Latin square experiment design in R

D Hodzic, A Hodzic and E Bajramovic

University of Bihac, Faculty of Technical Engineering, Irfana Ljubijankica bb, 77000 Bihac, Bosnia and Herzegovina

E-mail: damir.hodzic@unbi.ba

Abstract. In this paper we will describe design of experiment by latin square method. Latin square is statistical test which is used in planning of experiment and is one of most accurate method. Programming language R is a tool which can be used for statistical tests and graphics. It has a wide range of applications both in research and practice. A practical example will be presented in this paper and its processing in the classical approach and in the programming language R.

1. Introduction

Design of Experiment is one of the new scientific disciplines whose object is the scientific-research experiment. The word experiment is inspired by the Latin word of the experiment, which translates into an opus or an excuse. There are several different definitions of experiments, but from the engineering point of view we can say that the experiment presented a scientifically designed opus that includes the operation system, algorithm and experimental technique, to test an object under well-defined regimes and conditions. The experiment is used in the final phase of the research when it is the key criterion for verifying the truth of theory and hypothesis. Experiment planning is a discipline associated with sample theory. Theory of samples evaluates population characteristics based on observation of individual parts of these populations without any change of population while planning of experiments affects one part of the population to assess the effect of these activities on the entire population. One experiment can have different meanings. Thus, in engineering analyzes, the experiment may be the preparation of pieces of a material and the measurement of the characteristics of that piece. It may also mean studying the moon by sending monthly footage or studying the Earth's Ozone Layer. Also, the experiment may also mean studying the effects of certain drugs in the treatment of certain diseases in humans. In all of these studies it is assumed that the visible part of the population or data is representative.

A well-planned experiment allows us to get clear interpretations and avoid complicated analyzes. The poorly planned experiment gives us some wrong conclusions of a process. Design of Experiment is of particular importance in all quantitative studies, especially when multiple factors are studied at the same time to examine and compare their effects.

The basic terms we will use in this paper and which are the basis of each experiment planning are factors, treatments and blocks. Factors represent different types of influences that apply to
experimental units. Treatment is a specific combination of levels of different factors and the blocks are homogeneous groups by which we compare the treatments.

2. Latin square method
If the experiment is performed so that another factor is presented beside the effect of the treatments and the blocks, and if there is no interaction between these three factors, then we can use Latin square method. The basic feature of the randomized block design, which is a method also used for design of experiment and is similar to Latin square method, is that at each block one treatment occurs at least once. In the Latin square method in each row and column every treatment occurs only once. In this way, it is often achieved that the variations of these basic groups are small, so it is a small experimental error.

For the statistical analysis of the Latin square, we will characterize the row with $i$, column with $j$ and treatment with $k$. The results of measurement occurring under the $i$th row, $j$th column and $k$th treatment is denoted by $y_{ijk}$ where $1 \leq i, j, k \leq p$.

In general The Latin square with dimensions $p \times p$ is a quadratic scheme in which Latin letters $p$ means $p$ rows and $p$ columns, indicating treatments and each letter is repeated $p$ times. Thus each letter occurs only once in each row and each column [2], [3].

Some of examples of Latin squares are:

$$
\begin{align*}
A & B & C & D & A & D & B & E & C & A & D & C & E & B & F \\
B & A & D & C & D & A & C & B & E & B & A & E & C & F & D \\
C & D & B & A & C & B & E & D & A & C & E & D & F & A & B \\
D & C & A & B & B & E & A & C & D & D & C & F & B & E & A \\
& E & C & D & A & B & F & B & A & D & C & E & & E & F & B & A & C & D 
\end{align*}
$$

**Figure 1.** Examples of Latin Square ($4x4$, $5x5$, $6x6$)

Latin squares are used in many fields of research work (agriculture, industry, medicine, sociology,...). The advantage of the Latin squares is that three factors can be studied in a small experiment. In addition to contributing to a reduction in the experimental error, the statistical analysis procedure in the Latin square is relatively simple. The lack of a Latin squares is that a greater number of treatments require a greater number of repetitions, making the experiment uneconomical and often less precise. Therefore, the squares are rarely applied with square dimensions greater than $10 \times 10$. Also, Latin squares with dimensions less than $4 \times 4$ are not being used because of the small number of degrees of freedom at the sum of squares for error.

Results can be displayed in several ways, some of which are:

$$
\begin{align*}
8_A & 10_C & 12_B & 9_D & A+8 & C+10 & B+12 & D+9 & 8 & 10 & 12 & 9 \\
7_B & 8_D & 10_A & 11_C & B+7 & D+8 & A+10 & C+11 & 7 & 8 & 10 & 11 \\
9_D & 10_A & 8_C & 9_B & D+9 & A+10 & C+8 & B+9 & 9 & 10 & 8 & 9 \\
12_C & 11_B & 9_D & 13_A & C+12 & B+11 & D+9 & A+13 & 12 & 11 & 9 & 13 
\end{align*}
$$

**Figure 2.** Examples of displaying latin square results

In the statistical analysis of the Latin square, the Analysis of Variance (ANOVA) with three factors is applied, which will be presented in the next section.
3. Practical example of Latin square

We will assume that we have an example of four workers who have processed different workpieces in four periods. The work was graded from 1 to 15. The experiment was performed by random selection of the Latin square shown in Figure 3.

\[
\begin{array}{cccc}
C & B & D & A \\
D & C & A & B \\
A & D & B & C \\
B & A & C & D \\
\end{array}
\]

**Figure 3.** Random selection of Latin square for example

and we got following results (Table 1).

| Worker | Period | Worker | Period |
|--------|--------|--------|--------|
| u₁     | t₁     | u₁     | t₁     |
| 12     | 9      | 12     | 9      |
| 9      | 8      | 8      | 7      |
| u₂     | t₂     | u₂     | t₂     |
| 6      | 7      | 6      | 9      |
| 7      | 6      | 6      | 9      |
| u₃     | t₃     | u₃     | t₃     |
| 5      | 10     | 6      | 9      |
| 10     | 6      | 6      | 9      |
| u₄     | t₄     | u₄     | t₄     |
| 11     | 8      | 7      | 8      |
| 11     | 8      | 7      | 8      |

**Table 1.** Results for the example

It is necessary to examine whether there is a significant influence of the worker, the time at which the workpiece was processed or the processing method to the experimental results if \( \alpha = 0.05 \).

First, we will determine the sum of columns and rows and arithmetic means for the given data.

**Table 2.** Sum of columns and rows and arithmetic means for the example

\[
\begin{array}{cccc|c|c}
Period & t₁ & t₂ & t₃ & t₄ & Vᵢ &  \bar{y}_{i,} \\
\hline
u₁ & 12 & 9 & 8 & 7 & 36 & 9  \\
u₂ & 6  & 7 & 6 & 9 & 28 & 7  \\
u₃ & 5  & 10& 6 & 9 & 30 & 7,5
\end{array}
\]

\[
\begin{array}{cccc|c|c}
Period & t₁ & t₂ & t₃ & t₄ & \bar{y}_{.,} \\
\hline
Kᵢ & 34 & 34 & 27 & 33 & 128 & 8,5  \\
\bar{y}_{.,} & 8,5 & 8,5 & 6,75 & 8,25 & 8,\bar{j}  \\
\end{array}
\]

According to ANOVA calculation we have to calculate sum of squares

\[
q = \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} (\bar{y}_{ijk} - \bar{y})^2 = (8 - 12)^2 + (8 - 9)^2 + \ldots + (8 - 9)^2 = 56
\]  \hspace{1cm} (1)

\[
q_{\nu} = p \sum_{i=1}^{p} (\bar{y}_{i,} - \bar{y})^2 = 4(1^2 + 1^2 + 0,5^2 + 0,5^2) = 10
\]  \hspace{1cm} (2)
\[ q_K = p \sum_{i=1}^{k} (\bar{y}_{i,j} - \bar{y})^2 = 4(0,5^2 + 0,5^2 + 1,25^2 + 0,25^2) = 8,5 \]  \hspace{1cm} (3)

where:

\( q \) – Sum of squares total
\( q_v \) – Sum of squares of rows
\( q_K \) – Sum of squares of columns

Now we will show results according to treatments (Table 3).

**Table 3. Results according to treatments**

| Treatment | A   | B   | C   | D   |
|-----------|-----|-----|-----|-----|
|           | 7   | 9   | 12  | 8   |
|           | 6   | 9   | 7   | 6   |
|           | 5   | 6   | 9   | 10  |
|           | 8   | 11  | 7   | 8   |
| Total     | 26  | 35  | 35  | 32  |

\[ \bar{y}_{.,k} \]

| \( \bar{y}_{.,k} \) | 6,5 | 8,75 | 8,75 | 8   |

\[ q_T = p \sum_{k} (\bar{y}_{.,k} - \bar{y})^2 = 4(1,5^2 + 0,75^2 + 0,75^2 + 0) = 13,5 \]  \hspace{1cm} (4)

\[ q_K = q - q_v - q_K = 56 - 10 - 8,5 - 13,5 = 24 \]  \hspace{1cm} (5)

where:

\( q_T \) – Sum of squares of treatments
\( q_K \) – Sum of squares of error

Now we calculate \( F \) values for rows, columns and treatments:

\[ F_v = \frac{(p-1)(p-2)q_v}{p-1} \frac{p}{q_R} = \frac{3 \cdot 2}{3} \cdot \frac{10}{24} = 0,83 \]  \hspace{1cm} (6)

\[ F_K = \frac{(p-1)(p-2)q_K}{p-1} \frac{p}{q_R} = \frac{3 \cdot 2}{3} \cdot \frac{8,5}{24} = 0,71 \]  \hspace{1cm} (7)

\[ F_T = \frac{(p-1)(p-2)q_T}{p-1} \frac{p}{q_R} = \frac{3 \cdot 2}{3} \cdot \frac{13,5}{24} = 1,125 \]  \hspace{1cm} (8)

where \( p \) is number of rows and columns.

Now we can make ANOVA table with all datas (Table 4)

**Table 4. Analysis of Variance for Latin square example**

| Source of variation | Sum of squares | Degrees of freedom | Mean square | \( F \) value |
|---------------------|---------------|--------------------|------------|--------------|
| Rows                | 10            | 3                  | 3,33       | 0,83         |
| Columns             | 8,5           | 3                  | 2,83       | 0,71         |
| Treatments          | 13,5          | 3                  | 4,5        | 1,125        |
| Error               | 24            | 6                  | 4          |              |
| Total               | 56            | 15                 | 3,73       |              |
$F$ values calculated in equations (6), (7) and (8) we compare with $F$ from distribution table for $\alpha = 0.05$ (Figure 4). We choose $F$ according to degrees of freedom 3 and 6 from Table 4.

![F Distribution](image)

**Figure 4.** $F$ distribution for $\alpha = 0.05$

Tabular $F$ value is 4.76 and we can see that this value is greater then all calculated $F$ values ($F_K < F; F_V < F; F_T < F$). It means that the effects of columns, rows and treatments are not significant, there is no significant influence on the time of processing, the worker and the way of processing on the experimental results.

4. Practical example of Latin square in R
R is the programming language and environment for statistical calculations and visualization. It is free to use and provides a wide range of statistical methods for linear and nonlinear modeling, classical statistical tests and others. In this paper, it will be shown that the programming language R can easily solve an example of a previously performed classic experiment. First it is necessary to input values and names for rows, columns and treatments then to group the values and analyze the variance (Figure 5) [4-6].

# Latin square
# Damir Hodzic, Atif Hodzic, Esad Bajramovic - Faculty of Technical Engineering Bihac
# Input values
worker <- c(rep("worker1",1), rep("worker2",1), rep("worker3",1), rep("worker4",1))
period <- c(rep("periodA",4), rep("periodB",4), rep("periodC",4), rep("periodD",4))
letter <- c("C","D","A","B","B","C","D","A","D","A","B","C","A","B","C","D")
value <- c(12,6,5,11,9,7,10,8,8,6,6,7,7,9,9,8)

# Grouping the values
mydata <- data.frame(worker, period, letter, value)
matrix(mydata$letter, 4,4)
matrix(mydata$value, 4,4)
myfit <- lm(values ~ worker+period+letter, mydata)

# Analysis of variance
anova(myfit)
par(mfrow=c(2,2))
plot(value ~ worker+period+letter, mydata)

Figure 5. Program code in R for One way ANOVA

After R gets input values it will generate and arrange them into the matrix (Figure 6).

> mydata
  worker period letter value
  1 worker1 periodA C   12
  2 worker2 periodA D   6
  3 worker3 periodA A   5
  4 worker4 periodA B  11
  5 worker1 periodB B   9
  6 worker2 periodB C   7
  7 worker3 periodB D  10
  8 worker4 periodB A   8
  9 worker1 periodC D   8
 10 worker2 periodC A   6
 11 worker3 periodC B   6
 12 worker4 periodC C   7
 13 worker1 periodD A   7
 14 worker2 periodD B   9

Figure 6. Generating the values in R

By executing the ANOVA command (Figure 7) we will get the results in the R which are identical to the calculated values in Table 4.

> anova(myfit)
Analysis of Variance Table
Response: value
  DF  Sum Sq Mean Sq F value Pr(>F)
worker 3 10.000  3.3333 0.8333 0.5226
period 3  8.500  2.8333 0.7083 0.5813
letter 3  13.500  4.5000 1.1250 0.4108
Residuals 6  24.000  4.0000

Figure 7. Results in R

Results can also been presented graphically.

5. Conclusion
Design of Experiment is an important scientific field in engineering disciplines. The Latin squares method is one of the methods that has a wide range of applications and is interesting for analysis in new software packages. Software program R is a free code package that is accessible to everyone and
is therefore used for that reason. It is relatively easy to reach conclusions for a particular experiment and much faster than the classic way outlined in this paper.

References
[1] Montgomery D C 2015 Design and Analysis of Experiments (8th Edition), Arizona State University, John Wiley and Sons
[2] Kovač P 2016 Metode planiranja i obrade eksperimenta, Univerzitet u Novom Sadu, Fakultet tehničkih nauka
[3] Petrović Lj 2013 Teorija uzoraka i planiranje eksperimenta, Univerzitet u Beogradu, Ekonomski fakultet
[4] Hodzic D 2017 Planiranje eksperimenta metodom ANOVA u programskom jeziku R, 11th International Scientific Conference on Production Engineering RIM 2017, Sarajevo, Bosnia and Herzegovina
[5] Kerns G J 2012 Introduction to Probability and Statistics using R, GNU Free Documentation
[6] Gerrard P, Johnson R M 2015 Mastering Scientific Computing with R, Packt Publishing, Birmingham-Mumbai