# Repeated Measurement Designs under Subject Dropout

Shi Zhao,XEROX
Dibyen Majumdar,Univ. of Illinois,Chicago

2009 Quality and Production Research Conference

June 4, 2009

# Outlines

**Repeated Measurement Design (Crossover Design)**: A number of experimental units (subjects) are used to compare treatments of interest and each subject is assigned to a sequence of treatments over a number of time periods. Heavily used in various practices, ranging from psychology and human factor engineering to medical and agricultural applications.

**Example**

| | | | |
|---|---|---|---|
| 0 | 1 | 2 | 3 |
| 1 | 2 | 3 | 0 |
| 3 | 0 | 1 | 2 |
| 2 | 3 | 0 | 1 |

$n$: the number of subjects; $t$: the number of treatments; $p$: the number of periods.

$\Omega(t, n, p)$: the class of designs with t treatments, n subjects and p periods.

**Traditional fixed-effects model**:

$$y_{ij} = \alpha_i + \beta_j + \tau_{d(i,j)} + \rho_{d(i-1,j)} + \epsilon_{ij}$$

$i = 1, 2, \cdots, p$ and $j = 1, 2, \cdots, n$.
$\alpha_i$: period effect,
$\beta_j$: subject effect,
$\tau_{d(i,j)}$: treatment direct effect,
$\rho_{d(i-1,j)}$: carryover effect,
$\epsilon_{ij}$: random error with $E(\epsilon_{ij}) = 0$ and $Var(\epsilon_{ij}) = \sigma^2$,
$Cov(\epsilon_{ij}, \epsilon_{i'j'}) = 0$ for any $(i, j) \neq (i', j')$.

For any design $d$:

$$Y^d = X_1^d\theta_1 + X_2^d\theta_2 + \varepsilon.$$

The Fisher Information Matrix:

$$C^d = (X_1^d)'(Pr^{\perp}X_2^d)X_1^d = (X_1^d)'X_1^d - (X_1^d)'X_2^d((X_2^d)'X_2^d)^-(X_2^d)'X_1^d.$$

$A-$criterion: $tr(C^d)^+$.

**Theorem**

*(Kiefer (1975)) Suppose a class $\{C^d, d \in \Omega(t,n,p)\}$ of matrices contains a $C^{d^*}$ for which*

1. $C^{d^*}$ *is completely symmetric,*
2. $trC^{d^*} = \max_{d\in\Omega(t,n,p)} trC^d$.

*Then $d^*$ is universally optimal in $\Omega(t,n,p)$.*

There is an **implicit, but critical**, assumption that the experiment will yield all the planned observations.

**Example**

| | | | |
|---|---|---|---|
| 0 | 1 | 2 | 3 |
| 1 | 2 | 3 | 0 |
| 3 | 0 | 1 | * |
| * | 3 | * | * |

* means that observation is not available

1. What statistical properties are maintained in the design at the end of the experiment? Here we choose to study UBRMD.
2. Whether and how can we measure the goodness of a planned design in terms of precision loss due to subject dropouts? We suggest using two quantities: the expected precision loss and the maximum precision loss.

**Assumption**:
Subjects drop out independently with a probability $\lambda$.
**The case**:
Dropouts occur in the final period.
$d_{plan}$: the starting design,
$d_{min}$: the starting design without the final period,
$d_{imp}$: the design actually observed; inbetween $d_{plan}$ and $d_{min}$.
The loss of $d_{imp}$ with respect to $d_{plan}$ is measured by

$$L_{d_{imp}:d_{plan}} = \frac{H^{d_{plan}} - H^{d_{imp}}}{H^{d_{plan}}} = 1 - \frac{H^{d_{imp}}}{H^{d_{plan}}}$$

where $H^d = \frac{t-1}{tr(C^d)^+}$.

**Uniformly Balanced Repeated Measure Designs (UBRMDs)**:
A design is **uniform** if

- for each subject, each treatment is allocated to the same number of periods
- for each period, each treatment is allocated to the same number of subjects

A design is called **balanced for carryover effects (balanced, in short)** if, in the order application, each treatment is preceded by every other treatment the same number of times and is not preceded by itself.

### Example

A UBRMD with $t = 4$, $p = 4$ and $n = 4$:

$$\begin{matrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 0 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \end{matrix}$$

It is universally optimal in $\Omega(4, 4, 4)$.
The construction of UBRMDs are provided in Williams (1949), called **Williams Designs (WDs)**:

- (Family One) t is even, a UBRMD with $n = p = t$
- (Family Two) t is odd, a UBRMD with $n = 2t$ and $p = t$

We suggest a type of design: **UBRMD Balanced for Loss**, denoted by $d^*$.

- $d^*$ is a UBRMD,
- $d^*$ covers all ordered pairs of distinct treatments the same number times in the last two periods.

This requires $n \geq t(t-1)$.

**Example**

A UBRMD Balance for Loss with $t = p = 4$, $n = 12$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 1 | 0 | 3 | 2 | 2 | 3 | 0 | 1 | 3 | 2 | 1 | 0 |
| 2 | 3 | 0 | 1 | 3 | 2 | 1 | 0 | 1 | 0 | 3 | 2 |
| 3 | 2 | 1 | 0 | 1 | 0 | 3 | 2 | 2 | 3 | 0 | 1 |

This UBRMD is universally optimal in $\Omega(4, 12, 4)$.

## Theorem

*In the class of designs* $\Omega(t, n, t-1)$, *for which* $z_d \geq [\frac{2n}{t(t-2)}]$ *or* $z_d = 0$, *n is a multiple of* $t(t-1)$, $d^*_{min}$ *is universally optimal.*

## Corollary

*If* $n = t(t-1)$ *and* $t \geq 5$, $d^*_{min}$ *is universally optimal in the class of designs* $\Omega(t, n, t-1)$ *for which* $z_d \geq 2$ *or* $z_d = 0$. *When* $t = 4$, $d^*_{min}$ *is universally optimal in* $\Omega(4, 12, 3)$ *for which* $z_d \geq 3$ *or* $z_d = 0$. *When* $t = 3$, $d^*_{min}$ *is universally optimal in* $\Omega(3, 6, 2)$ *for which* $z_d \geq 4$ *or* $z_d = 0$.

Using the bound in Stufken(1991) that is generalized to the entire class by Hedayat and Yang (2004), the lower bound to efficiency of $d^*_{min}$

| Order t | $d^*_{min}$ | $WD$ |
|---|---|---|
| 4 | 0.9967427 | * |
| 5 | 0.9995917 | 0.898608 |
| 6 | 0.9999083 | 0.8953254 |
| 7 | 0.9999718 | 0.973486 |
| 8 | 0.9999894 | 0.9631566 |
| 9 | 0.9999954 | 0.9881278 |
| 10 | 0.9999978 | 0.9811 |

Proved that the lower bound to efficiency of $d^*_{min} \geq 99.63\%$ for $t \geq 4$. Thus $d^*_{min}$ is nearly optimal.

Maximum Loss Table:

| Order t= | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|
| $d^*$ | 0.41 | 0.28 | 0.21 | 0.18 | 0.15 | 0.13 | 0.11 |
| $WD$ | * | 0.35 | 0.30 | 0.20 | 0.18 | 0.14 | 0.13 |

Proved the maximum loss of $d^*$ is an decreasing function of $t$.

$$Expected\ \ Loss = \sum_{d_{imp}} L_{d_{imp}:d_{plan}} \times P(d_{imp}),$$

where $P(d_{imp})$ is the probability to observe $d_{imp}$.

## Example

A UBRMD with $t = p = 5$ and $n = 20$

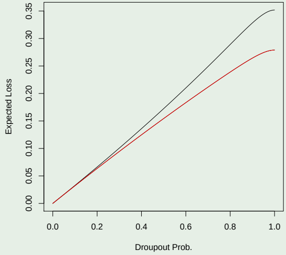



Figure: t=5, $d^*$ (red) and two copies of $WD$(black)

## Definition

The subject dropout pattern of a implemented design is a $t \times t$ matrix with $(i, j)$ entry 1 if treatment $i$ lost due to subject dropout in the last period is preceded by treatment $j$; 0 otherwise. It is denoted by $DPM_{d_{imp}}$. Two implemented designs $d_{imp1}$ and $d_{imp2}$ from the same planned design which is a *UBRMD* is said to have the same dropout pattern if there exits a permutation matrix $P$ such that $DPM_{d_{imp1}} = P \cdot DPM_{d_{imp2}} \cdot P'$.

## Theorem

*If two implemented designs from the same planned design which is a UBRMD have the same subject dropout pattern, then they have the same information loss.*

Constructions of UBRMD Balanced for Loss with $n = t(t-1)$:

1. Complete set of mutually orthogonal Latin squares.
2. Union of WDs using permutations.
3. Case $t = 6$ (Amusement in Mathematics, 1958).
4. Methods by Bose and Bagchi(Utilitas Mathematica).

# MOLS Method

Suppose $x$ is a primitive root of the field $GF(t)$. Express all the elements in the field as a power of $x$:
$\alpha_0 = 0, \alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2, \cdots, \alpha_{t-1} = x^{t-2}$. The following $t-1$ pairwise orthogonal squares $L_1, \cdots, L_{t-1}$ are constructed as: $L_i = (l_{jk}) = (\alpha_j + \alpha_i \alpha_k)$ for $i = 1, \cdots, t-1; j = 0, \cdots, t-1; k = 0, \cdots, t-1$. To get $L$, juxtapose $L_1^{'}, \cdots, L_{t-1}^{'}$ in the following way:

$$\left[ L_1^{'} \; L_2^{'} \; \cdots \; L_{t-1}^{'} \right].$$

**Theorem**

*L is a UBRMD Balanced for Loss.*

### Example

$t = 4$. Use an irreducible polynomial of order 2 which is $x^2 + x + 1$ on $\mathbf{Z}_2$ to construct a $GF(4)$ having elements $0, 1, x, 1 + x$. Relabel $0, 1, x, 1 + x$ by $0, 1, 2, 3$, the following $4 \times 12$ array is our type of design:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 1 | 0 | 3 | 2 | 2 | 3 | 0 | 1 | 3 | 2 | 1 | 0 |
| 2 | 3 | 0 | 1 | 3 | 2 | 1 | 0 | 1 | 0 | 3 | 2 |
| 3 | 2 | 1 | 0 | 1 | 0 | 3 | 2 | 2 | 3 | 0 | 1 |

# Union of WDs Using Permutation Method

1. Start with a *WD* of order $t$, call it $W_0$.
2. Find a permutation matrix of order $t$ $P_1$, such that $W_1$ obtained from $W_0$ by applying $P_1$, and $W_0$ covers totally different pairs of treatments in the last two rows.
3. Find a permutation matrix of order $t$ again $P_2$, such that $W_2$ obtained from $W_0$ by applying $P_2$, covers pairs totally different from those covered by $W_1$ and $W_0$ in the last two rows.
4. Continue in this fashion till all designs obtained together constitute a design with $t(t-1)$ sequences and every pair of distinct treatments is represented once in the final two rows.

## Example

$t = 5$,

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 3 | 4 | 0 | 1 | 2 | 0 | 3 | 1 | 4 | 2 | 4 | 2 | 0 | 3 | 1 |
| 1 | 2 | 3 | 4 | 0 | 2 | 3 | 4 | 0 | 1 | 3 | 1 | 4 | 2 | 0 | 1 | 4 | 2 | 0 | 3 |
| 4 | 0 | 1 | 2 | 3 | 4 | 0 | 1 | 2 | 3 | 2 | 0 | 3 | 1 | 4 | 2 | 0 | 3 | 1 | 4 |
| 2 | 3 | 4 | 0 | 1 | 1 | 2 | 3 | 4 | 0 | 1 | 4 | 2 | 0 | 3 | 3 | 1 | 4 | 2 | 0 |
| 3 | 4 | 0 | 1 | 2 | 0 | 1 | 2 | 3 | 4 | 4 | 2 | 0 | 3 | 1 | 0 | 3 | 1 | 4 | 2 |

$L_2$(right WD) obtained from $L_1$(left WD) by permutating treatments. Permutation is $(0, 3, 1, 4, 2)$

# Case $t = 6$ and Methods by Bose and Bagchi

In Dudeney (1958), a *UBRMD* Balanced for Loss with $t = 6, 10$ was provided. One can construct *UBRMD*s Balanced for Loss for $t = 3, 4, 5, 7, 8, 9, 11$ using either *MOLS* method or permutation algorithm. For $t = 12$, the existence of a *UBRMD* Balanced for Loss is still unclear. In Bose and Bagchi (2006), a different construction method for *UBRMD*s Balanced for Loss was suggested for $t$ is an odd prime. They started with the initial columns of *WD* and expanded each column $mod(t)$ to get a *UBRMD* Balanced for Loss in $\Omega(t, n = t(t-1), p = t)$.

## Random Subject Effect Model:

$$E(y_{ij}) = \alpha_i + \tau_{d(i,j)} + \rho_{d(i-1,j)},$$
$$Var(y_{ij}) = \sigma^2 + \sigma_\beta^2,$$

and

$$Cov(y_{ij}, y_{i'j'}) = \begin{cases} 0 & \text{if } j \neq j' \\ \sigma_\beta^2 & \text{if } j = j' \text{ and } i \neq i' \end{cases}$$

$i = 1, 2, \cdots, p$ and $j = 1, 2, \cdots, n$. Let $\theta = \frac{\sigma_\beta^2}{\sigma^2}$

## Example

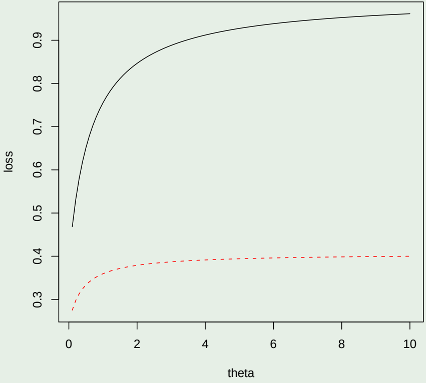



Figure: Maximum loss for t=4,$d^*$ (red) and $WD$(black)

### Theorem

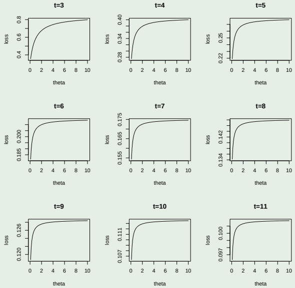
*The maximum loss of a UBRMD Balanced for Loss is monotonically increasing in $\theta$, for any given $t \geq 3$.*

### Theorem

*The maximum loss for WD is monotonically increasing in $\theta$, for any given $t \geq 3$.*

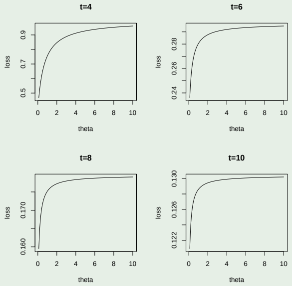
## Example

Maximum information loss for $d^*$:



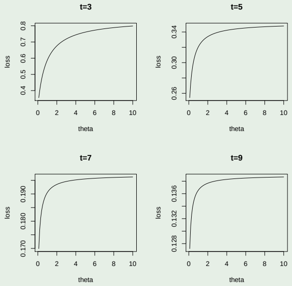
## Example

Maximum information loss for *WD*:



## Example

Maximum information loss for *WD*:



## Autoregressive Variance Model:

$$E(y_{ij}) = \alpha_i + \tau_{d(i,j)} + \rho_{d(i-1,j)},$$
$$Var(y_{ij}) = \sigma^2 + \sigma_\beta^2,$$

$$Cov(y_{ij}, y_{i'j'}) = \begin{cases} 0 & \text{if } j \neq j' \\ \sigma_\beta^2 \rho^{|i-i'|} & \text{if } j = j' \text{ and } i \neq i' \end{cases}$$

$i = 1, 2, \cdots, p$ and $j = 1, 2, \cdots, n$.

$$Var(Y_{\cdot j}) = \sigma_\beta^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^{p-1} \\ \rho & 1 & \cdots & \rho^{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & 1 \end{pmatrix}$$

## Example

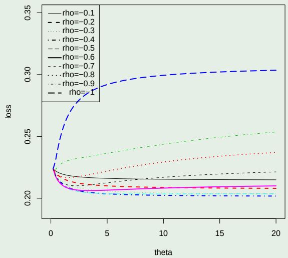



Figure: Maximum Loss of WD $t = 5$, $\rho < 0$

## Example

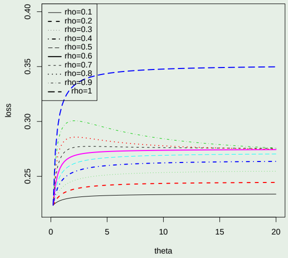



Figure: Maximum Loss of WD $t = 5$, $\rho > 0$

## Example

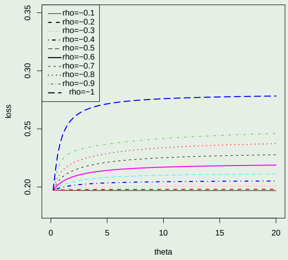



Figure: Maximum Loss of $d^*$ $t = 5$, $\rho < 0$

## Example

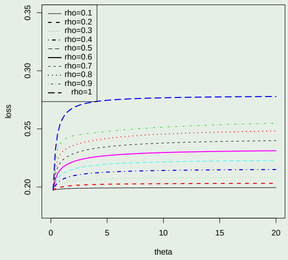



Figure: Maximum Loss of $d^*$ $t = 5$, $\rho > 0$

**Thank you!**