S 09

Data Analytics for Data Scientists

Design of Experiments (DoE)

Suggested solutions for Exercise 09: Factorial Designs

2025

Prof. Dr. Jürg Schwarz
MSc Adrian Bieri

Suggested solution 01

Properties of designs

There are full factorial and fractional factorial designs.

Please give one example of a full factorial design and one example of a fractional factorial design and explain the advantages and disadvantages of each.

Suggested answers to the questions

Full factorial design	**Fractional factorial design**
Description: A full factorial design is an experimental setup in which all possible combinations of factor levels are tested. **For example**, if there are three factors, A, B, and C, each with two levels (low and high), a full factorial design would include 2 x 2 x 2 = 8 experimental runs: (A low, B low, C low), (A low, B low, C high), (A low, B high, C low), (A low, B high, C high), (A high, B low, C low), (A high, B low, C high), (A high, B high, C low), and (A high, B high, C high).	**Description**: A fractional factorial design is a type of experimental design in which only a subset of all possible combinations of factor levels is tested. **For example**, if there are three factors, A, B, and C, each with two levels, instead of running all $2^3 = 8$ combinations, a fractional factorial design might include only 4 selected combinations to reduce the experimental burden: (A low, B low, C low), (A low, B high, C high), (A high, B low, C high), and (A high, B high, C low).
Advantages: - Allows the estimation of all main effects and interaction effects between factors. - Provides a complete picture of how all factors influence the response variable. - Suitable for studying complex systems where interactions between factors are important.	**Advantages**: - More efficient in terms of the number of runs required, especially when there are many factors. - Useful when there are resource constraints or when a preliminary study is being conducted to identify significant factors. - Can provide valuable information with fewer experimental runs if the assumption of negligible interactions holds.

Disadvantages: - Requires a large number of experimental runs, especially as the number of factors or levels increases. - Can be very resource-intensive in terms of time, cost, and materials.	**Disadvantages**: - Does not allow estimation of all interactions between factors, as only a subset of combinations is tested. - Risk of confounding effects, where some factor effects are indistinguishable from others, making interpretation more difficult. - May not provide a complete understanding of the system if interactions are actually significant.

Note

The note in "Task 01" in the "Tasks for Exercise 09: Factorial Designs"– see below – was, in this case, a trap intended to make you think.

The differences are precisely symmetrical in the case of full factorial vs. fractional factorial.

Note: When comparing methods, things, etc., advantages and disadvantages can often be presented in terms of symmetry.

But this can trivialize the comparisons → see the example small car vs. luxury limousine.

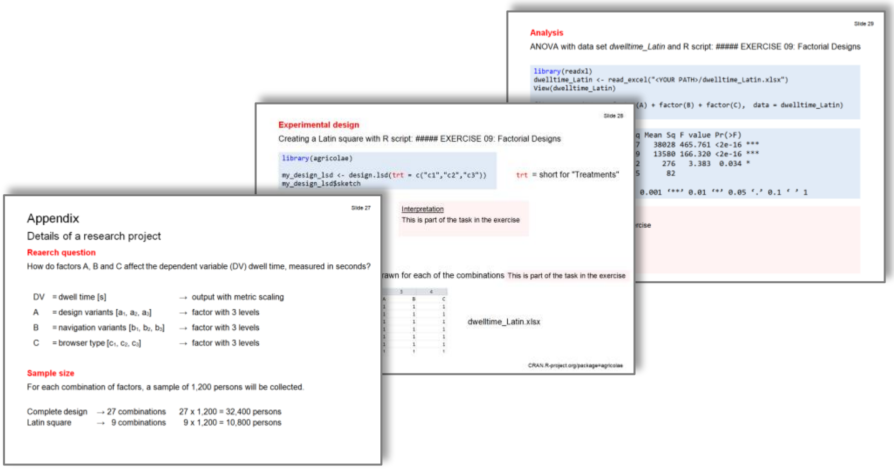
Avoid such comparisons and find **distinct** advantages and disadvantages.

Small car	**Luxury limousine**
Advantages → Small and agile	Advantages → High comfort
Disadvantages → Low comfort	Disadvantages → Big and bulky

Suggested solution 02

Research project – Latin square with R & Analysis

In the appendix of the lecture notes you will find details of a research project:



First, read through all the 3 slides in the appendix.

Use R to create the missing Latin square on Slide 28 and interpret its output.
Use R to replicate the ANOVA shown on Slide 29 and interpret its output.

Use key words to interpret the Latin square and the ANOVA.
Describe the results – insert also the R-code, the R-output and if necessary, R-plots.

Suggested answers to the questions – Experimental design

```
library(agricolae)

my_design_lsd <- design.lsd(trt = c("c1","c2","c3"))
my_design_lsd$sketch
```

The column names of the Latin square ([,1] [,2] [,3]) correspond to the levels of factor A.
The row names of the Latin square ([1,] [2,] [3,]) correspond to the levels of factor B.
The content in the cells of the Latin square ("c1", "c2", "c3") correspond to the levels of factor C.

```
     [,1] [,2] [,3]
[1,] "c3" "c2" "c1"
[2,] "c2" "c1" "c3"
[3,] "c1" "c3" "c2"
```

[,1] [,2] [,3]	→ Levels of factor A
[1,] [2,] [3,]	→ Levels of factor B
"c1","c2","c3"	→ Levels of factor C

Suggested answers to the questions – Sampling

A sample with n = 1,200 is drawn for each of the combinations [a_1, b_1, c_3], [a_2, b_1, c_2], … .

Suggested answers to the questions – Analysis

```
                Df Sum Sq Mean Sq F value Pr(>F)
factor(A)        2  76057   38028 465.761 <2e-16 ***
factor(B)        2  27159   13580 166.320 <2e-16 ***
factor(C)        2    552     276   3.383  0.034 *
Residuals    10338 844075      82
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

There is a main effect of A (levels 1, 2, 3) on DV, $F(2, 10338) = 465.761$, $p = .000$.
There is a main effect of B (levels 1, 2, 3) on DV, $F(2, 10338) = 166.320$, $p = .000$.
There is a main effect of C (levels 1, 2, 3) on DV, $F(2, 10338) = 3.383$, $p = .034$.

<u>Note</u>
With $p = .034$ the factor C is relatively close to the significance limit of $p = .050$.