Thailand Statistician
July 2010; 8(2) : 123-142
http://statassoc.or.th
Contributed paper

Comparison of Model Selection Criteria for Classical Designed Experiments

Natthasurang Yasongnoen [a], John J. Borkowski [b] and Kamon Budsaba* [a]

[a] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand.

[b] Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717 USA

*Author for correspondence; e-mail: kamon@mathstat.sci.tu.ac.th

Received: 7 December 2009
Accepted: 10 May 2010

Abstract

Model selection procedures play an important role in many researches especially quantitative research. In science research and Biological sciences, the analysis and model selection of experiments are often used. In this article, we studied a 2^2 and 3^2 factorial design. Only qualitative factors were considered. These designs are used in this study for model selections by using Akaike's Information Criterion (AIC), using p-value from ANOVA (P-Value Method), and Stepwise Procedure. In additional, there are two types under this study of the model specified: Type I and Type II corresponding to a model having "relatively few" important effects and a model having "many" important effects respectively**.** The results of this study will be presented for the model selection performance and for comparisons of selected models that match between methods. For each design and types of models, a model is selected and the proportion of selected models that are the true or "correct" model is recorded. These results will point to which model selection methods select the correct model more often than others. Furthermore, the final models of two model selection methods will be compared using the proportion of selected models that match for each of the 500 simulated replications. These results indicate which methods tend to select the same or

different final model. The conclusions for a 2^2 factorial design, the model selection methods that selected the correct models and the model that match related with replicates per experimental treatment and the value of parameter for both model types. For a 3^2 factorial design, the P-value and Stepwise Methods selected the corrected model often the other method and also selected the same model more often than the other pair of method for all replicates per experimental treatment (n=2, 4), all values of the parameters and both model types.

Keywords: AIC, Akaike's information criterion, factorial design, model selection criteria, P-Value, stepwise.

1. Introduction

Model selection methods are procedure selecting the variable to be in statistical model and function that show relationship between independent and dependent variables. The model selections are commonly used in the presence composed of using p-value from an ANOVA, stepwise procedure and the Akaike's Information Criterion (AIC and AICc). Akaike's Information Criterion is alternative method for comparing models based on information theory. In this article, we presented the comparison of model selected of three model selection methods. The analysis is driven by the principles of effect heredity or heredity principle [1] to define the set of potential models that could result from application of that model. The possible models that can be fitted from the simulated experimental data for each design based on heredity principle.

The goals of this study are as follows:

(1) To determine if the final conclusions reached that is, the final model selected are the true models for each of the three approaches to model selection resulting from the analysis of experimental design.

(2) To assess if the choices of the model effects on model selection are similar or different across the three approaches to model selection resulting from the analysis of experimental design data.

For each goal, there are also two related objectives: For models with relatively few important effects, the models contain less than or equal to 50% of the effect or parameter terms (that do not include β_0 or μ) in the full model, called model type I. For models with many important effects, the model contains more than to 50% of the effect or parameter terms in the full model, called model type II (see Table 1). Here, we

investigate by using simulation study. We conclude two results-the proportion of correct model selected and the proportion of matching models selection.

2. Methodology

2.1 Model Selection using p-values from an ANOVA (P-Value Method)

We shall describe the algorithm for model selection based on the p-value procedure in terms of t-test results for individual model effects. The t-statistics are computed as $t^* = \frac{\hat{\beta}_i}{s\{\hat{\beta}_i\}}$ where $s\{\hat{\beta}_i\}$ is the standard error of the regression coefficient $\hat{\beta}_i$. For each statistic t^* is an associated p-value. The p-value procedure routine first fits a "full" regression model which contains the full set of potential model effects determined by the researcher. Then consider the set of p-values from the ANOVA. If the largest p-value is greater than a predetermined α (a non-significant effect) assuming "weak heredity" or "strong heredity", the effect is removed from the model.

2.2 Akaike's Information Criterion (AIC)

Akaike established a relationship between the maximum likelihood and the Kullback-Leibler information [2]. Furthermore, he developed an information criterion for comparing models known as Akaike's information criterion (AIC). It is defined as AIC = -2(log-likelihood) +2*K*, where *K* is the number of estimated parameters included in the model. In the special case in which analyses are based on least square estimation assuming a normally distributed error, AIC can be expressed as AIC = $n \log(\hat{\sigma}^2) + 2K$ where $\hat{\sigma}^2 = \frac{\Sigma(\hat{\varepsilon}_i)^2}{n}$, and the $\hat{\varepsilon}_i$ are the estimated residuals from the fitted model. In this case, *K* equals the total number of parameters in the model, including the intercept and $\hat{\sigma}^2$.

If the design size *n* is small compared to *K*, the second-order Akaike information Criterion (AICc) is calculated as the following:

$$\text{AICc} = -2(\text{log-likelihood}) + 2K + \frac{2K(K+1)}{(n-K-1)}$$

$= \text{AIC} + \frac{2K(K+1)}{(n-K-1)}$, where *n* is the sample size. It is recommended that AICc should be used when *n/K* is less than 40 for the model with the largest value of *K*. As *n* increases, the AICc values converge to the AIC.

2.3 Model Selection using AIC

When using AIC as a model selection criterion, the researcher proposes a set of candidate models with the goal of selecting the best from that candidate set. Suppose there are *R* models in the candidate set of models. The AIC values are then calculated for each of the *R* models. Unfortunately, there are no clear interpretations of individual AIC values, and they are not directly used to compare models. However, two measures associated with AIC or AICc that can be used to compare models are the delta AIC (denoted ΔAIC) values and the Akaike model weights (denoted w_i). ΔAIC is calculated as $\Delta_i = \text{AICi} - \min(\text{AIC})$ where AIC_i is the AIC value for model i, and min (AIC) is the minimum of the *R* different AIC_i values. The simple rules of thumb to assess the relative worth of a model are as follows: Any model which has $\Delta_i \leq 2$ infers strong evidence supporting that model. $4 \leq \Delta_i \leq 7$ indicates that the model has considerably less support while $\Delta_i \geq 10$ indicate the model has essentially no support [2].

The Akaike model weight model offers another measure of weight of evidence for each model. The Akaike model weight (w_i) for model *i* (*i* = 1,2,…*R*) is represented as the ratio involving Δ_i and the whole set of ΔAIC values for the *R* candidate models :

$$w_i = \frac{\exp(-\Delta_i / 2)}{\sum_{r=1}^{R} \exp(-\Delta_i / 2)}$$

The weights sum to 1, and weight w_i is interpreted as the probability model *i* is the best among the set of candidate models.

Note that if we need to evaluate the importance of the individual model effect, the researcher can also calculate an effect weight. The effect weight is sum of all AIC or AICc weights for the subset of candidate models that contain that effect [2-4].

2.4 Model Selection using a Stepwise Procedure

The stepwise procedure is a modification of forward selection where at each step all independent variables entered into the model previously are reassessed via their partial F statistics. This procedure requires two cutoff values, one for entering (say F_{in})

and one for removing (say F_{out}) variables. An independent variable added at an earlier step may now be redundant because of the relationship between it and independent variables now in the equation. If the partial F statistic for a variable is less than an F_{out} value, that variable is dropped from the model [5].

2.5 Models Studied

The analysis of the models in this study can be formulated in the form of an effects model for a 2^2 and as a regression model for a 3^2 two-factorial design. The "full" models (assuming $\varepsilon \sim N(0,1)$) are defined to be

- for 2^2 design: $y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon$
- for 3^2 design: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$

Two types (Type I and Type II) of model will be considered when a model is specified.

- Type I is a model having "relatively few" important effects.
- Type II is a model having "many" important effects.

The specific Type I and Type II models are shown in Table 1

Table 1. Two types of model for each design

Design	Type I	Type II
2^2	$y = \mu + \alpha_i + \varepsilon$	$y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon$
3^2	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$

2.6 Possible Fitted Models

When considering any model selection method, it is necessary to define the set of potential models that could result from application of that method. For example, one might allow a model that contains any possible subset of terms of the full model. In this study, the possible models that can be fitted from the simulated experimental data for each design based on heredity principles. Specifically, two classes of the models based on the strong and weak heredity principles [1] were considered. In this article,

- Weak heredity implies that an interaction ($x_i x_j$) effect can be in the model only if x_i or x_j are in the model, and a quadratic (x_i^2) effect can be in the model only if x_i is in the model.

- Strong heredity implies that an interaction ($x_i x_j$) effect can be in the model only if both x_i and x_j are in the model, and a quadratic (x_i^2) effect can be in the model only if x_i is in the model.

2.7 Possible Fitted Models for a 2^2 factorial design.

The full model is $y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon$.

Table 2. Parameters being in possible models for a 2^2 factorial design.

Model i	α_i	β_j	$(\alpha\beta)_{ij}$
1	/	/	/
2	/	/	
3	/		
4		/	
5	/		/
6		/	/

In Table 2, Models (1)-(4) are possible models based on strong heredity. Models (1)-(6) are possible models based on weak heredity.

2.8 Possible Fitted Models for a 3^2 factorial design

The full model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$.

Table 3. Parameter being in possible models for a 3^2 factorial design.

Model i	β_1	β_2	β_{11}	β_{22}	β_{12}
1	/	/	/	/	/
2	/	/			/
3	/	/	/		/
4	/	/		/	/
5	/	/	/	/	
6	/	/	/		
7	/	/		/	
8	/	/			
9	/		/		
10		/		/	
11	/				
12		/			
13	/		/		/
14	/				/
15		/		/	/
16		/			/

In Table 3, Models (1)-(12) are possible models based on strong heredity. Models (1)-(16) are possible models based on weak heredity.

3. Simulation study

For each experimental design, data are simulated for the Type I and Type II models to assess the objectives concerning the selection of correct models and the proportion of selected models that match across the selection methods. Data for the Type I and Type II models were simulated for the following cases: For each experimental design, a model with μ = 10 for the effect model or β_0 = 4 for the regression model were generated. The choices for μ and β_0 are arbitrary. These have no impact on any results in this study. Prior values are then assigned to the model effects or parameters. The values of the effects or parameters are varied across 0, 0.25, 0.5, 1, and 2, $\varepsilon \sim N(0,1)$ and the number of experimental treatment replicates is *n* = 2, 4. The SAS® system for windows 9.0 was used for the simulation of data and for data analysis. For each Type I and Type II model and for each choice of model parameter values, 500 data sets were simulated. For each of these 500 data sets, a final model was determined for the following model selection methods: Model selection using Akaike's information criterion (AIC) and the P-Value Method that select a final model from among all possible models assuming strong heredity (SH) and weak heredity (WH).However, the stepwise procedure that selects a final model from all possible models and does not depend upon principle heredity.

4. Results

The proportions of correct models and the proportions of models that match were calculated. The results of the study will now be summarized:

4.1 The proportion of correct model for 2^2 factorial design.

Table 4. The proportions of correct models selected for 2^2 factorial design, Model Type I and n = 2 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
1	0.25	0.53	0.488	0.098	0.098	0.094
2	0.5	0.704	0.642	0.214	0.214	0.208
3	1	0.896	0.828	0.63	0.63	0.608
4	2	0.912	0.844	0.944	0.944	0.914

Table 5. The proportions of correct models selected for 2^2 factorial design, Model Type I and n = 4 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
1	0.25	0.578	0.528	0.122	0.122	0.12
2	0.5	0.766	0.704	0.43	0.43	0.408
3	1	0.856	0.79	0.9	0.9	0.866
4	2	0.856	0.79	0.952	0.952	0.914

Table 6. The proportions of correct models selected for a 2^2 factorial design, Model Type II and n = 2 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	$\beta_1 = -\beta_2$	$(\alpha\beta)_{11} = -(\alpha\beta)_{12} = (\alpha\beta)_{22} = -(\alpha\beta)_{21}$	AIC SH	AIC WH	P-Value SH	P-Value WH	Step wise
1	0.25	0.25	0.25	0.018	0.002	0.006	0.006	0.004
2	0.5	0.25	0.25	0.02	0.01	0.012	0.012	0
3	0.5	0.25	0.5	0.048	0.014	0.018	0.018	0
4	0.5	0.5	0.25	0.024	0.012	0.016	0.016	0.007
5	0.5	0.5	0.5	0.054	0.02	0.028	0.028	0.006
6	1	0.25	0.25	0.02	0.01	0.012	0.012	0.004
7	1	0.25	0.5	0.054	0.02	0.026	0.026	0.004
8	1	0.25	1	0.22	0.046	0.054	0.054	0.014
9	1	0.5	0.25	0.028	0.016	0.024	0.024	0.002
10	1	0.5	0.5	0.076	0.042	0.056	0.056	0.008
11	1	0.5	1	0.254	0.098	0.126	0.126	0.026
12	1	1	0.25	0.042	0.034	0.044	0.044	0.004
13	1	1	0.5	0.112	0.092	0.114	0.114	0.018
14	1	1	1	0.324	0.242	0.292	0.292	0.03
15	2	0.25	0.25	0.02	0.01	0.012	0.012	0.01
16	2	0.25	0.5	0.054	0.02	0.028	0.028	0.012
17	2	0.25	1	0.22	0.05	0.06	0.06	0.034
18	2	0.25	2	0.77	0.06	0.074	0.074	0.046
19	2	0.5	0.25	0.028	0.016	0.026	0.026	0.006
20	2	0.5	0.5	0.08	0.05	0.07	0.07	0.032
21	2	0.5	1	0.256	0.116	0.154	0.154	0.078
22	2	0.5	2	0.786	0.158	0.196	0.196	0.116
23	2	1	0.25	0.048	0.04	0.052	0.052	0.028
24	2	1	0.5	0.122	0.108	0.134	0.134	0.072
25	2	1	1	0.38	0.326	0.382	0.382	0.114
26	2	1	2	0.872	0.496	0.548	0.548	0.186
27	2	2	0.25	0.056	0.056	0.064	0.064	0.038
28	2	2	0.5	0.146	0.146	0.17	0.17	0.104
29	2	2	1	0.508	0.508	0.558	0.558	0.192
30	2	2	2	0.964	0.952	0.968	0.968	0.038

Table 7. The proportions of correct models selected for a 2^2 factorial design, Model Type II and n = 4 replicates per experimental treatment.

Case	$\alpha_1 =$ $-\alpha_2$	$\beta_1 =$ $-\beta_2$	$(\alpha\beta)_{11} =$ $-(\alpha\beta)_{12} =$ $(\alpha\beta)_{22} =$ $-(\alpha\beta)_{21}$	AIC SH	AIC WH	P-Value SH	P-Value WH	Step wise
1	0.25	0.25	0.25	0.106	0.026	0.004	0.004	0.004
2	0.5	0.25	0.25	0.126	0.042	0.014	0.014	0.01
3	0.5	0.25	0.5	0.432	0.102	0.046	0.046	0.036
4	0.5	0.5	0.25	0.176	0.106	0.036	0.036	0.024
5	0.5	0.5	0.5	0.492	0.248	0.124	0.124	0.098
6	1	0.25	0.25	0.138	0.062	0.026	0.026	0.02
7	1	0.25	0.5	0.458	0.15	0.08	0.08	0.072
8	1	0.25	1	0.954	0.238	0.16	0.16	0.156
9	1	0.5	0.25	0.204	0.148	0.056	0.056	0.048
10	1	0.5	0.5	0.56	0.372	0.212	0.212	0.194
11	1	0.5	1	0.974	0.588	0.406	0.406	0.392
12	1	1	0.25	0.254	0.246	0.128	0.128	0.126
13	1	1	0.5	0.626	0.612	0.43	0.43	0.412
14	1	1	1	0.984	0.96	0.894	0.894	0.884
15	2	0.25	0.25	0.138	0.062	0.026	0.026	0.022
16	2	0.25	0.5	0.458	0.15	0.08	0.08	0.074
17	2	0.25	1	0.954	0.238	0.16	0.16	0.158
18	2	0.25	2	1	0.24	0.16	0.16	0.16
19	2	0.5	0.25	0.204	0.154	0.058	0.058	0.05
20	2	0.5	0.5	0.56	0.378	0.216	0.216	0.198
21	2	0.5	1	0.974	0.596	0.42	0.42	0.416
22	2	0.5	2	1	0.6	0.426	0.426	0.426
23	2	1	0.25	0.254	0.254	0.136	0.136	0.136
24	2	1	0.5	0.626	0.624	0.45	0.45	0.444
25	2	1	1	0.984	0.98	0.938	0.938	0.936
26	2	1	2	1	0.994	0.968	0.968	0.968
27	2	2	0.25	0.254	0.254	0.136	0.136	0.136
28	2	2	0.5	0.626	0.626	0.462	0.462	0.462
29	2	2	1	0.984	0.984	0.968	0.968	0.968
30	2	2	2	1	1	1	1	1

The results in Tables 4 -5, the proportion of correct models selected of 2^2 factorial design for model Type I, the correct model is $y = \mu + \alpha_i + \varepsilon$ .In Table 4 for n=2 replicates per experimental treatment, when small $\alpha_i \leq 0.5$, the AIC SH Methods perform the best. For $\alpha_i \geq 1$, the AIC SH Methods had the best model selection performance. These were similar to small α_i but slightly better for the P-Value and Stepwise Methods. In Table 5 for n=4 replicates per experimental treatment, the AIC SH

Methods perform the best when $\alpha_i \leq 0.5$. For $\alpha_i \geq 1$, the P-Value Methods had the best model selection performance. The results in Tables 6-7, the proportion of correct models selected for model Type I, the correct model is $y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon$. For n=2 replicates per experimental treatment presented in table 6, the performance results for all model selection methods are very poor.

The Stepwise Methods performed the worst for all effect values. For $\alpha_i \geq 1$, $\beta_j \leq 1$ and $(\alpha\beta)_{ij} \geq \beta_j$, the AIC SH Method had the best model selection performance. For large α_i and β_j (=2), the P-Value Method had the best model selections performance. In Table 7 for n=4 replicates per experimental treatment, the AIC SH Method perform the best for all combinations of effect values.

4.2 The proportion of matching models selection for 2^2 factorial design.

Table 8. The proportions of matching models selection for 2^2 factorial design, Model Type I and n = 2 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	AIC SH with P-Value SH	AIC WH With P-Value WH	AIC SH with Stepwise	AIC WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	0.25	0.144	0.14	0.136	0.134	0.288	0.286
2	0.5	0.258	0.252	0.234	0.234	0.416	0.414
3	1	0.654	0.602	0.808	0.746	0.582	0.582
4	2	0.964	0.896	0.938	0.93	0.962	0.962

Table 9. The proportions of matching models selection for 2^2 factorial design, Model Type I and n = 4 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AIC WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	0.25	0.16	0.164	0.158	0.164	0.336	0.342
2	0.5	0.43	0.4	0.42	0.404	0.624	0.628
3	1	0.856	0.804	0.844	0.836	0.942	0.946
4	2	0.904	0.838	0.892	0.874	0.958	0.958

Table 10. The proportions of matching models selection for 2^2 factorial design, Model Type II and n = 2 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	$\beta_1 = -\beta_2$	$(\alpha\beta)_{11} = -(\alpha\beta)_{12} = (\alpha\beta)_{22} = -(\alpha\beta)_{22}$	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AIC WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	0.25	0.25	0.25	0.148	0.146	0.128	0.136	0.344	0.342
2	0.5	0.25	0.25	0.252	0.254	0.218	0.233	0.422	0.418
3	0.5	0.25	0.5	0.216	0.25	0.178	0.2	0.398	0.382
4	0.5	0.5	0.25	0.334	0.334	0.276	0.294	0.482	0.482
5	0.5	0.5	0.5	0.246	0.286	0.188	0.228	0.466	0.452
6	1	0.25	0.25	0.612	0.594	0.564	0.574	0.764	0.746
7	1	0.25	0.5	0.534	0.494	0.442	0.512	0.656	0.642
8	1	0.25	1	0.28	0.382	0.136	0.354	0.458	0.434
9	1	0.5	0.25	0.574	0.564	0.494	0.51	0.73	0.72
10	1	0.5	0.5	0.502	0.482	0.396	0.438	0.622	0.604
11	1	0.5	1	0.334	0.388	0.17	0.302	0.43	0.41
12	1	1	0.25	0.632	0.628	0.472	0.472	0.694	0.69
13	1	1	0.5	0.574	0.562	0.38	0.39	0.622	0.618
14	1	1	1	0.438	0.468	0.162	0.172	0.416	0.362
15	2	0.25	0.25	0.606	0.594	0.908	0.896	0.598	0.622
16	2	0.25	0.5	0.948	0.734	0.824	0.872	0.834	0.832
17	2	0.25	1	0.764	0.36	0.404	0.794	0.47	0.432
18	2	0.25	2	0.17	0.628	0.038	0.504	0.04	0.248
19	2	0.5	0.25	0.92	0.866	0.876	0.902	0.93	0.93
20	2	0.5	0.5	0.89	0.778	0.794	0.862	0.854	0.854
21	2	0.5	1	0.784	0.488	0.512	0.76	0.566	0.538
22	2	0.5	2	0.272	0.66	0.128	0.642	0.37	0.396
23	2	1	0.25	0.878	0.866	0.826	0.824	0.946	0.936
24	2	1	0.5	0.87	0.84	0.762	0.766	0.872	0.87
25	2	1	1	0.796	0.708	0.528	0.55	0.626	0.604
26	2	1	2	0.574	0.782	0.208	0.396	0.396	0.324
27	2	2	0.25	0.98	0.98	0.738	0.738	0.756	0.756
28	2	2	0.5	0.956	0.956	0.656	0.656	0.69	0.69
29	2	2	1	0.902	0.902	0.388	0.388	0.436	0.436
30	2	2	2	0.964	0.958	0.038	0.038	0.056	0.056

Table 11. The proportions of matching models selection for 2^2 factorial design, Model Type II and n = 4 replicates per experimental treatment.

Case	$\alpha_1 = -\alpha_2$	$\beta_1 = -\beta_2$	$(\alpha\beta)_{11} = -(\alpha\beta)_{12} = (\alpha\beta)_{22} = -(\alpha\beta)_{22}$	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Step wise	AIC WH with Step wise	P-Value SH with Step wise	P-Value WH with Step wise
1	0.25	0.25	0.25	0.18	0.172	0.172	0.194	0.436	0.45
2	0.5	0.25	0.25	0.398	0.35	0.368	0.376	0.616	0.62
3	0.5	0.25	0.5	0.27	0.27	0.238	0.352	0.446	0.504
4	0.5	0.5	0.25	0.436	0.41	0.414	0.412	0.744	0.736
5	0.5	0.5	0.5	0.364	0.358	0.324	0.4	0.602	0.634
6	1	0.25	0.25	0.77	0.632	0.734	0.708	0.876	0.876
7	1	0.25	0.5	0.54	0.386	0.474	0.692	0.59	0.606
8	1	0.25	1	0.198	0.292	0.176	0.85	0.194	0.322
9	1	0.5	0.25	0.666	0.602	0.648	0.63	0.902	0.904
10	1	0.5	0.5	0.552	0.49	0.514	0.628	0.732	0.742
11	1	0.5	1	0.43	0.492	0.41	0.76	0.432	0.528
12	1	1	0.25	0.828	0.828	0.824	0.832	0.986	0.986
13	1	1	0.5	0.784	0.784	0.766	0.778	0.958	0.958
14	1	1	1	0.908	0.916	0.898	0.92	0.918	0.942
15	2	0.25	0.25	0.812	0.664	0.776	0.758	0.892	0.892
16	2	0.25	0.5	0.576	0.38	0.502	0.736	0.596	0.596
17	2	0.25	1	0.204	0.17	0.18	0.9	0.194	0.194
18	2	0.25	2	0.16	0.17	0.16	0.92	0.16	0.17
19	2	0.5	0.25	0.7	0.632	0.686	0.666	0.928	0.928
20	2	0.5	0.5	0.574	0.498	0.538	0.662	0.74	0.74
21	2	0.5	1	0.446	0.434	0.436	0.81	0.452	0.452
22	2	0.5	2	0.426	0.43	0.426	0.826	0.426	0.433
23	2	1	0.25	0.862	0.862	0.862	0.862	1	1
24	2	1	0.5	1	0.812	1	0.806	0.988	0.988
25	2	1	1	0.954	0.954	0.952	0.956	0.97	0.97
26	2	1	2	0.954	0.954	0.954	0.958	0.938	0.938
27	2	2	0.25	0.882	0.882	0.882	0.882	1	1
28	2	2	0.5	0.836	0.836	0.836	0.836	1	1
29	2	2	1	0.984	0.984	0.984	0.984	1	1
30	2	2	2	1	1	1	1	1	1

The results in Tables 8 -9, the proportion of matching models selection of 2^2 factorial design for model Type I, the model is $y = \mu + \alpha_i + \varepsilon$ .The results for n = 2 replicates per experimental treatment presented in Table 8, the AIC and P-Value Methods had the highest proportion of selected models that match for α_i =2 and β_j =2. For α_i =1, the P-Value SH and Stepwise Methods had the highest proportion of selected models that match. For small $\alpha_i \leq 0.5$, the P-Value and Stepwise Methods had the

highest proportion of selected models that match. In Table 9 for n = 4 replicates per experimental treatment, when α_i =2,

- The P-Value and Stepwise Methods had the highest proportion of models that match for $\beta_i \geq (\alpha\beta)_{ij}$.
- All models performed very high for $\beta_j \geq 1$, but P-Value and Stepwise Methods had the highest proportion of selected models that match for β_i =2.
- The AIC WH and Stepwise Methods had the highest proportion of selected models that match for $\beta_i < (\alpha\beta)_{ij}$.

For α_i =1, the P-Value and Stepwise Methods had the highest proportion of selected models that match for $\beta_i \geq (\alpha\beta)_{ij}$. Moreover, the AIC WH and Stepwise Methods had the highest proportion of selected models that match for $\beta_i < (\alpha\beta)_{ij}$.For small $\alpha_i \leq 0.5$, the P-Value and Stepwise Methods had the highest proportion of selected models that match for $\beta_i \geq (\alpha\beta)_{ij}$.

The results in Tables 10 -11, the proportion of matching models selection of 2^2 factorial design for model Type II, the model is $y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon$. For $\alpha_i \leq$ 0.5 the P-Value SH and Stepwise Methods had the highest proportion of models that match for n=2.

For n = 4 replicates per experimental treatment, the P-Value SH and Stepwise Methods had the highest proportion of models that match for all combinations of effect values.

4.3 The proportion of correct model for 3^2 factorial design.

Table 12. The proportions of correct models selected for 3^2 factorial design, Model Type I and n = 2 replicates per experimental treatment.

Case	β_1	β_2	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
1	2	2	0.462	0.462	0.868	0.868	0.872
2	2	1	0.458	0.458	0.78	0.78	0.782
3	2	0.5	0.328	0.328	0.328	0.328	0.328
4	1	1	0.45	0.45	0.702	0.702	0.7
5	1	0.5	0.32	0.32	0.296	0.296	0.294
6	0.5	0.5	0.214	0.214	0.144	0.144	0.138

Table 13. The proportions of correct models selected for 3^2 factorial design, Model Type I and n = 4 replicates per experimental treatment.

Case	β_1	β_2	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
1	2	2	0.498	0.498	0.838	0.838	0.838
2	2	1	0.498	0.498	0.838	0.838	0.838
3	2	0.5	0.414	0.414	0.554	0.554	0.554
4	1	1	0.498	0.498	0.836	0.836	0.836
5	1	0.5	0.411	0.411	0.552	0.552	0.552
6	0.5	0.5	0.348	0.348	0.376	0.376	0.372

Table 14. The proportions of correct models selected for 3^2 factorial design, Model Type II and n = 2 replicates per experimental treatment.

Case	β_1	β_2	β_{11}	β_{22}	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
1	2	2	2	2	0.728	0.728	0.89	0.89	0.888
2	2	2	2	1	0.58	0.58	0.376	0.376	0.372
3	2	2	2	0.5	0.288	0.288	0.136	0.136	0.136
4	2	2	1	1	0.452	0.452	0.19	0.19	0.184
5	2	2	1	0.5	0.234	0.234	0.068	0.068	0.066
6	2	2	0.5	0.5	0.128	0.128	0.024	0.024	0.024
7	2	1	2	1	0.58	0.58	0.336	0.336	0.33
8	2	1	2	0.5	0.284	0.284	0.12	0.12	0.118
9	2	1	1	1	0.452	0.452	0.166	0.166	0.152
10	2	1	1	0.5	0.23	0.23	0.058	0.058	0.054
11	2	1	0.5	0.5	0.126	0.126	0.018	0.018	0.018
12	2	0.5	2	0.5	0.248	0.248	0.058	0.058	0.054
13	2	0.5	1	0.5	0.196	0.196	0.034	0.034	0.03
14	2	0.5	0.5	0.5	0.122	0.122	0.01	0.01	0.008
15	1	1	1	1	0.452	0.452	0.156	0.156	0.138
16	1	1	1	0.5	0.23	0.23	0.058	0.058	0.05
17	1	1	0.5	0.5	0.124	0.124	0.018	0.018	0.018
18	1	0.5	1	0.5	0.198	0.198	0.034	0.034	0.028
19	1	0.5	0.5	0.5	0.1	0.1	0.01	0.01	0.008
20	0.5	0.5	0.5	0.5	0.084	0.084	0.006	0.006	0.004

Table 15. The proportions of correct models selected for 3^2 factorial design, Model Type II and n = 4 replicates per experimental treatment.

Case	β_1	β_2	β_{11}	β_{22}	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
1	2	2	2	2	0.786	0.786	0.95	0.95	0.95
2	2	2	2	1	0.728	0.728	0.742	0.742	0.742
3	2	2	2	0.5	0.042	0.042	0.254	0.254	0.254
4	2	2	1	1	0.664	0.664	0.584	0.584	0.584
5	2	2	1	0.5	0.372	0.372	0.194	0.194	0.192
6	2	2	0.5	0.5	0.21	0.21	0.06	0.06	0.06
7	2	1	2	1	0.728	0.728	0.74	0.74	0.738

Case	β_1	β_2	β_{11}	β_{22}	AIC SH	AIC WH	P-Value SH	P-Value WH	Stepwise
8	2	1	2	0.5	0.42	0.42	0.254	0.254	0.254
9	2	1	1	1	0.664	0.664	0.582	0.582	0.582
10	2	1	1	0.5	0.372	0.372	0.194	0.194	0.192
11	2	1	0.5	0.5	0.21	0.21	0.06	0.06	0.06
12	2	0.5	2	0.5	0.414	0.414	0.174	0.174	0.174
13	2	0.5	1	0.5	0.366	0.366	0.136	0.136	0.134
14	2	0.5	0.5	0.5	0.204	0.204	0.04	0.04	0.04
15	1	1	1	1	0.664	0.664	0.582	0.582	0.582
16	1	1	1	0.5	0.372	0.372	0.194	0.194	0.192
17	1	1	0.5	0.5	0.21	0.21	0.06	0.06	0.06
18	1	0.5	1	0.5	0.366	0.366	0.136	0.136	0.134
19	1	0.5	0.5	0.5	0.204	0.204	0.04	0.04	0.04
20	0.5	0.5	0.5	0.5	0.202	0.202	0.028	0.028	0.028

The results in Tables 12 -13, the proportion of correct models selected of 3^2 factorial design for model Type I, the correct model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. In Table 12 for n=2 replicates per experimental treatment**,** the P-Value and Stepwise Methods performed the best for large values of β_1 and β_2 (≥ 1). For small either or both β_1 and β_2 $(= 0.5)$, the AIC Methods perform the best but were still not very good. The result in Table 13, The P-Value and Stepwise Methods had the best model selection performance but they were still not very good for small values either or both β_1 and β_2 .The results in Tables 14 -15, the proportion of correct models selected of 3^2 factorial design for model Type II, the correct model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$ .In Table 14 for n=2 replicates per experimental treatment, the AIC Methods performed the best for all combinations of the parameter values except for $\beta_1, \beta_2, \beta_{11}$ and β_{22} =1; they were still not very good because the values are still very low. For n=4 replicates per experimental treatment presented in Table 15, The AIC Methods performed the best except for $\beta_1 = \beta_{11} = 2$ and $\beta_{22} \geq 0.5$, but these results were still not very good.

The proportion of matching models selection for 3^2 factorial design.

Table 16. The proportions of matching models selection for 3^2 factorial design, Model Type I and n = 2 replicates per experimental treatment.

Case	β_1	β_2	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AIC WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	2	2	0.556	0.556	0.552	0.552	0.996	0.996
2	2	1	0.52	0.52	0.516	0.518	0.986	0.986
3	2	0.5	0.442	0.418	0.434	0.416	0.956	0.956
4	1	1	0.48	0.48	0.47	0.47	0.94	0.944
5	1	0.5	0.406	0.386	0.388	0.374	0.838	0.842
6	0.5	0.5	0.264	0.25	0.192	0.182	0.43	0.448

Table 17. The proportions of matching models selection for 3^2 factorial design, Model Type I and n = 4 replicates per experimental treatment.

Case	β_1	β_2	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AIC WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	2	2	0.614	0.614	0.614	0.614	1	1
2	2	1	0.612	0.612	0.612	0.612	0.998	0.998
3	2	0.5	0.53	0.51	0.53	0.516	0.976	0.976
4	1	1	0.61	0.61	0.61	0.61	0.998	0.998
5	1	0.5	0.528	0.508	0.528	0.514	0.976	0.976
6	0.5	0.5	0.438	0.416	0.414	0.4	0.79	0.794

Table 18. The proportions of matching models selection for 3^2 factorial design, Model Type II and n = 2 replicates per experimental treatment.

Case	β_1	β_2	β_{11}	β_{22}	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AI C WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	2	2	2	2	0.746	0.746	0.742	0.742	0.996	0.996
2	2	2	2	1	0.486	0.486	0.476	0.476	0.99	0.99
3	2	2	2	5	0.58	0.58	0.578	0.578	0.998	0.998
4	2	2	1	1	0.364	0.364	0.354	0.354	0.99	0.99
5	2	2	1	0.5	0.438	0.438	0.432	0.432	0.99	0.99
6	2	2	0.5	0.5	0.438	0.438	0.434	0.434	0.996	0.996
7	2	1	2	1	0.44	0.438	0.428	0.426	0.94	0.94
8	2	1	2	0.5	0.534	0.534	0.528	0.528	0.972	0.978
9	2	1	1	1	0.328	0.328	0.31	0.31	0.934	0.936
10	2	1	1	0.5	0.408	0.408	0.394	0.394	0.962	0.964
11	2	1	0.5	0.5	0.404	0.404	0.398	0.4	0.97	0.97

Case	β_1	β_2	β_{11}	β_{22}	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AI C WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
12	2	0.5	2	0.5	0.426	0.412	0.418	0.404	0.884	0.906
13	2	0.5	1	0.5	0.336	0.328	0.326	0.318	0.88	0.888
14	2	0.5	0.5	0.5	0.304	0.29	0.294	0.286	0.872	0.876
15	1	1	1	1	0.304	0.304	0.266	0.266	0.79	0.79
16	1	1	1	0.5	0.376	0.376	0.348	0.348	0.86	0.86
17	1	1	0.5	0.5	0.374	0.374	0.356	0.356	0.9	0.9
18	1	0.5	1	0.5	0.306	0.298	0.28	0.27	0.73	0.732
19	1	0.5	0.5	0.5	0.276	0.266	0.256	0.25	0.758	0.76
20	0.5	0.5	0.5	0.5	0.18	0.176	0.134	0.138	0.394	0.404

Table 19. The proportions of matching models selection for 3^2 factorial design, Model Type II and n = 4 replicates per experimental treatment.

Case	β_1	β_2	β_{11}	β_{22}	AIC SH with P-Value SH	AIC WH with P-Value WH	AIC SH with Stepwise	AI C WH with Stepwise	P-Value SH with Stepwise	P-Value WH with Stepwise
1	2	2	2	2	0.836	0.836	0.836	0.836	1	1
2	2	2	2	1	0.71	0.71	0.708	0.708	0.998	0.998
3	2	2	2	5	0.616	0.616	0.614	0.614	0.998	0.998
4	2	2	1	1	0.62	0.62	0.62	0.62	1	1
5	2	2	1	0.5	0.526	0.526	0.522	0.522	0.996	0.996
6	2	2	0.5	0.5	0.476	0.476	0.476	0.476	0.998	0.998
7	2	1	2	1	0.708	0.708	0.706	0.706	0.996	0.996
8	2	1	2	0.5	0.614	0.614	0.612	0.612	0.998	0.998
9	2	1	1	1	0.618	0.618	0.618	0.618	0.988	0.998
10	2	1	1	0.5	0.524	0.524	0.52	0.52	0.996	0.996
11	2	1	0.5	0.5	0.476	0.476	0.476	0.476	0.998	0.998
12	2	0.5	2	0.5	0.502	0.488	0.5	0.486	0.908	0.912
13	2	0.5	1	0.5	0.43	0.418	0.426	0.414	0.908	0.908
14	2	0.5	0.5	0.5	0.388	0.376	0.388	0.376	0.91	0.912
15	1	1	1	1	0.618	0.618	0.388	0.376	0.996	0.996
16	1	1	1	0.5	0.522	0.522	0.518	0.518	0.994	0.994
17	1	1	0.5	0.5	0.474	0.474	0.474	0.474	0.998	0.998
18	1	0.5	1	0.5	0.428	0.416	0.422	0.41	0.902	0.902
19	1	0.5	0.5	0.5	0.386	0.374	0.386	0.374	0.91	0.912
20	0.	0.5	0.5	0.5	0.298	0.29	0.284	0.276	0.698	0.704

The results in Tables 16 -19, the proportion of matching models selection of 3^2 factorial design for model Type I , the model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ and model Type II ,

the model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$. The P-Value and Stepwise Methods had the highest proportion of selected models that match for all combinations of parameter values for n=2 and n=4 replicates per experimental treatment and both types of model.

5. Conclusions

In this section, the results of the proportions of correct models selected are interpreted in the term of method that selected the correct model more often than other. The proportions of matching model selections are interpreted in the term of match of model selection methods that selected the same final model. From above results, we can conclude as follow:

5.1 The Model Selections for a 2^2 Factorial Design

For the proportions of correct models selected, Model Type I, for n=2 replicates per experimental treatment, the AIC SH Method selected the correct model more often than for the other methods. For n=4 replicates per experimental treatment, the AIC SH Method still selected the correct model more often than for the other methods for α_i is small (≤ 0.5). The P-Value Methods selected the correct model more often than for the other methods when $\alpha_i \geq 1$.

Model Type II, for n=2 replicates per experimental treatment, the AIC SH Method selected the correct model more often than for the other methods for $\alpha_i \geq 1$, $\beta_j \leq 1$ and $(\alpha\beta)_{ij} \geq \beta_j$. The P-Value Methods selected the correct model more often than for the other methods when $\beta_j = 2$. For n=4 replicates per experimental treatment, the AIC SH Method selected the correct model more often than for the other methods.

For the proportions of matching models selections, Model Type I, for $\alpha_i \leq 0.5$ the P-Value SH and Stepwise Methods selected the same model more often than other pair of methods for n=2 replicates per experimental treatment. The P-Value SH and Stepwise Methods selected the same model more often than other pair of methods when n= 4.

Model Type II, for n=2 replicates per experimental treatment, the AIC and P-Value Methods selected the same model more often than the other pairs of methods for large values of parameters α_i and β_i (=2). If small α_i (≤ 1), the P-Value and Stepwise Methods selected the same model more often than the other pairs of methods. For n=4

replicates per experimental treatment, all methods selected the same model always for large (α_i =2) and β_j (=2). The P-Value and Stepwise Methods selected the same model more often than the other pairs of methods for $\alpha_i \geq 1$ and $\beta_j \geq (\alpha\beta)_{ij}$. The AIC WH and Stepwise Methods also selected the same model more often than the other pairs of methods for $\alpha_i \geq 1$ and $\beta_j < (\alpha\beta)_{ij}$.

5.2 The Model Selections for a 3^2 Factorial Design

For the proportions of correct models selected, Model Type I and Model Type II, the P-Value and Stepwise Methods selected the correct model often than the other methods for n=2, 4 replicates per experimental treatment.

For the proportions of matching model selections, Model Type I and Model Type II, the P-Value and Stepwise Methods selected the same model more often than the other pair of methods for n=2, 4 replicates per experimental treatment.

6. Recommendations

In this article, we focused on the three model selection Methods and two model types for each experimental design. Respectively, the two Model Type I and Type II correspond to models with relatively few important effects and to models with many important effects. The sets of potential models that could be selected using AIC or using P-values were based on either the strong heredity or weak heredity principle. In this article, the 'full' model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$. Model Type I, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ is a strong heredity model. Model Type II, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$ is also a strong heredity model. We could consider other models, such as the Type I: $y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \varepsilon$ which is also a strong heredity model or the Type II: $y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + + \beta_{12} x_1 x_2 + \varepsilon$ and $y = \beta_0 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$ which weak heredity models are. The difference in the proportions of correct models selected is highly dependent on whether the true model is a strong heredity model or a weak heredity model. If a model is a weak heredity model, then the proportion of correct models selected would be 0 if the model selection Method was based on the strong heredity principle (either the AIC SH Method or the P-Value SH Method). Therefore, if the researchers suspect that a weak heredity model is

likely to be the true model, then they should not consider model selection using the AIC SH Method and the P-Value SH Method.

It should be noted that although many scientists use the Stepwise Method to select a model, the results of this study do not support the use of the Stepwise Method because of its poor performance in selecting the correct model. One reason for the poor performance is that model selection using the Stepwise Method is not based on the strong or weak heredity principle. This means the final model may not be consistent with the underlying scientific knowledge or have a reasonable scientific explanation. Therefore, the recommendation is not to use the Stepwise Method for model selection for designed experiments.

The results in Tables 16 -19, the proportion of matching models selection of 3^2 factorial design for model Type I , the model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ and model Type II, the model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$. The P-Value and Stepwise Methods had the highest proportion of selected models that match for all combinations of parameter values for n=2 and n=4 replicates per experimental treatment and both types of model**.**

Future studies with other types of experimental designs, larger numbers of replicates, and other true model forms should be considered to further validate the recommendations from this study.

7. Acknowledgements

We would like to thank Thammasat University for partial support this research.

References

[1] Chipman, H. Bayesain Variable Selection With Related Predictor. *Canadian Journal of Statistics*, 1996;24: 17-36.

[2] Burnham, K.P., & Anderson, D.R. Multimodal Inference Understanding AIC and BIC in Model Selection. *Sociological Method&Research,* 2004;33: 216-304

[3] Bozdogan, H. Model selection and Akaike's Information Criteria (AIC): the general theory and its analytical extensions. *Psychometrika,* 1987;52: 345-370.

[4] Farrell, S., & Wagenmakers, Eric-Jan. AIC model selection using Akaike weights. *Psychomic Bullen & Review,* 2004;11: 192-196.

[5] Draper, N.R., & Smith, H. *Applied regression analysis* (3rd ed.). New York: Wiley & Sons, Inc. 1998.