Construction of a randomization test for the linear contrast of treatment effects

Surendra Prasad Sinha
Instituto de Estadística Aplicada y Computación
Universidad de Los Andes
Av. Las Américas, Núcleo La Liria
Mérida, Venezuela

In this article, the problem of construction of a randomization test for the linear contrast of treatment effects for the one-way design is considered without any normality assumption. It is observed that the general method of construction of permutation and randomization tests proposed by Welch (1990) can not be applied in the present work. Therefore it became necessary to transform the linear model of the one-way design by using a reparameterization. The transformed linear model of the one-way design satisfies the invariance property, which is required for the construction of the randomization test.

KEY WORDS: Randomization; Permutation; One-way design; Linear contrast; Reparameterization.

1. Introduction

Randomization tests are considered as special cases of permutation tests and are known since they were introduced by Fisher (1935). However the use of randomization tests is not yet widespread due to a great amount of computation involved in its application. Only recently, due to the advent of high-speed computers, the randomization tests are being considered feasible and practical.

Hoeffding (1952) proposed a general method of construction of permutation test. Welch (1990) developed a method of construction of permutation tests using the properties of invariance and sufficiency. Edgington (1987) published a book on randomization tests and provided methods and computer programs to analyze experimental data for several experimental designs which are frequently used in applications. The randomization tests do not require the usual normality assumption, which may be considered unrealistic in many types of experimental data.

The purpose of this article is to develop a randomization test for the hypothesis about the linear contrast in the one-way experimental designs. The construction of a randomization test in general requires that the permutations of the vector of observations be invariant in distribution under the null hypothesis. The distribution of permutations of the vector of observations in this case was found to be not invariant under the null hypothesis except for the case when the number of treatments is equal to 2. Therefore the method proposed by Welch (1990) can not be used directly in this work. However a reparameterization of the design model enables the fulfillment of the requirement of invariance of the distribution by permuting a linear combination of observations and this will be shown in the section 4.

2. Construction of a randomization test for a linear contrast in a completely randomized design

Let

$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$; $i = 1, \ldots, k$; $j = 1, \ldots, n$; (2.1)

$E(\varepsilon_{ij}) = 0$; $Var(\varepsilon_{ij}) = \sigma^2$; $Cov(\varepsilon_{ij}, \varepsilon_{gh}) = 0$, if $(i,j) \neq (g,h)$;

$i, g = 1, \ldots, k$; $j,h = 1, \ldots, n$;

be the linear model of the completely randomized design, where

Y_{ij} is the observation for i-th treatment and j-th replication, μ is the mean effect, τ_i is the effect due to i-th treatment, ε_{ij} is the random error for i-th treatment and j-th replication.

We will need another additional assumption in the above model that the distribution functions corresponding to the k - populations in the model are continuous and may differ only in their respective location parameters.

Using the matrix notation, model given by (2.1) can be written as:

$\mathbf{Y} = \mathbf{X} b + e$ (2.2)

where

$\mathbf{Y} = (Y_1', Y_2', \ldots, Y_k')'$ is the vector of observations, in which $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ for $i = 1, 2, \ldots, k$; $\mathbf{X}$ is the design matrix of dimension $kn \times (k+1)$. The first column of the matrix $\mathbf{X}$ has all its elements equal to 1. The elements of other columns of $\mathbf{X}$ can be determined by the ordering of the elements as shown in the vector $\mathbf{Y}$. $b = (\mu, \tau_1, \tau_2, \ldots, \tau_k)'$ is the vector of parameters.

e is the vector of random errors such that $E(e) = \mathbf{0}$; $Cov(e) = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $\sigma^2 > 0$.

The hypothesis of a linear contrast in a completely randomized design can be specified as:

$H_0 : \sum_{i=1}^{k} C_i \tau_i = 0$; $H_A : \sum_{i=1}^{k} C_i \tau_i \neq 0$; (2.3) where $\sum_{i=1}^{k} c_i = 0$; (by the definition of a linear contrast)

For $k = 2$, H_0 is true $\Rightarrow \tau_1 = \tau_2$; Therefore in the case of k=2, the distribution of any permutation of the vector of observations **Y** will be invariant. However, in general for $k > 2$, such a property of invariance does not exist and consequently the method of construction given by Welch (1990) can not be used. Nevertheless, the model (2.2) can be transformed to another model by reparameterization such that the transformed model fulfills the invariance property required for the construction of randomization test in order to use the methods given by Welch (1990) and Hoeffding (1952). The theorem given below proves this result.

3. A transformation of the linear model of the completely randomized design

Theorem 3.1:

Let

$\mathbf{Y} = \mathbf{X}\,b + e$ be the linear model of a completely randomized design with k treatments and n replications under the specifications given by (2.2). Then there exists a real matrix denoted by **A** such that $\mathbf{AX}\,b = \mathbf{X}^*\,b^*$, where $\mathbf{X}^*$ is a matrix of full rank by columns and b^* is a linear transformation of b .

Proof:

The notation defined in the section 2 for the matrix representation of the model (2.2) will be used in this proof. We will consider the hypothesis of the linear contrast defined by (2.3). We assume that

$c_i > 0$ for $i = 1,2,...,r$; where $1 \leq r < k$ and $c_i \leq 0$ for $i = r+1, ..., k$; (3.1)

For the inequalities in (3.1) to be true, it may be necessary to rename the treatments by using integers 1 to k.

Let

$$C^+ = \sum_{i=1}^{r} c_i \;; \quad C^- = \sum_{i=r+1}^{k} c_i \;; \quad \tau^+ = \sum_{i=1}^{r} c_i\,\tau_i \;; \quad \tau^- = \sum_{i=r+1}^{k} c_i\,\tau_i \;; \qquad (3.2)$$

We observe that $C^+ + C^- = 0$.

We also note that $\tau^{+} + \tau^{-} = 0$ represents the null hypothesis specified by (2.3) . Let the design matrix **X** be partitioned in submatrices as given by

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n \end{bmatrix} \qquad (3.3)$$

In (3.3), $\mathbf{1}_n$ represents a column vector of n elements, all of which are equal to 1, and $\mathbf{0}_n$ is also a column vector of n elements, all of which are equal to 0 . Let

$$\mathbf{A} = \begin{bmatrix} C_1 I_n & C_2 I_n & ,\ldots, & C_r I_n & \mathbf{0}_{nxn} & \mathbf{0}_{nxn} & ,\ldots, & \mathbf{0}_{nxn} \\ \mathbf{0}_{nxn} & \mathbf{0}_{nxn} & ,\ldots, & \mathbf{0}_{nxn} & -C_{r+1} I_n & -C_{r+2} I_n & ,\ldots, & -C_k I_n \end{bmatrix}$$

Where $\mathbf{I}_n$ is the identity matrix of order n and $\mathbf{0}_{nxn}$ indicates a square matrix of order n with all its elements equal to 0 . It can be verified that

$$\mathbf{AX}\beta = \begin{bmatrix} (C^{+}\mu + \tau^{+})\mathbf{1}_n \\ (-C^{-}\mu - \tau^{-})\mathbf{1}_n \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} (C^{+}\mu + \tau^{+}) \\ (-C^{-}\mu - \tau^{-}) \end{bmatrix} = \mathbf{X}^{*}\beta^{*} \qquad (3.4)$$

where $\mathbf{X}^{*}$ and b^{*} are defined by the matrix and the vector respectively which appear on the left side of the last equality in (3.4). Therefore the transformation of the model $\mathbf{Y} = \mathbf{X}\, b + e$ is given by

$\mathbf{Y}^{*} = \mathbf{X}^{*} b^{*} + e^{*}$ in which $\mathbf{Y}^{*} = \mathbf{AY}$ and $e^{*} = \mathbf{A}\, e$.

4. Discussion

We consider the problem of construction of a randomization test for testing the hypothesis given by (2.3) en section 2.

H_0 is true $\Rightarrow (C^{+} + C^{-})\mu + \tau^{+} + \tau^{-} = 0 \Rightarrow C^{+}\mu + \tau^{+} = -C^{-}\mu - \tau^{-} \Rightarrow$ The distribution of every permutation of the elements of vector $\mathbf{Y}^{*}$ is invariant under the null hypothesis H_0 .

Consequently, the methods given by Hoeffding (1952) and Welch (1990) can be applied in the case of the model $\mathbf{Y}^* = \mathbf{X}^* b^* + e^*$ to construct a randomized test for the hypothesis given by (2.3).

Let $\mu^*_1 = c^+\mu + \tau^+$ and $\mu^*_2 = -c^-\mu - \tau^-$. We can divide the elements of $\mathbf{Y}^*$ en two groups using the following rule of assignment:

An element y^* in $\mathbf{Y}^*$, is assigned to the i-th group if $E(y^*) = \mu^*_i$, for i = 1, 2.

We observe that the elements of the $\mathbf{Y}^*$ are the linear combinations of the elements of $\mathbf{Y}$. Let

$$H_0 : \mu_1^* = \mu_2^* ; \qquad H_A : \mu_1^* \neq \mu_2^* ; \tag{4.1}$$

We note that the hypothesis (4.1) and (2.3) are equivalent. The hypothesis $H_0 : \mu^*_1 = \mu^*_2$ can be tested by using a randomization test for the Analysis of variance problem in the case of two populations and can be consulted in the book of Edgington (1987). Edgington has provided methods and computer programs to analyze experimental data using randomization test for several experimental designs, which are commonly used in applications. The reference set in this case will contain the transformed observations, which can be obtained as the permutations of the elements of the vector $\mathbf{Y}^*$. Since the vector $\mathbf{Y}^*$ has 2n elements which can be assigned to 2 groups, the number of the data permutations which will be used in the reference set is equal to $(2n)!/(n!)^2$.

5. Conclusion

A transformation of the one-way design model is used for the construction of a randomization test for the linear contrast of treatment effects. The distribution of the permutations of the elements of the transformed response vector is shown to be invariant under the null hypothesis, thereby enabling the application of the method proposed by Welch (1990). The method used in this work can be easily adapted and applied in the case of a two-way experimental design.

References

Edgington, E. S. , Randomization Tests (2nd. ed.) , Marcel Dekker, New York, 1987.

Fisher, R. A., The Design of Experiments (8th ed.), Oliver & Boyd, Edinburgh, 1966.

Hoeffding, W. , The large sample power of tests based on permutations of observations, The Annals of Mathematical Statistics, 23, 169-192, 1952.

Welch, W. J., Construction of permutation tests, Journal of American Statistical Association, 85, 693-698, 1990.

E-Mail address of the author: sinha@faces.ula.ve