

Index of Lecture 8: Experimental design

Page	Title
<hr/>	
1	Practical information
2–3	Introduction to experimental design
4	A checklist for planning experiments
5	Block design versus completely randomised design
6	Statistical analysis of block designs
7	(Unreplicated) Block design – Example
8	Latin square design
9	Latin square data example
10	Mangold data results
11	Incomplete block designs
12	Balanced incomplete block design (BIBD)
13	BIBD examples
14	Youden square

PRACTICAL INFORMATION

Today's lecture:

- principles of experimental design, with brief overview of ideas and methods,
- a range of classical block designs (many more exist, also in Oehlert textbook),
- more focus on data analysis than on design construction.¹

Guidelines for textbook reading:

- **skip** entirely: discussions of efficiency (13.2.3, 13.3.5, pp. 362₇—368¹¹) and advanced sections (13.5, 14.3—7),
- **skip** for now: references to Chapters 11-12, incl. Hasse diagrams,
- **postpone** (Session 9): cross-over designs (13.3.1, 13.3.6).

Other news:

- home assignment 3 due on Monday (March 3),
- definitely time to make progress on your project (proposal).

¹ More detailed references, in particular for design construction:

- * Cochran & Cox: *Experimental Designs* (UPEI library),
- * Box, Hunter & Hunter: *Statistics for Experimenters* (UPEI library),
- * Dean, Voss & Draguljic: *Design and Analysis of Experiments*, 2nd ed., 2017 (UPEI library).

INTRODUCTION TO EXPERIMENTAL DESIGN

What characterises an experiment?²

the experimenter imposes **treatments** onto the **experimental units** from which responses are measured.

Advantages of an experiment:

- allows (potentially) to **conclude about causation**,³
- can be designed to give **small error** in comparisons,
- can be designed to **avoid/minimise bias** in comparisons.

Plain glossary for experimental design:

- **treatments** (short form: tx's): the different procedures to compare,
- **experimental units** (EUs): the “things/subjects” to which treatments are applied,
- **responses**: outcomes of interest to compare, observed at EUs after treatments are applied,
- **randomisation**: random assignment of treatments to experimental units (to avoid unexpected patterns),

² As opposed to observational studies.

³ The reasoning goes: “If experimental units differ only (ideally) by their treatments, any differences beyond random fluctuations in their responses must be caused by the treatments”.

- **experimental error**: variation between equally treated units (both population and observed values),
- **measurement units**: the actual objects on which responses are measured (may be smaller than the EUs),
- **blinding**: response evaluators do not know the treatment allocation,
- **control**: treatment consisting of “no treatment”, possibly **placebo** (inactive treatment),⁴
- **factor**: (controlled) explanatory, categorical variable,
- **level/category**: specific value of factor/treatment,
- **replication**: multiple “identical” experimental units (sometimes, confusingly, repetitions of the entire experiment⁵),⁶
- **blocking**: division of EUs into homogeneous groups (discussed from 8L–5 onward),
- **balancedness** = (generally) all treatment levels occur the same number of times,
 - * **desirable** property of a design, but not indispensable,⁷
 - * gives **same precision** on all treatment estimates and comparisons, and simple computational formulae,
- (technical) **confounding**: the effect of one factor/treatment cannot be distinguished from that of another factor/ treatment (different meaning than in epidemiology!).

⁴ The main purpose of a control group is to control lurking variables/confounders.

⁵ Repetition of an entire experiment would usually be considered as a blocking factor.

⁶ The purpose of replication is to enable estimation of experimental error.

⁷ Balancedness is necessary for some ANOVA methods – however, (general) **linear model methods** always apply.

A CHECKLIST FOR PLANNING EXPERIMENTS

(from Dean & Voss, Chapter 2, pages 7–14)⁸

- (a) Define the objectives of the experiment.
- (b) Identify all sources of variation, including:
 - (i) treatment factors and their levels,
 - (ii) experimental units,
 - (iii) blocking factors, noise factors, and covariates.
- (c) Choose a rule for assigning the experimental units to the treatments (the narrow meaning of experimental design).
- (d) Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.
- (e) Run a pilot experiment.
- (f) Specify the (statistical) model.
- (g) Outline the (statistical) analysis.
- (h) Calculate the number of observations that need to be taken.
- (i) Review the above decisions. Revise if necessary.

⁸ Chapter 2 is also available at the VHM 802 Moodle account.

BLOCK DESIGN VERSUS COMPLETELY RANDOMISED DESIGN

Blocks = groups of homogeneous experimental units (EUs), so that

- EUs are **more alike within than between groups**, before and during experiment, possibly because:
 - * EUs within blocks **share inherent characteristics** that cannot be assigned to them (e.g., gender, location, unquantifiable “management”).

Contrasting block and completely randomised designs:⁹

Characteristic	Block design	Completely randomised design
Randomization of tx's	within EUs per block	among all EUs
Replication of tx's	possible, not necessary ≤ 1 or ≥ 1 EU per tx per block	necessary > 1 EU per tx
Error variance	among EUs per block (if no tx's)	among all EUs (if no tx's)
Balancedness	different meanings, simplest: same # EU per tx per block	same # EU per tx
Model impact	include block effects	none

How to distinguish completely randomized and block designs?

— (GO) follow the randomisation!

⁹ Basic introductions to the two designs were given in Lecture 2 of VHM 801.

STATISTICAL ANALYSIS OF BLOCK DESIGNS

Statistical model for observations (y_{ij}) on treatments ($i = 1, \dots, g$) in blocks ($j = 1, \dots, r$):

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \text{where}$$

- α_i and $\beta_j \sim$ effects of treatment i and block j , respectively,
- the errors $\varepsilon_{11}, \dots, \varepsilon_{gr}$ are i.i.d. and $\sim N(0, \sigma^2)$.
- block effects are (usually) **additive** to treatments:
 - * **assumes** that tx effects do not depend on blocks (“should” be true?),
 - * most useful when tx by block interaction believed to be absent or weak,¹⁰
 - * **non-additive block effects**: only possible with replicated treatments within blocks or a factorial treatment structure (and omitted interactions).

Statistical analysis follows usual ANOVA/linear model principles, except that

- block effects often of less interest (for interpretation),
- blocks are usually not randomized \Rightarrow (GO) tests for block effects could be omitted entirely,
- possible to compute efficiency of block design relative to completely randomized design (“did the block design work?”), see GO for details (not part of the course).

¹⁰ Ignoring an existing tx by block interaction \Rightarrow inflated σ^2 and loss of power.

(UNREPLICATED) BLOCK DESIGN – EXAMPLE

Mealybug example: (GO Example 13.1)

- changes (before minus after treatment) in counts of mealybugs,
- 3 treatments (water, spores, oil), 5 plants (\sim blocks),

○ **data layout:**

Treatment \ Plant	1	2	3	4	5
water	-9	18	10	9	-6
	-6	5	9	0	13
spores	-4	29	4	-2	11
	7	10	-1	6	-1
oil	4	29	14	14	7
	11	36	16	18	15

○ **experimental unit:**
branches of cycad plants,

○ **measurement units:**
two patches per branch
treated the same way,

○ **outcome:** average change across the two patches per branch.

ANOVA table:

Source	DF	SS	MS	<i>F</i>	<i>P</i> -value
Treatments	2	432.0	216.0	12.2	.004
Plants	4	686.4	171.6	(9.7)	(.004)
Error	8	141.8	17.7		
Total	14	1260.2			

* **clear treatment effects:**
oil treatment has larger
change than the others:

water (4.3)^b, spores (5.9)^b, oil (16.4)^a,

* plant effects also clearly significant (but less important).

LATIN SQUARE DESIGN

- **definition:** designs with
 - * g treatments,
 - * two blocking variables,
 - * g^2 experimental units arranged in a square,
 and every treatment **exactly once in each row / column**,

- **examples:** (cyclic Latin squares, plus permutations)

$g=2$:

A	B
B	A

B	A
A	B

$g=3$:

A	B	C
B	C	A
C	A	B

A	C	B
C	B	A
B	A	C

$g=4$:

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

- **advantage:** takes into account two blockings
 \Rightarrow potentially more accurate (lower experimental error),
- **disadvantage:** specific requirements for dimensions of design (number of blocks, number of experimental units),
- **randomisation** between squares: not easy to do properly¹¹;
 — a simple and “crude” method randomly permutes rows, columns and symbols,
- **modelling:** additive effects of rows and columns,
- **balanced** in the treatments \Rightarrow simple analysis (due to orthogonal factors).

¹¹ Fisher-Yates procedure based on **standard** squares tabled, see GO p. 327.

LATIN SQUARE DATA EXAMPLE

Mangold (wurzel) example: (Mercer & Hall 1911; field trial)

- yield of plots located in a 5×5 Latin square and with 5 (unspecified) treatments,

y_{ijk} = weight of mangold in plot with treatment i at location (j, k) ,

$i = A, B, C, D, E \sim$ treatments,

$j, k = 1, 2, 3, 4, 5 \sim$ rows and columns,

technical note: only some sets (i, j, k) occur,

- **data layout:**
(treatments and yields)

D 376	E 371	C 355	B 356	A 335
B 316	D 338	E 336	A 356	C 332
C 326	A 326	B 335	D 343	E 330
E 317	B 343	A 330	C 327	D 336
A 321	C 332	D 317	E 318	B 306

- **statistical model** (additive treatments and blocks):

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk},$$

where the ε_{ijk} 's are i.i.d. and $\sim N(0, \sigma^2)$.

Note: Latin squares can be extended to allow for two treatments (still two blockings)
→ Graeco-Latin squares (GO Section 13.4).

MANGOLD DATA RESULTS

ANOVA table

for additive model:

Source	DF	SS	MS	<i>F</i>	<i>P</i> -value
Rows	4	4240.2	1060.1	(7.25)	(.003)
Columns	4	701.8	175.5	(1.20)	(.36)
Treatments	4	330.2	82.6	0.56	.69
Error	12	1754.3	146.2		
Total	24	7026.6			

- **no indication** of treatment effects, but D highest; we explore the contrast comparing D to the average of others:

$$\hat{w} = 4\hat{\alpha}_D - \hat{\alpha}_A - \hat{\alpha}_B - \hat{\alpha}_C - \hat{\alpha}_E$$

$$= 4\bar{y}_{D..} - \bar{y}_{A..} - \bar{y}_{B..} - \bar{y}_{C..} - \bar{y}_{E..} = 34.4,$$

$$SE(\hat{w}) = \sqrt{MS_E (4^2 + 1^2 + 1^2 + 1^2 + 1^2)/5} = 24.18,$$

$$t = \hat{w}/SE(\hat{w}) = 34.4/24.2 = 1.42 \quad \sim \text{clearly NS even if pre-planned,}$$

$$SS(\hat{w}) = t^2 MS_E = 295.84 \quad \sim 90\% (295.84/330.24) \text{ of variation,}$$

- **no column effects**, but **clear row effects**:

row 1 > rows 2–4 > row 5 (maybe not particularly interesting).

Further analyses:

- **model reduction**: possible to drop non-significant effects, but not necessary,
- **diagnostics**: all look good.

INCOMPLETE BLOCK DESIGNS

Incomplete: the blocks do not comprise all treatments (combinations of tx factors).

Why of **potential interest**?

- maybe limited block sizes (e.g., litters or twins),
- maybe save experimental units (economy),
- missing values \Rightarrow incomplete blocks.

Main points about incomplete block designs:

- **unbalanced** (in an extreme form) \Rightarrow order of testing for factors is important: partial SS \neq sequential SS,¹²
- “**balanced**” **incomplete block design** (BIBD):
balanced in the sense that all pairs of treatments occur equally often in the same block \Rightarrow same precision on all treatment comparisons,
- many specialised designs exist with certain properties (GO 14.3–7; not in course),
- **statistical analysis:** using methods for (general) linear models, in particular **least squares means**.¹³

¹² Recall that **partial/adjusted SS** \sim removing effect while keeping all others, and **sequential SS** \sim removing effects sequentially (bottom up in ANOVA table).

¹³ Recall that **least squares means** are adjusted for other factors by giving their levels/categories equal weights; this would seem the correct approach when unequal representation occurs by design.

BALANCED INCOMPLETE BLOCK DESIGN (BIBD)

A **BIBD** with g treatments and b blocks must meet the requirements,

- all treatments occur the same number of times, r ,
- all blocks are of the same size, k ,
- every pair of treatments “meet” the same number of times, λ ,

Then the following relations hold:

$$rg = bk \quad \text{and} \quad \lambda(g - 1) = r(k - 1).$$

Construction of BIBD's:

- designs do not always exist (even if above relations satisfied),
- designs tabled in textbooks, e.g. GO or Cochran & Cox.

Statistical model has additive treatment and block effects:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, \dots, g \ (\sim \text{treatments}), \quad j = 1, \dots, b \ (\sim \text{blocks}),$$

and where not all pairs (i, j) occur.

Technical notes on analysis:

- treatment SE computable by hand, using “effective sample size” ($\lambda g/k$) instead of r ,
- “intrablock” analysis \sim linear model; “interblock” analysis \sim random effects model.

BIBD EXAMPLES

Simplest BIBD:

$(g=3, b=3, r=2, k=2, \lambda=1)$:

Block	1	2	3
Treat-	A	B	C
ment	B	C	A

Dish detergent example: (GO Example 14.2)

- count plates before foam disappears, in “sessions” with 3 simultaneous operators and 3 basins,

y_{ij} = number of plates for detergent i in session j

$i = A, B, C, D, E, F, G, H, J \sim$ detergents, treatments

$j = 1, \dots, 12 \sim$ sessions, blocks,

- experimental layout:

Block	1	2	3	4	5	6	7	8	9	10	11	12
Treat-	A	D	G	A	B	C	A	B	C	A	B	C
ment	B	E	H	D	E	F	E	F	D	F	D	E
	C	F	J	G	H	J	J	G	H	H	J	G

- BIBD with $g=9, b=12, r=4, k=3, \lambda=1$,

- further treatment structure:

- * A,B,C,D \sim detergent I + doses 3,2,1,0 of additive,
- * E,F,G,H \sim detergent II + doses 3,2,1,0 of additive,
- * J \sim control.

YOUDEN SQUARE

- **incomplete block design** with **two blocking factors**, e.g. Mangold data with last column omitted:

D 376	E 371	C 355	B 356	(A .)
B 316	D 338	E 336	A 356	(C .)
C 326	A 326	B 335	D 343	(E .)
E 317	B 343	A 330	C 327	(D .)
A 321	C 332	D 317	E 318	(B .)

- **design properties**:¹⁴
 - * first blocking factor (“rows”) and treatments form a BIBD (example: $g = 5$, $b = 5$, $r = 4$, $k = 4$, $\lambda = 3$),
 - * second blocking factor (“columns”) has every treatment once per block (\sim complete and balanced blocks),
- **advantages**: retains “nice” analysis for treatments from BIBD while allowing for two blocking factors,
- **note**: a Youden square is **not** a square(!), so name probably originates from the relation to Latin squares (dropping one row or one column).

¹⁴ These two properties may also be taken as the definition of more general Youden squares, see GO Section 14.2.