MAT20306 - Advanced Statistics

Lecture 10: Two-way ANOVA: pairwise comparisons & blocking

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Map of two-way ANOVA

Pairwise comparisons

Type I, II and III of SS

Blocking, RCBD

RE of CRD vs. RCBD

Three or more factors

What after significant F-tests in two-way ANOVA?

We perform pairwise comparisons of means.

Which table of means we inspect depends upon the test for interaction.

- Interaction important, i.e. significant

  compare means for combinations of levels of factors, i.e. one table of means of combinations of the two factors (via one way ANOVA)

- Interaction not important, but (some) main effect(s) important

  compare (marginal) means for levels of individual factors, i.e. two tables of (marginal) means, one table for each factor

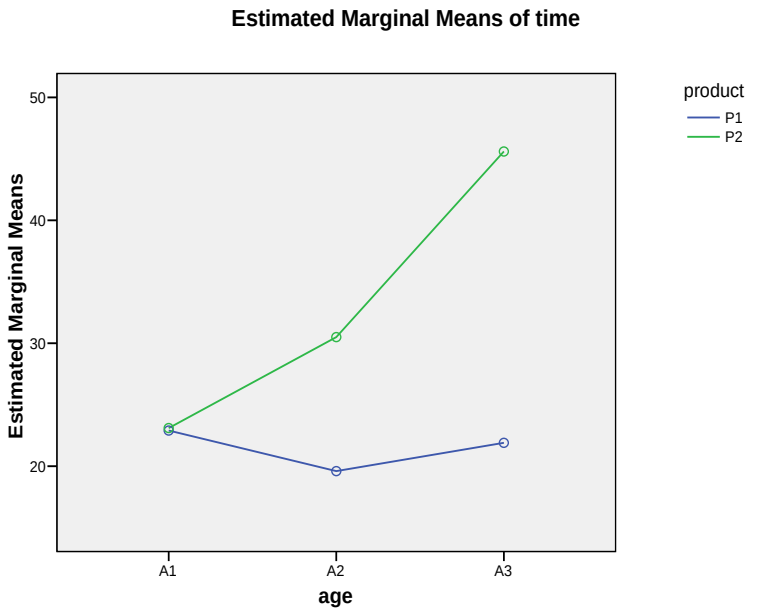
Similar to one-way ANOVA use e.g. Fisher's LSD or Tukey's procedure.

Further examples, attention span commercial

y = attention span of a child for a television commercial

3 age classes (A_1: 5 - 6, A_2: 7 - 8, A_3: 9 - 10 years)
2 types of product (P_1: cereal, P_2: video game)

10 representative children of each age class, randomly assigned to products



← clear interaction

Two factors:

A (age at 3 levels)

P (product at 2 levels)

6 treatments

Attention span, SPSS

Tests of Between-Subjects Effects

Dependent Variable: time

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4705.733[a]	5	941.147	6.398	.000
Intercept	44608.267	1	44608.267	303.228	.000
age	1303.033	2	651.517	4.429	.017
product	2018.400	1	2018.400	13.720	.001
age * product	1384.300	2	692.150	4.705	.013
Error	7944.000	54	147.111		
Total	57258.000	60			
Corrected Total	12649.733	59			

a. R Squared = .372 (Adjusted R Squared = .314)

interaction Age and Product significant (P-value = 0.013)

do not look at main effects!

Example: high blood pressure

135 patients with high blood pressure

y = average oral body temperature of a patient

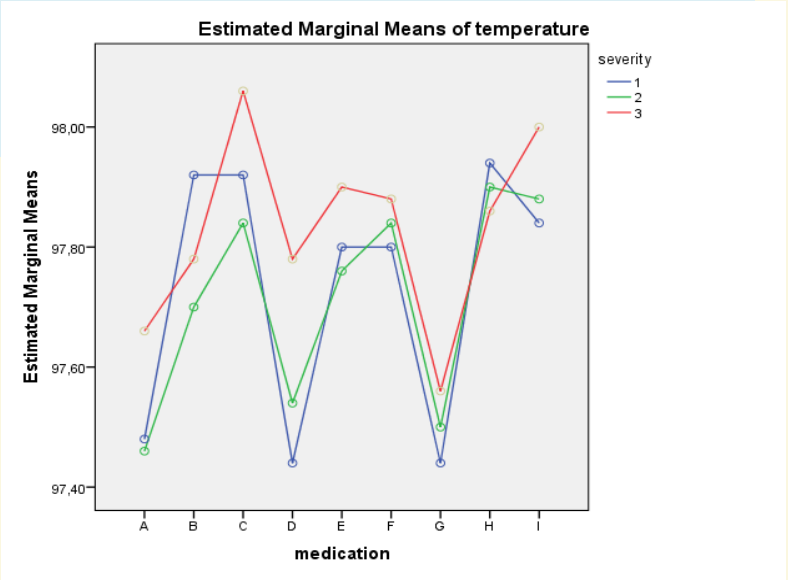
9 medications (factor at 9 levels)

3 levels of severity of blood pressure disorder (factor at 3 levels)

9 medications randomly assigned to 45 random patients from each severity group

Lines roughly parallel:

interaction possibly not important



High blood pressure

Tests of Between-Subjects Effects

Dependent Variable: Body Temperature

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.376[a]	26	.168	6.854	.000
Intercept	1290158.272	1	1290158.272	5.3E+07	.000
severity	.363	2	.181	7.388	.001
med	3.512	8	.439	17.876	.000
severity * med	.501	16	.031	1.276	.226
Error	2.652	108	.025		
Total	1290165.300	135			
Corrected Total	7.028	134			

a. R Squared = .623 (Adjusted R Squared = .532)

balanced scheme, so *SS* are unique

interaction not significant, P-value = 0.23 > 0.05

main effects medication significant, P-value = 0.000 < 0.05

main effects severity significant, P-value = 0.001 < 0.05

High blood pressure

Compare means of medications pairwise (9 * 8 / 2 = 36 pairs):

Tukey's $W = q * \sqrt{\hat{\sigma}_\epsilon^2 / n} = q * \sqrt{MSE/n} = 4.47 * \sqrt{\frac{0.02456}{15}} = 0.181$

t = 9, α = 0.05, df = 108 (from *SSE*), so (table 10 O&L) q is about 4.47

compare means with yardstick *W*

medications can be simply grouped here: {A, D, G} and {B, C, E, F, H, I},

two groups of medications with relatively low or high oral temperature:

A^a B^b C^b D^a E^b F^b G^a H^b I^b

SPSS / R can do it for you !!!

Map of two-way ANOVA

Pairwise comparisons

Type I, II and III of SS

Blocking, RCBD

RE of CRD vs. RCBD

Three or more factors

Some design issues

4. Balanced vs non-balanced design

A configuration of data (the design) is balanced when each treatment (factor level combination) is repeated equally often.

Advantages of balanced designs are that

- estimated effects are simple expressions in terms of sample means
- sums of squares for effects in the model are unique (more to come ...)

When e.g. observations are lost, or the study is observational, the design will commonly not be balanced.

Then:

- sums of squares will depend on the order by which factors are introduced in the model.
- estimated effects (and se's) have more complex structure (but can still be obtained from SPSS, or other software)

Type I and type II SS

In the construction of the F-test, extra *SS* in the numerator is always *SSB* for *B* after *A* and *SSA* for *A* after *B*.

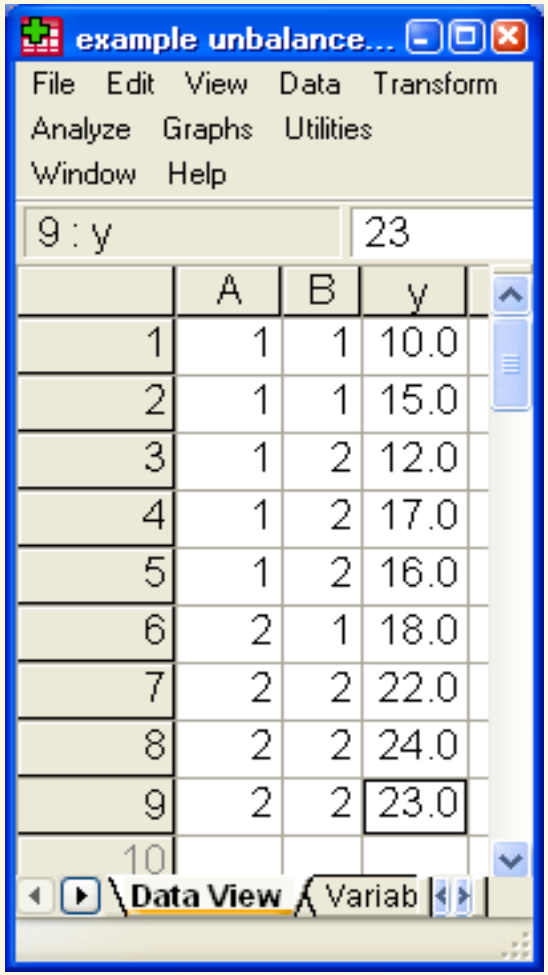
These extra *SS* are called type II *SS*.

For testing, rule of thumb is: use type II, do not use type I.

For a balanced design, type I and type II *SS* are the same.

Unbalanced design, type I, type II *SS*

A non-orthogonal design:



Dependent Variable: y

Source	Type I SS	df
Corr. Model	159.7	3
Intercept	2738.8	1
A	133.5	1
B	23.4	1
A * B	2.9	1
Error	28.5	5
Total	2927.0	9
Corr. Total	188.2	8

B after A

Dependent Variable: y

Source	Type I SS	df
Corr. Model	159.7	3
Intercept	2738.8	1
B	43.6	1
A	113.3	1
A * B	2.9	1
Error	28.5	5
Total	2927.0	9
Corr. Total	188.2	8

A after B

Dependent Variable: y

Source	Type II SS	df
Corr. Model	159.7	3
Intercept	2738.8	1
B	23.4	1
A	113.3	1
A * B	2.9	1
Error	28.5	5
Total	2927.0	9
Corr. Total	188.2	8

SSA+SSB +SSAB +SSE ≠SST (= 188.2)

Summary type I and type II SS

- one-way ANOVA: $SST = SSTreat + SSE$
- in effects model *SSTreat* split into *SS* for interaction & main effects, e.g. *SSA, SSB, SSAB*
- these *SS* can be constructed sequentially and are called type I *SS*
- for balanced designs split is unique, *SS* are not order dependent
- for unbalanced designs split usually not unique, *SS* are order dependent
- use type II (or type III, to be discussed next) *SS* for testing

Type III *SS*

Type III *SS* allows to test for main effects in presence of interaction.

Mathematically sounds good, but tricky in practice.

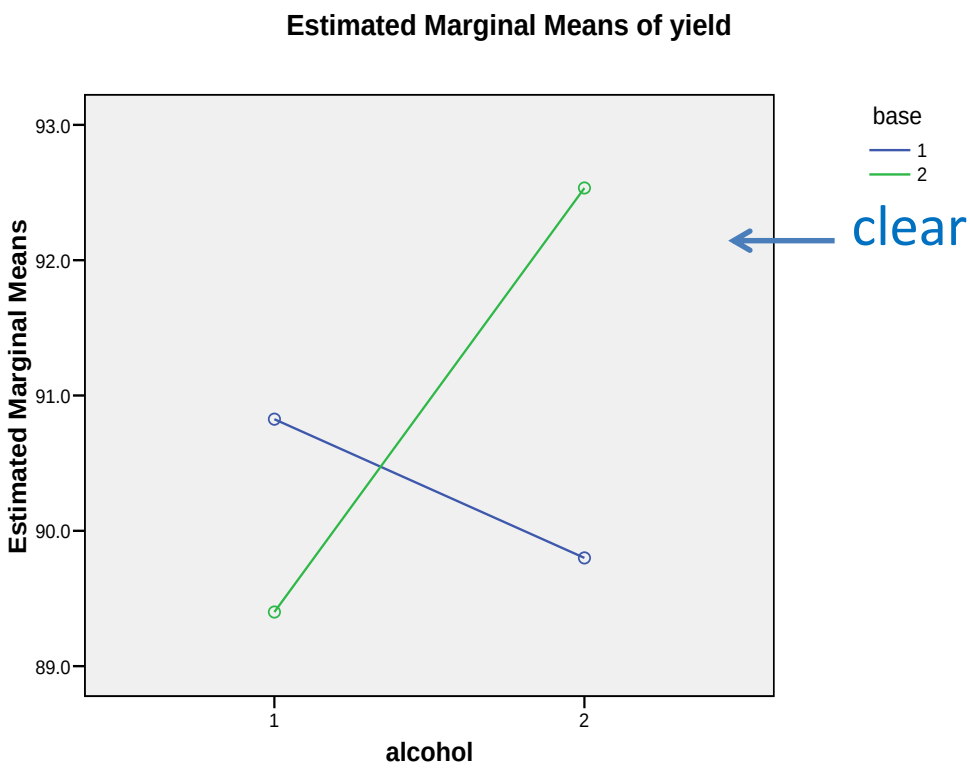
May give misleading results; think of base and alcohol example (next two slides).

Further examples, Alcohol and base

y = % yield of a chemical process

2 types of alcohol (A_1, A_2)

2 types of base (B_1, B_2)



clear interaction

Two experimental factors:

A (alcohol at two levels)

B (base at two levels)

4 treatments

Alcohol and base, SPSS

Tests of Between-Subjects Effects

Dependent Variable: yield

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	18.295[a]	3	6.098	2.460	.123
Intercept	114898.921	1	114898.921	46341.110	.000
alcohol	2.006	1	2.006	.809	.389
base	1.467	1	1.467	.592	.460
alcohol * base	14.821	1	14.821	5.978	.035
Error	24.794	10	2.479		
Total	114942.010	14			
Corrected Total	43.089	13			

a. R Squared = .425 (Adjusted R Squared = .252)

interaction Alcohol and Base significant (P-value = 0.035)

do not look at main effects (in this case very misleading)!

Type II and type III *SS*

Type II gives you sums of squares and test results for main effects that are appropriate when interaction can be ignored.

Remember, type II *SSA* and *SSB* are derived by comparing the additive model with main effects for *A* and *B,* but without interaction, with models with *A* or *B* only, i.e. *A+B* versus *A* (for *SSB*) and *A+B* versus *B* (for *SSA*).

Type III gives you sums of squares for main effects of *A* and *B* that are derived from the model with main effects and interaction for *A* and *B*.

Main effects for e.g. *A* obtained by averaging over the levels of *B*.

Map of two-way ANOVA

Pairwise comparisons

Type I, II and III of SS

Blocking, RCBD

RE of CRD vs. RCBD

Three or more factors

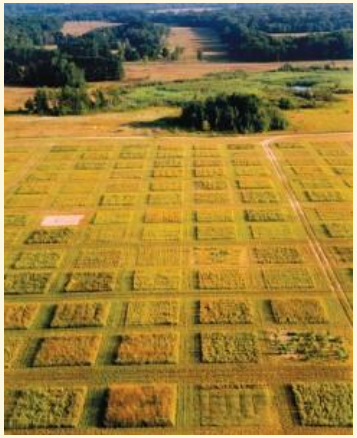
Example: field experiment

Consider a field experiment where expected yield of three potato varieties A, B, C is to be compared.

The field consists of 12 plots, numbered 1 ... 12.

Plots are the experimental units.

How should we design this experiment?



1	2	3	4
5	6	7	8
9	10	11	12

Example: field experiment

Option 1

allocate varieties to plot numbers in some systematic way

A	A	B	C
A	B	B	C
A	B	C	C

Which design do you prefer?

A	A	A	A
B	B	B	B
C	C	C	C

Option 2

Let chance decide which variety goes where and use randomisation.

A	B	B	A
B	A	C	B
C	A	C	C

Randomization, CRD

- With a systematic design treatment differences may be confounded with systematic differences between units.
- Randomization protects against confounders.
- Guarantees that assumption of independent error terms in e.g. one- and two-way ANOVA is valid.

Completely Randomized Design = CRD

Assign treatments randomly to experimental units.

Here with equal number of replicates (4 plots for each variety)

Use of extra qualitative information

Suppose there is a gradient in fertility in the field and we know its direction.

A	C	B	C
A	A	B	B
C	A	C	B

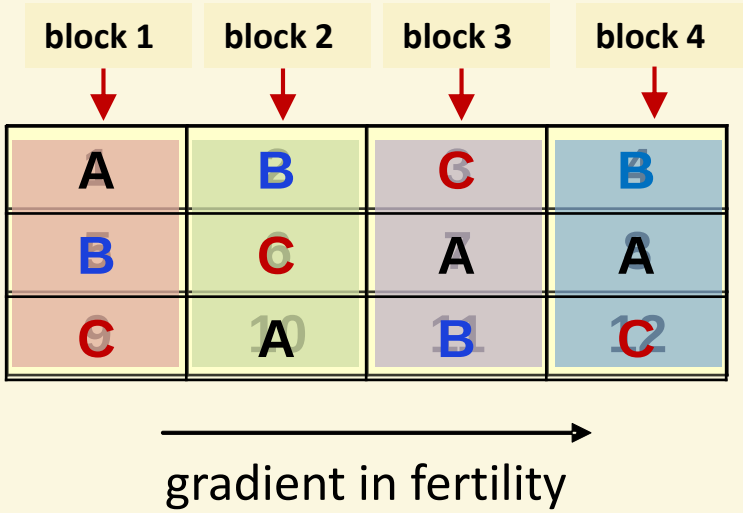
- gradient in fertility +

Do you like this result of randomization in view of the extra information?

Do you accept a design with A more on the left, B more on the right, when we know about a fertility gradient?

Changing the design, we can increase accuracy of comparison between treatments, and reduce our worries at the same time.

Blocking



- Construct blocks = groups of similar plots.
- Randomize treatments within each block.
- Comparison between treatments will be more accurate, because treatments are compared using similar plots within blocks.
- Example of a Randomised Complete Block Design (RCBD).

RCBD - 1

- A block is a group of homogeneous experimental units, e.g.

  plots with similar soil condition in a field experiment
  plants in the same greenhouse,
  people with similar physical condition, ...

- Treatments randomly assigned to the units within each block.

- In an RCBD each treatment appears an equal number of times, usually once, in every block.

  So, RCBD is a balanced design; no worries about sums of squares. In case of missing values use type II (or type III) *SS*.

RCBD - 2

Additive model for factors treatment and block.

Typically, no interaction between treatments and blocks.

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij},$$

with τ_i main effects for treatments, β_j main effects for blocks.

Commonly only a single observation per treatment per block (otherwise add extra index k).

RCBD, sums of squares

Consider RCBD where t treatments appear once in each of b blocks. So, total number of observations is $N = b * t$.

In the ANOVA table:

SSBlock
SSTreat
SSE
SST

$SST = SSBlock + SSTreat + SSE$

Degrees of freedom: $(N-1) = (b-1) + (t-1) + (t-1)(b-1)$

Compared with CRD, in RCBD part (*SSBlock*) is taken out of the *SSE*, to get a more accurate comparison between treatments.

We are filtering part of the variation out of the residual error.

RCBD, F-tests

In the ANOVA table there will be two F-tests:

one for hypothesis

$H_0: \tau_1 = \tau_2 = \cdots = \tau_t = 0$ test statistic: *F = MSTreat / MSE*

(H_0: no differences between treatments)

and one for hypothesis

$H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$ test statistic: *F = MSBlock / MSE*

(H_0: no differences between blocks, i.e. blocking was not useful)

Paired data, a special case of an RCBD

When each block consists of two experimental units, with two treatments in each block, we have paired observations.

The two-sided paired t-test will give the same P-value as the F-test in the additive model with treatment and block as factors, and $F = t^2$.

So, we could consider the field experiment example as a generalization from pairs to quadruples.

5. Estimate treatment means (additive model)

Here we assume that we have one observation for each treatment – block combination. For treatment means and their differences we show estimator and se. Estimates for μ_{ij}'s are in general not useful.

$$\hat{\mu}_i := \hat{\mu}_{i.} = \bar{y}_{i.}$$

$$se(\bar{y}_{i.}) = \hat{\sigma} / \sqrt{n_{i.}} = \hat{\sigma} / \sqrt{b}$$

$$\hat{\mu}_i - \hat{\mu}_j = \bar{y}_{i.} - \bar{y}_{j.}$$

$$se(\bar{y}_{i.} - \bar{y}_{j.}) = \hat{\sigma} \sqrt{\frac{1}{n_{i.}} + \frac{1}{n_{j.}}} = \hat{\sigma} \sqrt{2/b}$$

With SPSS, means are usually estimated using the EMM table and the differences using a Post Hoc method.

The student should know when mean and / or st. error could be easily calculated by hand, and when not.

Example: insecticides and string beans

3 insecticides (treatments) each applied once in each of 4 plots (blocks)

response y = number of seedlings that emerge out of 100 seeds of string beans

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} ,$$

$$\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2) \quad \text{independent},$$

$$\tau_i = \text{effect of } i-\text{th insecticide}, \; i = 1, 2, 3 \; (\tau_3 = 0),$$

$$\beta_j = \text{effect of } j-\text{th plot}, \; j = 1, 2, 3, 4 \; (\beta_4 = 0).$$

Tests of Between-Subjects Effects

Dependent Variable: number of seedlings

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2270.000[a]	5	454.000	104.769	.000
Intercept	67500.000	1	67500.000	15576.923	.000
INSECT	1832.000	2	916.000	211.385	.000
PLOT	438.000	3	146.000	33.692	.000
Error	26.000	6	4.333		
Total	69796.000	12			
Corrected Total	2296.000	11			

a. R Squared = .989 (Adjusted R Squared = .979)

significant differences among treatments (= insecticides)

significant differences among blocks (=plots)

Insecticides, 0.95 CI for $\mu_1 - \mu_2 = \tau_1 - \tau_2$

Derive an estimate, associated standard error and 95% confidence interval for the difference between the population means of insecticides 1 and 2.

Insecticide

Dependent Variable: Number of seedlings

Insecticide	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	58,000	1,041	55,453	60,547
2	87,000	1,041	84,453	89,547
3	80,000	1,041	77,453	82,547

$\hat{\tau}_1 - \hat{\tau}_2 = \bar{y}_{1.} - \bar{y}_{2.} = 58 - 87 = -29$

$se(\hat{\tau}_1 - \hat{\tau}_2) = se(\bar{y}_{1.} - \bar{y}_{2.}) = \sqrt{MSE * 2/b} = \sqrt{2 * 4.333/4} = \sqrt{2.166} = 1.47$

0.95 confidence interval : $-29 \pm t_6(0.025) \times 1.47$, with $t_6(0.025) = 2.45$

→ 0.95 confidence interval for $\tau_1 - \tau_2$ is: (-32.6, -25,4)

Insecticides, continued (in class)

Suppose that we compare the insecticides pairwise by Fisher's LSD.

We know that b=4, t=3, $\hat{\sigma}^2 = 4.3$, so $\hat{\sigma}=2.1$. The 3 observed treatment means are 58 87 and 80, respectively

- Calculate the *LSD* for $\alpha = 0.05$.

  Remember: $LSD = t_{crit}\sqrt{(2s_\varepsilon^2 / n)}$, here: $LSD = t_{crit}\sqrt{(2*MSE / b)}$

- Compare the insecticide means and draw the conclusions.

Multiple Comparisons

Dependent Variable: Number of seedlings

LSD

(I) Insecticide	(J) Insecticide	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-29.00*	1.472	.000	-32.60	-25.40
	3	-22.00*	1.472	.000	-25.60	-18.40
2	1	29.00*	1.472	.000	25.40	32.60
	3	7.00*	1.472	.003	3.40	10.60
3	1	22.00*	1.472	.000	18.40	25.60
	2	-7.00*	1.472	.003	-10.60	-3.40

Map of two-way ANOVA

Pairwise comparisons

Type I, II and III of SS

Blocking, RCBD

RE of CRD vs. RCBD

Three or more factors

Insecticides, relative efficiency of RCBD versus CRD

When plots (blocks) are substantially different, we expect a RCBD to be more efficient than a CRD.

Relative efficiency is:

$$RE = \frac{(b-1)\,MSBlock + b(t-1)\,MSE}{(bt-1)\,MSE}$$

Insecticides:
b = 4, *MSBlock* =146.000, t = 3, *MSE* = 4.333, so: *RE* = 9.9.

You need about 10 times more observations for a CRD to have the same precision as the RCBD in the insecticides example.

Relative efficiency, rough & ready alternative - 1

Rough and ready alternative:

fit RCBD and fit CRD separately on the data,

save *MSE* from each output,

RE is ratio of *MSE*'s of CRD and RCBD.

$$RE = MSE_{\text{CRD}} / MSE_{\text{RCBD}}$$

or

$$RE = \frac{(b-1)\,MSBlock + b(t-1)\,MSE}{(bt-1)\,MSE}$$

Relative efficiency, rough & ready alternative - 2

RE is ratio of *MSE* 's of CRD and RCBD.

For insecticide data:

Tests of Between-Subjects Effects

Dependent Variable: number of seedlings

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2270.000[a]	5	454.000	104.769	.000
Intercept	67500.000	1	67500.000	15576.923	.000
INSECT	1832.000	2	916.000	211.385	.000
PLOT	438.000	3	146.000	33.692	.000
Error	26.000	6	4.333		
Total	69796.000	12			
Corrected Total	2296.000	11			

a. R Squared = .989 (Adjusted R Squared = .979)

$$RE = \frac{(b-1)\,MSBlock + b(t-1)\,MSE}{(bt-1)\,MSE}$$

- for RCBD: *MSE* = 4.333 (read from output SPSS)

  MSBlock = 146; b-1 = 3; t-1 = 2

- RE = (3*146 + 4*2*4.33) / 11*4.33 = 9.9

Map of two-way ANOVA

Pairwise comparisons

Type I, II and III of SS

Blocking, RCBD

RE of CRD vs. RCBD

Three or more factors

Factorial structure and blocks, protein in bread

y = % protein in a loaf of bread

3 types of flour (factor A, 3 levels)

3 temperatures (factor B, 3 levels)

3 days (blocking factor), 9 loaves per day, 27 loaves in total

Treatments	*SSTreat*	df = 9-1	= 8
Blocks (days)	*SSBlock*	df = (3-1)	= 2
Residual (error)	*SSE*	df = (9-1)*(3-1)	= 16
Total	*SST*	df = 27 – 1	= 26

split up *SST*reat

main effects *A* *SSA*	df = (3-1)	= 2	
main effects *B* *SSB*	df = (3-1)	= 2	
interaction *AB* *SSAB*	df = (3 - 1) * (3 - 1)	= 4	2 + 2 + 4 = 8

Three factors or more

- There can be more than two factors, e.g. in bread example two treatment factors and one block factor, no interaction assumed between block and treatment factors.
- Respect hierarchy of model terms:

  with 3 factors & all interactions, first *ABC*, then *AB*, *AC*, *BC*, then *A*, *B*, *C*

  in testing model terms, peel off the model like an onion from outside:

  - 1st layer: 3rd order interactions *ABC*,
  - 2nd layer: 2nd order interactions *AB*, *AC*, *BC*,
  - and finally main effects *A*, *B*, *C*,

  So, e.g. do not look at main effect for *B*, when *AB,* is significant.
- In balanced designs all *SS* commonly taken from same full ANOVA table.

Summary of the design in ANOVA

- In observational study cause-effect relationship generally cannot be concluded, whereas in experimental study it can
- Make distinction between experimental and measurement units, otherwise multiple measurements per experimental unit lead to pseudo-replication
- In experimental design pay attention to replication, randomization and reducing noise (by blocking or use of covariates = Lecture 11)
- Completely Randomized Design (CRD):
  treatments fully randomized over all experimental units
- Randomized Complete Block Design (RCBD):
  treatments randomized over experimental units within each of several blocks; a block is a group of similar experimental units. .