How much better is an RCBD? Approximate answers

A) relative efficiency:

efficiency $> 1 \Rightarrow$ RCBD more efficient

Requires fewer observations per trt to provide same precision

focuses on error variance and/or se_{diff} and/or sample sizes for comparable precision

Plant study (3 trt, 10 blocks), $MS_{RCBD} = 3.7$, $MS_{CRD} = 5.2$

1a) using se_{diff}

RCBD: 10 obs/trt, $\Rightarrow se_{diff} = \sqrt{3.7 \times 2/10} = 0.86$

CRD: 10 obs/trt, $\Rightarrow se_{diff} = \sqrt{5.2 \times 2/10} = 1.02$

$$\text{efficiency, RCBD vs CRD} = \left(\frac{se_{diff,CRD}}{se_{diff,RCBD}}\right)^2 = (1.02/0.86)^2 = 1.41$$

1b) using sample sizes that provide the same se_{diff}

RCBD, current design, 10 obs/trt, 30 obs total $\Rightarrow se_{diff} = 0.86$

What n in a CRD (MSE = 5.2) gives an $se_{diff} = 0.86$?

$0.86 = \sqrt{5.2 \times 2/n}$, $n = 5.2 \times 2/0.86^2 = 14.1$

CRD with same precision requires $n = 14.1$ obs/trt to provide $se_{diff} = 0.86$

$$\text{efficiency, RCBD vs CRD} = \frac{n_{crd}}{n_{rcbd}} = \frac{14.1}{10} = 1.41$$

2) directly using MSE from two ANOVA tables

$$\text{efficiency, RCBD vs CRD} = \frac{MSE_{crd}}{MSE_{rcbd}} = \frac{5.2}{3.7} = 1.41$$

How much better is an RCBD: More correct answers

We're using data collected from an RCBD to infer what a CRD would do

Because we're calculating MSE from a CRD using data from an RCBD

there are more possible randomizations in the CRD

A) Include two correction factors in the relative efficiency computation

a) Correction based on error df for both RCBD and CRD

$$\text{efficiency, RCBD vs CRD} = \left[\frac{df_{rcbd}+1}{df_{rcbd}+3}\right]\left[\frac{df_{crd}+3}{df_{crd}+1}\right]\left[\frac{MSE_{crd}}{MSE_{rcbd}}\right] = \left[\frac{19}{21}\right]\left[\frac{30}{28}\right]1.41 = 1.37$$

b) Correction to the SS to account for the additional randomizations in the CRD

$$MSE_{crd} = \frac{SS_{blocks} + SSE_{rcbd} + (df_{trt}-1)MSE_{rcbd}}{n_{blocks}\, n_{trt} - 1}$$

Absolutely needed when studying relative efficiency for different types of studies

My sense: the approximate computations are good enough for a rough answer

B) Relative Performance: focuses on tests and confidence intervals

they depend on the error df

RCBD has fewer error df, hence larger T quantiles, than does CRD

when df large, e.g. ≥ 60, RE $\approx$ RP because T quantiles are similar

when df small, RP < RE ⇒ RCBD not as good as RE shows
because RCBD has fewer error df
Look at the Doncaster et al optional reading if interested

Practical experience with blocks:
Almost all Agronomy studies and most studies in other areas use blocks
Sometimes only because they are a convenient way to organize the experiment
Typical RCBD efficiency in Agronomy field studies is 1.10
Blocks appear in the ANOVA table, but testing block effects is meaningless
You included them because you believe blocks differed

Quick evaluation of whether blocking was useful
Context: have used blocking in a study, have the ANOVA table
Can you say whether blocking was useful without computing RE or RP
Calculate the F statistic for blocks:

$$F = \frac{MS_{blocks}}{MS_{error}}, \quad \begin{array}{l} F > 1 \Rightarrow \text{RE } > 1 \\ F < 1 \Rightarrow \text{RE } < 1 \end{array}$$

Plant example:

RCBD					CRD			
Source	df	SS	MS	F	Source	df	SS	MS
Trt	2	51.3	25.6		Trt	2	51.3	25.6
Blocks	9	74.8	8.31	2.26				
Error	18	62.3	3.46		Error	27	66.6	2.47

F for blocks > 1, RCBD more precise
2nd example:

RCBD					CRD			
Source	df	SS	MS	F	Source	df	SS	MS
Trt	2	51.3	25.6		Trt	2	51.3	25.6
Blocks	9	23.5	2.61	0.75				
Error	18	62.3	3.46		Error	27	85.8	3.18

F for blocks < 1, CRD more precise
p-value for blocks is irrelevant.
$p < 0.05$ typically requires $F > 2$ to 3

What to do when Blocks are not helpful ($F_{blocks} < 1$)
Multiple opinions

Here is what I believe is the most common practice in the US
Do not drop blocks from the model
because the analysis needs to follow the design
design had blocks ⇒ analysis uses blocks
When designing a **new** experiment
1) find a better way to block
2) consider not using blocks

Practical advice about blocking:

1. Paired data are blocks with 2 trts per block
   Can analyze either as
   paired data set (differences within each block)
   or as a block + treatment data set
2. before you design, think
   how can eu's be grouped to make groups of similar eu's?
3. when have blocks, use them again and again
   Thought example: Field experiment, 4 blocks
   Need to sample over multiple days, may be a week or more apart
   "Use them again" ⇒ deliberately confound field block and sampling day
   Sample an entire block in a day (or 2 full blocks)
   Never sample 1/2 a block in a day
   Block effects = field differences + day differences
   don't care that they are can't be separated
4. Field blocks do not need to be the same shape or even contiguous
   similarity of eu's is all that matters

Should blocks be modeled as a fixed effect or a random effect?
Many folks automatically treat blocks as random
Because the design is called Randomized ...
"Randomized" is because treatments are randomized
has nothing to do with fixed or random
Strong opinions for both options
My view: Blocks are a mechanism to improve precision
so how does the choice affect conclusions about treatments?

The facts:
1. when $\hat{\sigma}^2_{blocks} > 0$ (equiv. to MS blocks > MS error) and blocks are complete,
   choice is irrelevant: same inference about trt differences
   (a) when unequal # obs per block, slight difference
   (b) Big difference is the se of a treatment mean

$$\text{F:} = \sqrt{\frac{\hat{\sigma}^2_{error}}{n}}, \quad \text{R:} = \sqrt{\frac{\hat{\sigma}^2_{error} + \hat{\sigma}^2_{blocks}}{n}}$$

Illustration, using mead53b data (10 blocks)
Mead: MSE = $\hat{\sigma}^2_{error} = 3.7$, MS blocks = 8.3, $\hat{\sigma}^2_{blocks} = 1.24$

Data set	se of	fixed blocks	random blocks
Mead	trt mean	$\sqrt{3.7/10} = 0.61$	$\sqrt{(3.7 + 1.24)/10} = 0.70$
	trt diff	$\sqrt{3.7 * 2/10)} = 0.857$	$\sqrt{3.7 * 2/10} = 0.857$

Why the difference in se trt mean, even when $\sigma^2_{blocks} > 0$?
inference with Fixed blocks is about treatments in the blocks used in the study
Random blocks is about treatments in new blocks from the same population
requires more assumptions

2. when MS blocks < MSerror, Random with REML $\Rightarrow \hat{\sigma}^2 = 0$
   MSE is wrong
   Fixed blocks $\Rightarrow$ correct inferences about trt diff's
   Random blocks $\Rightarrow$ wrong inferences about trt diff's
   Unless allow negative estimates of the block variance component
   Repeat with "bad" data:
   Bad: MSE = 3.46, $\hat{\sigma}^2_{error} = 2.46$, $\hat{\sigma}^2_{blocks} = 0$

Data set	se of	fixed blocks	random blocks
Bad	trt mean	$\sqrt{3.46/10} = 0.59$	$\sqrt{(2.46 + 0)/10} = 0.50$
	trt diff	$\sqrt{3.46 * 2/10} = 0.83$	$\sqrt{2.46 * 2 * /10} = 0.70$

Why the differences when σ^2_{blocks} forced = 0?
Already seen that "pulling up" a VC to = 0 (i.e. REML) changes σ^2_{error}
Smaller se mean under Random model is a cause for concern
More general inference (to new blocks) should have larger se

3. when # blocks is small, $\hat{\sigma}^2_{blocks}$ is poorly estimated
   So really don't have a good idea about σ^2_{blocks}
   Matters a lot of random blocks, less so for fixed blocks

4. Random assumes blocks are a random sample from a population
   Will make sense for some studies
   a) randomly select farms in IA, identify plots on those farms
   assign treatments to those plots
   b) Prepare "soil beds"
   Assign treatments within each soil bed
   What about blocks in these studies?
   a) Marsden Farm - one particular area of the Ag Experimental fields
   divided into 4 blocks. That's that entire area of Marsden Farm
   b) Animal nutrition studies: common to construct blocks from initial weight
   e.g., if have 5 treatments,
   smallest 5 animals = block 1, next smallest 5 = block 2, $\cdots$ largest 5 = last block
   Blocks are not random samples from some (hypothetical) population of blocks

5. One reason to use random blocks: "recover interblock information"
   Example study:
   Randomly choose students = blocks
   Collect baseline data, apply a treatment, collect followup data
   Estimate change as difference = followup - baseline
   When no missing data, average difference is meaningful
   What if some students don't show up one of the measurements?
   They still provide information about followup - baseline!
   Imagine an unpaired study where every student only provides partial information
   Can still estimate the change; less precise because of variability among students
   Fixed blocks: change only based on complete data
   Students with both baseline and followup $\Rightarrow$ difference for that student
   Random blocks: combines precise estimate (change within a student)
   and information from students missing one of the measurements
   = "interblock information"
   Most useful when block variance component is small
   Critical assumption: The two estimates of change are identical
   Change in the paired data is the same as change in the unpaired data
   Huge assumption about why individuals are missing one of the measurements
   Basically: missingness is not informative

Blocking to deal with more than one source of variability:
Context: Two (or more) sources of unwanted variability
e.g., field plot and harvesting day
Deliberately confound them if possible
Not always possible: Brome grass study
Field experiment, 5 treatments
plots vary by wetness (wet - dry) and distance to a major roadway
Field picture: define very small blocks by combination of wetness and distance

Latin Square:
Arrange treatments very carefully (see picture of 3 x 3 LS)
Very common in animal nutrition:
blocks = cows and period - see picture
# treatments = # row blocks = # column blocks
model includes additive effects of each blocking variable

$$Y_{i(jk)} = \mu + \tau_i + r_j + c_k + \varepsilon_{ijk}$$

ANOVA table

Source	DF for General	3 x 3 LS
Treatment	T-1	2
Rows	R - 1	2
Columns	C - 1	2
Error		2
c.Total	T^2 - 1	2

T = R = C for the standard Latin Square
  Is blocking useful? Separately look at rows and at columns
    Is F > 1? If so, blocking by (rows, columns) is good
  Error df often small, unless many treatments or multiple squares

Latin Rectangles and multiple Latin Squares
  Increase error df by replicating the LS
  Two ways to replicate:
  1. Latin Rectangle: see picture
       Squares have a row (or a column) in common
  2. Multiple Latin Squares: see picture
       Nothing shared - need care writing model
       e.g., 2 3x3 LS (3 treatments, total of 6 rows, 6 columns)

Wrong		Right	
Source	df	Source	df
Rows	5	Square	1
Columns	5	Row(Square)	4
Treatments	2	Col(Square)	4
		Treatments	2
Error	5	Error	6
c.total	17	c.total	17

Randomizing a Latin Square
  The preferred approach (randomize where ever possible)
    Randomly choose a unit square from a book of designs
    Permute the rows
    Permute the columns
    Randomly assign treatments to the LS "letters"

Carryover
  One of the blocking variables is time
    e.g., period in the cow and period LS
    Assumption in a LS is that current response only due to current treatment
  May be violated when treatments are sequential

- Solutions:
  - Washout: time between experimental periods
  - Balance for carryover ((Williams designs)
    - Each treatment preceded equally often by all other treatments
    - Requires an extra period

Strip plot designs (see picture)

- Has rows and columns but is not a LS
  - Treatments assigned to rows and to columns
  - Will see how to analyze after studying split plot designs