

[15] Optimizing Attribute Responses using Design of Experiments (DOE), Part 2, Available online in August 23, 2012, Available from: https://blog.minitab.com/en/statistics-in-the-field/optimizing-attribute-responses-using-design-of-experiments-doe-part-2