UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Design and Analysis of Experiments

11 - Blocking

Version 2.11

Felipe Campelo
http://orcslab.cpdee.ufmg.br/

Graduate Program in Electrical Engineering

Belo Horizonte
May 2015

CC BY NC SA

"*Science doesn't always go forwards. It's a bit like doing a Rubik's cube. You sometimes have to make more of a mess with a Rubik's cube before you can get it to go right.*"

Jocelyn Bell Burnell
1943 -
Irish astrophysicist

Image: http://www.topbritishinnovations.org/PastInnovations/DiscoveryofPulsars.aspx

Blocking

Nuisance factors

In the completely randomized design (CRD) introduced in the last chapter, the observations are assigned - as the name implies -in a completely random manner.

Implicitly, we assume that the experimental conditions are homogeneous, i.e., that no other important effects are present besides those of the experimental factor.

In some situations, however, the experiment may present other factors that are not of immediate interest, but that can influence the response variable. This was already discussed when we introduced the concept of *pairing* in Chapter 7.

Blocking

Nuisance factors

The generalization of the *paired design* for an arbitrary number of factor levels is called *blocking*, which is an elegant way configure the data collection in order to enable the modeling and exclusion of the effects of nuisance factors.

In fact, it can be argued that the systematic use of blocking is beneficial even in the absence of known nuisance factors, as it helps boosting statistical power by excluding the between-replicates variability from the residual term.

> "*Block what you can, randomize what you cannot.*"
>
> George E.P. Box (1919–2013)
> British Statistician

Image: `https://www.engr.wisc.edu/eday/eday1990.html`

Blocking

Blocking *versus* Randomizing

The idea behind randomization is trying to prevent unknown factors to bias our observations;

Blocking comes into play whenever we know from the beginning that certain factors can influence our response variable, but for some reason we are not interested in their effects[a], for instance:

- the effect of different batches of raw material when comparing the performance of chemical reactions;
- the effect of different benchmark problems when comparing the performance of computer algorithms.

[a] If we're interested in their effects, they are no longer nuisance factors, but additional experimental factors. We'll deal with factorial designs in the next chapter.

Example: algorithm comparison

Problem definition

A Ph.D. student decides to compare a standard optimization algorithm with six modified versions for the solution of a certain family of Vehicle Routing Problems (VRPs). His intention is to verify whether any of the modified versions is systematically better than the standard one.

The algorithms are applied for the solution of 180 problem instances, divided in 36 groups of five homogeneous instances. The algorithms are run 30 times on each instance, and the cost value found after each run is recorded.

The student wants an experiment with $\alpha = 0.05$ and power $\pi = 1 - \beta = 0.8$ for a minimally interesting effect $\delta^* = 50$.

Image: PhD Comics by Jorge Cham:

http://www.phdcomics.com/comics/archive.php?comicid=1139

Example: algorithm comparison

Nuisance factor

Besides the effect of the experimental factor (*Algorithm*), we have a possible nuisance effect due to the variability between experimental units (instance groups).

If we employ a CRD, the residual term would contain all the between-instances variability, which would almost certainly mask any between-algorithms effect of interest.

Since we can control the allocation of algorithm runs within each instance, one possibility is to do it in a systematic way so that we can model the instance effects (in addition to the factor effect).

Example: algorithm comparison

Randomized complete block design

Within each *block*, the order of execution should in principle be randomized, so as to prevent other unknown factors from affecting our analysis[b].

This experimental setup is known as a *randomized complete block design* (RCBD) with one blocking factor.

[b] For this kind of computer experiment the within-block randomization is unnecessary. It is very important, however, when dealing with systems that are less controllable.

Randomized complete block design

Assumptions

The RCBD assumes:

- *one replicate per block*[c];
- independent blocks;
- independent within-block randomization.

Failure to account for dependency structures (i.e., using blocks that are not independent) can result in pseudoreplication, i.e., the use of the incorrect number of degrees-of-freedom in the test statistics, which usually results in an inflation of the type-I error rate.

[c] For multiple within-block replicates the appropriate design is known as *Generalized randomized block design* (GRBD), which can be analyzed using a different model.

Example: algorithm comparison

RCBD

To comply with these requirements, we consider as our response variable the average performance of an algorithm on a particular *instance group*.

The performance of each algorithm will be averaged within each individual instance (mean of 30 runs), and then within each group (mean of 5 instances), so that each algorithm has 36 independent observations.

Example: algorithm comparison

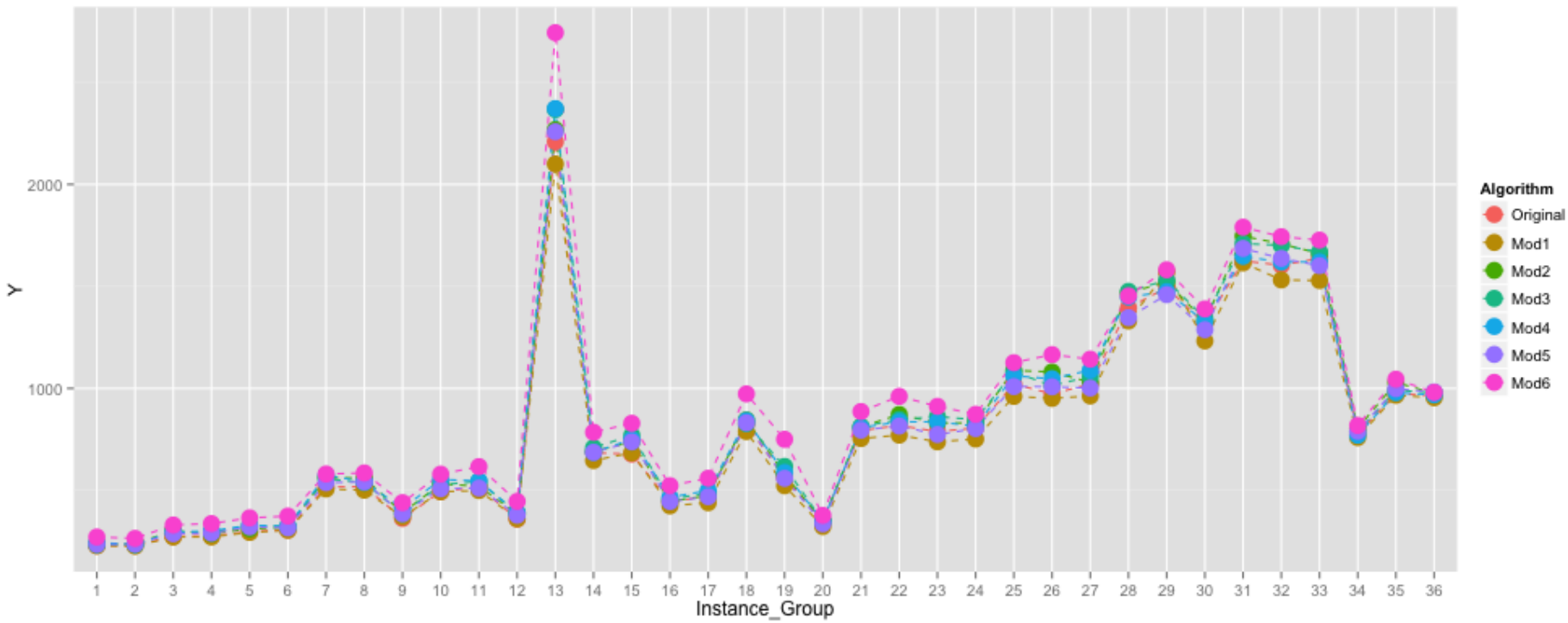
Data loading and preprocessing

```
# Load data
> data <- read.table("../data files/algo.csv", header = TRUE)
# Aggregate data (algorithm means by instance group)
> aggdata <- with(data, aggregate(x   = Result,
                                  by  = list(Algorithm, Group),
                                  FUN = mean))
# Rename columns and coerce factor variables
> names(aggdata) <- c("Algorithm", "Instance_Group", "Y")
> for (i in 1:2) aggdata[, i] <- as.factor(aggdata[, i])

> levels(aggdata$Algorithm) <- c("Original",
                                 unlist(lapply(X   = "Mod",
                                               FUN = paste0, 1:6)))
> summary(aggdata)
    Algorithm  Instance_Group       Y
 Original:36   1      :  7    Min.   : 227.3
 Mod1    :36   2      :  7    1st Qu.: 440.0
 Mod2    :36   3      :  7    Median : 756.8
 Mod3    :36   4      :  7    Mean   : 821.0
 Mod4    :36   5      :  7    3rd Qu.:1020.5
 Mod5    :36   6      :  7    Max.   :2743.6
 Mod6    :36   (Other):210$
```

Example: algorithm comparison

Exploratory data analysis

```
library(ggplot2)
> p <- ggplot(aggdata, aes(x = Instance_Group, y = Y,
                           group = Algorithm, colour = Algorithm))

> p + geom_line(linetype=2) + geom_point(size=5)
```



Analysis of the RCBD

Statistical model

In the general case, we have a levels of the experimental factor, and b levels for the blocking factor. The statistical model is given as:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases}, \text{ assuming } \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right)$$

Similarly to the CRD we have, by construction:

$$\sum_{i=1}^{a} \tau_i = 0 \qquad \sum_{j=1}^{b} \beta_j = 0$$

Analysis of the RCBD

Statistical model

In the RCBD, we are interested only on the effect of the experimental factor. Consequently, the test hypotheses refer only to its coefficients:

$$\begin{cases} H_0 : \tau_i = 0, \quad \forall i = 1, \cdots, a \\ H_1 : \exists\, \tau_i \neq 0 \end{cases}$$

The partition of the total sample variability is given by:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} \left(y_{ij} - \bar{y}_{..}\right)^2$$

$$= \underbrace{b\sum_{i=1}^{a} \left(\bar{y}_{i.} - \bar{y}_{..}\right)^2}_{SS_{\text{levels}}} + \underbrace{a\sum_{j=1}^{b} \left(\bar{y}_{.j} - \bar{y}_{..}\right)^2}_{SS_{\text{blocks}}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b} \left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2}_{SS_E}$$

Analysis of the RCBD

Statistical model

The mean squares are given by:

$$MS_{\text{levels}} = \frac{SS_{\text{levels}}}{a-1} \qquad MS_{\text{blocks}} = \frac{SS_{\text{blocks}}}{b-1} \qquad MS_E = \frac{SS_E}{(a-1)(b-1)}$$

and their expected values are:

$$E\left[MS_{\text{levels}}\right] = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1} \qquad E\left[MS_{\text{blocks}}\right] = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E\left[MS_E\right] = \sigma^2$$

Analysis of the RCBD

Statistical model

$$E\left[MS_{\text{levels}}\right] = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1} \qquad E\left[MS_E\right] = \sigma^2$$

Under the null hypotheses we have that the test statistic

$$F_0 = \frac{MS_{\text{levels}}}{MS_E}$$

is a distributed according to an F distribution with $(a-1)$ numerator degrees-of-freedom and $(a-1)(b-1)$ denominator DoFs.

The critical region for the test is therefore given by

$$f_0 > F_{\alpha;(a-1),(a-1)(b-1)}$$

Analysis of the RCBD

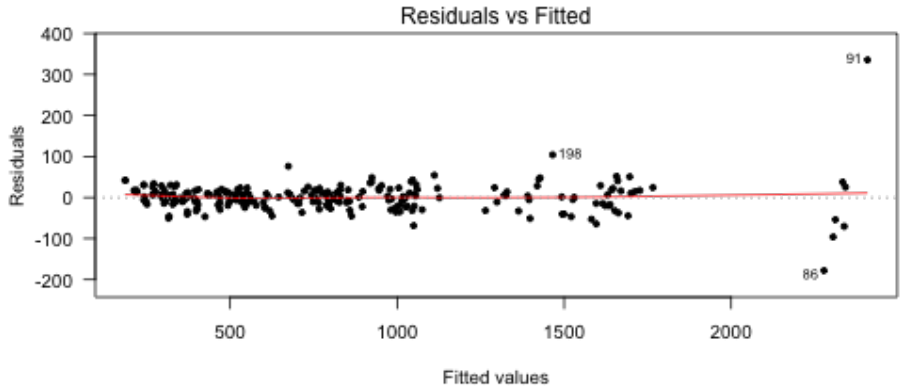
Statistical model

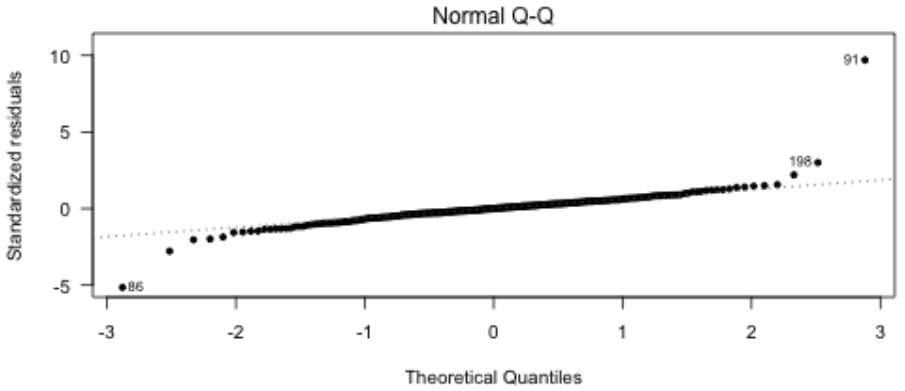
```
# First model
> model <- aov(Y~Algorithm+Instance_Group, data=aggdata)

> summary(model)
                Df   Sum Sq Mean Sq F value Pr(>F)
Algorithm        6   359949   59991   41.75 <2e-16 ***
Instance_Group  35 60438639 1726818 1201.86 <2e-16 ***
Residuals      210   301725    1437
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary.lm(model)$r.squared
[1] 0.9950618
```

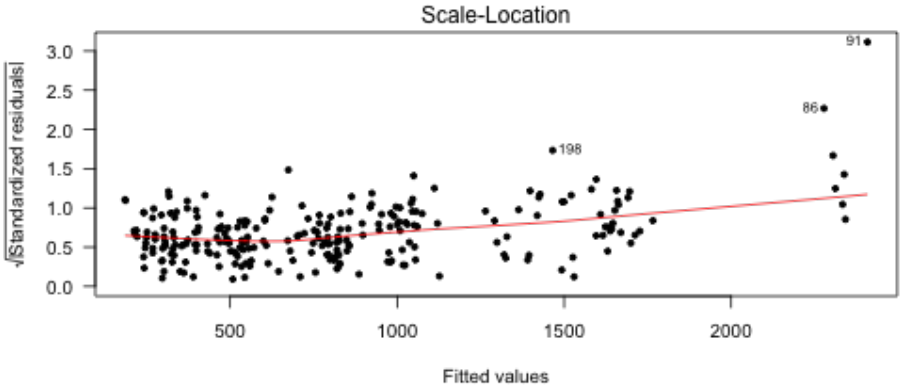
Analysis of the RCBD

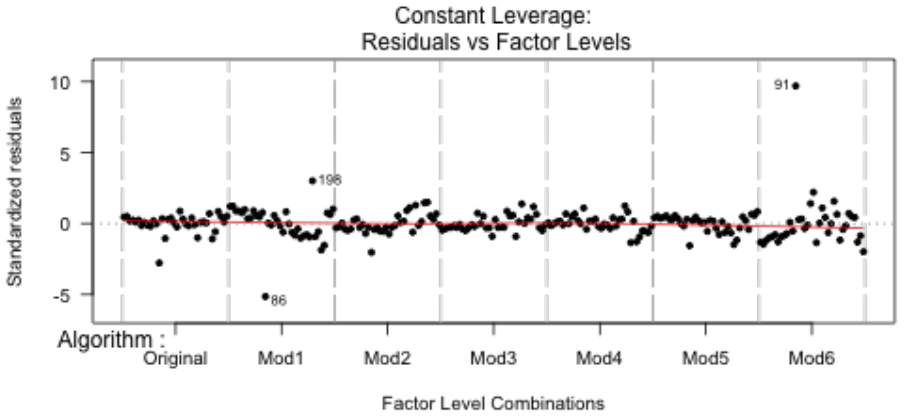
Statistical model - residuals

```
> par(mfrow = c(2, 2))
> plot(model, pch = 20, las = 1)
```









Analysis of the RCBD

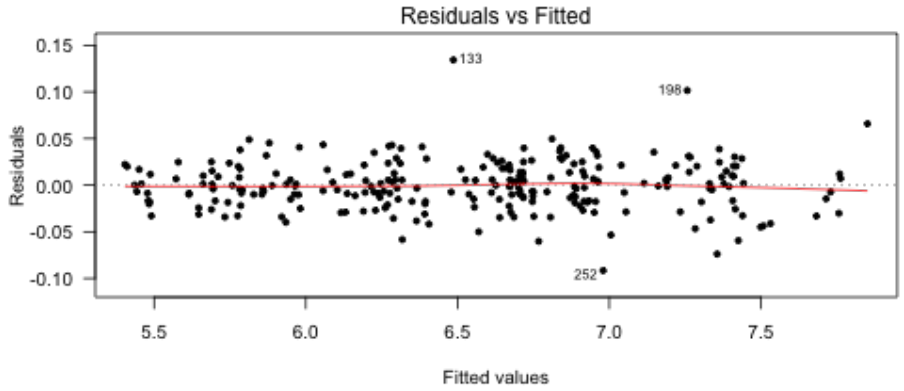
Statistical model - data transformation

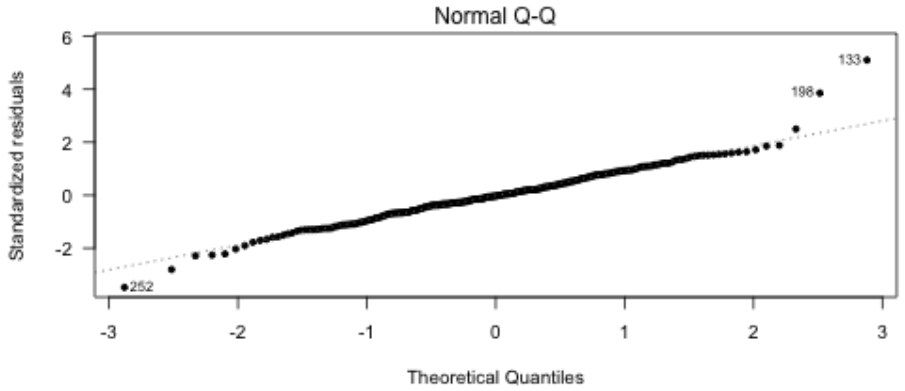
```
# Trying with log-transformed response variable
> model2 <- aov(log(Y)~Algorithm+Instance_Group, data=aggdata)

> summary(model2)
                Df Sum Sq Mean Sq F value Pr(>F)
Algorithm        6   0.60  0.1002   120.4 <2e-16 ***
Instance_Group  35  89.92  2.5690  3089.0 <2e-16 ***
Residuals      210   0.17  0.0008
---

> summary.lm(model2)$r.squared
[1] 0.9980742
```

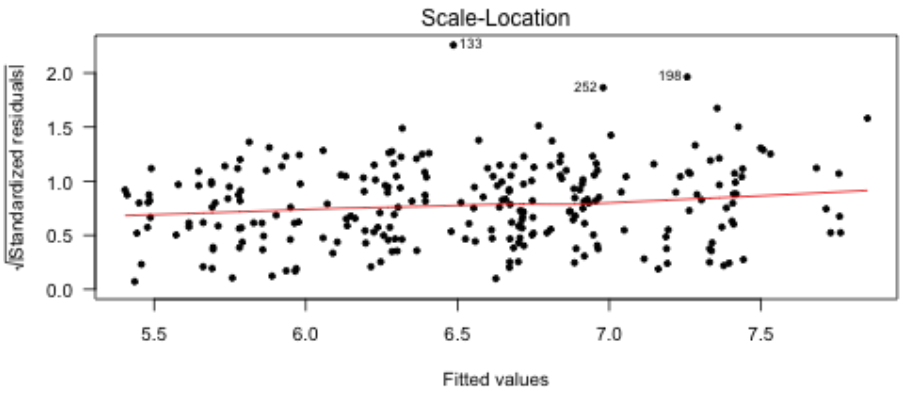
Analysis of the RCBD

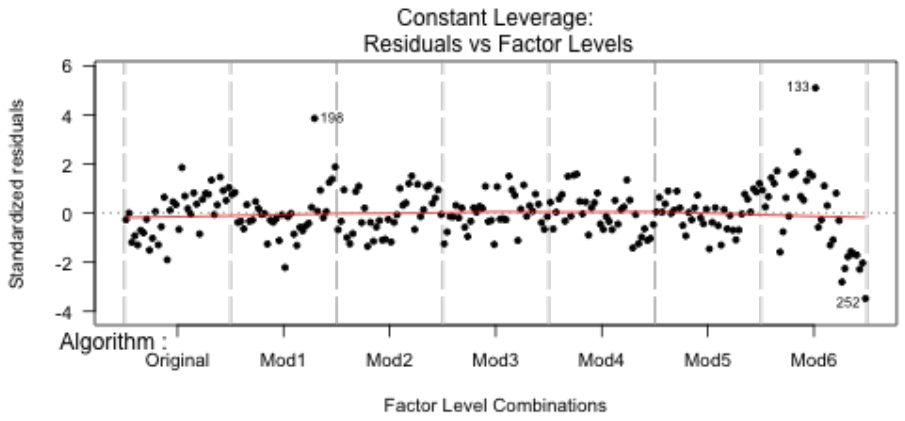
Statistical model - residuals

```
> par(mfrow = c(2, 2))
> plot(model2, pch = 20, las = 1)
```









Analysis of the RCBD

Multiple blocking factors

If there is more than one factor that need to be blocked out of the analysis, it is possible to simply add new terms to the statistical model. The sum of squares for this new term is similar to the ones already shown for the single block case;

Analysis of the RCBD

Blocking efficiency

After an experiment is performed and the ANOVA model is fit to the data, it is possible to calculate the *relative blocking efficiency* (E), which quantifies how much larger a CRD would have to be in order to attain the same power as the corresponding RCBD.

The relative efficiency can be calculated from the mean squares returned in the ANOVA table of the RCBD as:

$$E = \frac{(b-1)MS_{\text{Blocks}} + b(a-1)MS_{\text{E}}}{(ab-1)MS_{\text{E}}}$$

A value of 1.3, for instance, would indicate that the CRD would have required 30% more observations to achieve the same power. For the running example, this value can be calculated from the summary table: $E = 431.6$

Analysis of the RCBD

Sample size calculations

The determination of the required sample size for an RCBD follows the same procedures as the CRD: the number of replicates calculated using the CRD formulas represent the number of blocks required (since we have one replicate per block).

The within-group variability term necessary for calculating the required sample size corresponds to the estimated residual variance, i.e., the unexplained variability after accounting for the experimental and blocking factors, which can be somewhat challenging to estimate without a information on previous studies (or a pilot experiment).

Analysis of the RCBD

Multiple comparisons

For the post-hoc multiple comparison procedures, we also follow the same general guidelines used for the CRD, with some minor differences.

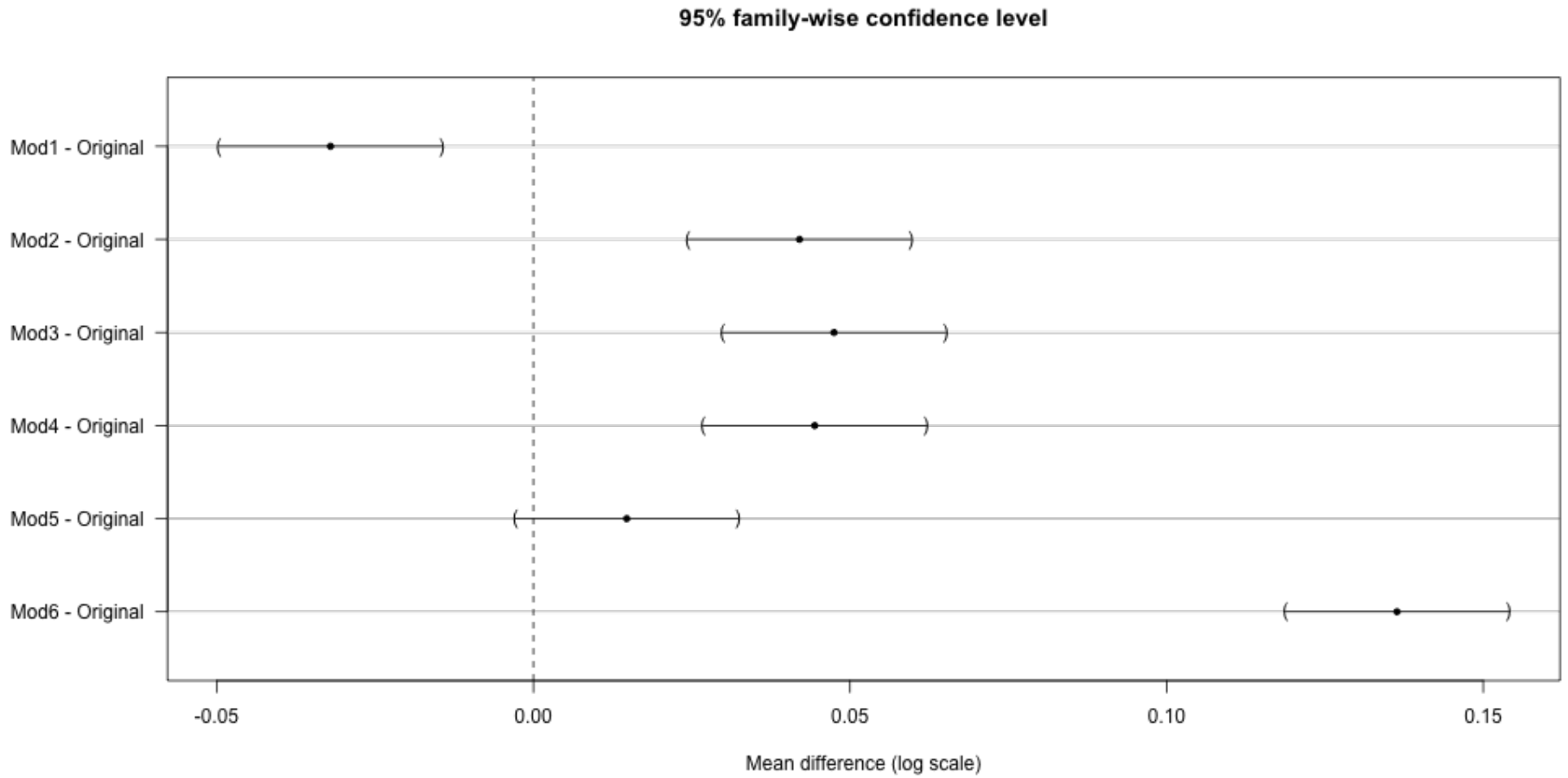
Paired t-tests are used for the pairwise comparisons, with $df_E = (a-1)(b-1)$ as the number of degrees of freedom used for the reference distributions;

As with the CRD, it may be convenient to use the post-hoc pairwise comparison procedures as the basis for sample size calculations, particularly in cases where the cost of observations is relatively low.

Analysis of the RCBD

Multiple comparisons

```
> library(multcomp)
> duntest     <- glht(model2, linfct = mcp(Algorithm = "Dunnett"))
> duntestCI   <- confint(duntest)
> par(mar = c(5, 8, 4, 2), las = 1)
> plot(duntestCI, xlab = "Mean difference (log scale)")
```



Randomized complete block designs

Final considerations

The RCBD is a generalization of the pairing concept used with t-tests, and provides an elegant way to remove nuisance effects from the analysis;

The consequence of using a blocking design is an increased sensitivity of the test, which means smaller sample sizes.

In practice, running an experiment as a RCBD results in little experimental overhead, and should be preferred over the CRD whenever possible.

The design and analysis of incomplete block designs (e.g., with missing observations) are also relatively simple[e].

[e] See Montgomery (2005) for details.

Bibliography

Required reading

1. D.C. Montgomery, *Design and Analysis of Experiments*, Ch. 4-5, 5th ed., Wiley, 2005;
2. M. Christman, R. Littell, *STA166 - Statistical Methods Research I: RCBD ANOVA Notes II* - `http://goo.gl/KrpbKL`

Recommended reading

1. R. Dawkins, *Unweaving the Rainbow: Science, Delusion and the Appetite for Wonder*, Mariner Books, 2000.
2. J. Morgan and C. Albon, *Partially Derivative* (podcast) - `http://www.partiallyderivative.com`

About this material

Conditions of use and referencing

This work is licensed under the Creative Commons CC BY-NC-SA 4.0 license (Attribution Non-Commercial Share Alike International License version 4.0).

http://creativecommons.org/licenses/by-nc-sa/4.0/

Please reference this work as:

Felipe Campelo (2015), *Lecture Notes on Design and Analysis of Experiments*.
Online: https://github.com/fcampelo/Design-and-Analysis-of-Experiments
Version 2.11, Chapter 11; Creative Commons BY-NC-SA 4.0.

```
@Misc{Campelo2015-01,
  title={Lecture Notes on Design and Analysis of Experiments},
  author={Felipe Campelo},
  howPublished={\url{https://github.com/fcampelo/Design-and-Analysis-of-Experiments}},
  year={2015},
  note={Version 2.11, Chapter 11; Creative Commons BY-NC-SA 4.0.},
}
```