

**PRELUCRAREA DATELOR
EXPERIMENTALE
Partea a 2-a**

**EXPERIMENTAL DATA PROCESSING
Part 2**

Wilhelm LAURENZI

Conf.dr.ing – TRANSILVANIA University in Braşov – Faculty of Wood Engineering

Adresa/Address: B-dul Eroilor nr. 29, 500036 Braşov, România

E-mail: willy@unitbv.ro

Rezumat:

Acest articol reprezintă partea doua a studiului privind prelucrarea datelor experimentale mono-factoriale, în care se prezintă programul original de prelucrare a datelor experimentale conceput de autor. Utilizând metodele şi relaţiile de calcul consacrate, programul permite stabilirea numărului de probe, generarea planului de experimentare, introducerea datelor măsurate şi salvarea acestora, identificarea datelor afectate de erori aberante, verificarea caracterului aleator, verificarea normalităţii repartiţiei de date, calculul principalilor parametri statistici şi exportul datelor experimentale în Excel sau în alte programe de prelucrare statistică a datelor.

Cuvinte cheie: cercetare experimentală, metode statistico-matematice, program prelucrare date experimentale.

INTRODUCERE

Datele experimentale obţinute în urma unor experimentări necesită prelucrarea acestora. În multe cazuri acestea sunt reprezentate numai grafic, interpretate vizual şi concluziile rezultate sunt bazate mai mult pe experienţa cercetătorului. Îmbunătăţirea calităţii prelucrării datelor experimentale se poate face prin utilizarea statisticii matematice care pune la dispoziţia cercetătorului metode şi teste obiective pentru interpretarea şi tragerea concluziilor corecte.

Există în prezent foarte multe soft-uri statistice de firmă sau realizate de diverşi autori, mai simple sau mai complexe, mai ieftine, mai scumpe sau câteodată gratuite. Cele gratuite nu conţin de obicei toate elementele dorite, foarte rar sunt Open Source pentru a putea fi modificate sau adaptate, iar cele de firmă conţin de obicei foarte multe facilităţi dar nu sunt dedicate unui anumit scop.

Ținând cont de aceste realități și pe baza informațiilor prezentate în prima parte a articolului (Laurenzi 2010), s-a conceput un program dedicat prelucrării statistice a datelor experimentale monofactoriale.

**PROGRAMUL DE PRELUCRARE STATISTICA A
DATELOR EXPERIMENTALE MONOFACTORIALE**

Programul de prelucrare statistica a datelor experimentale monofactoriale permite realizarea următoarelor operații importante (Laurenzi 2000,

Abstract:

This paper represents the second part of a study regarding the processing of experimental mono-factorial data, and it presents the original program developed by the author for processing experimental data. Using consecrated methods and relations, this program allows establishing the number of samples, generating the experimental plan, entering and saving the measured data, identifying the data corrupted by aberrant errors, verifying the randomness, verifying the normality of data repartition, calculating the main statistical parameters and exporting the experimental data to Excel or to other programs for statistical data processing.

Key words: experimental research, statistical-mathematical methods, program for processing experimental data.

INTRODUCTION

The experimental data obtained after experimentation must be processed. In most cases, they are only represented in diagrams and visually interpreted and the resulting conclusions are mainly based on the researcher's experience. The quality of processing experimental data can be improved by using mathematical statistics, which provide the researcher with objective methods and tests for interpreting data and deriving correct conclusions.

At present, there are several statistical software programs on the market, some developed by companies, some by programmers, some simple, some more complex, some low priced, some expensive and some even free. The free ones usually lack some elements and are rarely Open Source and modifiable, and the branded programs usually have a lot of features but are not specifically developed for a single purpose.

Considering these realities and based on the information provided within the first part of the article (Laurenzi 2010), a program for the statistical processing of mono-factorial experimental data was developed.

**PROGRAM FOR STATISTICAL PROCESSING OF
MONOFACTORIAL EXPERIMENTAL DATA**

The program for statistical processing of mono-factorial experimental data supports following important operations (Laurenzi 2000, www.labsmn.ro).

<http://www.labsmn.pub.ro/academic>);

- stabilirea numărului de probe;
- generarea planului de experimentare;
- salvarea planului de experimentare;
- introducerea datelor măsurate;
- salvarea datelor măsurate;
- deschiderea fișierelor cu date salvate;
- identificarea datelor afectate de erori aberante (testul Chauvenet);
- verificarea caracterului aleator (testul Young);
- verificarea normalității repartiției de date (testul Hi-pătrat, testul Massey și testul Shapiro-Wilk);
- calculul principalilor parametri statistici;
- exportul datelor experimentale în Excel sau în alte programe de prelucrare statistică a datelor.

În Fig. 1 este prezentat meniul principal al programului. Se poate observa că interfața grafică este organizată în mai multe ferestre de afișare, de introducere a informațiilor și de realizare a comenzilor. În partea de jos a interfeței grafice se afișează informații despre operațiile care se execută.

[pub.ro/academic](http://www.labsmn.pub.ro/academic));

- establishing the number of samples;
- generating the experimental schedule;
- saving the experimental schedule;
- entering the measured data;
- saving the measured data;
- opening the saved data file;
- identifying the data corrupted by aberrant errors (Chauvenet test);
- verifying the randomness (Young test);
- verifying the normality of data repartition (Hi-square test, Massey test and Shapiro-Wilk test)
- calculating the main statistical parameters;
- exporting the experimental data to Excel or to other programs for statistical data processing

Fig. 1 displays the main menu of the program. As can be seen, the graphical interface is organized into several windows for displaying information, entering data and carrying out commands. In the bottom part of the graphical interface, the user is given information about the current operations.

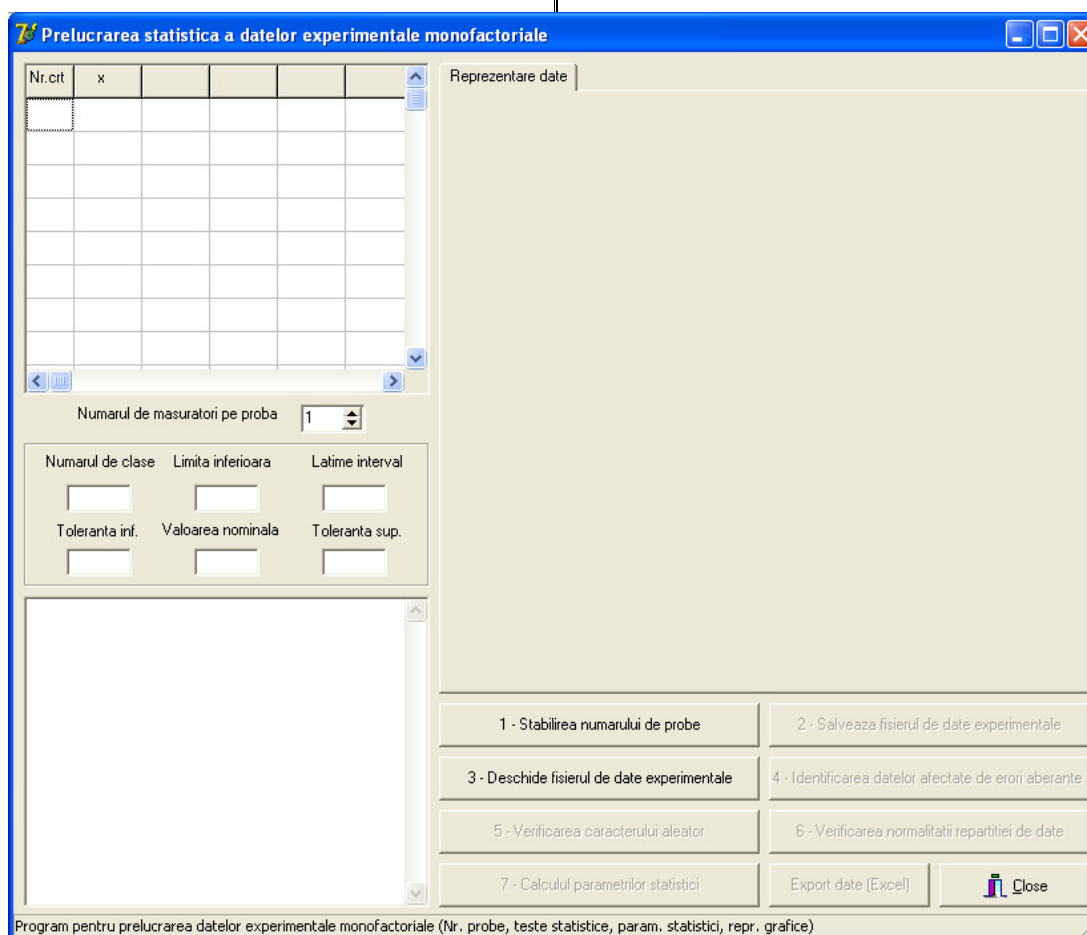


Fig. 1.
Meniul principal al programului / Main menu of the program.

Prezentarea programului

Programul este astfel conceput încât să fie foarte ușor de utilizat. Acesta ghidează utilizatorul astfel încât acesta să parcurgă fără greșeli toate testele. Pentru fiecare operație se afișează informații explicative în bara de stare din partea inferioară a interfeței grafice iar butoanele de comandă care lansează în execuție diferite activități sunt accesibile numai la momentul potrivit.

Pentru mai buna înțelegere, prezentarea programului se va face pe baza unui exemplu având ca scop prelucrarea statistică a datelor dintr-un studiu asupra preciziei grosimilor unui lot de 30 de plăci de PAL de 18 mm obținute prin presare.

După lansarea în execuție se pot face la alegere următoarele operații:

- se stabilește numărul de probe care se extrag din lot și se generează aleator planul experimental;
- se încarcă un fișier cu date salvate anterior.

Stabilirea numărului de probe și a planului experimental

Prin apăsarea butonului “Stabilirea numărului de probe” se deschide fereastra de dialog prezentată în Fig. 2.

Program Presentation

The program was conceived as a user-friendly application. It guides the user through all the necessary steps and thus ensures that the tests are completed without any mistakes. The status bar on the bottom on the graphical interface displays explanatory information for each step, and the control buttons for launching different operations are available only when needed.

For a better understanding, the program will be presented by means of a concrete example, namely the statistical processing of experimental data obtained in a study regarding the thickness variation for a batch of thirty 18-mm chipboards after pressing.

After initializing the program, the following operations can be carried out alternatively:

- establishing the number of samples extracted from the batch and randomly generating the experimental plan;
- loading a previously saved data file.

Establishing the Number of Samples and the Experimental Plan

A button is available for “Establishing the number of samples”. When pressing it, the dialogue box presented in Fig. 2 is opened.

Nr. crt.	Nr. proba	x1	x2	x3	x4	x5	x6	x7	x8
1	3								
2	4								
3	5								
4	6								
5	7								
6	8								
7	9								
8	10								

Fig. 2.

Stabilirea numărului de probe și a planului experimental / Establishing the sample size and the experimental plan.

Numărul de probe se stabilește automat sau se introduce de către utilizator în fereastra de editare “Numarul de probe”. Dacă se dorește stabilirea automată a numărului de probe este necesar să se introducă mărimea lotului și precizia în % cu care rezultatele măsurătorilor constituie o estimare corectă a întregului lot. După introducerea acestor informații se apasă butonul “Determinare număr probe”. În cazul unui lot de 30 de plăci și a unei precizii de 5% rezultă 28 de probe.

Înainte generării planului de experimentare se introduce numărul de măsurători pe probă (placă) care conform SR EN 324/1 este de 8 măsurători pe placă (Fig. 3)

The number of samples is established either automatically or by entering the number into the respective editing window. In order to determine the number automatically, the user must enter the size of the batch and the precision (in %) with which the results of the measurements represent a correct estimation of the entire batch. After entering this information, the user must press the button for “Determining the number of samples”. For a batch of 30 boards and a 5% precision, the result is of 28 samples.

Before generating the experimental plan, the number of measurements per sample (board) must be entered. According to SR EN 324/1 the number of measurements per chipboard is 8 (Fig. 3).

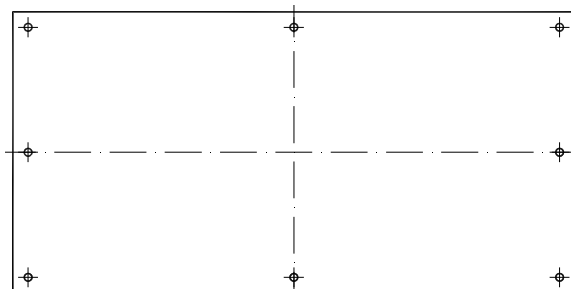


Fig. 3.
Punctele de măsurare / Measuring points.

Planul experimental se generează aleator prin apăsarea butonului "Generare". Acest plan, pentru maxim 8 măsurători (Fig. 2., dreapta), este afișat în fereastra din dreapta și poate fi salvat sub formă de fișier de tip text. Fișierul se tipărește și acest formular va fi completat cu valorile măsurate.

Desfășurarea experimentului

Înainte desfășurării experimentului se numerează toate probele din lot în ordine crescătoare, după care se extrag probele conform planului experimental (valorile aleatoare din coloana 2 – vezi Fig. 2). După extragerea probelor se realizează măsurătorile cu un micrometru pentru fiecare probă conform planului de măsurare. Valorile măsurate se înscriu pe formularul tipărit sau se introduc direct în program în fereastra din stânga sus a meniului principal (Fig. 4). De asemenea se completează liniile de editare "Numarul de masuratori pe proba", "Valoare nominala", "Toleranta inf." și "Toleranta sup.". Toate aceste informații se salvează cu ajutorul butonului "Salveaza fișierul de date experimentale".

The experimental plan is generated randomly when pressing the button "Generate". This schedule, for maximum 8 measurements (Fig. 2, right), is displayed in the right window and can be saved as a text file. The file is printed and the resulting form is filled out with the measured values.

Experiment Procedure

Before executing the experiment, all the batch samples are numbered in ascending order. Then, the samples are extracted according to the experimental plan (random values from column 2 – see Fig. 2). The extracted samples are all measured with a micrometer according to the measuring schedule. The measured values are registered in the printed form or are directly entered into the program through the upper left window of the main menu (Fig. 4). Also, the editing lines regarding the number of samples, nominal value, lower and upper tolerances must be filled out. All this informations are saved when pressing the "Save" button.

Nr. crt.	x1	x2	x3	x4	x5
1	17.8	18.8	18.6	17.9	18.3
2	18.4	19.0	18.3	18.6	18.1
3	17.5	18.3	17.8	17.5	17.2
4	18.9	18.9	18.9	18.8	17.6
5	17.8	17.5	17.7	17.1	17.5
6	18.5	18.5	17.7	17.3	18.5
7	17.6	18.4	18.2	17.1	17.9
8	18.4	17.1	19.0	17.4	17.8

x5	x6	x7	x8	V.med
18.3	18.4	18.6	18.4	
18.1	18.3	17.6	18.6	
17.2	17.7	18.3	17.2	
17.6	17.1	18.9	18.8	
17.5	17.3	18.8	17.7	
18.5	18.2	18.9	18.2	
17.9	17.2	17.8	17.4	
17.8	17.4	17.7	18.3	

Numarul de masuratori pe proba: 8

Numarul de clase: Limita inferioara: Latime interval:

Toleranta inf.: 0.4 Valoarea nominala: 18 Toleranta sup.: 0.4

Fig. 4.

Modul de introducere a datelor experimentale, valorii nominale si toleranțelor admisibile /
Entering mode of experimental data, normal value and admissible tolerances.

Încărcarea datelor experimentale salvate

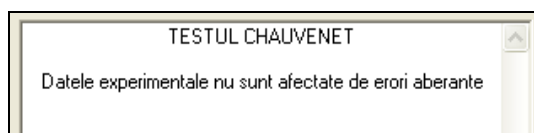
Datele experimentale salvate anterior pot fi vizualizate oricând în fereastra din stânga sus a meniului principal. Pentru aceasta se apasă butonul "Deschide fișierul de date experimentale" și se alege fișierul de date dorit din baza de date (subdirectorul "Date experimentale"). După încărcarea datelor experimentale se poate trece la realizarea testelor statistice.

Loading Previously Saved Experimental Data

Previously saved experimental data can be viewed anytime in the upper left window of the main menu. In order to do so, the user must press the button "Open file" and choose the appropriate data file from the database (subdirectory "Experimental data"). After loading the experimental data, the statistical tests can be performed.

Testul de identificare a datelor afectate de erori aberante

Erorile aberante afectează rezultatele prelucrărilor statistice și în consecință datele afectate de aceste erori trebuie identificate și eliminate. Aceasta operație se realizează prin apăsarea butonului "Identificarea datelor afectate de erori aberante". Acest buton lansează în execuție testul Chauvenet care ordonează șirul de date și în cazul în care una din valorile extreme este afectată de erori aberante le elimină și reia procesul până când nu mai există valori aberante. În cazul în care apar erori aberante afișează mesajul "Valoarea *** este afectată de erori aberante" altfel afișează mesajul "Datele experimentale nu sunt afectate de erori aberante". Pentru exemplul nostru rezultatul testului este că datele experimentale nu sunt afectate de erori aberante (Fig. 5, stânga).



Identification Test for Data Corrupted by Aberrant Errors

Aberrant errors affect the results of statistical processing. Therefore, the data corrupted by these errors must be identified and eliminated. This is done by pressing the button "Identify data corrupted by aberrant errors". This button launches the Chauvenet test, which arranges the data in the correct order and, if one of the extreme values is corrupted by aberrant errors, eliminates it and repeats the process until there are no more aberrant values left. If there are any aberrant errors present, the message "Value *** is corrupted by aberrant errors" is displayed; if not, then the message "The experimental data are not corrupted by aberrant errors" is shown on the screen. In our example, the result of the test is that the experimental data are not corrupted by aberrant errors (Fig. 5, left).

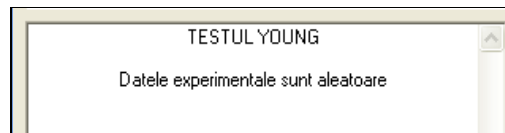


Fig. 5.

**Rezultatul testelor Chauvenet și Young /
Result of the Chauvenet and Young tests.**

Testul de verificare a caracterului aleator

În vederea obținerii unor rezultate corecte este necesar ca datele experimentale să fie aleatoare. Acest lucru poate fi verificat cu ajutorul testului Young. Programul realizează această operație prin apăsarea butonului "Verificarea caracterului aleator". După analizarea datelor experimentale se afișează mesajul "Datele experimentale sunt aleatoare" sau "Datele experimentale nu sunt aleatoare. Refaceți experimentul". Pentru exemplul nostru rezultă că datele experimentale sunt aleatoare (Fig. 5, dreapta).

Randomness Verification Test

In order to obtain correct results, the experimental data must be chosen randomly. The randomness can be verified through the Young test. The program executes this operation when pressing the button "Verify the randomness character". After the experimental data have been analyzed, the message "The experimental data are not random. Please repeat the experiment" is displayed. In our example, the result is that the experimental data are random (Fig. 5, right).

Testul de normalitate

Funcție de precizia de prelucrare, de reglarea mașinilor și de erorile de măsurare datele experimentale sunt împrăștiate mai mult sau mai puțin în jurul valorii medii. În mod normal datele experimentale trebuie să aibă o repartiție normală (Gauss) dar din diverse motive sunt repartizate mai mult sau mai puțin normal. Toate aceste lucruri pot fi studiate prin realizarea testelor de normalitate. Programul utilizează testul Massey pentru mai puțin de 32 de valori, testul χ^2 (Hi)-pătrat pentru mai mult de 32 de valori și testul Shapiro-Wilk pentru 7-500 de valori.

Normality Test

Depending on the precision of processing, the calibration of the machines and the measuring errors, the experimental data are more or less spread around the mean value. As a rule, the experimental data should be distributed normally (Gauss), but for one reason or another they can be distributed more or less normally. This can be studied by carrying out normality tests. The program uses the Massey test for less than 32 values, the test χ^2 (Hi)-square for more than 32 values and the Shapiro-Wilk test for 7-500 values.

Testele de normalitate sunt însoțite de afișarea valorilor și a graficelor frecvențelor absolute și relative grupate pe valori (Fig. 6) și pe clase (Fig. 7), a histogramei și a graficelor repartiției de date experimentale (curba neagră) și a celei normale (curba roșie) (Fig. 8). Vizualizarea acestor grafice se face prin apăsarea butoanelor corespunzătoare din

The normality tests are accompanied by displaying the values and the graphical representations of absolute and relative frequencies, grouped in values (Fig. 6) and classes (Fig. 7), the histogram and the graphical representations of the experimental data distribution (black curve) and the normal distribution (red curve) (Fig. 8). These graphs can be viewed when pressing the

fereastra din dreapta sus.

Datele pentru exemplul nostru sunt repartizate normal (testele Massay și Shapiro Wilk) și se încadrează în domeniul media $\pm 3\sigma$. Curba de repartiție este puțin asimetrică spre stânga și cu aproximativ 30% mai turtită decât repartiția normală. Acest lucru se datorează faptului că histograma nu este simetrică și nu are un singur vârf în centrul domeniului 6σ .

corresponding buttons in the upper right window.

The data in our example are distributed normally (Massey and Shapiro Wilk tests) and fall into the range of mean $\pm 3\sigma$. The distribution curve is a little asymmetrical to the left and approximately 30% more flat than the normal distribution. This is due to the fact that the histogram is not symmetrical and displays more than one peak in the center of the range 6σ .

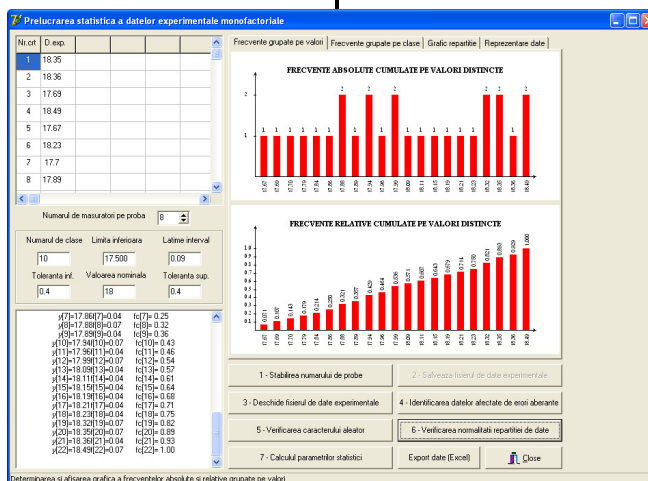


Fig. 6.

**Frecvențe absolute și relative grupate pe valori /
Absolute and relative frequencies grouped on values.**

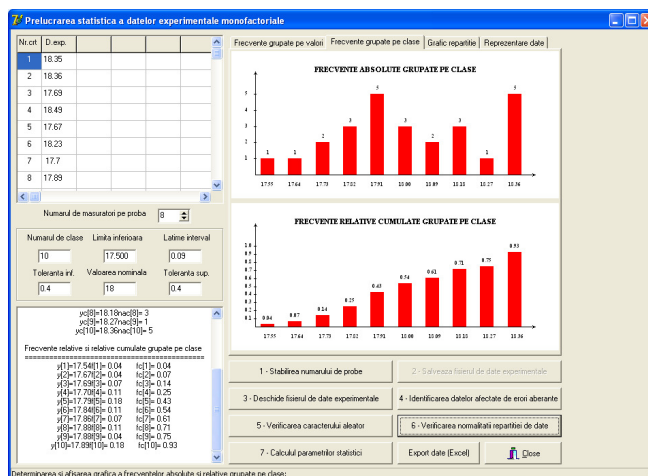


Fig. 7.

**Frecvențe absolute și relative grupate pe clase /
Absolute and relative frequencies grouped on classes.**

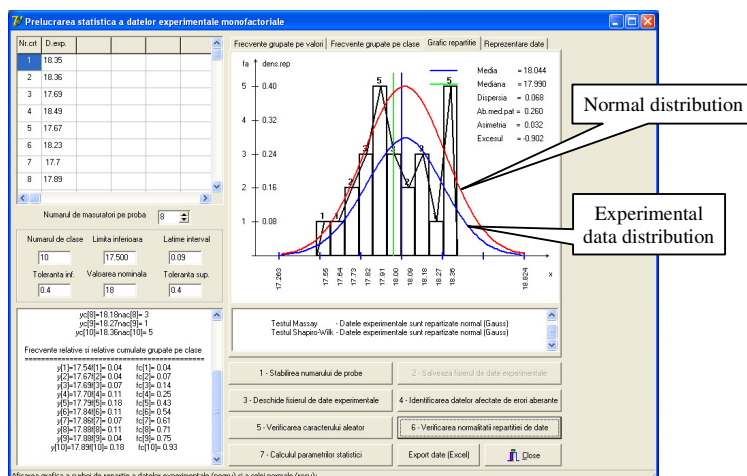


Fig. 8.

Curba de repartiție a datelor experimentale și a celei normale / Experimental data and normal distribution

Parametrii statistici

Programul calculează principalii parametri statistici pentru datele experimentale cum ar fi: media, mediana, dispersia, abaterea medie pătratică, coeficientul de variație, asimetria, excesul și numărul de valori cuprinse în domeniul media $\pm 3\sigma$ (Fig. 9). Aceste date, împreună cu graficele frecvențelor, a histogramei și a curbelor de repartiție și testele de normalitate, oferă informații despre normalitatea datelor experimentale. O repartiție de date trebuie să fie cât mai simetrică (asimetria=0), să nu fie turtită (excesul=0), mediana să fie egală cu media, abaterea medie pătratică σ să fie cât mai mică, datele experimentale să se încadreze în domeniul media $\pm 3\sigma$, histograma să fie simetrică și să aibă un singur vârf etc.

Statistical Parameters

The program calculates the main statistical parameters for experimental data, such as: mean value, median value, dispersion, mean square deviation, variation coefficient, asymmetry, excess and number of values in the range of mean value $\pm 3\sigma$ (Fig. 9). These data, along with the frequency graphs, the histogram, the distribution curves and the normality tests, provide information about the normality of the experimental data. The data repartition must be as symmetrical as possible (asymmetry = 0), must not be flat (excess = 0), the median value must be equal to the mean value, the mean square deviation σ must be as small as possible, the experimental data must fall within the range of mean value $\pm 3\sigma$, the histogram must be symmetrical and single-peaked etc.

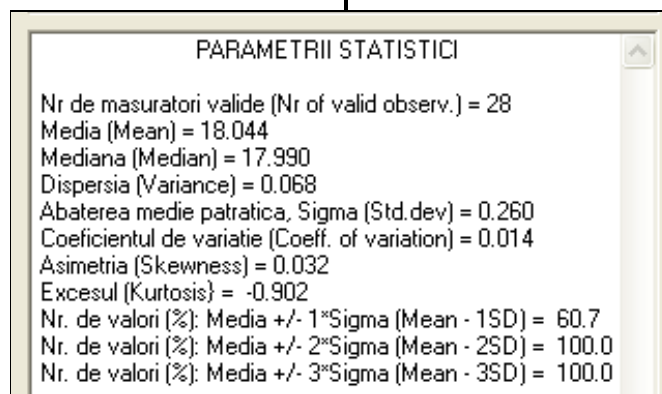


Fig. 9.

Afișarea principalilor parametri statistici / Displaying the most important statistic parameters.

Verificarea preciziei de măsurare, prelucrare și reglare

Grosimea plăcilor de PAL depinde în mare măsură de precizia de presare și de corectitudinea

Verifying the Measuring, Processing and Adjustment Precision

The thickness of the chipboards is largely dependent on the precision of the pressing and the

reglării preselor. Grosimile rezultate după presare trebuie să se încadreze într-un domeniu admisibil care pentru plăci de PAL de interior de 18 mm este conform STAS 6438-83 de $\pm 4\text{mm}$. Prin reprezentarea grafică a valorilor datelor experimentale se pot observa și analiza aceste aspecte și se pot trage concluziile necesare (Fig.10).

Dacă datele experimentale se încadrează în domeniul admisibil totul este în regulă. Dacă datele ies în ambele părți din domeniul admisibil înseamnă că a scăzut precizia de presare sau sunt erori de măsurare. Dacă datele experimentale nu depășesc domeniul $\pm$ toleranța dar depășesc limita inferioară sau superioară înseamnă că presa este dereglată. În cazul de față se poate observa că numai câteva valori depășesc domeniul admisibil ceea ce constituie semne de atenționare.

Exportul datelor experimentale

Datele experimentale pot fi exportate în Excel sau în alte programe de aplicație pentru completarea testelor statistice sau pentru verificarea calculului și testelor statistice utilizate în acest program (Fig. 11). În acest sens toate testele au fost verificate cu ajutorul programului Statistica [Statistica 2008].

right adjustment of the pressing machines. It must fall within the admissible range, which for 18 mm interior chipboards is of $\pm 4\text{mm}$. Through the graphical representation of the experimental data values, these aspects can be observed and analyzed and the necessary conclusions can be drawn (Fig.10).

If the experimental data fall within the admissible range, then everything is in order. If the data overstep both limits of the admissible range, it means that the precision of pressing has decreased or that the measurements show errors. If the experimental data fall within the range $\pm$ tolerance, but overstep the lower or upper limit, then the press is out of tune. In the present case it is noticeable that only a few values exceed the admissible range, which can be viewed as a warning sign.

Export of Experimental Data

Experimental data can be exported to Excel or to other applications in order to complete the statistical tests or to verify the calculations and the statistical tests used for this program (Fig. 11). In our example, all the tests have been verified with the program Statistica (Statistica 2008).

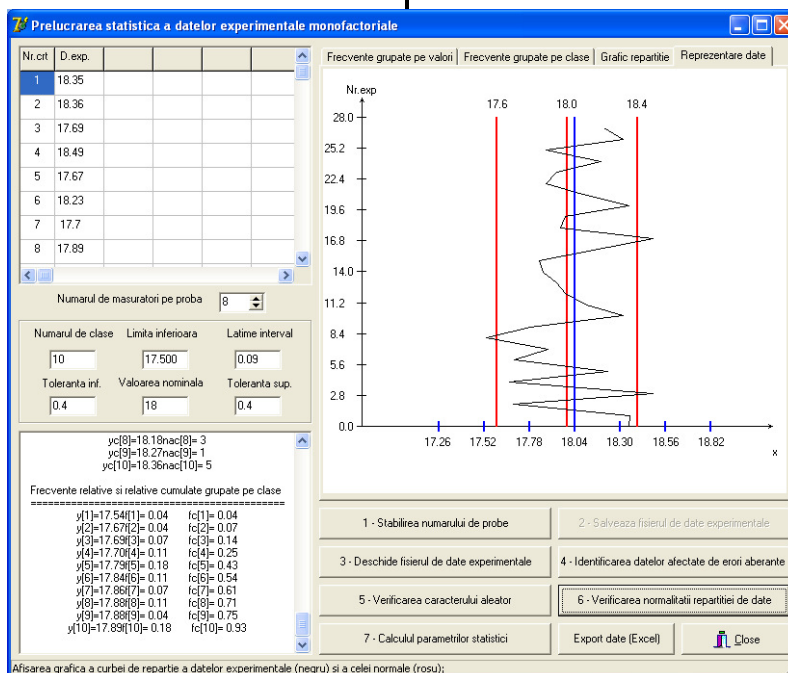


Fig. 10.
Reprezentarea datelor experimentale și limitele domeniului admisibil /
Representation of the experimental data and the limits of the admissible domain.

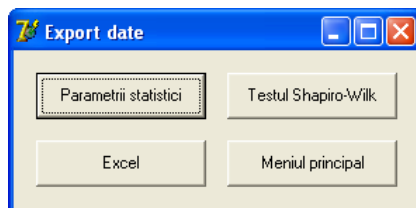


Fig. 11.
Exportul datelor experimentale / Export of experimental data.

Butonul "Excel" permite exportul datelor într-un worksheet Excel unde acestea pot fi utilizate pentru calcule statistice, reprezentări grafice etc. Butoanele "Parametrii statistici" și "Testul Shapiro-Wilk" permit accesarea unor aplicații Excel pentru calculul unor parametrii statistici și pentru realizarea testului de normalitate Shapiro-Wilk.

CONCLUZII

Programul de prelucrare statistică a datelor experimentale monofactoriale este util atât în activitatea didactică cât și în activitatea de cercetare științifică. Programul este ușor de utilizat și constituie un ajutor în analiza statistică a datelor experimentale cu un singur factor și tragerea concluziilor corecte. Programul prezentat constituie prima versiune și sigur va fi dezvoltat în viitor.

The Excel button exports the data to an Excel worksheet where they can be used for statistical calculations, diagrams etc. The buttons "Statistical Parametres" and "Shapiro-Wilk Test" open Excel applications for calculating statistical parameters and for carrying out the Shapiro-Wilk normality test.

CONCLUSION

The program for statistical processing of experimental single-factorial data is useful both for teaching purposes and for research activities. The program is a user-friendly and useful tool for the statistical analysis of single factorial experimental data and for drawing the correct conclusions. This first version presented above will surely be further developed in the future.

BIBLIOGRAFIE / REFERENCES

LAURENZI, W. (2000). Contribuții la modelarea și optimizarea așchierii lemnului cu pânze circulare în vederea conducerii procesului de tăiere cu ajutorul calculatorului (Contributions for modelling and optimization of wood cutting with circular saw blades in order to lead the process with a computer). PhD Thesis. TRANSILVANIA University of Brașov..

LAURENZI, W. (2010). Experimental Data Processing. Part 1. PRO LIGNO 6(4):55 - 63.

*** Statistica 8 (1984-2007). Statsoft Inc. <http://www.statsoft.com/>

*** <http://www.labsmn.pub.ro/academic>