

Index of Lecture 13: Experimental design

| Page | Title |
|------|--|
| 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 | Introduction to cross-over designs |
| 15 | Simplest design: AB/BA |
| 16 | Analysis of AB/BA design |
| 17 | More treatments and periods |
| 18 | Combining Latin squares |
| 19 | Latin square cross-over trial examples |

PRACTICAL INFORMATION

Today's lecture:

- principles of experimental design, with brief overview of ideas and methods,
- a range of classical block designs, including cross-over designs,
- 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, 14.2-7), as well as refs to Chapters 11-12, incl. Hasse diagrams,
- **cross-over designs**: GO Sections 13.3.1, 13.3.4, 13.3.6, 16.7.²

Other news:

- lab and presentation sessions next week scheduled,
- project report added in Moodle (deadline April 28, worth 60% of course mark).

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: *Design and Analysis of Experiments* (UPEI library e-book collections; link via course homepage).

² Lecture notes also incorporate material from:

[1] Senn (2005), Crossover Designs, In: *Encyclopedia of Biostatistics* (available electronically at UPEI library),

[2] Toutenburg (2002), *Statistical Analysis of Designed Experiments*, Chapter 9 (also electronic access at UPEI library).

INTRODUCTION TO EXPERIMENTAL DESIGN

What characterises an experiment?³

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

Advantages of an experiment:

- allows (potentially) to **conclude about *causation***,⁴
- 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),

³ As opposed to observational studies.

⁴ 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),⁵
- **factor**: (controlled) explanatory, categorical variable,
- **level/category**: specific value of factor/treatment,
- **replication**: multiple “identical” experimental units (sometimes, confusingly, repetitions of the entire experiment⁶),⁷
- **blocking**: division of EUs into homogeneous groups (discussed from 7L–5 onward),
- **balancedness** = (generally) all treatment levels occur the same number of times,
 - * **desirable** property of a design, but not indispensable,⁸
 - * 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!).

⁵ The main purpose of a control group is to control lurking variables/confounders.

⁶ Repetition of an entire experiment would usually be considered a blocking factor.

⁷ The purpose of replication is to enable estimation of experimental error.

⁸ 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)⁹

- (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.

⁹ 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:¹⁰

| 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!

¹⁰ 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,¹¹
 - * **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 course).

¹¹ 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 | F | P-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
 $\left\{ \begin{array}{l} * \text{ } g \text{ treatments,} \\ * \text{ two blocking variables,} \\ * \text{ } g^2 \text{ experimental units arranged in a square,} \end{array} \right.$

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 properly¹²;
 — 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).

¹² 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:

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

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

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

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

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

- **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,¹³
- **“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 14.2–7; not in course),
- **statistical analysis:** using methods for (general) linear models, in particular **least squares means**.¹⁴

¹³ Recall that **partial/adjusted SS** \sim removing effect while keeping all others, and **sequential SS** \sim removing effects sequentially (bottom up in ANOVA table).

¹⁴ 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.

INTRODUCTION TO CROSS-OVER DESIGNS

Definition of a cross-over trial:

“a trial in which individual subjects are given sequences of treatments with the object of studying differences between individual treatments (or subsequences of treatments)” [1]

- **idea:** each subject forms a block (or one's own control)
⇒ reduced variation (by eliminating between-subject variation),
- **advantages:**
 - * (potentially) (much) higher efficiency than completely randomized design,
 - * individuals' reactions to treatments can be studied,
- **drawbacks:**
 - * longer execution time ⇒ logistical challenges, greater risk of subject dropouts,
 - * more complex data analysis (due to added risk of bias caused by carry-over effects), and really a special case of **repeated measures**,
- simple 3×3 **Latin square** example: Bioequivalence (GO Example 13.6).

Common terminology:

- **periods:** occasions where subjects are treated (and measured),
- **carry-over effect:** residual effect from one period to the next,
- **wash-out period:** interval between tx periods to eliminate/reduce carry-over effects.

SIMPLEST DESIGN: AB/BA

- two treatments (A and B),
- two periods (1 and 2),
- two groups of subjects:

| | Period 1 | Period 2 |
|-----------------|----------|----------|
| subject group 1 | A | B |
| subject group 2 | B | A |

Data example: Scents and learning¹⁵ — can pleasant (floral) aromas help a student learn better?

- completion times for pencil and paper mazes for 21 subjects with scented (S) and unscented (U) masks,
- 11 subjects used unscented masks first, 10 subjects scented masks first (1 subject excluded):

| Subject | Sequence | mean U | mean S | diff U–S | diff 1–2 |
|---------|----------|--------|--------|----------|----------|
| 1 | US | 30.6 | 38.0 | -7.4 | -7.4 |
| 2 | SU | 48.4 | 51.6 | -3.2 | 3.2 |
| 3 | US | 60.8 | 56.7 | 4.1 | 4.1 |
| 4 | SU | 36.1 | 40.5 | -4.4 | 4.4 |
| 5 | US | 68.5 | 49.0 | 19.5 | 19.5 |
| 6 | SU | 32.4 | 43.2 | -10.8 | 10.8 |
| ... | ... | ... | ... | | |

¹⁵ Reduced data set (Mazes and smells) available at: <https://dasl.datadescription.com>.

ANALYSIS OF AB/BA DESIGN

Assuming **no carry-over effects**, simple approaches will work:

- compare treatments by two-sample analysis (e.g. *t*-test) for differences 1–2,¹⁶
- compare periods by two-sample analysis (e.g. *t*-test) for differences A–B.

Combined analysis of response y_{ijk} in period j for k 'th subject with treatment sequence i , using the model

$$y_{ijk} = \mu_{ij} + s_{ik} + \varepsilon_{ijk}, \quad \text{where}$$

- * μ_{ij} is the mean for sequence i in period j , given by:

| sequence i | period 1 ($j=1$) | period 2 ($j=2$) |
|--------------|----------------------------|---|
| AB | $\mu + \alpha_A + \beta_1$ | $\mu + \alpha_B + \beta_2 + \lambda_{AB}$ |
| BA | $\mu + \alpha_B + \beta_1$ | $\mu + \alpha_A + \beta_2 + \lambda_{BA}$ |

$\alpha_A, \alpha_B \sim$ **treatment** effects; $\beta_1, \beta_2 \sim$ **period** effects; $\lambda_{AB}, \lambda_{BA} \sim$ **carry-over** effects,

- * s_{ik} is the effect of subject k with treatment sequence i ,
- * ε_{ijk} is the error term $\sim N(0, \sigma^2)$,
- * **same inference** as above if $\lambda_{AB} = \lambda_{BA} = 0$,
- * **complex analysis** ([1],[2]) if λ 's $\neq 0$.

¹⁶ Treatment comparison by a paired two-sample analysis for measurements for A and B (effectively a one-sample analysis for differences A–B) is only valid when no period effects exist.

MORE TREATMENTS AND PERIODS

Examples of extensions of AB/BA design:

- **2 tx, > 2 periods:** enables modelling of carry-over effects in analysis, e.g. for sequences AABB and BBAA:

| Carry-over model | Sequence | | | | Sequence | | | |
|-------------------|----------|----------------|----------------|----------------|----------|----------------|----------------|----------------|
| | A | A | B | B | B | B | A | A |
| “prev. period” | – | λ_A | λ_A | λ_B | – | λ_B | λ_B | λ_A |
| “change only” | – | 0 | λ_{AB} | 0 | – | 0 | λ_{BA} | 0 |
| “prev. + present” | – | λ_{AA} | λ_{AB} | λ_{BB} | – | λ_{BB} | λ_{BA} | λ_{AA} |

- **3 tx, 3 periods:** to ensure balancedness of tx’s in periods traditionally laid out in Latin squares, e.g.,

| | |
|-------|-------|
| A B C | A C B |
| B C A | C B A |
| C A B | B A C |

where rows \sim periods, columns \sim subjects, symbols \sim tx,

- * simple analysis when no carry-over effects,
- * desirable to include all sequences by combining two different Latin squares (as shown above),

- **g tx, g periods:** use (multiple) $g \times g$ Latin squares,
- **g tx, $k < g$ periods:** use incomplete block design, preferably BIBD (g, b, k, r, λ) .

COMBINING LATIN SQUARES

Multiple Latin squares in same design:

- increases the error degrees of freedom \Rightarrow larger power,
- several extra options for modelling, depending on data context.

Notation/Model:
$$\begin{cases} y_{ijkl} = \text{outcome for tx } i \text{ in row } j \text{ and column } k \text{ in square } l, \\ y_{ijkl} = \mu + \alpha_i + \beta_{j(l)} + \gamma_{k(l)} + \varepsilon_{ijkl}, \end{cases}$$

- row ($\beta_{j(l)}$) and column ($\gamma_{k(l)}$) effects “nested in” (separate for) squares \sim different effects across squares,
- a block effect may be assumed the same in all squares, e.g. for periods:
 - * **same effects** if all subjects go through same periods,
 - * **different effects** if periods are not the same (e.g. due to different ages of subjects),

Modelling refinements:

- **square type**¹⁷ interactions: carry-over effects may show up as interactions between square type and periods,
- **residual effects** may be modelled directly to split each tx effect into “direct” and “residual” effects.

¹⁷ The square type is determined by the carry-over combinations it contains.

LATIN SQUARE CROSS-OVER TRIAL EXAMPLES

Bioequivalence trial (GO Example 13.10) with 12 subjects in 4 Latin squares.

Milk production example ~ cross-over trial (GO Example 13.12):

- o milk yield of cows during three periods with different diets,

y_{ijkl} = yield for cow k in square l in period j on diet i
 i = A,B,C ~ diets (roughage, limited grain, full grain)
 j = 1, 2, 3 ~ period (for each cow),
 k = 1, ..., 3 ~ cow number (within squares)
 l = 1, ..., 6 ~ Latin square number.

| | | Cow | | | Cow | | | Cow | | |
|--------|--|-----|---|---|-----|----|----|-----|----|----|
| Period | | 1 | 2 | 3 | 7 | 8 | 9 | 13 | 14 | 15 |
| 1 | | A | B | C | A | B | C | A | B | C |
| 2 | | B | C | A | B | C | A | B | C | A |
| 3 | | C | A | B | C | A | B | C | A | B |
| | | Cow | | | Cow | | | Cow | | |
| Period | | 4 | 5 | 6 | 10 | 11 | 12 | 16 | 17 | 18 |
| 1 | | A | B | C | A | B | C | A | B | C |
| 2 | | C | A | B | C | A | B | C | A | B |
| 3 | | B | C | A | B | C | A | B | C | A |

- o 6 separate Latin squares,
 - * two types of Latin squares (top/bottom), 3 replicates of each,
 - * top ~ diet order AB, BC, CA; bottom ~ AC, CB, BA,
- o basic statistical model (additive, no square effects),

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_{k(l)} + \varepsilon_{ijkl}, \text{ or}$$

$$y_i = \mu + \alpha_{\text{diet}(i)} + \beta_{\text{per}(i)} + \gamma_{\text{cow}(i)} + \varepsilon_i,$$

where the errors are i.i.d. and $\sim N(0, \sigma^2)$.