# Replications

Calculate the number of replications needed for the Exercise 2



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cleaner | Machines | Workers | x1 | x2 | Mean | Confidence interval (alfa=0.05) | n\*  5% | n\*  6% |
| - | - | - | 20,885 | 20,261 | 20,57 | 3,964 | 29,71 | 20,628 |
| - | - | + | 33,836 | 36,368 | 35,10 | 16,08 | 168,01 | 116,67 |
| - | + | - | 9,9099 | 142 | 75,95 | 839,18 | 97.654,01 | 67815,28 |
| - | + | + | 17,766 | 131,13 | 74,45 | 720,21 | 74.869,67 | 51992,82 |
| + | - | - | 42,759 | 0,0402 | 21,40 | 271,39 | 128.673,31 | 89356,46 |
| + | - | + | 5,7025 | 2,327 | 4,01 | 21,44 | 22.825,53 | 15851,06 |
| + | + | - | 10,481 | 8,7404 | 9,61 | 11,05 | 1.059,13 | 735,50 |
| + | + | + | 5,9775 | 5,1167 | 5,55 | 5,46 | 777,56 | 539,97 |

# Bank model

## A bank is planning the requirements for its ATMs, in the frame of a new expansion. There are places for up to six ATMs, not all these places have to be used. You can buy three types of ATM: ATM generals (to give cash, balances, mini statements and change PIN), cash machines for the payment of money and ATMs that provide the full statements of the account. The Bank has a policy that customers would not have to wait more than 5 minutes in *most* cases (usually performed with the 99% of cases).

For the problem of the Bank, we select the following variables as factors, determining the next factorial design. We are interested in analyzing the service time of the ATM.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | General ATM.  (A) | Income ATM.  (B) | Management ATM.  (C) | Answer |
| E1 | - | - | - | 8.7 |
| E2 | - | - | + | 8.7 |
| E3 | - | + | - | 8.7 |
| E4 | - | + | + | 8.7 |
| E5 | + | - | - | 1.4 |
| E6 | + | - | + | 1.4 |
| E7 | + | + | - | 1.4 |
| E8 | + | + | + | 1.4 |

Where "+" means 2 servers while "-" one.

Calculate the effects of (A), (B) and (C) using direct formulas and using Yates algorithm.

# 9 factors and 20K€

We have a limited budget to analyze the 9 different factors that can be considered to improve the behavior of the system. Each individual experiment costs 100€ and we have a total budget of 20.000€ to be destined to experimentation. Define an experiment design, with this constraint, considering that we need at least 3 replications for each experiment.

# SEO Part one

We are going to face with a SEO problem; we try to understand what are the main factors that determines the position of a website on a search result. The main concern is the huge volume of data we need to consider. We are facing a table with more than 300.000 observations (that represents the results for a specific search), and more than 200 variables (factors) in each individual (observation).

The structure of the data is similar to the presented on this table.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Keyword | Pos | url | Doc\_length (words) | AdSense | chars | (..) |
| anniversary gift | 1 | <http://url1.com> | 1451 | 0 | 42786 | .. |
| anniversary gift | 2 | <http://url2.com> | 1887 | 1 | 46354 | .. |
| anniversary gift | 3 | <http://url3.com> | 1202 | 3 | 18632 | .. |
| anniversary gift | 4 | <http://url4.com> | 1503 | 0 | 59325 | .. |
| anniversary gift | 5 | <http://url5.com> | 1832 | 2 | 65323 | .. |
| .. | .. | .. | .. | .. | .. | .. |
| Halloween costumes | 1 | <http://url11.com> | 1451 | 0 | 42786 | .. |
| Halloween costumes | 2 | <http://url12.com> | 1887 | 1 | 46354 | .. |
| Halloween costumes | 3 | <http://url13.com> | 1202 | 3 | 18632 | .. |
| Halloween costumes | 4 | <http://url14.com> | 1503 | 0 | 59325 | .. |
| Halloween costumes | 5 | <http://url15.com> | 1832 | 2 | 65323 | .. |
| .. | .. | .. | .. | .. | .. | .. |

We want to analyze it to understand if is possible to detect what are the main factors (and the values needed) in order to improve our position in a search engine regarding a specific keyword.

What are the techniques you are going to use here? **Detail clearly every step to be done**.

# SEO Part two

With the previous information, you will be able to define a model but, a detailed analysis makes clear that is needed to define a complex simulation model to obtain accurate results. This model needs more than one hour to complete a single replication.

With this model, we can parametrize our websites to try to find the best combinations to improve the position of the website in the search engines. Although the number of factors in this simulation model is just 10 (not the 200 previously analyzed) the amount of experiments to be conducted will be huge.

Following these two assumptions:

1. For the distinct factors two levels are enough.
2. 3 replications are enough for a preliminary analysis.

What is the experimental design we need define to obtain results in less than 5 weeks? We have just a single computer.

Discuss the proposed solution.

# 10 factors model

We have a limited budget to analyze the different 10 factors to consider on our model. Each individual experiment costs 1000€ and we have a total budget of 56.000€ to be destined to experimentation. Define an experiment design, with this constraint, considering that we need at least 2 replications for each experiment. We do not want to consider nothing but the main effects.

# The ceramic strength

We want to determine the effect of machining factors on ceramic strength, our response variable is the ceramic strength.

We have 3 factors and, for each factor, different values.

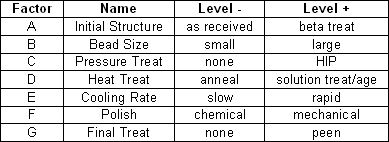
* Factor 1, the table speed is going from .025 m/s to .125 m/s, a real value.
* Factor 2, the down feed rate is going from .05 mm to .125 mm, a real value.
* Factor 3, the direction has two levels, longitudinal and transverse.

Define a DOE to determine what is the best scenario regarding our response variable. How do you deal with the randomness of the experiment? What are you going to apply to determine the interactions and main effects? Justify your answers.

# Life testing of weld-repaired

Consider a life testing of weld-repaired. The objective of the test is to identify the key factors that affect the life and to improve the product life. There are seven factors that may affect the life. A two-level full factorial design will require 27 = 128 runs. It will be time-consuming and costly.

For this example, the seven factors are:



Compare the alternative of a full factorial design with other less costly alternatives. Discuss the pros and the cons of the considered alternatives.

A maximum of 64 scenarios can be analyzed.

# Improving the factory

We want to determine the effects of improving or changing several machines on a factory. Our goal is to minimize the time needed to produce a piece.

We have 3 machines that can be changed or improved in its mean service time.

1. Machine 1, we can change this machine, the time of the old machine is 10 seconds, the new machine needs 5 seconds. The price for the new machine is 8000€.
2. Machine 2, we can improve the software, in that case with the new software the machine needs 30 seconds to complete its operation, 50 seconds with the old software. The price to improve the software is 3000 €.
3. Machine 3, on this machine we can use more oil. Using more oil, we reduce the operation time 5 seconds, from 10 to 5 seconds to operation. The price of the oil is 10€ by day.

Define a DOE to determine what is the best scenario regarding our response variable, the total time needed to complete the piece.

We know that the function that defines the total time of operation is:

Where *exp* is an exponential distribution with the parameter defined by the factor.

****

Consider that (because you don’t have enough time to calculate a full factorial design) you need to work with just 4 experiments.

How do you deal with the randomness of the experiment? what is the number of replications needed?

Considering that only 2 replications are enough, explain the results and calculate the investment needed to improve the industry results during a year.

# Manufacturing process

Consider a manufacturing process for a new computer chipset. The objective is to improve the overall production process. Mangers detects four factors that are key elements on the process development.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Factor** | **Name** | **Unit** | **Low** | **High** |
| A | Aperture Setting | - | Small: value 1 | Large: value 2 |
| B | Exposure Time | min | 20 | 30 |
| C | Develop Time | s | 30 | 45 |
| D | Mask quality | - | Average: value 10 | Good: value 20 |

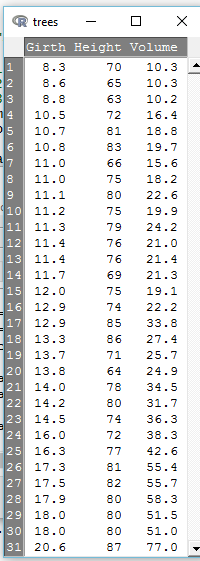
We want to determine the effects of these factors on the quality of the final chipset. This quality is measured in number of errors, measured on the experiment (less errors is better).

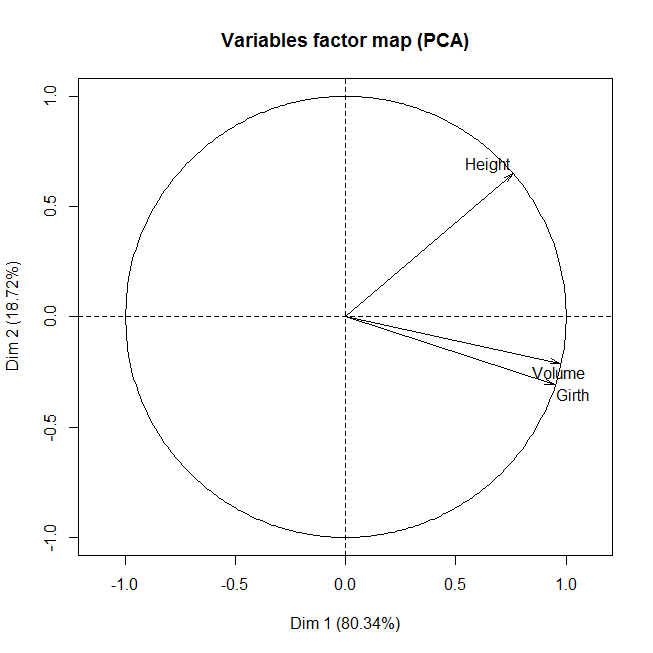
Define a DOE to determine what is the best scenario regarding our response variable. How do you deal with the randomness of the experiment? What are you going to apply to determine the best scenario? Justify your answers.

To obtain the answer of our experiments we can use this function:

# Sequoias

P1 (0,5 points). We want to analyze the relation of the Girth, Height and the volume in large trees (sequoias). Discuss the nature of the information we have. See the table.

P2 (1 point). Analyzing the next diagram, what can you understand of the Volume of a tree? Describe the information you can obtain form the analysis.



P3 (2 points). We want to conduct an experiment to understand if we can modify the Height of the trees considering other factors, like climatic condition (mainly humidity, sun radiation), and the exposure to some gases (mainly Co2 and No2). Define an experimental design to analyze all these factors (Volume, Girth, Height, Humidity, Sun radiation, Co2, NO2). The main idea is to plant different trees under the selected conditions and analyze the evolution of those trees during 5 years. We have the chance to control only **8 different environments** (greenhouses).

What will be the experimental design to be used? Justify your answers.

How are you going to deal with randomness?

# Tire pressure on fuel consumption

The treatment to be considered is the effect of tire pressure on fuel consumption Cox (1958, [[COX1](http://www.statsref.com/HTML/latin_sqaures.html#cox1)]). Four different tire pressures (levels) are to be tested, A, B, C, D. The design uses four buses and is carried out over four days. To eliminate variations between buses, and between days, these factors provide the row and column blocking elements of the design.

# Latin square example

We want to test 4 treatments; we select for 4 cows and during 4 periods. To do so, we define a Latin square experiment as is shown next.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Cow 1 | Cow 2 | Cow 3 | Cow 4 |
| Period 1 | T4 | T1 | T3 | T2 |
| Period 2 | T1 | T4 | T2 | T3 |
| Period 3 | T3 | T2 | T1 | T4 |
| Period 4 | T2 | T3 | T4 | T1 |

After the randomizations and the executions, we obtain the values of the next table. We also add the means.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cow 1 | Cow 2 | Cow 3 | Cow 4 |  |  |  |  |
| Period 1 | 192 | 195 | 292 | 249 | 232 |  | T1 | 191 |
| Period 2 | 190 | 203 | 218 | 210 | 205,25 |  | T2 | 206,75 |
| Period 3 | 214 | 139 | 245 | 163 | 190,25 |  | T3 | 217 |
| Period 4 | 221 | 152 | 204 | 134 | 177,75 |  | T4 | 190,5 |
|  | 204,25 | 172,25 | 239,75 | 189 | 201,3125 |  |  |  |

Calculate the ANOVA table that allows to detect the effects of the experimental design.

# NZEBs

Here we want to define the best parameters for a building in order to **improve its energetic behavior**. The building is the (e)CO building that is going to compete in the prestigious solar Decathlon competition, see Figure 1.

****

Figure 1. (e)CO 2012 Solar Decathlon building.

**QUESTION 1: What are the steeps that are going to be done in this analysis?**

You will be able to define a model but, a detailed analysis makes clear that is needed to define a complex simulation model to obtain accurate results.

We work with four walls in two modules, **North Wall, West Wall, South Wall** and **East Wall** for north models and south modules, and the **ceiling**. Also, we discuss regarding the **external supply** use (Yes or Not). For the walls and the ceiling, we can use different materials that goes from **0.212** to **0.174** of **thermal insulation**.

We can build a simulation model that taking all those parameters provides an answer in about one hour (to complete a single replication for a specific parametrization).

If 3 replications are enough for the pilot test analysis…

**QUESTION 2: What is the experimental design we need define to obtain results in less than 5 weeks? We have just a single computer. How are you going to deal with randomness (replications)? Discuss the proposed solution.**

Once we have the experimental design to be executed completed, is needed to execute the simulation model that be able to obtain the answers for each scenario. This simulation model takes care of two elements, the arrival of a new energy demand to the building, and the time that this energy demand last.

We are going to consider only the factor that defines if we are using an external energy supply. If a new energy demand arrives to the building, we can use an external power supply or we can put this demand on a queue and wait.

**QUESTION 3: Model the two scenarios using an event scheduling approach. The data to be used is on the next table. We consider that when a new demand arrives, the solar power of the building is enough, but if two demands arrives (at the same time) then, if the power supply exist we are going to use it, if not, this demand must wait to be served in a queue.**

|  |  |  |
| --- | --- | --- |
| Element | Demand | Until |
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 2.5 | 2 |
| 4 | 3 | 1 |
| 5 | 6 | 2 |

Once you execute the model and you put the results in the DOE table you obtain the next answers (we consider only a subset of the scenarios).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **North MODULE isolation** | | |  |  |  |  |
| **North Wall** | **South Wall** | **External supply** | VALUES |  |  | External supply use |
| - | - | - | 0,174 | 0,174 | 0 | 2,87 |
| - | - | + | 0,174 | 0,174 | 1 | 0,74 |
| - | + | - | 0,174 | 0,212 | 0 | 2,59 |
| - | + | + | 0,174 | 0,212 | 1 | 0,72 |
| + | - | - | 0,212 | 0,212 | 0 | 2,36 |
| + | - | + | 0,212 | 0,212 | 1 | 0,70 |
| + | + | - | 0,212 | 0,212 | 0 | 2,36 |
| + | + | + | 0,212 | 0,212 | 1 | 0,70 |

**QUESTION 4: What is the effect of the insulation or the external power use in this NZEB?**

# NZEB parameters

Consider a NZEB where we can change several parameters in order to improve the overall behavior (energy consumption) of the building.

The next table defines the energy needs depending on 3 factors.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Factor** | **Name** | **Unit** | **Low** | **High** |
| A | Apertures (windows) | - | Small window: 1 | Large window:  2 |
| B | Roof insulation | range | 20 | 30 |
| C | Floor insulation | range | 30 | 45 |
| D | Wall insulation | range | 10 | 25 |

We want to determine the effects of these factors on the energy consumption of the final building. This quality is measured in the energy consumption, measured on the experiment (less energy is better). To start we consider that we FIX the wall with the cheaper material (less insulation i.e 10 value)

Define a DOE to determine what is the best scenario regarding our response variable. What is the more important factor from the analyzed?

Justify your answers.

To obtain the answer of our experiments we can use this function (as usual this is a simplification, the expression will come from real experiments conducted on a real building, or as is done more a more frequently, from a simulator that is like a black box).

Once we finish this analysis, we want to see what the effect of D on the results FIXING all the other values is (with the best candidates). Can you define an expression to explain this *maybe* linear relation?

Discuss the results, considering that you have money to analyze 4 more scenarios.