Index of Lecture 8

Practical Information

Experimental design in this course:

- brief overview of ideas and methods,
- more focus on data analysis than on design construction,
- more detailed references – textbook extras, and
  - Cochran & Cox: Experimental Designs (UPEI library),
  - Box, Hunter & Hunter: Statistics for Experimenters (UPEI lib.),
  - Dean & Voss: Design and Analysis of Experiments[1].

Today's lecture:

- follow-up from lab?
- principles of experimental design,
- a range of classical block designs.

Guidelines for textbook reading:

- skip entirely: discussions of efficiency (13.2.3, 13.3.5, pp. 362_7–368^{11}) and advanced sections (13.5, 14.3-7),
- skip for now: refs to Ch. 11-12, incl. Hasse diagrams,
- postpone (Session 9): cross-over designs (13.3.1, 13.3.6).

Other news:

- home assignments: #3 due on Monday (29/2),
- definitely time to make progress on your project.

[1] Available through UPEI library e-book collections; link via course homepage.

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: the different procedures to compare,
- experimental units: the "things/subjects/individuals" to which treatments are applied,
- responses: outcomes of interest to compare, observed at the units after treatments are applied,
- randomisation: random assignment of treatments to experimental units (to avoid unexpected patterns),
- experimental error: variation between equally treated units (both population and observed values),

[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".

- measurement units: the actual objects on which responses are measured (may be smaller than the experimental units),
- 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 experiment[5]),[6]
- blocking: division of experimental units into homogeneous groups (in order to reduce experimental error),
- balancedness
  = all treatment levels occur the same number of times,
  - *desirable* property of a design, but *not indispensable*,
  - gives same precision on all treatment estimates and comparisons, and simple computational formulae,
  - necessary for some ANOVA methods – however, (general) linear model methods always apply,
- (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 a blocking factor.
[6] The purpose of replication is to enable estimation of experimental error.

A Checklist for Planning Experiments

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

(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.

[7] Chapter 2 is also available at the VHM 802 Moodle account.

Completely Randomised Design

In a completely randomised design, all treatments are allocated at random among the experimental units (using a method for random selection). In all other respects, the units are treated as equally as possible.
$\Rightarrow$ (idea/rationale:)
Differences in response must be due to either treatment effects or play of chance in random assignment of units.

Comments:

+ simple/easy to understand, carry out and analyse,
+ flexible (allows any number of levels and replications),
− all experimental units need to be "homogeneous", otherwise the random variation will be large,
− if a good grouping of experimental units exists (either in their state before treatments are applied or in general conditions during the experiment), other designs will be more efficient (give greater precision),

- primarily for small designs with no obvious grouping,
- completely randomised designs with a large number of replications (and DF_E) are "suspect" (rarely realistic).

Randomisation may be achieved by randomly reordering experimental units (easy in statistical software).

(Randomised) Block Design

Blocks = groups of homogeneous experimental units, i.e., units are more alike within than between groups, before and during experiment.

In a randomised block design, treatments are assigned randomly to the units within each block; typically, each treatment occurs once (≥ 1: *complete* design) per block.
$\Rightarrow$ (idea/rationale:)
More accurate to compare similar units (in same block) and to aggregate such comparisons across blocks.

Examples of factors used to form blocks:

- agriculture: areas in fields,
- animal science: litters, groups (age, weight, sex), environment (herd),
- human medicine: twins (matched pairs design), family, groups (as above + health, social status, lifestyle...),
- general: time, operator (surgeon, technician).

Comments on block designs:

\+ improvement in accuracy if efficient groups available,

+/− minor added complexity in design and analysis,

− less flexible (block size should match no. treatments),

Randomisation: by randomly reordering experimental units within each block.

Statistical Analysis of Block Designs

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

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

where

- α_i and $\beta_j \sim$ effects of treatment i and block j, resp.,
- the errors $\varepsilon_{11}, \ldots, \varepsilon_{gr}$ are i.i.d. and $\sim \mathrm{N}(0, \sigma^2)$.
- block effects are (usually) additive to treatments:
  - assumes that tx effects do not depend on blocks,
  - most useful when tx by block interaction believed to be absent or weak,[8]
  - 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 $\Rightarrow$ tests for block effects could be omitted entirely (GO),
- (technical) possible to compute *efficiency* of block design relative to compl. rand. design ("did the block design work?"), see GO for details (not part of course).

[8] 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 – after treatment) in counts of mealybugs,
- 3 treatments (water, spores, oil), 5 plants ($\sim$ blocks),
- experimental unit: branches of cycad plants,
- measurement units: two patches per branch treated the same way,
- 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

- outcome: average change across two patches per branch.

ANOVA table for mealybug data:

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 treatments effect: 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: design 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

$g = 4$:

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

$g = 3$:

A	B	C
B	C	A
C	A	B

A	C	B
C	B	A
B	A	C

- advantage: takes into account two blockings
  $\Rightarrow$ potentially more accurate (lower exper. error),
- disadvantage: specific requirements for dimensions of design (no. of blocks, no. of experimental units),
- randomisation between squares: not easy to do properly[9]; a simple and "crude" method randomly permutes rows, columns and symbols,
- modelling: additive effects of rows and columns,
- balanced in the treatments: "nice", simple analysis.

[9] 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,

$$\begin{aligned} y_{ijk} &= \text{weight of mangold in plot with} \\ &\quad \text{treatment } i \text{ at location } (j,k), \\ i &= \text{A,B,C,D,E} \sim \text{treatments}, \\ j,k &= 1,2,3,4,5 \sim \text{rows and columns}, \end{aligned}$$

  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 $\sim \text{N}(0, \sigma^2)$.

Extension of design to allow for two treatments: Graeco-Latin squares (Section 13.4).

Mangold Data Results

ANOVA table for additive model:

Source	DF	SS	MS	F	P-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; contrast comparing D to average of others:

$$\begin{aligned}
\hat{w} &= 4\hat{\alpha}_D - \hat{\alpha}_A - \hat{\alpha}_B - \hat{\alpha}_C - \hat{\alpha}_E \\
&= 4\bar{y}_{D\cdot\cdot} - \bar{y}_{A\cdot\cdot} - \bar{y}_{B\cdot\cdot} - \bar{y}_{C\cdot\cdot} - \bar{y}_{E\cdot\cdot} = 34.4, \\
\mathrm{SE}(\hat{w}) &= \sqrt{\mathrm{MS}_E\,(4^2+1^2+1^2+1^2+1^2)/5} = 24.18, \\
t &= \hat{w}/\mathrm{SE}(\hat{w}) = 34.4/24.2 = 1.42, \\
\mathrm{SS}(\hat{w}) &= t^2\,\mathrm{MS}_E = 295.84,
\end{aligned}$$

  - accounts for 90% (295.84/330.24) of variation,
  - clearly non-significant even if pre-planned,

- no column effects, but clear row effects: row 1 > rows 2−4 > row 5 (maybe not particularly interesting).

Further analyses:

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

Incomplete Block Designs

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

Why of interest?

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

Main points about incomplete block designs:

- unbalanced (in an extreme form) ⇒ order of testing for factors is important: partial SS ≠ successive SS,[10]
- "balanced" incomplete block design (BIBD): balanced in the sense that all pairs of treatments occur equally often in the same block ⇒ same precision on all treatment comparisons,
- many specialised designs exist with certain properties (GO Sections 14.3-7; not part of course),
- statistical analysis: using methods for (general) linear models, in particular *least squares means*[11].

[10] Recall that partial/adjusted SS ∼ removing effect while keeping all others, and sequential SS ∼ removing effects sequentially (bottom up).

[11] 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 balanced incomplete block design with g treatments and b blocks must meet the requirements,

- all treatments occur the same no. of times, r,
- all blocks are of the same size, k,
- every pair of treatments "meet" the same no. 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 relations satisfied),
- designs tabled in textbooks, e.g. GO or Cochran & Cox.

Statistical model has additive treatments and blocks:

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

where $i = 1, \ldots, g \sim$ treatments and $j = 1, \ldots, b \sim$ blocks, and not all pairs (i, j) occur. Alternative model formulation:

$$y_i = \mu + \alpha_{\text{treat}(i)} + \beta_{\text{block}(i)} + \varepsilon_i.$$

Technical notes on analysis:

- treatment SE computable by hand, using "effective sample size" $(\lambda g/k)$ instead of r.
- "intrablock" analysis $\sim$ usual linear model, and "interblock" analysis $\sim$ random effects model (Session 9).

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 no. of plates before foam disappears, in "sessions" with 3 simultaneous operators and 3 basins,

  y_{ij} = no. of plates for detergent i in session j

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

  j = $1, \ldots, 12$ $\sim$ 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:[12]
  - 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 relation to Latin squares (dropping one row or one column).

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