A SEQUENTIAL MULTIPLE COMPARISON PROCEDURE FOR DETECTING A LOWEST DOSE HAVING INTERACTION IN A DOSE-RESPONSE TEST

Tomohiro Nakamura* and Hideyuki Douke†

ABSTRACT

In this study, we propose a multiple comparison procedure for detecting sequentially a lowest dose level having interaction based on two dose sample means on two treatments with increasing dose levels in a dose-response test. We apply a group sequential procedure in order to realize our method that tests sequentially the null hypotheses of no interaction based on tetrad differences. If we can first detect a dose level having interaction at an early stage in the sequential test, since we can terminate the procedure with just the few observations up to that stage, the procedure is useful from an economical point of view.

In the procedure, we present an integral formula to determine the repeated confidence boundaries for satisfying a predefined type I familywise error rate. Furthermore, we show how to decide a required sample size in each cell so as to guarantee the power of the test. In the simulation studies, we evaluate the superiority among the procedures based on three \(\alpha\) spending functions in terms of the power of the test and the required sample size for various configurations of population means.

1. Introduction

In a factorial experiment, one important purpose is to detect significant interactions by using a replicated two-way classification. If we wish to investigate which level combination (cell) of two factors has an interaction, we can use a multiple comparison procedure based on all cell-wise interaction contrasts in a two-way table.

Hochberg and Tamhane (1987) were concerned with two multiple comparison procedures based on a family of interaction contrasts to detect simultaneously the interactions between two factors. One type of interaction contrast is based on the family of tetrad differences taken on the four cell population means in any \(2 \times 2\) subtable in a two-way table. Another type is based on the family of interaction residuals under the assumption in which the expectation of an observation is given by an additive model in each cell.

However, it seems difficult to detect interactions effectively by using these procedures with a large number of levels of the two factors. Also, these procedures must carry out the test based on the replicated observations (group observations) across all cells for detecting these significant interactions. If high costs are involved for obtaining the group observations in each cell in the factorial experiment, it is meaningful to develop a multiple comparison procedure to sequentially detect the interaction with as few observations as possible.

*Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan E-mail: tnakamura@megabank.tohoku.ac.jp
†Tokai University, 1117, Kitakaname, Hiratsuka 259-1292, Japan

Key words: Multiple comparison; Lowest dose having interaction; Group sequential procedure
In this study, we focus on the problem of detecting the lowest dose level at which an interaction first appears from two dose sample means on two treatments with increasing dose levels in the dose-response test. Then, we propose a sequential multiple comparison procedure for detecting an interaction based on the group observations obtained by a cell-wise sequential dose experiment in the two-way table. We apply a group sequential procedure in order to realize our method that tests sequentially the null hypotheses of no interaction based on tetrad differences. If we can first detect an interaction at an early stage in the sequential test, since we can terminate the procedure with the few observations up to the corresponding cell, the procedure is useful from an economical point of view.

The group sequential procedure by Pocock (1977) is a method to sequentially test the hypothesis of no difference between two population means on two treatments based on the group observations obtained at each of the sequential stage. The procedure is characterized by the way, constructing the repeated confidence boundaries under a pre-specified significance level. Lan and DeMets (1983) proposed a unique method that constructs the repeated confidence boundaries by using an \( \alpha \) spending function. In this study, we adopt three \( \alpha \) spending functions from Lan and DeMets (1983).

Also, since our procedure is a step-wise method for detecting a lowest dose level having interaction by sequentially comparing the difference between two population means at each of a set of increasing doses with that at a control dose, we can consider our sequential procedure as an application of Dunnett’s procedure (1955). By applying Dunnett’s procedure (1955), Nakamura and Douke (2007a), and Nakamura, Douke and Yamamoto (2007b) developed a sequential multiple comparison procedure to sequentially identify a minimum effective dose having greater effect than a control dose in a dose-response test. Thus, we will devise our sequential procedure by referring to these studies. The contents of our paper are as follows:

In Section 2, we set up the null hypotheses against the alternative hypotheses based on the tetrad differences and give the test statistics in our sequential test. We present an integral formula to determine the repeated confidence boundaries for satisfying a predefined type I familywise error rate (Type I FWER). Furthermore, we show the procedure to decide a required sample size in each cell so as to guarantee the power of the test. In the simulation studies, Section 3 compares the superiority among these procedures based on three \( \alpha \) spending functions in terms of the power of the test and the required sample size for various configurations of population means having interaction. In Section 4, we apply our sequential procedure in a case study. Finally, Section 5 gives the conclusions.

2. Sequential Multiple Comparison Procedure

Here, we will develop our procedure for detecting sequentially a lowest dose level having interaction by applying the group sequential procedure by Pocock (1977).

2.1. Hypotheses for Sequential Test

Here, we set up the null hypotheses against the alternative hypotheses for the sequential test.

Let \( A_1, A_2 \) be two treatments (two levels) and let \( B_1, B_2, \cdots, B_b \) be dose levels that have a natural order (i.e., \( B_1 < B_2 < \cdots < B_b \)). Here, \( B_1 \) is a control dose level. We assume that the groups of observations of equal sample size \( r \) at each dose level are successively obtained by a cell-wise sequential dose experiment. Furthermore, we assume that the \( m(m = 1, \cdots, r) \)th observation in cell \((A_i, B_j)(i = 1, 2, \; j = 1, \cdots, b)\) \( X_{ijm} \) is a random variable that is distributed according to a normal distribution with a population mean \( \mu_{ij} \) and a
common unknown variance $\sigma^2$. We can write this distribution as

$$X_{ijm} \sim N(\mu_{ij}, \sigma^2).$$

(2.1)

The sequential dose experiment is carried out in order of cells $(A_1, B_1)$ and $(A_2, B_1)$, $(A_1, B_2)$ and $(A_2, B_2)$, $\cdots$, $(A_1, B_b)$ and $(A_2, B_b)$. The grouped observations obtained by the sequential experiment are shown in a $2 \times b$ two-way table in Table 1.

**Table 1: $2 \times b$ two-way table**

| Treatments | Dose levels | $B_2$ | $\cdots$ | $B_b$ |
|------------|------------|-------|-----------|-------|
| $A_1$      | $X_{111}$  | $X_{121}$ | $\cdots$ | $X_{1b1}$ |
|            | $X_{112}$  | $X_{122}$ | $\cdots$ | $X_{1b2}$ |
|            | $\vdots$   | $\vdots$ | $\cdots$ | $\vdots$ |
|            | $X_{11r}$  | $X_{12r}$ | $\cdots$ | $X_{1br}$ |
| $A_2$      | $X_{211}$  | $X_{221}$ | $\cdots$ | $X_{2b1}$ |
|            | $X_{212}$  | $X_{222}$ | $\cdots$ | $X_{2b2}$ |
|            | $\vdots$   | $\vdots$ | $\cdots$ | $\vdots$ |
|            | $X_{21r}$  | $X_{22r}$ | $\cdots$ | $X_{2br}$ |

A dose sample mean in cell $(A_i, B_j)$ is

$$\bar{X}_{ij} = \frac{\sum_{m=1}^{r} X_{ijm}}{r}, \quad i = 1, 2, \quad j = 1, \ldots, b.$$  

(2.2)

Then, $\bar{X}_{ij}$ can be denoted as

$$\bar{X}_{ij} \sim N \left( \mu_{ij}, \frac{\sigma^2}{r} \right).$$

(2.3)

Here, the tetrad difference is given as

$$(\mu_{1j} - \mu_{2j}) - (\mu_{11} - \mu_{21}) = \gamma_j - \gamma_1, \quad j = 2, \ldots, b.$$  

(2.4)

Since we consider a problem for detecting sequentially a lowest dose level having interaction, we will give the steps to test sequentially the null hypotheses of no interaction in any of four cells $H_0^{(k)}(k = 1, \cdots, b - 1)$ against the alternative hypotheses in which the interaction exists in at least one of the four cells $H_1^{(k)}(k = 1, \cdots, b - 1)$ as follows:

**step 1** $H_0^{(1)}: \gamma_2 = \gamma_1, \quad H_1^{(1)}: \gamma_2 \neq \gamma_1.$

If $H_0^{(1)}$ is rejected, we see that the interaction exists at least one of $(A_1, B_1)$, $(A_2, B_1)$, $(A_1, B_2)$, $(A_2, B_2)$.

**step 2** $H_0^{(2)}: \gamma_3 = \gamma_1, \quad H_1^{(2)}: \gamma_3 \neq \gamma_1.$
If $H_0^{(2)}$ is rejected, we see that the interaction exists at least one of $(A_1, B_3), (A_2, B_3)$.  

\[ \text{step } k \quad H_0^{(k)} : \gamma_{k+1} = \gamma_1, \quad H_1^{(k)} : \gamma_{k+1} \neq \gamma_1. \]

If $H_0^{(k)}$ is rejected, we see that the interaction exists at least one of $(A_1, B_{k+1}), (A_2, B_{k+1})$.  

\[ \text{step } (b - 1) \quad H_0^{(b-1)} : \gamma_b = \gamma_1, \quad H_1^{(b-1)} : \gamma_b \neq \gamma_1. \]  

(2.5)

If $H_0^{(b-1)}$ is rejected, we see that the interaction exists at least one of $(A_1, B_b), (A_2, B_b)$.

### 2.2. Test Statistics

Here, we give the test statistic at each step. We first set up

\[ Y_k = (\bar{X}_{1,k+1} - \bar{X}_{2,k+1}) - (\bar{X}_{11} - \bar{X}_{21}), \quad k = 1, \ldots, b - 1. \]

Then, $Y_k$ is denoted as

\[ Y_k \sim N \left( \frac{4\sigma^2}{r} \right). \]  

(2.7)

Thus, we can denote that

\[ Z_k = \sqrt{\frac{r}{4\sigma^2}} Y_k \sim N \left( \sqrt{\frac{r}{4\sigma^2}} (\gamma_{k+1} - \gamma_1), 1 \right), \quad k = 1, \ldots, b - 1. \]  

(2.8)

Under $H_0^{(k)}$, (2.8) can be denoted as

\[ Z_k \sim N(0, 1). \]  

(2.9)

On the other hand, the pooled sample variance in cell $(A_1, B_j)$ and $(A_2, B_j)$ is given by

\[ S_j = \frac{\sum_{i=1}^{r} \sum_{m=1}^{r} (X_{ijm} - \bar{X}_{ij})^2}{2(r-1)}, \quad j = 1, \ldots, b. \]

(2.10)

Then, we can write as

\[ V_j = \frac{2(r-1)S_j}{\sigma^2} \sim \chi^2(2(r-1)), \]  

(2.11)

where $\chi^2(2(r-1))$ is distributed according to the $\chi^2$ distribution with $2(r-1)$ degrees of freedom. Also, we can denote as

\[ U_k = V_1 + \cdots + V_{k+1} \sim \chi^2(2(r-1)(k+1)). \]  

(2.12)

Thus, under $H_0^{(k)}$, we give the test statistic at step $k(k = 1, \ldots, K)$ as

\[ T_k = \frac{Z_k}{\sqrt{\frac{U_k}{2(r-1)(k+1)}}}, \quad k = 1, \ldots, b - 1. \]  

(2.13)
decision rules

When the repeated confidence boundaries \(d_1, \ldots, d_{b-1}\) were determined in advance and the \(t\)-values \(t_1, \ldots, t_{b-1}\) of \(T_1, \ldots, T_{b-1}\) were sequentially obtained, the decision rules in our sequential procedure are as follows:

(i) step \((k = 1, \ldots, b - 2)\)

(a) If \(|t_k| < d_k\), then one continues to test the hypothesis \(H_0^{(k+1)}\) based on \(t_{k+1}\) at step \((k+1)\),

(b) If \(|t_k| > d_k\), then one rejects the hypothesis \(H_0^{(k)}\) and terminates the sequential test.

(ii) step \((b - 1)\)

(a) If \(|t_{b-1}| \leq d_{b-1}\), then one accepts the hypothesis \(H_0^{(b-1)}\),

(b) If \(|t_{b-1}| > d_{b-1}\), then one rejects the hypothesis \(H_0^{(b-1)}\).

2.3. Repeated Confidence Boundaries

The sequential test is performed by constructing the repeated confidence boundaries. We will first define the Type I FWER for a pre-specified significance level \(\alpha\) under \(\gamma_1 = \cdots = \gamma_b\) as

\[
\alpha = \Pr \left\{ (H_0^{(1)} \text{ is rejected} | \gamma_1 = \gamma_2) \cup (H_0^{(2)} \text{ is rejected} | \gamma_1 = \gamma_2 = \gamma_3) \cup \cdots \cup (H_0^{(k)} \text{ is rejected} | \gamma_1 = \cdots = \gamma_k = \gamma_{k+1} \text{ are not rejected}) \cup \cdots \cup (H_0^{(b-2)} \text{ are not rejected}) \cup (H_0^{(b-1)} \text{ is rejected} | \gamma_1 = \cdots = \gamma_{b-1} = \gamma_b) \right\}
\]

\[
= \Pr\left\{ (|T_1| > d_1) \cup (|T_1| \leq d_1, |T_2| > d_2) \cup \cdots \cup (|T_1| \leq d_1, \ldots, |T_{k-1}| \leq d_{k-1}, |T_k| > d_k) \right. 
\]

\[
+ \cdots + \left. \Pr\left(|T_1| \leq d_1, \ldots, |T_{b-2}| \leq d_{b-2}, |T_{b-1}| > d_{b-1}\right) \right\}.
\]

(2.14)

To determine the repeated confidence boundaries \(d_1, \ldots, d_{b-1}\), we denote the following relations by using Armitage, McPherson and Rowe (1969) on (2.14).

\[
\alpha_1^* = \Pr(|T_1| > d_1),
\]

\[
\vdots
\]

\[
\alpha_k^* = \Pr(|T_1| \leq d_1, \ldots, |T_{k-1}| \leq d_{k-1}, |T_k| > d_k),
\]

\[
\vdots
\]

\[
\alpha_{b-1}^* = \Pr(|T_1| \leq d_1, \ldots, |T_{b-2}| \leq d_{b-2}, |T_{b-1}| > d_{b-1}).
\]

(2.15)

The nominal significance level \(\alpha_k^*(k = 1, \ldots, b-1)\) is the probability of crossing the boundary that terminates the test by rejecting first \(H_0^{(k)}\) at step \(k\). Furthermore, \(\alpha_k^*\) must satisfy

\[
\alpha = \sum_{k=1}^{b-1} \alpha_k^*.
\]

(2.16)

Here we can denote

\[
1 - \alpha = 1 - \sum_{k=1}^{b-1} \alpha_k^* = \Pr(|T_1| \leq d_1, \ldots, |T_{b-1}| \leq d_{b-1})
\]

(2.17)

and

\[
1 - (\alpha_1^* + \cdots + \alpha_k^*) = \Pr(|T_1| \leq d_1, |T_2| \leq d_2, \cdots, |T_k| \leq d_k), \quad k = 1, \ldots, b - 1.
\]

(2.18)
We show (2.18) by an integral formula as follows:
We can set \( \gamma_1 = \cdots = \gamma_b = 0 \) without loss of generality to calculate \( d_1, \cdots, d_{b-1} \) under \( H_0^{(1)}, \cdots, H_0^{(b-1)} \). Here, we give

\[
\tilde{Y}_1 = \sqrt{\frac{r}{4\sigma^2}} (X_{11} - X_{21}), \quad \tilde{Y}_{k+1} = \sqrt{\frac{r}{4\sigma^2}} (X_{1,k+1} - X_{2,k+1}), \quad k = 1, \cdots, b - 1. \quad (2.19)
\]

Then \( \tilde{Y}_1, \cdots, \tilde{Y}_b \) are independently distributed as

\[
\tilde{Y}_j \sim N \left( 0, \frac{1}{2} \right), \quad j = 1, \cdots, b. \quad (2.20)
\]

Thus, (2.18) can be written as

\[
1 - (\alpha_1^* + \cdots + \alpha_k^*) = \Pr(|T_1| \leq d_1, |T_2| \leq d_2, \cdots, |T_k| \leq d_k) = \Pr \left[ \left| \tilde{Y}_2 - \tilde{Y}_1 \right| \leq d_1 \sqrt{\frac{U_1}{6(r-1)}}, \left| \tilde{Y}_3 - \tilde{Y}_2 \right| \leq d_2 \sqrt{\frac{U_2}{6(r-1)}}, \cdots, \left| \tilde{Y}_{k+1} - \tilde{Y}_k \right| \leq d_k \sqrt{\frac{U_k}{2(r-1)(k+1)}} \right] \]

\[
= \int_{-\infty}^{\infty} \cdots \int_{0}^{\infty} \Pr \left( \tilde{Y}_1 - d_1 \sqrt{\frac{U_1}{6(r-1)}} \leq \tilde{Y}_2 \leq \tilde{Y}_1 + d_1 \sqrt{\frac{U_1}{6(r-1)}} \right) \times \Pr \left( \tilde{Y}_2 - d_2 \sqrt{\frac{U_2}{6(r-1)}} \leq \tilde{Y}_3 \leq \tilde{Y}_2 + d_2 \sqrt{\frac{U_2}{6(r-1)}} \right) \times \cdots \times \Pr \left( \tilde{Y}_k - d_k \sqrt{\frac{U_k}{2(r-1)(k+1)}} \leq \tilde{Y}_{k+1} \leq \tilde{Y}_k + d_k \sqrt{\frac{U_k}{2(r-1)(k+1)}} \right) \times g(v_1) \cdots g(v_{k+1}) f (\tilde{y}_1) dv_1 \cdots dv_{k+1} d\tilde{y}_1
\]

\[
= \int_{-\infty}^{\infty} \cdots \int_{0}^{\infty} \left( \int_{\tilde{y}_1 - d_1 \sqrt{\frac{U_1}{6(r-1)}}}^{\tilde{y}_1 + d_1 \sqrt{\frac{U_1}{6(r-1)}}} f (\tilde{y}_2) d\tilde{y}_2 \right) \left( \int_{\tilde{y}_2 - d_2 \sqrt{\frac{U_2}{6(r-1)}}}^{\tilde{y}_2 + d_2 \sqrt{\frac{U_2}{6(r-1)}}} f (\tilde{y}_3) d\tilde{y}_3 \right) \times \cdots \times \left( \int_{\tilde{y}_k - d_k \sqrt{\frac{U_k}{2(r-1)(k+1)}}}^{\tilde{y}_k + d_k \sqrt{\frac{U_k}{2(r-1)(k+1)}}} f (\tilde{y}_{k+1}) d\tilde{y}_{k+1} \right) g(v_1) \cdots g(v_{k+1}) f (\tilde{y}_1) dv_1 \cdots dv_{k+1} d\tilde{y}_1
\]

(2.21)

where \( f (\tilde{y}_1), f (\tilde{y}_2), \cdots, f (\tilde{y}_{k+1}) \) are the probability density functions (p.d.f.s) of \( N(0, 1/2) \) and \( g(v_1), g(v_2), \cdots, g(v_{k+1}) \) are the p.d.f.s of \( \chi^2 \)-distribution with \( 2(r - 1) \) degrees of freedom. After, \( \alpha_1^*, \cdots, \alpha_k^* \) are given, we can determine the repeated confidence boundaries \( d_1, \cdots, d_k \) by (2.21).

2.4. \( \alpha \) Spending Function

To handle the problem of multiplicity in the sequential tests, we use the \( \alpha \) spending functions (\( \alpha \) S.F.s) in Lan and DeMets (1983) to decide \( \alpha_1^*, \cdots, \alpha_{b-1}^* \). So, we adopt the following three \( \alpha \) S.F.s in Lan and DeMets (1983).
(1) Normal

$$\alpha_k = 2 \left\{ 1 - \Phi \left( z_{\frac{1}{2} \sqrt{\frac{k}{b-1}}} \right) \right\}, \quad k = 1, \cdots, b-1,$$

(2.22)

where $\Phi$ is the distribution function of $N(0,1)$ and $z_{\frac{1}{2}}$ is the upper $100 \times \frac{\alpha}{2}$ % point of $N(0,1)$.

(2) Log

$$\alpha_k = \alpha \times \log \left\{ 1 + (e - 1) \frac{k}{b-1} \right\}, \quad k = 1, \cdots, b-1.$$

(2.23)

(3) $\alpha t$

$$\alpha_k = \alpha \times \frac{k}{b-1}, \quad k = 1, \cdots, b-1.$$

(2.24)

In Table 2, we compare three $\alpha$ S.F.s for $\alpha = 0.05, b-1 = 4$.

Fig.1 shows the change of three $\alpha$ S.F.s with increasing steps from Table 2.

| $\alpha$ S.F. | step 1 | step 2 | step 3 | step 4 |
|---------------|--------|--------|--------|--------|
| Normal        | 0.00009| 0.00557| 0.02363| 0.05   |
| $\alpha t$    | 0.01250| 0.02500| 0.03750| 0.05   |
| Log           | 0.01787| 0.03101| 0.04140| 0.05   |

$(\alpha = 0.05, b-1 = 4)$

Fig. 1: Change of three $\alpha$ S.F.s ($\alpha = 0.05, b-1 = 4$)
After we selected one of the three $\alpha$ S.F.s, we can decide $\alpha_k^*(k = 1, \cdots, b - 1)$ at each step so as to satisfy

$$\alpha_k^* = \alpha_k - \alpha_{k-1}, \quad k = 1, \cdots, b - 1, \quad (2.25)$$

where $\alpha_0 = 0.0$.

### 2.5. Power of Test and Required Sample Size

We will define the power of the test and give the procedure to decide a required sample size in each cell. Here we give a configuration of population means in which an interaction first exists at step $(n-1)$ and some interactions may exist from step $n$ on as

$$\gamma_1 = \cdots = \gamma_{n-1} \neq \gamma_n, \quad \gamma_{n+1} \neq^* \gamma_1, \quad \cdots, \quad \gamma_l \neq^* \gamma_1, \quad \cdots, \quad \gamma_b \neq^* \gamma_1, \quad (2.26)$$

where $\gamma_l \neq^* \gamma_1$ means $\gamma_l = \gamma_1$ or $\gamma_l \neq \gamma_1$. Under (2.26), $\tilde{Y}_j$ can be denoted by

$$\tilde{Y}_j \sim N\left( \sqrt{\frac{r}{4\sigma^2}} \gamma_j, \frac{1}{2} \right), \quad j = 1, \cdots, b. \quad (2.27)$$

Then we denote the probability $\pi_k^*$ of first rejecting $H_0^{(k)}$ at step $k (k = 1, \cdots, b - 1)$ as

$$\pi_k^* = \Pr(|T_1| \leq d_1, \cdots, |T_{k-1}| \leq d_{k-1}, |T_k| > d_k | \gamma_1 = \cdots = \gamma_{n-1} \neq \gamma_n, \gamma_{n+1} \neq^* \gamma_1, \cdots, \gamma_b \neq^* \gamma_1), \quad k = 1, \cdots, b - 1. \quad (2.28)$$

We give the probability

$$\pi_{b-1} = \sum_{k=1}^{b-1} \pi_k^*$$

$$= 1 - \Pr(|T_1| \leq d_1, \cdots, |T_{b-2}| \leq d_{b-2}, |T_{b-1}| \leq d_{b-1} | \gamma_1 = \cdots = \gamma_{n-1} \neq \gamma_n, \gamma_{n+1} \neq^* \gamma_1, \cdots, \gamma_b \neq^* \gamma_1)$$

$$= 1 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left( \int_{\tilde{y}_1 = -d_1}^{\tilde{y}_1 + d_1} \sqrt{\frac{U_1}{r (r-1)}} f(\tilde{y}_2) d\tilde{y}_2 \right) \left( \int_{\tilde{y}_1 = -d_1}^{\tilde{y}_1 + d_1} \sqrt{\frac{U_2}{r (r-1)}} f(\tilde{y}_3) d\tilde{y}_3 \right) \cdots$$

$$\times \left( \int_{\tilde{y}_1 = -d_1}^{\tilde{y}_1 + d_1} \sqrt{\frac{U_{b-1}}{2(r-1)b}} f(\tilde{y}_b) d\tilde{y}_b \right) g(v_1) \cdots g(v_b) f(\tilde{y}_1) dv_1 \cdots dv_b d\tilde{y}_1, \quad (2.29)$$

where $f(\tilde{y}_1), f(\tilde{y}_2), \cdots, f(\tilde{y}_b)$ are the p.d.f.s of $N(\gamma_j \sqrt{r/(2\sigma)}, 1/2)(j = 1, \cdots, b)$. Here, after deciding the repeated confidence boundaries by (2.21) based on a required sample size, we call the probability obtained by (2.29) based on these boundaries the power of the test $\pi$.

**procedure to decide a required sample size** $s$

We show the procedure to decide a required sample size $s$ at each cell as follows:

(i) Specify $b, \sigma^2, \alpha$ and the power of the test $\pi$ in advance.
(ii) Select an $\alpha$ S.F. and decide $\alpha_1^*, \cdots, \alpha_{b-1}^*$.
(iii) Set up a configuration of population means.
(iv) Give a tentative sample size $s_1$.
(v) Put $i = 1$.
(vi) Decide the tentative $d_i^{(s_1)}, \cdots, d_{b-1}^{(s_1)}$ by (2.21).
(vii) Calculate the probability $\pi_{b-1}^{(s_i)}$ of (2.29) based on $d_1^{(s_i)}, \ldots, d_{b-1}^{(s_i)}$.
(viii) • if $i = 1$, change $s_1$ to $s_2(s_1 \leq s_2)$, put $i = 2$ and go to (vi).
   • if $i > 1$, go to (ix).
(ix) • if $\pi_{b-1}^{(s_i-1)} < \pi < \pi_{b-1}^{(s_i)}$, we adopt a minimum value $s_i$ that satisfies the condition as the required sample size $s_i$.
   • if $\pi_{b-1}^{(s_i-1)} < \pi_{b-1}^{(s_i)} < \pi$, change $s_i$ to $s_{i+1}(s_i \leq s_{i+1})$, put $i = i + 1$ and go to (vi).

3. Simulation Studies

The purpose of the simulation studies is to compare the superiority among the multiple comparison procedures based on three $\alpha$ S.F.s. Since we can suppose that the power of the test is influenced by the shape of the repeated confidence boundaries, we investigate the boundaries based on three $\alpha$ S.F.s. Also, we compare the procedures based on three $\alpha$ S.F.s in terms of the power of the test and the required sample size for various configurations of population means having interaction.

We calculate the repeated confidence boundaries based on three $\alpha$ S.F.s under $b = 4, \alpha = 0.05, r = 5, 10, 15$. Table 3 shows only the upper repeated confidence boundaries across all steps for each of $r = 5, 10, 15$.

| $\alpha$ S.F.s | step | $r$ |
|----------------|------|-----|
|                |      | 5   | 10  | 15  |
| Normal         | 1    | 4.194 | 3.715 | 3.595 |
|                | 2    | 2.595 | 2.490 | 2.469 |
|                | 3    | 2.171 | 2.122 | 2.109 |
| Log            | 1    | 2.522 | 2.382 | 2.344 |
|                | 2    | 2.525 | 2.431 | 2.413 |
|                | 3    | 2.543 | 2.473 | 2.454 |
| $\alpha t$     | 1    | 2.673 | 2.511 | 2.468 |
|                | 2    | 2.510 | 2.418 | 2.399 |
|                | 3    | 2.420 | 2.357 | 2.340 |

$(\alpha = 0.05, \ r = 5, 10, 15)$

Fig.2 ($r = 5$) illustrates the shapes of the repeated confidence boundaries based on three $\alpha$ S.F.s from Table 3.

Table 3 and Fig.2 show the following results:

(i) The repeated confidence boundaries based on $\alpha t$ and Log are almost flat across all steps, and the boundary based on $\alpha t$ becomes slightly wider than that based on Log at step 1, but the two boundaries are switched at step 3. The boundary based on Normal becomes wider at step 1 and becomes narrower at step 3 than the boundaries based on the other $\alpha$ S.F.s.

(ii) The repeated confidence boundaries based on three $\alpha$ S.F.s become narrower with increasing $r$.

Now, we show the patterns of configurations of population means as $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ for $b = 4$ and suppose here 13 patterns. Table 4 shows the powers of the test based on three $\alpha$ S.F.s for each pattern under $b = 4, \alpha = 0.05, \sigma = 1, r = 5, 10, 15$.

---

Fig. 2: The repeated confidence boundaries \((b = 4, \alpha = 0.05, r = 5)\)

Table 4: The powers of the test for each pattern \((b = 4)\)

| No. | pattern | Normal 5 | Normal 10 | Normal 15 | Log 5 | Log 10 | Log 15 | at 5 | at 10 | at 15 |
|-----|---------|----------|----------|----------|-------|--------|--------|------|-------|-------|
| 1   | (0,0,0,1) | 0.170 | 0.309 | 0.442 | 0.119 | 0.218 | 0.329 | 0.133 | 0.246 | 0.365 |
| 2   | (0,0,0,2) | 0.539 | 0.854 | 0.962 | 0.412 | 0.766 | 0.928 | 0.454 | 0.799 | 0.942 |
| 3   | (0,0,1,0) | 0.116 | 0.216 | 0.325 | 0.123 | 0.230 | 0.343 | 0.125 | 0.233 | 0.347 |
| 4   | (0,0,1,1) | 0.205 | 0.379 | 0.532 | 0.170 | 0.324 | 0.471 | 0.183 | 0.345 | 0.495 |
| 5   | (0,0,2,1) | 0.425 | 0.771 | 0.926 | 0.432 | 0.782 | 0.933 | 0.440 | 0.787 | 0.935 |
| 6   | (0,0,1,2) | 0.542 | 0.854 | 0.962 | 0.425 | 0.773 | 0.929 | 0.464 | 0.803 | 0.942 |
| 7   | (0,1,0,0) | 0.055 | 0.071 | 0.101 | 0.129 | 0.245 | 0.367 | 0.113 | 0.215 | 0.328 |
| 8   | (0,1,1,0) | 0.118 | 0.225 | 0.341 | 0.178 | 0.341 | 0.493 | 0.167 | 0.323 | 0.473 |
| 9   | (0,1,0,1) | 0.172 | 0.315 | 0.453 | 0.175 | 0.335 | 0.486 | 0.175 | 0.334 | 0.485 |
| 10  | (0,1,1,1) | 0.206 | 0.382 | 0.537 | 0.211 | 0.399 | 0.562 | 0.212 | 0.400 | 0.563 |
| 11  | (0,2,1,0) | 0.153 | 0.408 | 0.675 | 0.439 | 0.794 | 0.940 | 0.396 | 0.760 | 0.925 |
| 12  | (0,1,2,0) | 0.398 | 0.760 | 0.925 | 0.434 | 0.783 | 0.933 | 0.435 | 0.785 | 0.935 |
| 13  | (0,1,1,2) | 0.542 | 0.853 | 0.961 | 0.436 | 0.777 | 0.929 | 0.470 | 0.804 | 0.941 |

\((\alpha = 0.05, \sigma = 1, r = 5, 10, 15)\)

Fig. 3 \((r = 5)\) illustrates the relations among these patterns and the powers of the test based on three \(\alpha\) S.F.s from Table 4.

Table 4 and Fig.3 show the following results:

(i) The powers of the test on the patterns 2,5,6,11,12,13 that have a large interaction \((\gamma = 2)\) become higher as \(r\) becomes larger.

(ii) Log on the patterns 7,8,9,10,11,12 in which the interaction appears at an early dose level \((\text{dose level 2})\) becomes more powerful than the other \(\alpha\) S.F.s, because the repeated confidence boundaries based on Log become narrower than the other \(\alpha\) S.F.s at the early
Sequential Multiple Comparisons for Detecting Interaction

Fig. 3: The powers of the test for each pattern \(b = 4, \alpha = 0.05, r = 5\)

dose level. On the other hand, Normal on the patterns 2,6,13 in which a large interaction \(\gamma = 2\) appears at the dose level 4 and on the patterns 1,4 in which the interactions appear in the latter half of the dose levels becomes more powerful than the other \(\alpha\) S.F.s, because the repeated confidence boundaries based on Normal become narrower than the other \(\alpha\) S.F.s in the latter half of the dose levels.

Next we calculate a required sample size \(s\) in each cell so as to guarantee the power of the test \(\pi = 0.8\) under \(b = 4, \alpha = 0.05, \sigma^2 = 1\).

Table 5 shows the required sample size \(s\) based on three \(\alpha\) S.F.s for 13 patterns.

Table 5: The required sample size \(s\) for each pattern \((b = 4)\)

| No. | pattern       | Normal | Log  | \(\alpha t\) |
|-----|---------------|--------|------|--------------|
| 1   | (0,0,0,1)     | 35     | 43   | 40           |
| 2   | (0,0,0,2)     | 9      | 11   | 11           |
| 3   | (0,0,1,0)     | 43     | 41   | 40           |
| 4   | (0,0,1,1)     | 28     | 31   | 30           |
| 5   | (0,0,2,1)     | 11     | 11   | 11           |
| 6   | (0,0,1,2)     | 9      | 11   | 10           |
| 7   | (0,1,0,0)     | 72     | 39   | 42           |
| 8   | (0,1,1,0)     | 39     | 30   | 31           |
| 9   | (0,1,0,1)     | 33     | 30   | 29           |
| 10  | (0,1,1,1)     | 28     | 26   | 26           |
| 11  | (0,2,1,0)     | 19     | 11   | 11           |
| 12  | (0,1,2,0)     | 11     | 11   | 11           |
| 13  | (0,1,1,2)     | 9      | 11   | 10           |

\((\pi = 0.8, \alpha = 0.05, \sigma = 1)\)
From Table 5, we can obtain the following results:

(i) The required sample sizes $s$ based on three $\alpha$ S.F.s on the patterns 2,5,6,11,12,13 that have a large interaction ($\gamma = 2$) take small values and the differences of values of $s$ among three $\alpha$ S.F.s are small.

(ii) Log and $\alpha t$ on the patterns 7,8,9,10,11 take smaller values of $s$ than Normal. But, Normal on the patterns 1,2,4,6,13 takes smaller values of $s$ than the other $\alpha$ S.F.s.

4. Case study

Johnson and Leone (1977) showed data from an experiment on the warping of copper plates. In the two-way layout experiment, two factors are the temperature and the copper content of the plates, and the response variable is the amount of warping. Table 6 indicates the data on the warping for the levels $(50^\circ C,75^\circ C)$ of the temperature and the levels $(40\%, 60\%, 80\%, 100\%)$ of the copper content.

| temperatures($^\circ C$) | copper contents(%) |
|--------------------------|------------------|
|                          | 40   | 60   | 80   | 100  |
| 50                       | 17   | 16   | 24   | 28   |
| 75                       | 12   | 18   | 17   | 27   |

In our problem, we deal with as the observations obtained by a cell-wise sequential dose experiment in a two-way table by quoting the data. Here, we suppose that the sequential experiment is performed in order of the levels $40\%$(level 1), $60\%$(level 2), $80\%$(level 3), $100\%$(level 4) of the copper content for each of levels $(50^\circ C,75^\circ C)$ of the temperature.

First, the repeated confidence boundaries based on three $\alpha$ S.F.s under $b = 4$, $r = 2$, $\alpha = 0.05$ are as follows:

Normal $d_1 = 9.460$, $d_2 = 3.326$, $d_3 = 2.476$,

Log $d_1 = 3.606$, $d_2 = 3.157$, $d_3 = 3.005$,

$\alpha t$ $d_1 = 3.961$, $d_2 = 3.142$, $d_3 = 2.826$.

**step 1** $H_0^{(1)} : \gamma_2 = \gamma_1$, $H_1^{(1)} : \gamma_2 \neq \gamma_1$

The results of the sequential test based on the observations on cells $(50^\circ C,40\%)$,$(75^\circ C,40\%)$ and $(50^\circ C,60\%)$,$(75^\circ C,60\%)$ are as follows:

Normal $|t_1| = |-1.213| \leq d_1 = 9.460$, then $H_0^{(1)}$ is not rejected and go to step 2,

Log $|t_1| = |-1.213| \leq d_1 = 3.606$, then $H_0^{(1)}$ is not rejected and go to step 2,

$\alpha t$ $|t_1| = |-1.213| \leq d_1 = 3.961$, then $H_0^{(1)}$ is not rejected and go to step 2.

**step 2** $H_0^{(2)} : \gamma_3 = \gamma_1$, $H_1^{(2)} : \gamma_3 \neq \gamma_1$

The results of the sequential test based on the observations on cells $(50^\circ C,40\%)$, $(75^\circ C,40\%)$ and $(50^\circ C,80\%)$,$(75^\circ C,80\%)$ are as follows:

Normal $|t_2| = |0.124| \leq d_2 = 3.326$, then $H_0^{(2)}$ is not rejected and go to step 3,

Log $|t_2| = |0.124| \leq d_2 = 3.157$, then $H_0^{(2)}$ is not rejected and go to step 3.
Sequential Multiple Comparisons for Detecting Interaction

\[ \alpha |t_2| = |0.124| \leq d_2 = 3.142, \text{ then } H_0^{(2)} \text{ is not rejected and go to step 3.} \]

**step 3** \[ H_0^{(3)} : \gamma_4 = \gamma_1, \quad H_1^{(3)} : \gamma_4 \neq \gamma_1 \]

The results of the sequential test based on the observations on cells \((50^\circ C, 40\%), (75^\circ C, 40\%)\) and \((50^\circ C, 100\%), (75^\circ C, 100\%)\) are as follows:

- Normal \(|t_3| = |−2.517| > d_3 = 2.476, \text{ then } H_0^{(3)} \text{ is rejected,} \]
- Log \(|t_3| = |−2.517| \leq d_3 = 3.005, \text{ then } H_1^{(3)} \text{ is accepted,} \]
- \(\alpha |t_3| = |−2.517| \leq d_3 = 2.826, \text{ then } H_1^{(3)} \text{ is accepted.} \]

Thus, we can detect the lowest dose level having interaction at step 3 by using the procedure based on Normal. Fig. 4 illustrates the sample means of warping versus the levels of temperature by levels of copper content.

![Fig. 4: Means of warping vs. temperature by copper content (40%, 60%, 80%, 100%)](image)

### 5. Conclusions

In this study, we developed a sequential multiple comparison procedure for detecting a lowest dose level having interaction. First, we made clear the theoretical problems to construct our procedure as follows:

Specifically, we set up the null hypotheses against the alternative hypotheses based on the tetrad differences and gave the test statistics in our sequential test. Also, we defined a unique Type I FWER and presented an integral formula to determine the repeated confidence boundaries for satisfying the error rate. Furthermore, we formulated the power of the test and showed the procedure to decide a required sample size in each cell.

Next, in the simulation studies, we compared the superiority among these procedures based on three \(\alpha\) spending functions in terms of the power of the test and the required sample size for various configurations of population means. From the studies, we obtained the following results:

If we can guess that the interaction appears at an early dose level for a specified value of \(b\), the procedure based on Log is preferable. On the other hand, if we can guess that the
interaction appears in the latter half of the dose levels, the procedure based on Normal is preferable.

However, there still remain some problems. In simulation studies, the programs on the multiple integrations to determine the repeated confidence boundaries and a required sample size were made by the software of Mathematica. The necessary times were about 10 minutes and 1.5 hours to determine the repeated confidence boundaries and a required sample size under $b - 1 = 3$ by using a PC (64bit, Core i5, 16GB). Also, the necessary times were about 20 minutes and 3 hours to determine the repeated confidence boundaries and a required sample size under $b - 1 = 4$. Thus, as the number of dose level for each of two factors increases, it is generally difficult to calculate the repeated confidence boundaries and the required sample size. In the future, we will develop a sequential multiple comparison procedure with unequal sample size at each dose level in the problem.

Acknowledgments

The authors are deeply grateful to the referees for their helpful advice and useful comments for the paper. Special thanks are due to the editors for their constructive comments and suggestions.

REFERENCES

Armitage, P., McPherson, C. K. and Rowe, B. C. (1969). Repeated significance tests on accumulating data. Journal of the Royal Statistical Society. Series A, 132, 235–244.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 50, 1096–1121.

Hochberg, Y. and Tamhane, A. C. (1987). Multiple Comparison Procedures. John Wiley & Sons, Inc.

Johnson, J. L. and Leone, F. C. (1977). Statistics and Experimental Design in Engineering and the Physical Sciences. John Wiley & Sons.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. Biometrika, 70, 659–663.

Nakamura, T. and Douke, H. (2007a). Development of Sequential Multiple Comparison Procedure for Dose Response Test, Biometrical Journal, 49(1), 30–39.

Nakamura, T., Douke, H. and Yamamoto, Y. (2007b). A comparative study and development of sequential multiple comparison procedure for dose response test, Bulletin of the Computational Statistics of Japan, 20(1-2), 33–47.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. Biometrika, 64, 191–199.

(Received: June 24, 2014, Accepted: February 8, 2015)