The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 2019-2023

A NOTE ON PARTITIONING A SET OF NORMAL POPULATIONS BY THEIR LOCATIONS WITH RESPECT TO TWO CONTROLS

BY PAUL SEEGER

University of Uppsala

The problem of partitioning a set of normal populations with respect to one control has been considered by Tong (1969). This note points at the possibility of using two controls which gives a partitioning into three disjoint groups. The procedure utilizes Tong's tables. It is also stated that the procedure can be used for mixed randomized block designs with a certain symmetry. An extension to more than two controls is also mentioned.

1. Introduction. Consider the set

$$\Omega = (\pi_1, \pi_2, \cdots, \pi_k)$$

of k normal populations with vector of means $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_k)'$, and two normal control populations π_{01} and π_{02} with means μ_{01} and μ_{02}, satisfying $\mu_{01} < \mu_{02}$. All the means are assumed to be unknown parameters. The populations are supposed to have a common variance σ^2. For arbitrary but fixed constants δ_{11}^*, δ_{21}^*, δ_{12}^* and δ_{22}^*, such that $\delta_{11}^* < \delta_{21}^*$ and $\delta_{12}^* < \delta_{22}^*$, we define five disjoint and exhaustive subsets:

$$\begin{aligned}
\Omega_B &= (\pi_i : \mu_i \leqq \mu_{01} + \delta_{11}^*) , \\
\Omega_{I1} &= (\pi_i : \mu_{01} + \delta_{11}^* < \mu_i < \mu_{01} + \delta_{21}^*) , \\
\Omega_M &= (\pi_i : \mu_{01} + \delta_{21}^* \leqq \mu_i \leqq \mu_{02} + \delta_{12}^*) , \\
\Omega_{I2} &= (\pi_i : \mu_{02} + \delta_{12}^* < \mu_i < \mu_{02} + \delta_{22}^*) , \\
\Omega_G &= (\pi_i : \mu_i \geqq \mu_{02} + \delta_{22}^*) .
\end{aligned}$$

Random samples of size n are drawn from each of the $k + 2$ populations. Their observed means are denoted by

$$\bar{X}_{01}, \bar{X}_{02}, \bar{X}_1, \cdots, \bar{X}_k .$$

On the basis of these sample means the set of populations will be divided into three disjoint subsets S_B, S_M and S_G.

A decision is said to be correct (CD) if

$$\Omega_B \subset S_B ; \qquad \Omega_M \subset S_M ; \qquad \Omega_G \subset S_G ;$$
$$\Omega_{I1} \subset S_M \cup S_B ; \qquad \Omega_{I2} \subset S_M \cup S_G .$$

When $\mu_{01} + \delta_{21}^* > \mu_{02} + \delta_{12}^*$ the decision is correct if

$$\Omega_B \subset S_B \qquad \text{and} \qquad \Omega_G \subset S_G .$$

The purpose of this note is to obtain a classification procedure R, for the case

Received May 17, 1971; revised February 11, 1972.

of known σ^2, so that the probability of correct decision has at least a prescribed value P^*, i.e.,

$$P(\mathrm{CD} \mid \boldsymbol{\mu}, \sigma^2; R) \geqq P^* \qquad \text{for every mean vector } \boldsymbol{\mu}. \tag{1}$$

2. Procedure. Following Tong (1969) the procedure is to divide the population in the following three sets:

$$\begin{aligned} S_B &= (\pi_i : \bar{X}_i \leqq \bar{X}_{01} + d_1) , \\ S_M &= (\pi_i : \bar{X}_{01} + d_1 < \bar{X}_i \leqq \bar{X}_{02} + d_2) , \\ S_G &= (\pi_i : \bar{X}_i > \bar{X}_{02} + d_2) , \end{aligned}$$

where

$$d_1 = (\delta_{11}^* + \delta_{21}^*)/2 \qquad \text{and} \qquad d_2 = (\delta_{12}^* + \delta_{22}^*)/2 .$$

It is assumed that the probability for $\bar{X}_{01} + d_1 > \bar{X}_{02} + d_2$ is very low.

The problem now is to determine the number, n, of independent observations to be made on each of the $k + 2$ populations so that (1) is satisfied. It is clear that the minimum of $P(\mathrm{CD})$ is obtained when $\mu_{01} + \delta_{21}^* = \mu_{02} + \delta_{12}^*$, and $\boldsymbol{\mu} = \boldsymbol{\mu}^0$ has the components

$$\begin{aligned} \mu_i &= \mu_{01} + \delta_{11}^* ; \quad i \in A_1 = \{1, 2, \cdots, r_1\} ; \\ \mu_i &= \mu_{01} + \delta_{21}^* ; \quad i \in A_2 = \{r_1 + 1, \cdots, k - r_2\} ; \\ \mu_i &= \mu_{02} + \delta_{22}^* ; \quad i \in A_3 = \{k - r_2 + 1, \cdots, k\} ; \end{aligned}$$

for some values of r_1 and r_2.

We then have

$$\begin{aligned} P(\mathrm{CD} \mid \boldsymbol{\mu}^0, \sigma^2; R) = P(&\bar{X}_i - \bar{X}_{01} \leqq d_1; \bar{X}_j - \bar{X}_{01} > d_1; \bar{X}_j - \bar{X}_{02} \leqq d_2; \\ &\bar{X}_t - \bar{X}_{02} > d_2; i \in A_1; j \in A_2; t \in A_3 \mid \boldsymbol{\mu}^0, \sigma^2) . \end{aligned}$$

Following Tong we make the transformation;

$$\begin{aligned} \bar{X}_i - \bar{X}_{01} - d_1 &= \sigma(2/n)^{\frac{1}{2}} Y_i - a_1 ; \quad i \in A_1 ; \\ \bar{X}_j - \bar{X}_{01} - d_1 &= \sigma(2/n)^{\frac{1}{2}} (-Y_j) + a_1 ; \quad j \in A_2 ; \\ \bar{X}_j - \bar{X}_{02} - d_2 &= \sigma(2/n)^{\frac{1}{2}} Y_j' - a_2 ; \quad j \in A_2 ; \\ \bar{X}_t - \bar{X}_{02} - d_2 &= \sigma(2/n)^{\frac{1}{2}} (-Y_t') + a_2 ; \quad t \in A_3 ; \end{aligned}$$

where

$$a_1 = (\delta_{21}^* - \delta_{11}^*)/2 \qquad \text{and} \qquad a_2 = (\delta_{22}^* - \delta_{12}^*)/2 .$$

This yields

$$\begin{aligned} P(\mathrm{CD} \mid \boldsymbol{\mu}^0, \sigma^2; R) = P(&Y_i \leqq (a_1/\sigma)(n/2)^{\frac{1}{2}}; Y_j' \leqq (a_2/\sigma)(n/2)^{\frac{1}{2}}; \\ &i \in A_1 \cup A_2; j \in A_2 \cup A_3) , \end{aligned}$$

and, if $a_1 = a_2 = a$,

$$P(\mathrm{CD} \mid \boldsymbol{\mu}^0, \sigma^2; R) = P(Y_s \leqq (a/\sigma)(n/2)^{\frac{1}{2}}; s = 1, 2, \cdots, 2k - r_1 - r_2) ,$$

where $\mathbf{Y} = (Y_1, \cdots, Y_{2k-r_1-r_2})'$ follows a multivariate normal distribution with mean vector $= \mathbf{0}$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_3 \\ \mathbf{\Sigma}_3' & \mathbf{\Sigma}_2 \end{bmatrix},$$

where

$$\mathbf{\Sigma}_1 = \left[\begin{array}{c:c} \overbrace{\begin{matrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{matrix}}^{r_1} & \overbrace{\begin{matrix} & & \\ & -\frac{1}{2} & \\ & & \end{matrix}}^{k - r_1 - r_2} \\ \hdashline \begin{matrix} & & \\ & -\frac{1}{2} & \\ & & \end{matrix} & \begin{matrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{matrix} \end{array}\right], \qquad \mathbf{\Sigma}_2 = \left[\begin{array}{c:c} \overbrace{\begin{matrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{matrix}}^{k - r_1 - r_2} & \overbrace{\begin{matrix} & & \\ & -\frac{1}{2} & \\ & & \end{matrix}}^{r_2} \\ \hdashline \begin{matrix} & & \\ & -\frac{1}{2} & \\ & & \end{matrix} & \begin{matrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{matrix} \end{array}\right],$$

and

$$\mathbf{\Sigma}_3 = \left[\begin{array}{c:c} \mathbf{0} & \mathbf{0} \\ \hdashline \begin{matrix} -\frac{1}{2} & & 0 \\ & \ddots & \\ 0 & & -\frac{1}{2} \end{matrix} & \mathbf{0} \end{array}\right].$$

If we introduce the notation

$$(a/\sigma)(n/2)^{\frac{1}{2}} = b\,, \qquad P(Y_s \leqq b;\, s = 1, 2, \cdots, 2k - r_1 - r_2) = H^{\mathbf{\Sigma}}_{2k-r_1-r_2}(b)\,,$$

we then have the requirement

$$H^{\mathbf{\Sigma}}_{2k-r_1-r_2}(b) = P^*\,. \tag{2}$$

It is possible, at least in principle, to solve (2) for b. This would give us the required sample size.

It is, however, easier to use Tong's Theorem 1.2 from which it follows that

$$H^{\mathbf{\Sigma}_1}_{k-r_2}(b') + H^{\mathbf{\Sigma}_2}_{k-r_1}(b') = 1 + P^*\,. \tag{3}$$

This equation yields $b' \geqq b$ and thus a conservative n-value. Some numerical results given by Tong indicate that the difference between b' and b is not too large. Because of the symmetry inf $P(\mathrm{CD})$ is obtained when $|r_1 - r_2| \leqq 1$ which gives

$$H^{\mathbf{\Sigma}_1}_{k-r}(b') + H^{\mathbf{\Sigma}_2}_{k-r}(b') = 1 + P^*\,, \tag{4}$$

if $r = r_1 = r_2$, or

$$H^{\mathbf{\Sigma}_1}_{k-r+1}(b') + H^{\mathbf{\Sigma}_2}_{k-r}(b') = 1 + P^*\,, \tag{5}$$

if $r = r_1 = r_2 + 1$.

Tong has shown that the largest b'-value is obtained when the off-diagonal submatrices of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are as near square matrices as possible. Thus we have for (4)

$$H_{k-[k/3]}^{\mathbf{\Sigma}'}(b') = (1 + P^*)/2 ,$$

where $[k/3]$ is the largest integer $\leqq k/3$ and $\mathbf{\Sigma}'$ the matrix corresponding to Tong's matrix (1.8) for which there is a table of b'-values. In (5) the largest b'-value is obtained for $r = [k/3] + 1$, thus

$$H_{k-[k/3]}^{\mathbf{\Sigma}_1}(b') + H_{k-[k/3]-1}^{\mathbf{\Sigma}_2}(b') = 1 + P^* .$$

If we solve b'' from

$$H_{k-[k/3]}^{\mathbf{\Sigma}'}(b'') = (1 + P^*)/2 ,$$

it is obvious that $b'' \geqq b' \geqq b$.

Thus, to obtain n we enter Tong's Table 1 with $k - [k/3]$ and $(1 + P^*)/2$ and find b. Then the required sample size is obtained from $n \geqq 2b^2(\sigma^2/a^2)$.

The case of $c > 2$ controls can be dealt with by a straightforward but a bit tedious extension of the above methods as long as $k > c$. Equation (2) is the same and the matrix $\mathbf{\Sigma}$ looks the same but with $(2c)^2$ submatrices instead of 4^2. Tong's Theorem 1.2 is easily generalized to c events. This gives a conservative solution with the result that we should enter Tong's Table 1 (or an extension of it) with $k - (c - 1)[k/(c + 1)]$ and $(c - 1 + P^*)/c$. However, in practice we will not often have $c > 2$.

I suspect that such a modification can also be made to Tong's two-stage and sequential procedures.

3. Randomized block designs. The procedure was given by Tong for independent observations. It is, however, also possible to use the procedure with certain randomized block designs. Let us consider the model equation

$$X_{ij} = u_{ij} + e_{ij} ; \qquad i = 01, 02, 1, 2, \cdots, k; j = 1, 2, \cdots, n ,$$

or, in matrix notation

$$\mathbf{X}_j = \mathbf{u}_j + \mathbf{e}_j ,$$

where the random vector $\mathbf{e}_j$ follows a multivariate normal distribution $(\mathbf{0}, \sigma_e^2\mathbf{I})$. The random vector $\mathbf{u}_j$ follows independently of $\mathbf{e}_j$, a multivariate normal distribution $(\boldsymbol{\mu}, \mathbf{\Sigma})$, where

$$\boldsymbol{\mu} = (\mu_{01}, \mu_{02}, \mu_1, \cdots, \mu_k)' ,$$

and $\mathbf{\Sigma} = (\alpha_i + \alpha_{i'} + \lambda\delta_{ii'})$.

The covariance matrix $\mathbf{\Sigma}$ is of a certain symmetrical type discussed by Huynh and Feldt (1970).

For this case the covariance matrix of the variables Y_s in (2) is easily shown to be unchanged. Then the common variance σ^2 is taken to be $\sigma_e^2 + \lambda$, where λ is the variance of the effects of interactions between the $k + 2$ populations and the blocks. Thus the procedure to find the required number of blocks is the same as in Section 2.

REFERENCES

[1] HUYNH, H. and FELDT, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *J. Amer. Statist. Assoc.* **65** 1582–1589.

[2] TONG, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. *Ann. Math. Statist.* **40** 1300–1324.

THE AGRICULTURAL COLLEGE OF SWEDEN
DEPARTMENT OF ECONOMICS AND STATISTICS
S-750 07 UPPSALA 7, SWEDEN