9th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2020
Belgrade, Serbia, 15 – 16 October 2020

FACTORIAL ANOVA EXPERIMENTAL DESIGN IN R

DAMIR HODŽIĆ
University of Bihać, Faculty of Technical Engineering, Bihać, damir.hodzic@unbi.ba

FADIL ISLAMOVIĆ
University of Bihać, Faculty of Technical Engineering, Bihać, fadil.islamovic@unbi.ba

Abstract: *In this paper we will describe design of experiment by factorial analysis of variance (ANOVA) method. Experiments where the effects of more then one one factors are considered together are called factorial experiments and may be analyzed with the use of factorial ANOVA. Programming software 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 sofware R.*

Keywords: *ANOVA, factors, experiment, software package R*

1. INTRODUCTION

Design of Experiment is one of the new scientific disciplines whose object is the scientific-research experiment. 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 [1].

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.

Design of Experiment (DOE) is a powerful tool for achieving significant product quality improvements and process efficiency. For the design of the experiment it can also be said that it is the methodology of applying statistics in the process of performing the experiment. As such discipline is not just a engineering problem but is used in different social sciences.

During design of experiment various statistical softwers can be used. In this paper we will use software R for statistical analysis.

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 modelling, 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 [2,3].

2. ANOVA

ANOVA is short of Analysis of Variance. There are two types of ANOVA, one way and factorial. Main purpose of one way ANOVA is to test if two or more groups differ from each other significantly in one or more characteristics. Factorial ANOVA compares means across two or more independent variables.

To complete ANOVA test it is necessary to calculate sum of squares (SS), degrees of freedom (dF) and mean of squares (MS) for each factor, interaction between factors and error of experiment. After that it is needed to calculate Fisher value (F) and to compare it with F value from the table for definied factor of significance α. If calculated F value is greater than F value from the table we will reject null hypothesis which says that means of input values are equal.

In this paper we will analyse an example of 2^2 factorial design in which ANOVA parameters will be calculated and the same example will be presented in R software.

3. FACTORIAL DESIGN 2^2

Factorial design is used in experiments which involve several factors where it is necessary to study the joint effect of the factors on response.

Factorial design 2^2 is from the 2^k design series which means that there are two factors A and B, each run at two levels. The levels of the factors are usually called „low“ and „high“ [1].

In our example we will take two factors A and B at two levels and levels are denoted by (-) „low“ level and (+) „high“ level. The experiment is replicated three times, so there are 12 runs. The order in which the runs are made is random so we can conclude this is a completely randomized experiment.

The output data obtained are as shown in table 1.

Table 1. Experiment datas

Factor		Replicate of experiment			Total
A	B	I	II	III	
(-)	(-)	18	15	17	50
(+)	(-)	26	22	22	70
(-)	(+)	8	9	13	30
(+)	(+)	21	20	19	60

The four treatment combinations in this design can be shown graphically, Figure 1.

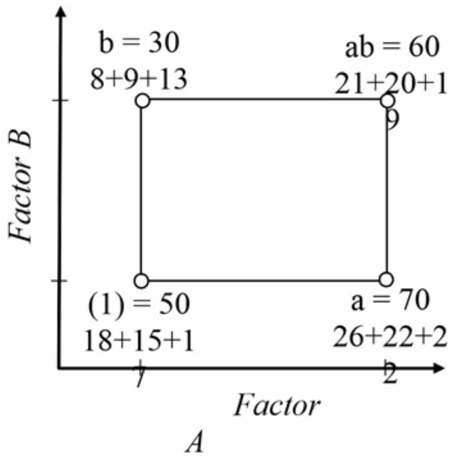

Figure 1. Treatment combinations

First, we will calculate main effects of factors A and B and interaction effect between factors, where n is number of replicates, in our example $n = 3$.

$$A = \frac{1}{2n}\left[ab + a - b - (1)\right] = \frac{1}{6}\left[60 + 70 - 30 - 50\right] = 8{,}33 \quad (1)$$

$$B = \frac{1}{2n}\left[ab + b - a - (1)\right] = \frac{1}{6}\left[60 + 30 - 70 - 50\right] = -5 \quad (2)$$

$$AB = \frac{1}{2n}\left[ab + (1) - a - b\right] = \frac{1}{6}\left[60 + 50 - 70 - 30\right] = 1{,}67 \quad (3)$$

Now we can calculate degrees of freedom (dF), as follows.

$$dF_A = bn_A - 1 = 2 - 1 = 1 \quad (4)$$

$$dF_B = bn_B - 1 = 2 - 1 = 1 \quad (5)$$

$$dF_{AB} = (bn_A - 1)\cdot(bn_B - 1) = 1 \cdot 1 = 1 \quad (6)$$

$$dF_{Error} = N - bn_A \cdot bn_B = 12 - 4 = 8 \quad (7)$$

$$dF_{Total} = N - 1 = 12 - 1 = 13 \quad (8)$$

In equations (4) to (8) bn_A and bn_B are number of levels of each factor and N is total number of runs.

Important parametar in ANOVA calculation is sum of squares. In following equations we will calculate sum of squares for each factor, interaction between factors, error and total sum of squares.

$$SS_A = \frac{\left[ab + a - b - (1)\right]^2}{N} = \frac{50^2}{12} = 208{,}33 \quad (9)$$

$$SS_B = \frac{\left[ab + b - a - (1)\right]^2}{N} = \frac{(-30)^2}{12} = 75 \quad (10)$$

$$SS_{AB} = \frac{\left[ab + (1) - a - b\right]^2}{N} = \frac{10^2}{12} = 8{,}33 \quad (11)$$

$$SS_{Total} = 18^2 + \ldots + 19^2 - \frac{(50 + 70 + 30 + 60)^2}{12} = 323 \quad (12)$$

$$SS_{Error} = SS_{Total} - SS_A - SS_B - SS_{AB} = 31{,}34 \quad (13)$$

Now, to determine mean of squares for factors, interaction, error and total we have to calculate ratio between sum of squares and degrees of freedom, as follows.

$$MS_A = \frac{SS_A}{dF_A} = \frac{208{,}33}{1} = 208{,}33 \quad (14)$$

$$MS_B = \frac{SS_B}{dF_B} = \frac{75}{1} = 75 \quad (15)$$

$$MS_{AB} = \frac{SS_{AB}}{dF_{AB}} = \frac{8{,}33}{1} = 8{,}33 \quad (16)$$

$$MS_{Error} = \frac{SS_{Error}}{dF_{Error}} = \frac{31{,}34}{8} = 3{,}91 \quad (17)$$

$$MS_{Total} = \frac{SS_{Total}}{dF_{Total}} = \frac{323}{13} = 24{,}84 \quad (18)$$

The complete ANOVA is summarized in Table 2.

Table 2. ANOVA table for example

Source of variation	*SS*	*DF*	*MS*	*F*
Factor *A*	208,33	1	208,33	53,14
Factor *B*	75	1	75	19,18
Interaction	8,33	1	8,33	2,13
Error	31,34	8	3,91	
Total	323	11	24,84	

In the last column of Table 2 there is calculated *F* value. It is necessary to calculate F value for factors and interaction between factors.

$$F_A = \frac{MS_A}{MS_{Error}} = \frac{208{,}33}{3{,}91} = 53{,}14 \quad (19)$$

$$F_B = \frac{MS_B}{MS_{Error}} = \frac{75}{3{,}91} = 19{,}18 \quad (20)$$

$$F_{AB} = \frac{MS_{AB}}{MS_{Error}} = \frac{8{,}33}{3{,}91} = 2{,}13 \quad (21)$$

To determine significance of factors and interactions between factors we will take level of significance α = 0,05 and from Fisher table determine that $F_{table}(1,8) = 5{,}317$.

We can conclude that calculated F values for factors A and B are greater than F value from the table which means that factors are statistically significant. Also we can conclude that calculated F value for interaction between factors is less than F value from table which means there is no interaction between factors.

4. FACTORIAL DESIGN 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 modelling, 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 [4,5].

At the beginning of the program code in the software R it is necessary to enter basic measurements data.

```
# Factorial design 2^2
# Input values
yield<-
matrix(c(18,15,17,26,22,22,8,9,13,21,20,19)
,byrow=T,ncol=3)
A <- rep(c(-1,1),2)
B <- rep(c(-1,-1,1,1),1)
Total <- apply(yield,1,sum)
n <- 3
```

After the previous entry following matrix will be shown.

```
     A  B AB Total
[1] -1 -1  1    50
[2]  1 -1 -1    70
[3] -1  1 -1    30
[4]  1  1  1    60
```

Next step is calculating effects of factors.

```
Abeff <- (Total %*% AB)/(2*n)
Effects <- t(Total) %*% cbind(A,B,AB)/(2*n)
Summary <- rbind( cbind(A,B,AB),Effects )
Summary
```

Display after this step should show effects of factors and interaction between factors.

```
            A      B       AB
[1] -1.000000 -1  1.000000
[2]  1.000000 -1 -1.000000
[3] -1.000000  1 -1.000000
[4]  1.000000  1  1.000000
[5]  8.333333 -5  1.666667
```

At the end of the program it is necessary to write codes for analysis of variance.

```
yield.vec <- c(t(yield))
Af <- rep(as.factor(A),rep(3,4))
Bf <- rep(as.factor(B),rep(3,4))
options(contrasts=c(„contr.sum“,“contr.poly
“))
yield.lm <- lm(yield.vec ~ Af*Bf)
options(contrasts=c(„contr.sum“,“contr.poly
“))
yield.lm <- lm(yield.vec ~ Af*Bf)
summary(yield.lm)
```

After running last line of code ANOVA table should be shown.

```
Analysis of Variance Table
Response: yield.vec
          Df  Sum Sq Mean Sq F value
A          1 208.333 208.333 53.1915
B          1  75.000  75.000 19.1489
A:B        1   8.333   8.333  2.1277
Residuals  8  31.333   3.917
```

If compare ANOVA table in R with Table 2 we can see that the output values are the same.

Finally we can writte whole program code for this example.

```
# Factorial Design 2^2

# Input values
yield<-
matrix(c(28,25,27,36,32,32,18,19,23,31,30,2
9),byrow=T,ncol=3)
A <- rep(c(-1,1),2)
B <- rep(c(-1,-1,1,1),1)
Total <- apply(yield,1,sum)
n <- 3
Aeff <- (Total %*% A)/(2*n)
Beff <- (Total %*% B)/(2*n)
AB <- A*B
cbind(A,B,AB,Total)

# Effects
ABeff <- (Total %*% AB)/(2*n)
Effects <- t(Total) %*% cbind(A,B,AB)/(2*n)
Summary <- rbind( cbind(A,B,AB),Effects )
Summary

# ANOVA table
yield.vec <- c(t(yield))
A <- rep(as.factor(A),rep(3,4))
B <- rep(as.factor(B),rep(3,4))
options(contrasts=c("contr.sum","contr.poly
"))
yield.lm <- lm(yield.vec ~ A*B)
options(contrasts=c("contr.sum","contr.poly
"))
yield.lm <- lm(yield.vec ~ A*B)
summary(yield.lm)
anova(yield.lm)
```

As we can see from the previous it is simple to calculate ANOVA for any factorial design example in R and it is much faster than calculating in clasical way.

5. CONCLUSION

Design of experiment is an important scientific field in engineering disciplines. There are various methods in designing of experiment. Engineers and researchers often

use factorial experimental design. For this methods it is important to know and implement calculating of analysis of variance. It is preferably to use softwares to calculate output parameters. In this paper an example of factorial design in software R was presented and as shown it is simple and fast to calculate all necessary parameters for factorial design.

References

[1] Montgomery, D.C., *Design and Analysis of Experiments (8th Edition)*, Arizona State University, John Wiley and Sons.

[2] Hodžić, D., Hodžić, A., Bajramović, E., *Latin square experiment design in R*, IOP Conference Series Materials Science and Engineering. 477. 012019. DOI 10.1088/1757-899X/477/1/012019.

[3] Kerns, G.,J., *Introduction to Probability and Statistics using R*, GNU Free Documentation.

[4] Gerrard, P, Johnson, R.M., *Mastering Scientific Computing with R*, Packt Publishing, Birmingham-Mumbai, 2015

[5] Hurem, N., Hodžić, A., Hodžić, D., *Application of Comparative Experiment in Analysis of Wood Strength*, International Conference on Applied Science ICAS 2020, Hunedoara, Romania, 2020.