Linear model analysis for ordinal response in a mixture experiment

W Andani, A H Wigena*, U D Syafitri
Department of Statistics, Bogor Agricultural University, Bogor, Indonesia
Level 4 Wing 22 Meranti Street, Dramaga, 16680

*ajiwigena@ymail.com

Abstract. Mixing of ingredients is commonly applied in food and beverage industry activities such as making cookies. An applicable method to get an optimal formula for mixing ingredients is a mixture experiment. The mixture experiment is an experiment in which levels are the proportions of the factors which are ingredients. The standard mixture design is developed based on linear models. However, the response for organoleptic test in food industry is ordinal scale. The goal of the research was to evaluate some mixture designs of linear model for ordinal logistic model which is more appropriate model for ordinal scale. The mixture designs were the D-optimal and I-optimal design. Based on the D-optimality criterion of ordinal model and D-efficiencies of the D-optimal design compared to other designs, the D-optimal design was the best design.

1. Introduction
A mixture experiment is an experiment that suit for mixing ingredients. The levels of each ingredients are proportions which lie between zero and one and the sum of the level of all ingredients is one [1]. These cause any change in a proportion of a ingredient will change the proportions of other ingredients and also the design region becomes a \((q-1)\)-dimensional regular simplex [2]. These changes will cause multi-collinear in the mixture experiment. Consequently, the standard polynomial model cannot be used. Scheffé [3] developed several models of mixture experiments by reparameterization of standard polynomials called canonical polynomial models.

The designs of mixture experiments that commonly used are a simplex-lattice and a simplex-centroid design [1]. Unlike other designs, the designs of mixture experiments depend on model used. These designs are easy and simple to construct when the experimental region is a regular simplex. However, the composition of design is more difficult to construct when the experimental region is the irregular one. One alternative solution to solve the problems is constructing the design by an optimal design approach. The optimal design is a branch of an experimental design in which the researchers have flexibility to construct the design based on the real problem [2]. Optimal designs are seeking a design based on a certain criterion.

A criterion which the largest attention in literature is D-optimality criterion [2]. D-optimality criterion focuses on the accuracy of parameter estimation obtained by maximizing the determinant of the information matrix from variance-matrix of the estimator. Another design that has an attention recently is an I-optimal criterion which is based on precision of prediction variance [4]. From computation point of view, I-optimality criterion is more complex than D-optimality criterion.
The Scheffé models are widely used in practice. These models are based on linear model which the responses are numerical data. Most of designs which had ordinal responses were analyzed based on linear model [5]. Lately research which was done by [6] discovered that the design based on linear model was biased if the response of the design was an ordinal scale.

This study focused on evaluating the mixture designs based on linear model that would be used for ordinal response. There were two designs that generated based on linear model, and then the designs would be evaluated by D-optimality criterion of ordinal response.

2. Methods
The study case used in this paper was mixing three ingredients; namely $x_1$, $x_2$, and $x_3$. The constraints of the proportion of each ingredient are found in Table 1. Producing a mixture of ingredients that provide optimal flavor was the ultimate purpose of this study. Taste was the response in this study that would be measured in a likert scale of 1-3. The likert scale shows that 1 represents for unpleasant, 2 represents for moderate, and 3 represents for tasty.

| Ingredients | Constraints                |
|-------------|----------------------------|
| 1           | $0.0033 \leq x_1 \leq 0.008$ |
| 2           | $0.4 \leq x_2 \leq 0.5$    |
| 3           | $0.4 \leq x_3 \leq 0.5$    |

The constraints affected the experimental region. As the range of constraints were very small so the experimental region was a dot in a whole simplex. However, if the dot is enlarged, the experimental region is a trapezoidal. Figure 1 shows the experimental region of the case.

The steps of the research on this paper are:
- The model used in this paper was a second-order cheffe model. The model is written as

\[
\text{Quadratic} : y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i x_j + \epsilon \tag{1}
\]

The response of this study was consumer preference which was measured by likert scale form 1-6.
- Creating D and I-optimal design using the coordinate exchange algorithm.
D-optimal design is seeking a design which minimizing the determinant of the inverse of information matrix or maximizing the determinant of information matrix. D-optimality criterion is a criterion which focuses on the accuracy of parameter estimation and is written as

\[
D = |(X'X)^{-1}| = |X'X| \tag{2}
\]

where \( X \) is a \((n \times p)\) model matrix with \( p \) is the number of parameter model and \( n \) is the number of observations. On the other hand, I-optimal design is seeking a design that minimizing the average prediction of variance. In contrary with D-optimality criterion, I-optimality criterion focuses on precision on prediction variance instead of precision on parameter estimation. The formula of I-optimality criterion is written as

\[
I_{opt} = \frac{1}{\int_{\mathbb{R}^{n-1}}^d} tr\left(\left(\int \frac{1}{f(x)}f(x)dx\right)^{-1}\left(\int X'X'f(x)dx\right)\right) \tag{3}
\]

To find D and I-optimal design, coordinate exchange algorithm for mixture experiments is used [7, 8, 9]. In this study, 12 runs were set.

- Counting D-optimality criterion of both designs based on linear and ordinal model.

The two designs were optimal based on D-optimality and I-optimality criterion, respectively. The question is if the design is optimal based on linear model, which is the design also optimal for ordinal response? D-optimality criterion based on linear model was calculated by Equation (2) while the D-optimality criterion based on ordinal model was obtained by Equation (4). Ordinal response belongs to generalized linear model (GLM) and the information matrix of GLM is stated as follows:

\[
M(i_n) = X'WX \tag{4}
\]

with \( W \) is the weight matrix [10]. According to [6], the \( W \) matrix for each observation \( i \) is as follows:

\[
W_i = \begin{bmatrix}
           u_{i1} & -b_{i2} & 0 & \cdots & 0 \\
           -b_{i2} & u_{i2} & -b_{i3} & \cdots & 0 \\
            0 & -b_{i3} & \ddots & \ddots & 0 \\
            \vdots & \vdots & \ddots & \ddots & \ddots \\
            0 & 0 & 0 & \ddots & -b_{ij-1} u_{ij-1}
\end{bmatrix} \tag{5}
\]

where

\[
u_{i} = g_{ij}^2 \left( \pi_{i-1} + \pi_{i+j+1}^{-1} \right) \tag{6}
\]

\[
b_{ij+1} = -g_{ij} g_{ij+1} \pi_{j+1}^{-1} \]

where

\[
g_{ij} = (g^{-1})(\eta_{ij}), \quad (g^{-1})(\eta_{ij}) = (g^{-1}(\eta_{ij})) (1 - (g^{-1}(\eta_{ij}))),
\]

\[
(g^{-1})(\eta_{ij}) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})},
\]

\[
\eta_{ij} = \log \frac{p(y \leq j)}{p(y > j)} = \alpha_j + \beta' x.
\]

To calculate \( u_{ij} \) and \( b_{ij} \), the parameter model of parameter model of ordinal model should be estimated in advance. Ordinal model was used in this study was the proportional odds models. As no
information about the parameter model before hands, the parameter model was assumed following Uniform distribution between $a$ and $b$ $(a,b)$.

- Determining D-efficiency of two designs for evaluating designs. D-efficiency is achieved by comparing the information matrix of two designs, for example the first design with the $M_1$ information matrix and the second design with the $M_2$ information matrix. D-efficiency is calculated by

$$D_{eff} = \left( \frac{|M_1|}{|M_2|} \right)^{\frac{1}{p}}$$  (7)

where $p$ is the number of parameters in the model. If D-efficiency $> 1$, the first design is better than the second design.

3. Results and Discussion

3.1. Mixture Experiment Models
Unlike other designs, in mixture experiments, the model should be stated in advance because the design points depend on the model. The model used in this research was second-order Scheffe model. The second-order Scheffe model with three three ingredients is written as

$$y = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_1 x_2 + \hat{\beta}_5 x_1 x_3 + \hat{\beta}_6 x_2 x_3$$  (8)

3.2. Optimal Design
The D-optimal and I-optimal design of 12 runs are shown in Table 2. Both design resulted seven distinct design points. Five design points of D-optimal design are same as five design points of I-optimal design. Figure 2 shows the final points of D-optimal and I-optimal design. The design points of the D-optimal design lie on the vertices and mid of the edges. Unlike the D-optimal design, the I-optimal design consisted of two design points which lie not on the vertex or mid of the edge. A design point lies on the centroid and another design point lie on the edge. Furthermore, the distribution of the replications of design points was different between both optimal designs. In D-optimal design, the distribution of the replications was almost equally among the design points. However, it was different in I-optimal design.

| Design | Design Points | Design | Design Points |
|--------|---------------|--------|---------------|
|       | $X_1$ | $X_2$ | $X_3$ | $X_1$ | $X_2$ | $X_3$ |
| D-optimal | 0,0033 | 0,4967 | 0,5000 | 0,0033 | 0,4977 | 0,499 |
|         | 0,0033 | 0,4967 | 0,5000 | 0,006 | 0,494 | 0,5 |
|         | 0,0033 | 0,5000 | 0,4967 | 0,006 | 0,494 | 0,5 |
|         | 0,0060 | 0,4940 | 0,5000 | 0,006 | 0,497 | 0,497 |
|         | 0,0060 | 0,5000 | 0,4940 | 0,006 | 0,5 | 0,494 |
|         | 0,0080 | 0,4920 | 0,5000 | 0,006 | 0,5 | 0,494 |
|         | 0,0080 | 0,4920 | 0,5000 | 0,008 | 0,492 | 0,5 |
|         | 0,0080 | 0,4960 | 0,496 | 0,008 | 0,496 | 0,496 |
|         | 0,0080 | 0,5000 | 0,492 | 0,008 | 0,5 | 0,492 |
|         | 0,0080 | 0,5000 | 0,492 | 0,008 | 0,5 | 0,492 |
| I-optimal | 0,0080 | 0,5000 | 0,492 | 0,008 | 0,5 | 0,492 |
3.3. Design Evaluation
To evaluate D-optimal design compared to I-optimal design, D-efficiency was used.

3.3.1. D-efficiency based on Linear Model. D-optimality criterion of D-optimal design was larger than D-optimality of I-optimal design. The D-efficiency of D-optimal design compared to I-optimal design was 1.342. It means that the D-optimal design performed well compared to the I-optimal design in terms of precision parameter model. It make sense because the D-optimal design was constructed using the D-optimality criterion which focus on precision parameter model.

| Design  | D-optimality criterion | D-efficiency |
|---------|------------------------|--------------|
| D-optimal | 2.26195E-37           | 1.342        |
| I-optimal | 3.85585E-38           |

*Comparison

3.3.2. D-efficiency based on Nonlinear Model. In this study, the initial estimated parameter value was calculated using the Bayes method approach assuming that $\beta_i \sim U(a, b)$. It is better to assume a (prior) statistical distribution for parameters instead of a single point value [6]. The intercept parameters were set 0 and 2, respectively. The basic design used in this case was generated by D-optimality criterion. The uniform distributions were generated 10000 times and each time was calculated the determinant of the information matrix of the proportional odds model. The lower and upper bounds of the largest determinant was pointed as the estimated of parameter model. The lower (a) and upper (b) bound of Uniform distribution of each parameter and the best value can be seen in Table 4. The determinant based on the values of Table 4 was 2.343359E-30.

| Parameter | Distribution | Parameter values |
|-----------|--------------|------------------|
| $\alpha_1$ | 0.00000      | 0.000000         |
| $\alpha_2$ | 2.00000      | 2.000000         |
| $\beta_1$ | $U \sim (-10, 10)$ | 6.556114 |
| $\beta_2$ | $U \sim (-10, 10)$ | 0.993388 |
| $\beta_3$ | $U \sim (-10, 10)$ | -1.578350 |
\[
\begin{align*}
\beta_4 & \quad U \sim (-10, 10) \quad -5.037284 \\
\beta_5 & \quad U \sim (-10, 10) \quad -6.560227 \\
\beta_6 & \quad U \sim (-10, 10) \quad 3.996400
\end{align*}
\]

D-optimality criterion for proportional odds model of D-optimal design for linear model was 7.098932e-38 meanwhile D-optimality criterion of I-optimal design for linear model was 1.700449e-38. The D-efficiency of the D-optimal design compared to I-optimal design was 1.20. It means that the D-optimal design was also well performed compared to I-optimal design in terms of precision of parameter model for ordinal model. The number of parameter in ordinal model is 8 while the number of parameter in linear model is 6. Two extra parameter in ordinal model are the intercepts (\( \alpha_1 \) and \( \alpha_2 \)).

**Table 5.** D-optimal criterion dan D-efficiency.

| Design     | D-optimality criterion | D-efficiency |
|------------|------------------------|--------------|
| D-optimal  | 7.098932e-38           | 1.20         |
| I-optimal  | 1.700449e-38           |              |

a based design

4. Conclusion

In conclusion, the D-optimal design of linear model was better than the I-optimal design for linear model as well as ordinal model in terms of precision of parameter model. In addition, the D-optimal design of linear model could be used as well for ordinal model. On other words, Furthermore, D-efficiency calculations based on linear and nonlinear models show the same results in determining the optimal design. In this case the optimal design is D-optimal design. The results of the design evaluation have not been able to show the bias from the misuse of the linear model for ordinal scale responses.

Acknowledgments

The authors thank to Ministry of Research, Technology, and Higher Education who gave funding for the research. The authors also thank to partner company for research collaboration.

References

[1] Cornell J A 2002 *Experiment with Mixtures Designs: Designs, Models and The Analysis of Mixture Data, 3rd edition* (New York: A John Wiley and Sons, Inc. Publication)

[2] Atkinson A C, Donev A N and Tobias R D 2007 *Optimum Experimental Designs, with SAS* (Oxford: Oxford University Press)

[3] Scheffé H 1958 Experiments with mixtures *Journal of the Royal Statistical Society* (Ser B-20:344–360)

[4] Goos P and Jones B 2011 *Design of Experiments: A Case Study Approach* (Wiley)

[5] Piepel G F and A Khuri 2006 50 years of mixture experiment research: 1955-2004 *Response Surface Methodology and Related Topics*

[6] Mancenido M V 2016 Categorical Responses in Mixture Experiments [Disertation] Arizona State University, United State

[7] Meyer R K and Nachtsheim C J 1995 The coordinate-exchange algorithm for constructing exact optimal experimental design *Technometrics* 37: 60–69.

[8] Piepel G, Cooley S and Jones B 2005 Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm *Engineering* 17: 579–594.

[9] Smith W 2005 *Experimental Design for Formulation* (Philadelphia: Siam)