STATISTICAL PROCESSING OF EXPERIMENTAL DATA USING ANALYSIS OF VARIANCE

Ioan MILOSAN

"Transilvania" University, Brasov, Romania (milosan@unitbv.ro)

DOI: 10.19062/2247-3173.2016.18.1.67

Abstract: *The differences between samples of observations made by statistics is called analysis of variance (ANOVA). Being an extension of analysis of variance, the two-way ANOVA compares the mean differences between groups that have been split on two independent variables. This method is similar to the process of comparing the statistical difference between two samples, in that it invokes the concept of hypothesis testing. The paper presents an example of calculation for two-way analysis of variance repeated measures applied for the results of strength (Rm) and impact strength (KC) of Ni-Cu and Ni-Cu-Cr austempered ductile iron. Two-factor ANOVA / Two-way ANOVA: is an experiment with two independent variables, call them factor 1 (in our case, the auxiliary alloying elements: Ni-Cu and Ni-Cu-Cr) and factor 2 (in our case, the maintained time at the isothermal level,* t_{iz}*) that has three levels:* t_{iz} *= 5, 30 and 60 minutes. It was selected the level of significance (the default is 5% or 0.05).*

Keywords: *statistics, analysis of variance, strength, impact strength*

1. INTRODUCTION

Any technological process operates simultaneously on several factors, random and systematic, each having an influence on the process performance.

By analysis of variance (ANOVA) are separate effects of random variation factors of the effects of systematic factors (technological parameters), the separation taking place by decomposition of the total variance components and estimate their dispersions [1].

Analysis of variance usually refers to statistical analysis involving simultaneous comparison of multiple sets of observations, not just the comparison of two averages [2].

ANOVA creates a way to test several null hypothesis at the same time. The logic behind this procedure has to do with how much variance there is in the population [3].

In situations where the influences are tested simultaneously to two factors and their interactions on the performance of a technological process, using two factors variance analysis (Two-way ANOVA). Based on research results are selected significant factors and interactions that come as variables in the mathematical model of the process [4, 5].

The paper presents an example of calculation for two-way analysis of variance repeated measures applied for the results of strength (Rm) and impact strength (KC) of Ni-Cu and Ni-Cu-Cr austempered ductile iron.

2. THE STEPS OF THE TWO-WAY ANALYSIS OF VARIANCE

Solving Two-Way Analysis of Variance is done in the following steps [1, 2]:

(1) Enter experimental data

↓

(2) Calculation of the Sum of squares for A factor, SS(A), (i.e. sum of the squared deviations from the mean) is done with the relationship:

$$SS(A) = r * b \sum_{i=1}^{a} (\bar{y}i - \bar{y})^2$$

↓

(3) Calculation of the Sum of squares for B factor, SS(B), (i.e. sum of the squared deviations from the mean) is done with the relationship:

$$SS(B) = r * a \sum_{j=1}^{b} (\bar{y}i - \bar{y})^2$$

↓

(4) Calculation of the Sum of squares for AB factors, SS(AB) is done with the relationship:

$$SS(AB) = r * \sum_{i=1}^{a} * \sum_{j=1}^{b} (\bar{y}ij - \bar{y}i - \bar{y}j + \bar{y})^2$$

↓

(5) Calculation of the captures variability within each group SS(Within) is done with the relationship:

$$SS(Within) = \sum_{i=1}^{a} * \sum_{j=1}^{b} * \sum_{k=1}^{r} (\bar{y}ijk - \bar{y}ij)^2$$

↓

(6) Calculation of the Total sum of squares, SS(Total), is done with the relationship:

$$SS(Total) = \sum_{i=1}^{a} * \sum_{j=1}^{b} * \sum_{k=1}^{r} (\bar{y}ijk - \bar{y})^2$$

↓

a

a

↓

(7) Calculation of the Source of Variation for factor A, is done with the relationship:

$MS_A=SS_A/df_A$; $F=MS_A/MS_w$; $df_A=a-1$

↓

(8) Calculation of the Source of Variation for factor B, is done with the relationship:

$MS_B=SS_B/df_B$; $F=MS_B/MS_w$; $df_B=b-1$

↓

(9) Calculation of the Source of Variation for AB interaction, is done with the relationship:

$MS_{AB}=SS_{AB}/df_{AB}$; $F=MS_{AB}/MS_w$; $df_{AB}= (a-1)(b-1)$

↓

(10) Calculation of the Source of Variation for Within groups, is done with the relationship:

$MS_w=SS_w/df_w$; $df_w=ab(r-1)$

↓

(11) Comparing the two values of the *F* test statistic: calculate value (*F*) and critical value (F-crit).

where:

Factor A has 1, 2, ..., a levels;

Factor B has 1, 2, ..., b levels;

SS(A) = Sum of squares for A factor (i.e. sum of the squared deviations from the mean);

SS(B)= Sum of squares for B factor (i.e. sum of the squared deviations from the mean);

SS (AB)= Sum of squares for AB factors (i.e. sum of the squared deviations from the mean);

AB = the interaction between A and B.

SS(Within)= captures variability within each group;

SS(Total)= Total sum of squares

a = levels for A factor;

b = levels for B factor;

r = total number of observations for each interaction;
abr = n = the total number of observations;
Y = process performance;
$\bar{Y}$ = average of process variable;
i = levels of A factor;
j = levels of B factor;
k = levels of interactions;
df = the degrees of freedom is equal to the sum of the individual degrees of freedom for each sample.
abr = n = the total number of observations;
SS = Sum of squares (i.e. sum of the squared deviations from the mean);
MS = Mean square;
F = the calculated value of the Fischer criteria;
F-crit = is the critical value of the Fischer criteria: for the level of significance, α = 0.05 (the default is 5% or 0.05.), the critical value for F;
P-value is determined from the F ratio and the two values for degrees of freedom;

3. EXPERIMENTAL INVESTIGATION

3.1. Materials

The first studied materials (Lot A), was a Ni-Cu cast iron with the following composition (% in weight): 3.80 %C; 2.60 %Si; 0.45 %Mn; 0.005 %P; 0.001 %S; 0.072 %Mg; 0.40 %Ni and 0.18 %Cu. The second materials (Lot B), was a Ni-Cu-Cr cast iron with the following composition (% in weight): 3.61% C; 2.67% Si; 0.53 % Mn; 0.011%P; 0.005%S; 0.06%Mg; 0.45% Ni; 0.22%Cu and 0.20% Cr.

This materials were subjected to a heat treatment whose parameters have been:

- the austenizing temperature, T_A = 900 [°C];
- the maintained time at austenizing temperature, τ_A = 30 [min];
- the temperature at isothermal level, T_{iz} = 300 [°C];
- the maintained time at the isothermal level, τ_{iz} = 5, 30 and 60 [min], corresponding to the first stage for obtaining bainitic structure.

All these experimental specimens, were performed at isothermal maintenance in salt-bath (55% KNO3+45% NaNO3), being the cooling after the isothermal maintenance was done in air.

From this material, 36 typical test specimens were done: 18 typical test specimens for strength (Rm) determination and 18 typical test specimens for impact strength (KC) determination. For each Rm and KC determination it was done three parallel determinations (r = 3).

3.2. Results

The values of the mechanical results are presented in tables 1.

It can be certainly observed a normal evolution of the values for mechanical characteristics [6, 7]:

-when maintaining time at the isothermal level for both Lots (A and B) is growing, then strength (R_m) decreases and impact strength (KC) (unnotched samples) increases;

-when maintaining time at the same temperature of the isothermal level is increasing, then Rm decreases and KC increases;

Table 1. Data Analysis of R_m and KC values for T_A= 900°C and T_{iz}=300°C

Materials	τ_{iz} [min]	R_m [N /mm ^{2}] parallel observations			KC [J /cm^2] parallel observations		
		Obs. 1	Obs. 2	Obs. 3	Obs. 1	Obs. 2	Obs. 3
Ni-Cu (Lot A)	5	1420	1425	1423	56	57	56
	30	1390	1390	1393	61	60	61
	60	1300	1309	1311	67	68	68
Ni-Cu-Cr (Lot B)	5	1454	1450	1450	47	46	46
	30	1425	1425	1421	53	54	52
	60	1395	1390	1390	57	56	56

- comparing the 2 groups, we see that for the same parameters of heat treatment, Lot B (Ni-Cu-Cr) has sensible superior values for strength and inferior values for impact strength comparing with Lot A (Ni-Cu).

This evolution of the mechanical properties is determined by the structural constituents for each chemical composition, which can be constituted of inferior bainite, residual austenite and martensite. By allying with additional Cr, the complexes carbides made, will induce higher values sensitive to strength (R_m) and lower for impact strength (KC).

4. CALCULATION OF TWO-WAY ANOVA

Two-factor ANOVA / Two-way ANOVA: is an experiment with two independent variables, call them factor 1 (in our case, the auxiliary alloying elements: Ni-Cu and Ni-Cu-Cr) and factor 2 (in our case, the maintained time at the isothermal level, t_{iz}) that has three levels: t_{iz} = 5, 30 and 60 minutes.

It was selected the level of significance (the default is 5% or 0.05). The R_m and KC Summary Table of the ANOVA- Two Factor With Replication are presented in Tables 2 and 3.

Table 2.Summary Table of the R_m ANOVA- Two Factor With Replication

SUMMARY	**Sample 5**	**Sample 30**	**Sample 60**	**Total**
Ni-Cu (lot A)				
Count	3	3	3	9
Sum	4268	4173	3920	12361
Average	1422,667	1391	1306,667	1373,444
Variance	6,333333	3	34,33333	2707,278
SUMMARY	**Sample 30**	**Sample 60**	**Sample 90**	
Ni-Cu-Cr (lot B)				
Count	3	3	3	9
Sum	4354	4271	4175	12800
Average	1451,333	1423,667	1391,667	1422,222
Variance	5,333333	5,333333	8,333333	673,4444
TOTAL	**Sample 30**	**Sample 60**	**Sample 90**	
Count	6	6	6	
Sum	8622	8444	8095	
Average	1437	1407,333	1349,167	
Variance	251,2	323,4667	2184,567	

Table 3.Summary Table of the KC ANOVA -Two Factor With Replication

SUMMARY	**Sample 5**	**Sample 30**	**Sample 60**	**Total**
Ni-Cu (lot A)				
Count	3	3	3	9
Sum	169	182	203	554
Average	56,33333	60,66667	67,66667	61,55556
Variance	0,333333	0,333333	0,333333	24,77778
SUMMARY	**Sample 30**	**Sample 60**	**Sample 90**	
Ni-Cu-Cr (lot B)				
Count	3	3	3	9
Sum	139	159	169	467
Average	46,33333	53	56,33333	51,88889
Variance	0,333333	1	0,333333	19,86111
TOTAL	**Sample 30**	**Sample 60**	**Sample 90**	
Count	6	6	6	
Sum	308	341	372	
Average	51,33333	56,83333	62	
Variance	30,26667	18,16667	38,8	

The Source of Variation of the ANOVA- Two Factor With Replication are presented in Tables 4 and 5, where:

F is the calculated value of the Fischer criteria; $F_{Sample} = MS_{Sample} / MS_{Within}$; $F_{Columns} = MS_{Columns} / MS_{Within}$; $F_{Interaction} = MS_{Interaction} / MS_{Within}$;

F-crit = is the critical value of the Fischer criteria: for the level of significance, α = 0.05 (the default is 5% or 0.05.), the critical value for F;

F-crit for Sample with df (1, 16) is 4.493998;

F-crit for Columns with df (3, 16) is 3.238872;

F-crit for Interaction with df (3, 16) is 3.238872.

Table 4.Source of variation of the R_m ANOVA, T_A = 900°C and T_{iz}=300°C

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F-crit
Sample	10706,72	1	10706,72	1025,112	5,44E-13	4,747225
Columns	23956,33	2	11978,17	1146,846	1,99E-14	3,885294
Interaction	2964,111	2	1482,056	141,8989	4,46E-09	3,885294
Within	125,3333	12	10,44444			
Total	37752,5	17				

Table 5.Source of variation of the KC ANOVA, T_A = 900°C and T_{iz}=300°C

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F-crit
Sample	420,5	1	420,5	946,125	8,75E-13	4,747225
Columns	341,4444	2	170,7222	384,125	1,32E-11	3,885294
Interaction	10,33333	2	5,166667	11,625	0,001556	3,885294
Within	5,333333	12	0,444444			
Total	777,6111	17				

In figure 1 is presented the influence of the source of variation over the values of the Fischer Criteria.

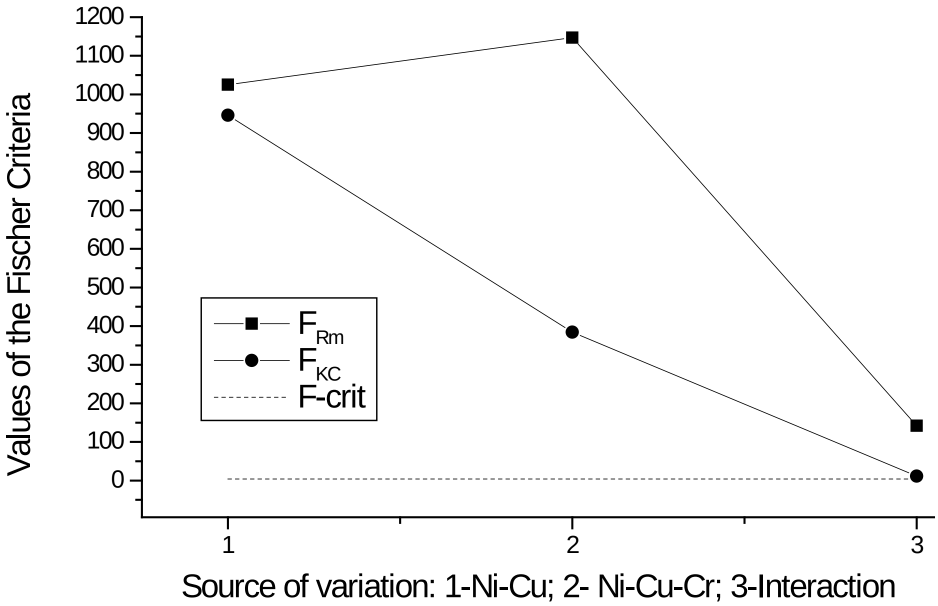

FIG. 1. The influence of the source of variation over the values of the Fischer Criteria.

The following observations can be made after analyzing the results presented in table 4, 5 and figure 1:

(a) From table 4, source of variation of the R_m ANOVA- Two Factor With Replication, T_A = 900°C and T_{iz}=300°C, it was compares the two values of the F test statistic: calculate value (F) and critical value (F-crit). Note that in this case the calculated value (F) is greater than the critical value (F-crit), i.e.:

- analyzing the Lot A, is observed for Ni-Cu alloying, that the calculated value F = 1025.112 is greater than the critical value (F-crit) = 4.747225 (1025.112 > 4.747225), therefore, the effect of the first independent variable is significant;

- analyzing the Lot B, observed for Ni-Cu-Cr alloying, that the calculated value F = 1146.846 is greater than the critical value (F-crit) = 3.885294 (1146.846 > 3.885294), therefore, the effect of the first independent variable is also significant;

- analyzing the effect of interaction between the two lots, we see that for Rm, the calculated value F = 141.8989 is greater than the critical value (F-crit) = 3.885294 (141.8989 > 3.885294), therefore, so both lots are the parameters for the technological process.

(b) From table 5, source of variation of the KC ANOVA- Two Factor With Replication, T_A = 900°C and T_{iz}=300°C, it was compares the two values of the F test statistic: calculate value (F) and critical value (F-crit). Note that in this case the calculated value (F) is greater than the critical value (F-crit), i.e.:

- analyzing the Lot A, is observed for Ni-Cu alloying, that the calculated value F = 944.125 is greater than the critical value (F-crit) = 4.747225 (944.125 > 4.747225), therefore, the effect of the first independent variable is significant;

- analyzing the Lot B, observed for Ni-Cu-Cr alloying, that the calculated value F = 384.125 is greater than the critical value (F-crit) = 3.885294 (384.125 > 3.885294), therefore, the effect of the first independent variable is also significant;

- analyzing the effect of interaction between the two lots, we see that for Rm, the calculated value F = 11.625 is greater than the critical value (F-crit) = 3.885294 (11.625 > 3.885294), therefore, so both lots are the parameters for the technological process.

(c) From table 4, 5 and figure 1, the following observations can be made:

- the values of the Fischer Criteria for the Rm properties is bigger for all the source of variation comparing with the values of the Fischer Criteria for the KC properties;

- both values of the Interaction are bigger than the critical value from the Fischer Criteria (F-crit = 3,885294), so both variables are technological factors for the studied materials (lot A and lot B).

CONCLUSIONS

By studying all the data presented in this paper following remarkable conclusions:

(a) Analysis of variance usually refers to statistical analysis involving simultaneous comparison of multiple sets of observations, not just the comparison of two averages;

(b) In situations where the influences are tested simultaneously to two factors and their interactions on the performance of a technological process, using two factors variance analysis (Two-way ANOVA).

(c) The evolution of the mechanical properties is determined by the structural constituents for each chemical composition, which can be constituted of inferior bainite, residual austenite and martensite. By allying with additional Cr, the complexes carbides made, will induce higher values sensitive to strength (R_m) and lower for impact strength (KC).

(d) Analyzing the effect of interaction between the two independent variables, we see that:

- the values of the Fischer Criteria for the Rm properties is bigger for all the source of variation comparing with the values of the Fischer Criteria for the KC properties;

- both values of the Interaction are bigger than the critical value from the Fischer Criteria (F-crit = 3,885294), so both variables are technological factors for the studied materials (lot A and lot B).

REFERENCES

[1] K. Hae-Young, Statistical notes for clinical researchers: Two-way analysis of variance (ANOVA)-exploring possible interaction between factors, *Restor Dent Endod.*, vol 39, no. 2, pp. 143-147, 2014;

[2] R.A. Armstrong, F. Eperjesi and B. Gilmartin, The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthalmic and Physiological Optics*, vol. 22, no. 3, pp. 248-256, 2002;

[3] D. Taloi, *Optimization of technological processes*, E.A. Bucharest, 1984;

[4] J. Budde, Variance analysis and linear contracts in agencies with distorted performance measures. *Management Accounting Research*, vol. 20, no. 3, pp. 166-176, 2009;

[5] K.T. Wing, Using enhanced cost models in variance analysis for better control and decision making, *Management Accounting Quarterly*, vol. 4, no. 2, pp. 27-35, 2005;

[6] R.A. Harding, The production, properties and automotive applications of austempered ductile iron. *Kovove Materialy*, vol. 45, pp.1-16, 2007;

[7] A. Rimmer, DI solutions aid vehicle design, *The Foundry Trade Journal*, vol. 2, pp. 54-56, 2004;