Linear statistical models
Experimental design

Notes by Yao-ban Chan and Owen Jones

Experimental design

When we conduct an experiment involving two variables A and B, we seek to understand how A and B are related.

Does A cause B (causality)? Or vice versa? Or does something else affect both of them?

An **observational study** can not determine causality, only correlation. Instead you need a **designed experiment/trial**. Even then, we can never be sure.

Hill's criteria for determining causal links

- Strength of association
- Consistency of association (from one study to another)
- Consistent with existing knowledge
- * Monotonic response (increasing A makes B more likely)
- * Temporal relationship (A must come before B)
- * Plausibility of alternatives (is there another explanation?)
- Predictive value of link (can you use it?)

[A.B. Hill 1897–1991]

A designed experiment can help with the *'s.

Examples: false causality

Temporal:

- The more firemen, the bigger the fire, therefore firemen cause fires.
- The more tourists waiting, the sooner Old Faithful will erupt, therefore tourists cause eruptions.

Inconsistency of association:

- Since the 1600's there has been global warming and a decrease in pirates, therefore pirates help stop global warming.

Examples: false causality

Confounding factor:

- Sleeping with shoes on causes headaches?
- Heavy drinking causes lung cancer?
- Sleeping with the light on as a child increases the risk of myopia?

In each case there is a better explanation, based on a confounding factor:

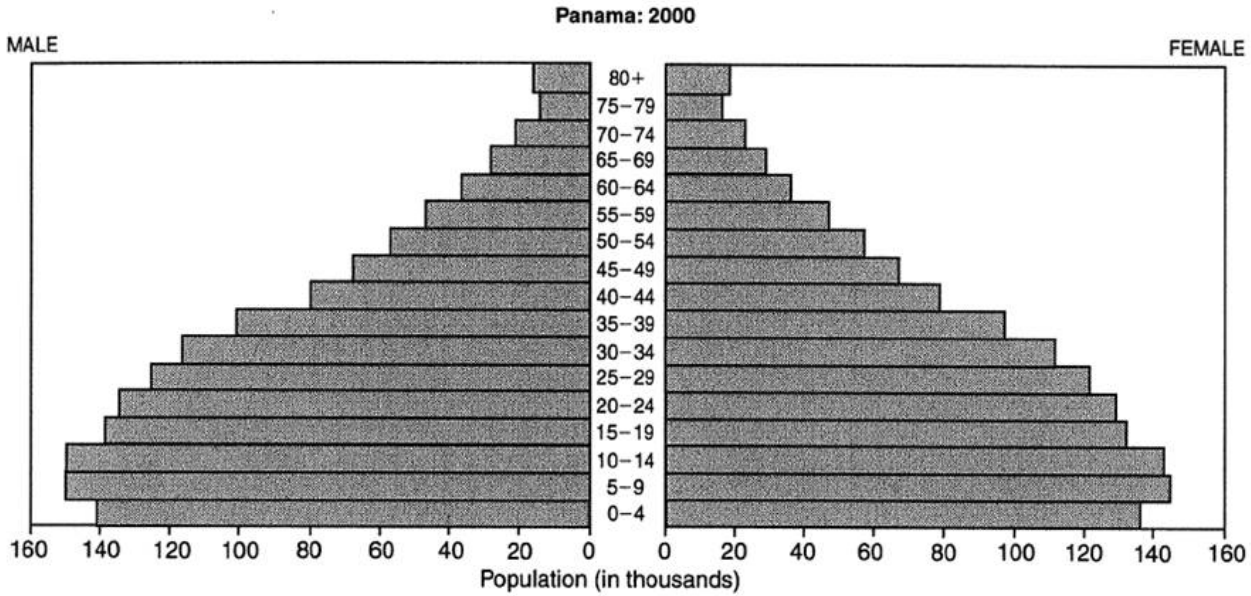
- Drinking the night before;
- Passive smoking;
- Myopic parents.

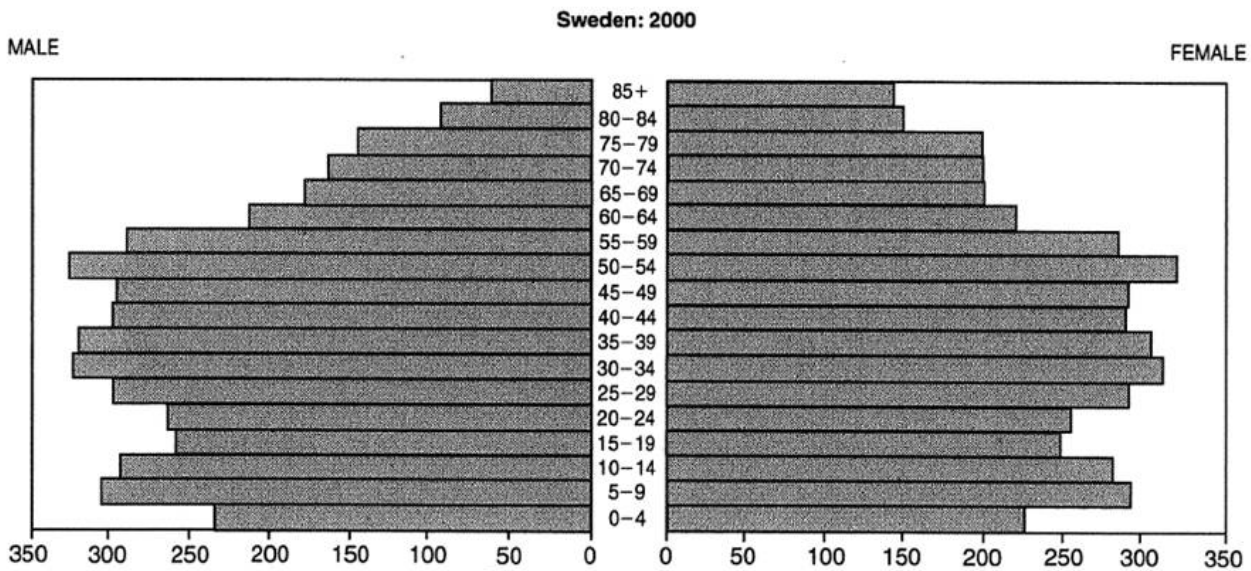
Example: life expectancy in Sweden and Panama

"Common sense" tells us that the residents of Sweden should have lower death rates than the residents of Panama.

But each year, a greater proportion of Swedish residents die. Why?

Example: life expectancy in Sweden and Panama





Example: life expectancy in Sweden and Panama

Older people die at a greater rate than younger people.

For individuals of the same age, the death rate among Swedes is less than the death rate among Panamanians.

BUT Sweden has a population that is older than that in Panama, so a greater proportion of Swedes die in any one year, despite the lower death rates within the age categories.

Age is a confounding factor when measuring death rates.

Example: smoking and cancer

Consider the following mortality data, summarised from a study that looked at the smoking habits of a group of female residents of Whickham, UK, in the period 1972-1974, and then tracked their survival over the next twenty years.

smoker	non-smoker	total
139/582	230/732	369/1314
(24%)	(31%)	(28%)

Only 24% of the women who were smokers at the time of the intial survey died during the 20-year follow-up period, whereas 31% of the non-smokers died in the same period.

Does this difference indicate that women who were smokers fared better than women who were not smokers?

Example: smoking and cancer

Here is the same data *stratified* by age (at interview).

age	smoker	non-smoker	total
18–24	2/55 (4%)	1/62 (2%)	3/117 (3%)
25–34	3/124 (2%)	5/157 (3%)	8/281 (3%)
35–44	14/109 (13%)	7/121 (6%)	21/230 (9%)
45–54	27/130 (21%)	12/78 (15%)	39/208 (19%)
55–64	51/115 (44%)	40/121 (33%)	91/236 (39%)
65–74	29/36 (81%)	101/129 (78%)	130/165 (79%)
75+	13/13 (100%)	64/64 (100%)	77/77 (100%)
total	139/582 (24%)	230/732 (31%)	369/1314 (28%)

The death rate for smokers is higher for all age groups except 25-34 (where it is very close), but the age distribution is different for smokers and non-smokers.

Age is a confounding variable in this case.

Example: vitamin A during pregnancy

To study the relation between the diet of pregnant women and the development of birth defects in their offspring, a US study interviewed more than 22,000 pregnant women early in their pregnancies.

The women were divided into cohorts according to the amount of vitamin A in their diet, from food or from supplements:

daily vitamin A	birth defects
0–5000 IU	51/11083 (0.0046%)
5001–8000 IU	54/10585 (0.0051%)
8001–10000 IU	9/763 (0.0118%)
10001+ IU	7/317 (0.0221%)

Example: vitamin A during pregnancy

These data indicate that the prevalence of these defects increased with increasing intake of vitamin A.

But there is a confounding factor: vitamin A increases the probability of a foetus with defects surviving to full term.

Principles of experimental design

Experimental design aims to avoid the effects of confounding factors, known and unknown, by:

- **Control and comparison**
- **Blocking**
- **Randomisation**
- **Blind and double blind testing**

Control

Keep everything the same except the variables you know about. That is, compare groups where you vary just known variables.

For example, all test subjects are non-smoking males aged 18–25, and we compare the effect of two types of diet supplement. The consequence is that the test results only apply to this subpopulation.

Often a group with no treatment is used as a basis for comparison. This is called a **control group**.

Blocking

Given a known confounding factor, e.g. gender or age, partition the population into blocks which are homogeneous in the confounding variables.

Then, within each block assign treatment and control units, so the effect of the treatment variable can be judged within each block.

Randomisation

There may be *lurking* factors: confounding factors we are unaware of.

The solution is to randomly assign units to different treatments (possibly within blocks).

Examples: randomisation

A self-selected or voluntary sample is invariably biased.

First Salk polio vaccine trial: In this trial, parents could choose if their children were to participate. Poor families were less likely to participate, but poor children were naturally more likely to get polio.

Literary Digest poll: Readers were asked if they would vote for Roosevelt. Responders were self-selected, and were biased.

Survey	Sample size	Roosevelt's %
Literary Digest	2,400,000	43%
Gallup	50,000	56%
Election		62%

Examples: randomisation

Choosing from a list: Twenty people are selected to trial a new drug. We need to assign ten to the control group and ten to the treatment group.

Their last names are: Chang, Chao, Cheng, Chou, Chu, Gordon, Hsu, Huang, Hu, Li, Liu, MacGregor, MacIntosh, MacKenzie, MacMillan, Munro, Murray, Shannon, Stewart, Urquhart.

Choosing names alphabetically would introduce a cultural bias (for example, diet, which depends on your cultural background, could affect the drug's performance). Randomly assigning people to each group avoids this.

Blind and double blind testing

In a blind experiment, the patients do not know if they are being treated or not.

In a double blind experiment, neither the patients nor those administering treatments know which treatment is being given to whom.

We do this to avoid response bias and the placebo effect.

Example: mild polio can be confused with flu, and in the first Salk polio vaccine trial, doctors were more likely to diagnose children who had not been vaccinated.

Blind and double blind testing

Example: gastric freezing was a treatment proposed for ulcer patients. The idea was to reduce acid secretion by cooling the stomach and so relieve ulcers. This was achieved by pumping a freezing liquid into a balloon in the stomach.

An experiment reported in the Journal of the American Medical Association showed that gastric freezing did reduce acid secretion and relieve ulcer pain.

The treatment was safe and easy and was widely used for several years.

Blind and double blind testing

Unfortunately, the reported effect was just a placebo effect.

A better-designed experiment, done several years later, divided ulcer patients into two groups. One group was treated by gastric freezing as before (the treatment group), while the other group received a placebo treatment in which the solution in the balloon was at body temperature rather than freezing (a control group).

The results: 34% of the 82 patients in the treatment group improved versus 38% of the 78 patients in the control group.

Replication

Greater precision is achieved by **replication**.

We replicate treatment combinations to minimise the variance of our estimators. Doing this optimally leads to **balanced designs**.

Types of design

Suppose we have a factor whose effect is of interest — the **treatment** — and zero or more confounding factors, which are dealt with by blocking. Any lurking factors are dealt with by randomisation.

- Completely randomised design (CRD)
  no confounding factors (one-way classification)
- (Randomised) complete block design (CBD)
  one confounding/blocking factor (two-way classification)
- Latin squares
  two confounding/blocking factors (three-way classification)
  more efficient than combining the confounding factors
- Balanced incomplete block design (BIBD)
  one confounding factor but CBD impossible

Completely randomised design (CRD)

We use a completely randomised design for a single factor (treatment) with k levels.

For example:

- no drug; small amount of drug A; large amount of drug A;
- no drug; drug A; drug B;
- fertiliser A; fertiliser B.

Given n_i test units for factor level i, we assign them randomly.

To analyse, we use a one-way classification model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} = \mu_i + \varepsilon_{ij}.$$

Choosing test units at random

```
> n <- c(5,6,4)
> nsum <- sum(n)
> x <- sample(nsum, nsum)
> (j1 <- x[1:n[1]])

[1]  2 12  6  9  3

> (j2 <- x[n[1]+(1:n[2])])

[1] 11  1  7 10  5  4

> (j3 <- x[(n[1]+n[2]+1):nsum])

[1] 13 15 14  8
```

Choosing test units at random

$$\begin{bmatrix} y_2 \\ y_{12} \\ y_6 \\ y_9 \\ y_3 \\ y_{11} \\ y_1 \\ y_7 \\ y_{10} \\ y_5 \\ y_4 \\ y_{13} \\ y_{15} \\ y_{14} \\ y_8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_2 \\ \varepsilon_{12} \\ \varepsilon_6 \\ \varepsilon_9 \\ \varepsilon_3 \\ \varepsilon_{11} \\ \varepsilon_1 \\ \varepsilon_7 \\ \varepsilon_{10} \\ \varepsilon_5 \\ \varepsilon_4 \\ \varepsilon_{13} \\ \varepsilon_{15} \\ \varepsilon_{14} \\ \varepsilon_8 \end{bmatrix}$$

$$\mathbf{y} = X \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Completely randomised design (CRD)

In a one-factor model, we have already seen that $\hat{\mu}_i = \bar{y}_i$, the average of the responses in group i.

What is its variance? We have (for example)

$$\hat{\mu}_1 = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix} \mathbf{b}$$

$$\text{var } \hat{\mu}_1 = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix} (X^T X)^c \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sigma^2$$

Completely randomised design (CRD)

$$\begin{aligned}
\text{var } \hat{\mu}_1 &= \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_k} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sigma^2 \\
&= \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{n_1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sigma^2 \\
&= \frac{\sigma^2}{n_1},
\end{aligned}$$

and it is easy to extend this to var $\hat{\mu}_i = \frac{\sigma^2}{n_i}$.

Completely randomised design (CRD)

Theorem 8.1

In a completely randomised design with n test units, the allocation of test units to factor levels which minimises

$$\sum_{i=1}^{k} \text{var } \hat{\mu}_i = \sigma^2 \sum_{i=1}^{k} \frac{1}{n_i}$$

is

$$n_i = \frac{n}{k}$$

(assuming n is a multiple of k).

Completely randomised design (CRD)

Proof: We use Lagrangian multipliers to deal with the constraint in sample size. Take

$$f(n_1, \ldots, n_k, \lambda) = \sigma^2 \sum_{i=1}^{k} \frac{1}{n_i} + \lambda \left(\sum_{i=1}^{k} n_i - n \right).$$

We minimise this function with respect to all variables; the equation $\frac{\partial f}{\partial \lambda} = 0$ ensures that the total sample size is constrained to n.

Completely randomised design (CRD)

The remaining equations give

$$\begin{aligned} \frac{\partial f}{\partial n_i} &= -\frac{\sigma^2}{n_i^2} + \lambda = 0 \\ n_i^2 &= \frac{\sigma^2}{\lambda}. \end{aligned}$$

This does not depend on i, so in order to satisfy these equations we must choose all n_i equal.

The same method can be used to calculate *optimal allocations* when we wish to minimise the variance of other statistics.

Completely randomised design (CRD)

Example. Suppose we have 4 treatments and want to study the treatment contrasts $\tau_2 - \tau_1$, $\tau_3 - \tau_1$ and $\tau_4 - \tau_1$. We have

$$\text{var } \widehat{\tau_i - \tau_1} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_i} \right),$$

so we minimise

$$f(n_1, n_2, n_3, n_4, \lambda) = \sigma^2 \left(\frac{3}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4} \right) + \lambda \left(\sum_{i=1}^{4} n_i - n \right).$$

Completely randomised design (CRD)

We get (for $i \neq 1$):

$$\begin{aligned}
\frac{\partial f}{\partial n_1} &= -3\frac{\sigma^2}{n_1^2} + \lambda = 0 \\
n_1^2 &= 3\frac{\sigma^2}{\lambda} \\
\\
\frac{\partial f}{\partial n_i} &= -\frac{\sigma^2}{n_i^2} + \lambda = 0 \\
n_i^2 &= \frac{\sigma^2}{\lambda} \\
n_1^2 &= 3n_i^2 \\
n_1 &= \sqrt{3}n_i.
\end{aligned}$$

Completely randomised design (CRD)

Therefore

$$\begin{aligned} n_1 + 3\frac{n_1}{\sqrt{3}} &= n \\ n_1 &= \frac{n}{1+\sqrt{3}} \\ n_i &= \frac{n}{\sqrt{3}(1+\sqrt{3})}. \end{aligned}$$

For example if we have $n = 20$, then rounding gives $n_1 = 8$ and $n_2 = n_3 = n_4 = 4$.

This is a reflection of the fact that all the contrasts involve τ_1, so it is more important to estimate that accurately.

(Randomised) Complete block design (CBD)

We can use a complete block design when we have one factor of interest (the treatment) and one nuisance/confounding factor (the blocking factor).

For example:

- fertilizer (treatment); position on slope (blocking)
- drug (treatment); age and/or gender (blocking).

We partition the experimental units into blocks of size k (where k is the number of treatments), which are homogeneous in the blocking factor. In each block, we apply one treatment to each experimental unit, chosen randomly.

(Randomised) Complete block design (CBD)

Blocks which have a size which is an integer multiple of k also work; an equal number of units in each block receive each treatment.

To analyse, we use a two-way classification model:

$$y_{ijk} = \mu + \beta_i + \tau_j + \varepsilon_{ijk},$$

where the β_i are the block effects and the τ_j are the treatment effects.

Assigning test units at random

Suppose we have $k = 3$ treatment levels and $b = 4$ blocks, and one sample from each treatment/block combination. Then we have $n = 3 \times 4 = 12$.

Without loss of generality, suppose units 1, 2, 3 are in block 1, units 4, 5, 6 are in block 2, etc.

```
> sample(1:3, 3)

[1] 2 1 3

> sample(4:6, 3)

[1] 4 5 6

> sample(7:9, 3)

[1] 7 9 8

> sample(10:12, 3)

[1] 10 11 12
```

Assigning test units at random

$$
\begin{bmatrix} y_2 \\ y_1 \\ y_3 \\ \hline y_4 \\ y_5 \\ y_6 \\ \hline y_7 \\ y_9 \\ y_8 \\ \hline y_{10} \\ y_{11} \\ y_{12} \end{bmatrix}
=
\left[\begin{array}{c|ccc|ccc|cc}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ \hline
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}\right]
\begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}
+
\begin{bmatrix} \varepsilon_2 \\ \varepsilon_1 \\ \varepsilon_3 \\ \hline \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \hline \varepsilon_7 \\ \varepsilon_9 \\ \varepsilon_8 \\ \hline \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \end{bmatrix}
$$

$$
\mathbf{y} = X \boldsymbol{\beta} + \boldsymbol{\varepsilon}
$$

(Randomised) Complete block design (CBD)

We are not interested in μ and the β_i, only in the treatment effects τ_j. Therefore we consider the treatment effects and nuisance parameters separately. We write

$$\mathbf{y} = \left[\; X_1 \; \middle| \; X_2 \; \right] \left[\begin{array}{c} \boldsymbol{\mu} \\ \boldsymbol{\beta} \\ \hline \boldsymbol{\tau} \end{array} \right] + \boldsymbol{\varepsilon}.$$

(Randomised) Complete block design (CBD)

For the example, this gives

$$X_1 = \left[\begin{array}{c|cccc} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{array}\right], \quad X_2 = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

Reduced normal equations

Theorem 8.2

In the general linear model, write

$$\mathbf{y} = \left[\begin{array}{c|c} X_1 & X_2 \end{array} \right] \left[\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}_2} \right] + \boldsymbol{\varepsilon}.$$

Then $\left[\begin{array}{c} \mathbf{b}_1 \\ \mathbf{b}_2 \end{array} \right]$ *is a solution to the normal equations only if* $\mathbf{b}_2$ *is a solution to the* **reduced normal equations**

$$X_2^T [I - H_1] X_2 \mathbf{b}_2 = X_2^T [I - H_1] \mathbf{y},$$

where

$$H_1 = X_1 (X_1^T X_1)^c X_1^T.$$

Reduced normal equations

Proof. The normal equations are

$$X^T X\mathbf{b} = X^T\mathbf{y}$$

$$\begin{aligned} X_1^T X_1\mathbf{b}_1 + X_1^T X_2\mathbf{b}_2 &= X_1^T\mathbf{y} \\ X_2^T X_1\mathbf{b}_1 + X_2^T X_2\mathbf{b}_2 &= X_2^T\mathbf{y} \end{aligned}$$

$$\begin{aligned} &X_2^T X_1(X_1^T X_1)^c X_1^T X_1\mathbf{b}_1 \\ &\quad + X_2^T X_1(X_1^T X_1)^c X_1^T X_2\mathbf{b}_2 = X_2^T X_1(X_1^T X_1)^c X_1^T\mathbf{y} \end{aligned}$$

$$X_2^T X_1\mathbf{b}_1 + X_2^T H_1 X_2\mathbf{b}_2 = X_2^T H_1\mathbf{y}$$

$$X_2^T[I - H_1]X_2\mathbf{b}_2 = X_2^T[I - H_1]\mathbf{y}.$$

Reduced normal equations

Because $I - H_1$ is symmetric and idempotent, the reduced normal equations are the same as the normal equations for the model

$$\mathbf{y} = [I - H_1]X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

This is itself a linear model!

We write $X_{2|1} = [I - H_1]X_2$, the design matrix for this model.

Reduced normal equations for CBD

Let's look closer at the reduced normal equations for CBD.
Remember we have k treatments and b blocks, each of size k.

Let J_k be the $k \times k$ matrix of 1's, and $\mathbf{1}_k$ the $k \times 1$ vector of 1's.
Then we have

$$X_1 = \begin{bmatrix} \mathbf{1}_k & \mathbf{1}_k & & 0 \\ \vdots & & \ddots & \\ \mathbf{1}_k & 0 & & \mathbf{1}_k \end{bmatrix}$$

$$X_1^T X_1 = \begin{bmatrix} n & k & \dots & k \\ k & k & & 0 \\ \vdots & & \ddots & \\ k & 0 & & k \end{bmatrix}$$

Reduced normal equations for CBD

$$
\begin{aligned}
H_1 &= \begin{bmatrix} \mathbf{1}_k & \mathbf{1}_k & & 0 \\ \vdots & & \ddots & \\ \mathbf{1}_k & 0 & & \mathbf{1}_k \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{k} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \frac{1}{k} \end{bmatrix} \begin{bmatrix} \mathbf{1}_k^T & \cdots & \mathbf{1}_k^T \\ \mathbf{1}_k^T & & 0 \\ & \ddots & \\ 0 & & \mathbf{1}_k^T \end{bmatrix} \\
&= \frac{1}{k} \begin{bmatrix} \mathbf{1}_k & \mathbf{1}_k & & 0 \\ \vdots & & \ddots & \\ \mathbf{1}_k & 0 & & \mathbf{1}_k \end{bmatrix} \begin{bmatrix} 0 & \cdots & 0 \\ \mathbf{1}_k^T & & 0 \\ & \ddots & \\ 0 & & \mathbf{1}_k^T \end{bmatrix} \\
&= \frac{1}{k} \begin{bmatrix} J_k & 0 & \cdots & 0 \\ 0 & J_k & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & J_k \end{bmatrix}.
\end{aligned}
$$

Reduced normal equations for CBD

Also

$$
\begin{aligned}
X_2 &= \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \\
X_{2|1} &= [I - H_1]X_2 \\
&= \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} - \frac{1}{k} \begin{bmatrix} J_k \\ \vdots \\ J_k \end{bmatrix}.
\end{aligned}
$$

Reduced normal equations for CBD

This gives

$$
\begin{aligned}
X_{2|1}^T X_{2|1} &= \begin{bmatrix} I_k & \dots & I_k \end{bmatrix} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} - \frac{1}{k} \begin{bmatrix} I_k & \dots & I_k \end{bmatrix} \begin{bmatrix} J_k \\ \vdots \\ J_k \end{bmatrix} \\
&\quad - \frac{1}{k} \begin{bmatrix} J_k & \dots & J_k \end{bmatrix} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} + \frac{1}{k^2} \begin{bmatrix} J_k & \dots & J_k \end{bmatrix} \begin{bmatrix} J_k \\ \vdots \\ J_k \end{bmatrix} \\
&= bI_k - \frac{b}{k} J_k - \frac{b}{k} J_k + \frac{b}{k^2} k J_k \\
&= b \left[I_k - \frac{1}{k} J_k \right].
\end{aligned}
$$

Reduced normal equations for CBD

Now

$$\begin{aligned}\left(X_{2|1}^T X_{2|1}\right)\left(\frac{1}{b} I_k\right)\left(X_{2|1}^T X_{2|1}\right) &= b\left[I_k - \frac{1}{k} J_k\right]\left[I_k - \frac{1}{k} J_k\right] \\ &= bI_k - \frac{2b}{k} J_k + \frac{b}{k^2} k J_k \\ &= bI_k - \frac{b}{k} J_k \\ &= X_{2|1}^T X_{2|1},\end{aligned}$$

so

$$(X_{2|1}^T X_{2|1})^c = \frac{1}{b} I_k.$$

Reduced normal equations for CBD

Thus a solution to the reduced normal equations is

$$\begin{aligned}
\mathbf{b}_2 &= (X_{2|1}^T X_{2|1})^c X_{2|1}^T \mathbf{y} \\
&= \frac{1}{b} I_k \left[\begin{bmatrix} I_k & \dots & I_k \end{bmatrix} - \frac{1}{k} \begin{bmatrix} J_k & \dots & J_k \end{bmatrix} \right] \begin{bmatrix} y_{11} \\ \vdots \\ y_{1k} \\ \vdots \\ y_{b1} \\ \vdots \\ y_{bk} \end{bmatrix} \\
&= \begin{bmatrix} \bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot} \\ \vdots \\ \bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot} \end{bmatrix}.
\end{aligned}$$

Comparison of CBD and CRD

Suppose that we have a treatment with k levels, a confounding factor with b levels, and $n = kb$ experimental units.

We could use a CBD as above, or we could just ignore the confounding factor and use a CRD. That is, assign b units to each treatment level at random.

We will show that this actually does not change the estimating equations for the treatment effects, and they are still unbiased, but it does increase their variance.

Comparison of CBD and CRD

In a CBD, if $\mathbf{t}^T\boldsymbol{\tau}$ (involving only τ parameters) is estimable, then

$$\begin{aligned}
\mathbf{t}^T(X_{2|1}^T X_{2|1})^c X_{2|1}^T X_{2|1} &= \mathbf{t}^T \\
\Rightarrow \mathbf{t}^T \frac{1}{b} I_k \cdot b \left[I_k - \frac{1}{k} J_k \right] &= \mathbf{t}^T \\
\Rightarrow \mathbf{t}^T \left[I_k - \frac{1}{k} J_k \right] \mathbf{1}_k &= \mathbf{t}^T \mathbf{1}_k \\
\Rightarrow 0 &= \mathbf{t}^T \mathbf{1}_k.
\end{aligned}$$

That is, $\mathbf{t}$ *must* be a contrast.

Comparison of CBD and CRD

Thus, if $\mathbf{t}^T\boldsymbol{\tau}$ is estimable, its estimate is

$$\mathbf{t}^T \begin{bmatrix} \bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot} \\ \vdots \\ \bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot} \end{bmatrix} = \mathbf{t}^T \begin{bmatrix} \bar{y}_{\cdot 1} \\ \vdots \\ \bar{y}_{\cdot k} \end{bmatrix}.$$

That is, the estimate is the sum over all treatments i, of t_i times the mean of those responses with treatment i.

This is exactly what we would get if we ignored the blocks and treated the experiment as a CRD.

Comparison of CBD and CRD

What is the variance of the estimator?

$$\begin{aligned}
\text{var } \mathbf{t}^T\mathbf{b}_2 &= \text{var } \mathbf{t}^T(X_{2|1}^T X_{2|1})^c X_{2|1}^T \mathbf{y} \\
&= \mathbf{t}^T(X_{2|1}^T X_{2|1})^c X_{2|1}^T \sigma^2 I X_{2|1}(X_{2|1}^T X_{2|1})^c \mathbf{t} \\
&= \mathbf{t}^T(X_{2|1}^T X_{2|1})^c \mathbf{t}\sigma^2 \\
&= \mathbf{t}^T \frac{1}{b} I_k \mathbf{t}\sigma^2 \\
&= \frac{1}{b}\sum_{j=1}^{k} t_j^2 \sigma^2.
\end{aligned}$$

Comparison of CBD and CRD: CBD

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}, \qquad i = 1, \ldots, b, \quad j = 1, \ldots, k$$

$$\begin{aligned} \widehat{\mathbf{t}^T \boldsymbol{\tau}} &= \sum_{j=1}^{k} t_j \bar{y}_{\cdot j} \\ \mathbb{E}\widehat{\mathbf{t}^T \boldsymbol{\tau}} &= \mathbf{t}^T \boldsymbol{\tau} \\ \text{var } \widehat{\mathbf{t}^T \boldsymbol{\tau}} &= \frac{1}{b} \sum_{j=1}^{k} t_j^2 \sigma^2 \end{aligned}$$

where σ^2 is the variance of ε_{ij}.

Comparison of CBD and CRD: CRD

$$y_{ij} = \mu + \tau_i + \varepsilon'_{ij}$$

where ε'_{ij} corresponds to a randomly chosen $\beta_u + \varepsilon_{uv}$.
(Note the treatments now correspond to the first index. There is a bit of hand-waving here, as the ε'_{ij} are not i.i.d.)

$$\begin{aligned}
\widehat{\mathbf{t}^T\boldsymbol{\tau}} &= \sum_{i=1}^{k} t_i \bar{y}_i \\
\mathbb{E}\widehat{\mathbf{t}^T\boldsymbol{\tau}} &= \mathbf{t}^T\boldsymbol{\tau} \\
\text{var } \widehat{\mathbf{t}^T\boldsymbol{\tau}} &= \frac{1}{b}\sum_{j=1}^{k} t_j^2 \sigma^{2\prime}
\end{aligned}$$

where $\sigma^{2'}$ is the variance of ε'_{ij}.

Comparison of CBD and CRD: CRD

That is,

$$\begin{aligned} \sigma^{2\prime} &= \sigma^2 + \frac{1}{b}\sum_{j=1}^{b}(\beta_j - \bar{\beta}.)^2 \\ &\geq \sigma^2. \end{aligned}$$

Thus, by using a CRD, we might have increased the variance of our estimators.

Comparison of CBD and CRD

Note that in the CRD s^2 has $b \times k - k$ d.f.,
while in the CBD s^2 has $b \times k - (b + k - 1)$ d.f.

So if there is really no blocking effect, then the CRD is marginally better than the CBD, but if there is a effect due to blocks then ignoring it means your estimators are less accurate than they could be.

Design principle: segment the study population into homogeneous groups to reduce variation.

Example

Suppose we want to compare two drugs and a control (three treatment levels) using mice.

We could take 30 mice and assign 10 randomly to each treatment (a CRD).

Alternatively, we could take 10 groups (blocks), each of three mice from the same family (siblings), then assign all three treatments to the mice in each group. Within each group the treatments are assigned randomly.

This is a CBD. The genetic similarity of the blocks will reduce the error variance.

Fluid example

Recall our data on the time to dissolution of two types of capsule in different body fluids.

```
> dissolve <- data.frame(time = c(39.5, 31.2, 47.4, 44.0),
+                        fluid = c("G", "D", "G", "D"),
+                        capsule = c("A", "A", "B", "B"))
> model1 <- lm(time ~ fluid, dissolve)
> model2 <- lm(time ~ fluid + capsule, dissolve)
```

Fluid example

```
> summary(model1)

Call:
lm(formula = time ~ fluid, data = dissolve)

Residuals:
    1     2     3     4
-3.95 -6.40  3.95  6.40

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   40.525      3.760  10.777   0.0085 **
fluid1        -2.925      3.760  -0.778   0.5181
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.521 on 2 degrees of freedom
Multiple R-squared:  0.2323, Adjusted R-squared:  -0.1516
F-statistic: 0.605 on 1 and 2 DF,  p-value: 0.5181
```

Fluid example

```
> summary(model2)

Call:
lm(formula = time ~ fluid + capsule, data = dissolve)

Residuals:
     1      2      3      4
 1.225 -1.225 -1.225  1.225

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   40.525      1.225  33.082   0.0192 *
fluid1        -2.925      1.225  -2.388   0.2525
capsule1      -5.175      1.225  -4.224   0.1480
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.45 on 1 degrees of freedom
Multiple R-squared:  0.9593, Adjusted R-squared:  0.8778
F-statistic: 11.77 on 2 and 1 DF,  p-value: 0.2018
```

Fluid example

For the CRD, we have $\hat{\tau}_2 - \hat{\tau}_1 = 5.850$ and $s = 7.521$.

For the CBD, $\hat{\tau}_2 - \hat{\tau}_1$ is the same, but s has decreased to 2.45.

As a result, the fluid factor is more significant: the p-value has come down from 0.5181 to 0.2525.

Latin square designs

Latin square designs are used when there are two confounding factors.

Examples:

- treatment is type of fertiliser, confounding/blocking factors are the type of tillage and variety of seed;
- treatment is variety of seed, confounding/blocking factors are the soil water content and soil salt content.

Latin square designs

In theory, we can deal with this using blocking; for example, a CBD that treats the two confounding factors A and B as a single confounding factor C, where the levels of C are given by all possible combinations of levels from A and B.

This can become onerous when there are many levels of the confounding factors, or when there are more than two. In this case we can use Latin squares, which use fewer experimental units than a CBD.

However they require the treatment and both blocking factors to have the same number of levels, t say.

Latin square designs

We would like a design that treats each confounding factor like a CBD.

Imagine the experimental units laid out in a grid, where blocking factor one has levels A, B, C, and blocking factor two has levels U, V, W:

block A	one of each treatment
block B	one of each treatment
block C	one of each treatment

block U	block V	block W
one of each treatment	one of each treatment	one of each treatment

Latin square designs

The solution is a Latin square: each treatment appears exactly once in each row and each column.

```
> library(magic)
> rlatin(5)
```

```
     [,1] [,2] [,3] [,4] [,5]
[1,]    3    5    2    4    1
[2,]    5    3    1    2    4
[3,]    1    2    4    5    3
[4,]    2    4    3    1    5
[5,]    4    1    5    3    2
```

Latin square designs

```
> rlatin(5)
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	3	5	2	4
[2,]	3	5	2	4	1
[3,]	2	4	1	3	5
[4,]	4	1	3	5	2
[5,]	5	2	4	1	3

Latin square designs

Choosing a Latin square at random helps avoid the effect of lurking variables.

Given a Latin square, a random permutation of the rows and columns gives another Latin square.

This procedure will not generate all Latin squares of a given size, but is generally sufficient.

Latin square designs

An initial $k \times k$ Latin square can always be found by putting entry (i, j) equal to $(i + j - 2 \text{ mod } k) + 1$.

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

Latin square designs

The model for a Latin square design is an additive three-way classification model:

$$y_{ijk} = \mu + \beta_i + \gamma_j + \tau_k + \varepsilon_{ijk},$$

where $1 \leq i, j \leq t$ and $k = k(i, j)$ is determined by i and j.

Here μ is the overall mean, β_i is the effect of level i of the first confounding factor, γ_j is the effect of level j of the second confounding factor, and τ_k is the k-th treatment effect.

Note: the experimental unit is determined by i and j; there is only one k for each (i, j) pair.

Latin square designs

A Latin square design of size t can be written as

$$\mathbf{y} = X_1 \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon},$$

where

$$X_1 = \begin{bmatrix} \mathbf{1}_t & \mathbf{1}_t & 0 & \cdots & 0 & I_t \\ \mathbf{1}_t & 0 & \mathbf{1}_t & \cdots & 0 & I_t \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \mathbf{1}_t & 0 & 0 & \cdots & \mathbf{1}_t & I_t \end{bmatrix}, \quad X_2 = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_t \end{bmatrix},$$

and each P_i is a permutation matrix (zero except for a single 1 in each row and column), such that

$$\sum_{i=1}^{t} P_t = J_t.$$

Latin square designs

What is the reduced design matrix for this model? We have

$$
\begin{aligned}
H_1 X_2 &= X_1 (X_1^T X_1)^c X_1^T X_2 \\
X_1^T X_2 &= \begin{bmatrix} \mathbf{1}_t^T & \dots & \mathbf{1}_t^T \\ \mathbf{1}_t^T & & 0 \\ & \ddots & \\ 0 & & \mathbf{1}_t^T \\ I_t & \dots & I_t \end{bmatrix} \begin{bmatrix} P_1 \\ \vdots \\ P_t \end{bmatrix} \\
&= \begin{bmatrix} t\mathbf{1}_t^T \\ J_t \\ J_t \end{bmatrix}.
\end{aligned}
$$

Latin square designs

However we also have

$$
\begin{aligned}
X_1^T \frac{1}{t} J_{t^2 \times t} &= \frac{1}{t} \begin{bmatrix} \mathbf{1}_t^T & \dots & \mathbf{1}_t^T \\ \mathbf{1}_t^T & & 0 \\ & \ddots & \\ 0 & & \mathbf{1}_t^T \\ I_t & \dots & I_t \end{bmatrix} \begin{bmatrix} J_t \\ \vdots \\ J_t \end{bmatrix} \\
&= \begin{bmatrix} t\mathbf{1}_t^T \\ J_t \\ J_t \end{bmatrix}.
\end{aligned}
$$

Latin square designs

Therefore

$$\begin{aligned} H_1 X_2 &= X_1 (X_1^T X_1)^c X_1^T X_2 \\ &= X_1 (X_1^T X_1)^c X_1^T \frac{1}{t} J_{t^2 \times t} \\ &= H_1 \frac{1}{t} J_{t^2 \times t} \\ &= \frac{1}{t} J_{t^2 \times t} \end{aligned}$$

which follows because the columns of $J_{t^2 \times t}$ are contained in the columnspace of X_1.

Latin square designs

Thus the reduced design matrix for this model is

$$X_{2|1} = \begin{bmatrix} P_1 \\ \vdots \\ P_t \end{bmatrix} - \frac{1}{t} \begin{bmatrix} J_t \\ \vdots \\ J_t \end{bmatrix}.$$

It follows that

$$\begin{aligned} X_{2|1}^T X_{2|1} &= t \left[I_t - \frac{1}{t} J_t \right] \\ (X_{2|1}^T X_{2|1})^c &= \frac{1}{t} I_t \end{aligned}$$

Latin square designs

Thus a solution to the reduced normal equations is

$$\begin{aligned}
\mathbf{b}_2 &= (X_{2|1}^T X_{2|1})^c X_{2|1}^T \mathbf{y} \\
&= \frac{1}{t}\left[\begin{bmatrix} P_1^T & \dots & P_t^T \end{bmatrix} - \frac{1}{t}\begin{bmatrix} J_t & \dots & J_t \end{bmatrix}\right]\mathbf{y} \\
&= \begin{bmatrix} \bar{y}_{..1} - \bar{y}_{...} \\ \vdots \\ \bar{y}_{..t} - \bar{y}_{...} \end{bmatrix}
\end{aligned}$$

This is the same as the CRD!

Latin square example

We wish to analyse the productivity of 5 kinds of fertilizer, 5 kinds of tillage, and 5 kinds of seed. The data are organized in a Latin square design:

	treatA	treatB	treatC	treatD	treatE
fertilizer1	"A42"	"C47"	"B55"	"D51"	"E44"
fertilizer2	"E45"	"B54"	"C52"	"A44"	"D50"
fertilizer3	"C41"	"A46"	"D57"	"E47"	"B48"
fertilizer4	"B56"	"D52"	"E49"	"C50"	"A43"
fertilizer5	"D47"	"E49"	"A45"	"B54"	"C46"

The three factors are: fertilizer (fertilizer1:5), tillage (treatA:E), and seed (A:E). The numbers are the productivity in cwt/year.

Latin square example

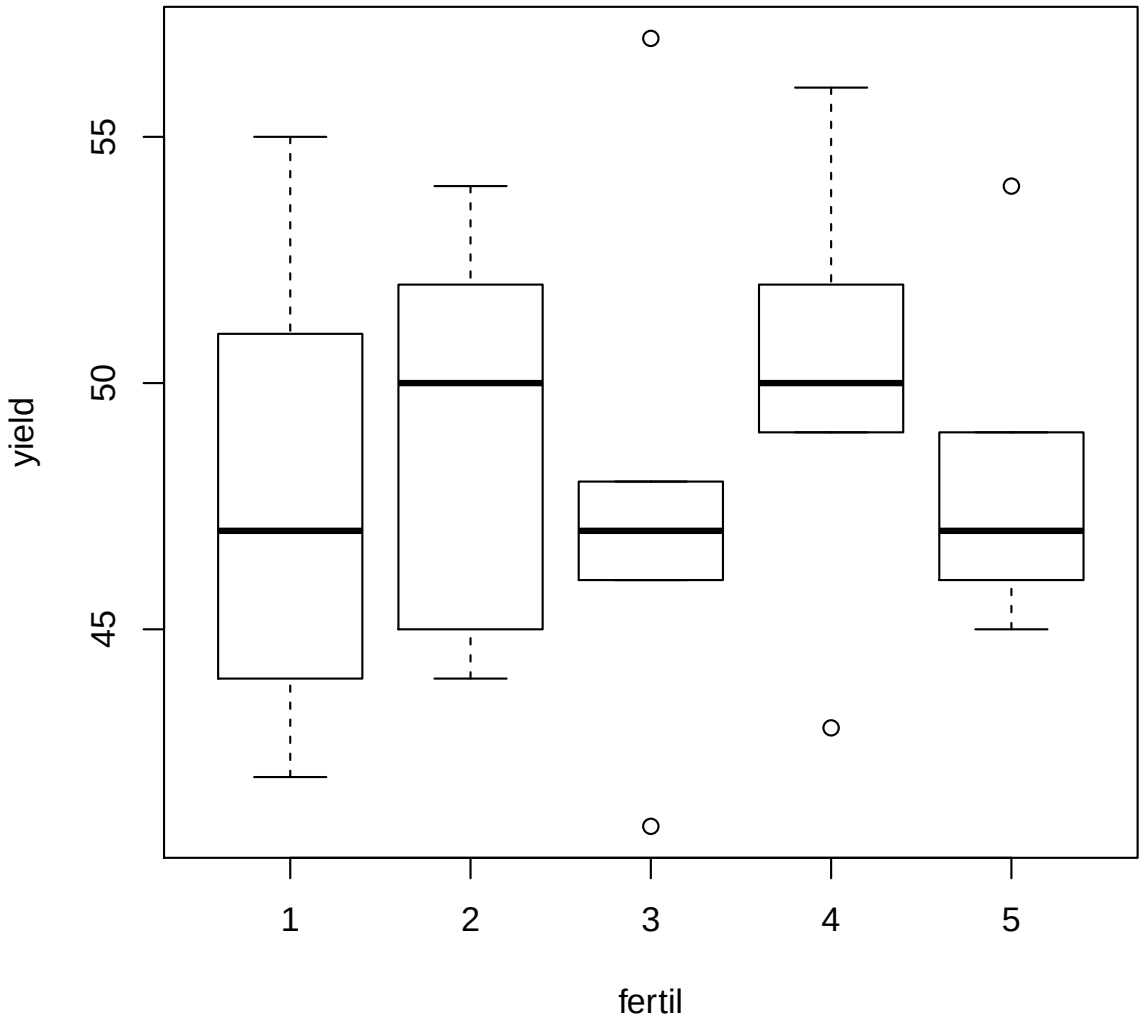
```
> fertil <- c("1", "2", "3", "4", "5")
> tillage <- c(rep("A",5), rep("B",5),
+              rep("C",5), rep("D",5), rep("E",5))
> seed <- c("A","E","C","B","D", "C","B","A","D","E",
+           "B","C","D","E","A", "D","A","E","C","B",
+           "E","D","B","A","C")
> yield <- c(42,45,41,56,47, 47,54,46,52,49, 55,52,57,49,45,
+            51,44,47,50,54, 44,50,48,43,46)
> latindata <- data.frame(tillage, fertil, seed, yield)
```

Latin square example

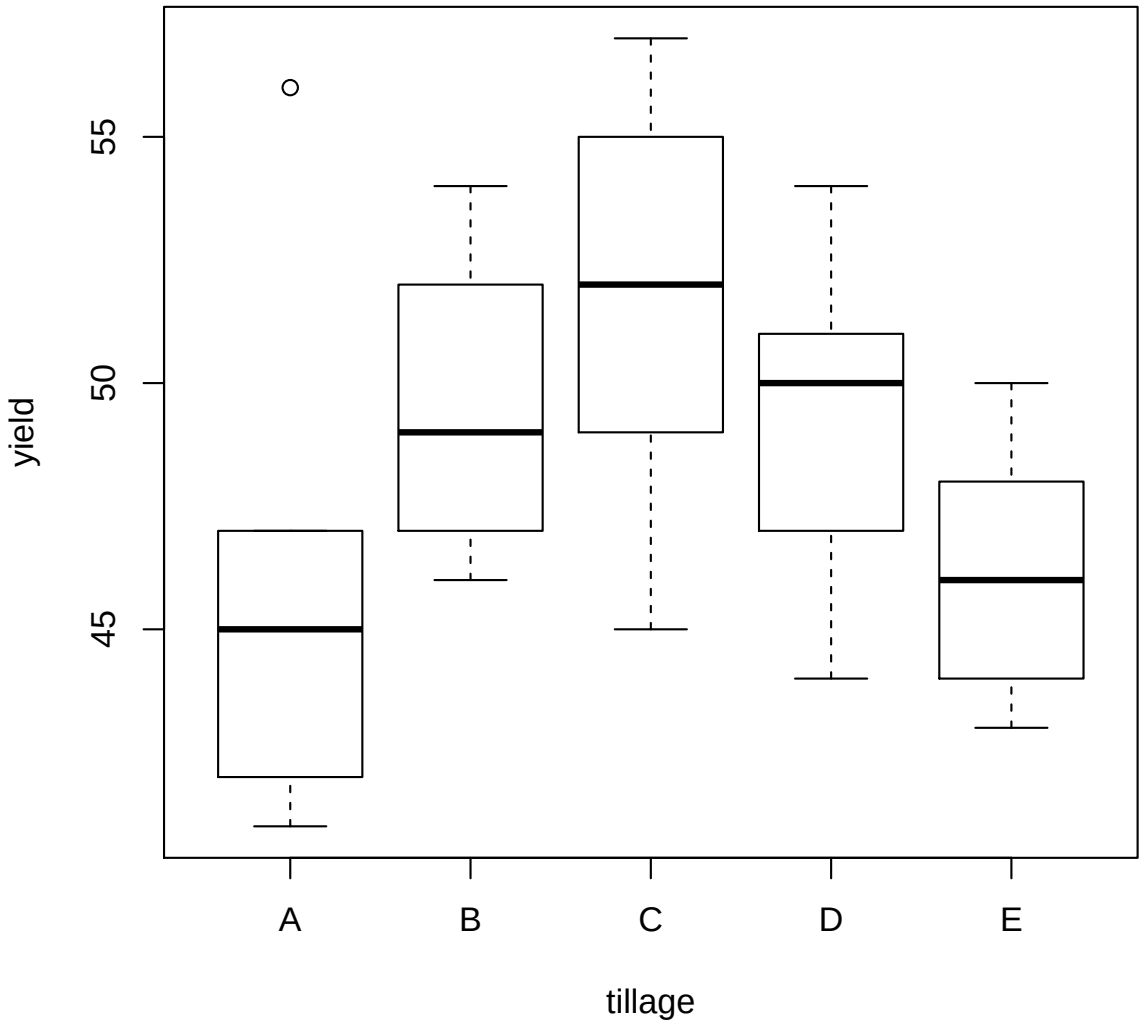
```
> matrix(latindata$seed, 5,5)
     [,1] [,2] [,3] [,4] [,5]
[1,] "A"  "C"  "B"  "D"  "E"
[2,] "E"  "B"  "C"  "A"  "D"
[3,] "C"  "A"  "D"  "E"  "B"
[4,] "B"  "D"  "E"  "C"  "A"
[5,] "D"  "E"  "A"  "B"  "C"
```

```
> matrix(latindata$yield, 5,5)
     [,1] [,2] [,3] [,4] [,5]
[1,]   42   47   55   51   44
[2,]   45   54   52   44   50
[3,]   41   46   57   47   48
[4,]   56   52   49   50   43
[5,]   47   49   45   54   46
```

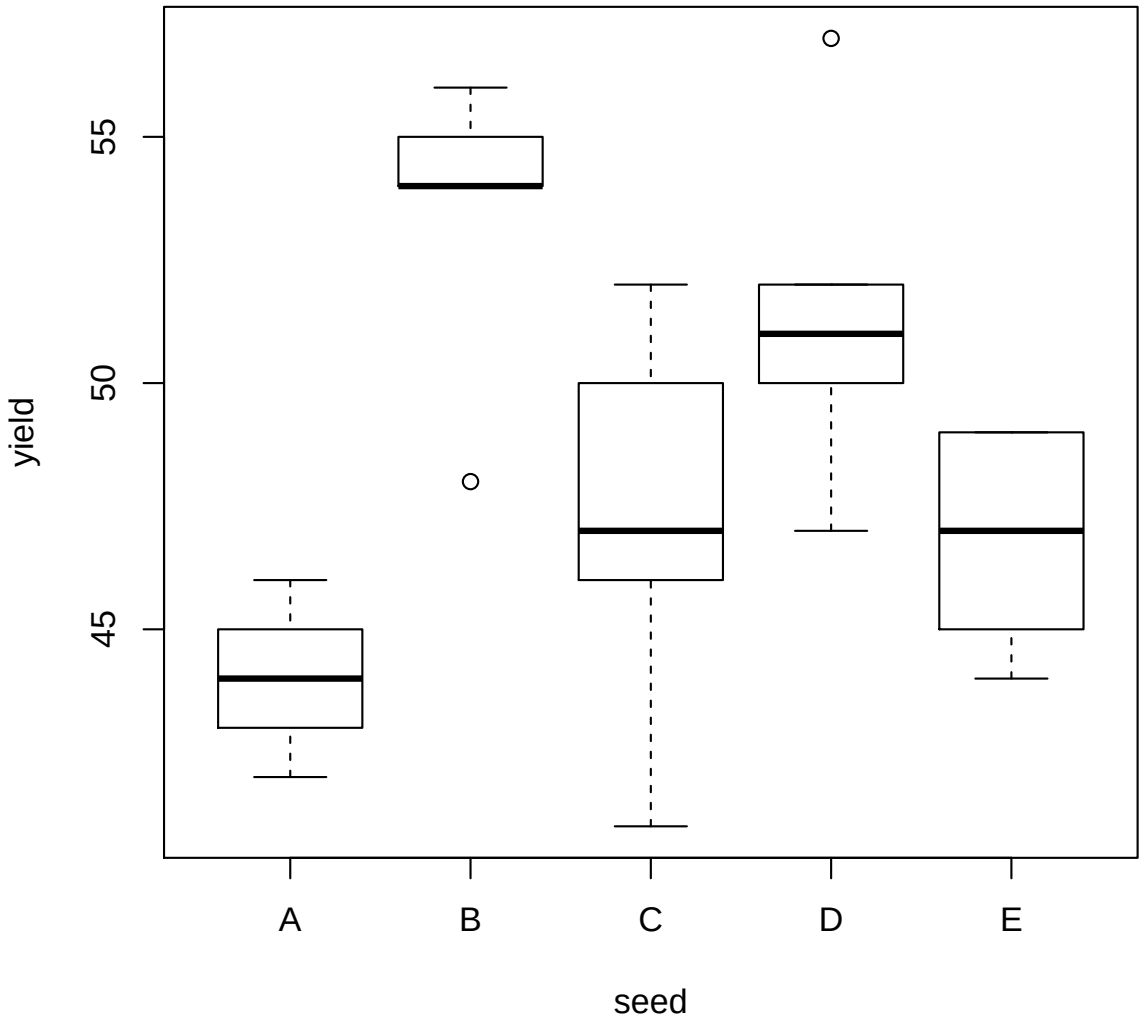
Latin square example



Latin square example



Latin square example



Latin square example

We see that the differences due the fertilizer are small; those due the tillage are medium, and those due to the seed are very high.

We confirm these graphical observations with an ANOVA:

```
> model <- lm(yield ~ fertil+tillage+seed, latindata)
```

Latin square example

```
> summary(model)

Call:
lm(formula = yield ~ fertil + tillage + seed, data = latindata)

Residuals:
   Min     1Q Median     3Q    Max
 -3.08  -1.08   0.12   1.12   3.52

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.5600     0.4722 102.847  < 2e-16 ***
fertil1      -0.7600     0.9443  -0.805 0.436582
fertil2       0.4400     0.9443   0.466 0.649598
fertil3      -0.7600     0.9443  -0.805 0.436582
fertil4       1.4400     0.9443   1.525 0.153194
tillage1     -2.3600     0.9443  -2.499 0.027958 *
tillage2      1.0400     0.9443   1.101 0.292355
tillage3      3.0400     0.9443   3.219 0.007365 **
tillage4      0.6400     0.9443   0.678 0.510794
seed1        -4.5600     0.9443  -4.829 0.000413 ***
seed2         4.8400     0.9443   5.125 0.000251 ***
seed3        -1.3600     0.9443  -1.440 0.175389
seed4         2.8400     0.9443   3.007 0.010914 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.361 on 12 degrees of freedom
Multiple R-squared:  0.8607, Adjusted R-squared:  0.7214
F-statistic: 6.179 on 12 and 12 DF,  p-value: 0.001794
```

Latin square example

```
> anova(model)
Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value   Pr(>F)
fertil     4  17.76   4.440  0.7967 0.549839
tillage    4 109.36  27.340  4.9055 0.014105 *
seed       4 286.16  71.540 12.8361 0.000271 ***
Residuals 12  66.88   5.573
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Latin square example

```
> model2 <- lm(yield ~ tillage+seed, latindata)
> summary(model2)

Call:
lm(formula = yield ~ tillage + seed, data = latindata)

Residuals:
   Min     1Q Median     3Q    Max
 -3.84  -1.04  -0.44   1.16   4.96

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    48.56       0.46 105.565  < 2e-16 ***
tillage1       -2.36       0.92  -2.565 0.020754 *
tillage2        1.04       0.92   1.130 0.274955
tillage3        3.04       0.92   3.304 0.004478 **
tillage4        0.64       0.92   0.696 0.496627
seed1          -4.56       0.92  -4.957 0.000143 ***
seed2           4.84       0.92   5.261 7.76e-05 ***
seed3          -1.36       0.92  -1.478 0.158754
seed4           2.84       0.92   3.087 0.007070 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.3 on 16 degrees of freedom
Multiple R-squared:  0.8237, Adjusted R-squared:  0.7356
F-statistic: 9.346 on 8 and 16 DF,  p-value: 9.237e-05
```

Latin square example

```
> anova(model2)

Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value    Pr(>F)
tillage    4 109.36   27.34  5.1682  0.007233 **
seed       4 286.16   71.54 13.5236 5.287e-05 ***
Residuals 16  84.64    5.29
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Latin square example

Suppose we are just interested in the effect of tillage on yield.

If we treat the experiment as a CRD, with treatment tillage, we get the same estimates for the tillage that we get from the Latin square, but the treatment is no longer significant.

This is because the variation due to the seed is not taken into account.

Latin square example

```
> CRDfit <- lm(yield ~ tillage, latindata)
> summary(CRDfit)

Call:
lm(formula = yield ~ tillage, data = latindata)

Residuals:
   Min     1Q Median     3Q    Max
  -6.6   -2.6   -0.2    2.4    9.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.5600     0.8612  56.389   <2e-16 ***
tillage1     -2.3600     1.7223  -1.370   0.1858
tillage2      1.0400     1.7223   0.604   0.5527
tillage3      3.0400     1.7223   1.765   0.0928 .
tillage4      0.6400     1.7223   0.372   0.7141
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.306 on 20 degrees of freedom
Multiple R-squared:  0.2278, Adjusted R-squared:  0.07331
F-statistic: 1.475 on 4 and 20 DF,  p-value: 0.2472
```

Latin square example

```
> anova(CRDfit)

Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value Pr(>F)
tillage    4 109.36   27.34  1.4746 0.2472
Residuals 20 370.80   18.54
```

Balanced incomplete block designs (BIBD)

Suppose we have t treatment levels, and b blocks of size $k < t$.

This can happen if there is a natural block size, e.g.:

- twins (2);
- car tyres (2 or 4);
- stamen from a flower (?).

In this case, we cannot use a CBD. Instead we must use a *balanced incomplete block design* (BIBD).

Balanced incomplete block designs (BIBD)

A design is called a BIBD if

- Each treatment occurs at most once in a block;
- Each treatment occurs exactly $r = bk/t$ times (first order balance);
- Each pair of treatments occurs in the same number of blocks, λ say (second order balance).

  There are $t(t-1)/2$ different pairs of treatments and $bk(k-1)/2$ available slots, so we must have

$$\lambda = \frac{bk(k-1)}{t(t-1)} = r\frac{k-1}{t-1}.$$

Balanced incomplete block designs (BIBD)

Given t and k, we can always find a BIBD with $b = \binom{t}{k}$ blocks, by taking all possible subsets of size k.

In this case,

$$r = \binom{t}{k}\frac{k}{t} = \binom{t-1}{k-1} \text{ and } \lambda = \binom{t-1}{k-1}\frac{k-1}{t-1} = \binom{t-2}{k-2}.$$

Sometimes smaller designs are possible.

Steiner systems

A Steiner system $S(t, k, n)$ is an n-element set S, together with a set of k-element subsets of S (called blocks) with the property that each t-element subset of S appears in exactly one block.

Steiner systems with $t = 2$ can be used to construct BIBDs with $\lambda = 1$.

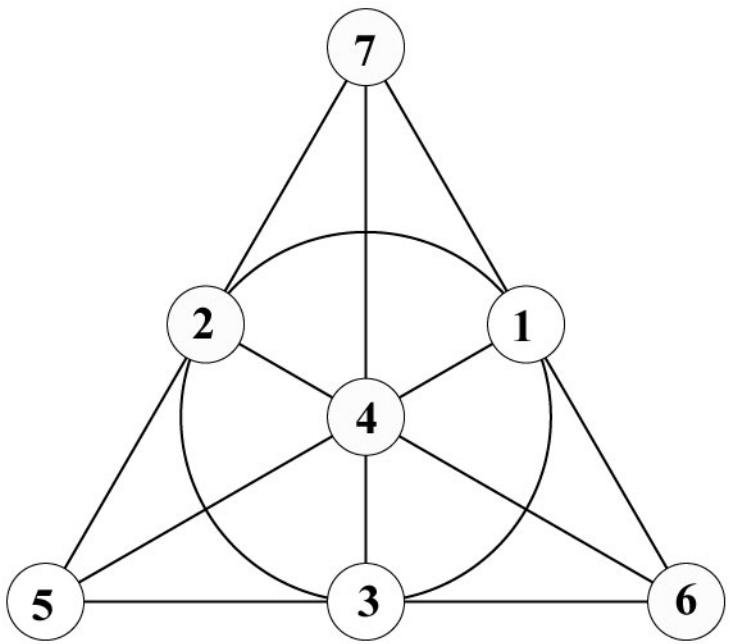
More generally, Steiner systems with general t can be used to construct BIBDs with $\lambda = \binom{n-2}{t-2}$.

Steiner systems

For example, consider the following Steiner triple system with 7 treatments and 7 blocks of size 3.

1	1	1	2	2	3	3
2	4	6	4	5	4	5
3	5	7	6	7	7	6

This system can be represented graphically, with treatments as points and blocks as lines. (This is known as the Fano plane.)



Balanced incomplete block designs (BIBD)

The model for a BIBD is:

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} \text{ for } 1 \leq i \leq b,\ j \in S(i),$$

where $S(i)$ are the treatments in block i.

Balanced incomplete block designs (BIBD)

Writing

$$\mathbf{y} = X_1 \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\beta} \end{bmatrix} + X_2 \boldsymbol{\tau} + \boldsymbol{\varepsilon},$$

we have

$$X_1 = \begin{bmatrix} \mathbf{1}_k & \mathbf{1}_k & 0 & \cdots & 0 \\ \mathbf{1}_k & 0 & \mathbf{1}_k & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{1}_k & 0 & 0 & \cdots & \mathbf{1}_k \end{bmatrix}, \quad X_2 = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_b \end{bmatrix}.$$

Here X_2 has a single 1 in each row, r 1's in each column, and each T_i has at most one 1 in each column.

Balanced incomplete block designs (BIBD)

The reduced normal equations for this model can be written as

$$\frac{\lambda t}{k}\left[I_t - \frac{1}{t}J_t\right]\mathbf{b}_2 = \mathbf{t} - X_2^T X_1 \mathbf{b} =: \mathbf{q},$$

where

$$\mathbf{t} = (y_{\cdot 1}, \ldots, y_{\cdot t})^T$$

are the treatment totals, and

$$\mathbf{b} = (y_{1\cdot}, \ldots, y_{b\cdot})^T$$

are the block totals.

Balanced incomplete block designs (BIBD)

As before, we can show that

$$(X_{2|1}^T X_{2|1})^c = \frac{k}{\lambda t} I_t,$$

so a solution to the reduced normal equations is

$$\mathbf{b}_2 = \frac{k}{\lambda t}\mathbf{q}.$$

Here $\mathbf{q}$ can be thought of as the "adjusted" treatment totals after the effect of the blocks is taken into account.

For a CBD, the presence of the blocks does not affect $\mathbf{b}_2$, but this is not the case for a BIBD.

Orthogonally blocked designs

Consider two designs that each assign n_j units to treatment j, for $j = 1, \ldots, t$.

The two designs may have different block structures, however.

Write the two designs, using the same X_2, as

$$\begin{aligned} \mathbf{y} &= X_1^A \boldsymbol{\beta}_A + X_2 \boldsymbol{\tau} + \boldsymbol{\varepsilon}_A, \\ \mathbf{y} &= X_1^B \boldsymbol{\beta}_B + X_2 \boldsymbol{\tau} + \boldsymbol{\varepsilon}_B. \end{aligned}$$

Equivalent blocking

If

$$\begin{aligned} H_1^A X_2 &= X_1^A (X_1^{AT} X_1^A)^c X_1^{AT} X_2 \\ &= H_1^B X_2, \end{aligned}$$

then

$$\begin{aligned} X_{2|1}^A &= (I - H_1^A) X_2 \\ &= (I - H_1^B) X_2 \\ &= X_{2|1}^B, \end{aligned}$$

and thus the reduced normal equations are the same for both designs.

Equivalent blocking

It follows that for any estimable combination of $\boldsymbol{\tau}$, the estimates are the same.

In this case, we say the blocking is *equivalent* for the two designs.

If a design is equivalent to a CRD, then we say the treatments are *orthogonal* to the blocks.

Exercise: Show that the CRD, CBD and Latin square designs have equivalent blocking.

BIBD vs. CRD

Theorem 8.3

Let

$$X_2 = \begin{bmatrix} \mathbf{1}_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{n_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{n_t} \end{bmatrix}.$$

If we have a CRD

$$\mathbf{y} = \mathbf{1}\mu + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}'$$

and $\mathbf{c}$ *is a contrast, then*

$$\text{var}\,\widehat{\mathbf{c}^T\boldsymbol{\tau}} = \sigma^{2\prime}\sum_{i=1}^{t} c_i^2/n_i,$$

where $\text{var}\,\boldsymbol{\varepsilon} = \sigma^{2\prime} I$.

BIBD vs. CRD

If we have some block structure

$$\mathbf{y} = X_1\boldsymbol{\beta} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon},$$

then

$$\text{var}\,\widehat{\mathbf{c}^T\boldsymbol{\tau}} = \sigma^2 \sum_{i=1}^{t} c_i^2/n_i + \sigma^2 \mathbf{c}^T Q \mathbf{c},$$

where $\text{var}\,\boldsymbol{\varepsilon} = \sigma^2 I$ *and*

$$Q = (X_2^T X_2)^{-1} X_2^T X_1 [X_1^T (I - H_2) X_1]^c X_1^T X_2 (X_2^T X_2)^{-1}.$$

BIBD vs. CRD

Suppose we wish to estimate $\mathbf{c}^T\boldsymbol{\tau}$, and we have n_j observations of treatment j, $j = 1, \ldots, t$.

The difference in the variance of our estimate using the model

$$\mathbf{y} = \mathbf{1}\mu + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}'$$

and the model

$$\mathbf{y} = X_1\boldsymbol{\beta} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}$$

is

$$(\sigma^{2\prime} - \sigma^2)\sum_{i=1}^{t} c_i^2/n_i - \sigma^2\mathbf{c}^T Q\mathbf{c}.$$

BIBD vs. CRD

When the second model has treatments orthogonal to the blocks (CBD), the $\mathbf{c}^T Q \mathbf{c}$ term vanishes.

However, BIBD does *not* have treatments orthogonal to the blocks, so it is only worth doing if the difference between $\sigma^{2'}$ and σ^2 outweighs the $\mathbf{c}^T Q \mathbf{c}$ term.