UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Design and Analysis of Experiments

12 - Factorial Designs

Version 2.11

Felipe Campelo
http://www.cpdee.ufmg.br/~fcampelo

Graduate Program in Electrical Engineering

Belo Horizonte
May 2015

"*We did not evolve to understand or comprehend reality. We evolved to survive it. For understanding we need science.*"

Mark A. Crislip
1952-
American infectologist

Image: `http://boards.medscape.com/.29f3af03/`

Factorial Designs

Basic definitions

Many experiments involve more than a single *factor* of interest - that is, multiple independent variables that can influence a response variable.

In general, an effective way to explore the main effects and interactions of multiple factors is the use of a *factorial design* in which all level combinations are evaluated at each experimental replicate;

In this context, the **main effect** of a factor quantifies the mean change in the response variable due to changing between the levels of that factor;

An **interaction effect** represents the mean change in the response variable due to the simultaneous change of levels of two or more factors.

Factorial Designs

Example: Electrical current in motors

Two engineers wish to investigate factors that may affect the electrical current demanded by the single-phase motors used for ventilation in an industrial chicken coop.

Previous observations suggest that the current varies considerably from motor to motor, and process knowledge suggests two likely candidates for explaining this variability: the *Manufacturer* (A, B or C) and the *State* (original or rewinded) of each motor.

To investigate this question, the engineers decide to sample 40 motors from each manufacturer, with 20 in the original state and 20 being rewinded motors.

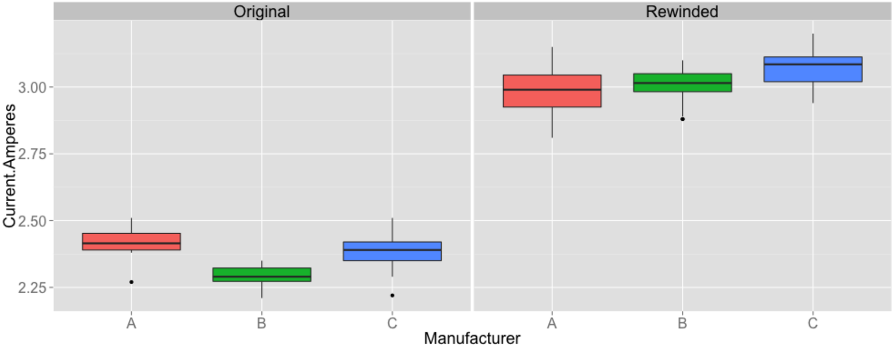
Adapted from M.H.Costa and T.L. Vieira's course project for the Design and Analysis of Experiments Course, PPGEE-UFMG, November 2013. The data used in this example is not necessarily the original one.

Image: `http://refrigelms.com.br/ventilador-para-aviario-qla85-grade-p-1734.html`

Exploratory data analysis

Example: Electrical current in motors

```
> 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) + ...
```

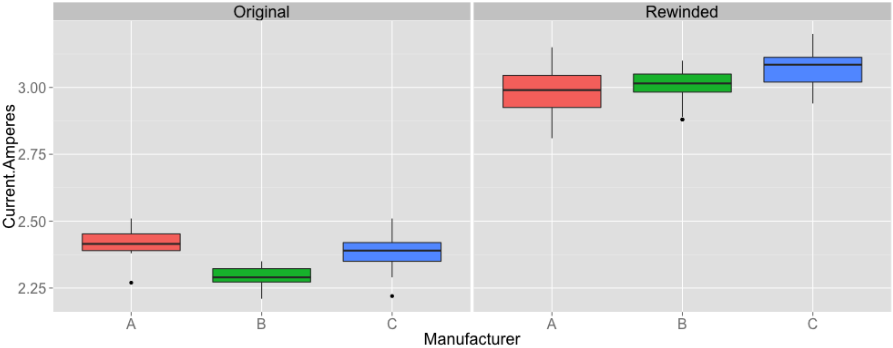


Factorial Designs

Example: Electrical current in motors

The exploratory plot suggests a relatively large effect for the *State* factor, but is inconclusive with regards to the *Manufacturer* effect. Any interaction effect is also likely to by small.

Lets assume for this example that the engineers want $\alpha = 0.05$, $\beta = 0.2$ and $\delta^* = 0.1A$.



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 effects model for a set of observations collected following this design can be expressed as:

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

Factorial Designs

Statistical model for two factors

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

As before, effects are treated as deviations from the grand mean. By construction:

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

$$\sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$$

The factorial design emerges whenever we wish to model both the main and the interaction effects of multiple factors. This means that, for the two-factor case, the hypotheses that can be tested are:

$$\text{Factor } A\text{, main effect:} \begin{cases} H_0 : \tau_i = 0, \ \forall i \\ H_1 : \exists \tau_i \neq 0 \end{cases}$$

$$\text{Factor } B\text{, main effect:} \begin{cases} H_0 : \beta_j = 0, \ \forall j \\ H_1 : \exists \beta_j \neq 0 \end{cases}$$

$$\text{Interaction effect, } AB\text{:} \begin{cases} H_0 : (\tau\beta)_{ij} = 0, \ \forall i, j \\ H_1 : \exists (\tau\beta)_{ij} \neq 0 \end{cases}$$

Factorial Designs

Statistical model for two factors

The test statistics for these hypotheses will, as usual, be derived from the partition of the total variability into specific components:

$$
\begin{aligned}
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\bar{y}_{\dots}\right)^2 \\
&= \underbrace{bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{\dots})^2}_{SS_A} + \underbrace{an\sum_{j=1}^{b}\left(\bar{y}_{.j.}-\bar{y}_{\dots}\right)^2}_{SS_B} \\
&\quad + \underbrace{n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{\dots}\right)^2}_{SS_{AB}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\bar{y}_{ij.}\right)^2}_{SS_E}
\end{aligned}
$$

To distinguish SS_{AB} from SS_E we need $n \geq 2$.

Factorial Designs

Statistical model for two factors

The mean squares are also calculated as usual:

$$MS_A = \frac{SS_A}{a-1} \qquad E\left[MS_A\right] = \sigma^2 + \frac{bn\sum\limits_{i=1}^{a}\tau_i^2}{a-1}$$

$$MS_B = \frac{SS_B}{b-1} \qquad E\left[MS_B\right] = \sigma^2 + \frac{an\sum\limits_{j=1}^{b}\beta_j^2}{b-1}$$

$$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} \qquad E\left[MS_{AB}\right] = \sigma^2 + \frac{n\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$MS_E = \frac{SS_E}{ab(n-1)} \qquad E\left[MS_E\right] = \sigma^2$$

Factorial Designs

Statistical model for two factors

If the usual assumptions (ϵ_{ijk} i.i.d. $\mathcal{N}(0, \sigma^2)$) hold, the fractions:

$$F_0^{(A)} = \frac{MS_A}{MS_E}$$

$$F_0^{(B)} = \frac{MS_B}{MS_E}$$

$$F_0^{(AB)} = \frac{MS_{AB}}{MS_E}$$

are distributed under their respective null hypotheses as F variables (each with their respective degrees of freedom), and the hypotheses can be tested in the usual manner (i.e., comparing the obtained value of F_0 against the critical value of $F_{\alpha;df_1;df_2}$).

Example: Electrical current in motors

Statistical model for two factors

```
> 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
```

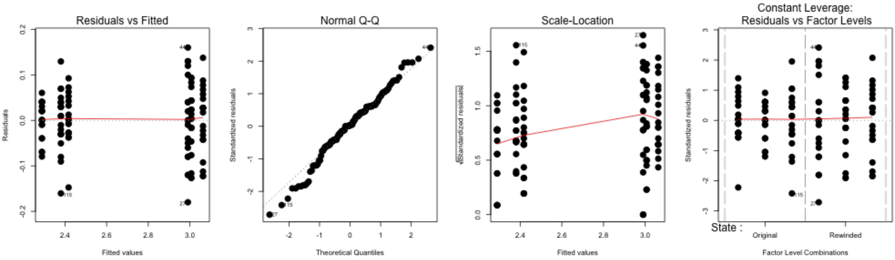
Example: Electrical current in motors

Statistical model for two factors

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
```



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:

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

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

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.

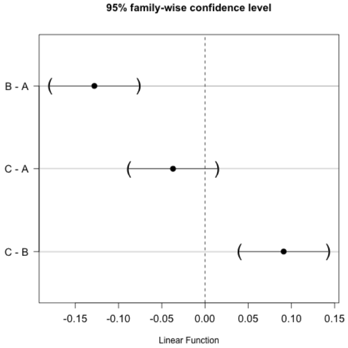
Factorial designs

Multiple Comparisons

The usual routines for performing multiple comparisons in *R* are applicable. For instance, performing *all vs all* comparisons using Tukey's method yields, for the *Manufacturer* factor:

```
> mcp.manuf<-glht(model,
+                 linfct = mcp(Manufacturer = "Tukey"))
Warning message:
In mcp2matrix(model, linfct = linfct) :
  covariate interactions found
  -- default contrast might be inappropriate

> plot(confint(mcp.manuf),
+      cex.axis   = 1.2,
+      cex        = 2)
```



Bibliography

Required reading

1.
2.

Recommended reading

1.
2.

About this material

Conditions of use and referencing

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.},
}
```