"HENRI COANDA"
AIR FORCE ACADEMY
ROMANIA

"GENERAL M.R. STEFANIK"
ARMED FORCES ACADEMY
SLOVAK REPUBLIC

INTERNATIONAL CONFERENCE of SCIENTIFIC PAPER
AFASES 2015
Brasov, 28-30 May 2015

MATHEMATICAL MODELING BY USING A C++ SOFTWARE

Ioan Milosan*

*Faculty of Materials Science and Engineering, Transilvania University of Brasov, Romania

Abstract: *The aim of the study is to achieve a easy calculation of the linear regression (mathematical modeling) using of a C++ software. This application focused on the fitting and checking of linear regression models, using small and large data sets, with computers. Were constructed logical steps of the procedure and software based on their achievements C++ was made. The performance of regression analysis methods in practice depends on the form of the data generating process, and how it relates to the regression approach being used. It was used some statistical criteria as: Cochran, Student and Fischer criteria. After solving statistical analysis of the linear regression models, in the end there was obtained an applied statistical analysis of the linear regression model through the use of C++ software.*

Keywords: *analytical models; computer aided software engineering, heat treatment, input variables, mechanical properties.*
MSC2010: *15A06, 49K05, 68N17, 97P30, 97Q60, 97R30.*

1. INTRODUCTION

Mathematical modeling aims to de- scribe the different aspects of the real world, their interaction, and their dynamics through mathematics [1].

Today, mathematical modeling has an important key role in industrial fields and especially in complex metallurgical processes.

Linear regression is a statistical technique that is used to learn more about the relationship between a dependent variable and one or more independent variables [2, 3].

Linear regression equations with one or more variables were obtained under the assumption that the formula $\tilde{Y} = f(X_1, X_2 \ldots X_n)$ was known in relationships physical analysis of the problem. There are many situations in industrial processes in which this assumption not confirmed; therefore equations obtained by regression analysis are subject to a statistical analysis of the regression equation to determine whether or not consistent with experimental data [4, 5, 6, 7].

The aim of the study is to achieve a easy calculation of the linear regression (mathematical modeling) using of a C++ software for all industrial processes and in particularly for the statistical analysis of the regression equation applied to the mechanical results of a special S.G. cast iron.

2. METHOD OF CALCULATING

Solving statistical analysis of the linear regression models is done in the following steps:

(1) Enter experimental data

(2) Linear regression model building, $\tilde{y} = b_0 + b_1(x_i - \bar{x})$;

(3) Calculating the b_0 and b_1 coefficients from the regression of the mathematical equation, with the relationships: $b_0 = \frac{1}{K}\sum_{i=1}^{K}\bar{y}_i$;

$b_1 =$ $\sum_{i=1}^{K} x_i\bar{y}_i - \bar{x}\cdot\sum_{i=1}^{K}\bar{y}_i$

(4) With coefficients b_0 and b_1 can calculate the mathematical model linear regression model building,

$$\tilde{y} = b_0 + b_1(x_i - \bar{x});$$

(5) Determining the experimental dispersion, S_i^2, with the relationships: $S_i^2 = \frac{\sum_{i=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{\nu_i}$

$\nu_i = n_i - 1$

(6) The value of the Cochran criterion for the parallel determinations (y_1, y_2 and y_3), is calculated:

$$G_c = \frac{Si^2_{\max}}{\sum_{i=1}^{K} S_i^2}$$

(7) Verifying the homogeneity of the dispersions with the help of the Cochran criteria: $G_T = G_{\alpha:\nu:k}$

A

B

A

B

(8.a) If $G_C < G_T$, the dispersions are homogeneous

(8.b) If $G_C > G_T$, the dispersions are not homogeneous, *calculation stops*

(9) Will determine the reproducibility dispersion $S_0{}^2$, with the relationship:

$$S_0{}^2 = \frac{1}{k} \cdot \sum_{i=1}^{K} S_i^2$$

(10) The value of the Student criteria is calculated:

$$t_{b0} = \frac{|b_0|}{Sb_0};\ t_{b1} = \frac{|b_1|}{Sb_1};\ Sb_0 = \sqrt{Sb_0^2};\ Sb_1 = \sqrt{Sb_1^2};\ Sb_0^2 = \frac{S^2}{N};$$

$$S^2 = \frac{1}{N-l} \cdot \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(y_{ij} - \tilde{y}_i\right)^2;\ N = k \cdot n_i;\ Sb_1^2 = \frac{S^2}{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}$$

(11) Verifying the coefficient signification b_o and b_1 with the help of the critical value of the Student criteria, t_T :

$$t_T = t_{\alpha,\nu}$$

C

D

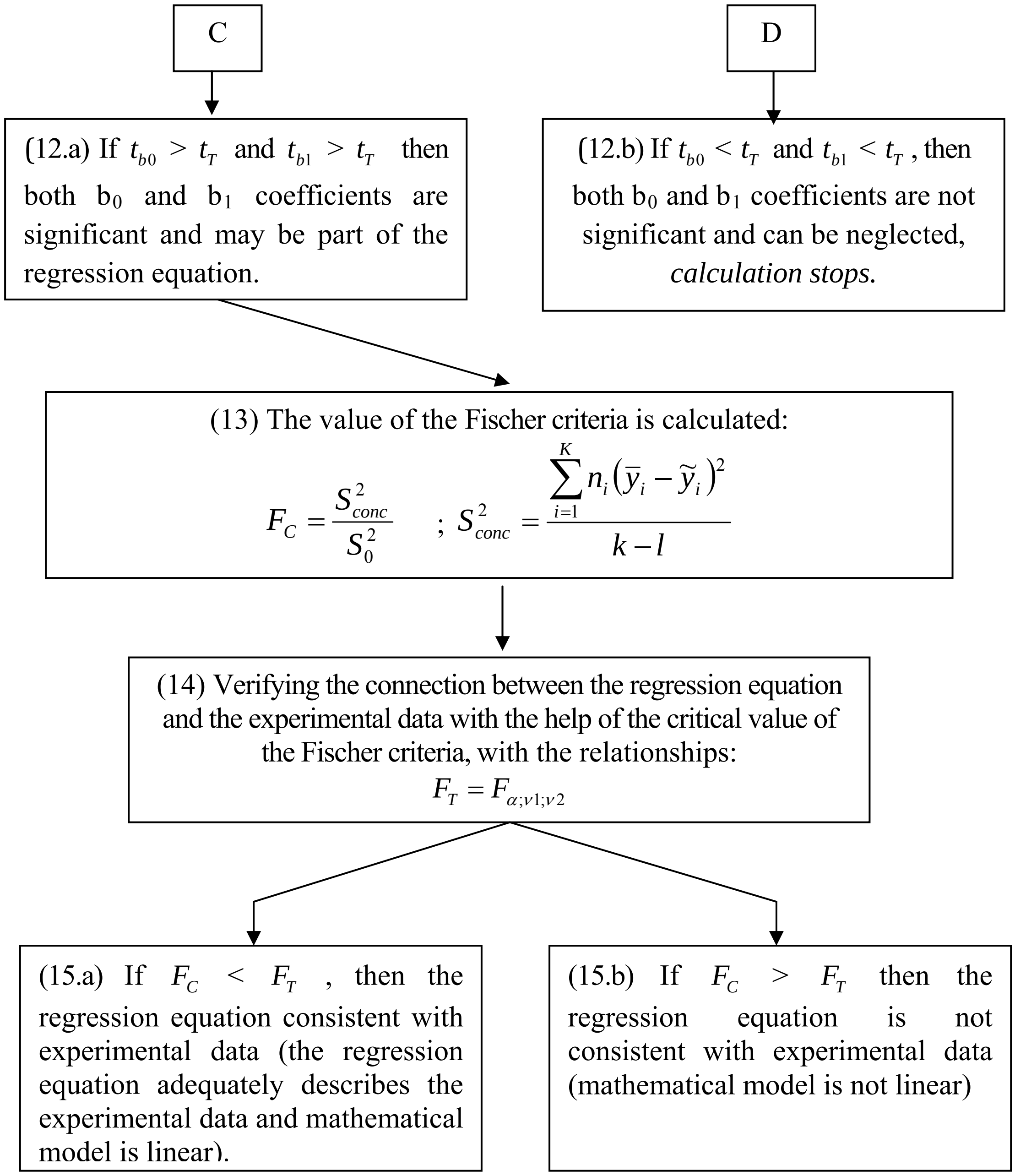


Fig. 1. Steps for calculating the statistical analysis of the linear regression models.

where:

- $\tilde{y}$ is the mathematical model;
- x_i is the process variable;
- $\bar{x}$ is the average of process variables;
- b_0 is the slope coefficient from the regression of the mathematical equation, indicating the magnitude and direction of that relation;
- b_1 is the intercept coefficient from the regression of the mathematical equation, indicating the status of the dependent variable when the independent variable is absent.
- b_0 is the slope coefficient from the regression of the mathematical equation, indicating the magnitude and direction of that relation;
- b_1 is the intercept coefficient from the regression of the mathematical equation, indicating the status of the dependent variable when the independent variable is absent;
- K is the number of experimental points;
- $\bar{y}_i$ is the average of process performances;
- S_i^2 is the experimental dispersion;

- y_{ij} is the process performances;
- $\bar{y}_i$ is the average of process performances;
- ν is the number of degrees of freedom;
- n_i is the number of parallel determinations performed each experimental points k.
- G_c is the calculated value of the Cochran criteria;
- G_T is the critical value of the Cochran criteria;
- $S^2_{i\,max}$ is the maximal experimental dispersion from the K number of experimental points;
- S_i^2 is the experimental dispersion;
- $\alpha = 0{,}05$ is the coefficient of statistical significance level;
- ν is the number of degrees of freedom, $\nu = n_i - 1$;
- n_i is the number of parallel determinations performed each experimental points k.
- Sb_0^2 is the dispersion of b_0 coefficient;
- Sb_1^2 is the dispersion of b_1 coefficient;
- S^2 is the theoretical dispersion;
- *N* is the total number of determinations;
- *l* is the number of coefficients of the regression equation (l = 2)
- $\tilde{y}_i$ is the mathematical model for each "i" point;
- y_{ij} is the process performances;
- t_T is the critical value of the Student criteria;
- $\alpha = 0{,}05$ is the coefficient of statistical significance level;
- ν is the number of degrees of freedom, ν = N-2;
- S^2_{conc} is the dispersion according;
- n_i is the number of parallel determinations performed each experimental points k.
- $\bar{y}_i$ is the average of process performances;
- $\tilde{y}_i$ is the mathematical model for each "i" point;
- F_T is the critical value of the Fischer criteria;
- $\alpha = 0{,}05$ is the coefficient of statistical significance level;
- ν_1 is the number of degrees of freedom, ν_1 = k-l;
- ν_2 is the number of degrees of freedom, ν_2 = k;

3. EXPERIMENTAL RESEARCHES

The results were obtained after performing some KCU properties tests on samples pieces in the case of an austempered ductile iron. The studied cast iron has the following chemical composition (% in weight): 3.75% C; 2.14% Si; 0.4 % Mn; 0.012%P; 0.003%S; 0.05%Mg; 0.40% Ni, 0.42%Cu. This cast iron was made in an induction furnace.

The parameters of the heat treatment done were the following: the austenizing temperature, t_A = 900[°C], the maintained time at austenizing temperature was, τ_A = 30 [min]; the temperature at isothermal level, t_{iz} = 300[°C]; the maintained time at the isothermal level, τ_{iz} = 10, 20, 30, 40, 50 and 60 [min]. The experimental samples were performed at isothermal maintenance in salt-bath, being the cooling after the isothermal maintenance was done in air.

After the heat treating there were used 18 samples for each determining the impact strength (KCU).

The values of experimental points, the process variable, the average of process variable and the process performances (the values of impact strength, KCU) are presented in table 1. The process variable (xi) was the maintaining time at isothermal heat treatment and for each six samples it was made three parallel determinations (y_1, y_2 and y_3).

Table 1. The values of experimental points, the process variable, the average of process variable and the process performances (the values of impact strength)

Experimental points	Process variable	Average of process variables	Process performances, KCU [J / cm^2]		
k	x_i	$\bar{X}_i$	y_1	y_2	y_3
1	10		22	24	23
2	20		28	26	27
3	30		32	30	31
4	40	35	35	36	35
5	50		38	39	39
6	60		41	43	43

For easy calculation of the statistical analysis regression equation will work spreadsheet, considering the above six steps work, which is presented in table 2.

Table 2. Tabular presentation of data taken into account

k	$\bar{y}_i$	$x_i\bar{y}_i$	$(x_i-\bar{X}_i)^2$	$(y_{ij}-\bar{y}_i)^2$	S_i^2	$\hat{y}_i$	$(y_{ij}-\hat{y}_i)^2$	$(\bar{y}_i-\hat{y}_i)^2$
1	23	230	625	2.0000	1.0000	3.699	23.1746	0.0305
2	27	540	225	2.0000	1.0000	4.375	27.0603	0.0036
3	31	930	25	2.0000	1.0000	5.051	30.9460	0.0029
4	35.3333	1413.3333	25	0.6667	0.3333	5.727	34.8317	0.2516
5	38.6667	1933.3333	225	0.6667	0.3333	6.403	38.7175	0.0026
6	42.3333	2540	625	2.6667	1.3333	7.079	42.6032	0.0728
Σ	197.333	7586.6666	1750	9.3334	4.6666	32.334	197.3333	0.3640

Solving statistical analysis of the linear regression models in this case is done using the 15 steps above.

4. CALCULATING THE STATISTICAL ANALYSIS OF THE REGRESSION EQUATION WITH C++ SOFTWARE

Solving statistical analysis of the linear regression models with C++ software is done in the following steps:

(1) Input data into the program, actually presented in figure 2.

(2) Output data into the program, actually presented in figure 3.

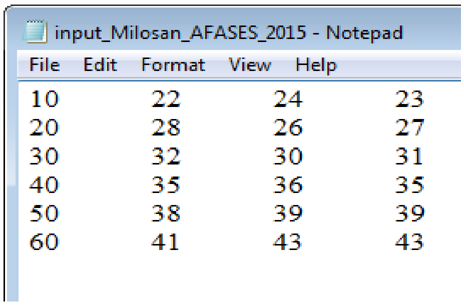
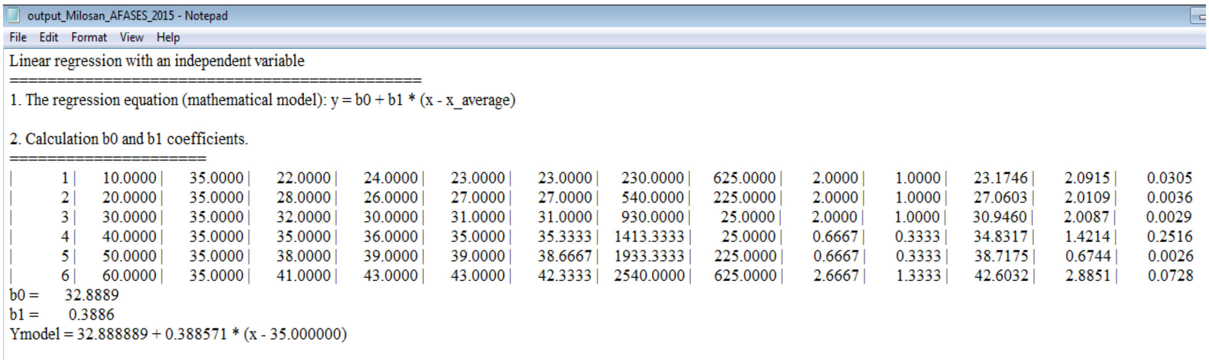


Fig. 2. Input data into the program

(3) Running Windows Commander data to determine all values in accordance with the working steps of the method, presented in figure 4.



output_Milosan_AFASES_2015 - Notepad

File Edit Format View Help

Linear regression with an independent variable

==

1. The regression equation (mathematical model): y = b0 + b1 * (x - x_average)

2. Calculation b0 and b1 coefficients.

=====================

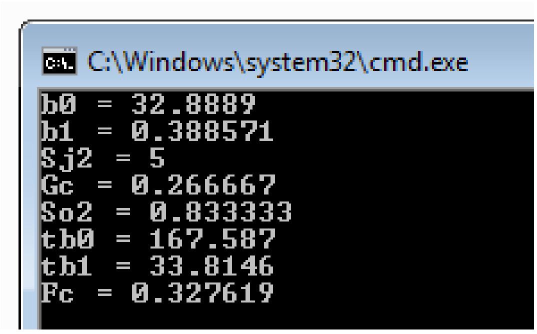
1	10.0000	35.0000	22.0000	24.0000	23.0000	23.0000	230.0000	625.0000	2.0000	1.0000	23.1746	2.0915	0.0305
2	20.0000	35.0000	28.0000	26.0000	27.0000	27.0000	540.0000	225.0000	2.0000	1.0000	27.0603	2.0109	0.0036
3	30.0000	35.0000	32.0000	30.0000	31.0000	31.0000	930.0000	25.0000	2.0000	1.0000	30.9460	2.0087	0.0029
4	40.0000	35.0000	35.0000	36.0000	35.0000	35.3333	1413.3333	25.0000	0.6667	0.3333	34.8317	1.4214	0.2516
5	50.0000	35.0000	38.0000	39.0000	39.0000	38.6667	1933.3333	225.0000	0.6667	0.3333	38.7175	0.6744	0.0026
6	60.0000	35.0000	41.0000	43.0000	43.0000	42.3333	2540.0000	625.0000	2.6667	1.3333	42.6032	2.8851	0.0728

b0 = 32.8889

b1 = 0.3886

Ymodel = 32.888889 + 0.388571 * (x - 35.000000)

Fig.3. Output data into the program.



C:\Windows\system32\cmd.exe

```
b0 = 32.8889
b1 = 0.388571
Sj2 = 5
Gc = 0.266667
So2 = 0.833333
tb0 = 167.587
tb1 = 33.8146
Fc = 0.327619
```

Fig. 4. Running Windows Commander data.

Solving statistical analysis of the linear regression models in this case is done in the following six steps, considering the data presented in figure 2 and calculated with the C++ software:

(1) Linear regression model building,

$\tilde{y} = b_0 + b_1(x_i - \bar{x})$;

(2) After calculating the b_0 and b_1 coefficients from the regression of the mathematical equation, there has been obtained the following results: $b_0 = 32.8889$; $b_1 = 0.388571$, and the mathematical model is:

$\tilde{y} = 19.288904 - 0,388571 \cdot x_i$

(3) After calculating the experimental dispersion, S_i^2 there has been obtained the following results: $S_i^2 = 5$;

(4) After calculating the homogeneity of the dispersions with the help of the Cochran criteria, there has been obtained the following results: $G_c = 0.266667$ and $G_c < G_T$ respectively $0.4266667 < 0.6161$. Because $G_C < G_T$ the dispersions are homogeneous and the dispersion value was determined: $S_0{}^2 = 0.833333$.

(5) After verifying the coefficient signification b_o and b_1 with the help of the Student criteria, there has been obtained the following results: $t_{b0} > t_T$ result $167.587 > 2.12$ and $t_{b1} > t_T$, result $33.8146 > 2.12$, and then both b_0 and b_1 coefficients are significant and may be part of the regression equation.

(6) After verifying the connection between the regression equation and the experimental data with the help of the Fischer criteria, there has been obtained the following results: $F_C < F_T$ result $0.327619 < 4.53$, therefore the regression equation consistent with experimental data (the regression equation adequately describes the experimental data and mathematical model is linear).

5. CONCLUSIONS

(a) The regression analysis is one of the most widely used statistical tools to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.

(b) The performance of regression analysis methods in practice depends on the form of the data generating process, and how it relates to the regression approach being used.

(c) After verifying the connection between the regression equation and the experimental data with the help of the Fischer criteria, it is noted that the regression equation consistent with experimental data (the regression equation adequately describes the experimental data and mathematical model is linear).

(d) By using C++ software we obtained more accurate results and the application time was reduced by several hours (for the classical calculation) to 2-3 minutes.

REFERENCES

1. Quarteroni, A., Mathematical Models in Science and Engineering, *American Mathematical Society*, 56, 1, 10-19 (2009).
2. Evans, J.R., Olson, D.L. *Statistics, Data Analysis and Decision Mo*deling, 2nd edition, Prentice Hall, New Jersey, USA (2003).
3. Akaike, H., A New Look at Statistical Modek Identification, *IEEE Trans. Auto Control*, 19, 716-723 (1974).
4. Lind, D.A., Marchal, W.G., Mason, R.D., *Statistical Techniques in Busines & Economics,* 11th edition, Mc Graw Inc., New York, USA (2005).
5. Kenkel, J.L., *Introductory Statistics for Management and Economics*. 4th edition. Duxbury Press, Wadsworth Publishing Company, New York, USA (1996).
6. Taloi, D., *Metallurgical Process Optimization- Applications in metallurgy,* E.D.P. Publishing Company, Bucharest, Romania (1984).
7. Crossman, A., Linear Regression Analysis. Linear Regression and Multiple Linear Regression, *Source* [online]. Available from: http://sociology.about.com/od/Statistics/a/Linear-Regression-Analysis.htmm (March, 2015).
8. Eijndhoven, S. Mathematical models in industrial context, Design of mathematical models, *Source* [online]. Available: https://static.tue.nl/uploads/media/2.1_Mathematical_models_in_industrial_context.pdf , (March, 2015).