

Experiment Design for Computer Sciences (0AL0400)

Topic 07 - Experimental Factors

Claus Aranha

caranha@cs.tsukuba.ac.jp

University of Tsukuba, Department of Computer Sciences

Version 2022.1 (Updated June 17, 2022)

In this material, we study simple strategies for handling experiments with **multiple factors**.

- One Variable at a Time Design (OVAT)
- Blocking Design
- Factorial Design

Part I – One Variable At a Time (OVAT)

Factors and Levels

- **Factors:** Input variables in the experiment / "control variables";
- **Levels:** The values that a factor can assume / "treatments";

Example: Carrying Speed Experiment

I want to measure how the weight of my luggage affects my speed. I measure how far I walk in 5 minutes, when carrying 10, 20 and 40kgs of luggage.

- Response Variable: distance walked in 5 minutes;
- Factors: Weight of luggage;
- Levels: 10kgs, 20kgs, 40kgs;

Experiments with One Factor

Until now, we considered experiments with 1 factor and 1 return variable;

- **t-test**: 1 factor, two levels
- **ANOVA**: 1 factor, many levels

Example

We want to compare how long it takes to train a neural network until a certain error value is reached. We compare four optimization algorithms: SGD, Adam, Adagrad, CMA-ES.

- Response Variable: Time until error threshold is reached;
- Factor: Optimization Algorithm;
- Levels: 1- SGD, 2- Adam, 3- Adagrad, 4- CMA-ES

Experiment with Two Factors

We add a new factor to the experiment: The number of layers in the network (2, 5 or 10)

- Response Variable: Time until error threshold is reached;
 - Factor 1, 4 levels: Optimization Algorithm: SGD, Adam, Adagrad, CMA-ES;
 - Factor 2, 3 levels: 2, 5, 10 layers;
-
- **Experiment with 1 factor:** Only consider the effect of the levels on the response variable.
 - **Experiment with 2 factors:** Consider main effect and interference effects:
 - **Main Effect:** Effect of the levels of each factor in the response variable.
 - **Interaction Effects:** Combined effects of levels of both factors in the response variable.

Number of Factors and Experiment Complexity

The higher the number of factors and levels, the more complex becomes the experiment.

- 2 factors, 4 levels each: 16 total combinations; 2nd order interactions;
- 3 factors, 3 levels each: 27 total combinations; 3rd order interactions;
- 4 factors, 3 levels each: 81 total combinations; 4th order interactions;
- ...

Because of this, it is desirable to keep experiments simple, even if there are statistical techniques to detect interaction effects.

OVAT – One Variable at a Time

The simplest design to use with multiple factors is the OVAT Design.

- For N factors, perform N different experiments, one for each factor.
- The levels of the other factors are fixed at a "standard" value.
- For each factor, you perform the statistical analysis independently (t-test, anova, visualizations, etc);

OVAT – Example

Fine tuning Cr and F for Differential Evolution (DE)

DE is an effective optimization algorithm where the performance depends strongly on two parameters: Cr and F.

To find the optimal values for these parameters, you perform a preliminary experiment on a synthetic benchmark.

OVAT Design:

- Choose standard values for Cr and F from the literature.
- Analyze Cr using an experiment. Fix the value of F and try 10 different values (levels) for CR.
- Analyze F using another experiment. Fix the value of Cr and try 10 different values (levels) for F.
- The execution and analysis of these two experiments follow the methods that we already studied.

Part II - Blocking Design

Completely Randomized Design (CRD)

Until now, we studied the following situation:

- One input factor (for example, algorithm type)
- One response variable (for example, solution speed)
- An **homogeneous experimental condition**.

Homogeneous experimental condition

This means that the experiment environment is guaranteed to be the same for all observations, and there are no significant error sources.

In this situation, we execute the experiment in **random order** to guarantee that any error sources are equally distributed. We call this a **Completely Randomized Design (CRD)**

Block Design

Noise / Nuisance Factors are input variables that can have a significant impact in the experiment output, but they are not relevant to the scientific question.

In a previous lecture, we discussed the use of **Pairing** to remove the effect of *one* noise factor. **Blocking** is the generalization of the Paired Design approach for multiple noise factors.

Block Design vs Completely Randomized Design

We use Randomization (CRD) to prevent the effect of *unknown* noise factors in our experiment.

We use Blocking or Pairing when we know from the beginning that certain factors affect the output variable, but for some reason we are not interested in their effects.

Example of Block Design

Algorithm Comparison

A student decides to compare a standard optimization algorithm against size variants to solve a certain family of Vehicle Routing Problems (VRPs).

He wants to know if any variant returns a systematically lower cost value, when applied to 180 problem instances. These instances can be divided in **36 groups**, with 5 similar instance in each group.

Experiment Design Variables

- confidence: $\alpha = 0.05$
- power: $\pi = 1 - \beta = 0.8$
- minimally interesting effect: $\delta^* = 50$



Example of Block Design: Factors and levels

For this problem, we can identify the following variables:

- **Input Factor:** Algorithm; **Levels:** Original and 6 variants;
- **Return Variable:** Cost; **Levels:** Integer value;
- **Noise Factor:** Instance type; **Levels:** 36 groups;

If we employ a CRD (ignoring or randomizing the Noise Factor), the difference in result among Instance types would have an effect in the residual, and possibly mask the difference between algorithms.

To avoid this, we organize the experiment in 36 blocks (one block per level of the Noise Factor), and randomize the execution of all 7 algorithms inside each block. This is the **Randomized Complete Block Design (RCBD)**

Assumptions of the RCBD

The RCBD assumes the following characteristics for the experiment:

- 1 one replicate per block;
- 2 the blocks are independent;
- 3 randomization inside the block is independent;

These assumptions are important. If we fail to guarantee the independence between the blocks (for example, some of the blocks are dependent), we introduce something called **pseudoreplication**, which may result in an inflation of type-I error.

Example: Pre-processing the data

In the algorithm example, we obtain the following experimental data:

- 7 algorithm variants,
- 180 total problem instances, 36 instance group (5 instances per group)
- 30 repetitions, per algorithm, per instance ($7 \times 30 \times 180 = 37800$ observations)

However, as explained in the last slide, the RCBD requires **1 replicate per block**. To satisfy this requirement, we pre-process the data as follows:

- The performance of each algorithm, per instance, is averaged (average of 30 repetitions)
- The performance of each algorithm, per group, is averaged (average of 5 instances)

This gives us 36 observations per algorithm (1 observation per block)

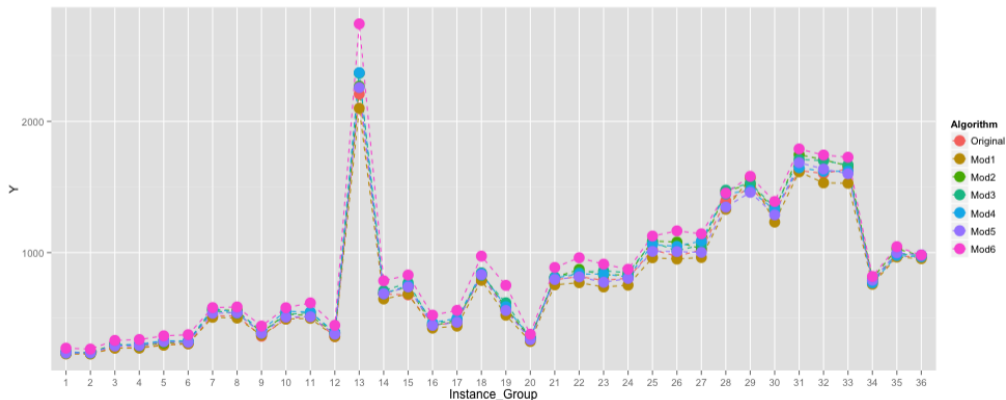
Example: Pre-processing the data (R code)

```
> 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: Pre-visualization of the Data (R code)

```
> p <- ggplot(aggdata, aes(x = Instance_Group, y = Y,  
                           group = Algorithm, colour = Algorithm))  
> p + geom_line(linetype=2) + geom_point(size=5)
```



Statistical Analysis of the CRBD

The Statistical analysis of the CRBD is similar to the ANOVA analysis we studied in a previous class. For a derivation of the model, see Campelo in the recommended readings.

However, note that the number of observations is based on the number of blocks and treatment levels.

```
# Anova Model: Treatments+Blocks
> 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
```

Statistical Analysis of the CRBD – Residuals

```
> summary.lm(model)$r.squared
```

```
[1] 0.9950618
```

```
> par(mfrow = c(2, 2))
```

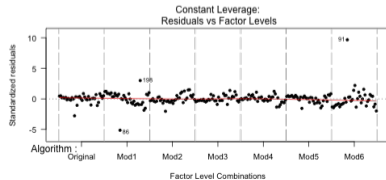
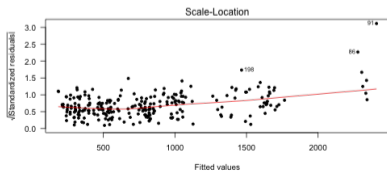
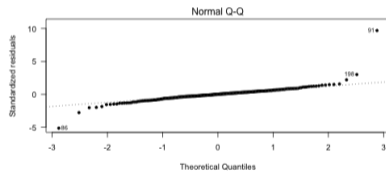
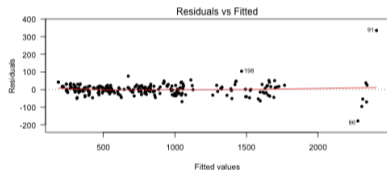
```
> plot(model, pch = 20, las = 1)
```

```
# It's important to check the residuals
```

```
# for anomalies. For example, here
```

```
# The residual plot shows several outliers
```

```
# that could affect the result.
```



Statistical Analysis of the CRBD – Log Transformation

An analysis of the residual model showed some irregularities in the residual plot, so we try a log transformation on the response variable to see if we can smooth it out.

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

> summary(model2)
```

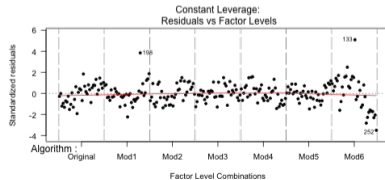
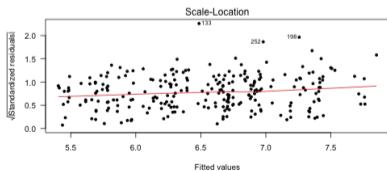
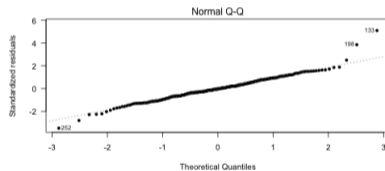
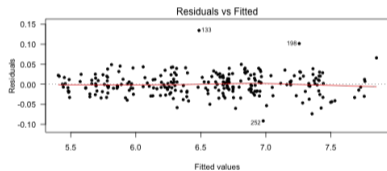
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Algorithm	6	0.60	0.1002	120.4	<2e-16	***
Instance_Group	35	89.92	2.5690	3089.0	<2e-16	***
Residuals	210	0.17	0.0008			

Statistical Analysis of the CRBD – Log Transformation

```

> summary.lm(model2)$r.squared      # The log transformation smoothed out
[1] 0.9980742                      # the outliers from our CRBD model.
> par(mfrow = c(2, 2))            #
> plot(model2, pch = 20, las = 1)  #

```



Statistical Analysis of the CRBD – final comments

The post-hoc analysis of the CRBD follows the same principles of the post-hoc analysis of the ANOVA.

- The type and number of comparisons should be defined a priori.
 - In this example, two-sample comparisons between the original algorithm and each variant would be appropriate.
- The number of observations should be based on the number of blocks.
- Alpha-correction for the paired observations should be applied as necessary.

Further Reading

You should do further reading for these two models that are closely related to the CRBD:

- "Incomplete Block Design" – When some of the observations (blocks) are missing;
- "Generalized Block Design" – When each block has multiple replicates;

Part III - Factorial Design

Factorial Designs

Many experiments involve more than one factor of interest. Sometimes we want to control **multiple independent variables** that influence the response of the experiment.

One effective way to explore the main effects and interactions of multiple factors is the **Factorial Design**. In a factorial design, all level combinations are explored.

Main Effect and Interaction

- **Main Effect** of a factor: The mean change of the response variable when we change the level of a factor;
- **Interaction** between factors: The mean change of the response variable when we change the level of two or more factors at the same time.

Factorial Design Example



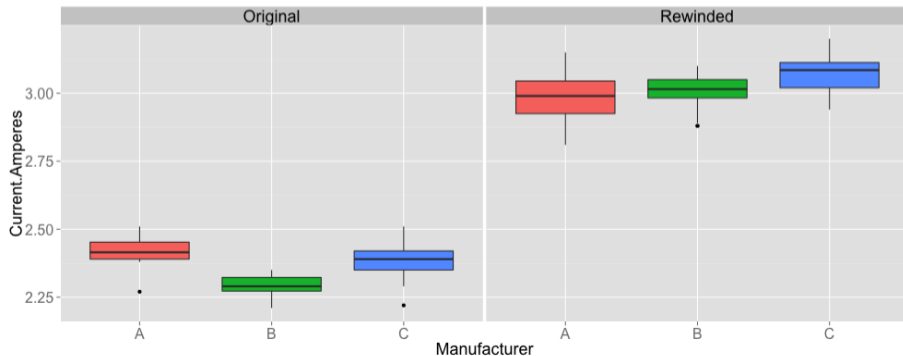
An engineer wants to investigate factors that affect the electrical current demanded by a motor used in the ventilation of a chicken coop. She identifies two factors to investigate regarding the current demand of a motor:

- **Factor 1:** The manufacturer of the motor (levels: A, B, C)
- **Factor 2:** The state of the motor (levels: original, rewinded)

To investigate this question, the engineers sample a 40 motors from each manufacturer, 20 in the original state, and 20 being rewinded motors. The current draw from each motor is registered. (See data file "motors.txt")

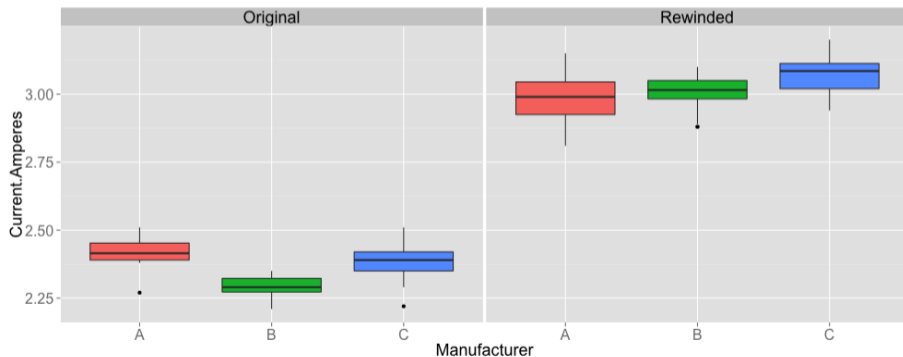
Example: Exploratory Data Analysis

```
> data <- read.table("../data files/motors.txt", header = TRUE)
> library(ggplot2)
> p <- ggplot(data, aes(x = Manufacturer, y = Current.Amperes,
                        fill = Manufacturer))
> p + geom_boxplot() + facet_grid(.~State) + ...
```



Example: Exploratory Data Analysis

The exploratory plot suggests a large main effect for the *State* factor, and is inconclusive for the effect of the *Manufacturer* factor. It is unclear if there are interaction effects or not.



Factorial Design Model

In the general case for a completely randomized factorial design we have:

- a levels for factor **A**;
- b levels for the factor **B**;
- n replicates within each combination of levels;
- Completely randomized collection of observations;

The **additive effects model** for a set of observations collected following this design can be expressed as follows. From this model, we can construct null/alternate hypotheses in a similar fashion as in ANOVA.

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1, \dots, a \\ j = 1, \dots, b \\ k = 1, \dots, n \end{cases}$$

Statistical Model for Two Factors

The ANOVA gives us the linear model for the mean effect of each factor (State, Manufacturer), as well as the interaction effect (State x Manufacturer).

By analysing these effects, we can define our post-hoc analysis.

```
> model <- aov(Current.Amperes~State*Manufacturer,
+              data = data)
> summary(model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
State	1	12.956	12.956	2798.41	< 2e-16	***
Manufacturer	2	0.118	0.059	12.71	1.04e-05	***
State:Manufacturer	2	0.114	0.057	12.27	1.49e-05	***
Residuals	114	0.528	0.005			

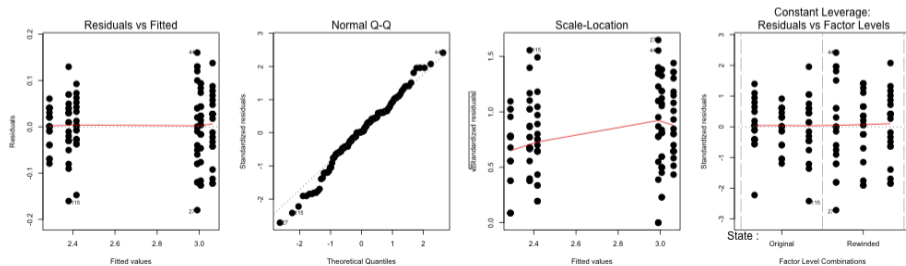
```
---
> summary.lm(model)$r.squared
[1] 0.9615174
```

Examining the Data Assumptions

As usual, the assumptions can be verified by means of residual analysis, like in the one-way ANOVA (except for a little adjustment needed for the Fligner-Killeen test)

```
> shapiro.test(model$residuals)
W = 0.9857, p-value = 0.2392
```

```
> fligner.test(Current.Amperes ~ interaction(State, Manufacturer),
+             data = data)
med chi-squared = 10.1721, df = 5, p-value = 0.0705
```



Factorial Design – Post-hoc comparison

If the ANOVA indicates the existence of significant effects, we can perform pairwise comparisons between levels to investigate specific differences; When the interaction

effect is not significant, the comparisons between factor levels can be done in a straightforward manner, using the estimated level means. For instance, the test statistic for comparing the means of levels 2 and 3 of factor A could be calculated as:

$$t_0 = \frac{\bar{y}_{2..} - \bar{y}_{3..}}{\sqrt{2 \frac{MS_E}{n'}}$$

where n' is the number of specific replicates for the comparison under consideration.

Factorial designs – Post-hoc comparison

More generally,

$$t_0 = \frac{\Delta \bar{y}}{\sqrt{2 \frac{MS_E}{n'}}$$

For comparisons of factor levels (main effects), the value of n' is the total number of observations under that level;

For comparisons of level combinations (interaction effects), it is the number of observations within each combination group;

```
> replications(Current ~ State*Manufacturer,
+             data = data)
              State      Manufacturer State:Manufacturer
              60              40              20
```

Also, the α value for the comparisons has to be adjusted to prevent inflation of the type-I error rate.

Recommended Reading

Block Design

- D.C. Montgomery, "Design and Analysis of Experiments", Ch4-5
- Felipe Campelo, "Design and Analysis of Experiments", Lecture 11

Factorial Design

- Felipe Campelo, "Design and Analysis of Experiments", Chapter 12
- Peter Hoff, "Design of Experiment Lecture Notes", Chapter 6

See links in manaba!

About these Slides

These slides were made by Claus Aranha, 2022. You are welcome to copy, re-use and modify this material.

These slides are a modification of "Design and Analysis of Experiments (2018)" by Felipe Campelo, used with permission.

Individual images in some slides might have been made by other authors. Please see the following references for those cases.

Image Credits I

[Page 13] PhD Student Image: PhD Comics by Jorge Cham,
<http://www.phdcomics.com/comics/archive.php?comid=1139>

[Page 27] Ventilator figure from <http://refrigelms.com.br>