Equivalence of factorial designs with qualitative and quantitative factors

Tena I. Katsaounis

The Ohio State University
Mansfield OH

DAE 2012

Factorial Design

There are

f **factors that affect the response variable:**

f_1 **qualitative factors with** $s_1 \geq 2$ **levels**

f_2 **quantitative factors with** $s_2 > 2$ **levels**
(for such factors, the distance between factor levels has a meaning)

i.e. $f = f_1 + f_2$

and n **out of total** $s_1^{f_1} \times s_2^{f_2}$ **level combinations are to be observed**

Equivalence

Two designs d_1 and d_2 are called cg-equivalent if one can be obtained from the other by

- **relabeling the factors of the same type**
- **relabeling the levels of one or more qualitative factors**
- **reversing the levels of one or more quantitative factors**
- **reordering the experimental runs.**

This is the randomization needed for a design before use.

Importance of Equivalence

Non-equivalent designs may have the same statistical properties for a particular model, but have different properties under a different model

Classification of designs into equivalence classes allows selection of a representative design from an equivalence class that fits the purpose of an experiment

Relationship with other types of equivalence

If d_1 and d_2 are cg-equivalent, then

their corresponding $n \times f_1$ subdesigns with qualitative factors only are *combinatorially* **equivalent (Clark and Dean, 2001)**

their corresponding $n \times f_2$ subdesigns with quantitative factors only are *geometrically* **equivalent (Cheng and Ye, 2004).**

Necessary condition for cg-equivalence

Example:

$$n = 4, \;\; f_1 = 3, \;\; f_2 = 1$$

$$\boldsymbol{T}_{d_1} = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{array}\right] \qquad \boldsymbol{T}_{d_2} = \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 \end{array}\right]$$

First 3 factorial effects are qualitative

If designs are cg-equivalent,

their corresponding 4×3 submatrices

are <u>combinatorially</u> equivalent, and

the last columns

are <u>geometrically</u> equivalent

However, the opposite might not hold!

Showing equivalence of two designs by direct application of the definition becomes computationally "hard" as the size of a design increases.

Requires checking up to

$n!$ **row permutations,**

$f_1! \times f_2!$ **column permutations, and**

$(s!)^{f_1} \times 2^{f_2}$ **column symbol changes.**

In this paper: necessary and sufficient conditions are given to verify equivalence or non-quivalence of two general factorial designs.

A brief history of combinatorial equivalence

- **Draper and Mitchell (1967,1968); wordlength pattern**
- **Chen and Lin (1991); wordlength pattern**
- **Chen (1992); mapping of defining relation; problem solved for regular designs**
- **Lin, Wallis and Zhu (1993); row J-characteristics**
- **Clark and Dean (2001; Hamming distance matrix; solved for regular and nongegular designs**
- **Stufken and Tang (2007); orthononal arrays**
- **Shrivastava, Ding (2007); graph based method; solved for regular designs**
- **Katsaounis and Dean (2008)-Hammming distance matrix; solved for general factorial designs**

A brief history of geometric equivalence

- **Clark and Dean (2001); Euclidean distance matrix; solved for general factorial designs**
- **Cheng and Ye (2004);** J**-characteristics; solved for general factorial designs**
- **Evangelaras, Koukouvinos, Dean and Dingus (2005);correlations**
- **Katsaounis Dingus and Dean (2007);Euclidean distance matrix; solved for general factorial designs**

A necessary and sufficient condition for equivalence of designs with qualitative and quantitative factors

Consider the *Hamming distance matrix* **of a design with qualitative factors, consisting of the Hamming distances of all pairs of runs**

The Hamming distance between two runs is the number of places where factor levels differ.

Notation: $\boldsymbol{H}_d^{\{i_1,\ldots,i_{p_1}\}}$ **is the Hamming distance matrix, based on columns** $i_1,\ldots,i_{p_1}$ **of an** $n \times f$ **design** d **with** $p_1 \le f_1$**.**

Consider the *absolute Euclidean distance matrix* $\boldsymbol{E}_d$ **of a design with quantitative factors, consisting of all pairwise absolute Euclidean distances between runs**

The absolute Euclidean distance between i^{th} **and** j^{th} **runs is:**

$$\Sigma_{k=1}^{f} \mid [\boldsymbol{T}_d]_{i,k} - [\boldsymbol{T}_d]_{j,k} \mid$$

where $[\boldsymbol{T}_d]_{i,k}$ **is the** $(ik)^{th}$ **element of** $n \times f$ **design matrix** $\boldsymbol{T}_d$**.**

Notation: $\boldsymbol{E}_d^{\{i_{p_1+1},\ldots,i_{p_2}\}}$ **is the absolute Euclidean distance matrix based on columns** $i_{p_1+1}, \ldots, i_{p_2}$ **of** $n \times f$ **design** d **with** $p_2 \leq f_2$ **quantitative factors.**

A necessary and sufficient condition for equivalence of designs with qualitative and quantitative factors

Two designs d_1 and d_2 with

f_1 **qualitative factors at** s_1 **levels and**

f_2 **quantitative factors at** s_2 **levels**

are equivalent if and only if there is

a column permutation $\{a_1, \ldots, a_{f_1}\}$ **of** $\{1, \cdots, f_1\}$**,**

a column permutation $\{a_{f_1+1}, \ldots, a_{f_2}\}$ **of** $\{f_{1+1}, \cdots, f_2\}$

and a *common* **row permutation matrix** $\boldsymbol{R}$ **such**

that for all $p_1 = 0, 1, \ldots, f_1$ **and** $p_2 = 0, 1, \ldots, f_2$

$1 \leq p = p_1 + p_2 \leq f, \;\; f = f_1 + f_2$**:**

$$\left[\boldsymbol{H}_{d_1}^{\{1,\ldots,p_1\}} \quad \boldsymbol{E}_{d_1}^{\{p_1+1,\cdots,p_2\}}\right] \;=\; \boldsymbol{R}\left[\boldsymbol{H}_{d_2}^{\{a_1,\ldots,a_{p_1}\}} \quad \boldsymbol{E}_{d_2}^{\{a_{p_1+1},\ldots,a_{p_2}\}}\right]\boldsymbol{R}'$$

Example (cont) $n = 4,\ \ f_1 = 3,\ \ f_2 = 1$

$$\boldsymbol{T}_{d_1} = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{array}\right] \qquad \boldsymbol{T}_{d_2} = \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 \end{array}\right]$$

$$\boldsymbol{H}_{d_1}^{(123)} = \begin{bmatrix} 0 & 3 & 2 & 2 \\ 3 & 0 & 2 & 3 \\ 2 & 2 & 0 & 3 \\ 2 & 3 & 3 & 0 \end{bmatrix} \qquad \boldsymbol{H}_{d_2}^{(123)} = \begin{bmatrix} 0 & 3 & 3 & 2 \\ 3 & 0 & 2 & 2 \\ 3 & 2 & 0 & 3 \\ 2 & 2 & 3 & 0 \end{bmatrix}$$

Example (cont)

$$\boldsymbol{T}_{d_1} = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{array}\right] \qquad \boldsymbol{T}_{d_2} = \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 \end{array}\right]$$

$$\boldsymbol{E}_{d_1}^{(4)} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{E}_{d_2}^{(4)} = \begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix}$$

Example (cont)

The following permutation matrix:

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

satisfies:

$$\boldsymbol{R}\ [\boldsymbol{H}_{d_2}^{(123)}\ \ \boldsymbol{E}_{d_2}^{(4)}]\ \boldsymbol{R}^{'} = [\boldsymbol{H}_{d_1}^{(123)}\ \ \boldsymbol{E}_{d_1}^{(4)}]$$

Thus, in $\mathbf{T}_{d_2}$:

- **no column permutation is needed**
- **permute the rows using** $(1\ 2\ 3\ 4) \rightarrow (2\ 1\ 4\ 3)$

- **Also, by inspection: relabel the levels in column 1 using** $(0\ 1\ 2) \rightarrow (1\ 2\ 0)$

- **reverse the levels in column 4 using** $(0\ 1\ 2) \rightarrow (2\ 1\ 0)$ **to obtain** $\mathbf{T}_{d_1}$

The two designs are equivalent.

An alternative method for the quantitative factors

is based on the concept of J-characteristic

(Tang, 2001):

For the $n \times f$ design matrix (of design d)

with columns $c_1, \cdots, c_f$ is given by:

$$J^d_{t_1,t_2,\ldots,t_f}(\boldsymbol{c}_1^{t_1}, \boldsymbol{c}_2^{t_2}, \ldots \boldsymbol{c}_f^{t_f}) \;=\; \mathbf{1}'\left(\boldsymbol{c}_1^{t_1} \circ \boldsymbol{c}_2^{t_2} \circ \cdots \circ \boldsymbol{c}_f^{t_f}\right)$$

for all $\boldsymbol{t} = (t_1, \ldots, t_f)$, $t_i = 0, 1, \ldots, s_2 - 1,\; i = 1, \ldots, f$

(Cheng and Ye, 2004).

An alternative necessary and sufficient condition for cg-equivalence

An extension of results by Cheng and Ye, 2004, and Katsaounis and Dean, 2007.

Two designs d_1 and d_2 with
f_1 qualitative factors at s_1 levels and
f_2 quantitative factors at s_2 levels
are equivalent if and only if there is
a column permutation $\{a_1, \ldots, a_{f_1}\}$ of $\{1, \cdots, f_1\}$,
a column permutation $\{a_{f_1+1}, \ldots, a_{f_2}\}$ of $\{f_1+1, \cdots, f_2\}$
a common row permutation matrix $\boldsymbol{R}$ and
an indicator vector $\boldsymbol{q} = (q_{f_1+1}, \ldots, q_{f_2})$
of 0's and 1's

such that for all $p_1 = 0, 1, \ldots, f_1$ **and** $p_2 = 0, 1, \ldots, f_2$**,** $1 \leq p = p_1 + p_2 \leq f, \quad f = f_1 + f_2$**:**
for the Hamming distance matrices of the sub-designs with the qualitative factors:

$$\boldsymbol{H}_{d_1}^{\{1,\ldots,p_1\}} \;=\; \boldsymbol{R}\ \boldsymbol{H}_{d_2}^{\{a_1,\ldots,a_{p_1}\}}\ \boldsymbol{R}'$$

for some permutation matrix $\boldsymbol{R}$

where $\boldsymbol{H}_{d}^{\{i_1,\ldots,i_{p_1}\}}$ **is the Hamming distance matrix based on columns** $i_1, \ldots, i_{p_1}$ **of design matrix** $\boldsymbol{T}_d$ **with** p_1 **qualitative factors,**

and for the $J-$characteristics of the sub-designs with quantitative factors:

$$J^{d_1}_{t_{f_1+1},\ldots,t_{f_2}} \;=\; \left(\;\Pi_{k=f_1+1}^{f_2} \quad (-1)^{q_k t_{a_k}} \;\right) \quad J^{d_2}_{t_{a_{f_1+1}},\ldots,t_{a_{f_2}}}$$

for a set of sign changes in factors indicated by vector q

for all possible $\boldsymbol{t} = (t_{f_1+1}, \ldots, t_{f_2})$**,**

$t_k = 0, 1, \ldots, s_2 - 1,\;\; k = f_1 + 1, \ldots, f_2$**,**

where $J^d_{t_{i_{p_1+1}},\ldots,t_{i_{p_2}}}$ **is the** J**-characteristic based on columns** $i_{p_1+1}, \ldots, i_{p_2}$ **of design matrix** $\boldsymbol{T}_d$ **with** p_2 **quantitative factors.**

Recall in previous example:

$$\boldsymbol{T}_{d_1} = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{array}\right] \qquad \boldsymbol{T}_{d_2} = \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 \end{array}\right]$$

$$\boldsymbol{H}_{d_1}^{(123)} = \begin{bmatrix} 0 & 3 & 2 & 2 \\ 3 & 0 & 2 & 3 \\ 2 & 2 & 0 & 3 \\ 2 & 3 & 3 & 0 \end{bmatrix} \qquad \boldsymbol{H}_{d_2}^{(123)} = \begin{bmatrix} 0 & 3 & 3 & 2 \\ 3 & 0 & 2 & 2 \\ 3 & 2 & 0 & 3 \\ 2 & 2 & 3 & 0 \end{bmatrix}$$

$\boldsymbol{R}\ \boldsymbol{H}_{d_2}^{(123)}\ \boldsymbol{R}' = \boldsymbol{H}_{d_1}^{(123)}$, with $\boldsymbol{R}$ as before.

Using the following orthonormal contrasts to re-code quantitative factors:

$$\boldsymbol{h}_{t_k}^{(k)} = \begin{cases} \frac{1}{\sqrt{3}}(1,1,1)' & \textbf{if } t_k = 0\,, \\ \frac{1}{\sqrt{2}}(-1,0,1)' & \textbf{if } t_k = 1\,, (linear\ effects) \\ \frac{1}{\sqrt{6}}(1,-2,1)' & \textbf{if } t_k = 2\,, (quadratic\ effects) \end{cases}$$

then, for the 4th column of T_{d_2} the

J-characteristic of the linear effect is $\frac{1}{\sqrt{2}}$, and

J-characteristic of the quadratic effect is $\frac{1}{\sqrt{6}}$

for the 4th column of T_{d_1} the

J-characteristic of the linear effect is $-\frac{1}{\sqrt{2}}$

J-characteristic of the quadratic effect is $\frac{1}{\sqrt{6}}$

Which suggests

reversal of the levels of the 4th column using

$(0\ 1\ 2) \rightarrow (2\ 1\ 0)$ **in** $\boldsymbol{T}_{d_2}$

So after applying R and above reversal of levels,

we see by inspection, that we need to

apply the permutation of levels

$(0\ 1\ 2) \rightarrow (1\ 2\ 0)$ **in the 1st column of** $\boldsymbol{T}_{d2}$

to obtain $\boldsymbol{T}_{d1}$

$cg - Deseq2$ Algorithm

$cg - Deseq2$ **implements the necessary and sufficient condition of the first theorem, and gives a column permutation and a row permutation that transform one design matrix to the other, if the two designs are equivalent; otherwise declares the designs non-equivalent.**

It is a modification of *Deseq2*, **by Clark and Dean (2001) for 2-level designs, and** *mDeseq2* **by Katsaounis and Dean (2008) for general symmetric designs with qualitative factors, obtained by replacing** $\boldsymbol{H}_{\boldsymbol{d}}^{\{1,\cdots,f\}}$ **with**

$$\left[\boldsymbol{H}_{\boldsymbol{d}}^{\{1,\cdots,f_1\}} \quad \boldsymbol{E}_{\boldsymbol{d}}^{\{f_1+1,\cdots,f\}}\right]$$

Evaluation of $cg - Deseq2$ Algorithm

Examples of 3-level and 4-level designs
showed that
$cg - Deseq2$ **algorithm can be slow**
depending on the size and structure
of the designs, however
is the only method that can detect
equivalence or non-equivalence of designs with
qualitative and quantitative factors

Screening methods for cg-non-equivalence

Two designs can be shown to be non-equivalent using a necessary only criterion for equivalence.

Such critera can be computationally faster.

Screening methods for cg-non-equivalence

(a) Combinatorial non-equivalence or geometric non-equivalence of the corresponding $n \times f_1$ and $n \times f_2$ subdesigns (of 2 $n \times f$ designs) implies cg-non-equivalence.

(b) Combinatorial non-eqivalence of two $n \times f$ designs, implies geometric non-equivalence

(a) and (b) imply that two $n \times f$ designs that are combinatorial non-equivalent are also cg-non-equivalent.

Thus, a screening method for detecting *combinatorial* non-equivalence can be used as a screening for cg-non-equivalence.

Screening methods for cg-non-equivalence

Advantage: Can use good existing methods for detecting combinatorial nonequivalence, such as:

- **Squared centered L_2 discrepancy (Ma, Fang and Lin, 2001)**

- $deseq1$
  Clark and Dean (2001) and Katsaounis and Dean (2008)

- **Moment aberration projection (Xu, 2003)**

All these methods detected non-equivalence fast for examples that cg-Deseq2 was slow.

Other existing methods can be used.

THANK YOU!

katsaounis.1@osu.edu

Equivalence of factorial designs with qualitative and quantitative factors, T.I. Katsaounis, Journal of Statistical Planning and Inference, 142 (2012), 79-85.