VARIATIONS ON THE RCBD, RELATIVE EFFICIENCY

VARIATIONS OF THE RCDB

RCBD with replication

If each treatment is replicated multiple times within each block (multiple experimental units per treatment within each block), we can model, and test for, interactions between treatments and blocks (Kuehl, p. 271). If y_{ijk} is the kth replicate in the combination of treatment i and block j,

$$y_{ijk} = \mu + \tau_i + \rho_j + (\tau\rho)_{ij} + e_{ijk},$$

where τ_i is the ith treatment effect, ρ_j is the jth block effect, $(\rho\tau)_{ij}$ is the effect of the interaction of treatment i and block j, and the e_{ijk}'s are random errors.

The computations of sums of squares and the tests for treatment and interaction effects are identical to those for two-way ANOVA (note that block can be treated as either fixed or random depending on situations).

If each treatment is replicated multiple times within each block (multiple experimental units per treatment within each block), we can model, and test for, interactions between treatments and blocks (Kuehl, p. 271).

With replication, the model equation for the RCBD experiment will be identical to that for the two-factor complete randomized experiments. As we discussed in the previous lecture, the key difference between the two designs are in the randomization schemes and some subtle implications are not be reflected in the model equation.

The computations of sums of squares and the tests for treatment and interaction effects are identical to those for two-way ANOVA. Note that block can be treated as either fixed or random depending on situations.

RCBD with subsamples

If two or more subsamples are taken per experimental unit, we can model the observation for the kth subsample from the experimental unit in the ith treatment and j th block as

$$y_{ijk} = \mu + \tau_i + \rho_j + e_{ij} + d_{ijk},$$

where the d_{ijk}'s are random effects for subsamples, assumed $N(0, \sigma_d^2)$.

Note that since the treatment and randomization procedure remain the same–we simply made more measurements on each experimental unit—we still cannot model the interaction between treatments and blocks.

The analysis is modified as described on pp. 160-163 for the single-factor design.

Easier: average over subsamples, and do an ordinary RCB analysis.

Like in other designs we have seen before, we need to be careful not to confuse subsamples with true replicates.

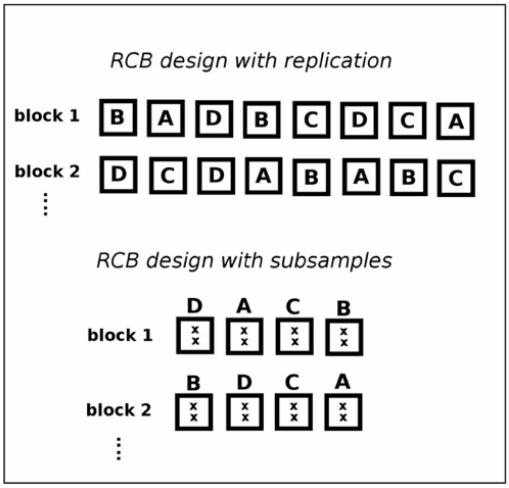
When there are subsamples or repeated measures, we will need an additional random effect term in the model equation for the subsamples.

When performing tests for treatment effects, we should use the mean square for the error terms not the mean square for the subsamples as the denominator in the F-tests.

Note that since the treatment and randomization procedure remain the same–we simply made more measurements on each experimental unit—we still cannot model the interaction between treatments and blocks.

As an easier alternative, one can average over subsamples, and do an ordinary RCB analysis on the averages.

RCBD with replication versus RCBD with subsamples



Possible layouts for RCBD with replication of treatments (upper) and RCBD with subsamples but no replication (lower). Four treatments (A–D) in each block.

The figure shows the difference between an RCBD with replication and an RCBD with subsamples.

The key is to follow the randomization of treatments, see how the treatments are assigned.

In the design with replication, in each block, each treatment is replicated two times.

In the design with subsamples, in each block, each treatment is replicated only once, and then multiple measurements were taken.

Block can be modeled as a random effect

Block can be modeled as a random effect (Kuehl, p. 275). If there is no treatment-block interaction, the sums of squares and F-tests are calculated as before.

As in the case of fixed block effects, it can be shown that

$$Var(\overline{y}_{i\cdot} - \overline{y}_{k\cdot}) = \sqrt{\sigma^2(1/r + 1/r)}.$$

In other words, treating block as random will not affect the inference on **contrasts of main-effect means**.

The variance of $\overline{y}_{i\cdot}$ has one more component as compared to with fixed-effect blocks:

$$Var(\overline{y}_{i\cdot}) = \frac{1}{r}(\sigma^2 + \sigma_b^2).$$

Block can sometimes be modeled as a random effect. If there is no treatment-block interaction, the sums of squares and F-tests are calculated as before.

Treating block as random will not affect the inference on contrasts of main-effect means. The formula is the same as before for variance of the difference of the sample means of two treatment groups.

The variance of the sample mean for a single treatment group, however, will have one more component as compared to with fixed-effect blocks.

Factorial experiments can be replicated in blocks

For a two-factor factorial with blocking, the model is:

$$y_{ijk} = \mu + \rho_k + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk},$$

where

- ρ_k is the effect of the kth block ($k = 1, \ldots, r$),
- α_i is the effect of the ith level of factor A ($i = 1, \ldots, a$), and
- β_j is the effect of the jth level of factor B ($j = 1, \ldots, b$).

Block can be modeled as either a fixed or random effect.

Tests of hypotheses about A, B, and AB are done as in the usual factorial design. Note that we cannot test for interactions involving block, since there is no replication of the factorial within blocks.

The design within each block does not have to be completely randomized design. For example, factorial experiments can be replicated in blocks.

Here we give the model equation for a two-factor factorial with blocking. Block can be modeled as either a fixed or random effect.

Tests of hypotheses about A, B, and AB are done as in the usual factorial design. Note that we cannot test for interactions involving block, since there is no replication of the factorial within blocks.

ANOVA table

Table 8.12 Analysis of variance for a two-factor treatment design in a randomized complete block experiment design

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*	*Expected Mean Square*
Total	$rab-1$	*SS* Total		
Blocks	$r-1$	*SS* Blocks		
A	$a-1$	*SSA*	MSA	$\sigma^2+rb\theta_a^2$
B	$b-1$	*SSB*	MSB	$\sigma^2+ra\theta_b^2$
AB	$(a-1)(b-1)$	*SS*(*AB*)	$MS(AB)$	$\sigma^2+r\theta_{ab}^2$
Error	$(ab-1)(r-1)$	*SSE*	MSE	σ^2

*See Chapter 6 for computational formulae of SSA, SSB, and $SS(AB)$. SSE is obtained by subtraction. SS Blocks $= ab\Sigma(\bar{y}_{..k}-\bar{y}_{...})^2$.

This is the ANOVA table for a two-factor factorial experiment in a randomized complete design.

Within each block, a two-factor completed randomized design is used.

With this design, we can test for the A, B main effects or their interaction, but not the interactions between these treatment effects and the block effect: the between block variation is viewed as error terms under this model.

Meta-analysis

"Meta-analysis" means combining results from multiple experiments.

If all the experiments combined follow the same design, the combined set of experiments can be thought as an RCBD experiment: each experiment is a block.

"Meta-analysis" means combining results from multiple experiments.

If all the experiments combined follow the same design, the combined set of experiments can be thought as an RCBD experiment. Each experiment is a block.

Example

"Meta-analysis" of 11 plankton community experiments by different investigators. The goal was to determine whether plankton biomass is controlled from the 'top down' (by fish predation), or from the 'bottom up' (by nutrient uptake).

Each experiment is a block. Within each block, a factorial design was used: fish (High/Low), nutrient (High/Low)

Two responses were of interest: phytoplankton and zooplankton biomass.

In homework **Problem 5.3**, you are asked to analyze this data set.

- The model equation and ANOVA table for factorial experiments replicated in blocks on previous pages might be useful when you work on that homework problem.

In this example, a researcher wants to do a "mata-analysis" of 11 plankton community experiments by different investigators. The goal was to determine whether plankton biomass is controlled from the 'top down' (by fish predation), or from the 'bottom up' (by nutrient uptake).

Each experiment is a block. Within each block, a factorial design was used: fish (High/Low), nutrient (High/Low)

Two responses were of interest: phytoplankton and zooplankton biomass.

In homework **Problem 5.3**, you are asked to analyze this data set.

The model equation and ANOVA table for factorial experiments replicated in blocks on previous pages might be useful when you work on that homework problem.

RELATIVE EFFICIENCY

Relative efficiency of the randomized block design

The rationale behind the RCBD is that by blocking, we hope to reduce within-block variation and increase the precision for treatment comparisons. But how do we know whether the blocking is actually effective? To quantify the potential benefit of blocking, we introduce the concept of "relative efficiency".

The ***relative efficiency*** of the RCB design is the increase in replications that is required if a completely randomized design is used, as compared to a randomized block, to achieve the same power.

The rationale behind the RCBD is that by blocking, we hope to reduce within-block variation and increase the precision for treatment comparisons. But how do we know whether the blocking is actually effective? To quantify the potential benefit of blocking, we introduce the concept of "relative efficiency".

The relative efficiency of the RCB design is the increase in replications that is required if a completely randomized design is used, as compared to a randomized block, to achieve the same power.

Relative efficiency of the randomized block design

The precision for mean comparisons depends on the **error variance**. Let

- σ_{cr}^2 = error variance in completely randomized design,
- σ_{rcb}^2 = error variance in RCB design

Estimate σ_{rcb}^2 and σ_{cr}^2 by

- s_{rcb}^2 = MSE from randomized block design.
- $s_{cr}^2 = \frac{\text{SS Blocks}+r(t-1)\text{MSE}}{rt-1} = \frac{(r-1)\text{MS Blocks}+r(t-1)\text{MSE}}{(rt-1)\text{MSE}}$

Then the relative efficiency

- $RE = \frac{s_{cr}^2}{s_{rcb}^2} = \frac{(r-1)\text{MS Blocks}+r(t-1)\text{MSE}}{(rt-1)\text{MSE}} = \cdots = k + (1-k)\cdot H,$

where $k = r(t-1)/(rt-1)$ and $H = \text{MS Blocks/MSE}$.

H is the F-statistic for blocks that would appear in the ANOVA table, and:

$$H < 1 \Leftrightarrow RE < 1, \quad H = 1 \Leftrightarrow RE = 1, \quad H > 1 \Leftrightarrow RE > 1$$

The precision for mean comparisons depends on the error variance.

Using data from a completed RCBD experiment, the error variance form the RCBD model can be estimated by the MSE from the fitted RCBD model.

The error variance of a completely randomized model can be estimated as

(SS Block + r times (t-1) times MSE) divided by (rt -1).

which is also a weighted average of MS Block and MSE.

The relative efficiency is defined as the ratio of the two estimated error variances for the completely randomized model and for the RCBD model.

The relative efficiency is a linear increasing function of H, the ratio of MS Blocks over MSE.

Note that this H is the F-statistic for blocks that would appear in the ANOVA table, even though we said earlier that the assumptions for F-test for block effects may not hold in an RCBD experiment.

Note that H is less than, equal to, or greater than one exactly when the relative efficiency is. That's why earlier we said that F-test for blocks can be used to see whether blocking is effective.

Notes on RE

Intuitively, the results on the previous page indicates that the RCBD will more efficient than a randomized design when $MS_{Blocks} > MSE$, i.e., when between-block variance is greater than within-block variance.

In other words, blocking will be effective if experimental units from the same block are more homogeneous (in terms of responses to treatments) than experiment units from different blocks.

(A more complicated formula for RE, on the bottom of Kuehl's p. 273, involves an adjustment for the fact that we estimate σ^2 by s^2.)

Intuitively, the results on the previous page indicates that the RCBD will more efficient than a randomized design when (MS Blocks) is greater than MSE, i.e., when between-block variance is greater than within-block variance.

In other words, blocking will be effective if experimental units from the same block are more homogeneous (in terms of responses to treatments) than experiment units from different blocks.

Our textbook gives a more complicated formula for the relative efficiency which involves an adjustment for the fact that we estimate (sigma squared) by (s squared).

Exercise

Recall the nitrogen content of wheat example, an RCBD experiment with 6 treatments replicated in 4 blocks ($r = 4$, $t = 6$). From the ANOVA table fitted to the experimental data, we get $\text{MS Blocks} = 65.67$ and $MSE = 7.2$.

Since MS Blocks is much greater than MSE, we know that blocking is effective in this case.

Can you compute the relative efficiency RE?

Recall the nitrogen content of wheat example, which is an RCBD experiment with 6 treatments replicated in 4 blocks.

From the ANOVA table fitted to the experimental data, we get values of MS Blocks and MSE.

Can you compute the relative efficiency RE?

Exercise (solution)

Nitrogen content of wheat example: $r = 4, t = 6$.

From the ANOVA table, we get MS Blocks $= 65.67$ and $MSE = 7.2$.

Compute the error variance estimates and RE:

- $s^2_{rcb} = MSE = 7.2$,
- $s^2_{cr} = \frac{(r-1)\text{MS Blocks}+r(t-1)\text{MSE}}{(rt-1)} = \frac{(4-1)\cdot 65.67+4\cdot 5\cdot 7.2}{(4\cdot 6-1)} = 14.83.$
- $RE = \frac{s^2_{cr}}{s^2_{rcb}} = 2.06.$

(Or, $H = \text{MS Blocks/MSE} = 65.67/7.20 = 9.121$, $k = r(t-1)/(rt-1) = 4 \cdot 5/23 = 0.870$, $\text{RE} = k + (1-k) \cdot H = 2.06$)

That is, we'd need twice as many replicates to achieve the same power with a completely randomized design that we achieved with this randomized complete block design.

With the values of MS Blocks and MSE, we can compute the error variance estimates for a RCB design and for a completely randomized design.

The value of RE is the ratio of the two and equals 2.06.

Note that the experiment with lower estimated error variance is more efficient.

The relative efficiency of 2.06 says that we'd need twice as many replicates to achieve the same power with a completely randomized design that we achieved with this randomized complete block design.

The cost of blocking

- error d.f. in the completely randomized design: $t(r-1)$
- error d.f. in the randomized complete block design: $(t-1)(r-1)$
- $\text{d.f. lost} = t(r-1) - (t-1)(r-1) = r-1.$

This is usually not too worrisome, except in very small experiments, in which the lost degrees of freedom are a big proportion of the total error degrees of freedom. (Recall that Hurlbert suggests that blocking is a good idea in small experiments, where the chances of imbalances in confounding factors across treatments are relatively high.)

As usual, the loss of error degrees of freedom is balanced against the reduction in SSE that is accomplished by blocking. (The MSE may increase or decrease.)

With blocking we lose (r-1) degrees of freedom for error estimation, since we need to estimation the block means.

This is usually not too worrisome, except in very small experiments, in which the lost degrees of freedom are a big proportion of the total error degrees of freedom. (Recall that Hurlbert suggests that blocking is a good idea in small experiments, where the chances of imbalances in confounding factors across treatments are relatively high.)

As usual, the loss of error degrees of freedom is balanced against the reduction in SSE that is accomplished by blocking. (The MSE may increase or decrease.)

Computing using R

In example8.1.html, there is a section on computing RE.

In example8.1.html, there is a section on computing RE.

Summary

Variations of RCBD design

- RCBD with replication
- RCBD with subsamples
- RCBD with random blocks

Extension of RCBD design

- Factorial experiments can be replicated in blocks

Relative efficiency

- Blocking is most effective if experimental units within a block are more homogeneous than experimental units from different blocks

In this lecture, we discussed a few variations of RCBD design such as RCBD with replication, RCBD with subsamples, and RCBD with random blocks.

Factorial experiments can be replicated in blocks too.

Blocking is most effective if experimental units within a block are more homogeneous than experimental units from different blocks

I am not afraid to say that RCBD type designs are my favorite: as mentioned in the previous lecture, by blocking, it increase precision for treatment comparison; by replicating over blocks, it increase generalizability.