APPLICATION OF THE CHAUVENET CRITERION TO DETECTION OF THE ABERRANT DATA OBTAINED IN THE INDUSTRIAL PROCESSES

Ioan MILOSAN

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

DOI: 10.19062/2247-3173.2017.19.1.47

Abstract: *In scientific research, there are cases of experimental errors that can cause problems in interpreting the results. Different statistical criteria are used to identify and correct these errors. In this paper, experimental hardness results of a heat treated Cr-Cu cast iron , are analyzed using the Chauvenet criterion.*

Keywords: *Chauvenet criterion, aberrant data, cast iron, heat treatment, hardness*

1. INTRODUCTION

Statistical verification of experimental data in any area of research is one of the researchers' priorities.

Experimental data to be used in mathematical modulation may also contain abnormal results (errors, disparate results, aberrant results, etc.) that, if not identified and removed, can lead to mathematical models that do not correspond to reality and That is, different tests for the validation of experimental data, tests that fall within the field of statistical processing of experimental data, are used.

One of the most criterion for the validation of experimental data is the Chauvenet criterion. The Chauvet abnormal outcomes criterion is based on the consideration that any value whose probability of occurrence is less than a limit value that depends on the "n" number of results, should be eliminated [1, 2, 4, 5, 7, 8].

This paper presents the main steps in the correct application of the detection of abnormal results and their removal from the experimental data base.

2. APPLICATION OF THE CHAUVENET CRITERION

Application of the Chauvenet criterion to detection and eliminating aberrant results obtained in the industrial processes, is carried out with the help of the statistical processing of the research results, as follows [1, 5, 8]:

(1) Data grouping is performed and extreme values are determined: the minimum (x_{min}) and maximum (x_{max}) of data experimental values;

(2) Arithmetic average is calculated [2]:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + .. + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

(3) Dispersion is calculated [2]:

$$s^2=\frac{1}{n-1}[(x_1-\bar{x})^2+..+(x_n-\bar{x})^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad (2)$$

(4) Standard deviation is calculated [2]:

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (3)$$

(5) Standard deviation from the mean for all suspected outliers is calculated [2]:

$$\tau_{sus}=\frac{|x_{sus}-\bar{x}\ |}{s} \qquad (4)$$

(6) Choosing the critical value of the Chauvenet criterion (τ_{crit}), according to the coefficient of statistical confidence level (α) and the number of determinations (n);

(7) Comparing the number of standard deviation of Chauvenet criterion for the suspect outliers (τ_{sus}), with the critical value (τ_{crit}) of the Chauvenet criterion.

where:

n = number of experimental data;

$\bar{x}$ = arithmetic average of the experimental data;

x_{sus} = the suspected date to be abnormal (outlier) from the experimental determinations;

s^2 = the dispersion of the experimental data;

s = the standard deviation of the experimental data;

τ_{sus}= the number of standard deviation from the mean for all suspected outliers;

τ_{crit} = the critical value (limit value) of the Chauvenet criterion, that depends on the "n" number of results [3];

α = the confidence coefficient; $\alpha = 0.95$ [7];

After comparing the two values, there may be two Chauvenet criterion specific situations:

(a) If $\tau_{sus} > \tau_{crit}$, the values of the x_{sus} is affected by aberrant errors, that value is eliminated from the sample (x_{sus} is an outlier);

In this care, we recalculate the values of the average ($\bar{x}$) and standard deviation (s) for the remained values and we start again with verifying the (a) condition, the algorithm is applied until that condition is no longer verified for any of the two external values of the sample (x_{min} or x_{max}).

b) If $\tau_{sus} < \tau_{crit}$, the values of the x_{sus} it is a noticeable value for the experimental data set.

3. MATERIALS AND RESULTS

3.1. Materials

The material on which experimental research has been performed has the following chemical composition: 3.61 %C; 2.22 %Si; 0.44 %Mn; 0.012%P; 0.003 %S; 0.06 %Mg; 0.42 %Cr and 0.21 %Cu.

The specific parameters of the heat treatment applied were:

- the austenizing temperature, t_A = 900 [°C];
- the maintained time at austenizing temperature, τ_A = 60 [min];
- the temperature at isothermal level, t_{iz} = 300[°C] for lot A and 400[°C] for lot B;

- the maintained time at the isothermal level, τ_{iz} = 60 [min], corresponding to the first stage for obtaining bainitic structure.

Isothermal maintenance was carried out in a salt bath (55% KNO_3+45% $NaNO_3$), subsequent cooling was done in the air.

From this material, 2 typical test specimens for hardness (HB) (100x10x10mm) were done: 1 for lot A and 1 for lot B.

3.2. Results

In table 1 are presented the mechanical results.

Table 1. Mechanical results of hardness [HB] values

Lot	t_A [°C]	τ_A [min]	t_{iz} [°C]	τ_{iz} [min]	Hardness [HB] (parallel determinations)
A	900	60	300	60	404; 426; 415; 398; 363; 390; 420;415;
B			400		331; 323; 302; 331; 315; 375; 315; 302;

It is noted that it is the structure that determines the values of the mechanical properties, specific to each thermal treatment parameter [3, 6]:

- specific to lot A (t_{iz} = 300 °C), is the structure consisting of inferior bainite, residual austenite and martensite, constituents that determine higher values of hardness (HB);

- specific to lot B (t_{iz} = 400 °C), is the structure consisting of superior bainite, and residual austenite, constituents that determine low values of hardness (HB);

- together with increasing the level of the isothermal maintenance temperature inside the structure will appear the superior bainite and the martensite will disappear (lot B).

4. APPLICATION OF THE CHAUVENET CRITERION TO DETECTION OF THE ABERRANT DATA

Experimental data of hardness (HB) with sixth parallel determinations of each experimental lots (A and B), presented in table 1, are checked using Chauvenet criterion [1, 2, 4, 5, 7, 8]. Chauvenet criterion for data sets are presented in table 2.

Table 2. Chauvenet criterion for data sets

Lot	Parallel determination [HB]	n	X	$\bar{x}$	s^2	s	τ_{sus}	τ_{crit}	Values remains Yes/No
A	404; 426; 415; 398; 363; 390; 420; 415;	8	x_{max}=426	403.875	413.554	20.336	1.088	1.863	Yes
			x_{min}=363				*2.010*		*No*
	404; 426; 415; 398; 390; 420; 426	7	x_{max}=426	409.714	164.238	12.816	1.271	1.803	Yes
			x_{min}=390				1.538		Yes
B	331; 323; 302; 331; 315; 375; 315; 302;	8	*x_{max}=375*	324.250	547.071	23.389	*2.169*	1.863	*No*
			x_{min}=302				0.951		Yes
	331; 323; 302; 331; 315; 315; 302;	7	x_{max}=331	317.000	147.667	12.152	1.152	1.863	Yes
			x_{min}=302				1.234		Yes

Comparing the τ_{sus} with the critical value τ_{crit} and analyzing the results for the Chauvenet criterion, two situations can occur:

a) If $\tau_{sus} < \tau_{crit}$, the values of the x_{sus} it is a noticeable value for the experimental data set and in table 2 it is noted with "Yes";

(b) If $\tau_{sus} > \tau_{crit}$, the values of the x_{sus} is affected by aberrant errors, in table 2 it is noted with "*No"* and that value is eliminated from the sample (x_{sus} is an outlier);

In this case, we recalculate the values of the average ($\bar{x}$) and standard deviation (s) for the remained values and we start again with verifying the (a) condition, the algorithm is applied until that condition is no longer verified for any of the two external values of the sample (x_{min} or x_{max}).

A general remark comes from studying the data presented in table 2: with "YES", note the values that are experimental results accepted by the criterion and with "NO" note the that that are rejected by the criterion:

- the value of x_{min} = 363, from the lot A with 8 number of the parallel determinations, τ_{sus} = 2.010, τ_{crit} = 1.863 so $\tau_{sus} > \tau_{crit}$. In this care x_{min} = 363 is an outlier, so we recalculate the values of the average ($\bar{x}$) and standard deviation (s) for the 7 remained values. Criterion Chauvenet was recalculated, and all values were found to be noticeable values for the experimental data.

- the value of x_{max} = 375, from the lot B with 8 number of the parallel determinations, τ_{sus} = 2.169, τ_{crit} = 1.863 so $\tau_{sus} > \tau_{crit}$. In this care x_{max} = 375 is an outlier, so we recalculate the values of the average ($\bar{x}$) and standard deviation (s) for the 7 remained values. Criterion Chauvenet was recalculated, and all values were found to be noticeable values for the experimental data.

CONCLUSIONS

(a) It is the structure that determines the values of the mechanical properties, specific to each thermal treatment parameter;

(b) Application of the Chauvenet criterion to detection and eliminating aberrant results obtained in the industrial processes, is carried out with the help of the statistical processing;

(c) After identifying and removing abnormal values, the Chauvenet criterion is again applied until abnormal values are no longer found;

(d) With Chauvenet's criterion you can be sure that your experimental research are free of aberrant data.

REFERENCES

[1] L. Lin and P.D. Sherman, *Cleaning Data the Chauvenet Way*, SESUS proceedings, Paper SA11, pp. 1-11, Available at http://analytics.ncsu.edu/sesug/2007/SA11.pdf, accessed on 24 April 2017;

[2] M. R. Costescu, *Identifying Data Affected by Aberrant Errors. Applied Program*, Economic Informatics, vol. 12, no. 1, pp. 52 - 22, 2008;

[3] M. Mărușteri, *Biostatistică, 2005,* Available at http://sorana.academicdirect.ro/pages/doc/Doc2012/_materiale/Biostatistica_curs.pdf, accessed on 24 April 2017;

[4] *** STATISTICS HOW TO, *Chauvenet's Criterion,* Practically Cheating Statistics Handbook, Available at http://www.statisticshowto.com/chauvenets-criterion/, accessed on 20 April 2017;

[5] N. M. Halfina, *Applications of the Chauvenet Test to Detection of Overliers in Observations Connected into a Homogeneus Markov Chain,* Journal of Mathematical Sciences, vol. 118, no. 6, pp. 5667–5672, 2003, Available at http://link.springer.com/article/10.1023/A:1026159026782, accessed on 20 April 2017;

[6] S. Sabade and D. M. H. Walker, *Evaluation of statistical outlier rejection methods for IDDQ limit setting*, Conference Paper, pp. 1-7, 2002, ResearchGate, Available at https://www.researchgate.net/profile/Sagar_Sabade/publication/3943191_Evaluation_of_statistical_outlier_rejection_methods_for_IDDQ_limit_setting/links/5408775f0cf23d9765b2dab8/Evaluation-of-statistical-outlier-rejection-methods-for-IDDQ-limit-setting.pdf, accessed on 20 April 2017;

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

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